



INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

REVSTAT

Statistical Journal



REVSTAT

Statistical Journal

Catálogo Recomendada

REVSTAT. Lisboa, 2003-
Revstat : statistical journal / ed. Instituto Nacional
de Estatística. - Vol. 1, 2003- . - Lisboa I.N.E.,
2003- . - 30 cm
Trimestral. - Continuação de : Revista de Estatística =
ISSN 0873-4275. - edição exclusivamente em inglês
ISSN 1645-6726

CREDITS

- EDITOR-IN-CHIEF

- *Isabel Fraga Alves*

- CO-EDITOR

- *Giovani L. Silva*

- ASSOCIATE EDITORS

- *Marília Antunes*
- *Barry ARNOLD (2019)*
- *Narayanaswamy Balakrishnan*
- *Jan Beirlant*
- *Graciela Boente (2019-2020)*
- *Paula Brito*
- *Vanda Inácio de Carvalho*
- *Arthur Charpentier*
- *Valérie Chavez-Demoulin*
- *David Conesa*
- *Charmaine Dean*
- *Jorge Milhazes Freitas*
- *Alan Gelfand*
- *Stéphane Girard*
- *Wenceslao Gonzalez-Manteiga*
- *Marie Kratz*
- *Victor Leiva*
- *Maria Nazaré Mendes-Lopes*
- *Fernando Moural*
- *John Nolan*
- *Paulo Eduardo Oliveira*
- *Pedro Oliveira*
- *Carlos Daniel Paulino (2019-2021)*
- *Arthur Pewsey*
- *Gilbert Saporta*
- *Alexandra M. Schmidt*
- *Julio Singer*

- *Manuel Scotto*

- *Lisete Sousa*

- *Milan Stehlik*

- *María Dolores Ugarte*

- EXECUTIVE EDITOR

- *José A. Pinto Martins*

- FORMER EXECUTIVE EDITOR

- *Maria José Carrilho*

- *Ferreira da Cunha*

- SECRETARY

- *Liliana Martins*

- PUBLISHER

- *Instituto Nacional de Estatística, I.P. (INE, I.P.)*
Av. António José de Almeida, 2
1000-043 LISBOA
PORTUGAL
Tel.: + 351 21 842 61 00
Fax: + 351 21 845 40 84
Web site: <http://www.ine.pt>
Customer Support Service
+ 351 218 440 695

- COVER DESIGN

- *Mário Bouçadas, designed on the stain glass window at INE by the painter Abel Manta*

- LAYOUT AND GRAPHIC DESIGN

- *Carlos Perpétuo*

- PRINTING

- *Instituto Nacional de Estatística, I.P.*

- EDITION

- *140 copies*

- LEGAL DEPOSIT REGISTRATION

- *N.º 191915/03*

- PRICE [VAT included]

- *€ 9,00*



INDEX

Classical and Bayesian Componentwise Predictors for Non-Compact Correlated ARH(1) Processes	
<i>M. Dolores Ruiz-Medina and Javier Álvarez-Liébana</i>	265
On weighted Kullback–Leibler Divergence for Doubly Truncated Random Variables	
<i>Rajesh Moharana and Suchandan Kayal</i>	297
The Beta Marshall–Olkin Lomax Distribution	
<i>Claudio J. Tablada and Gauss M. Cordeiro</i>	321
The CUSUM Median Chart for Known and Estimated Parameters	
<i>Philippe Castagliola, Fernanda Otilia Figueiredo and Petros E. Maravelakis</i>	345
AP-Optimum Designs for Minimizing the Average Variance and Probability-Based Optimality	
<i>N.M. Kilany and W.A. Hassanein</i>	371
An Information Theoretical Method for Analyzing Unreplicated Designs with Binary Response	
<i>Krystallenia Drosou and Christos Koukouvinos</i>	383
Prediction Intervals of the Record-Values Process	
<i>Amany E. Aly, H.M. Barakat and Magdy E. El-Adll</i>	401

Abstracted/indexed in: *Current Index to Statistics, DOAJ, Google Scholar, Journal Citation Reports/Science Edition, Mathematical Reviews, Science Citation Index Expanded[®], SCOPUS and Zentralblatt für Mathematic.*

CLASSICAL AND BAYESIAN COMPONENTWISE PREDICTORS FOR NON-COMPACT CORRELATED ARH(1) PROCESSES

Authors: M. DOLORES RUIZ-MEDINA
– Department of Statistics and Operation Research, University of Granada,
18071 Granada, Spain
mruiz@ugr.es

JAVIER ÁLVAREZ-LIÉBANA
– Department of Statistics and Operation Research, University of Granada,
18071 Granada, Spain
javialvaliebana@ugr.es

Received: May 2016

Revised: January 2017

Accepted: March 2017

Abstract:

- A special class of standard Gaussian Autoregressive Hilbertian processes of order one (Gaussian ARH(1) processes), with bounded linear autocorrelation operator, which does not satisfy the usual Hilbert–Schmidt assumption, is considered. To compensate the slow decay of the diagonal coefficients of the autocorrelation operator, a faster decay velocity of the eigenvalues of the trace autocovariance operator of the innovation process is assumed. As usual, the eigenvectors of the autocovariance operator of the ARH(1) process are considered for projection, since, here, they are assumed to be known. Diagonal componentwise classical and bayesian estimation of the autocorrelation operator is studied for prediction. The asymptotic efficiency and equivalence of both estimators is proved, as well as of their associated componentwise ARH(1) plug-in predictors. A simulation study is undertaken to illustrate the theoretical results derived.

Key-Words:

- *asymptotic efficiency; autoregressive Hilbertian processes; bayesian estimation; classical moment-based estimation; functional prediction; non-compact bounded autocorrelation operators.*

AMS Subject Classification:

- 60G10, 60G15, 60F99, 60J05, 65F15.

1. INTRODUCTION

Functional time series theory plays a key role in the analysis of high-dimensional data (see, for example, Aue, Norinho and Hörmann, 2014; Bosq, 2000; Bosq and Blanke, 2007). Inference for stochastic processes can also be addressed from this framework (see Álvarez-Liébana, Bosq and Ruiz-Medina, 2016, in relation to functional prediction of the Ornstein–Uhlenbeck process, in an ARH(1) process framework). Bosq (2000) addresses the problem of infinite-dimensional parameter estimation and prediction of ARH(1) processes, in the cases of known and unknown eigenvectors of the autocovariance operator. Alternative projection methodologies have been adopted, for example, in Antoniadis and Sapatinas (2003), in terms of wavelet bases, and, in Besse and Cardot (1996), in terms of spline bases. The book by Bosq and Blanke (2007) provides a general overview on statistical prediction, including Bayesian predictors, inference by projection and kernel methods, empirical density estimation, and linear processes in high-dimensional spaces (see also Blanke and Bosq, 2015, on Bayesian prediction for stochastic processes). Recently, Bosq and Ruiz-Medina (2014) have derived new results on asymptotic efficiency and equivalence of classical and Bayes predictors for l^2 -valued Poisson process, where, as usual, l^2 denotes the Hilbert space of square summable sequences. Classical and bayesian componentwise parameter estimators of the mean function and autocovariance operator, characterizing Gaussian measures in Hilbert spaces, are also compared in terms of their asymptotic efficiency, in that paper.

We first recall that the class of processes studied here could be of interest in applications, for instance, in the context of anomalous physical diffusion processes (see, for example, Meerschaert *et al.*, 2002; Gorenflo and Mainardi, 2003, and Metzler and Klafter, 2004, and the references therein). An interesting example of our framework corresponds to the case of spatial fractal diffusion operator, and regular innovations. Specifically, the class of standard Gaussian ARH(1) processes studied have a bounded linear autocorrelation operator, admitting a weak-sense diagonal spectral representation, in terms of the eigenvectors of the autocovariance operator. The sequence of diagonal coefficients, in such a spectral representation, displays an accumulation point at one. The singularity of the autocorrelation kernel is compensated by the regularity of the autocovariance kernel of the innovation process. Namely, the key assumption here is the summability of the quotient between the eigenvalues of the autocovariance operator of the innovation process and of the ARH(1) process. Under suitable conditions, the asymptotic efficiency and equivalence of the studied diagonal componentwise classical and bayesian estimators of the autocorrelation operator are derived (see Theorem 4.1 below). Under the same setting of conditions the asymptotic efficiency and equivalence of the corresponding classical and bayesian ARH(1) plug-in predictors are proved as well (see Theorem 4.2 below). Although both

theorems only refer to the case of known eigenvectors of the autocovariance operator, as illustrated in the simulation study undertaken in Álvarez-Liébana, Bosq and Ruiz-Medina (2017) (see also Ruiz-Medina and Álvarez-Liébana, 2017), a similar performance is obtained for the case of unknown eigenvectors, in comparison with other componentwise, kernel-based, wavelet-based penalized and nonparametric approaches adopted in the current literature (see Antoniadis and Sapatinas, 2003; Besse and Cardot, 1996; Bosq, 2000; Guillas, 2001; Mas, 1999).

Note that, for θ being the unknown parameter, in order to compute $E(\theta|X_1, \dots, X_n)$, with X_1, \dots, X_n denoting the functional sample, we suppose that $\theta_j \perp (X_{i,j'}, i \geq 1, j' \neq j)$, which leads to

$$\langle E(\theta|X_1, \dots, X_n), v_j \rangle_H = E(\theta_j|X_1, \dots, X_n) = E(\theta_j|X_{1,j}, \dots, X_{n,j}).$$

Here, for each $j \geq 1$, $\theta_j = \langle \theta, v_j \rangle_H$, and $X_{i,j} = \langle X_i, v_j \rangle_H$, $i = 1, \dots, n$, with $\langle \cdot, \cdot \rangle_H$ being the inner product in the real separable Hilbert space H . Note that $\{v_j, j \geq 1\}$ denotes an orthonormal basis of H , diagonalizing the common autocovariance operator of X_1, \dots, X_n . We can then perform an independent computation of the respective posterior distributions of the projections θ_j , $j \geq 1$, of parameter θ , with respect to the orthonormal basis $\{v_j, j \geq 1\}$ of H .

Finally, some numerical examples are considered to illustrate the results derived on asymptotic efficiency and equivalence of moment-based classical and beta-prior-based Bayes diagonal componentwise parameter estimators, and the associated ARH(1) plug-in predictors.

2. PRELIMINARIES

The preliminary definitions and results needed in the subsequent development are introduced in this section. We first refer to the usual class of standard ARH(1) processes introduced in Bosq (2000).

Definition 2.1. Let H be a real separable Hilbert space. A sequence $Y = (Y_n, n \in \mathbb{Z})$ of H -valued random variables on a basic probability space (Ω, \mathcal{A}, P) is called an autoregressive Hilbertian process of order one, associated with (μ, ε, ρ) , if it is stationary and satisfies

$$(2.1) \quad X_n = Y_n - \mu = \rho(Y_{n-1} - \mu) + \varepsilon_n = \rho(X_{n-1}) + \varepsilon_n, \quad n \in \mathbb{Z},$$

where $\varepsilon = (\varepsilon_n, n \in \mathbb{Z})$ is a Hilbert-valued white noise in the strong sense (i.e., a zero-mean stationary sequence of independent H -valued random variables with $E\|\varepsilon_n\|_H^2 = \sigma^2 < \infty$, for every $n \in \mathbb{Z}$), and $\rho \in \mathcal{L}(H)$, with $\mathcal{L}(H)$ being the space of linear bounded operators on H . For each $n \in \mathbb{Z}$, ε_n and X_{n-1} are assumed to be uncorrelated.

If there exists a positive $j_0 \geq 1$ such that $\|\rho^{j_0}\|_{\mathcal{L}(H)} < 1$, then, the ARH(1) process X in (2.1) is standard, and there exists a unique stationary solution to equation (2.1) admitting a MAH(∞) representation (see Theorem 3.1 in Bosq, 2000, p. 74).

The autocovariance and cross-covariance operators are given by

$$(2.2) \quad \begin{aligned} C &= E[X_n \otimes X_n] = E[X_0 \otimes X_0], \quad n \in \mathbb{Z}, \\ D &= E[X_n \otimes X_{n+1}] = E[X_0 \otimes X_1], \quad n \in \mathbb{Z}, \end{aligned}$$

where, for $f, g \in H$,

$$f \otimes g(h) = f \langle g, h \rangle_H, \quad \forall h \in H,$$

defines a Hilbert–Schmidt operator on H . Operator C is assumed to be in the trace class. In particular, $E\|X_n\|_H^2 < \infty$, for all $n \in \mathbb{Z}$. It is well-known that, from equations (2.1) and (2.2), for all $h \in H$, $D(h) = \rho C(h)$ (see, for example, Bosq, 2000). However, since C is a nuclear or trace operator, its inverse operator is an unbounded operator in H . Different methodologies have been adopted to overcome this problem in the current literature on ARH(1) processes. In particular, here, we consider the case where $C(H) = H$, under **Assumption A2** below, since C is assumed to be strictly positive. That is, its eigenvalues are strictly positive and the kernel space of C is trivial. In addition, they are assumed to have multiplicity one. Therefore, for any $f, g \in H$, there exist $\varphi, \phi \in H$ such that $f = C(\varphi)$ and $g = C(\phi)$, and

$$\langle C^{-1}(f), C^{-1}(g) \rangle_H = \langle C^{-1}(C(\varphi)), C^{-1}(C(\phi)) \rangle_H = \langle \varphi, \phi \rangle_H.$$

In particular, $\|C^{-1}(f)\|_H^2 < \infty$, for every $f \in H$.

Assumption A1. The operator ρ in (2.1) is self-adjoint with $\|\rho\|_{\mathcal{L}(H)} < 1$.

Assumption A2. Operator C is strictly positive, and its positive eigenvalues have multiplicity one. Furthermore, C and ρ admit the following diagonal spectral decompositions: For all $f, g \in H$,

$$(2.3) \quad C(g)(f) = \sum_{k=1}^{\infty} \lambda_k(C) \langle \phi_k, g \rangle_H \langle \phi_k, f \rangle_H$$

$$(2.4) \quad \rho(g)(f) = \sum_{k=1}^{\infty} \rho_k \langle \phi_k, g \rangle_H \langle \phi_k, f \rangle_H,$$

where $\{\lambda_k(C), k \geq 1\}$ and $\{\rho_k, k \geq 1\}$ are the respective systems of eigenvalues of C and ρ , and $\{\phi_k, k \geq 1\}$ is the common system of orthonormal eigenvectors of the autocovariance operator C .

Remark 2.1. As commented before, we consider here the case where the eigenvectors $\{\phi_k\}_{k \geq 1}$ of the autocovariance operator C are known. Thus, under **Assumption A2**, the natural way to formulate a componentwise estimator of the autocorrelation operator ρ is in terms of the respective estimators of its diagonal coefficients $\{\rho_k\}_{k \geq 1}$, computed from the respective projections of the observed functional data, X_0, \dots, X_T , into $\{\phi_k, k \geq 1\}$. We adopt here a moment-based classical and Beta-prior-based bayesian approach in the estimation of such coefficients $\{\rho_k\}_{k \geq 1}$.

From Cauchy–Schwarz inequality, applying Parseval identity,

$$\begin{aligned} |\rho(g)(f)|^2 &\leq \sum_{k=1}^{\infty} |\rho_k| [\langle \phi_k, g \rangle_H]^2 \sum_{k=1}^{\infty} |\rho_k| [\langle \phi_k, f \rangle_H]^2 \\ &\leq \sum_{k=1}^{\infty} [\langle \phi_k, g \rangle_H]^2 \sum_{k=1}^{\infty} [\langle \phi_k, f \rangle_H]^2 = \|g\|_H^2 \|f\|_H^2 < \infty. \end{aligned}$$

Thus, equation (2.4) holds in the weak sense.

From **Assumption A2**, the projection of X_n into the common eigenvector system $\{\phi_k, k \geq 1\}$ leads to the following series expansion in $\mathcal{L}_H^2(\Omega, \mathcal{A}, P)$:

$$(2.5) \quad X_n = \sum_{k=1}^{\infty} \sqrt{\lambda_k(C)} \eta_k(n) \phi_k, \quad \eta_k(n) = \frac{1}{\sqrt{\lambda_k(C)}} \langle X_n, \phi_k \rangle_H,$$

and, for each $j, p \geq 1$, and $n > 0$,

$$\begin{aligned} E[\eta_j(n) \eta_p(n)] &= E \left[\frac{1}{\sqrt{\lambda_j(C)}} \langle X_n, \phi_j \rangle_H \frac{1}{\sqrt{\lambda_p(C)}} \langle X_n, \phi_p \rangle_H \right] \\ (2.6) \quad &= \frac{1}{\sqrt{\lambda_j(C)}} \frac{1}{\sqrt{\lambda_p(C)}} C(\phi_j)(\phi_p) \\ &= \frac{1}{\sqrt{\lambda_j(C)}} \frac{1}{\sqrt{\lambda_p(C)}} \lambda_j(C) \langle \phi_j, \phi_p \rangle_H = \delta_{j,p}, \end{aligned}$$

where $\delta_{(\cdot, \cdot)}$ denotes the Kronecker delta function, and the last equality is obtained from the orthonormality of the eigenvectors $\{\phi_k, k \geq 1\}$. Hence, under **Assumptions A1** and **A2**, the projection of equation (2.1) into the elements of the common eigenvector system $\{\phi_k, k \geq 1\}$ leads to the following infinite-dimensional system of equations:

$$(2.7) \quad \sqrt{\lambda_k(C)} \eta_k(n) = \rho_k \sqrt{\lambda_k(C)} \eta_k(n-1) + \varepsilon_k(n), \quad k \geq 1,$$

or equivalently,

$$(2.8) \quad \eta_k(n) = \rho_k \eta_k(n-1) + \frac{\varepsilon_k(n)}{\sqrt{\lambda_k(C)}}, \quad k \geq 1,$$

where $\varepsilon_k(n) = \langle \varepsilon_n, \phi_k \rangle_H$, for $k \geq 1$, and $n \in \mathbb{Z}$. Thus, for each $j \geq 1$, $\{a_j(n) = \sqrt{\lambda_j(C)}\eta_j(n), n \in \mathbb{Z}\}$ defines a standard AR(1) process. Its Moving Average representation of infinite order is given by

$$(2.9) \quad a_j(n) = \sum_{k=0}^{\infty} [\rho_j]^k \varepsilon_j(n-k), \quad n \in \mathbb{Z}.$$

Specifically, under **Assumption A2**,

$$(2.10) \quad \begin{aligned} E[a_j(n)a_p(n)] &= \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} [\rho_j]^k [\rho_p]^l E[\varepsilon_j(n-k)\varepsilon_p(n-l)] \\ &= \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} [\rho_j]^k [\rho_p]^l \delta_{k,l} \delta_{j,p} = 0, \quad j \neq p, \\ E[a_j(n)a_p(n)] &= \sum_{k=0}^{\infty} \sigma_j^2 [\rho_j]^{2k}, \quad j = p, \end{aligned}$$

where $\sigma_j^2 = E[\varepsilon_j(n-k)]^2 = E[\varepsilon_j(0)]^2$.

From equation (2.10), under **Assumptions A1–A2**,

$$(2.11) \quad \begin{aligned} E\|X(n)\|_H^2 &= \sum_{j=1}^{\infty} E[a_j(n)]^2 = \sum_{j=1}^{\infty} \sigma_j^2 \sum_{k=0}^{\infty} [\rho_j]^{2k} \\ &= \sum_{j=1}^{\infty} \sigma_j^2 \left[\frac{1}{1 - [\rho_j]^2} \right] = \sum_{j=1}^{\infty} \lambda_j(C) < \infty, \end{aligned}$$

with, as before,

$$\sum_{j=1}^{\infty} \sigma_j^2 = E\|\varepsilon(n)\|_H^2 < \infty.$$

Equation (2.11) leads to the identity

$$(2.12) \quad \lambda_j(C) = \left[\frac{\sigma_j^2}{1 - \rho_j^2} \right], \quad j \geq 1,$$

from which, we obtain

$$(2.13) \quad \rho_k = \sqrt{1 - \frac{\sigma_k^2}{\lambda_k(C)}}, \quad \sigma_k^2 = E[\langle \phi_k, \varepsilon_t \rangle_H]^2, \quad \forall t \in \mathbb{Z}, k \geq 1.$$

Under (2.12), equation (2.8) can also be rewritten as

$$(2.14) \quad \eta_k(n) = \rho_k \eta_k(n-1) + \sqrt{1 - \rho_k^2} \frac{\varepsilon_k(n)}{\sigma_k}, \quad k \geq 1.$$

Assumption A2B. The sequences $\{\sigma_k^2\}_{k \geq 1}$ and $\{\lambda_k(C)\}_{k \geq 1}$ satisfy

$$(2.15) \quad \begin{aligned} \frac{\sigma_k^2}{\lambda_k(C)} &\leq 1, \quad k \geq 1, \quad \lim_{k \rightarrow \infty} \frac{\sigma_k^2}{\lambda_k(C)} = 0. \\ \frac{\sigma_k^2}{\lambda_k(C)} &= \mathcal{O}(k^{-1-\gamma}), \quad \gamma > 0, \quad k \rightarrow \infty. \end{aligned}$$

Equation (2.15) means that $\{\sigma_k^2\}_{k \geq 1}$ and $\{\lambda_k(C)\}_{k \geq 1}$ are both summable sequences, with faster decay to zero of the sequence $\{\sigma_k^2\}_{k \geq 1}$ than the sequence $\{\lambda_k(C)\}_{k \geq 1}$, leading, from equations (2.12) and (2.13), to the definition of $\{\rho_k^2\}_{k \geq 1}$ as a sequence with accumulation point at one.

Remark 2.2. Under **Assumption A2B**, **Assumption A3** below holds.

For each $k \geq 1$, from equations (2.7)–(2.9),

$$(2.16) \quad \begin{aligned} \sum_{n=1}^T [\eta_k(n-1)]^2 &= \frac{1}{\lambda_k(C)} \left[\sum_{n=1}^T [\varepsilon_k(n-1)]^2 \right. \\ &\quad \left. + \sum_{n=1}^T \sum_{l=1}^{\infty} \sum_{p=1}^{\infty} [\rho_k]^l [\rho_k]^p \varepsilon_k(n-1-l) \varepsilon_k(n-1-p) \right] \\ &= \frac{1}{\lambda_k(C)} \left[\sum_{n=1}^T [\varepsilon_k(n-1)]^2 + S(T, k) \right], \end{aligned}$$

where $S(T, k) = \sum_{n=1}^T \sum_{l=1}^{\infty} \sum_{p=1}^{\infty} [\rho_k]^l [\rho_k]^p \varepsilon_k(n-1-l) \varepsilon_k(n-1-p)$. Hence, $\sum_{n=1}^T [\varepsilon_k(n-1)]^2 + S(T, k) \geq 0$, for every $T \geq 1$, and $k \geq 1$.

Assumption A3. There exists a sequence of real-valued independent random variables $\{\widetilde{M}(k)\}_{k \geq 1}$ such that

$$(2.17) \quad \begin{aligned} &\inf_{T \geq 1} \sqrt{\left| \frac{S(T, k)}{T \left(\sum_{n=1}^{T-1} [\varepsilon_k(n)]^2 + [\varepsilon_k(0)]^2 \right)} \right|} \\ &= \inf_{T \geq 1} \sqrt{\left| \frac{\sum_{n=1}^T \sum_{l=1}^{\infty} \sum_{p=1}^{\infty} [\rho_k]^l [\rho_k]^p \varepsilon_k(n-1-l) \varepsilon_k(n-1-p)}{T \left(\sum_{n=1}^{T-1} [\varepsilon_k(n)]^2 + [\varepsilon_k(0)]^2 \right)} \right|} \\ &\geq [\widetilde{M}(k)]^{-1} \quad \text{a.s.,} \quad \text{with} \quad \sum_{k=1}^{\infty} E[\widetilde{M}(k)]^l < \infty, \quad 1 \leq l \leq 4. \end{aligned}$$

Remark 2.3. Note that the mean value of

$$\sum_{n=1}^T \sum_{l=1}^{\infty} \sum_{p=1}^{\infty} [\rho_k]^l [\rho_k]^p \varepsilon_k(n-1-l) \varepsilon_k(n-1-p)$$

is of order $\frac{T\sigma_k^2}{1-(\rho_k)^2}$, and the mean value of

$$T \left(\sum_{n=1}^{T-1} [\varepsilon_k(n)]^2 + [\varepsilon_k(0)]^2 \right)$$

is of order $T(T-1)\sigma_k^2$. Hence, for the almost surely boundedness of the inverse of $\left| \frac{S(T,k)}{T(\sum_{n=1}^{T-1} [\varepsilon_k(n)]^2 + [\varepsilon_k(0)]^2)} \right|$, by a suitable sequence of random variables with summable l -moments, for $l = 1, 2, 3, 4$, the eigenvalues of operator ρ must be close to one but strictly less than one. As commented in Remark 2.2, from **Assumption A2B**, this condition is satisfied in view of equation (2.13).

Assumption A4. $E[\eta_j(m)\eta_k(n)] = \delta_{j,k}$, with, as before, $\delta_{j,k}$ denoting the Kronecker delta function, for every $m, n \in \mathbb{Z}$, and $j, k \geq 1$.

Remark 2.4. **Assumption A4** implies that the cross-covariance operator D admits a diagonal spectral decomposition in terms of the system of eigenvectors $\{\phi_k\}_{k \geq 1}$. Thus, under **Assumption A4**, the diagonal spectral decompositions (2.3) and (2.4) also hold.

The classical diagonal componentwise estimator $\hat{\rho}_T$ of ρ considered here is given by

$$\begin{aligned} \hat{\rho}_T &= \sum_{k=1}^{\infty} \hat{\rho}_{k,T} [\phi_k \otimes \phi_k] \\ \hat{\rho}_{k,T} &= \frac{\sum_{n=1}^T a_k(n-1)a_k(n)}{\sum_{n=1}^T [a_k(n-1)]^2} \\ (2.18) \quad &= \frac{\sum_{n=1}^T \langle X_{n-1}, \phi_k \rangle_H \langle X_n, \phi_k \rangle_H}{\sum_{n=1}^T [\langle X_{n-1}, \phi_k \rangle_H]^2} \\ &= \frac{\sum_{n=1}^T X_{n-1,k} X_{n,k}}{\sum_{n=1}^T X_{n-1,k}^2}, \quad k \geq 1. \end{aligned}$$

From equations (2.7)–(2.8) and (2.12), for each $k \geq 1$,

$$\begin{aligned} \hat{\rho}_{k,T} - \rho_k &= \frac{\sum_{n=1}^T X_{n-1,k} X_{n,k}}{\sum_{n=1}^T [X_{n-1,k}]^2} - \rho_k \\ &= \frac{\sum_{n=1}^T \rho_k [\eta_k(n-1)]^2 + (\eta_k(n-1)\varepsilon_k(n))/\sqrt{\lambda_k(C)}}{\sum_{n=1}^T [\eta_k(n-1)]^2} - \rho_k \\ &= \rho_k + \frac{\sum_{n=1}^T \eta_k(n-1)\varepsilon_k(n)}{\sqrt{\lambda_k(C)} \sum_{n=1}^T [\eta_k(n-1)]^2} - \rho_k \\ (2.19) \quad &= \frac{\sum_{n=1}^T \eta_k(n-1)\varepsilon_k(n)}{\sqrt{\sigma_k^2/(1-\rho_k^2)} \sum_{n=1}^T [\eta_k(n-1)]^2} \end{aligned}$$

$$(2.20) \quad = \sqrt{1-\rho_k^2} \frac{\sum_{n=1}^T \eta_k(n-1)[\varepsilon_k(n)/\sigma_k]}{\sum_{n=1}^T [\eta_k(n-1)]^2}.$$

Remark 2.5. It is important to note that, for instance, unconditional bases, like wavelets, provide the spectral diagonalization of an extensive family of operators, including pseudodifferential operators, and in particular, Calderón–Zygmund operators (see Kyriazis and Petrushev, 2001; Meyer and Coifman, 1997). Therefore, the diagonal spectral representations (2.3) and (2.4), in **Assumption A2**, hold for a wide class of autocovariance and cross-covariance operators, for example, in terms of wavelets. When the autocovariance and the cross-covariance operators are related by a continuous function, the diagonal spectral representations (2.3) and (2.4) are also satisfied (see Dautray and Lions, 1985, pp. 119, 126 and 140). **Assumption A2** has been considered, for example, in Theorem 8.5, on pp. 215–216, and in Theorem 8.7, on p. 221, in Bosq (2000), to establish strong consistency, although, in this book, a different setting of conditions is assumed. Thus, **Assumption A1** and **A2** already have been used (e.g., in Bosq, 2000; Álvarez-Liébaná, Bosq, Ruiz-Medina, 2017, and Ruiz-Medina and Álvarez-Liébaná, 2017), and **Assumptions A2B**, **A3** and **A4** appear in Ruiz-Medina, Romano and Fernández-Pascual (2016). **Assumptions A2B** is needed since the usual assumption on the Hilbert–Schmidt property of ρ , made by several authors, is not considered here. At the same type, as commented before, **Assumptions A2B** implies **Assumption A3**.

The following lemmas will be used in the derivation of the main results of this paper, Theorems 4.1 and 4.2, obtained in the Gaussian ARH(1) context.

Lemma 2.1. Let \mathcal{X}_i , $i = 1, \dots, n$, be the values of a standard zero-mean autoregressive process of order one (AR(1) process) at times $i = 1, 2, \dots, n$, and $\hat{\rho}_n = \frac{\sum_{i=1}^n \mathcal{X}_{i-1} \mathcal{X}_i}{\sum_{i=1}^n \mathcal{X}_{i-1}^2}$, with \mathcal{X}_0 representing the random initial condition. Assume that $|\rho| < 1$, and that the innovation process is white noise. Then, as $n \rightarrow \infty$,

$$(2.21) \quad \sqrt{n} \frac{\hat{\rho}_n - \rho}{\sqrt{1 - \rho^2}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

The proof of Lemma 2.1 can be found in Hamilton (1994, p. 216).

Lemma 2.2. Let \mathcal{X}_1 and \mathcal{X}_2 be two normal distributed random variables having correlation $\rho_{\mathcal{X}_1 \mathcal{X}_2}$, and with means μ_1 and μ_2 , and variances σ_1^2 and σ_2^2 , respectively. Then, the following identities hold:

$$(2.22) \quad \begin{aligned} E[\mathcal{X}_1 \mathcal{X}_2] &= \mu_1 \mu_2 + \rho_{\mathcal{X}_1 \mathcal{X}_2} \sigma_1 \sigma_2 \\ \text{Var}(\mathcal{X}_1 \mathcal{X}_2) &= \mu_1^2 \sigma_2^2 + \mu_2^2 \sigma_1^2 + \sigma_1^2 \sigma_2^2 + 2\rho_{\mathcal{X}_1 \mathcal{X}_2} \mu_1 \mu_2 \sigma_1 \sigma_2 + \rho_{\mathcal{X}_1 \mathcal{X}_2}^2 \sigma_1^2 \sigma_2^2 \end{aligned}$$

(see, for example, Aroian, 1947; Ware and Lad, 2003).

Lemma 2.3. For each $k \geq 1$, the following limit is obtained:

$$(2.23) \quad \lim_{T \rightarrow \infty} TE[\widehat{\rho}_{k,T} - \rho_k]^2 = 1 - \rho_k^2, \quad k \geq 1$$

(see, for example, Bartlett, 1946).

3. BAYESIAN DIAGONAL COMPONENTWISE ESTIMATION

Let us now denote by R the functional random variable on the basic probability space (Ω, \mathcal{A}, P) , characterized by the prior distribution for ρ . In our case, we assume that R is of the form

$$(3.1) \quad R(f)(g) = \sum_{k=1}^{\infty} R_k \langle \phi_k, f \rangle_H \langle \phi_k, g \rangle_H, \quad \forall f, g \in H, \quad \text{a.s.}$$

where, for $k \geq 1$, R_k is a real-valued random variable such that $R(\phi_j)(\phi_k) = \delta_{j,k} R_k$, almost surely, for every $j \geq 1$. In the following, R_k is assumed to follow a beta distribution with shape parameters $a_k > 0$ and $b_k > 0$, i.e., $R_k \sim B(a_k, b_k)$, for every $k \geq 1$. We also assume that R is independent of the functional components of the innovation process $(\varepsilon_n, n \in \mathbb{Z})$, and that the random variables $R_k, k \geq 1$, are globally independent. That is, for each $f, g \in H$,

$$(3.2) \quad \begin{aligned} \varphi_R^{f,g}(t) &= E \left[\exp \left(it \sum_{k=1}^{\infty} R_k \langle \phi_k, f \rangle_H \langle \phi_k, g \rangle_H \right) \right] \\ &= \prod_{k=1}^{\infty} E [\exp (it R_k \langle \phi_k, f \rangle_H \langle \phi_k, g \rangle_H)] = \prod_{k=1}^{\infty} \varphi_{R_k} (t \langle \phi_k, f \rangle_H \langle \phi_k, g \rangle_H). \end{aligned}$$

Thus,

$$\varphi_R(t) = \prod_{k=1}^{\infty} \varphi_{R_k} (t (\phi_k \otimes \phi_k)),$$

where the last identity is understood in the weak-sense, i.e., in the sense of equation (3.2). In the definition of R from $\{R_j, j \geq 1\}$, we can then apply Kolmogorov extension theorem under the condition

$$\sum_{j=1}^{\infty} \frac{a_j b_j}{(a_j + b_j + 1)(a_j + b_j)^2} < \infty$$

(see, for example, Khoshnevisan, 2007).

As in the real-valued case (see Appendix A), considering $b_j > 1$, for each $j \geq 1$, the Bayes estimator of ρ is defined by (see Case 2 in Appendix A)

$$(3.3) \quad \tilde{\rho}_n = \sum_{j=1}^{\infty} \tilde{\rho}_{j,n} \phi_j \otimes \phi_j,$$

with, for every $j \geq 1$,

$$\begin{aligned}
 \tilde{\rho}_{j,n} &= \frac{1}{2\beta_{j,n}} \left[(\alpha_{j,n} + \beta_{j,n}) \pm \sqrt{(\alpha_{j,n} - \beta_{j,n})^2 - 4\beta_{j,n}\sigma_j^2[2 - (a_j + b_j)]} \right] \\
 (3.4) \quad &= \frac{\left[\sum_{i=1}^n x_{i-1,j}x_{i,j} + x_{i-1,j}^2 \right]}{2 \sum_{i=1}^n x_{i-1,j}^2} \\
 &\pm \frac{\sqrt{\left[\sum_{i=1}^n x_{i-1,j}x_{i,j} - x_{i-1,j}^2 \right]^2 - 4\sigma_j^2 \left[\sum_{i=1}^n x_{i-1,j}^2 \right] [2 - (a_j + b_j)]}}{2 \sum_{i=1}^n x_{i-1,j}^2},
 \end{aligned}$$

where

$$(3.5) \quad \alpha_{j,n} = \sum_{i=1}^n x_{i-1,j}x_{i,j}, \quad \beta_{j,n} = \sum_{i=1}^n x_{i-1,j}^2, \quad j \geq 1, \quad n \geq 2.$$

4. ASYMPTOTIC EFFICIENCY AND EQUIVALENCE

In this section, sufficient conditions are derived to ensure the asymptotic efficiency and equivalence of the diagonal componentwise estimators of ρ formulated in the classical (see equation (2.18)), and in the bayesian (see equations (3.3)–(3.4)) frameworks.

Theorem 4.1. *Under conditions **A1**, **A2**, **A2B**, **A3** and **A4**, assume that the ARH(1) process X satisfies, for each $j \geq 1$, and, for every $T \geq 2$,*

$$(4.1) \quad \sum_{i=1}^T \varepsilon_j(i)X_{i-1,j} \geq 0, \quad \text{a.s..}$$

That is, $\{\varepsilon_j(i), i \geq 1\}$ and $\{X_{i-1,j}, i \geq 0\}$ are almost surely positive empirically correlated. In addition, for every $j \geq 1$, the hyper-parameters a_j and b_j of the beta prior distribution, $\beta(a_j, b_j)$, are such that $a_j + b_j \geq 2$. Then, the following identities are obtained:

$$(4.2) \quad \lim_{T \rightarrow \infty} TE \left[\|\tilde{\rho}_T - \rho\|_{\mathcal{S}(H)}^2 \right] = \lim_{T \rightarrow \infty} TE \left[\|\hat{\rho}_T - \rho\|_{\mathcal{S}(H)}^2 \right] = \sum_{k=1}^{\infty} \frac{\sigma_k^2}{\lambda_k(C)} < \infty,$$

where $\hat{\rho}_T$ is defined in equation (2.18), and $\tilde{\rho}_T$ is defined from equations (3.3)–(3.4), considering

$$(4.3) \quad \tilde{\rho}_{j,T} = \frac{1}{2\beta_{j,T}} \left[(\alpha_{j,T} + \beta_{j,T}) - \sqrt{(\alpha_{j,T} - \beta_{j,T})^2 - 4\beta_{j,T}\sigma_j^2[2 - (a_j + b_j)]} \right],$$

with, as before, for each $j \geq 1$, $X_{i,j} = \langle X_i, \phi_j \rangle_H$, $i = 0, \dots, T$, and $\alpha_{j,T}$ and $\beta_{j,T}$ are given in (3.5), for every $T \geq 2$.

Proof: Under **Assumptions A1–A2**, from Remark B.1 and Corollary B.1, in Appendix B, for each $j \geq 1$, and for T sufficiently large,

$$(4.4) \quad |\hat{\rho}_{j,T}| \leq 1, \quad \text{a.s.}$$

Also, under (4.1),

$$\sum_{i=1}^T \rho_j X_{i-1,j}^2 + \varepsilon_j(i) X_{i-1} \geq \sum_{i=1}^T \rho_j X_{i-1,j}^2, \quad \text{a.s.,}$$

which is equivalent to

$$(4.5) \quad \hat{\rho}_{j,T} = \frac{\sum_{i=1}^T \rho_j X_{i-1,j}^2 + \varepsilon_j(i) X_{i-1}}{\sum_{i=1}^T X_{i-1,j}^2} \geq \rho_j, \quad \text{a.s.,}$$

for every $j \geq 1$.

From (4.5), to obtain the following a.s. inequality:

$$(4.6) \quad \begin{aligned} 2|\tilde{\rho}_{j,T} - \rho_j| &= \left| \hat{\rho}_{j,T} - \rho_j + 1 - \rho_j - \sqrt{(\hat{\rho}_{j,T} - 1)^2 - \frac{4\sigma_j^2[2 - (a_j + b_j)]}{\beta_{j,T}}} \right| \\ &\leq 2|\hat{\rho}_{j,T} - \rho_j|, \quad \text{a.s.,} \quad j \geq 1, \end{aligned}$$

it is sufficient that

$$-\hat{\rho}_{j,T} + \rho_j \leq 1 - \rho_j - \sqrt{(\hat{\rho}_{j,T} - 1)^2 - \frac{4\sigma_j^2[2 - (a_j + b_j)]}{\beta_{j,T}}} \leq \hat{\rho}_{j,T} - \rho_j, \quad \text{a.s.,}$$

which is equivalent to

$$(4.7) \quad 0 \leq -\frac{2 - (a_j + b_j)}{\beta_{j,T}} \leq 4(\hat{\rho}_{j,T} - \rho_j)(1 - \rho_j) \frac{\beta_{j,T}}{4\sigma_j^2}, \quad \text{a.s..}$$

That is, keeping in mind that $\sigma_j^2 = \lambda_j(C)(1 - \rho_j^2) = \lambda_j(C)(1 + \rho_j)(1 - \rho_j)$, condition (4.7) can also be expressed as

$$0 \leq -\frac{2 - (a_j + b_j)}{\beta_{j,T}} \leq 4(\hat{\rho}_{j,T} - \rho_j)(1 - \rho_j) \frac{\beta_{j,T}}{4\lambda_j(C)(1 + \rho_j)(1 - \rho_j)}, \quad \text{a.s.}$$

i.e.,

$$(4.8) \quad 0 \leq -\frac{2 - (a_j + b_j)}{\beta_{j,T}} \leq (\hat{\rho}_{j,T} - \rho_j) \frac{\beta_{j,T}}{\lambda_j(C)(1 + \rho_j)}, \quad \text{a.s.,}$$

for $j \geq 1$. Since, for each $j \geq 1$, $\frac{\beta_{j,T}}{\lambda_j(C)(1 + \rho_j)} \geq \frac{\beta_{j,T}}{2\lambda_j(C)}$, it is sufficient that

$$(4.9) \quad 0 \leq -\frac{2 - (a_j + b_j)}{\beta_{j,T}} \leq (\hat{\rho}_{j,T} - \rho_j) \frac{\beta_{j,T}}{2\lambda_j(C)}, \quad \text{a.s.}$$

to hold to ensure that inequality (4.6) is satisfied. Furthermore, from Remark B.1 and Corollary B.1, in Appendix B, for each $j \geq 1$, $\beta_{j,T} \rightarrow \infty$, and

$$\beta_{j,T} = \mathcal{O}(T), \quad T \rightarrow \infty, \quad \text{a.s.}, \quad j \geq 1.$$

Also, we have, from such remark and corollary, that

$$(\widehat{\rho}_{j,T} - \rho_j) = \mathcal{O}(1), \quad T \rightarrow \infty, \quad \text{a.s.}, \quad j \geq 1.$$

Thus, for each $j \geq 1$, the upper bound, in (4.9), diverges as $T \rightarrow \infty$, which means, that, for T sufficiently large, inequality (4.6) holds, if $a_j + b_j \geq 2$, for each $j \geq 1$.

Now, from (4.6), under **Assumption A3**, for each $j \geq 1$,

$$(4.10) \quad \begin{aligned} T|\widehat{\rho}_{j,T} - \rho_j|^2 &\leq \widetilde{M}^2(j), \quad \text{a.s.} \\ T|\widetilde{\rho}_{j,T} - \rho_j|^2 &\leq T|\widehat{\rho}_{j,T} - \rho_j|^2 \leq \widetilde{M}^2(j), \quad \text{a.s.} \end{aligned}$$

Furthermore, for each $j \geq 1$, $\beta_{j,T} \rightarrow \infty$, and $\beta_{j,T} = \mathcal{O}(T)$, as $T \rightarrow \infty$, almost surely. Hence,

$$-\frac{4\sigma_j^2[2 - (a_j + b_j)]}{\beta_{j,T}} \rightarrow 0, \quad T \rightarrow \infty, \quad \text{a.s.}, \quad \forall j \geq 1.$$

From equation (4.3), we then have that, for each $j \geq 1$,

$$(4.11) \quad \begin{aligned} \lim_{T \rightarrow \infty} \left| \widetilde{\rho}_{j,T} - \widehat{\rho}_{j,T} \right| &= \lim_{T \rightarrow \infty} \left| \frac{1}{2} [(\widehat{\rho}_{j,T} + 1) - ((\widehat{\rho}_{j,T} - 1)^2 \right. \\ &\quad \left. - \frac{4}{\beta_{j,T}} \sigma_j^2 [2 - (a_j + b_j)])^{1/2}] - \widehat{\rho}_{j,T} \right| = \lim_{T \rightarrow \infty} |\widehat{\rho}_{j,T} - \widehat{\rho}_{j,T}| = 0, \end{aligned}$$

almost surely. Thus, the almost surely convergence, when $T \rightarrow \infty$, of $\widetilde{\rho}_{j,T}$ and $\widehat{\rho}_{j,T}$ to the same limit is obtained, for every $j \geq 1$.

From equation (4.10),

$$(4.12) \quad T|\widetilde{\rho}_{j,T} - \widehat{\rho}_{j,T}|^2 \leq 2T \left[\left(\widetilde{\rho}_{j,T} - \rho_j \right)^2 + \left(\widehat{\rho}_{j,T} - \rho_j \right)^2 \right] \leq 4\widetilde{M}^2(j), \quad \text{a.s.}$$

Since $E[\widetilde{M}^2(j)] < \infty$, applying Dominated Convergence Theorem, from equation (4.12), considering (2.23) we obtain, for each $j \geq 1$,

$$(4.13) \quad \lim_{T \rightarrow \infty} TE[\widetilde{\rho}_{j,T} - \rho_j]^2 = \lim_{T \rightarrow \infty} TE[\widehat{\rho}_{j,T} - \rho_j]^2 = 1 - \rho_j^2.$$

Under **Assumptions A3**, from (4.10), for each $j \geq 1$, and for every $T \geq 1$,

$$\begin{aligned} TE[\widehat{\rho}_{j,T} - \rho_j]^2 &\leq E[\widetilde{M}^2(j)] \\ TE[\widetilde{\rho}_{j,T} - \rho_j]^2 &\leq E[\widetilde{M}^2(j)] \end{aligned}$$

with $\sum_{j=1}^{\infty} E[M^2(j)] < \infty$. Applying again Dominated Convergence Theorem (with integration performed with respect to a counting measure), we obtain from (4.13), keeping in mind relationship (2.13),

$$\begin{aligned}
 \lim_{T \rightarrow \infty} \sum_{j=1}^{\infty} TE[\tilde{\rho}_{j,T} - \rho_j]^2 &= \sum_{j=1}^{\infty} \lim_{T \rightarrow \infty} TE[\tilde{\rho}_{j,T} - \rho_j]^2 \\
 (4.14) \qquad &= \sum_{j=1}^{\infty} \lim_{T \rightarrow \infty} TE[\hat{\rho}_{j,T} - \rho_j]^2 = \sum_{j=1}^{\infty} 1 - \rho_j^2 \\
 &= \sum_{j=1}^{\infty} \frac{\sigma_j^2}{\lambda_j(C)} = \lim_{T \rightarrow \infty} \sum_{j=1}^{\infty} TE[\hat{\rho}_{j,T} - \rho_j]^2 < \infty,
 \end{aligned}$$

in view of equation (2.15) in **Assumption A2B**. That is, equation (4.2) holds. □

Theorem 4.2. *Under the conditions of Theorem 4.1,*

$$\begin{aligned}
 \lim_{T \rightarrow \infty} TE [\|\tilde{\rho}_T(X_T) - \rho(X_T)\|_H^2] &= \lim_{T \rightarrow \infty} TE [\|\hat{\rho}_T(X_T) - \rho(X_T)\|_H^2] \\
 (4.15) \qquad &= \sum_{k=1}^{\infty} \lambda_k(C)(1 - \rho_k^2).
 \end{aligned}$$

Here,

$$\begin{aligned}
 \tilde{\rho}_T(X_T) &= \sum_{j=1}^{\infty} \tilde{\rho}_{j,T} \langle X_T, \phi_j \rangle_H \phi_j, \\
 \tilde{\rho}_{j,T} &= \frac{1}{2\beta_{j,T}} \left[(\alpha_{j,T} + \beta_{j,T}) - \sqrt{(\alpha_{j,T} - \beta_{j,T})^2 - 4\beta_{j,T}\sigma_j^2[2 - (a_j + b_j)]} \right], \quad j \geq 1, \\
 \hat{\rho}_T(X_T) &= \sum_{j=1}^{\infty} \hat{\rho}_{j,T} \langle X_T, \phi_j \rangle_H \phi_j, \quad \hat{\rho}_{j,T} = \frac{\sum_{i=1}^T X_{i-1,j} X_{i,j}}{\sum_{i=1}^T X_{i-1,j}^2}, \quad j \geq 1, \\
 \rho(X_T) &= \sum_{j=1}^{\infty} \rho_j \langle X_T, \phi_j \rangle_H \phi_j, \quad \rho_j = \rho(\phi_j)(\phi_j), \quad j \geq 1.
 \end{aligned}$$

Proof: From equation (4.11), for every $j, k \geq 1$,

$$(4.16) \qquad [(\tilde{\rho}_{j,T} - \hat{\rho}_{j,T})(\tilde{\rho}_{k,T} - \hat{\rho}_{k,T})]^2 \rightarrow 0, \quad T \rightarrow \infty, \quad \text{a.s.}$$

In addition, from equation (4.12), for every $j, k \geq 1$,

$$(4.17) \qquad [(\tilde{\rho}_{j,T} - \hat{\rho}_{j,T})(\tilde{\rho}_{k,T} - \hat{\rho}_{k,T})]^2 \leq 16 \frac{\tilde{M}^2(k)\tilde{M}^2(j)}{T^2} \leq 16\tilde{M}^2(k)\tilde{M}^2(j),$$

with $E[\widetilde{M}^2(k)\widetilde{M}^2(j)] = E[\widetilde{M}^2(k)]E[\widetilde{M}^2(j)] < \infty$, under **Assumption A3**. Applying Dominated Convergence Theorem from (4.17), the almost surely convergence in (4.16) implies the convergence in mean to zero, when $T \rightarrow \infty$. Furthermore, under **Assumption A3**, for $T \geq 2$,

$$(4.18) \quad \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} T^2 E[(\widetilde{\rho}_{j,T}^- - \widehat{\rho}_{j,T})(\widetilde{\rho}_{k,T}^- - \widehat{\rho}_{k,T})]^2 \leq 16 \left[\sum_{j,k; j \neq k}^{\infty} E[\widetilde{M}^2(j)]E[\widetilde{M}^2(k)] \right] + \left[\sum_{k=1}^{\infty} E[\widetilde{M}^4(k)] \right] < \infty.$$

From (4.18), for every $T \geq 2$,

$$(4.19) \quad T^2 E\|\widetilde{\rho}_T^- - \widehat{\rho}_T\|_{\mathcal{S}(H)}^4 = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} T^2 E[(\widetilde{\rho}_{j,T}^- - \widehat{\rho}_{j,T})(\widetilde{\rho}_{k,T}^- - \widehat{\rho}_{k,T})]^2 \leq 16 \left[\sum_{j,k; j \neq k}^{\infty} E[\widetilde{M}^2(j)]E[\widetilde{M}^2(k)] \right] + \left[\sum_{k=1}^{\infty} E[\widetilde{M}^4(k)] \right] < \infty.$$

Equation (4.19) means that the rate of convergence to zero, as $T \rightarrow \infty$, of the functional sequence $\{\widetilde{\rho}_T^- - \widehat{\rho}_T\}_{T \geq 2}$ in the space $\mathcal{L}_{\mathcal{S}(H)}^4(\Omega, \mathcal{A}, P)$ is of order T^{-2} .

From definition of the norm in the space bounded linear operators, applying Cauchy–Schwarz inequality, we obtain

$$(4.20) \quad \begin{aligned} E\|\widetilde{\rho}_T^-(X_T) - \widehat{\rho}_T(X_T)\|_H^2 &\leq E \left[\|\widetilde{\rho}_T^- - \widehat{\rho}_T\|_{\mathcal{L}(H)}^2 \|X_T\|_H^2 \right] \\ &\leq \sqrt{E \left[\|\widetilde{\rho}_T^- - \widehat{\rho}_T\|_{\mathcal{L}(H)}^4 \right]} \sqrt{E \left[\|X_T\|_H^4 \right]} \\ &\leq \sqrt{E \left[\|\widetilde{\rho}_T^- - \widehat{\rho}_T\|_{\mathcal{S}(H)}^4 \right]} \sqrt{E \left[\|X_T\|_H^4 \right]}. \end{aligned}$$

From the orthogonal expansion (2.5) of X_T , in terms of the independent real-valued standard Gaussian random variables $\{\eta_k(T)\}_{k \geq 1}$, we have

$$(4.21) \quad \begin{aligned} E \left[\|X_T\|_H^4 \right] &= \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \lambda_j(C)\lambda_k(C) E[\eta_j(T)\eta_k(T)]^2 = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \lambda_j(C)\lambda_k(C) 3\delta_{j,k} \\ &= 3 \sum_{k=1}^{\infty} \lambda_k^2(C) < \infty. \end{aligned}$$

From equations (4.19)–(4.21),

$$E\|\widetilde{\rho}_T^-(X_T) - \widehat{\rho}_T(X_T)\|_H^2 = \mathcal{O} \left(\frac{1}{T} \right), \quad T \rightarrow \infty.$$

Thus, $\widetilde{\rho}_T(X_T)$ and $\widehat{\rho}_T(X_T)$ have the same limit in the space $\mathcal{L}_H^2(\Omega, \mathcal{A}, P)$.

We now prove the approximation by trace $(C(I - \rho^2))$ of the limit, in equation (4.15). Consider

$$\begin{aligned}
 & E\|\widehat{\rho}_T(X_T) - \rho(X_T)\|_H^2 - \text{trace}(C(I - \rho^2)) \\
 (4.22) \quad &= \sum_{k=1}^{\infty} E[(\widehat{\rho}_{k,T} - \rho_k)^2 \eta_k^2(T)] \lambda_k(C) - \lambda_k(C)(1 - \rho_k^2),
 \end{aligned}$$

where $\text{trace}(C(I - \rho^2)) = \sum_{k=1}^{\infty} \lambda_k(C)(1 - \rho_k^2)$. From Lemmas 2.1 and 2.2 (see last identity in equation (2.22)), for each $k \geq 1$, and for T sufficiently large,

$$\begin{aligned}
 (4.23) \quad & E[(\widehat{\rho}_{k,T} - \rho_k)^2 \eta_k^2(T)] \simeq \text{Var}(\widehat{\rho}_{k,T} - \rho_k) \text{Var}(\eta_k) \\
 & \times (1 + 2[\text{Corr}(\widehat{\rho}_{k,T} - \rho_k, \eta_k(T))]^2).
 \end{aligned}$$

Under **Assumption A3**, from equations (2.17)–(2.20), for every $k \geq 1$,

$$(4.24) \quad T \text{Var}(\widehat{\rho}_{k,T} - \rho_k) \leq (1 - \rho_k^2) E[\widetilde{M}^2(k)]$$

From equations (4.22)–(4.24),

$$\begin{aligned}
 (4.25) \quad & TE\|\widehat{\rho}_T(X_T) - \rho(X_T)\|_H^2 - \text{trace}(C(I - \rho^2)) \\
 & \leq \sum_{k=1}^{\infty} \lambda_k(C)(1 - \rho_k^2) E[\widetilde{M}^2(k)] \\
 & \times [1 + 2[\text{Corr}(\widehat{\rho}_{k,T} - \rho_k, \eta_k(T))]^2] - \lambda_k(C)(1 - \rho_k^2) \\
 & \leq \sum_{k=1}^{\infty} 3\lambda_k(C) E[\widetilde{M}^2(k)] - \sum_{k=1}^{\infty} \lambda_k(C)(1 - \rho_k^2) < \infty,
 \end{aligned}$$

since $\sum_{k=1}^{\infty} \lambda_k(C)(1 - \rho_k^2) \leq \sum_{k=1}^{\infty} \lambda_k(C) < \infty$, by the trace property of C . Here, we have applied Cauchy–Schwartz inequality to obtain, for a certain constant $L > 0$,

$$\begin{aligned}
 (4.26) \quad & \sum_{k=1}^{\infty} 3\lambda_k(C) E[\widetilde{M}^2(k)] \leq 3 \sqrt{\sum_{k=1}^{\infty} \lambda_k^2(C) \sum_{k=1}^{\infty} [E[\widetilde{M}^2(k)]]^2} \\
 & \leq 3L \sqrt{\sum_{k=1}^{\infty} \lambda_k(C) \sum_{k=1}^{\infty} E[\widetilde{M}^2(k)]} < \infty,
 \end{aligned}$$

from the trace property of C , and since $\sum_{k=1}^{\infty} E[\widetilde{M}^2(k)] < \infty$, under **Assumption A3**. From equations (2.23) and (4.25), one can get, applying Dominated

Convergence Theorem,

$$\begin{aligned}
 \lim_{T \rightarrow \infty} TE \|\widehat{\rho}_T(X_T) - \rho(X_T)\|_H^2 &= \sum_{k=1}^{\infty} \lambda_k(C) \lim_{T \rightarrow \infty} TE [\widehat{\rho}_{k,T} - \rho_k]^2 \\
 &\quad \times \lim_{T \rightarrow \infty} [1 + [\text{Corr}(\widehat{\rho}_{k,T} - \rho_k, \eta_k(T))]^2] \\
 (4.27) \qquad \qquad \qquad &= \sum_{k=1}^{\infty} \lambda_k(C) \lim_{T \rightarrow \infty} TE [\widehat{\rho}_{k,T} - \rho_k]^2 \\
 &= \sum_{k=1}^{\infty} \lambda_k(C) (1 - \rho_k^2),
 \end{aligned}$$

where we have considered that

$$\begin{aligned}
 \lim_{T \rightarrow \infty} |\text{Cov}(\widehat{\rho}_{k,T} - \rho_k, \eta_k(T))|^2 &\leq \lim_{T \rightarrow \infty} E[\widehat{\rho}_{k,T} - \rho_k]^2 E[\eta_k(T)]^2 \\
 &= \lim_{T \rightarrow \infty} \frac{1 - \rho_k^2}{T} = 0. \qquad \square
 \end{aligned}$$

5. NUMERICAL EXAMPLES

This section illustrates the theoretical results derived on asymptotic efficiency and equivalence of the proposed classical and bayesian diagonal componentwise estimators of the autocorrelation operator, as well as of the associated ARH(1) plug-in predictors. Under the conditions assumed in Theorem 4.1, three examples of standard zero-mean Gaussian ARH(1) processes are generated, respectively corresponding to consider different rates of convergence to zero of the eigenvalues of the autocovariance operator. The truncation order k_T in Examples 1 and 2 is fixed, i.e., it does not depend on the sample size T (see equations (5.2) and (5.3) below). While in Example 3, k_T is selected such that

$$(5.1) \qquad \qquad \qquad \lim_{T \rightarrow \infty} C_{k_T} \sqrt{T} = \infty.$$

Specifically, in the first two examples, the choice of k_T is driven looking for a compromise between the sample size and the number of parameters to be estimated. With this aim the value $k_T = 5$ is fixed, independently of T . This is the number of parameters that can be estimated in an efficient way, from most of the values of the sample size T studied. In Example 3, the truncation parameter k_T is defined as a fractional power of the sample size. Note that Example 3 corresponds to the fastest decay velocity of the eigenvalues of the autocovariance operator. Hence, the lowest truncation order for a given sample size must be selected according to the truncation rule (5.1).

The generation of $N = 1000$ realizations of the functional values X_t , $t = 0, 1, \dots, T$, for $T = 250, 500, 750, 1000, 1250, 1500, 1750, 2000$, denoting as before the sample size, is performed, for each one of the ARH(1) processes, defined in the three examples below. Based on those generations, and on the sample sizes studied, the truncated empirical functional mean-square errors of the classical and Bayes diagonal componentwise parameter estimators of the autocorrelation operator ρ are computed as follows:

$$(5.2) \quad EFMSE_{\bar{\rho}_T} = \frac{1}{N} \sum_{\omega=1}^N \sum_{j=1}^{k_n} (\bar{\rho}_{j,T}^\omega - \rho_j)^2,$$

$$(5.3) \quad EFMSE_{\bar{\rho}_T(X_T)} = \frac{1}{N} \sum_{\omega=1}^N \sum_{j=1}^{k_n} (\bar{\rho}_{j,T}^\omega - \rho_j)^2 X_{T,j}^2,$$

where $\bar{\rho}_{j,T}^\omega$ can be the classical $\hat{\rho}_{j,T}$ or the Bayes $\tilde{\rho}_{j,T}$ diagonal componentwise estimator of the autocorrelation operator, and ω denotes the sample point $\omega \in \Omega$ associated with each one of the $N = 1000$ realizations generated of each functional value of the ARH(1) process X .

On the other hand, as assumed in the previous section, $\rho_k \sim B(a_k, b_k)$, with $a_k + b_k \geq 2$, $a_k > 0$ and $b_k > 1$, for each $k \geq 1$. Thus, parameters (a_k, b_k) are defined as follows:

$$(5.4) \quad b_k = 1 + 1/100, \quad a_k = 2^k, \quad k \geq 1,$$

where

$$(5.5) \quad \begin{aligned} E[\rho_k] &= \frac{a_k}{a_k + b_k} \rightarrow 1, \\ V[\rho_k] &= \frac{a_k b_k}{(a_k + b_k + 1)(a_k + b_k)^2} = \mathcal{O}\left(\frac{1}{2^{2k}}\right), \end{aligned} \quad k \rightarrow \infty,$$

with $\{\rho_k^2, k \geq 1\}$ being a random sequence such that its elements tend to be concentrated around point one, when $k \rightarrow \infty$. From (5.5), since

$$(5.6) \quad \sigma_k^2 = \lambda_k(C) (1 - \rho_k^2), \quad k \geq 1,$$

Assumption A2B is satisfied. In addition, Condition 4.1 is verified in the generations performed in the Gaussian framework.

5.1. Example 1

Let us assume that the eigenvalues of the autocovariance operator of the ARH(1) process X are given by

$$\lambda_k(C) = \frac{1}{k^{3/2}}, \quad k \geq 1.$$

Thus, C is a strictly positive and trace operator, where $\{\rho_k^2, k \geq 1\}$ and $\{\sigma_k^2, k \geq 1\}$ are generated from (5.4)–(5.6).

Tables 1 and 2 display the values of the empirical functional mean-square errors, given in (5.2)–(5.3), associated with $\hat{\rho}_T$ and $\tilde{\rho}_T$, and with the corresponding ARH(1) plug-in predictors, with, as before,

$$(5.7) \quad T = 250, 500, 750, 1000, 1250, 1500, 1750, 2000,$$

considering $k_T = 5$. The respective graphical representations are displayed in Figures 1–2, where, for comparative purposes, the values of the curve $1/T$ are also drawn for the finite sample sizes (5.7).

Table 1: Example 1. Empirical functional mean-square errors $EFMSE_{\hat{\rho}_T}$.

Sample size	Classical estimator $\hat{\rho}_T$	Bayes estimator $\tilde{\rho}_T$
250	2.13 e–003	2.23 e–003
500	1.24 e–003	1.04 e–003
750	8.44 e–004	7.13 e–004
1000	6.91 e–004	5.84 e–004
1250	5.97 e–004	4.72 e–004
1500	4.89 e–004	3.98 e–004
1750	4.13 e–004	3.06 e–004
2000	3.61 e–004	2.59 e–004

Table 2: Example 1. Empirical functional mean-square errors $EFMSE_{\hat{\rho}_T(X_T)}$.

Sample size	Classical predictor $\hat{\rho}_T(X_T)$	Bayes predictor $\tilde{\rho}_T(X_T)$
250	1.22 e–003	1.42 e–003
500	6.08 e–004	6.36 e–004
750	3.24 e–004	4.06 e–004
1000	3.05 e–004	2.77 e–004
1250	2.74 e–004	2.39 e–004
1500	2.07 e–004	1.78 e–004
1750	1.71 e–004	1.48 e–004
2000	1.64 e–004	1.42 e–004

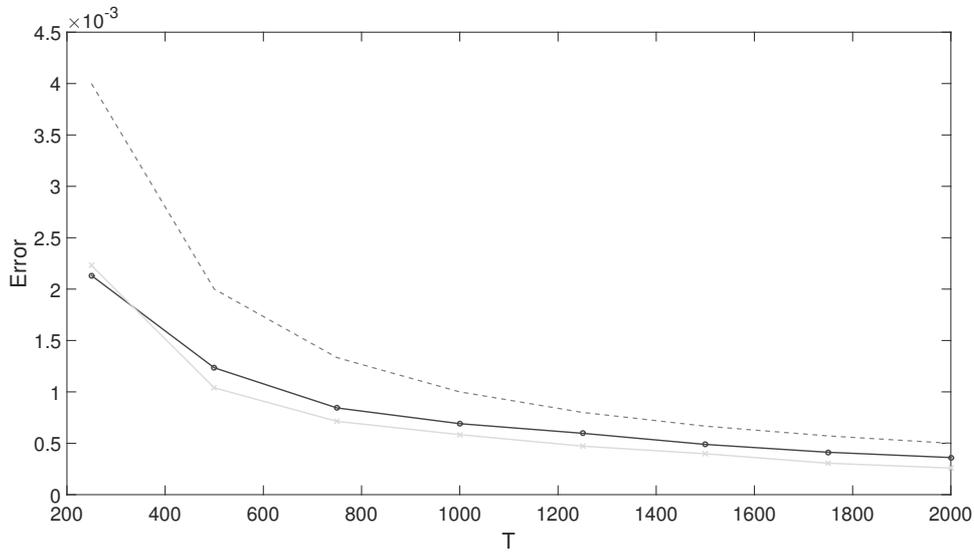


Figure 1: Example 1. Empirical functional mean-square errors of classical (blue circle line), and Bayes (green cross line) componentwise ARH(1) parameter estimators, with $k_T = 5$, for $N = 1000$ replications of the ARH(1) values, against the curve $1/T$ (red dot line), for $T = 250, 500, 750, 1000, 1250, 1500, 1750, 2000$.

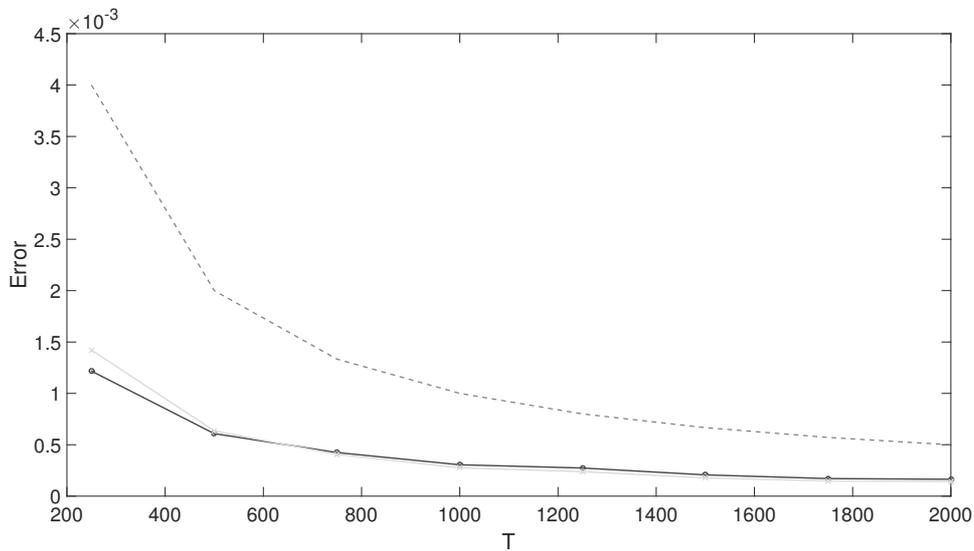


Figure 2: Example 1. Empirical functional mean-square errors of classical (blue circle line), and Bayes (green cross line) componentwise ARH(1) plug-in predictors, with $k_T = 5$, for $N = 1000$ replications of the ARH(1) values, against the curve $1/T$ (red dot line), for $T = 250, 500, 750, 1000, 1250, 1500, 1750, 2000$.

5.2. Example 2

In this example, a bit slower decay velocity, than in Example 1, of the eigenvalues of the autocovariance operator of the ARH(1) process is considered. Specifically,

$$\lambda_k(C) = \frac{1}{k^{1+1/10}}, \quad k \geq 1.$$

Thus, C is a strictly positive self-adjoint trace operator, where $\{\rho_k^2, k \geq 1\}$ and $\{\sigma_k^2, k \geq 1\}$ are generated, as before, from (5.4)–(5.6).

Tables 3 and 4 show the values of the empirical functional mean square errors, associated with $\hat{\rho}_T$ and $\tilde{\rho}_T$, and with the corresponding ARH(1) plug-in predictors, respectively. Figures 3–4 provide the graphical representations in comparison with the values of the curve $1/T$ for T given in (5.7), with, as before, $k_T = 5$.

Table 3: Example 2. Empirical functional mean-square errors $EFMSE_{\tilde{\rho}_T}$.

Sample size	Classical estimator $\hat{\rho}_T$	Bayes estimator $\tilde{\rho}_T$
250	4.18 e–003	6.09 e–003
500	2.20 e–003	2.30 e–003
750	1.52 e–003	1.39 e–003
1000	1.14 e–003	1.00 e–003
1250	9.55 e–004	7.97 e–004
1500	7.97 e–004	6.64 e–004
1750	7.01 e–004	5.37 e–004
2000	6.22 e–004	5.00 e–004

Table 4: Example 2. Empirical functional mean-square errors $EFMSE_{\tilde{\rho}_T(X_T)}$.

Sample size	Classical predictor $\hat{\rho}_T(X_T)$	Bayes predictor $\tilde{\rho}_T(X_T)$
250	3.25 e–003	3.18 e–003
500	1.59 e–003	1.40 e–003
750	9.47 e–004	8.19 e–004
1000	7.89 e–004	6.88 e–004
1250	7.24 e–004	6.10 e–004
1500	5.53 e–004	4.77 e–004
1750	5.31 e–004	4.49 e–004
2000	4.61 e–004	4.00 e–004

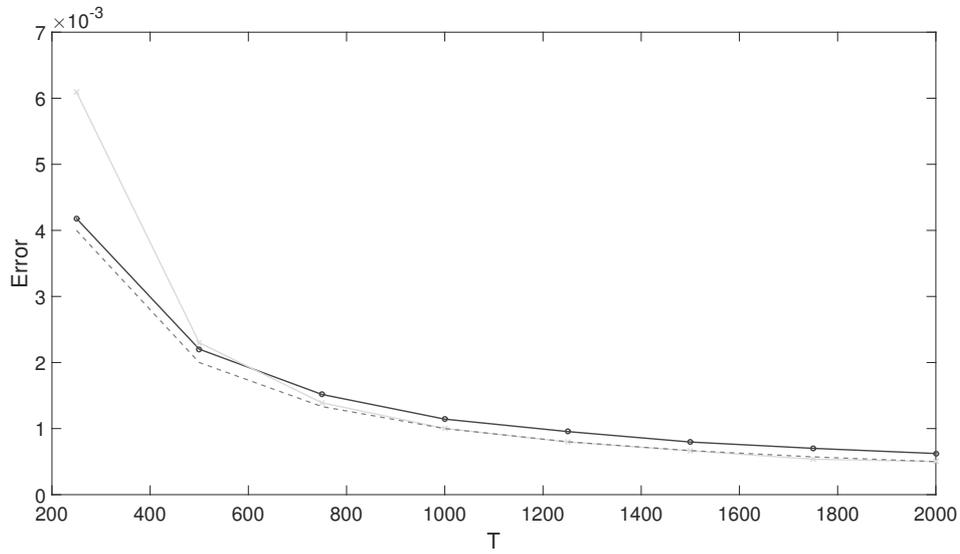


Figure 3: Example 2. Empirical functional mean-square errors of classical (blue circle line), and Bayes (green cross line) componentwise ARH(1) parameter estimators, with $k_T = 5$, for $N = 1000$ replications of the ARH(1) values, against the curve $1/T$ (red dot line), for $T = 250, 500, 750, 1000, 1250, 1500, 1750, 2000$.

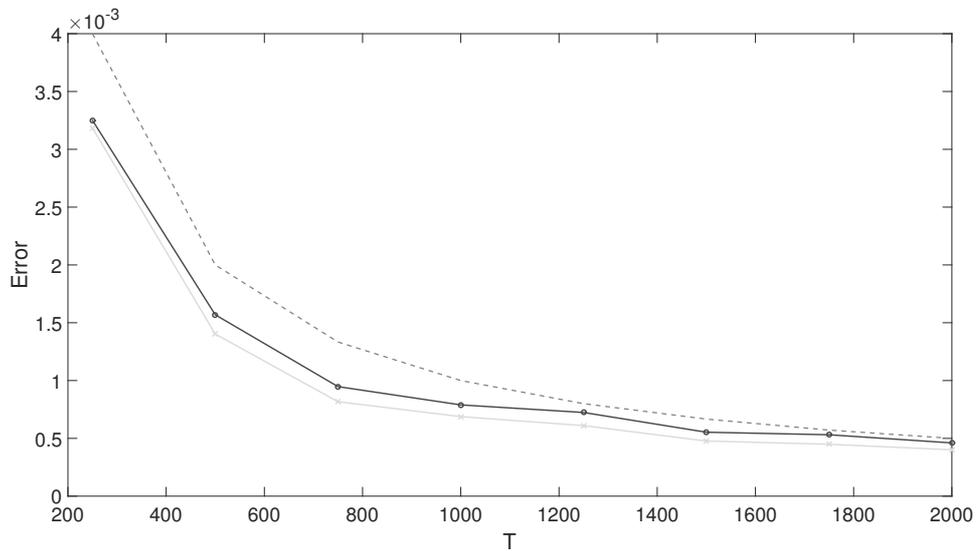


Figure 4: Example 2. Empirical functional mean-square errors of classical (blue circle line), and Bayes (green cross line) componentwise ARH(1) plug-in predictors, with $k_T = 5$, for $N = 1000$ replications of the ARH(1) values, against the curve $1/T$ (red dot line), for $T = 250, 500, 750, 1000, 1250, 1500, 1750, 2000$.

5.3. Example 3

It is well-known that the singularity of the inverse of the autocovariance operator C increases, when the rate of convergence to zero of the eigenvalues of C indicates a faster decay velocity, as in this example. Specifically, here,

$$(5.8) \quad \lambda_k(C) = \frac{1}{k^2}, \quad k \geq 1.$$

As before, $\{\rho_k^2, k \geq 1\}$ and $\{\sigma_k^2, k \geq 1\}$ are generated from (5.4)–(5.6). The truncation order k_T satisfies

$$(5.9) \quad k_T = \lceil T^{1/\alpha} \rceil, \quad \lim_{T \rightarrow \infty} k_T = \infty, \quad \lim_{T \rightarrow \infty} \sqrt{T} C_{k_T} = \infty$$

(see also the simulation study undertaken in Álvarez-Liébana, Bosq and Ruiz-Medina, 2017, for the case of ρ being a Hilbert–Schmidt operator). In particular, (5.9) holds for $\frac{1}{2} - \frac{2}{\alpha} > 0$. Thus, $\alpha > 4$, and we consider $\alpha = 4.1$, i.e., $k_T = \lceil T^{1/4.1} \rceil$.

Table 5: Example 3. Empirical functional mean-square errors $EFMSE_{\tilde{\rho}_T}$.

Sample size	k_n	Classical estimator $\hat{\rho}_T$	Bayes estimator $\tilde{\rho}_T$
250	3	1.73 e–003	1.52 e–003
500	4	9.72 e–004	1.01 e–003
750	5	6.98 e–004	7.10 e–004
1000	5	5.63 e–004	4.35 e–004
1250	5	4.49 e–004	2.84 e–004
1500	5	3.94 e–004	2.24 e–004
1750	6	3.31 e–004	1.84 e–004
2000	7	3.05 e–004	1.70 e–004

Table 6: Example 3. Empirical functional mean-square errors $EFMSE_{\tilde{\rho}_T(X_T)}$.

Sample size	k_n	Classical predictor $\hat{\rho}_T(X_T)$	Bayes predictor $\tilde{\rho}_T(X_T)$
250	3	1.92 e–003	1.31 e–003
500	4	8.24 e–004	5.75 e–004
750	5	5.60 e–004	4.08 e–004
1000	5	3.52 e–004	2.54 e–004
1250	5	2.62 e–004	1.45 e–004
1500	5	2.00 e–004	1.02 e–004
1750	6	1.37 e–004	9.57 e–005
2000	6	1.13 e–004	8.55 e–005

Tables 5–6 show the empirical functional mean square errors associated with $\widehat{\rho}_T$ and $\widetilde{\rho}_T$, and with the corresponding ARH(1) plug-in predictors, respectively. As before, Figures 5 and 6 provide the graphical representations, and the values of the curve $1/T$, for T in (5.7), with the aim of illustrating the rate of convergence to zero of the truncated empirical functional mean quadratic errors.

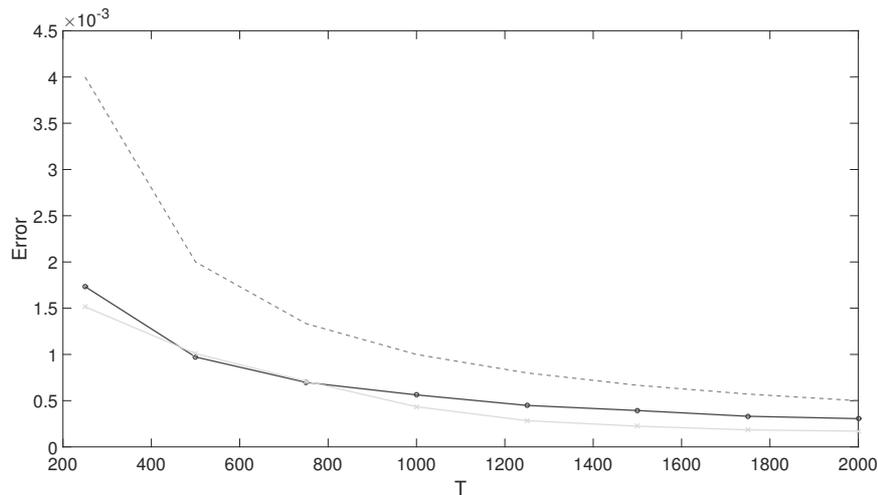


Figure 5: Example 3. Empirical functional mean-square errors of classical (blue circle line), and Bayes (green cross line) componentwise ARH(1) parameters estimators, with $k_T = \lceil T^{1/\alpha} \rceil$, $\alpha = 4.1$, for $N = 1000$ replications of the ARH(1) values, against the curve $1/T$ (red dot line), for $T = 250, 500, 750, 1000, 1250, 1500, 1750, 2000$.

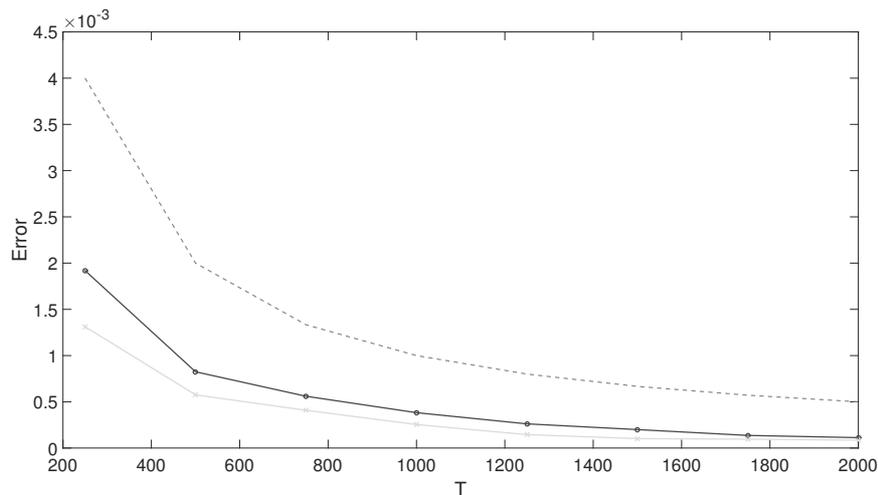


Figure 6: Example 3. Empirical functional mean-square errors of classical (blue circle line), and Bayes (green cross line) componentwise ARH(1) plug-in predictors, with $k_T = \lceil T^{1/\alpha} \rceil$, $\alpha = 4.1$, for $N = 1000$ replications of the ARH(1) values, against the curve $1/T$ (red dot line), for $T = 250, 500, 750, 1000, 1250, 1500, 1750, 2000$.

In Examples 1 and 2, where a common fixed truncation order is considered, we can observe that the biggest values of the empirical functional mean-square errors are located at the smallest sample sizes, for which the number $k_T = 5$ of parameters to be estimated is too large, with a slightly worse performance for those sample sizes, in Example 2, where a slower decay velocity, than in Example 1, of the eigenvalues of the autocovariance operator C is considered. Note that, on the other hand, when a slower decay velocity of the eigenvalues of C is given, a larger truncation order is required to explain a given percentage of the functional variance. For the fastest rate of convergence to zero of the eigenvalues of the autocovariance operator C , in Example 3, to compensate the singularity of the inverse covariance operator C^{-1} , a suitable truncation order k_T is fitted, depending on the sample size T , obtaining a slightly better performance than in the previous cases, where a fixed truncation order is studied.

6. FINAL COMMENTS

This paper addresses the case where the eigenvectors of C are known, in relation to the asymptotic efficiency and equivalence of $\hat{\rho}_{j,T}$ and $\tilde{\rho}_{j,T}^-$, and the associated plug-in predictors. However, as shown in the simulation study undertaken in Álvarez-Liébana, Bosq and Ruiz-Medina (2017), a similar performance is obtained in the case where the eigenvectors of C are unknown (see also Bosq, 2000, in relation to the asymptotic properties of the empirical eigenvectors of C).

In the cited references in the ARH(1) framework, the autocorrelation operator is usually assumed to belong to the Hilbert–Schmidt class. Here, in the absence of the compactness assumption (in particular, of the Hilbert–Schmidt assumption) on the autocorrelation operator ρ , singular autocorrelation kernels can be considered. As commented in the Introduction, the singularity of ρ is compensated by the regularity of the autocovariance kernel of the innovation process, as reflected in **Assumption A2B**.

Theorem 4.1 establishes sufficient conditions for the asymptotic efficiency and equivalence of the proposed classical and Bayes diagonal componentwise parameter estimators of ρ , as well as of the associated ARH(1) plug-in predictors (see Theorem 4.2). The simulation study illustrates the fact that the truncation order k_T should be selected according to the rate of convergence to zero of the eigenvalues of the autocovariance operator, and depending on the sample size T . Although, a fixed truncation order, independently of T , has also been tested in Examples 1 and 2, where a compromise between the rate of convergence to zero of the eigenvalues, and the rate of increasing of the sample sizes is found.

APPENDIX A — Bayesian estimation of real-valued autoregressive processes of order one

In this section, we consider the Beta-prior-based bayesian estimation of the autocorrelation coefficient ρ in a standard AR(1) process. Namely, the generalized maximum likelihood estimator of such a parameter is computed, when a beta prior is assumed for ρ . In the ARH(1) framework, we have adopted this estimation procedure in the approximation of the diagonal coefficients $\{\rho_k, k \geq 1\}$ of operator ρ with respect to $\{\phi_k \otimes \phi_k, k \geq 1\}$, in a bayesian componentwise context. Note that we also denote by ρ the autocorrelation coefficient of an AR(1) process, since there is no place for confusion here.

Let $\{X_n, n \in \mathbb{Z}\}$ be an AR(1) process satisfying

$$(A.1) \quad X_n = \rho X_{n-1} + \varepsilon_n, \quad n \in \mathbb{Z},$$

where $0 < \rho < 1$, and $\{\varepsilon_n, n \in \mathbb{Z}\}$ is a real-valued Gaussian white noise, i.e., $\varepsilon_n \sim \mathcal{N}(0, \sigma^2)$, $n \in \mathbb{Z}$, are independent Gaussian random variables, with $\sigma > 0$. Here, we will use the conditional likelihood, and assume that (x_1, \dots, x_n) are observed for n sufficiently large to ensure that the effect of the random initial condition is negligible. A beta distribution with shape parameters $a > 0$ and $b > 0$ is considered as a-priori distribution on ρ , i.e., $\rho \sim B(a, b)$. Hence, the distribution of (x_1, \dots, x_n, ρ) has density

$$\tilde{L} = \frac{1}{(\sigma\sqrt{2\pi})^n} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \rho x_{i-1})^2\right) \rho^{a-1} (1-\rho)^{b-1} \frac{\mathbb{I}_{\{0 < \rho < 1\}}}{\mathbb{B}(a, b)},$$

where

$$\mathbb{B}(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

is the beta function.

We first compute the solution to the equation

$$(A.2) \quad \begin{aligned} 0 &= \frac{\partial \log \tilde{L}}{\partial \rho} = \frac{\partial}{\partial \rho} \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \rho x_{i-1})^2 + (a-1) \log \rho + (b-1) \log(1-\rho) \right] \\ &= -\frac{1}{2\sigma^2} \sum_{i=1}^n (-2x_{i-1}(x_i - \rho x_{i-1})) + \frac{a-1}{\rho} - \frac{b-1}{1-\rho} \\ &= \frac{\alpha_n}{\sigma^2} - \frac{\rho}{\sigma^2} \beta_n + \frac{a-1}{\rho} - \frac{b-1}{1-\rho}, \end{aligned}$$

where $\alpha_n = \sum_{i=1}^n x_{i-1}x_i$ and $\beta_n = \sum_{i=1}^n x_{i-1}^2$. Thus, the following equation must be solved:

$$\begin{aligned} 0 &= \frac{\rho(1-\rho)\alpha_n}{\sigma^2} - \frac{\rho^2(1-\rho)}{\sigma^2} \beta_n + (a-1)(1-\rho) - \rho(b-1) \\ 0 &= \frac{\beta_n}{\sigma^2} \rho^3 - \frac{\alpha_n + \beta_n}{\sigma^2} \rho^2 + \left(\frac{\alpha_n}{\sigma^2} - [a+b] + 2 \right) \rho + (a-1). \end{aligned}$$

Case 1. Considering $a = b = 1$, and $\sigma^2 = 1$, we obtain the solution

$$\tilde{\rho}_n = \frac{\sum_{i=1}^n x_{i-1}x_i}{\sum_{i=1}^n x_{i-1}^2}.$$

Case 2. The general case where $b > 1$ is more intricate, since the solutions are $\tilde{\rho}_n = 0$, and

(A.3)

$$\begin{aligned} \tilde{\rho}_n &= \frac{1}{2\beta_n} \left[(\alpha_n + \beta_n) \pm \sqrt{(\alpha_n - \beta_n)^2 - 4\beta_n\sigma^2[2 - (a + b)]} \right] \\ &= \frac{[\sum_{i=1}^n x_{i-1}x_i + x_{i-1}^2] \pm \sqrt{[\sum_{i=1}^n x_{i-1}x_i - x_{i-1}^2]^2 - 4\sigma^2 [\sum_{i=1}^n x_{i-1}^2] [2 - (a + b)]}}{2 \sum_{i=1}^n x_{i-1}^2}. \end{aligned}$$

Case 3. For $\sigma^2 = a = 1$, we have

(A.4)

$$\begin{aligned} \tilde{\rho}_n &= \frac{1}{2\beta_n} \left[(\alpha_n + \beta_n) \pm \sqrt{(\alpha_n - \beta_n)^2 - 4\beta_n(1 - b)} \right] \\ &= \frac{1}{2 \sum_{i=1}^n x_{i-1}^2} \left[\sum_{i=1}^n x_{i-1}x_i + x_{i-1}^2 \right] \pm \sqrt{\left[\sum_{i=1}^n x_{i-1}x_i - x_{i-1}^2 \right]^2 - 4 \left[\sum_{i=1}^n x_{i-1}^2 \right] (1 - b)}. \end{aligned}$$

APPENDIX B — Strong-ergodic AR(1) processes

This section collects some strong-ergodicity results applied in this paper, for real-valued weak-dependent random sequences. In particular, their application to the AR(1) case is considered.

A real-valued stationary process $\{Y_n\}_{n \in \mathbb{Z}}$ is strongly-ergodic (or ergodic in an almost surely sense), with respect to $\mathbb{E}\{f(Y_0, \dots, Y_{n-1})\}$ if, as $n \rightarrow \infty$,

$$(B.1) \quad \frac{1}{n-k} \sum_{i=0}^{n-1-k} f(Y_i, \dots, Y_{i+k}) \xrightarrow{a.s.} \mathbb{E}\{f(Y_0, \dots, Y_{n-1})\}, \quad k \geq 0.$$

In particular, the following lemma provides sufficient condition to get the strong-ergodicity for all second-order moments (see, for example, Theorem 3.5.8 in Stout, 1974, and the results presented on p. 495, in Billingsley, 1995).

Lemma B.1. *Let $\{\tilde{\varepsilon}_n\}_{n \in \mathbb{Z}}$ be an i.i.d. sequence of real-valued random variables. If $f : \mathbb{R}^\infty \rightarrow \mathbb{R}$ is a measurable function, then*

$$(B.2) \quad Y_n = f(\tilde{\varepsilon}_n, \tilde{\varepsilon}_{n-1}, \dots), \quad n \in \mathbb{Z},$$

is a stationary and strongly-ergodic process for all second-order moments.

Lemma B.1 is now applied to the invertible AR(1) case, when the innovation process is white noise.

Remark B.1. If $\{Y_n\}_{n \in \mathbb{Z}}$ is a real-valued zero-mean stationary AR(1) process

$$(B.3) \quad Y_n = \rho Y_{n-1} + \tilde{\varepsilon}_n, \quad \rho \in \mathbb{R}, \quad |\rho| < 1, \quad n \in \mathbb{Z},$$

where $\{\tilde{\varepsilon}_n\}_{n \in \mathbb{Z}}$ is strong white noise, we can define the measurable (even continuous) function

$$(B.4) \quad f(a_0, a_1, \dots) = \sum_{k=0}^{\infty} \rho^k a_k,$$

such that, from Lemma B.1 and for each $n \in \mathbb{Z}$,

$$(B.5) \quad Y_n = \sum_{k=0}^{\infty} \rho^k \tilde{\varepsilon}_{n-k} = f(\tilde{\varepsilon}_n, \tilde{\varepsilon}_{n-1}, \dots),$$

is a stationary and strongly-ergodic process for all second-order moments.

In the results derived in this paper, Remark B.1 is applied, for each $j \geq 1$, to the real-valued zero-mean stationary AR(1) processes $\{X_{n,j} = \langle X_n, \phi_j \rangle_H, n \in \mathbb{Z}\}$, with $\{X_n, n \in \mathbb{Z}\}$ now representing an ARH(1) process.

Corollary B.1. *Under Assumptions A1–A2, for each $j \geq 1$, let us consider the real-valued zero-mean stationary AR(1) process $\{X_{n,j} = \langle X_n, \phi_j \rangle_H, n \in \mathbb{Z}\}$, such that, for each $n \in \mathbb{Z}$*

$$(B.6) \quad X_{n,j} = \rho_j X_{n-1,j} + \varepsilon_{n,j}, \quad \rho_j \in \mathbb{R}, \quad |\rho_j| < 1,$$

Here, $\{\varepsilon_{n,j}\}_{n \in \mathbb{Z}}$ is a real-valued strong white noise, for any $j \geq 1$. Thus, for each $j \geq 1$, $\{X_{n,j}\}_{n \in \mathbb{Z}}$ is a stationary and strongly-ergodic process for all second-order moments. In particular, for any $j \geq 1$, as $n \rightarrow \infty$,

$$(B.7) \quad \widehat{C}_{n,j} = \frac{1}{n} \sum_{i=1}^n X_{i-1,j}^2 \xrightarrow{a.s.} C_j = \mathbb{E} \{X_{i-1,j}^2\}, \quad i \geq 1$$

$$(B.8) \quad \widehat{D}_{n,j} = \frac{1}{n-1} \sum_{i=1}^n X_{i-1,j} X_{i,j} \xrightarrow{a.s.} D_j = \mathbb{E} \{X_{i-1,j} X_{i,j}\}, \quad i \geq 1.$$

ACKNOWLEDGMENTS

We would like to thank Professor Denis Bosq, for his valuable contribution to this paper. This work has been supported in part by project MTM2015–71839–P of the DGI, MINECO, Spain.

REFERENCES

- [1] ÁLVAREZ-LIÉBANA, J.; BOSQ, D. and RUIZ-MEDINA, M.D. (2016). Consistency of the plug-in functional predictor of the Ornstein–Uhlenbeck process in Hilbert and Banach spaces, *Statistics and Probability Letters*, **117**, 12–22.
- [2] ÁLVAREZ-LIÉBANA, J.; BOSQ, D. and RUIZ-MEDINA, M.D. (2017). Asymptotic properties of a componentwise ARH(1) plug-in predictor, *J. Mult. Anal.*, **155**, 12–34.
- [3] ANTONIADIS, A. and SAPATINAS, T. (2003). Wavelet methods for continuous-time prediction using Hilbert-valued autoregressive processes, *Journal of Multivariate Analysis*, **87**, 133–158.
- [4] AROIAN, L.A. (1947). The probability function of a product of two normal distributed variables, *Annals of Mathematical Statistics*, **18**, 256–271.
- [5] AUE, A.; NORINHO, D.D. and HÖRMANN, S. (2014). On the prediction of stationary functional time series, *Journal of the American Statistical Association*, **110**, 378–392 .
- [6] BARTLETT, M.S. (1946). On the theoretical specification and sampling properties of autocorrelated time series, *Journal of the Royal Statistical Society (Supplement)*, **8**, 27–41.
- [7] BESSE, P.C. and CARDOT, H. (1996). Approximation spline de la prévision d'un processus fonctionnel autorégressif d'ordre 1, *Canad. J. Statist.*, **24**, 467–487.
- [8] BILLINGSLEY, P. (1995). *Probability and Measure, 3rd edition*, John Wiley & Sons, Chicago.
- [9] BLANKE, D. and BOSQ, D. (2015). Bayesian prediction for stochastic processes: Theory and applications, *Sankhya A*, **77**, 79–105.
- [10] BOSQ, D. (2000). *Linear Processes in Function Spaces*, Springer-Verlag, New York.
- [11] BOSQ, D. and BLANKE, D. (2007). *Inference and Predictions in Large Dimensions*, John Wiley, Chichester, UK.
- [12] BOSQ, D. and RUIZ-MEDINA, M.D. (2014). Bayesian estimation in a high dimensional parameter framework, *Electron. J. Statist.*, **8**, 1604–1640.
- [13] DAUTRAY R. and LIONS J.-L. (1985). *Mathematical Analysis and Numerical Methods for Science and Technology*, Volume 3: Spectral Theory and Applications, Springer ISBN 978-3-540-66099-6.

- [14] GORENFLO, R. and MAINARDI, F. (2003). Fractional diffusion processes: Probability distribution and continuous time random walk, *Lecture Notes in Physics*, **621**, 148–166.
- [15] GUILLAS, S. (2001). Rates of convergence of autocorrelation estimates for autoregressive Hilbertian processes, *Statistics & Probability Letters*, **55**, 281–291.
- [16] HAMILTON, J.D. (1994). *Time Series Analysis*, Princeton University Press, Princeton N.J.
- [17] KHOSHNEVISAN, D. (2007). *Probability*, American Mathematical Society, UE.
- [18] KYRIAZIS, G. and PETRUSHEV, P. (2001). New bases for Triebel–Lizorkin and Besov spaces, *Transactions of the American Mathematical Society*, **354**, 749–776.
- [19] MAS, A. (1999). Normalité asymptotique de l’estimateur empirique de l’opérateur d’autocorrélation d’un processus ARH(1), *C. R. Acad. Sci. Paris Sér. I Math.*, **329**, 899–902.
- [20] MEERSCHAERT, M.M.; BENSON, D.A.; SCHEFFLER, H.-P. and BAEUMER, B. (2002). Stochastic solution of space-time fractional diffusion equations, *Phys. Rev. E*, **65**, 1103–1106.
- [21] METZLER, R. and KLAFTER, J. (2004). The restaurant at the end of the random walk: recent developments in the description of anomalous transport by fractional dynamics, *J. Physics A*, **37**, R161–R208.
- [22] MEYER, Y. and COIFMAN, R. (1997). *Wavelets, Calderón–Zygmund and Multilinear Operators*, Cambridge University Press.
- [23] RUIZ-MEDINA, M.D. and ÁLVAREZ-LIÉBANA, J. (2017). Consistent diagonal componentwise ARH(1) prediction. (Submitted).
- [24] RUIZ-MEDINA, M.D.; ROMANO, E. and FERNÁNDEZ-PASCUAL, R. (2016). Plug-in prediction intervals for a special class of standard ARH(1) processes, *Journal of Multivariate Analysis*, **146**, 138–150.
- [25] STOUT, W.F. (1974). *Almost sure convergence*, Academic, New York–London.
- [26] WARE, R. and LAD, F. (2003). Approximating the distribution for sums of product of normal variables, Research-Paper 2003-15, Department of Mathematics and Statistics (University of Canterbury – New Zealand).

ON WEIGHTED KULLBACK–LEIBLER DIVERGENCE FOR DOUBLY TRUNCATED RANDOM VARIABLES

Authors: RAJESH MOHARANA
– Department of Mathematics, National Institute of Technology Rourkela,
Rourkela-769008, India
rajeshmoharana31@gmail.com

SUCHANDAN KAYAL
– Department of Mathematics, National Institute of Technology Rourkela,
Rourkela-769008, India
suchandan.kayal@gmail.com , kayals@nitrkl.ac.in

Received: July 2016

Revised: January 2017

Accepted: March 2017

Abstract:

- In this communication, we study doubly truncated weighted Kullback–Leibler divergence (KLD) between two nonnegative random variables. The proposed measure is a generalization of the dynamic weighted KLD introduced by Yasaei Sekeh *et al.* (2013). In reliability theory and survival analysis, it plays a significant role to study several aspects of a system when lifetimes fall in a time interval. It is showed that under some conditions, the proposed measure determines the distribution function uniquely. Further, characterization theorems for various lifetime distributions are proved. The effect of the monotone transformation on the proposed measure is studied. Some inequalities and bounds in terms of useful measures are obtained and finally, few applications are provided.

Key-Words:

- *weighted Kullback–Leibler divergence; generalized failure rate; weighted geometric vitality function; proportional (reversed) hazard model; likelihood ratio order.*

AMS Subject Classification:

- 94A17, 62E10, 62N05, 60E15.

1. INTRODUCTION

Kullback–Leibler divergence (see Kullback and Leibler, 1951) is an important measure in information theory, which has proven to be useful in reliability analysis and other related fields. It measures similarity (closeness) between two statistical distributions. To be specific, let X and Y be two nonnegative absolutely continuous random variables associated with probability density functions (pdf) f and g , and cumulative distribution functions (cdf) F and G , respectively. Then the KLD between f and g is given by

$$(1.1) \quad D_{KL}(X||Y) = \int_0^{+\infty} f(x) \ln \left(\frac{f(x)}{g(x)} \right) dx = E_f \left(\ln \left(\frac{f(X)}{g(X)} \right) \right),$$

where “ln” stands for the natural logarithm. We remark that $D_{KL}(X||Y)$ is nonnegative, not symmetric in f and g , zero if the distributions match exactly. It is scale invariant, that is, for two nonnegative random variables $Z_1 = aX$ and $Z_2 = aY$ with $a > 0$, we have $D_{KL}(X||Y) = D_{KL}(Z_1||Z_2)$. Note that $D_{KL}(X||Y)$ given by (1.1), which is a special case of Csiszar’s ϕ -divergence measure can be viewed as a measure of the information loss in the fitted model relative to that in the reference model. For some recent development on KLD, we refer to Kasza and Solomon (2015) and Sankaran *et al.* (2016).

In recent past, there have been considerable interest to enlarge the concept of uncertainty by introducing nonnegative weight function. Belis and Guiasu (1968) first proposed the notion of (discrete) weighted entropy. It takes two kind of uncertainty into consideration. One of them is related to objective probability and other is related to utility. In analogy to Belis and Guiasu (1968), Di Crescenzo and Longobardi (2006) considered the weighted differential entropy for a nonnegative absolutely continuous random variable X as $S^w(X) = - \int_0^{+\infty} x f(x) \ln f(x) dx$. It is shift dependent, though the differential entropy $S(X) = - \int_0^{+\infty} f(x) \ln f(x) dx$ is not. Besides weighted differential entropy, several authors introduced and studied some other weighted information measures. In this direction, we refer to Suhov and Yasaei Sekeh (2015), Mirali *et al.* (2017), Nourbakhsh *et al.* (2016), and Rajesh *et al.* (2017).

Recently, based on the concept of weighted differential entropy, Yasaei Sekeh *et al.* (2013) considered weighted KLD as

$$(1.2) \quad D_{KL}^w(X||Y) = \int_0^{+\infty} x f(x) \ln \left(\frac{f(x)}{g(x)} \right) dx = E_f \left(X \ln \left(\frac{f(X)}{g(X)} \right) \right),$$

which takes into account the qualitative characteristic related to utility. To illustrate the importance of the weighted KLD, we consider the following example.

Example 1.1. Let X_1 and Y_1 be two nonnegative absolutely continuous random variables with pdfs $f_1(x) = 2x$, $0 < x < 1$ and $g_1(x) = 2(1 - x)$, $0 < x < 1$,

respectively. We consider another random variables X_2 and Y_2 with pdfs $f_2(x) = x/2$, $0 < x < 2$ and $g_2(x) = (2-x)/2$, $0 < x < 2$, respectively. From (1.1) we obtain $D_{KL}(X_1||Y_1) = 1$ and $D_{KL}(X_2||Y_2) = 1$. Also, from (1.2) we get $D_{KL}^w(X_1||Y_1) = 1$ and $D_{KL}^w(X_2||Y_2) = 2$. Thus, from the objective probability point of view, KLD measures are same. But, when we take the qualitative characteristics into consideration, they differ. Here, $D_{KL}^w(X_1||Y_1) < D_{KL}^w(X_2||Y_2)$.

Note that when the weight function “ x ” equals to 1, $D_{KL}^w(X||Y)$ coincides with the standard KLD given by (1.1). Yasaei Sekeh *et al.* (2013) considered weighted KLD between two residual (truncated from left) lifetime distributions and two past (truncated from right) lifetime distributions. But there exist several situations in real life, where statistical data are not only truncated from left or right side, but also from both sides. When data are truncated from left and right sides, we call it as doubly truncated. Doubly truncated data play a central role in various statistical analysis of survival data. Doubly truncated failure time occurs if the failure of an individual occurs within a certain interval. In medical science, the induction time data in AIDS are doubly truncated, since HIV was unknown to us before the year 1982. Also, during a survival experiment, sometimes it is required to collect data after an engineering system starts operating and before it fails. Let X denote the lifetime of a system. Then the conditional random variable $(X|x < X < y)$ is known as doubly truncated lifetime. That is, event time of an individual lies within a certain time interval (x, y) is only observed. Therefore, an individual is not observed if it’s event time does not fall in this predefined interval. Hence, information on the subject outside this interval is not available to the investigator. Misagh and Yari (2012) considered doubly truncated (truncated from both sides) KLD as

$$(1.3) \quad D_{KL}(X||Y; t_1, t_2) = \int_{t_1}^{t_2} \frac{f(x)}{\Delta F} \ln \left(\frac{f(x)/\Delta F}{g(x)/\Delta G} \right) dx,$$

where $\Delta F = F(t_2) - F(t_1)$, $\Delta G = G(t_2) - G(t_1)$ and $(t_1, t_2) \in \mathbb{D} = \{(x, y) | F(x) < F(y) \text{ and } G(x) < G(y)\}$. In this paper, we consider doubly truncated weighted KLD between f and g , which is given by

$$(1.4) \quad D_{KL}^w(X||Y; t_1, t_2) = \int_{t_1}^{t_2} x \frac{f(x)}{\Delta F} \ln \left(\frac{f(x)/\Delta F}{g(x)/\Delta G} \right) dx.$$

Note that $D_{KL}^w(X||Y; t_1, t_2)$ is a generalization of the measures considered by Yasaei Sekeh *et al.* (2013) in the sense that it reduces to the weighted KLD between two residual lives and two past lives, when t_2 tends to $+\infty$ and t_1 tends to 0, respectively. Mathematically,

$$\lim_{t_2 \rightarrow +\infty} D_{KL}^w(X||Y; t_1, t_2) = \int_{t_1}^{+\infty} x \frac{f(x)}{\bar{F}(t_1)} \ln \left(\frac{f(x)/\bar{F}(t_1)}{g(x)/\bar{G}(t_1)} \right) dx$$

and

$$\lim_{t_1 \rightarrow 0} D_{KL}^w(X||Y; t_1, t_2) = \int_0^{t_2} x \frac{f(x)}{F(t_2)} \ln \left(\frac{f(x)/F(t_2)}{g(x)/G(t_2)} \right) dx.$$

The doubly truncated weighted KLD given by (1.4) may take negative values, though the doubly truncated KLD always take nonnegative values. The expression given by (1.4) measures the weighted discrepancy between two systems with lifetimes X and Y , which have survived up to time t_1 and have seen to be down at time t_2 . In fact, (1.4) can be used to measure divergence between two distributions having different supports. We consider the following example to illustrate this.

Example 1.2. Let X and Y be two nonnegative absolutely continuous random variables with pdfs $f(x) = \frac{3}{4}(2x + x^2)$, $0 < x < 1$ and $g(x) = \frac{1}{8}(1 + 3y)$, $0 < y < 2$, respectively. As supports of the distributions are different, therefore, the expression given by (1.1) can not be used to compute the divergence between f and g . In this situation, one may use (1.4) for finding divergence. In particular, $D_{KL}^w(X||Y; 0.1, 0.5) = 0.033427$ and $D_{KL}^w(X||Y; 0.3, 0.9) = 0.014291$.

Again, the doubly truncated weighted KLD given by (1.4) can be expressed in terms of the doubly truncated weighted Shannon entropy and the doubly truncated weighted inaccuracy as

$$(1.5) \quad D_{KL}^w(X||Y; t_1, t_2) = -S^w(X; t_1, t_2) + I^w(X||Y; t_1, t_2),$$

where $S^w(X; t_1, t_2)$ and $I^w(X||Y; t_1, t_2)$ are defined in the next section.

The paper is arranged as follows. First, in Section 2, we recall some preliminary definitions. Few characterization results are proved in Section 3. Further, various lifetime distributions are characterized from the relationships among reliability measures. In Section 4, we analysis the effect of the affine transformations on the doubly truncated weighted KLD. Then, few inequalities and bounds are obtained in Section 5. Section 6 contains few examples in support of the results obtained in Section 5. Finally, some concluding remarks have been added in Section 7.

Throughout the paper, the random variables are taken to be nonnegative and absolutely continuous. The terms increasing and decreasing are used in non-strict sense. The differentiation, integration and expectation wherever used are assumed to exist.

2. PRELIMINARY RESULTS

In this section, we present some preliminary definitions and results which are useful for the rest of the paper. Let X and Y be two nonnegative absolutely continuous random variables with pdfs f and g , and cdfs F and G , respectively.

Then the generalized failure rate (GFR) functions of $(X|t_1 < X < t_2)$ and $(Y|t_1 < Y < t_2)$ are given by (see Navarro and Ruiz, 1996)

$$(2.1) \quad h_1^X(t_1, t_2) = \frac{f(t_1)}{\Delta F}, \quad h_2^X(t_1, t_2) = \frac{f(t_2)}{\Delta F}$$

and

$$(2.2) \quad h_1^Y(t_1, t_2) = \frac{g(t_1)}{\Delta G}, \quad h_2^Y(t_1, t_2) = \frac{g(t_2)}{\Delta G},$$

respectively for $(t_1, t_2) \in \mathbb{D}$. Note that when t_2 tends to $+\infty$, $h_1^X(t_1, t_2)$ reduces to the failure rate of X , and when t_1 tends to zero, $h_2^X(t_1, t_2)$ reduces to reversed failure rate of X . Similarly for the random variable Y . Navarro and Ruiz (1996) showed that the distribution function can be uniquely determined by GFR functions.

Definition 2.1. Let X be a nonnegative random variable with pdf f and cdf F . Then the generalized conditional mean (GCM) of a doubly truncated random variable $(X|t_1 < X < t_2)$ is given by

$$(2.3) \quad \mu_X(t_1, t_2) = E(X|t_1 < X < t_2) = \int_{t_1}^{t_2} \frac{xf(x)}{\Delta F} dx.$$

For some characterizations based on (2.3), one may refer to Ruiz and Navarro (1996).

Definition 2.2. Let X be a nonnegative random variable with pdf f and cdf F . Then the geometric vitality function for $(X|t_1 < X < t_2)$ is defined as

$$(2.4) \quad G_X(t_1, t_2) = E(\ln X|t_1 < X < t_2) = \int_{t_1}^{t_2} \frac{\ln xf(x)}{\Delta F} dx.$$

Note that $G_X(t_1, t_2)$ gives the geometric mean life of X truncated at two points t_1 and t_2 . Nair and Rajesh (2000) gave some applications of geometric vitality function. Sunoj *et al.* (2009) discussed few properties of this measure and showed that it determines the distribution function uniquely. The weighted version of the measure given by (2.4) is defined as follows.

Definition 2.3. The weighted geometric vitality function of a nonnegative random variable X with pdf f and cdf F is given by

$$(2.5) \quad G_X^w(t_1, t_2) = E(X \ln X|t_1 < X < t_2) = \int_{t_1}^{t_2} \frac{x \ln xf(x)}{\Delta F} dx.$$

Definition 2.4. For a nonnegative random variable X with pdf f and cdf F , the doubly truncated weighted Shannon entropy is given by

$$(2.6) \quad S^w(X; t_1, t_2) = - \int_{t_1}^{t_2} x \frac{f(x)}{\Delta F} \ln \left(\frac{f(x)}{\Delta F} \right) dx.$$

In the following we consider the definition of weighted inaccuracy measure between two doubly truncated random variables.

Definition 2.5. The doubly truncated weighted inaccuracy measure between two nonnegative random variables X and Y is given by

$$(2.7) \quad I^w(X||Y; t_1, t_2) = - \int_{t_1}^{t_2} x \frac{f(x)}{\Delta F} \ln \left(\frac{g(x)}{\Delta G} \right) dx.$$

Next we recall the following important definition from Shaked and Shanthikumar (2007).

Definition 2.6. Let X and Y be two random variables with pdfs f and g , and cdfs F and G , respectively. We say that X is larger than Y in likelihood ratio order, denoted by $X \geq^{lr} Y$ if $f(x)/g(x)$ is increasing in x .

Log-sum inequality: Let m be a sigma finite measure. If f and g are positive and m integrable, then

$$(2.8) \quad \int f \log \left(\frac{f}{g} \right) dm \geq \left[\int f dm \right] \log \left[\frac{\int f dm}{\int g dm} \right].$$

3. CHARACTERIZATIONS

In this section, we obtain some characterization results which may be used to describe probability distributions. The general characterization problem is to determine when the doubly truncated weighted KLD uniquely determines the distribution function. Yasaei Sekeh *et al.* (2013) showed that under some conditions, the weighted KLD for two residual and past lifetime distributions characterizes the distribution function uniquely. In the following theorem, we show that using relationship between doubly truncated weighted KLD and GCM, and under the condition on GFR functions, one can characterize one of the distributions when other is known.

Theorem 3.1. *Let X and Y be two absolutely continuous nonnegative random variables with pdfs f and g and cdfs F and G , respectively, such that*

$h_i^Y(t_1, t_2) \leq h_i^X(t_1, t_2)$, $i = 1, 2$. Then the doubly truncated weighted KLD given by (1.4) characterizes the distribution function G (or F), when F (or G) is known, provided $D_{KL}^w(X||Y; t_1, t_2) = \mu_X(t_1, t_2)$.

Proof: Differentiating (1.4) partially with respect to t_1 (for any fixed t_2) and t_2 (for any fixed t_1), we get after simplification

$$(3.1) \quad \frac{\partial D_{KL}^w(X||Y; t_1, t_2)}{\partial t_1} = h_1^X(t_1, t_2) \left[D_{KL}^w(X||Y; t_1, t_2) + \mu_X(t_1, t_2) + t_1 \ln \left(\frac{h_1^Y(t_1, t_2)}{h_1^X(t_1, t_2)} \right) \right] - h_1^Y(t_1, t_2) \mu_X(t_1, t_2)$$

and

$$(3.2) \quad \frac{\partial D_{KL}^w(X||Y; t_1, t_2)}{\partial t_2} = -h_2^X(t_1, t_2) \left[D_{KL}^w(X||Y; t_1, t_2) + \mu_X(t_1, t_2) + t_2 \ln \left(\frac{h_2^Y(t_1, t_2)}{h_2^X(t_1, t_2)} \right) \right] + h_2^Y(t_1, t_2) \mu_X(t_1, t_2).$$

Moreover, differentiating (2.3) with respect to t_1 and t_2 , we obtain

$$(3.3) \quad \frac{\partial \mu_X(t_1, t_2)}{\partial t_1} = h_1^X(t_1, t_2) \left[\mu_X(t_1, t_2) - t_1 \right]$$

and

$$(3.4) \quad \frac{\partial \mu_X(t_1, t_2)}{\partial t_2} = -h_2^X(t_1, t_2) \left[\mu_X(t_1, t_2) - t_2 \right],$$

respectively. Again, differentiating $D_{KL}^w(X||Y; t_1, t_2) = \mu_X(t_1, t_2)$ with respect to t_i (for fixed t_j , $j \neq i$), $i, j = 1, 2$ and using (3.1), (3.2), (3.3) and (3.4) we get

$$(3.5) \quad t_i h_i^X(t_1, t_2) \left[1 + \ln \left(\frac{h_i^Y(t_1, t_2)}{h_i^X(t_1, t_2)} \right) \right] + [h_i^X(t_1, t_2) - h_i^Y(t_1, t_2)] \mu_X(t_1, t_2) = 0.$$

The above equation given by (3.5) can be further written as

$$(3.6) \quad g_i(x) = t_i[1 + \ln x] + (1 - x)\mu_X(t_1, t_2) = 0, \quad i = 1, 2,$$

where $x = h_i^Y(t_1, t_2)/h_i^X(t_1, t_2)$ and $0 < x < 1$. Thus, for any fixed t_2 and arbitrary t_1 , $h_1^Y(t_1, t_2)/h_1^X(t_1, t_2)$ is a positive solution of the equation $g_1(x) = 0$. Also, for any fixed t_1 and arbitrary t_2 , $h_2^Y(t_1, t_2)/h_2^X(t_1, t_2)$ is a positive solution of the equation $g_2(x) = 0$. After some simple calculations, it is easy to show that both the solutions of $g_1(x) = 0$ and $g_2(x) = 0$ are unique. Hence, using the result that GFR functions uniquely determine the distribution function (see Navarro and Ruiz, 1996), the proof follows. This completes the proof of the theorem. \square

Theorem 3.2. *Let X and Y be two absolutely continuous nonnegative random variables with pdfs f and g and cdfs F and G , respectively. Assume $h_1^Y(t_1, t_2) = \theta h_1^X(t_1, t_2)$ and $D_{KL}^w(X||Y; t_1, t_2) > (<) (\theta - 1)\mu_X(t_1, t_2) - t_1 \log \theta$, where $\theta > 0$. If $D_{KL}^w(X||Y; t_1, t_2)$ is strictly increasing (decreasing) in t_1 for fixed t_2 , then $D_{KL}^w(X||Y; t_1, t_2)$ characterizes the distribution function uniquely.*

Proof: Under the given hypothesis, (3.1) is reduced to

$$(3.7) \quad \frac{\partial D_{KL}^w(X||Y; t_1, t_2)}{\partial t_1} = h_1^X(t_1, t_2) \left[D_{KL}^w(X||Y; t_1, t_2) + (1 - \theta)\mu_X(t_1, t_2) + t_1 \ln \theta \right].$$

The expression in (3.7) can be written further as

$$(3.8) \quad \begin{aligned} u(x) &= \frac{\partial D_{KL}^w(X||Y; t_1, t_2)}{\partial t_1} - x \left[D_{KL}^w(X||Y; t_1, t_2) + (1 - \theta)\mu_X(t_1, t_2) + t_1 \ln \theta \right] \\ &= 0, \end{aligned}$$

where $x = h_1^X(t_1, t_2)$ is a positive solution of $u(x) = 0$. Since $D_{KL}^w(X||Y; t_1, t_2) > (<)(\theta - 1)\mu_X(t_1, t_2) - t_1 \log \theta$, therefore from (3.8) it is easy to show that

$$(3.9) \quad \lim_{x \rightarrow +\infty} u(x) = -\infty \text{ } (+\infty).$$

Again, $D_{KL}^w(X||Y; t_1, t_2)$ is strictly increasing (decreasing) in t_1 for fixed t_2 . Therefore,

$$(3.10) \quad \lim_{x \rightarrow 0} u(x) = \frac{\partial D_{KL}^w(X||Y; t_1, t_2)}{\partial t_1} > (<) 0.$$

Differentiating (3.8) with respect to x we have $u'(x) = -[D_{KL}^w(X||Y; t_1, t_2) + (1 - \theta)\mu_X(t_1, t_2) + t_1 \ln \theta] < (>) 0$ implies $u(x)$ is a decreasing (increasing) function in $x > 0$. Hence, $x = h_1^X(t_1, t_2)$ is the only solution of $u(x) = 0$. This completes the proof of the theorem. □

Theorem 3.3. *Let X and Y be two absolutely continuous nonnegative random variables with pdfs f and g and cdfs F and G , respectively. Assume that for $\theta > 0$, $h_2^Y(t_1, t_2) = \theta h_2^X(t_1, t_2)$ and $D_{KL}^w(X||Y; t_1, t_2) < (>)(\theta - 1)\mu_X(t_1, t_2) - t_2 \ln \theta$. If $D_{KL}^w(X||Y; t_1, t_2)$ is strictly increasing (decreasing) in t_2 for fixed t_1 , then $D_{KL}^w(X||Y; t_1, t_2)$ characterizes the distribution function uniquely.*

Proof: Proof follows along the similar arguments of that of the Theorem 3.2. Hence it is omitted. □

It is noted that the conditions used in the above theorems are sufficient. Henceforth, we present characterization theorems for some useful continuous distributions. Let X and Y be two nonnegative absolutely continuous random variables with cdfs F and G , pdfs f and g , hazard rate functions λ_F and λ_G and reversed hazard rate functions r_F and r_G , respectively. Then X and Y are said to satisfy the proportional hazard rate model (PHRM) and proportional reversed hazard rate model (PRHRM) if for some $\theta > 0$,

$$(3.11) \quad \bar{G}(x) = [\bar{F}(x)]^\theta \text{ and } G(x) = [F(x)]^\theta,$$

respectively, where $\bar{F} = 1 - F$ and $\bar{G} = 1 - G$. The constant θ is known as proportionality constant. Several researchers used PHRM for survival data analysis. See, for instant, Cox (1972), Ebrahimi and Kirmani (1996) and Nair and Gupta (2007). On the other hand, for various results on PRHRM, we refer to Gupta and Gupta (2007) and Sankaran and Gleeja (2008). In the following consecutive theorems, we present characterizations of the first and second kind Pareto distributions.

Theorem 3.4. *Let X and Y be two absolutely continuous nonnegative random variables with pdfs f and g and cdfs F and G , respectively. Assume that F and G satisfies PHRM with proportionality constant $\theta > 0$. Then for $i = 1, 2$, the following relationship of the form*

$$(3.12) \quad \begin{aligned} D_{KL}^w(X||Y; t_1, t_2) + \mu_X(t_1, t_2) \left[\ln f(t_i) + (1 + \alpha\theta) \ln t_i + \ln \left(\frac{h_i^Y(t_1, t_2)}{h_i^X(t_1, t_2)} \right) \right] \\ = (1 + \alpha\theta) G_X^w(t_1, t_2) + E(X \ln f(X) | t_1 < X < t_2), \end{aligned}$$

holds if and only if X follows Pareto-I distribution with cdf $F(x) = 1 - (\beta/x)^\alpha$, $x > \beta > 0$, $\alpha > 0$.

Proof: The “if part” can be proved easily. To prove the “only if part”, we assume that (3.12) holds. Using (1.4) and after simplification, we get from (3.12)

$$(3.13) \quad \begin{aligned} \int_{t_1}^{t_2} x f(x) \ln \left(\frac{f(x)/\Delta F}{g(x)/\Delta G} \right) dx + \left[\ln f(t_i) + (1 + \alpha\theta) \ln t_i + \ln \left(\frac{g(t_i)}{\Delta G} \right) - \ln \left(\frac{f(t_i)}{\Delta F} \right) \right] \\ \times \int_{t_1}^{t_2} x f(x) dx = (1 + \alpha\theta) \int_{t_1}^{t_2} x f(x) \ln x dx + \int_{t_1}^{t_2} x f(x) \ln f(x) dx. \end{aligned}$$

Differentiating (3.13) with respect to t_i and then further calculations lead to

$$g(t_i) = k t_i^{-(\alpha\theta+1)}, \quad i = 1, 2 \text{ and } k > 0.$$

Hence the required result follows. This completes the proof. \square

Theorem 3.5. *Let X and Y be two absolutely continuous random variables as described in Theorem 3.4 and satisfying PHRM with proportionality constant $\theta > 0$. Then for $i = 1, 2$, the following relationship of the form*

$$(3.14) \quad D_{KL}^w(X||Y; t_1, t_2) + \mu_X(t_1, t_2) \left[\ln f(t_i) + (1 + \alpha\theta) \ln(t_i - \gamma + \beta) + \ln \left(\frac{h_i^Y(t_1, t_2)}{h_i^X(t_1, t_2)} \right) \right] = (1 + \alpha\theta)G_Z^w(t_1, t_2) + E(X \ln f(X)|t_1 < X < t_2),$$

where $G_Z^w(t_1, t_2) = E(X \ln(X - \gamma + \beta)|t_1 < X < t_2)$ holds if and only if X follows Pareto-II distribution with cdf $F(x) = 1 - [1 + (\frac{x-\gamma}{\beta})]^{-\alpha}$, $x > \gamma > 0$, $\alpha, \beta > 0$.

Proof: The “if part” is straightforward and hence omitted. To prove the “only if part”, let us assume that (3.14) holds. Then from (3.14) and (1.4) we obtain

$$(3.15) \quad \int_{t_1}^{t_2} x f(x) \ln \left(\frac{f(x)/\Delta F}{g(x)/\Delta G} \right) dx + \left[\ln f(t_i) + (1 + \alpha\theta) \ln(t_i - \gamma + \beta) + \ln \left(\frac{g(t_i)}{\Delta G} \right) - \ln \left(\frac{f(t_i)}{\Delta F} \right) \right] \int_{t_1}^{t_2} x f(x) dx = (1 + \alpha\theta) \int_{t_1}^{t_2} x f(x) \ln(x - \gamma + \beta) dx + \int_{t_1}^{t_2} x f(x) \ln f(x) dx.$$

Differentiating (3.15) with respect to $t_i, i = 1, 2$ and after some algebraic calculations, we get

$$g(t_i) = k(t_i - \gamma + \beta)^{-(1+\alpha\theta)}, \quad i = 1, 2 \text{ and } k > 0.$$

Hence the result follows. This completes the proof of the theorem. □

Here, below we present a characterization theorem for Weibull distribution.

Theorem 3.6. *Let X and Y be two absolutely continuous nonnegative random variables as mentioned in Theorem 3.4. Also, assume that they satisfy PHRM with proportionality constant $\theta > 0$. Then the following relationship of the form*

$$(3.16) \quad D_{KL}^w(X||Y; t_1, t_2) + \mu_X(t_1, t_2) \left[\lambda\theta t_i^p + \ln f(t_i) + (1 - p) \ln t_i + \ln \left(\frac{h_i^Y(t_1, t_2)}{h_i^X(t_1, t_2)} \right) \right] + \lambda\theta\mu_{X^{p+1}}(t_1, t_2) = (1 - p)G_{Z^*}^w(t_1, t_2) + E(X \ln f(X)|t_1 < X < t_2), \quad i = 1, 2,$$

where $\mu_{X^{p+1}}(t_1, t_2) = E(X^{p+1}|t_1 < X < t_2)$ and $G_{Z^*}^w(t_1, t_2) = E(X \ln(X - \alpha)|t_1 < X < t_2)$ holds if and only if X follows Weibull distribution with cdf $F(x) = 1 - \exp(-\lambda x^p)$, $x > 0, p > 0, \lambda > 0$.

Proof: The “if part” is straightforward. To prove the “only if part”, we first assume that (3.16) holds. Then using (1.4) in (3.16), we get

$$\begin{aligned}
 & \int_{t_1}^{t_2} x f(x) \ln \left(\frac{f(x)/\Delta F}{g(x)/\Delta G} \right) dx + \left[\lambda \theta t_i^p + \ln f(t_i) + (1-p) \ln t_i + \ln \left(\frac{g(t_i)}{\Delta G} \right) \right. \\
 & \left. - \ln \left(\frac{f(t_i)}{\Delta F} \right) \right] \int_{t_1}^{t_2} x f(x) dx = (1-p) \int_{t_1}^{t_2} x f(x) \ln(x-\alpha) dx \\
 (3.17) \qquad \qquad \qquad & + \int_{t_1}^{t_2} x f(x) \ln f(x) dx - \lambda \theta \int_{t_1}^{t_2} x^{p+1} f(x) dx.
 \end{aligned}$$

Differentiating (3.17) with respect to t_i , $i = 1, 2$, and after some algebraic calculations, we obtain

$$g(t_i) = c t_i^{(p-1)} e^{-\lambda \theta t_i^p}, \quad i = 1, 2 \text{ and } c > 0.$$

Hence the required result follows. This completes the proof. □

Remark 3.1. In particular, for $p = 1, 2$, the Theorem 3.6 provides characterization results of exponential distribution with cdf $F(x) = 1 - e^{-\lambda x}$, $x > 0$, $\lambda > 0$ and Rayleigh distribution with cdf $F(x) = 1 - e^{-\lambda x^2}$, $x > 0$, $\lambda > 0$, respectively.

Hereafter, we present results which characterize uniform and power distributions.

Theorem 3.7. *Let X and Y be two absolutely continuous nonnegative random variables as described in Theorem 3.4 and satisfying PRHRM with proportionality constant $\theta > 0$. Then the following relationship of the form*

$$\begin{aligned}
 & D_{KL}^w(X||Y; t_1, t_2) + \mu_X(t_1, t_2) \left[\ln f(t_i) + (1-\theta) \ln(t_i - \alpha) + \ln \left(\frac{h_i^Y(t_1, t_2)}{h_i^X(t_1, t_2)} \right) \right] \\
 (3.18) \qquad \qquad \qquad & = (1-\theta) G_{Z^*}^w(t_1, t_2) + E(X \ln f(X) | t_1 < X < t_2), \quad i = 1, 2,
 \end{aligned}$$

where $G_{Z^*}^w(t_1, t_2) = E(X \ln(X - \alpha) | t_1 < X < t_2)$ and $\alpha < t_1 < t_2 < \beta$ holds if and only if X follows uniform distribution in the interval (α, β) .

Proof: The “if part” is straightforward. To prove the “only if part”, assume that (3.18) holds for $i = 1$ and 2. Using (1.4), the above relation (3.18) further reduces to

$$\begin{aligned}
 & \int_{t_1}^{t_2} x f(x) \ln \left(\frac{f(x)/\Delta F}{g(x)/\Delta G} \right) dx + \left[\ln f(t_i) + (1-\theta) \ln(t_i - \alpha) + \ln \left(\frac{g(t_i)}{\Delta G} \right) \right. \\
 (3.19) \qquad \qquad \qquad & \left. - \ln \left(\frac{f(t_i)}{\Delta F} \right) \right] \int_{t_1}^{t_2} x f(x) dx = (1-\theta) \int_{t_1}^{t_2} x f(x) \ln(x-\alpha) dx \\
 & + \int_{t_1}^{t_2} x f(x) \ln f(x) dx,
 \end{aligned}$$

for $i = 1, 2$. Differentiating (3.19) with respect to $t_i, i = 1, 2$ and then simplifying further, we obtain

$$g(t_i) = c(t_i - \alpha)^{(\theta-1)}, \quad i = 1, 2 \text{ and } c > 0,$$

which gives the required result. This completes the proof. □

Theorem 3.8. *Let X and Y be two absolutely continuous nonnegative random variables as mentioned in Theorem 3.4 and satisfying PRHRM with proportionality constant $\theta > 0$. Then for $c > 0$, the following relationship of the form*

$$\begin{aligned} (3.20) \quad D_{KL}^w(X||Y; t_1, t_2) + \mu_X(t_1, t_2) & \left[\ln f(t_i) + (1 - c\theta) \ln t_i + \ln \left(\frac{h_i^Y(t_1, t_2)}{h_i^X(t_1, t_2)} \right) \right] \\ & = (1 - c\theta)G_X^w(t_1, t_2) + E(X \ln f(X)|t_1 < X < t_2), \quad i = 1, 2, \end{aligned}$$

where $G_X^w(t_1, t_2)$ is given by (2.5) holds if and only if X follows power distribution with cdf $F(x) = (\frac{x}{b})^c, 0 < x < b, c > 0$.

Proof: The “if part” is straightforward. To prove the “only if part”, let us assume that (3.20) holds. Using (1.4), (3.20) further reduces to

$$\begin{aligned} (3.21) \quad \int_{t_1}^{t_2} xf(x) \ln \left(\frac{f(x)/\Delta F}{g(x)/\Delta G} \right) dx + \left[\ln f(t_i) + (1 - c\theta) \ln t_i + \ln \left(\frac{g(t_i)}{\Delta G} \right) - \ln \left(\frac{f(t_i)}{\Delta F} \right) \right] \\ \times \int_{t_1}^{t_2} xf(x) dx = (1 - c\theta) \int_{t_1}^{t_2} xf(x) \ln x dx + \int_{t_1}^{t_2} xf(x) \ln f(x) dx. \end{aligned}$$

Differentiating (3.21) with respect to $t_i, i = 1, 2$, we get

$$g(t_i) = kt_i^{(c\theta-1)}, \quad i = 1, 2 \text{ and } k > 0,$$

which follows the required result. This completes the proof. □

4. MONOTONE TRANSFORMATIONS

In this section, we analysis the effect of the doubly truncated weighted KLD given by (1.4) under strictly monotone transformations. The following theorem is a generalization of the Theorem 4.13 of Yasaei Sekeh *et al.* (2013).

Theorem 4.1. *Let X and Y be two absolutely continuous nonnegative random variables with pdfs f and g , and cdfs F and G , respectively. Consider*

two bijective functions ϕ_1 and ϕ_2 , which are strictly monotone and differentiable. Then for all $0 \leq t_1 < t_2 < +\infty$, we have

$$D_{KL}^w(\phi_1(X)||\phi_2(Y); t_1, t_2) = \begin{cases} D_{KL}^{w,\phi_1}(X||\phi_1^{-1}(\phi_2(Y)); \phi_1^{-1}(t_1), \phi_1^{-1}(t_2)), \\ \quad \text{if } \phi_1 \text{ and } \phi_2 \text{ are strictly increasing,} \\ D_{KL}^{w,\phi_1}(X||\phi_1^{-1}(\phi_2(Y)); \phi_1^{-1}(t_2), \phi_1^{-1}(t_1)), \\ \quad \text{if } \phi_1 \text{ and } \phi_2 \text{ are strictly decreasing,} \end{cases}$$

where

$$(4.1) \quad D_{KL}^{w,\phi}(X||Y; t_1, t_2) = \int_{t_1}^{t_2} \phi(x) \frac{f(x)}{\Delta F} \ln \left(\frac{f(x)/\Delta F}{g(x)/\Delta G} \right) dx.$$

Proof: Assume that $\phi_1(x)$ and $\phi_2(x)$ are strictly increasing functions. Under this condition, the pdfs and cdfs of $\phi_1(X)$ and $\phi_2(Y)$ can be obtained as

$$(4.2) \quad f_{\phi_1}(x) = \frac{f(\phi_1^{-1}(x))}{\phi_1'(\phi_1^{-1}(x))} \quad \text{and} \quad F_{\phi_1}(x) = F(\phi_1^{-1}(x))$$

and

$$(4.3) \quad g_{\phi_2}(x) = \frac{f(\phi_2^{-1}(x))}{\phi_2'(\phi_2^{-1}(x))} \quad \text{and} \quad G_{\phi_2}(x) = G(\phi_2^{-1}(x)),$$

respectively. Moreover, the pdf and the cdf of $\phi_1^{-1}(\phi_2(X))$ are respectively given by

$$(4.4) \quad g_{\phi_1^{-1}(\phi_2)}(x) = \frac{g(\phi_2^{-1}(\phi_1(x)))\phi_1'(x)}{\phi_2'(\phi_2^{-1}(\phi_1(x)))} \quad \text{and} \quad G_{\phi_1^{-1}(\phi_2)}(x) = G(\phi_2^{-1}(\phi_1(x))).$$

Applying (4.2) and (4.3) in (1.4), we obtain

$$(4.5) \quad D_{KL}^w(\phi_1(X)||\phi_2(Y); t_1, t_2) = \int_{t_1}^{t_2} x \frac{f(\phi_1^{-1}(x))/\phi_1'(\phi_1^{-1}(x))}{\Delta F^{\phi_1}} \times \ln \left(\frac{f(\phi_1^{-1}(x))/(\phi_1'(\phi_1^{-1}(x))\Delta F^{\phi_1})}{g(\phi_2^{-1}(x))/(\phi_2'(\phi_2^{-1}(x))\Delta G^{\phi_2})} \right) dx,$$

where $\Delta F^{\phi_1} = F(\phi_1^{-1}(t_2)) - F(\phi_1^{-1}(t_1))$ and $\Delta G^{\phi_2} = G(\phi_2^{-1}(t_2)) - G(\phi_2^{-1}(t_1))$. Further, using the transformation $u = \phi_1^{-1}(x)$ in (4.5), we get

$$(4.6) \quad D_{KL}^w(\phi_1(X)||\phi_2(Y); t_1, t_2) = \int_{\phi_1^{-1}(t_1)}^{\phi_1^{-1}(t_2)} \phi_1(u) \frac{f(u)}{\Delta F^{\phi_1}} \times \ln \left(\frac{f(u)/(\phi_1'(u)\Delta F^{\phi_1})}{g(\phi_2^{-1}(\phi_1(u)))/(\phi_2'(\phi_2^{-1}(\phi_1(u)))\Delta G^{\phi_2})} \right) du.$$

Hence from (4.6), the first part of the theorem follows. The second part can be proved similarly and hence omitted. This completes the proof of the theorem. \square

Remark 4.1. Note that when $t_1 \rightarrow 0$ (for fixed t_2) and $t_2 \rightarrow \infty$ (for fixed t_1), Theorem 4.1 reduces to Theorem 4.13 of Yasaei Sekeh *et al.* (2013).

Remark 4.2. Consider $\phi_1(x) = F(x)$ and $\phi_2(x) = G(x)$. Here, both $F(x)$ and $G(x)$ are strictly increasing in their supports. Also, consider $\phi_1(x) = \bar{F}(x)$ and $\phi_2(x) = \bar{G}(x)$, which are strictly decreasing in supports. Clearly, $\phi_1(x)$ and $\phi_2(x)$ satisfy the assumptions of Theorem 4.1. Thus as an application of the Theorem 4.1, we get

$$(4.7) \quad D_{KL}^w(F(X)||G(Y); t_1, t_2) = D_{KL}^w(X||F^{-1}(G(Y)); F^{-1}(t_1), F^{-1}(t_2))$$

and

$$(4.8) \quad D_{KL}^w(\bar{F}(X)||\bar{G}(Y); t_1, t_2) = D_{KL}^w(X||\bar{F}^{-1}(\bar{G}(Y)); \bar{F}^{-1}(t_2), \bar{F}^{-1}(t_1)).$$

The following proposition is due to Theorem 4.1. It provides the effects of the doubly truncated weighted KLD under affine transformations.

Proposition 4.1. *Let X and Y be two nonnegative absolutely continuous random variables with pdfs f and g , and cdfs F and G , respectively. Define $\phi_1(X) = a_1X + b_1$ and $\phi_2(Y) = a_2Y + b_2$, where $a_1, a_2 > 0$ and $b_1, b_2 \geq 0$ are constants. Then for $t_1 > b_2$ and $b_2 \geq b_1$,*

$$(4.9) \quad D_{KL}^w(\phi_1(X)||\phi_2(Y); t_1, t_2) = D_{KL}^w\left(X||\frac{a_2}{a_1}Y + \frac{b_2 - b_1}{a_1}; \frac{t_1 - b_1}{a_1}, \frac{t_2 - b_1}{a_1}\right).$$

Remark 4.3. Under the assumptions as described in Proposition 4.1, the right hand side expression given by (4.9) can be written further as

$$(4.10) \quad \begin{aligned} D_{KL}^w(\phi_1(X)||\phi_2(Y); t_1, t_2) &= a_1 D_{KL}^w\left(X||\frac{a_2}{a_1}Y + \frac{b_2 - b_1}{a_1}; \frac{t_1 - b_1}{a_1}, \frac{t_2 - b_1}{a_1}\right) \\ &+ b_1 D_{KL}^w\left(X||\frac{a_2}{a_1}Y + \frac{b_2 - b_1}{a_1}; \frac{t_1 - b_1}{a_1}, \frac{t_2 - b_1}{a_1}\right). \end{aligned}$$

In particular, if we consider $\phi_1(x) = \phi_2(x) = \phi(x)$, then Theorem 4.1 reduces to the following result. We omit the proof as it follows from that of the Theorem 4.1.

Theorem 4.2. *Let X and Y be two absolutely continuous nonnegative random variables as described in Theorem 4.1. Assume that $\phi(x)$ is strictly monotone and differentiable function. Then for all $0 \leq t_1 < t_2 < +\infty$, we have*

$$D_{KL}^w(\phi(X)||\phi(Y); t_1, t_2) = \begin{cases} D_{KL}^w(X||Y; \phi^{-1}(t_1), \phi^{-1}(t_2)), & \text{if } \phi(x) \text{ is strictly increasing,} \\ D_{KL}^w(X||Y; \phi^{-1}(t_2), \phi^{-1}(t_1)), & \text{if } \phi(x) \text{ is strictly decreasing,} \end{cases}$$

where $D_{KL}^w(X||Y; t_1, t_2)$ is given by (4.1).

Proposition 4.2. *Let X and Y be two nonnegative absolutely continuous random variables with pdfs f and g , and cdfs F and G , respectively. Define $\phi_1(x) = \phi_2(x) = \phi(x) = ax + b$, where $a > 0$ and $b \geq 0$ are constants. Then for $t_1 > b$*

$$(4.11) \quad D_{KL}^w(\phi(X)||\phi(Y); t_1, t_2) = D_{KL}^{w,\phi}\left(X||Y; \frac{t_1 - b}{a}, \frac{t_2 - b}{a}\right).$$

5. INEQUALITIES AND BOUNDS

In this section, we obtain various inequalities and bounds for doubly truncated weighted KLD given by (1.4) in terms of other measures, which may be useful in mathematical statistics, ergodic theory and other scientific fields. Most of these depend on the measures (2.1)–(2.3) and (2.5)–(2.7). Several authors studied these measures and obtained various results. For some results on these measures, we refer to Navarro and Ruiz (1996), Misagh and Yari (2011), Sankaran and Sunoj (2004) and Kundu (2017).

Proposition 5.1. *Let X and Y be two nonnegative absolutely continuous random variables with pdfs f and g , and cdfs F and G , respectively. Suppose that $D_{KL}^w(X||Y; t_1, t_2)$ is increasing (decreasing) in t_1 (for fixed t_2) and t_2 (for fixed t_1). Then*

$$D_{KL}^w(X||Y; t_1, t_2) \geq (\leq) \left(\frac{h_1^Y(t_1, t_2)}{h_1^X(t_1, t_2)} - 1\right)\mu_X(t_1, t_2) - t_1 \ln\left(\frac{h_1^Y(t_1, t_2)}{h_1^X(t_1, t_2)}\right)$$

and

$$D_{KL}^w(X||Y; t_1, t_2) \leq (\geq) \left(\frac{h_2^Y(t_1, t_2)}{h_2^X(t_1, t_2)} - 1\right)\mu_X(t_1, t_2) - t_2 \ln\left(\frac{h_2^Y(t_1, t_2)}{h_2^X(t_1, t_2)}\right).$$

Proof: Under the given condition, the required inequalities follow from (3.1) and (3.2), and hence, omitted. \square

The following proposition provides bounds of (1.4) in terms of GCM, GFR functions and doubly truncated weighted inaccuracy measure.

Proposition 5.2. *Let X and Y be two random variables as described in Proposition 5.1. If $f(x)$ is increasing (decreasing) in $x > 0$, then*

$$D_{KL}^w(X||Y; t_1, t_2) \geq (\leq) \mu_X(t_1, t_2) \ln h_1^X(t_1, t_2) + I^w(X||Y; t_1, t_2)$$

and

$$D_{KL}^w(X||Y; t_1, t_2) \leq (\geq) \mu_X(t_1, t_2) \ln h_2^X(t_1, t_2) + I^w(X||Y; t_1, t_2).$$

Proof: Let $f(x)$ be increasing (decreasing) in $x > 0$. Then for $t_1 < x < t_2$, we have

$$(5.1) \quad \frac{f(t_1)}{\Delta F} \leq (\geq) \frac{f(x)}{\Delta F} \leq (\geq) \frac{f(t_2)}{\Delta F}.$$

Moreover, $g(x)/\Delta G$ is positive. Thus from (5.1) we get

$$(5.2) \quad \ln \left(\frac{f(t_1)}{\Delta F} / \frac{g(x)}{\Delta G} \right) \leq (\geq) \ln \left(\frac{f(x)}{\Delta F} / \frac{g(x)}{\Delta G} \right) \leq (\geq) \ln \left(\frac{f(t_2)}{\Delta F} / \frac{g(x)}{\Delta G} \right).$$

Multiplying (5.2) by $x \frac{f(x)}{\Delta F}$ and then integrating from t_1 to t_2 with respect to x , the required inequalities follow. \square

Proposition 5.3. *Let X and Y be two random variables as described in Proposition 5.1. If $g(x)$ is increasing (decreasing) in $x > 0$, then*

$$\begin{aligned} &D_{KL}^w(X||Y; t_1, t_2) \geq (\leq) -\mu_X(t_1, t_2) \ln h_2^Y(t_1, t_2) - S_X^w(t_1, t_2) \\ \text{and} \quad &D_{KL}^w(X||Y; t_1, t_2) \leq (\geq) -\mu_X(t_1, t_2) \ln h_1^Y(t_1, t_2) - S_X^w(t_1, t_2). \end{aligned}$$

Proof: Proof follows analogous to that of the Proposition 5.2. Hence omitted. \square

Below, in Proposition 5.4, we give new inequalities for $D_{KL}^w(X||Y; t_1, t_2)$ involving a pair of likelihood ratio ordered random variables.

Proposition 5.4. *Let X and Y be two random variables as described in Proposition 5.1. If $X \geq^{lr} Y$, then*

$$\mu_X(t_1, t_2) \ln \left(\frac{h_1^X(t_1, t_2)}{h_1^Y(t_1, t_2)} \right) \leq D_{KL}^w(X||Y; t_1, t_2) \leq \mu_X(t_1, t_2) \ln \left(\frac{h_2^X(t_1, t_2)}{h_2^Y(t_1, t_2)} \right).$$

Proof: Under the given condition, $f(x)/g(x)$ is increasing in $x > 0$. Then for $t_1 < x < t_2$, we have $\frac{f(t_1)}{g(t_1)} \leq \frac{f(x)}{g(x)} \leq \frac{f(t_2)}{g(t_2)}$. Thus from (1.4) we obtain

$$(5.3) \quad \begin{aligned} D_{KL}^w(X||Y; t_1, t_2) &\leq \int_{t_1}^{t_2} x \frac{f(x)}{\Delta F} \ln \left(\frac{f(t_2)/\Delta F}{g(t_2)/\Delta G} \right) dx \\ &= \mu_X(t_1, t_2) \ln \left(\frac{h_2^X(t_1, t_2)}{h_2^Y(t_1, t_2)} \right) \end{aligned}$$

and

$$(5.4) \quad \begin{aligned} D_{KL}^w(X||Y; t_1, t_2) &\geq \int_{t_1}^{t_2} x \frac{f(x)}{\Delta F} \ln \left(\frac{f(t_1)/\Delta F}{g(t_1)/\Delta G} \right) dx \\ &= \mu_X(t_1, t_2) \ln \left(\frac{h_1^X(t_1, t_2)}{h_1^Y(t_1, t_2)} \right). \end{aligned}$$

Combining (5.3) and (5.4), we obtain the required inequalities. \square

Proposition 5.5. *Let X and Y be two random variables as described in Proposition 5.1. Then*

$$D_{KL}^w(X||Y; t_1, t_2) \geq \mu_X(t_1, t_2) \ln \left(\frac{\mu_X(t_1, t_2)}{\mu_Y(t_1, t_2)} \right).$$

Proof: The result follows from the log-sum inequality and hence omitted. \square

Proposition 5.6. *Let X and Y be two random variables as described in Proposition 5.1. Then*

$$(5.5) \quad D_{KL}^w(X||Y; t_1, t_2) \leq E \left(X \frac{f(X)/\Delta F}{g(X)/\Delta G} \middle| t_1 < X < t_2 \right) - \mu_X(t_1, t_2).$$

Proof: The proof follows from the inequality $\ln x \leq x - 1$, for all $x > 0$. Hence it is omitted. \square

Hereafter, we consider three nonnegative random variables X_1 , X_2 and X_3 and obtain bounds of $D_{KL}^w(X_1||X_3; t_1, t_2)$, $D_{KL}^w(X_1||X_2; t_1, t_2)$ and $D_{KL}^w(X_2||X_3; t_1, t_2)$.

Proposition 5.7. *Let X_1 , X_2 and X_3 be three nonnegative absolutely continuous random variables with pdf's $f_1(x)$, $f_2(x)$ and $f_3(x)$, respectively. The corresponding cdf's are $F_1(x)$, $F_2(x)$ and $F_3(x)$. If $X_1 \geq^{lr} X_2$, then*

$$D_{KL}^w(X_1||X_3; t_1, t_2) \geq \mu_{X_1}(t_1, t_2) \ln \left(\frac{h_1^{X_1}(t_1, t_2)}{h_1^{X_2}(t_1, t_2)} \right) + ID^w(t_1, t_2)$$

and

$$D_{KL}^w(X_1||X_3; t_1, t_2) \leq \mu_{X_1}(t_1, t_2) \ln \left(\frac{h_2^{X_1}(t_1, t_2)}{h_2^{X_2}(t_1, t_2)} \right) + ID^w(t_1, t_2),$$

where $ID^w(t_1, t_2) = I^w(X_1||X_3; t_1, t_2) - I^w(X_1||X_2; t_1, t_2)$.

Proof: Using $X_1 \geq^{lr} X_2$ and $t_1 < x$, we obtain $f_1(x) \geq f_2(x)f_1(t_1)/f_2(t_1)$. Thus, from (1.4)

$$(5.6) \quad \begin{aligned} D_{KL}^w(X_1||X_3; t_1, t_2) &= \int_{t_1}^{t_2} x \frac{f_1(x)}{\Delta F_1} \ln \left(\frac{f_1(x)/\Delta F_1}{f_3(x)/\Delta F_3} \right) dx \\ &\geq \int_{t_1}^{t_2} x \frac{f_1(x)}{\Delta F_1} \ln \left(\frac{(f_2(x)f_1(t_1))/(\Delta F_2\Delta F_1)}{(f_3(x)f_2(t_1))/(\Delta F_3\Delta F_2)} \right) dx \\ &= \mu_{X_1}(t_1, t_2) \ln \left(\frac{h_1^{X_1}(t_1, t_2)}{h_1^{X_2}(t_1, t_2)} \right) + ID^w(t_1, t_2). \end{aligned}$$

The upper bound can be obtained similarly. This completes the proof. \square

Proposition 5.8. *Let X_1, X_2 and X_3 be three random variables as described in Proposition 5.7. If $X_2 \geq^{lr} X_3$, then*

$$D_{KL}^w(X_1||X_2; t_1, t_2) \geq \mu_{X_1}(t_1, t_2) \ln \left(\frac{h_2^{X_3}(t_1, t_2)}{h_2^{X_2}(t_1, t_2)} \right) + IS_1^w(t_1, t_2)$$

and

$$D_{KL}^w(X_1||X_2; t_1, t_2) \leq \mu_{X_1}(t_1, t_2) \ln \left(\frac{h_1^{X_3}(t_1, t_2)}{h_1^{X_2}(t_1, t_2)} \right) + IS_1^w(t_1, t_2),$$

where $IS_1^w(t_1, t_2) = I^w(X_1||X_3; t_1, t_2) - S^w(X_1; t_1, t_2)$.

Proof: Under $X_2 \geq^{lr} X_3$, for $x < t_2$, we have $f_2(x) \leq f_3(x)f_2(t_2)/f_3(t_2)$. Thus applying this inequality in (1.4) and after some simplifications, lower bound can be obtained. The upper bound can be obtained similarly. This completes the proof. \square

Proposition 5.9. *Let X_1, X_2 and X_3 be three random variables as described in Proposition 5.7. If $X_1 \geq^{lr} X_3$, then*

$$D_{KL}^w(X_2||X_3; t_1, t_2) \geq \mu_{X_2}(t_1, t_2) \ln \left(\frac{h_1^{X_1}(t_1, t_2)}{h_1^{X_3}(t_1, t_2)} \right) + IS_2^w(t_1, t_2)$$

and

$$D_{KL}^w(X_2||X_3; t_1, t_2) \leq \mu_{X_2}(t_1, t_2) \ln \left(\frac{h_2^{X_1}(t_1, t_2)}{h_2^{X_3}(t_1, t_2)} \right) + IS_2^w(t_1, t_2),$$

where $IS_2^w(t_1, t_2) = I^w(X_2||X_1; t_1, t_2) - S^w(X_2; t_1, t_2)$.

Proof: It is given that $X_1 \geq^{lr} X_3$. Therefore, for $x > t_1$, we have $f_3(x) \leq f_1(x)f_3(t_1)/f_1(t_1)$. Applying this in (1.4), we get the lower bound of $D_{KL}^w(X_2||X_3; t_1, t_2)$. The upper bound can be obtained similarly. This completes the proof. \square

6. NUMERICAL EXAMPLES

In this section, we consider examples for the verification of few of the results obtained in Section 5. To verify the Proposition 5.2, we consider the following example and present numerical values of the lower and upper bounds of the doubly truncated weighted KLD.

Example 6.1. Suppose that X follows power distribution and Y follows U-quadratic distribution in the interval $(0, 1)$ with pdfs $f(x) = cx^{c-1}$, $c > 0$, $0 < x < 1$ and $g(x) = 12(x - \frac{1}{2})^2$, $0 < x < 1$, respectively. Here, $f(x)$ is decreasing in x for $c < 1$ and increasing in x for $c > 1$. In Table 1 and Table 2, we present numerical values of the lower bounds (LB) and upper bounds (UB) of the doubly truncated weighted KLD for different values of t_1 and t_2 for $c = 0.5$ and 1.5 , respectively.

Table 1: Bounds of $D_{KL}^w(X||Y; t_1, t_2)$ for $c = 0.5$.

(t_1, t_2)	LB	$D_{KL}^w(X Y; t_1, t_2)$	UB	(t_1, t_2)	LB	$D_{KL}^w(X Y; t_1, t_2)$	UB
(0.1,0.4)	0.022421	0.077094	0.184155	(0.4,0.5)	0.405266	0.428923	0.455370
(0.1,0.5)	0.252138	0.321949	0.473062	(0.4,0.7)	0.467240	0.532709	0.619190
(0.1,0.9)	0.233992	0.360127	0.710058	(0.4,0.9)	0.546199	0.649186	0.802994
(0.2,0.5)	0.309069	0.369426	0.464263	(0.5,0.6)	0.445497	0.469392	0.495566
(0.2,0.7)	0.356237	0.449858	0.622275	(0.5,0.7)	0.439779	0.485807	0.540251
(0.2,0.8)	0.296777	0.406123	0.620246	(0.5,0.9)	0.426941	0.513659	0.629808

Table 2: Bounds of $D_{KL}^w(X||Y; t_1, t_2)$ for $c = 1.5$.

(t_1, t_2)	LB	$D_{KL}^w(X Y; t_1, t_2)$	UB	(t_1, t_2)	LB	$D_{KL}^w(X Y; t_1, t_2)$	UB
(0.1,0.4)	0.048524	0.185502	0.232704	(0.4,0.5)	0.455614	0.483096	0.505924
(0.1,0.5)	0.434464	0.634879	0.694851	(0.4,0.7)	0.362879	0.458389	0.518698
(0.1,0.9)	0.014652	0.519028	0.628185	(0.4,0.9)	0.280465	0.458343	0.550633
(0.2,0.5)	0.458314	0.570112	0.623685	(0.5,0.6)	0.395761	0.422765	0.445968
(0.2,0.7)	0.404517	0.620202	0.701468	(0.5,0.7)	0.340430	0.398162	0.441841
(0.2,0.8)	0.213175	0.487130	0.581533	(0.5,0.9)	0.227166	0.356106	0.435718

To illustrate Proposition 5.3, we consider the following example.

Example 6.2. Let X and Y be two random variables with pdfs $f(x) = 1, 0 < x < 1$ and $g(x) = b(1 - x)^{b-1}, 0 < x < 1, b > 0$, respectively. Note that $g(x)$ is increasing in x for $b < 1$ and decreasing in x for $b > 1$. In Table 3 and 4, we present the numerical values of the bounds of the doubly truncated weighted KLD for $b = 0.2$ and $b = 1.2$, respectively.

Table 3: Bounds of $D_{KL}^w(X||Y; t_1, t_2)$ for $b = 0.2$.

(t_1, t_2)	LB	$D_{KL}^w(X Y; t_1, t_2)$	UB	(t_1, t_2)	LB	$D_{KL}^w(X Y; t_1, t_2)$	UB
(0.1,0.4)	-0.042188	-0.006974	0.038905	(0.4,0.5)	-0.033416	-0.000816	0.032219
(0.1,0.5)	-0.074668	-0.012759	0.066401	(0.4,0.7)	-0.163019	-0.006694	0.141966
(0.1,0.9)	-0.532245	-0.044619	0.346645	(0.4,0.9)	-0.547117	-0.003054	0.384598
(0.2,0.5)	-0.068887	-0.007283	0.062713	(0.5,0.6)	-0.050187	-0.000755	0.047997
(0.2,0.7)	-0.193724	-0.020447	0.159375	(0.5,0.7)	-0.128847	-0.002608	0.116350
(0.2,0.8)	-0.315074	-0.027845	0.239444	(0.5,0.9)	-0.521607	0.005471	0.379679

Table 4: Bounds of $D_{KL}^w(X||Y; t_1, t_2)$ for $b = 1.2$.

(t_1, t_2)	LB	$D_{KL}^w(X Y; t_1, t_2)$	UB	(t_1, t_2)	LB	$D_{KL}^w(X Y; t_1, t_2)$	UB
(0.1,0.4)	-0.009386	0.002084	0.010888	(0.4,0.5)	-0.007930	0.000329	0.008478
(0.1,0.5)	-0.015747	0.004043	0.019521	(0.4,0.7)	-0.033325	0.003839	0.042921
(0.1,0.9)	-0.069415	0.028401	0.150307	(0.4,0.9)	-0.080490	0.016423	0.152438
(0.2,0.5)	-0.015039	0.002460	0.017861	(0.5,0.6)	-0.011771	0.000417	0.012774
(0.2,0.7)	-0.036351	0.008604	0.051923	(0.5,0.7)	-0.027794	0.001945	0.033504
(0.2,0.8)	-0.052349	0.014472	0.086279	(0.5,0.9)	-0.081046	0.012505	0.144275

The following example shows a case in which Proposition 5.4 is fulfilled.

Example 6.3. Consider two nonnegative random variables X and Y with pdfs $f(x) = \frac{2}{3}(1+x)$, $0 < x < 1$ and $g(x) = \frac{2}{3}(2-x)$, $0 < x < 1$, respectively. By straightforward calculations, it is not hard to verify that $X \geq^{lr} Y$. In Table 5, we present the numerical values of the lower and upper bounds of the doubly truncated weighted KLD between X and Y .

Table 5: Bounds of $D_{KL}^w(X||Y; t_1, t_2)$.

(t_1, t_2)	LB	$D_{KL}^w(X Y; t_1, t_2)$	UB	(t_1, t_2)	LB	$D_{KL}^w(X Y; t_1, t_2)$	UB
(0.1,0.3)	-0.028607	0.005274	0.278213	(0.5,0.7)	-0.080397	0.006264	0.081120
(0.1,0.6)	-0.126393	0.034815	0.122128	(0.5,0.8)	-0.131348	0.014560	0.134047
(0.1,0.8)	-0.229447	0.071593	0.225767	(0.5,0.9)	-0.189889	0.026835	0.196978
(0.3,0.4)	-0.023699	0.001387	0.023540	(0.6,0.7)	-0.043674	0.001614	0.043969
(0.3,0.7)	-0.136517	0.023788	0.136517	(0.6,0.8)	-0.094577	0.006683	0.096309
(0.3,0.9)	-0.248611	0.056871	0.255551	(0.6,0.9)	-0.153075	0.015624	0.158454

In this part of the paper, we provide an example in support of the Proposition 5.7.

Example 6.4. Let X and Y be two nonnegative random variables as described in Example 6.3. Consider another random variable Z with pdf $f_3(x) = \frac{1}{2\sqrt{1-x}}$, $0 < x < 1$. Here, $X \geq^{lr} Y$. The lower and upper bounds of $D_{KL}^w(X||Y; t_1, t_2)$ obtained in the Proposition 5.7 are presented in Table 6.

Table 6: Bounds of $D_{KL}^w(X||Y; t_1, t_2)$.

(t_1, t_2)	LB	$D_{KL}^w(X Y; t_1, t_2)$	UB	(t_1, t_2)	LB	$D_{KL}^w(X Y; t_1, t_2)$	UB
(0.1,0.2)	-0.011780	0.000238	0.009483	(0.4,0.6)	-0.074033	-0.001039	0.000091
(0.1,0.5)	-0.107631	0.000543	0.061938	(0.4,0.7)	-0.127975	-0.003167	0.094956
(0.1,0.7)	-0.230205	-0.004876	0.113178	(0.4,0.9)	-0.274740	-0.012652	0.175896
(0.2,0.4)	-0.046788	0.000172	0.035488	(0.6,0.7)	-0.045791	-0.000503	0.041852
(0.2,0.5)	-0.085598	-0.000282	0.058566	(0.6,0.8)	-0.103484	-0.002223	0.087403
(0.2,0.8)	-0.275617	-0.011008	0.146067	(0.6,0.9)	-0.172764	-0.004064	0.138765

Real Data: We consider two real data sets, which represent the failure times of the air conditioning system of two different air planes (see Bain and Engelhardt, 1991, p. 101). The data sets are given below:

Data Set I (Plane 7912): 1, 3, 5, 7, 11, 11, 11, 12, 14, 14, 14, 16, 16, 20, 21, 23, 42, 47, 52, 62, 71, 71, 87, 90, 95, 120, 120, 225, 246, 261.

Data Set II (Plane 7911): 33, 47, 55, 56, 104, 176, 182, 220, 239, 246, 320.

The above data sets, Data Set I and Data Set II can be fitted as exponential distributions with parameters (hazard rates) λ_1 and λ_2 , respectively. We assume that due to some reasons, the data in the interval $[50, 200]$ are observed. Based on this assumption, the unknown parameters can be estimated. For this purpose, we use the method of maximum likelihood. Here, the estimated values of the parameters are $\hat{\lambda}_1 = 0.026029$ and $\hat{\lambda}_2 = 0.005611$. From (1.4), we get $\hat{D}_{KL}^w(X||Y; 50, 200) = 2.5069$.

7. CONCLUDING REMARKS

In this paper, we consider a generalized discrimination measure, which is known as the doubly truncated weighted KLD. We obtain few characterization results based on the proposed measure. These results may be useful in studying various characteristics of a system when its lifetimes fall in an interval. Further, the effect of the affine transformations on the proposed discrimination measure is studied. Several inequalities and bounds are obtained. Finally, few applications with bounds in support of the results are presented.

ACKNOWLEDGMENTS

The authors wish to express their sincere thanks to an anonymous reviewer for some useful comments which have improved the initial version of this manuscript.

REFERENCES

- [1] BAIN, L.J. and ENGELHARDT, M. (1991). *Statistical Analysis of Reliability and Life-Testing Models*, 2nd. Edition, Marcel and Dekker, New York.
- [2] BELIS, M. and GUIASU, S. (1968). A quantitative-qualitative measure of information in cybernetic systems, *IEEE Transactions on Information Theory*, **4**, 593–594.
- [3] COX, D.R. (1972). Regression models and life tables (with Discussion), *Journal of the Royal Statistical Society, Series B*, **34**, 187–220.
- [4] DI CRESCENZO, A. and LONGOBARDI, M. (2006). On weighted residual and past entropies, *Scientiae Mathematicae Japonicae*, **64**, 255–266.

- [5] EBRAHIMI, N. and KIRMANI, S.N.U.A. (1996). A characterization of the proportional hazards model through a measure of discrimination between two residual life distributions, *Biometrika*, **83**, 233–235.
- [6] GUPTA, R.C. and GUPTA, R.D. (2007). Proportional reversed hazard rate model and its applications, *Journal of Statistical Planning and Inference*, **137**, 3525–3536.
- [7] KASZA, J. and SOLOMON, P. (2015). Comparing score-based methods for estimating Bayesian networks using Kullback–Leibler divergence, *Communications in Statistics – Theory and Methods*, **44**, 135–152.
- [8] KULLBACK, S. and LEIBLER, R.A. (1951). On information and sufficiency, *Annals of Mathematical Statistics*, **22**, 79–86.
- [9] KUNDU, C. (2017). On weighted measure of inaccuracy for doubly truncated random variables, *Communications in Statistics – Theory and Methods*, **46**, 3135–3147.
- [10] MIRALI, M.; BARATPOUR, S. and FAKOOR, V. (2017). On weighted cumulative residual entropy, *Communications in Statistics – Theory and Methods*, **46**, 2857–2869.
- [11] MISAGH, F. and YARI, G. (2011). On weighted interval entropy, *Statistics and Probability Letters*, **81**, 188–194.
- [12] MISAGH, F. and YARI, G. (2012). Interval entropy and informative distance, *Entropy*, **14**, 480–490.
- [13] NAIR, N.U. and GUPTA, R.P. (2007). Characterization of proportional hazard models by properties of information measures, *International Journal of Statistical Sciences*, **6**, 223–231.
- [14] NAIR, K.R.M. and RAJESH, G. (2000). Geometric vitality function and its applications to reliability, *IAPQR Transactions*, **25**, 1–8.
- [15] NAVARRO, J. and RUIZ, J.M. (1996). Failure rate functions for doubly truncated random variables, *IEEE Transactions on Reliability*, **45**, 685–690.
- [16] NOURBAKHSH, M.; YARI, G. and MEHRALI, Y. (2016). Weighted entropies and their estimations, *Communications in Statistics – Simulation and Computation*, <http://www.tandfonline.com/doi/abs/10.1080/03610918.2016.1140776>.
- [17] RAJESH, G.; ABDUL-SATHAR, E.I. and ROHINI, N.S. (2017). On dynamic weighted survival entropy of order α , *Communications in Statistics – Theory and Methods*, **46**, 2139–2150.
- [18] RUIZ, J.M. and NAVARRO, J. (1996). Characterizations based on conditional expectations of the doubles truncated distribution, *Annals of the Institute of Statistical Mathematics*, **48**, 563–572.
- [19] SANKARAN, P.G. and GLEEJA, V.L. (2008). Proportional reversed hazard and frailty models, *Metrika*, **68**, 333–342.
- [20] SANKARAN, P.G. and SUNOJ, S.M. (2004). Identification of models using failure rate and mean residual life of doubly truncated random variables, *Statistical Papers*, **45**, 97–109.
- [21] SANKARAN, P.G.; SUNOJ, S.M. and NAIR, N.U. (2016). Kullback–Leibler divergence: A quantile approach, *Statistics and Probability Letters*, **111**, 72–79.

- [22] SHAKED, M. and SHANTHIKUMAR, J.G. (2007). *Stochastic Orders*, Springer Verlag, New York.
- [23] SUHOV, Y. and YASAEI SEKEH, S. (2015). Weighted cumulative entropies: an extension of CRE and CE, <http://arxiv.org/abs/1507.07051v1>.
- [24] SUNOJ, S.M.; SANKARAN, P.G. and MAYA, S.S. (2009), Characterizations of life distributions using conditional expectations of doubly (interval) truncated random variables, *Communications in Statistics – Theory and Methods*, **38**, 1441–1452.
- [25] YASAEI SEKEH, S.; BORZADARAN, G.R.M. and ROKNABADI, A.H.R. (2013). On Kullback–Leibler dynamic information, <http://dx.doi.org/10.2139/ssrn.2344078>.

THE BETA MARSHALL–OLKIN LOMAX DISTRIBUTION

Authors: CLAUDIO J. TABLADA

- Graduate Program in Statistics, Universidade Federal de Pernambuco,
Recife, PE, Brazil
Department of Statistics, Universidade Federal da Paraíba,
João Pessoa, PB, Brazil
claudio@de.ufpb.br

GAUSS M. CORDEIRO

- Department of Statistics, Universidade Federal de Pernambuco,
Recife, PE, Brazil
gauss@de.ufpe.br

Received: September 2016

Revised: April 2017

Accepted: April 2017

Abstract:

- Compounding distributions is the most common method in lifetime analysis to obtain more flexible families of distributions. Based on the beta Marshall–Olkin generated family, we present a new four-parameter distribution, so-called the beta Marshall–Olkin Lomax, for lifetime applications. We obtain some of its properties from those of well-established distributions. We provide a simulation study to illustrate the performance of the maximum likelihood estimates. An application to uncensored data is carried out and we use some goodness-of-fit statistics to study the flexibility of the new distribution, proving empirically that this model can be appropriate for lifetime applications.

Key-Words:

- *beta-G family; generalized distributions; lifetime analysis; Lomax distribution; Marshall–Olkin extended family.*

AMS Subject Classification:

- 62E15, 62N02, 62N99, 65C05.

1. INTRODUCTION

In applications involving lifetime models such as survival analysis, demography, reliability, actuarial study and others, the distributions with positive real supports play a fundamental role. Because of this, in recent years, there is a growing interest in constructing new distributions to model aging phenomena [15, 14]. The method that has received most attention by researchers to generate new models is that one by compounding existing distributions, usually referred to as generalized G families of distributions [28]. The principal reason for this is the ability of these generalized distributions to be more flexible than the baseline G distribution to provide better fits to skewed data and good control of the tails [23]. The second reason is the powerful computational and analytical facilities available in several software packages, which facilitate handling and computing complex mathematical expressions. Some of the generalized G families best known are: the Marshall–Olkin extended (MOE) family [18], the exponentiated-generated (exp- G) family [13, 8], the beta-generated (beta- G) family [9], the Kumaraswamy-generated (Kw- G) family [7], the gamma-generated (gamma- G) families [29, 25, 22] and the McDonald-generated (Mc- G) family [2]. A detailed compilation of these families is given in [28].

In this paper, we adopt the beta Marshall–Olkin generated (BMO- G) family proposed by Alizadeh *et al.* [3] to define the new *beta Marshall–Olkin Lomax* (BMOL) distribution obtained by taking the Lomax distribution [17] as the baseline G model. Given that the proposed distribution has positive real support, our objective is to define a wide flexible distribution for real lifetime applications.

The paper unfolds as follows. In Section 2, we describe some preliminaries and introduce the BMOL distribution. In Section 3, we plot its density and hazard rate functions for some parameter values. In Section 4, we obtain an expansion for the BMOL density function as a linear combination of exp-Lomax and Lomax densities. In Sections 5–10, we present explicit expressions for the quantile function (qf), moments, generating function, mean deviations, Bonferroni and Lorenz curves, Shannon entropy and order statistics. Section 11 is devoted to the maximum likelihood estimates (MLEs) for complete samples and, in Section 12, we carry out a simulation study to study the performance of these estimates. In Section 13, we consider an application of the BMOL distribution and compare it with others related distributions and with the exponentiated Weibull (EW) distribution [20] based on some goodness-of-fit statistics. Finally, Section 14 concludes the paper.

2. THE NEW DISTRIBUTION

Marshall & Olkin [18] pioneered a method of introducing an additional parameter to a distribution. If $G(x; \boldsymbol{\xi})$ is a baseline distribution with parameter vector $\boldsymbol{\xi}$, then the cumulative distribution function (cdf) given by

$$(2.1) \quad F(x; c, \boldsymbol{\xi}) = \frac{G(x; \boldsymbol{\xi})}{c + (1 - c)G(x; \boldsymbol{\xi})}, \quad c > 0,$$

defines a new distribution with an extra shape parameter c . As commented by Marshall & Olkin [18], “By various methods, new parameters can be introduced to expand families of distributions for added flexibility or to construct covariate models”.

The cdf of the beta-G family (for $a, b > 0$) is defined by

$$(2.2) \quad F(x; a, b, \boldsymbol{\xi}) = \frac{B(G(x; \boldsymbol{\xi}); a, b)}{B(a, b)} = \frac{1}{B(a, b)} \int_0^{G(x; \boldsymbol{\xi})} w^{a-1} (1 - w)^{b-1} dw,$$

where $B(a, b) = \int_0^1 w^{a-1} (1 - w)^{b-1} dw$ is the beta function and $B(z; a, b) = \int_0^z w^{a-1} (1 - w)^{b-1} dw$ is the incomplete beta function. In this case, the generated distribution $F(x; a, b, \boldsymbol{\xi})$ has two extra shape parameters a and b . The beta G family was introduced by Eugene *et al.* [9], who studied the properties of the beta-normal distribution. If the baseline $G(x; \boldsymbol{\xi})$ in (2.2) is the Lomax distribution, we obtain the beta-Lomax (BL) distribution as defined in [24].

A generalization of these concepts, introduced in [4], follows by considering the $T-X$ method. Let $R(x; \boldsymbol{\gamma})$ be a cdf with support $[d, e]$ and density $r(x; \boldsymbol{\gamma})$. For a given baseline distribution $G(x; \boldsymbol{\xi})$, let $W(\cdot)$ be a function satisfying the following properties

$$\begin{cases} W[G(x; \boldsymbol{\xi})] \in [d, e], \\ W[G(x; \boldsymbol{\xi})] \text{ is differentiable and monotonically non-decreasing,} \\ \lim_{x \rightarrow -\infty} W[G(x; \boldsymbol{\xi})] = d, \quad \lim_{x \rightarrow \infty} W[G(x; \boldsymbol{\xi})] = e. \end{cases}$$

Then, the cdf

$$(2.3) \quad F(x; \boldsymbol{\delta}, \boldsymbol{\gamma}, \boldsymbol{\xi}) = \int_d^{W[G(x; \boldsymbol{\xi})]} r(t; \boldsymbol{\gamma}) dt$$

defines a new distribution, where the link function $W(\cdot) = W(\cdot; \boldsymbol{\delta})$ possibly depends on a parameter vector $\boldsymbol{\delta}$. We say that the distribution $R(x; \boldsymbol{\gamma})$ is ‘transformed’ by the ‘transformer’ $W[G(x; \boldsymbol{\xi})]$.

Following this idea, Alizadeh *et al.* [3] introduced the BMO-G family by considering in (2.3) the function $W(z) = z/[c + (1 - c)z]$, $c > 0$, and the beta

distribution as the ‘transformed’ distribution $R(x; \gamma)$. Notice that, in this case, the ‘transformer’ $W[G(x; \xi)]$ is given by (2.1).

In this paper, we study the BMOL distribution by considering the baseline $G(x; \xi)$ in (2.3) as the Lomax distribution [17], which has cdf given by

$$(2.4) \quad G(x; \alpha, \lambda) = 1 - \left(1 + \frac{x}{\lambda}\right)^{-\alpha}, \quad x \geq 0, \alpha > 0, \lambda > 0$$

and probability density function (pdf)

$$(2.5) \quad g(x; \alpha, \lambda) = \frac{\alpha}{\lambda} \left(1 + \frac{x}{\lambda}\right)^{-(\alpha+1)}.$$

For the sake of simplicity, we will write sometimes the Lomax distribution with cdf $G(x)$ and pdf $g(x)$, respectively, without explicit mention to the parameters α and λ .

It is clear that a generalized G distribution has more parameters than the baseline G distribution. Generally, the use of four parameters should be sufficient for most practical purposes. In addition, notice that if $X \sim \text{Lomax}(\alpha, \lambda)$, then $X/\lambda \sim \text{Lomax}(\alpha, 1)$ and, consequently, λ is just a scale parameter. Henceforth, we consider the BMOL distribution with only four parameters by taking, without loss of generality, $\lambda = 1$ in equations (2.4) and (2.5). Thus, if $\theta = (a, b, c, \alpha)^\top$ is the parameter vector, we define the BMOL cdf by

$$(2.6) \quad F(x; \theta) = \frac{B(W[G(x)]; a, b)}{B(a, b)} = \frac{1}{B(a, b)} \int_0^{W[G(x)]} w^{a-1} (1-w)^{b-1} dw,$$

where $W[G(x)]$ is given by (2.1). From equations (2.1) and (2.4) (with $\lambda = 1$), we have

$$(2.7) \quad W[G(x)] = \frac{(1+x)^\alpha - 1}{(1+x)^\alpha + c - 1}.$$

The BMOL pdf follows from (2.6) as

$$(2.8) \quad f(x; \theta) = \frac{1}{B(a, b)} g(x) w[G(x)] \{W[G(x)]\}^{a-1} \{1 - W[G(x)]\}^{b-1},$$

where $w(z) = W'(z) = c/[c + (1-c)z]^2$. Thus, we obtain the BMOL pdf from (2.4), (2.7) and (2.8) as

$$(2.9) \quad f(x; \theta) = \frac{\alpha c^b (1+x)^{-b\alpha-1} [1 - (1+x)^{-\alpha}]^{a-1}}{\{c + (1-c) [1 - (1+x)^{-\alpha}]\}^{a+b} B(a, b)}.$$

Hereafter, a random variable X with density function (2.9) will be denoted by $X \sim \text{BMOL}(a, b, c, \alpha)$.

In lifetime analysis, a very useful function is the hazard rate function (hrf) $h(x)$. Therefore, the hrf of $X \sim \text{BMOL}(a, b, c, \alpha)$ is given by

$$(2.10) \quad h(x) = \frac{\alpha c^b (1+x)^{-b\alpha-1} [1 - (1+x)^{-\alpha}]^{\alpha-1}}{\{c + (1-c) [1 - (1+x)^{-\alpha}]\}^{a+b} [B(a, b) - B(W[G(x)], a, b)]}.$$

A random variable X with pdf (2.9) is easily simulated as follows: if $U \sim \text{Beta}(a, b)$, then

$$X = \left[\left(\frac{1 - (1-c)U}{1-U} \right)^{1/\alpha} - 1 \right] \sim \text{BMOL}(a, b, c, \alpha).$$

For specific values of the parameters a , b and c , some known sub-models of the BMOL distribution are given in Table 1.

Table 1: Some BMOL sub-models. MOEL: Marshall–Olkin extended Lomax, Kw–GL: Kumaraswamy–Generalized Lomax, BL: beta Lomax.

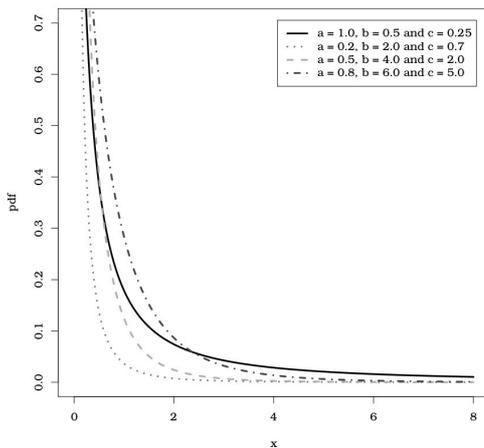
a	b	c	Model	Reference
1	1	1	Lomax($\alpha, 1$)	[17]
1	1	—	MOEL($c, 1, \alpha$)	[10]
1	—	1	Kw–GL($1, b, \alpha, 1$)	[27]
—	—	1	BL($a, b, \alpha, 1$) (with $\mu = 0$)	[24]

3. SHAPES OF THE DENSITY AND HAZARD RATE FUNCTIONS

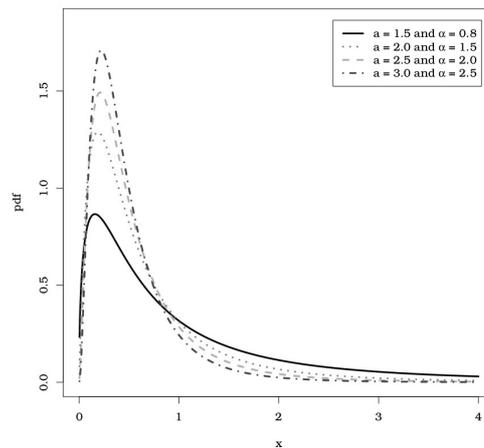
The shapes of the pdf (2.9) can be described analytically by examining the roots of the equation $f'(x) = 0$ and analyzing its limits in (2.9) when $x \rightarrow 0$ or $x \rightarrow \infty$. Clearly, since $f(x) \geq 0$ is integrable, then $\lim_{x \rightarrow \infty} f(x) = 0$. The behavior of $f(x)$ when $x \rightarrow 0$ is governed by the parameter a , which is inherited from the properties of the beta distribution. For $a \leq 1$, we have that $f(x)$ is convex and strictly decreasing. For $a = 1$, $\lim_{x \rightarrow 0} f(x) = b\alpha/c$ and, for $a < 1$, $\lim_{x \rightarrow 0} f(x) = \infty$. For $a > 1$, $f(0) = 0$ and it is unimodal with mode at

$$x_0 = -1 + \left\{ \frac{A_{a,b,c,\alpha} + \left[A_{a,b,c,\alpha}^2 - 4(c-1)(\alpha-1)(b\alpha+1) \right]^{1/2}}{2(b\alpha+1)} \right\}^{1/\alpha},$$

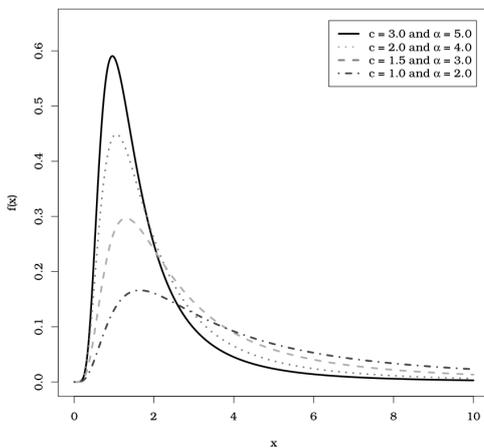
where $A_{a,b,c,\alpha} = 2 - c - \alpha + b\alpha + a c \alpha$. All parameters allow extensive control on the right tail, providing, when $a > 1$, more light or heavy tails, according to the parameters decrease or increase, respectively, and conversely when $a \leq 1$. Some plots in Figure 1 display possible shapes of the pdf for selected parameter values. These plots confirm the above analysis.



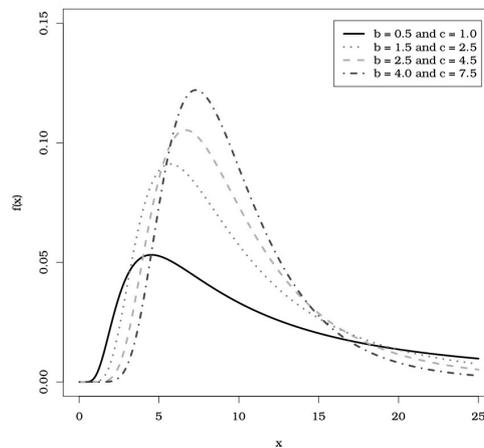
(a) $\alpha = 1.0$



(b) $b = 2.0, c = 0.8$



(c) $a = 7.0, b = 0.5$



(d) $a = 15.0, \alpha = 1.5$

Figure 1: Plots of the pdf (2.9) for selected parameters.

The corresponding hrf can have the classical shapes such as decreasing or unimodal, as shown in Figure 2. Therefore, the new distribution can be appropriate for different applications in lifetime analysis.

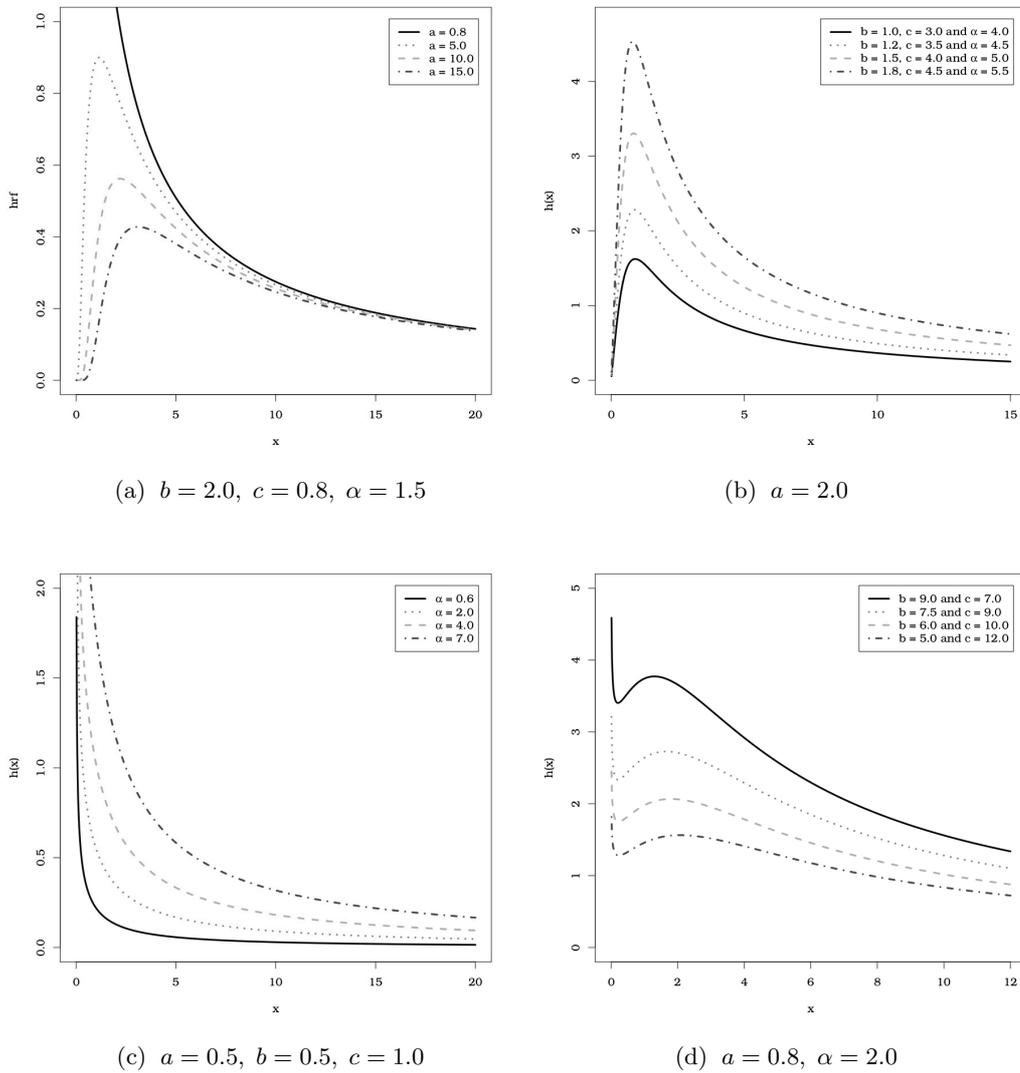


Figure 2: Plots of the hrf (2.10) for selected parameters.

4. USEFUL REPRESENTATION

Using the generalized binomial expansion, Alizadeh *et al.* [3] reveal that the cdf (2.6) admits the following power series

$$(4.1) \quad F(x) = \sum_{k=0}^{\infty} s_k G^k(x),$$

where $G(x)$ is the baseline cdf (2.4) (with $\lambda = 1$) and, for $k \geq 0$,

$$(4.2) \quad s_k = \sum_{i,j=0}^{\infty} \sum_{l=k}^{\infty} \frac{(-1)^{i+l+k} (1-c)^i \binom{b-1}{i} \binom{-a-i}{j} \binom{a+i+j}{l} \binom{l}{k}}{c^{a+i+j} (a+i) B(a,b)}.$$

We note that (4.2) is valid only for $c > 1$, it does not converge for $c < 1$ and it is not applicable for $c = 1$. Differentiating (4.1) term by term, we obtain

$$(4.3) \quad f(x) = \sum_{k=0}^{\infty} s_{k+1} h_{k+1}(x),$$

where $h_{k+1}(x) = (k+1) g(x) G^k(x)$ denotes the exp-G density function with power parameter $k + 1$. Therefore, from (4.3), several properties of the new model can be derived from those exp-G properties [13].

It is possible to go a step further in (4.1). Using the binomial expansion in (4.1) gives

$$F(x) = \sum_{k=0}^{\infty} \sum_{j=0}^k (-1)^j \binom{k}{j} s_k (1+x)^{-j\alpha}.$$

By exchanging the indices j and k in the sums, we can write

$$(4.4) \quad F(x) = \sum_{j=0}^{\infty} \sum_{k=j}^{\infty} (-1)^j \binom{k}{j} s_k (1+x)^{-j\alpha}.$$

Finally, differentiating (4.4) term by term, we obtain

$$(4.5) \quad f(x) = \sum_{j=0}^{\infty} p_j g(x; (j+1)\alpha, 1),$$

where $g(x; (j+1)\alpha, 1)$ is given in (2.5) and, for $j = 0, 1, \dots$,

$$(4.6) \quad p_j = \sum_{k=j+1}^{\infty} (-1)^j \binom{k}{j+1} s_k.$$

From equation (4.5), we note that $f(x)$ is given by a linear combination of Lomax densities. Therefore, several properties of the BMOL distribution can be obtained from those of the Lomax distribution [17].

5. QUANTILE FUNCTION

Let $Q_{a,b}(z)$ denote the qf of the beta distribution with parameters a and b . Then, the qf of the BMOL distribution is given by

$$(5.1) \quad Q(z) = \left[\frac{1 - (1-c) Q_{a,b}(z)}{1 - Q_{a,b}(z)} \right]^{1/\alpha} - 1.$$

An expansion up to third order about $z = 0$ for the beta qf is given by

$$Q_{a,b}(z) = \sum_{i=1}^3 q_i z^{i/a} + \mathcal{O}(z^{4/a}),$$

where $q_i = d_i [a B(a, b)]^{i/a}$, $i = 1, 2, 3$, with $d_1 = 1$,

$$d_2 = \frac{b-1}{a+1}, \quad d_3 = \frac{(b-1)(a^2 + 3ab - a + 5b - 4)}{2(a+1)^2(a+2)}.$$

The skewness and kurtosis measures are determined by $\alpha_3 = \mu_3/\sigma^3$ and $\alpha_4 = \mu_4/\sigma^4$, respectively, where μ_j is the j -th central moment and σ is the standard deviation. For some generalized distributions obtained by the T - X method defined by (2.3), as noted by Alzaatreh *et al.* [4], it could be difficult to determine the third and fourth moments. Alternative measures for the skewness and kurtosis based on the qf are sometimes more appropriate. The measures of skewness S of Bowley [12] and kurtosis K of Moors [19] are defined by

$$(5.2) \quad S = \frac{Q(6/8) + Q(2/8) - 2Q(4/8)}{Q(6/8) - Q(2/8)},$$

$$(5.3) \quad K = \frac{Q(7/8) - Q(5/8) + Q(3/8) - Q(1/8)}{Q(6/8) - Q(2/8)}.$$

These measures are more robust and they exist even for distributions without moments.

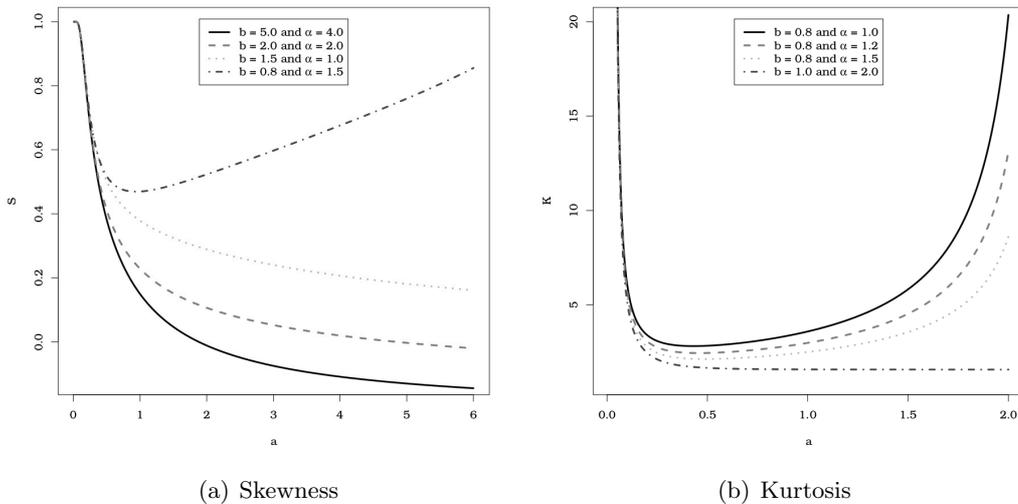


Figure 3: Plots of S of Bowley skewness (5.2) and of K of Moors kurtosis (5.3) measures for selected parameters ($c = 2.0$).

The plots in Figure 3 display the skewness (5.2) and kurtosis (5.3) as functions of the parameter a for some values of the parameters b, c and α . Note that, as pointed in Section 3, the BMOL pdf does not have mode when $a \leq 1$, which implies a greater skewness for these values of the parameter a , as illustrated in Figure 3(a). Similarly, note that the skewness increases when $b > 1$, obtaining negative values when $b, \alpha > 2$. In addition, note that the kurtosis decreases when the values of the parameters b and α increase, as illustrated in Figures 3(b), 1(c) and 1(d).

6. MOMENTS

The moments of X with cdf given by (2.6) can be expressed from the (r, k) -th probability weighted moment (PWM) of a random variable Y with baseline cdf $G(x)$ and pdf $g(x)$, which is defined, for $r, k = 0, 1, \dots$, by

$$\omega_{r,k} = \mathbb{E}\left[Y^r G^k(Y)\right] = \int_0^\infty y^r G^k(y) g(y) dy.$$

Setting $u = G(y)$, we obtain

$$(6.1) \quad \omega_{r,k} = \int_0^1 Q_G^r(u) u^k du,$$

where $Q_G(u)$ is the qf of $G(x)$.

The r -th ordinary moment of X , with $r \in \mathbb{N}$, follows from (4.3), for $c > 1$, as

$$\mu'_r = \mathbb{E}(X^r) = \sum_{k=0}^\infty \int_0^\infty x^r s_{k+1} h_{k+1}(x) dx,$$

where it is possible to exchange the infinite sum and the integral using the dominated convergence theorem. By using (6.1) and $h_{k+1}(x) = (k + 1) g(x) G^k(x)$, we obtain

$$(6.2) \quad \mu'_r = \sum_{k=0}^\infty (k + 1) s_{k+1} \int_0^1 Q_G^r(u) u^k du = \sum_{k=0}^\infty (k + 1) s_{k+1} \omega_{r,k}.$$

Equation (6.2) reveals that the moments of the BMOL distribution can be expressed as an infinite weighted sum of the baseline PWMs.

If $G(x)$ is the Lomax cdf (with $\lambda = 1$), we obtain, using the binomial expansion,

$$Q_G^r(z) = \left[\frac{1}{(1-z)^{1/\alpha}} - 1 \right]^r = \sum_{j=0}^r \binom{r}{j} \frac{(-1)^{r+j}}{(1-z)^{j/\alpha}}$$

and therefore, from equation (6.1),

$$(6.3) \quad \omega_{r,k} = \sum_{j=0}^r (-1)^{r+j} \binom{r}{j} \int_0^1 \frac{u^k}{(1-u)^{j/\alpha}} du.$$

As a result, from (6.2) and (6.3), we obtain that $\mu'_r < \infty$ for $r < \alpha$ and $\mu'_r = \infty$ for $0 < \alpha \leq r$, a condition that also holds for the Lomax distribution.

We can express the r -th ordinary moment of X as a linear combination of the r -th ordinary moments of Lomax random variables. In fact, for $j = 0, 1, \dots$, let $\alpha_j = (j+1)\alpha$. By applying the dominated convergence theorem and using equation (4.5), we can write, for $c > 1$,

$$\mu'_r = \sum_{j=0}^{\infty} p_j \int_0^{\infty} x^r g(x; \alpha_j, 1) dx = \sum_{j=0}^{\infty} p_j \mathbb{E}(Y_j^r),$$

where $Y_j \sim \text{Lomax}(\alpha_j, 1)$.

From the equality $\mathbb{E}(Y_j^r) = \Gamma(r+1)\Gamma(\alpha_j - r)/\Gamma(\alpha_j)$, for $r < \alpha_j$, (see [16]), we obtain

$$(6.4) \quad \mu'_r = \Gamma(r+1) \sum_{j=0}^{\infty} p_j \frac{\Gamma(\alpha_j - r)}{\Gamma(\alpha_j)}, \quad r < \alpha_j, \forall j.$$

Equations (6.2) and (6.4) are the main results of this section. However, the moments of X can be determined from (6.4) more easily than from (6.2).

7. GENERATING FUNCTION

A formula for the moment generating function (mgf) $M(t) = \mathbb{E}(e^{tX})$ of $X \sim \text{BMOL}(a, b, c, \alpha)$ follows from (4.3) as

$$(7.1) \quad M(t) = \sum_{k=0}^{\infty} (k+1) s_{k+1} \rho_k(t),$$

where

$$\rho_k(t) = \int_0^{\infty} e^{tx} g(x) G^k(x) dx.$$

We can obtain an expansion for $\rho_k(t)$, with $t < 0$ and $\alpha \in \mathbb{N}$, using the upper incomplete gamma function, which is defined as

$$(7.2) \quad \Gamma(v, z) = \int_z^{\infty} x^{v-1} e^{-x} dx, \quad v \in \mathbb{R}, z > 0.$$

In fact, setting $w = 1 + x$, we have

$$\rho_k(t) = \int_1^\infty e^{t(w-1)} g(w-1) G^k(w-1) dw = \alpha \int_1^\infty e^{t(w-1)} w^{-\alpha-1} (1-w^{-\alpha})^k dw.$$

Using the binomial expansion, we have

$$(1-w^{-\alpha})^k = \sum_{j=0}^k (-1)^j \binom{k}{j} w^{-j\alpha},$$

which leads to

$$\rho_k(t) = \alpha \sum_{j=0}^k (-1)^{\alpha_j-j+1} \binom{k}{j} t^{\alpha_j+1} e^{|t|} \int_1^\infty e^{-|t|w} (|t|w)^{-\alpha_j-1} dw, \quad t < 0, \alpha \in \mathbb{N},$$

Since $\alpha \in \mathbb{N}$, then $\alpha_j = (j+1)\alpha \in \mathbb{N}$ for all j , which ensures that the quantity $(-1)^{\alpha_j-j+1}$ in the above expression is a real number. Finally, using (7.2), we obtain

$$(7.3) \quad \rho_k(t) = \alpha \sum_{j=0}^k (-1)^{\alpha_j-j} \binom{k}{j} t^{\alpha_j} e^{|t|} \Gamma(-\alpha_j, |t|), \quad t < 0, \alpha \in \mathbb{N}.$$

Equations (7.1) and (7.3) are the main results of this section.

8. MEAN DEVIATIONS AND BONFERRONI AND LORENZ CURVES

As before, for $j = 0, 1, \dots$, let $Y_j \sim \text{Lomax}(\alpha_j, 1)$. The mean deviations of $X \sim \text{BMOL}(a, b, c, \alpha)$ about the mean, $\delta_1 = \mathbb{E}|X - \mu'_1|$ (with $1 < \alpha_j, \forall j$), and about the median, $\delta_2 = \mathbb{E}|X - M|$, can be expressed as

$$\delta_1 = 2\mu'_1 F(\mu'_1) - 2m_X^{(1)}(\mu'_1), \quad \delta_2 = \mu'_1 - 2m_X^{(1)}(M),$$

where μ'_1 is the first ordinary moment of X given by (6.4), $m_X^{(1)}(z) = \int_0^z x f(x) dx$ denotes the first incomplete moment of X , $M = Q(0.5)$ is the median of X and $Q(\cdot)$ is given by (5.1). The mean deviations δ_1 and δ_2 are used frequently as dispersion measures.

Using (4.5), we can write

$$(8.1) \quad m_X^{(1)}(z) = \sum_{j=0}^\infty p_j \int_0^z x g(x; \alpha_j, 1) dx = \sum_{j=0}^\infty p_j m_{Y_j}^{(1)}(z),$$

where $m_{Y_j}^{(1)} = \int_0^z x g(x; \alpha_j, 1) dx$ denotes the first incomplete moment of Y_j and p_j is given by (4.6). For computing δ_1 and δ_2 , we use (2.6), (6.4) and (8.1).

The incomplete moments can be applied to obtain the Bonferroni and Lorenz curves [1], which are useful in several areas. The Bonferroni and Lorenz curves are defined, respectively, by

$$B(\pi) = \frac{m_X^{(1)}(q)}{\pi\mu'_1}, \quad L(\pi) = \pi B(\pi),$$

where $q = Q(\pi)$ is evaluated from (5.1) for $0 < \pi < 1$.

9. ENTROPY

Entropy is a measure of disorder or uncertainty. Two variants of entropy are generally used, the Shannon and Rényi entropies [5]. The latter is a generalization of the first.

For a random variable $X \sim \text{BMOL}(a, b, c, \alpha)$, it is easier to obtain an explicit expression for the Shannon entropy than for the Rényi entropy. The Shannon entropy of an absolutely continuous random variable X with pdf $f(x)$ is defined by

$$\eta_X = \mathbb{E}_f\{-\log[f(X)]\} = -\int_0^\infty \log[f(x)] f(x) dx.$$

Considering that $W[G(x)]$ is an absolutely continuous distribution with density $g(x)w[G(x)]$, where $G(x)$ is the baseline distribution and $w(z) = W'(z)$ (see Section 2), it can be proved that the density $f(x)$ satisfies

$$\mathbb{E}_f\{\log(W[G(X)])\} = -\xi(a, b),$$

$$\mathbb{E}_f\{1 - \log(W[G(X)])\} = -\xi(b, a),$$

$$\mathbb{E}_f\{\log(w[G(X)])\} + \mathbb{E}_f\{\log[g(X)]\} - \mathbb{E}_U\{\log[w(U)]\} - \mathbb{E}_U\{\log(g[Q_G(U)])\} = 0,$$

where $\xi(a, b) = -\frac{\partial}{\partial a} \log[B(a, b)] = \psi(a+b) - \psi(a)$, $\psi(\cdot)$ denotes the digamma function and $U \sim \text{Beta}(a, b)$.

From the equalities $w(z) = c/[c + (1-c)z]^2$ (with $c \neq 1$) and $g(Q_G(u)) = \alpha(1-u)^{(\alpha+1)/\alpha}$, we obtain

$$\begin{aligned} \mathbb{E}_U\{\log[w(U)]\} &= \log c - 2\mathbb{E}_U\{\log[c + (1-c)U]\}, \\ \mathbb{E}_U\{\log(g[Q_G(U)])\} &= \log \alpha + \frac{\alpha+1}{\alpha} \mathbb{E}_U[\log(1-U)]. \end{aligned}$$

Further, we have

$$\begin{aligned} \mathbb{E}_U\{\log(1 - U)\} &= \frac{1}{B(a, b)} \int_0^1 \log(1 - u) u^{a-1} (1 - u)^{b-1} du \\ &= -\xi(b, a), \\ \mathbb{E}_U\{\log[c + (1 - c)U]\} &= \frac{1}{B(a, b)} \int_0^1 \log[c + (1 - c)u] u^{a-1} (1 - u)^{b-1} du \\ &= \log c - I_{a,b,c} {}_3F_2(1, 1, 1 + a; 2, 1 + a + b; \frac{c - 1}{c}), \end{aligned}$$

where $I_{a,b,c} = \frac{a(c-1)}{c(a+b)}$ and ${}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; z)$ is the generalized hypergeometric function.

Thus, we can write

$$\begin{aligned} \eta_X &= \log[B(a, b)] + (a - 1) \xi(a, b) + \left(b - 1 + \frac{\alpha + 1}{\alpha}\right) \xi(b, a) + \log c - \log \alpha \\ &\quad - 2 I_{a,b,c} {}_3F_2(1, 1, 1 + a; 2, 1 + a + b; \frac{c - 1}{c}). \end{aligned}$$

The Shannon entropy is relevant because it is related to other notions of entropy in various areas such as probability theory, computer sciences, dynamical systems and statistical physics.

10. ORDER STATISTICS

Let X_1, \dots, X_n be a random sample of size n from a distribution $F(x)$. Then, the pdf of the m -th order statistic, $X_{(m)}$, is given by [26, p. 218]

$$f_{(m)}(x) = K F^{m-1}(x)(1 - F(x))^{n-m} f(x),$$

where $K = n! / [(m - 1)!(n - m)!]$.

For $1 \leq m \leq n$, we obtain

$$f_{(m)}(x) = K f(x) \sum_{j=0}^{n-m} (-1)^j \binom{n-m}{j} F^{m+j-1}(x).$$

Based on (4.1) and (4.2) and using an expansion for power series raised to positive integer powers [11, p. 17], we have, for $c > 1$,

$$F^{m+j-1}(x) = \left(\sum_{k=0}^{\infty} s_k G^k(x) \right)^{m+j-1} = \sum_{k=0}^{\infty} v_{j,k} G^k(x),$$

where $G(x)$ is the baseline distribution given in (2.4) (with $\lambda = 1$), $v_{j,0} = s_0^{m+j-1}$ and, for $i \geq 1$,

$$v_{j,i} = \frac{1}{i s_0} \sum_{l=1}^i [(m+j)l - i] s_l v_{j,i-l}.$$

Therefore, we obtain

$$f_{(m)}(x) = K f(x) \sum_{j=0}^{n-m} \sum_{k=0}^{\infty} (-1)^j \binom{n-m}{j} v_{j,k} G^k(x),$$

where the density $f(x)$ is given in (2.9).

Considering the BMO-G family, Alizadeh *et al.* [3] propose other expansion for $f_{(m)}(x)$ given by

$$(10.1) \quad f_{(m)}(x) = \sum_{r,k=0}^{\infty} p_{r,k} h_{r+k+1}(x),$$

where $h_{r+k+1}(x)$ denotes the exp- G density function with parameter $r + k + 1$,

$$p_{r,k} = \frac{n!(r+1)(m-1)!s_{r+1}}{r+k+1} \sum_{j=0}^{n-m} \frac{(-1)^j v_{j,k}}{(n-m-j)!j!},$$

and s_r is given in (4.2) for $c > 1$.

Equation (10.1) reveals that, for the BMO-G family, the density function $f_{(m)}(x)$ of the m -th order statistic $X_{(m)}$ can be expressed as a linear combination of exp- G densities. Therefore, some structural properties of $X_{(m)}$ can be obtained from those of the exp- G distribution [13].

11. MAXIMUM LIKELIHOOD ESTIMATION

Several approaches for parameter estimation were proposed in the statistical literature but the maximum likelihood method is the most commonly employed. The MLEs enjoy desirable properties for constructing confidence intervals. In this section, we estimate the parameters of the BMOL distribution by maximum likelihood for complete data sets. Let $\mathbf{x} = (x_1, \dots, x_n)^\top$ be a sample of size n from $X \sim \text{BMOL}(a, b, c, \alpha)$ and $\boldsymbol{\theta} = (a, b, c, \alpha)^\top$ the parameter vector. The log-likelihood for $\boldsymbol{\theta}$ corresponding to the sample \mathbf{x} , denoted by $\ell_f(\boldsymbol{\theta}; \mathbf{x})$, is given by

$$\begin{aligned} \ell_f(\boldsymbol{\theta}; \mathbf{x}) = & -n \log[B(a, b)] + (a-1) \sum_{i=1}^n \log\{W[G(x_i)]\} \\ & + (b-1) \sum_{i=1}^n \log\{1 - W[G(x_i)]\} + \sum_{i=1}^n \log\{w[G(x_i)]\} + \ell_g(\alpha; \mathbf{x}), \end{aligned}$$

where $\ell_g(\alpha; \mathbf{x}) = \sum_{i=1}^n \log[g(x_i)]$ is the log-likelihood for the Lomax parameters (with $\lambda = 1$). From (2.4) and (2.7), we can write

$$\begin{aligned}\log\{W[G(x_i)]\} &= \log\left[\frac{(1+x)^\alpha - 1}{(1+x)^\alpha + c - 1}\right], \\ \log\{1 - W[G(x_i)]\} &= \log\left[\frac{c}{(1+x)^\alpha + c - 1}\right], \\ \log\{w[G(x_i)]\} &= \log\left\{\frac{c(1+x)^{2\alpha}}{[(1+x)^\alpha + c - 1]^2}\right\}.\end{aligned}$$

Then,

$$\begin{aligned}(11.1) \quad \ell_f(\boldsymbol{\theta}; \mathbf{x}) &= -n \log[B(a, b)] + (a-1) \sum_{i=1}^n \log\left[\frac{(1+x_i)^\alpha - 1}{(1+x_i)^\alpha + c - 1}\right] \\ &\quad + (b-1) \sum_{i=1}^n \log\left[\frac{c}{(1+x_i)^\alpha + c - 1}\right] \\ &\quad + \sum_{i=1}^n \log\left\{\frac{c(1+x_i)^{2\alpha}}{[(1+x_i)^\alpha + c - 1]^2}\right\} + \ell_g(\alpha; \mathbf{x}).\end{aligned}$$

The MLE $\hat{\boldsymbol{\theta}}_n$ of $\boldsymbol{\theta}$ can be obtained by maximizing (11.1) directly by using SAS (PROC NLMIXED), R (optim and MaxLik functions) or Ox program (sub-routine MaxBFGS). Details for fitting univariate distributions using maximum likelihood in R for censored or non censored data can be obtained at

<http://www.inside-r.org/packages/cran/fitdistrplus/docs/mledist>
[Accessed 28 02 2017].

Alternatively, we can obtain the components of the score vector $\mathbf{U}_\theta = (U_a, U_b, U_c, U_\alpha)^\top$ and set them to zero. They are given by

$$\begin{aligned}U_a &= \frac{\partial}{\partial a} \ell_f(\boldsymbol{\theta}; \mathbf{x}) = n[\psi(a+b) - \psi(a)] + \sum_{i=1}^n \log\left[\frac{(1+x_i)^\alpha - 1}{(1+x_i)^\alpha + c - 1}\right], \\ U_b &= \frac{\partial}{\partial b} \ell_f(\boldsymbol{\theta}; \mathbf{x}) = n[\psi(a+b) - \psi(b)] + \sum_{i=1}^n \log\left[\frac{c}{(1+x_i)^\alpha + c - 1}\right], \\ U_c &= \frac{\partial}{\partial c} \ell_f(\boldsymbol{\theta}; \mathbf{x}) = \frac{1}{c} \sum_{i=1}^n \frac{b[(1+x_i)^\alpha - 1] - c - a + 1}{(1+x_i)^\alpha + c - 1}, \\ U_\alpha &= \frac{\partial}{\partial \alpha} \ell_f(\boldsymbol{\theta}; \mathbf{x}) = \frac{n}{\alpha} - \sum_{i=1}^n \log(1+x_i) \\ &\quad + \sum_{i=1}^n \frac{\log(1+x_i)}{(1+x_i)^\alpha + c - 1} [2(c-1) - (b-1)(1+x_i)^\alpha] \\ &\quad + c(a-1) \sum_{i=1}^n \frac{(1+x_i)^\alpha \log(1+x_i)}{[(1+x_i)^\alpha - 1][(1+x_i)^\alpha + c - 1]}.\end{aligned}$$

The MLE $\hat{\boldsymbol{\theta}}_n$ is obtained by solving the equations $U_a = U_b = U_c = U_\alpha = 0$ simultaneously. Because they can not be solved in closed-form, numerical iterative Newton–Raphson type algorithms can be applied.

Under general regularity conditions, we have $(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \stackrel{a}{\sim} N_4(\mathbf{0}, K(\boldsymbol{\theta})^{-1})$, where $K(\boldsymbol{\theta})$ is the 4×4 expected information matrix and $\stackrel{a}{\sim}$ denotes asymptotic distribution. For n large, $K(\boldsymbol{\theta})$ can be approximated by the observed information matrix. This normal approximation for the MLE $\hat{\boldsymbol{\theta}}_n$ can be used for construing approximate confidence intervals and for testing hypotheses on the parameters a, b, c and α .

Suppose that the parameter vector is partitioned as $\boldsymbol{\theta} = (\boldsymbol{\psi}_1^\top, \boldsymbol{\psi}_2^\top)^\top$, where $\dim(\boldsymbol{\psi}_1) + \dim(\boldsymbol{\psi}_2) = \dim(\boldsymbol{\theta})$. The likelihood ratio (LR) statistic for testing the null hypothesis $\mathcal{H}_0 : \boldsymbol{\psi}_1 = \boldsymbol{\psi}_1^{(0)}$ against the alternative hypothesis $\mathcal{H}_1 : \boldsymbol{\psi}_1 \neq \boldsymbol{\psi}_1^{(0)}$ is given by $LR_n = 2 \{\ell_f(\hat{\boldsymbol{\theta}}_n) - \ell_f(\tilde{\boldsymbol{\theta}}_n)\}$, where $\hat{\boldsymbol{\theta}}_n = (\hat{\boldsymbol{\psi}}_1^\top, \hat{\boldsymbol{\psi}}_2^\top)^\top$, $\tilde{\boldsymbol{\theta}}_n = (\boldsymbol{\psi}_1^{(0)\top}, \tilde{\boldsymbol{\psi}}_2^\top)^\top$, $\hat{\boldsymbol{\psi}}_i$ and $\tilde{\boldsymbol{\psi}}_i$ are the MLE's under the alternative and null hypotheses, respectively, and $\boldsymbol{\psi}_1^{(0)}$ is a specified parameter vector. Based on the first-order asymptotic theory, we know that $LR \stackrel{a}{\sim} \chi_k^2$, where $k = \dim(\boldsymbol{\psi}_1)$. Thus, we can compute the maximum values of the unrestricted and restricted log-likelihoods to obtain LR statistics for testing some sub-models of the BMOL distribution (see Table 1).

12. SIMULATION STUDY

In this section, we perform a Monte Carlo simulation experiment to evaluate the behavior of the MLE $\hat{\boldsymbol{\theta}}_n = (\hat{a}_n, \hat{b}_n, \hat{c}_n, \hat{\alpha}_n)$ in finite samples and estimate the relative biases and mean squared errors (MSEs) of the estimates for different sample sizes n . We consider 10,000 Monte Carlo replications and use the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm with analytical derivatives to maximize the log-likelihood function (11.1). We set the parameter values $a = 0.5$, $c = 0.25$ and vary b and α . All computations are performed using the C programming language and the GNU Scientific Library (version 2.1).

The results given in Table 2 reveal that, generally, the relative bias and MSE values decrease when n increases, which is to be expected since the MLEs are asymptotically unbiased. The minimum absolute values for the relative biases and MSEs are equal to 0.003. In counterpart, the maximum absolute values for the relative biases and MSEs are, respectively, 0.930 and 2.182. Further, it can be noted in Table 2 that the parameter c was underestimated in some cases (negative relative bias).

Table 2: Relative bias and MSE values of the MLE $\hat{\theta}_n = (\hat{a}_n, \hat{b}_n, \hat{c}_n, \hat{\alpha}_n)$ (with $a = 0.5$ and $c = 0.25$).

b	α	n	relative bias				MSE			
			\hat{a}_n	\hat{b}_n	\hat{c}_n	$\hat{\alpha}_n$	\hat{a}_n	\hat{b}_n	\hat{c}_n	$\hat{\alpha}_n$
0.5	0.5	100	0.115	0.170	-0.003	0.326	0.035	0.184	0.037	0.268
		200	0.052	0.119	-0.017	0.165	0.008	0.099	0.016	0.118
		300	0.034	0.081	-0.013	0.114	0.004	0.059	0.010	0.075
	0.75	100	0.113	0.180	0.015	0.297	0.040	0.187	0.045	0.544
		200	0.051	0.118	-0.013	0.161	0.008	0.096	0.016	0.260
		300	0.033	0.084	-0.011	0.110	0.004	0.060	0.010	0.167
0.75	0.5	100	0.093	0.092	0.080	0.598	0.035	0.382	0.170	0.602
		200	0.044	0.060	-0.015	0.333	0.006	0.161	0.016	0.257
		300	0.029	0.059	-0.024	0.219	0.004	0.124	0.009	0.149
	0.75	100	0.090	0.113	0.118	0.544	0.036	0.424	0.241	1.166
		200	0.042	0.065	-0.009	0.312	0.006	0.160	0.016	0.533
		300	0.029	0.060	-0.021	0.209	0.004	0.122	0.009	0.323
1.0	0.5	100	0.089	0.032	0.183	0.930	0.027	0.577	0.175	1.131
		200	0.046	0.010	0.011	0.551	0.005	0.256	0.022	0.504
		300	0.031	0.010	-0.008	0.388	0.003	0.191	0.012	0.295
	0.75	100	0.083	0.071	0.251	0.835	0.024	1.279	0.404	2.182
		200	0.044	0.023	0.018	0.506	0.005	0.260	0.024	0.999
		300	0.030	0.015	-0.006	0.363	0.003	0.186	0.012	0.609

13. APPLICATION

In this section, the potentiality of the BMOL distribution is proved empirically by means of one lifetime application. We use an uncensored data set corresponding to 84 observations on service times for failed windshields [21, Table 16.11] and fit the BMOL distribution and its sub-models (see Table 1) to these data. All computations are performed using the R software (version 3.0.2, AdequacyModel package). The descriptive statistics for the current data are given in Table 3.

Table 3: Descriptive statistics for the service times data.

min.	1st quantile	median	mean	3rd quantile	max.
0.040	1.839	2.354	2.557	3.393	4.663

For maximizing the log-likelihood function (11.1), we use the BFGS algorithm with numerical derivatives. The MLEs are given in Table 4 (with standard errors in parentheses). For purposes of comparison, we compute some goodness-of-fit statistics: Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Hannan–Quinn Information Criterion (HQIC), Cramér–von Mises Criterion (W^*) and Anderson–Darling Criterion (A^*) [6]. In general, small values of these statistics suggest a better fit. We also include in the comparison the exponentiated-Weibull (EW) distribution [20], since it is a widely used lifetime model. Its cdf and pdf are given, respectively, by

$$R(x) = \left[1 - e^{-\left(\frac{x}{\alpha}\right)^\beta}\right]^\eta \quad \text{and} \quad r(x) = \frac{\beta\eta}{\alpha} \left(\frac{x}{\alpha}\right)^{\beta-1} \left[1 - e^{-\left(\frac{x}{\alpha}\right)^\beta}\right]^{\eta-1} e^{-\left(\frac{x}{\alpha}\right)^\beta},$$

where $x \geq 0$ and $\alpha, \beta, \eta > 0$.

Table 4: MLEs and standard errors for the service times data.

Distribution	MLE					$\hat{\eta}$
	\hat{a}	\hat{b}	\hat{c}	$\hat{\alpha}$	$\hat{\beta}$	
Lomax($\alpha, 1$)	—	—	—	0.824 (0.090)	—	—
MOEL($c, 1, \alpha$)	—	—	441.875 (242.694)	4.957 (0.424)	—	—
BL($a, b, \alpha, 1$)	6.664 (1.055)	38.687 (79.332)	—	0.133 (0.254)	—	—
Kw–GL($a, b, \alpha, 1$)	4.378 (0.517)	244.216 (213.820)	—	0.254 (0.083)	—	—
BMOL(a, b, c, α)	1.377 (0.356)	6.243 (5.526)	209.269 (143.799)	2.954 (0.627)	—	—
EW(α, β, η)	—	—	—	3.972 (0.136)	5.958 (0.255)	0.271 (0.036)

The goodness-of-fit values for the fitted distributions are listed in Table 5.

Based on the figures in Table 5, we note that the EW distribution presents the smaller values of the AIC, BIC and HQIC statistics. On the other hand, the BMOL distribution presents the smaller values of the W^* and A^* statistics. Since the BMOL and EW distributions are non-embedded models, a comparison between them is more appropriate by means of these statistics. Also, note that the BMOL model presents the smaller value of the AIC statistic among all its sub-models and the smaller values of the BIC and HQIC statistics comparatively with the Lomax, BL and Kw-GL distributions. Therefore, we can conclude that

the BMOL distribution gives the best fit to the current data. If a minimum number of parameters is taken into account, the MOEL or EW distributions can be chosen, since these also has less parameters.

Table 5: Goodness-of-fit statistics for the service times data.

Distribution	Statistic				
	AIC	BIC	HQIC	W*	A*
Lomax($\hat{\alpha}, 1$)	406.442	408.873	407.419	0.562	3.786
MOEL($\hat{c}, 1, \hat{\alpha}$)	266.987	271.849	268.942	0.068	0.650
BL($\hat{a}, \hat{b}, \hat{\alpha}, 1$)	312.806	320.098	315.737	0.553	3.737
Kw–GL($\hat{a}, \hat{b}, \hat{\alpha}, 1$)	282.938	290.230	285.869	0.175	1.463
BMOL($\hat{a}, \hat{b}, \hat{c}, \hat{\alpha}$)	265.694	275.417	269.602	0.048	0.487
EW($\hat{\alpha}, \hat{\beta}, \hat{\eta}$)	261.208	268.501	264.140	0.129	0.831

To analyze how significant are the parameters of the BMOL distribution in modeling the current data, we use the LR statistic, as discussed in Section 11, for testing the BMOL model versus its sub-models listed in Table 1. The results are given in Table 6. Based on the figures in this table, we note that the rejection of the null hypotheses for the Lomax, MOEL, BL and Kw–GL models (at the 10% significance level) is significant. So, we have evidence of the potential need for including the parameters a, b and c to model the current data.

Table 6: LR tests for the service times data.

Models	Hypotheses	LR statistic	p -value
Lomax vs. BMOL	$\mathcal{H}_0: a=b=c=1$ vs. $\mathcal{H}_1: \mathcal{H}_0$ is false	146.748	1.33×10^{-31}
MOEL vs. BMOL	$\mathcal{H}_0: a=b=1$ vs. $\mathcal{H}_1: \mathcal{H}_0$ is false	5.294	7.09×10^{-2}
BL vs. BMOL	$\mathcal{H}_0: c=1$ vs. $\mathcal{H}_1: \mathcal{H}_0$ is false	49.112	2.42×10^{-12}
Kw–GL vs. BMOL	$\mathcal{H}_0: a=c=1$ vs. $\mathcal{H}_1: \mathcal{H}_0$ is false	19.244	6.63×10^{-5}

The plots of the estimated densities for the EW, MOEL and BMOL distributions are displayed in Figure 4. Based on these plots, it is possible to assess the best overall fit of the BMOL distribution to the current data.

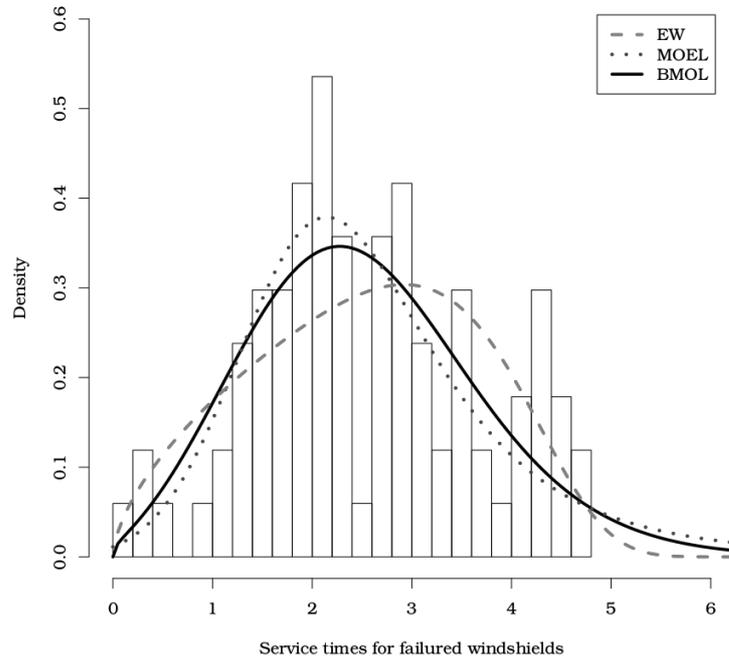


Figure 4: Comparison of the EW, MOEL and BMOL estimated densities for the service times data.

14. CONCLUSION AND FINAL REMARKS

In this chapter, we introduce a new four-parameter model, called the beta Marshall–Olkin Lomax (BMOL) distribution, as a member of the beta Marshall–Olkin generated (BMO-G) family [3] when the parent model is the Lomax distribution [17] (with $\lambda = 1$). Some sub-models of the BMOL distribution are presented. The new distribution has simple expressions for the cumulative and density functions. We study some of its mathematical and statistical properties. We demonstrate that the BMOL density can be expressed as linear combinations of Lomax and exponentiated-Lomax densities and therefore some of its structural properties can be obtained from those of these models. We present explicit expressions for the quantile function, moments, generating function, mean deviations, Bonferroni and Lorenz curves, Shannon entropy and order statistics. We obtain the maximum likelihood estimates for complete samples and perform a Monte Carlo simulation in order to evaluate the behavior of these estimates in finite samples. We compare the performance of the new model with other related distributions including the exponentiated Weibull model using classical goodness-of-fit statistics. The results confirm that the BMOL distribution is very appropriate for lifetime applications.

ACKNOWLEDGMENTS

This research was partially supported by CNPq and CAPES, Brazil.

REFERENCES

- [1] AABERGE, R. (2000). Characterizations of Lorenz curves and income distributions, *Social Choice and Welfare*, **17**(4), 639–653.
- [2] ALEXANDER, C.; CORDEIRO, G.M.; ORTEGA, E.M.M. and SARABIA, J.M. (2012). Generalized beta-generated distributions, *Computational Statistics and Data Analysis*, **56**(6), 1880–1897.
- [3] ALIZADEH, M.; CORDEIRO, G.M.; BRITO, E. and DEMÉTRIO, C.G.B. (2015). The beta Marshall–Olkin family of distributions, *Journal of Statistical Distributions and Applications*, **2**(1), 1–18.
- [4] ALZAATREH, A.; LEE, C. and FAMOYE, F. (2013). A new method for generating families of continuous distributions, *METRON*, **71**(1), 63–79.
- [5] BROMILEY, P.; THACKER, N. and BOUHOVA-THACKER, E. (2004). Shannon entropy, Rényi entropy, and information, *Statistics and Inf. Series (2004-004)*.
- [6] CHEN, G. and BALAKRISHNAN, N. (1995). A general purpose approximate goodness-of-fit test, *Journal of Quality Technology*, **27**(2), 154–161.
- [7] CORDEIRO, G.M. and DE CASTRO, M. (2011). A new family of generalized distributions, *Journal of Statistical Computation and Simulation*, **81**(7), 883–898.
- [8] CORDEIRO, G.M.; ORTEGA, E.M.M. and DA CUNHA, D.C.C. (2013). The exponentiated generalized class of distributions, *Journal of Data Science*, **11**(1), 1–27.
- [9] EUGENE, N.; LEE, C. and FAMOYE, F. (2002). Beta-normal distribution and its applications, *Communications in Statistics – Theory and Methods*, **31**(4), 497–512.
- [10] GHITANY, M.E.; AL-AWADHI, F.A. and ALKHALFAN, L.A. (2007). Marshall–Olkin extended Lomax distribution and its application to censored data, *Communications in Statistics – Theory and Methods*, **36**(10), 1855–1866.
- [11] GRADSHTEYN, I.S. and RYZHIK, I.M. (2007). *Table of integrals, series, and products*, Elsevier/Academic Press, Amsterdam, seventh ed.
- [12] GROENEVELD, R.A. and MEEDEN, G. (1984). Measuring skewness and kurtosis, *Journal of the Royal Statistical Society. Series D (The Statistician)*, **33**(4), 391–399.
- [13] GUPTA, R.C.; GUPTA, P.L. and GUPTA, R.D. (1998). Modeling failure time data by Lehmann alternatives, *Communications in Statistics – Theory and Methods*, **27**(4), 887–904.

- [14] LAI, C.D. (2013). Constructions and applications of lifetime distributions, *Applied Stochastic Models in Business and Industry*, **29**(2), 127–140.
- [15] LAI, C.D.; XIE, M. and BARLOW, R.E. (2006). *Stochastic Ageing and Dependence for Reliability*, Springer–Verlag New York, Inc.
- [16] LEMONTE, A.J. and CORDEIRO, G.M. (2013). An extended Lomax distribution, *Statistics*, **47**(4), 800–816.
- [17] LOMAX (1954). Business failures: Another example of the analysis of failure data, *Journal of the American Statistical Association*, **49**(268), 847–852.
- [18] MARSHALL, A.W. and OLKIN, I. (1997). A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families, *Biometrika*, **84**(3), 641–652.
- [19] MOORS, J.J.A. (1988). A quantile alternative for kurtosis, *Journal of the Royal Statistical Society. Series D (The Statistician)*, **37**(1), 25–32.
- [20] MUDHOLKAR, G.S. and SRIVASTAVA, D.K. (1993). Exponentiated Weibull family for analyzing bathtub failure-rate data, *IEEE Transactions on Reliability*, **42**(2), 299–302.
- [21] MURTHY, D.P.; XIE, M. and JIANG, R. (2004). *Weibull Models*, John Wiley & Sons.
- [22] NADARAJAH, S.; CORDEIRO, G.M. and ORTEGA, E.M.M. (2015). The Zografos–Balakrishnan–G family of distributions: Mathematical properties and applications, *Communications in Statistics – Theory and Methods*, **44**(1), 186–215.
- [23] PESCIM, R.R.; DEMÉTRIO, C.G.B.; CORDEIRO, G.M.; ORTEGA, E.M.M. and URBANO, M.R. (2010). The beta generalized half-normal distribution, *Computational Statistics and Data Analysis*, **54**(4), 945–957.
- [24] RAJAB, M.; ALEEM, M.; NAWAZ, T. and DANIYAL, M. (2013). On five parameter beta Lomax distribution, *Journal of Statistics*, **20**(1), 102–118.
- [25] RISTIĆ, M.M. and BALAKRISHNAN, N. (2012). The gamma-exponentiated exponential distribution, *Journal of Statistical Computation and Simulation*, **82**(8), 1191–1206.
- [26] SEVERINI, T. (2005). *Elements of Distribution Theory*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press.
- [27] SHAMS, T.M. (2013). The Kumaraswamy-Generalized Lomax distribution, *Middle-East Journal of Scientific Research*, **17**, 641–646.
- [28] TAHIR, M.H. and NADARAJAH, S. (2015). Parameter induction in continuous univariate distributions: Well-established G families, *Anais da Academia Brasileira de Ciências*, **87**(2), 539–568.
- [29] ZOGRAFOS, K. and BALAKRISHNAN, N. (2009). On families of beta- and generalized gamma-generated distributions and associated inference, *Statistical Methodology*, **6**(4), 344–362.

THE CUSUM MEDIAN CHART FOR KNOWN AND ESTIMATED PARAMETERS

- Authors: PHILIPPE CASTAGLIOLA
– Université de Nantes & LS2N UMR CNRS 6004,
Nantes, France
philippe.castagliola@univ-nantes.fr
- FERNANDA OTILIA FIGUEIREDO
– Faculdade de Economia da Universidade do Porto & CEAUL,
Porto, Portugal
otilia@fep.up.pt
- PETROS E. MARAVELAKIS
– Department of Business Administration, University of Piraeus,
Piraeus, Greece
maravel@unipi.gr

Received: July 2016

Revised: March 2017

Accepted: April 2017

Abstract:

- The usual CUSUM chart for the mean (CUSUM- \bar{X}) is a chart used to quickly detect small to moderate shifts in a process. In presence of outliers, this chart is known to be more robust than other mean-based alternatives like the Shewhart mean chart but it is nevertheless affected by these unusual observations because the mean (\bar{X}) itself is affected by the outliers. An outliers robust alternative to the CUSUM- \bar{X} chart is the CUSUM median (CUSUM- \tilde{X}) chart, because it takes advantage of the robust properties of the sample median. This chart has already been proposed by other researchers and compared with other alternative charts in terms of robustness, but its performance has only been investigated through simulations. Therefore, the goal of this paper is not to carry out a robustness analysis but to study the effect of parameter estimation in the performance of the chart. We study the performance of the CUSUM- \tilde{X} chart using a Markov chain method for the computation of the distribution and the moments of the run length. Additionally, we examine the case of estimated parameters and we study the performance of the CUSUM- \tilde{X} chart in this case. The run length performance of the CUSUM- \tilde{X} chart with estimated parameters is also studied using a proper Markov chain technique. Conclusions and recommendations are also given.

Key-Words:

- *average run length; CUSUM chart; estimated parameters; median; order statistics.*

AMS Subject Classification:

- 62G30, 62P30.

1. INTRODUCTION

A main objective for a product or a process is to continuously improve its quality. This goal, in statistical terms, may be expressed as variability reduction. Statistical Process Control (SPC) is a well known collection of methods aiming at this purpose and the control charts are considered as the main tools to detect shifts in a process. The most popular control charts are the Shewhart charts, the Cumulative Sum (CUSUM) charts and the Exponentially Weighted Moving Average (EWMA) charts. Shewhart type charts are used to detect large shifts in a process whereas CUSUM and EWMA charts are known to be fast in detecting small to moderate shifts.

The usual Shewhart chart for monitoring the mean of a process is the \bar{X} chart. It is very efficient for detecting large shifts in a process (see for example Teoh *et al.*, 2014). An alternative chart used for the same purpose is the median (\tilde{X}) chart. The median chart is simpler than the \bar{X} chart and it can be easily implemented by practitioners. The main advantages of the \tilde{X} chart over the \bar{X} chart is its robustness against outliers, contamination or small deviations from normality. This property is particularly important for processes running for a long time. Usually, in such processes, the data are not checked for irregular behaviour and, therefore, the \tilde{X} chart is an ideal choice.

The CUSUM- \bar{X} control chart has been introduced by Page (1954). It is able to quickly detect small to moderate shifts in a process. The CUSUM- \bar{X} control chart uses information from a long sequence of samples and, therefore, it is able to signal when a persistent special cause exists (see for instance Nenes and Tagaras (2006) or Liu *et al.* (2014)). However, since it is mean-based, the CUSUM- \bar{X} suffers from the inefficiency of the mean \bar{X} to correctly handle outliers, contamination or small deviations from normality. A natural alternative solution to overcome this problem is the CUSUM- \tilde{X} chart. This chart has already been proposed by Yang *et al.* (2010). In their paper they compare its performance with the Shewhart, EWMA and CUSUM charts for the mean under some contaminated normal distributions using only simulation procedures. This chart has also been considered by Nazir *et al.* (2013a), again in a simulation study with other CUSUM charts, in order to compare their performance for the phase II monitoring of location in terms of robustness against non-normality, special causes of variation and outliers.

In the last decades, different types of control charts for variable and count data based on CUSUM schemes, in univariate or multivariate cases, were proposed in the literature. Here we only mention some recent and less traditional works on robust and enhanced CUSUM schemes, nonparametric and adaptive CUSUM control charts and CUSUM charts for count and angular data. The

interested reader could also take into account the references in the paper that follow.

Ou *et al.* (2011, 2012) carried out a comparative study to evaluate the performance and robustness of some typical \bar{X} , CUSUM and SPRT type control charts for monitoring either the process mean or both the mean and variance, also providing several design tables to facilitate the implementation of the optimal versions of the charts. Nazir *et al.* (2013b), following the same methodology used in Nazir *et al.* (2013a), but for monitoring the dispersion, considered several CUSUM control charts based on different scale estimators, and analyzed its performance and robustness. Qiu and Zhang (2015) investigated the performance of some CUSUM control charts for transformed data in order to accommodate deviations from the normality assumption when monitoring the process data, and they compared its efficiency with alternative nonparametric control charts. To improve the overall performance of the CUSUM charts to detect small, moderate and large shifts in the process mean, Al-Sabah (2010) and Abujiya *et al.* (2015) proposed the use of special sampling schemes to collect the data, such as the ranked set sampling scheme and some extensions of it, instead of using the traditional simple random sampling. Saniga *et al.* (2006, 2012) discussed the economic advantages of the CUSUM versus Shewhart control charts to monitor the process mean when one or two components of variance exist in a process.

The use of nonparametric control charts has attracted the attention of researchers and practitioners. Chatterjee and Qiu (2009) proposed a nonparametric cumulative sum control chart using a sequence of bootstrap control limits to monitor the mean, when the data distribution is non-normal or unknown. Li *et al.* (2010b) considered nonparametric CUSUM and EWMA control charts based on the Wilcoxon rank-sum test for detecting mean shifts, and they discussed the effect of phase I estimation on the performance of the chart. Mukherjee *et al.* (2013) and Graham *et al.* (2014) proposed CUSUM control charts based on the exceedance statistic for monitoring the location parameter. Chowdhury *et al.* (2015) proposed a single distribution free phase II CUSUM control chart based on the Lepage statistic for the joint monitoring of location and scale. The performance of this chart was evaluated by analyzing some moments and percentiles of the run-length distribution, and a comparative study with other CUSUM charts was provided. Wang *et al.* (2017) proposed a nonparametric CUSUM chart based on the Mann–Whitney statistic and on a change point model to detect small shifts.

Wu *et al.* (2009) and Li and Wang (2010) proposed adaptive CUSUM control charts implemented with a dynamical adjustment of the reference parameter of the chart to efficiently detect a wide range of mean shifts. Ryu *et al.* (2010) proposed the design of a CUSUM chart based on the expected weighted run length (EWRL) to detect shifts in the mean of unknown size. Li *et al.* (2010a) and Ou *et al.* (2013) considered adaptive control charts with variable sampling intervals or variable sample sizes to overcome the detecting ability of the traditional

CUSUM. Liu *et al.* (2014) proposed an adaptive nonparametric CUSUM chart based on sequential ranks that efficiently and robustly detects unknown shifts of several magnitudes in the location of different distributions. Wang and Huang (2016) proposed an adaptive multivariate CUSUM chart, with the reference value changing dynamically according to the current estimate of the process shift, that performs better than other competitive charts when the location shift is unknown but falls within an expected range.

Some CUSUM charts for count data can be found in Saghir and Lin (2014), for monitoring one or both parameters of the COM-Poisson distribution, in He *et al.* (2014), for monitoring linear drifts in Poisson rates, based on a dynamic estimation of the process mean level, and in Rakitzis *et al.* (2016), for monitoring zero-inflated binomial processes. Recently, Lombard *et al.* (2017) developed and analyzed the performance of distribution-free CUSUM control charts based on sequential ranks to detect changes in the mean direction and dispersion of angular data, which are of great importance to monitor several periodic phenomena that arise in many research areas.

To update the literature review in Jensen *et al.* (2006) about the effects of parameter estimation on control chart properties, Psarakis *et al.* (2014) provided some recent discussions on this topic. We also mention the works of Gandy and Kvaloy (2013) and Saleh *et al.* (2016), that suggest the design of CUSUM charts with a controlled conditional performance to reduce the effect of the Phase I estimation, avoiding at the same time the use of large amount of data. Such charts are designed with an in-control ARL that exceeds a desired value with a predefined probability, while guaranteeing a reduced effect on the out of control performance of the chart.

In this paper we study the CUSUM- \tilde{X} chart with known and estimated parameters for monitoring the mean value of a normal process, a topic, as far as we know, not yet studied in the literature, apart from simulation. The CUSUM- \tilde{X} chart is the most simple alternative to the CUSUM- \bar{X} in terms of efficiency/robustness, when there is some chance of having small disturbances in the process. For instance, it is possible to have a small percentage of outliers or contamination along time, that does not affect the process location, and therefore the chart must not signal in such cases. It is important to note that the goal of this paper is not to monitor the capability of a capable but unstable process (as this has already been done in Castagliola and Vannman (2008) and Castagliola *et al.* (2009)), but to monitor the median of a process that must remain stable for ensuring the quality of the products. The paper has three aims. The first aim is to present the Markov chain methodology for the computation of the distribution and the moments of the run length for the known and the estimated parameters case. The second aim is to evaluate the performance of the CUSUM- \tilde{X} chart in the known and estimated parameters case in terms of the average run length and standard deviation of the run length when the process is in- and out-of-control.

The third aim is to help practitioners in the implementation of the CUSUM- \tilde{X} chart by giving the optimal pair of parameters for the chart with estimated parameters to behave like the one with known parameters.

The outline of the paper is the following. In Section 2, we present the definition and the main properties of the CUSUM- \tilde{X} chart when the process parameters are known, along with the Markov chain methodology dedicated to the computation of the run length distribution of the chart and its moments. In Section 3, we study the case of estimated parameters for the CUSUM- \tilde{X} chart and we also provide the modified Markov chain methodology for the computation of the run length properties. A comparison between CUSUM- \tilde{X} with known v.s. estimated parameters is provided in Section 4. Finally, a detailed example is given in Section 5, followed by some conclusions and recommendations in Section 6.

2. THE CUSUM- \tilde{X} CHART WITH KNOWN PARAMETERS

In this paper we will assume that $Y_{i,1}, \dots, Y_{i,n}$, $i = 1, 2, \dots$ is a Phase II sample of n independent normal $N(\mu_0 + \delta\sigma_0, \sigma_0)$ random variables where i is the subgroup number, μ_0 and σ_0 are the in-control mean value and standard deviation, respectively, and δ is the parameter representing the standardized mean shift, i.e. the process is assumed to be in-control (out-of-control) if $\delta = 0$ ($\delta \neq 0$).

The upper-sided CUSUM- \tilde{X} chart for detecting an increase in the process mean plots

$$(2.1) \quad Z_i^+ = \max(0, Z_{i-1}^+ + \bar{Y}_i - \mu_0 - k_z^+)$$

against i , for $i = 1, 2, \dots$ where \bar{Y}_i is the mean value of the quality variable for sample number i . The starting value is $Z_0^+ = z_0^+ \geq 0$ and k_z^+ is a constant. A signal is issued at the first i for which $Z_i^+ \geq h_z^+$, where h_z^+ is the upper control limit. The corresponding lower-sided CUSUM- \tilde{X} chart for detecting a decrease in the process mean plots

$$(2.2) \quad Z_i^- = \min(0, Z_{i-1}^- + \bar{Y}_i - \mu_0 + k_z^-)$$

against i , for $i = 1, 2, \dots$ where k_z^- is a constant and the starting value is $Z_0^- = z_0^- \leq 0$. The chart signals at the first i for which $Z_i^- \leq -h_z^-$, where $-h_z^-$ is the lower control limit. There is a certain way to compute the values of k_z^+ , k_z^- and h_z^+ , h_z^- which is related to the distribution of Y_i 's. The textbook of Hawkins and Olwell (1999) is an excellent reference on this subject.

Now, let \tilde{Y}_i be the sample median of subgroup i , i.e.

$$\tilde{Y}_i = \begin{cases} Y_{i,((n+1)/2)} & \text{if } n \text{ is odd} \\ \frac{Y_{i,(n/2)} + Y_{i,(n/2+1)}}{2} & \text{if } n \text{ is even} \end{cases}$$

where $Y_{i,(1)}, Y_{i,(2)}, \dots, Y_{i,(n)}$ is the ascendant ordered i -th subgroup. As the sample median is easier and faster to compute when the sample size n is an odd value, without loss of generality, we will confine ourselves to this case for the rest of this paper.

The upper-sided CUSUM- \tilde{X} chart for detecting an increase in the process median is given by

$$(2.3) \quad U_i^+ = \max(0, U_{i-1}^+ + \tilde{Y}_i - \mu_0 - k^+)$$

where i is the sample number and \tilde{Y}_i is the sample median. The starting value is $U_0^+ = u_0^+ \geq 0$ and k^+ is a constant. A signal is issued at the first i for which $U_i^+ \geq h^+$, where h^+ is the upper control limit. The corresponding lower-sided CUSUM- \tilde{X} chart for detecting a decrease in the process median plots

$$(2.4) \quad U_i^- = \min(0, U_{i-1}^- + \tilde{Y}_i - \mu_0 + k^-)$$

against i , for $i = 1, 2, \dots$ where k^- is a constant and the starting value is $U_0^- = u_0^- \leq 0$. The chart signals at the first i for which $U_i^- \leq -h^-$, where $-h^-$ is the lower control limit.

The mean value (*ARL*) and the standard deviation (*SDRL*) of the Run Length distribution are two common measures of performance of control charts that will be used in this work to design the CUSUM median chart. However we note that other methodologies recently appeared in the literature for the design of CUSUM charts. Li and Wang (2010), He *et al.* (2014) and Wang and Huang (2016), among others, suggested to design the CUSUM chart with the reference parameter dynamically adjusted according to the current estimate of the process shift, in order to improve the sensitivity of the chart to detect a wide range of shifts. Ryu *et al.* (2010) proposed the design of CUSUM charts based on the expected weighted run length, a measure of performance more appropriate than the usual ARL given that the magnitude of the shift is practically unknown. These interesting approaches are promising and will be explored in a future work.

As in the classical approach proposed by Brook and Evans (1972), the Run Length distribution of the upper-sided CUSUM- \tilde{X} chart with known parameters can be obtained by considering a Markov chain with states denoted as $\{0, 1, \dots, r\}$, where state r is the absorbing state (the computations for the lower-sided CUSUM- \tilde{X} chart can be done accordingly). The interval from 0 to h^+ is partitioned into r subintervals $(H_j - \Delta, H_j + \Delta]$, $j \in \{0, \dots, r-1\}$, each

of them centered in $H_j = (2j + 1)\Delta$ (the representative value of state j), with $\Delta = \frac{h^+}{2r}$. The Markov chain is in transient state $j \in \{0, \dots, r - 1\}$ for sample i if $U_i^+ \in (H_j - \Delta, H_j + \Delta]$, otherwise it is in the absorbing state.

Let \mathbf{Q} be the (r, r) submatrix of probabilities $Q_{j,k}$ corresponding to the r transient states defined for the upward CUSUM- \tilde{X} chart, i.e.

$$\mathbf{Q} = \begin{pmatrix} Q_{0,0} & Q_{0,1} & \cdots & Q_{0,r-1} \\ Q_{1,0} & Q_{1,1} & \cdots & Q_{1,r-1} \\ \vdots & \vdots & \vdots & \vdots \\ Q_{r-1,0} & Q_{r-1,1} & \cdots & Q_{r-1,r-1} \end{pmatrix}.$$

By definition, we have $Q_{j,k} = P(U_i^+ \in (H_k - \Delta, H_k + \Delta] | U_{i-1}^+ = H_j)$, where $j \in \{0, \dots, r - 1\}$ and $k \in \{1, \dots, r - 1\}$. This is actually equivalent to $Q_{j,k} = P(H_k - \Delta < \tilde{Y} + H_j - \mu_0 - k^+ \leq H_k + \Delta)$. This equation can be written as

$$\begin{aligned} Q_{j,k} &= P\left(\tilde{Y} \leq H_k - H_j + \Delta + k^+ + \mu_0\right) - P\left(\tilde{Y} \leq H_k - H_j - \Delta + k^+ + \mu_0\right) \\ &= F_{\tilde{Y}}\left(H_k - H_j + \Delta + k^+ + \mu_0 | n\right) - F_{\tilde{Y}}\left(H_k - H_j - \Delta + k^+ + \mu_0 | n\right), \end{aligned}$$

where $F_{\tilde{Y}}(\dots | n)$ is the cumulative distribution function (c.d.f.) of the sample median $\tilde{Y}_i, i \in \{1, 2, \dots\}$. For the computation of $Q_{j,0}, j \in \{0, \dots, r - 1\}$ we have that

$$\begin{aligned} Q_{j,0} &= P\left(\tilde{Y} \leq -H_j + \Delta + k^+ + \mu_0\right) \\ &= F_{\tilde{Y}}\left(-H_j + \Delta + k^+ + \mu_0 | n\right). \end{aligned}$$

The c.d.f. of the sample median \tilde{Y} is given by

$$F_{\tilde{Y}}(y | n) = F_{\beta}\left(\Phi\left(\frac{y - \mu_0}{\sigma_0} - \delta\right) \middle| \frac{n + 1}{2}, \frac{n + 1}{2}\right),$$

where $\Phi(x)$ and $F_{\beta}(x | a, b)$ are the c.d.f. of the standard normal distribution and the beta distribution with parameters (a, b) (here, we have $a = b = \frac{n+1}{2}$), respectively.

Let $\mathbf{q} = (q_0, q_1, \dots, q_{r-1})^T$ be the $(r, 1)$ vector of initial probabilities associated with the r transient states $\{0, \dots, r - 1\}$, where

$$q_j = \begin{cases} 0 & \text{if } U_0^+ \notin (H_j - \Delta, H_j + \Delta] \\ 1 & \text{if } U_0^+ \in (H_j - \Delta, H_j + \Delta] \end{cases}.$$

Using this method, the Run Length (RL) properties of the CUSUM- \tilde{X} chart with known parameters can be accurately evaluated if the number r of subintervals in matrix \mathbf{Q} is sufficiently large. In this paper, we have fixed $r = 200$. Using the results in Neuts (1981) or Latouche and Ramaswami(1999) concerning

the fact that the number of steps until a Markov chain reaches the absorbing state is a *Discrete PHase-type* (or DPH) random variable, the probability mass function (p.m.f.) $f_{RL}(\ell)$ and the c.d.f. $F_{RL}(\ell)$ of the RL of the CUSUM- \tilde{X} chart with known parameters are respectively equal to

$$\begin{aligned} f_{RL}(\ell) &= \mathbf{q}^T \mathbf{Q}^{\ell-1} \mathbf{c}, \\ F_{RL}(\ell) &= 1 - \mathbf{q}^T \mathbf{Q}^{\ell} \mathbf{1}, \end{aligned}$$

where $\mathbf{c} = \mathbf{1} - \mathbf{Q}\mathbf{1}$ with $\mathbf{1} = (1, 1, \dots, 1)^T$. Using the moment properties of a DPH random variable also allows to obtain the mean (ARL), the second non-central moment $E2RL = E(RL^2)$ and the standard-deviation ($SDRL$) of the RL

$$\begin{aligned} ARL &= \nu_1 \\ E2RL &= \nu_1 + \nu_2 \\ SDRL &= \sqrt{E2RL - ARL^2}, \end{aligned}$$

where ν_1 and ν_2 are the first and second factorial moments of the RL , i.e.

$$\begin{aligned} \nu_1 &= \mathbf{q}^T (\mathbf{I} - \mathbf{Q})^{-1} \mathbf{1}, \\ \nu_2 &= 2\mathbf{q}^T (\mathbf{I} - \mathbf{Q})^{-2} \mathbf{Q}\mathbf{1}. \end{aligned}$$

3. THE CUSUM- \tilde{X} CHART WITH ESTIMATED PARAMETERS

In real applications the in-control process mean value μ_0 and the standard deviation σ_0 are usually unknown. In such cases they have to be estimated from a Phase I data set, having $i = 1, \dots, m$ subgroups $\{X_{i,1}, \dots, X_{i,n}\}$ of size n . Following Montgomery's (2009, p. 193 and p. 238) recommendations, these subgroups must be formed from observations taken in a time-ordered sequence in order to allow the estimation of between-sample variability, i.e., the process variability over time. The observations within a subgroup must be taken at the same time from a single and stable process, or at least as closely as possible to guarantee independence between them, to allow the estimation of the within-sample variability, i.e., the process variability at a given time. Here we assume that there is independence within and between subgroups, and also that $X_{i,j} \sim N(\mu_0, \sigma_0)$. The estimators that are usually used for μ_0 and σ_0 are

$$(3.1) \quad \hat{\mu}_0 = \frac{1}{m} \sum_{i=1}^m \bar{X}_i,$$

$$(3.2) \quad \hat{\sigma}_0 = \frac{1}{c_4(n)} \left(\frac{1}{m} \sum_{i=1}^m S_i \right),$$

where \bar{X}_i and S_i are the sample mean and the sample standard deviation of subgroup i , respectively. Constant $c_4(n) = E(\frac{S_i}{\sigma_0})$ can be computed for different

sample sizes n under normality. Although these estimators are usually used in the mean (\bar{X}) chart, they are not a straightforward choice with the median chart. Keeping in mind that the median chart is based on order statistics, a more typical selection of estimators based on order statistics is the following

$$(3.3) \quad \hat{\mu}'_0 = \frac{1}{m} \sum_{i=1}^m \tilde{X}_i,$$

$$(3.4) \quad \hat{\sigma}'_0 = \frac{1}{d_2(n)} \left(\frac{1}{m} \sum_{i=1}^m R_i \right),$$

where \tilde{X}_i and $R_i = X_{i,(n)} - X_{i,(1)}$ are the sample median and the range of subgroup i , respectively, and $d_2(n) = E\left(\frac{R_i}{\sigma_0}\right)$ is a constant tabulated assuming a normal distribution. Instead of the range we could have considered an estimator for the standard deviation based on quantiles to achieve higher level of robustness against outliers. The analysis of the properties of such CUSUM median chart is cumbersome and we will apply this approach in a future work.

The standardised versions of the lower-sided and the upper-sided CUSUM- \tilde{X} chart with estimated parameters are given by

$$(3.5) \quad G_i^- = \min \left(0, G_{i-1}^- + \frac{\tilde{Y}_i - \hat{\mu}'_0}{\hat{\sigma}'_0} + k_g^- \right),$$

$$(3.6) \quad G_i^+ = \max \left(0, G_{i-1}^+ + \frac{\tilde{Y}_i - \hat{\mu}'_0}{\hat{\sigma}'_0} - k_g^+ \right),$$

respectively, where $G_0^- = g_0^- \leq 0$, $G_0^+ = g_0^+ \geq 0$ with k_g^- and k_g^+ being two constants to be fixed. For the lower-sided (upper-sided) CUSUM- \tilde{X} chart with estimated parameters a signal is issued at the first i for which $G_i^- \leq h_g^-$ ($G_i^+ \geq h_g^+$), where h_g^- (h_g^+) is the lower (upper) control limit.

Equations (3.5) and (3.6) can be equivalently written as

$$(3.7) \quad G_i^- = \min \left(0, G_{i-1}^- + \frac{\frac{\tilde{Y}_i - \mu_0}{\sigma_0} + \frac{\mu_0 - \hat{\mu}'_0}{\sigma_0}}{\frac{\hat{\sigma}'_0}{\sigma_0}} + k_g^- \right),$$

$$(3.8) \quad G_i^+ = \max \left(0, G_{i-1}^+ + \frac{\frac{\tilde{Y}_i - \mu_0}{\sigma_0} + \frac{\mu_0 - \hat{\mu}'_0}{\sigma_0}}{\frac{\hat{\sigma}'_0}{\sigma_0}} - k_g^+ \right).$$

If we define the random variables V and W as $V = \frac{\hat{\mu}'_0 - \mu_0}{\sigma_0}$ and $W = \frac{\hat{\sigma}'_0}{\sigma_0}$,

both G_i^+ and G_i^- can be written as

$$(3.9) \quad G_i^+ = \max \left(0, G_{i-1}^+ + \frac{\tilde{Y}_i - \mu_0}{\sigma_0} - V - k_g^+ \right),$$

$$(3.10) \quad G_i^- = \min \left(0, G_{i-1}^- + \frac{\tilde{Y}_i - \mu_0}{\sigma_0} + V + k_g^- \right).$$

Apparently, the decision about when a process is declared as out of control does not change.

Both $\hat{\mu}'_0$ and $\hat{\sigma}'_0$ are random variables, therefore V and W are also random variables. Assuming that $\hat{\mu}'_0$ and $\hat{\sigma}'_0$ have fixed values, which actually means that V and W have fixed values, the conditional p.m.f. (denoted as $\hat{f}_{RL}(\ell)$) of RL , the conditional c.d.f. (denoted as $\hat{F}_{RL}(\ell)$) of RL and the conditional factorial moments (denoted as $\hat{\nu}_1$ and $\hat{\nu}_2$) can be computed through the equations given in section 2. Therefore, if the joint p.d.f. $f_{(V,W)}(v, w|m, n)$ of V and W is known, then the unconditional p.d.f. $f_{RL}(\ell)$ and the unconditional c.d.f. $F_{RL}(\ell)$ of the Run Length of the upper-sided CUSUM- \tilde{X} chart with estimated parameters are equal to

$$(3.11) \quad f_{RL}(\ell) = \int_{-\infty}^{+\infty} \int_0^{+\infty} f_{(V,W)}(v, w|m, n) \hat{f}_{RL}(\ell) dw dv,$$

$$(3.12) \quad F_{RL}(\ell) = \int_{-\infty}^{+\infty} \int_0^{+\infty} f_{(V,W)}(v, w|m, n) \hat{F}_{RL}(\ell) dw dv.$$

Now we are ready to compute the unconditional ARL that is equal to

$$(3.13) \quad ARL = \int_{-\infty}^{+\infty} \int_0^{+\infty} f_{(V,W)}(v, w|m, n) \hat{\nu}_1 dw dv.$$

The unconditional $SDRL$ is derived using the well known relationship

$$(3.14) \quad SDRL = \sqrt{E2RL - ARL^2},$$

where

$$E2RL = \int_{-\infty}^{+\infty} \int_0^{+\infty} f_{(V,W)}(v, w|m, n) (\hat{\nu}_1 + \hat{\nu}_2) dw dv.$$

Assuming normality, it is known that \bar{X}_i and S_i^2 are two independent statistics. Consequently, $\hat{\mu}_0$ and $\hat{\sigma}_0$ in equations (3.1) and (3.2) are also independent statistics. However, \tilde{X}_i and R_i are dependent statistics and so are $\hat{\mu}'_0$ and $\hat{\sigma}'_0$ in equations (3.3) and (3.4). Hogg (1960) proved that “an odd location statistic like the sample median and an even scale location-free statistic like the sample range are uncorrelated when sampling from a symmetric distribution”. On account of the fact that we assume we are sampling from a normal distribution, the sample median \tilde{X}_i and the sample range R_i are uncorrelated statistics. Since $\hat{\mu}'_0$ and $\hat{\sigma}'_0$

are averaged quantities of \tilde{X}_i and R_i respectively, the central limit theorem can be used to conclude that their joint distribution asymptotically converges to a bivariate normal distribution as m increases. Moreover, the fact that these statistics are uncorrelated, leads us to the conclusion that the statistics $\hat{\mu}'_0$ and $\hat{\sigma}'_0$ are asymptotically independent (and so are V and W). Therefore, the joint p.d.f. $f_{(V,W)}(v, w|m, n)$ in equations (3.11)–(3.14) is well approximated by the product of the marginal p.d.f. $f_V(v|m, n)$ of V and $f_W(w|m, n)$ of W , i.e.

$$(3.15) \quad f_{(V,W)}(v, w|m, n) \simeq f_V(v|m, n) \times f_W(w|m, n).$$

To evaluate how large n has to be for equation (3.15) to hold, or instead, to get approximately independence between the median and the range statistics in case of normal data, we did some simulations, using the following algorithm.

1. We generated 100000 samples of size $n = 3, 5, 7, 9, 11, 13, 15$. Each observation X_{ij} ($i = 1, \dots, 100000$, $j = 1, \dots, n$) follows a $N(0,1)$ distribution;
2. Then, we computed the median (\tilde{X}) and the range (R) for these 100000 samples, i.e., we got the values \tilde{X}_i and R_i , $i = 1, \dots, 100000$.
3. Afterwards, we estimated the c.d.f. of the statistics \tilde{X} and R , and the joint c.d.f. of (\tilde{X}, R) , i.e., the functions
 - $F_{\tilde{X}}(xm) = P(\tilde{X} \leq xm)$, for several values of xm ,
 - $F_R(xr) = P(R \leq xr)$, for several values of xr ,
 - $F_{\tilde{X},R}(xm, xr) = P(\tilde{X} \leq xm \cap R \leq xr)$, for the combinations of (xm, xr) .
4. Finally, we computed the difference $|F_{\tilde{X},R}(xm, xr) - F_{\tilde{X}}(xm) \times F_R(xr)|$ for all the combinations of (xm, xr) , and we kept the maximum of these differences.

Table 1: Maximum difference between the joint c.d.f. and the product of the two marginal c.d.f. for some small sample sizes n .

Sample size n	$\max F_{\tilde{X},R}(xm, xr) - F_{\tilde{X}}(xm) \times F_R(xr) $
3	0.01134
5	0.00838
7	0.00640
9	0.00529
11	0.00481
13	0.00402
15	0.00402

The obtained results are presented in Table 1. As we can see, the difference between the joint c.d.f. and the product of the two marginal c.d.f.'s is very small and it gets smaller as n increases. This is not a proof of independence between \tilde{X} and R , but for sure these statistics seem to be almost independent for small values of n . We also notice that the estimates for the nominal values of the process are the average of the medians and of the ranges, and consequently, the convergence to independence is faster.

For the computation of $f_V(v|m, n)$ and $f_W(w|m, n)$ there is no known closed-form, however, suitable approximations with satisfactory results are provided in Castagliola and Figueiredo (2013):

- The marginal p.d.f. $f_V(v|m, n)$, can be computed through the equation

$$f_V(v|m, n) \simeq \frac{b}{\sqrt{(v - \delta)^2 + d^2}} \phi \left(b \sinh^{-1} \left(\frac{v - \delta}{d} \right) \right),$$

where $\phi(x)$ is the p.d.f. of the standard normal distribution, and

$$b = \sqrt{\frac{2}{\ln(\sqrt{2(\gamma_2(V) + 2)} - 1)}},$$

$$d = \sqrt{\frac{2\mu_2(V)}{\sqrt{2(\gamma_2(V) + 2)} - 2}},$$

with

$$\mu_2(V) \simeq \frac{1}{m} \left(\frac{\pi}{2(n + 2)} + \frac{\pi^2}{4(n + 2)^2} + \frac{\pi^2 \left(\frac{13}{24}\pi - 1 \right)}{2(n + 2)^3} \right),$$

$$\gamma_2(V) \simeq \frac{2(\pi - 3)}{m(n + 2)};$$

- The marginal p.d.f. $f_W(w|m, n)$, can be computed through the equation

$$f_W(w|m, n) \simeq \frac{2\nu d_2^2(n)w}{c^2} f_{\chi^2} \left(\frac{\nu d_2^2(n)w^2}{c^2} | \nu \right),$$

where $f_{\chi^2}(x|\nu)$ is the p.d.f. of the χ^2 distribution with ν degrees of freedom with

$$\nu \simeq \left(-2 + 2 \sqrt{1 + 2 \left(\frac{B}{A^2} + \frac{\left(-2 + 2\sqrt{1 + \frac{2B}{A^2}} \right)^3}{16} \right)} \right)^{-1},$$

$$c \simeq A \left(1 + \frac{1}{4\nu} + \frac{1}{32\nu^2} - \frac{1}{128\nu^3} \right),$$

and $A = d_2(n)$, $B = \frac{d_3^2(n)}{m}$ where $d_2(n)$ and $d_3(n)$ are constants tabulated for the normal case.

4. A COMPARISON

In this work we compute the mean and the standard deviation of the Run Length distribution under the assumption that the scheme starts at the specified initial value, and we denote them as $ARL_{0,m}$ and $SDRL_{0,m}$ when the process is in-control, and as $ARL_{1,m}$ and $SDRL_{1,m}$, when the process is out-of-control, where m is the number of samples used in Phase I. We compare several control charts implemented with known and estimated parameters, all with the same in-control ARL value, here assumed equal to $ARL_{0,m} = 370.4$. The chart that exhibits the best performance to detect a specific shift size δ among its counterparts is the one that has the smaller $ARL_{1,m}$ value for this specific shift. Due to the symmetry of the Gaussian distribution, the performance of the charts are similar to detect either an upward or a downward shift of the same magnitude in the process mean value. Therefore we only concentrate our analysis on the performance of the charts for $\delta \geq 0$.

Using equations (3.13) and (3.14) we computed the $ARL_{1,m}$ and $SDRL_{1,m}$ values for several combinations of the sample size n , the number of samples m , and the shift size δ . These values ($ARL_{1,m}, SDRL_{1,m}$) are present in Table 2, together with the optimal set of parameters (H, K) for the specific n and δ values. The value $m = +\infty$ is associated with the known parameters case. In the estimated parameters case the number of subgroups considered in the estimation is $m = 5, 10, 20, 50, 100$. The pairs (H, K) given in each line are optimal in the sense that, among all the possible values of H and K , the noted pair gives the smallest $ARL_{1,m}$ value for the case $m = +\infty$. Setting $h^- = h^+ = H$ and $k^- = k^+ = K$ in the CUSUM- \tilde{X} charts defined in (2.3) and (2.4) allow us to obtain a control chart with the ARL behavior described in Table 2. From Table 2 the following conclusions are easily observed:

- The $ARL_{1,m}$ and $SDRL_{1,m}$ values in the known and estimated parameters cases are significantly different when the shift size δ or the number of subgroups m is small. For example, in case of $n = 3$ and $\delta = 0.1$, the $ARL_{1,m}$ and $SDRL_{1,m}$ values are larger than 10^5 if $m = 5$, but for $m = \infty$ we have $ARL_{1,\infty} = 98.7$ and $SDRL_{1,\infty} = 69.9$.
- If we take $m = 20$ subgroups of size $n = 5$ for the estimation of the unknown process parameters, as usually happens, only for $\delta \geq 1$ we get $ARL_{1,m} \simeq ARL_{1,\infty}$.
- For δ small, even if m is relatively large, the $ARL_{1,m}$ values are larger than the ones obtained in the case of known parameters. See, for example, the case of $m = 100$ subgroups of size $n = 9$ (an overall sample of size $n \times m = 900$ observations): for $\delta = 0.1$, we get an $ARL_{1,m}$ value about 50% larger than the corresponding $ARL_{1,\infty}$.

Table 2: $ARL_{1,m}$, $SDRL_{1,m}$, H , K values for different combinations of n , m and δ .

δ	(H, K)	$n = 3$					
		$m = 5$	$m = 10$	$m = 20$	$m = 50$	$m = 100$	$m = \infty$
0.1	(8.003, 0.0501)	(> 10 ⁵ , > 10 ⁵)	(> 10 ⁵ , > 10 ⁵)	(> 10 ⁵ , > 10 ⁵)	(636.9, 60171.3)	(165.7, 659.3)	(98.7, 69.9)
0.2	(6.003, 0.0999)	(> 10 ⁵ , > 10 ⁵)	(> 10 ⁵ , > 10 ⁵)	(2723.1, > 10 ⁵)	(76.7, 407.1)	(55.9, 66.2)	(46.5, 30.4)
0.3	(4.813, 0.1497)	(> 10 ⁵ , > 10 ⁵)	(> 10 ⁵ , > 10 ⁵)	(98.3, 14407.4)	(34.4, 49.9)	(30.3, 25.5)	(27.5, 17.1)
0.5	(3.444, 0.2489)	(> 10 ⁵ , > 10 ⁵)	(63.6, 62917.8)	(17.7, 47.5)	(14.5, 11.3)	(13.9, 9.2)	(13.3, 7.7)
0.7	(2.666, 0.3478)	(1039.6, > 10 ⁵)	(11.8, 165.9)	(9.1, 8.6)	(8.4, 5.5)	(8.2, 4.9)	(8.0, 4.4)
1.0	(1.965, 0.4951)	(8.3, 2204.2)	(5.3, 5.7)	(4.9, 3.3)	(4.7, 2.7)	(4.7, 2.6)	(4.6, 2.4)
1.5	(1.319, 0.7432)	(2.8, 3.3)	(2.6, 1.7)	(2.5, 1.4)	(2.5, 1.3)	(2.5, 1.3)	(2.5, 1.2)
2.0	(0.934, 0.9963)	(1.7, 1.1)	(1.6, 0.9)	(1.6, 0.8)	(1.6, 0.8)	(1.6, 0.8)	(1.6, 0.8)

δ	(H, K)	$n = 5$					
		$m = 5$	$m = 10$	$m = 20$	$m = 50$	$m = 100$	$m = \infty$
0.1	(5.903, 0.0500)	(> 10 ⁵ , > 10 ⁵)	(> 10 ⁵ , > 10 ⁵)	(> 10 ⁵ , > 10 ⁵)	(275.2, 8232.2)	(115.9, 281.0)	(79.1, 54.7)
0.2	(4.269, 0.0999)	(> 10 ⁵ , > 10 ⁵)	(> 10 ⁵ , > 10 ⁵)	(266.2, 49881.9)	(48.3, 107.4)	(39.9, 38.2)	(35.1, 22.2)
0.3	(3.349, 0.1496)	(> 10 ⁵ , > 10 ⁵)	(829.8, > 10 ⁵)	(37.1, 474.7)	(23.4, 23.6)	(21.6, 16.0)	(20.2, 12.1)
0.5	(2.329, 0.2487)	(1035.0, > 10 ⁵)	(16.5, 316.0)	(11.3, 12.8)	(10.1, 6.9)	(9.8, 6.0)	(9.5, 5.3)
0.7	(1.767, 0.3473)	(14.5, 3462.5)	(6.9, 10.9)	(6.2, 4.5)	(5.9, 3.5)	(5.8, 3.2)	(5.7, 3.0)
1.0	(1.270, 0.4949)	(4.1, 11.7)	(3.6, 2.6)	(3.4, 2.0)	(3.3, 1.8)	(3.3, 1.7)	(3.3, 1.7)
1.5	(0.812, 0.7467)	(1.9, 1.3)	(1.8, 1.0)	(1.8, 0.9)	(1.8, 0.9)	(1.8, 0.9)	(1.7, 0.9)
2.0	(0.511, 0.9990)	(1.3, 0.6)	(1.2, 0.5)	(1.2, 0.5)	(1.2, 0.5)	(1.2, 0.5)	(1.2, 0.4)

δ	(H, K)	$n = 7$					
		$m = 5$	$m = 10$	$m = 20$	$m = 50$	$m = 100$	$m = \infty$
0.1	(4.749, 0.0500)	(> 10 ⁵ , > 10 ⁵)	(> 10 ⁵ , > 10 ⁵)	(27291.2, > 10 ⁵)	(170.3, 2633.7)	(90.9, 166.3)	(67.0, 45.5)
0.2	(3.348, 0.0999)	(> 10 ⁵ , > 10 ⁵)	(17933.4, > 10 ⁵)	(100.8, 5616.9)	(36.2, 54.4)	(31.6, 27.1)	(28.6, 17.8)
0.3	(2.586, 0.1496)	(> 10 ⁵ , > 10 ⁵)	(123.7, 36763.2)	(24.0, 102.8)	(18.0, 15.6)	(17.0, 11.8)	(16.2, 9.5)
0.5	(1.762, 0.2487)	(60.7, 50167.3)	(10.5, 42.8)	(8.5, 7.4)	(7.8, 5.0)	(7.7, 4.5)	(7.5, 4.1)
0.7	(1.316, 0.3472)	(7.0, 114.4)	(5.1, 4.9)	(4.7, 3.1)	(4.6, 2.6)	(4.5, 2.5)	(4.5, 2.3)
1.0	(0.926, 0.4963)	(3.0, 3.3)	(2.7, 1.8)	(2.6, 1.5)	(2.6, 1.4)	(2.6, 1.3)	(2.6, 1.3)
1.5	(0.557, 0.7484)	(1.5, 0.8)	(1.4, 0.7)	(1.4, 0.7)	(1.4, 0.6)	(1.4, 0.6)	(1.4, 0.6)
2.0	(0.286, 0.9998)	(1.1, 0.3)	(1.1, 0.3)	(1.1, 0.3)	(1.1, 0.3)	(1.1, 0.3)	(1.1, 0.3)

δ	(H, K)	$n = 9$					
		$m = 5$	$m = 10$	$m = 20$	$m = 50$	$m = 100$	$m = \infty$
0.1	(4.007, 0.0500)	(> 10 ⁵ , > 10 ⁵)	(> 10 ⁵ , > 10 ⁵)	(7728.6, > 10 ⁵)	(123.1, 1171.1)	(75.7, 115.1)	(58.7, 39.3)
0.2	(2.770, 0.0999)	(> 10 ⁵ , > 10 ⁵)	(2714.7, > 10 ⁵)	(58.7, 1396.9)	(29.4, 35.5)	(26.4, 21.0)	(24.3, 14.9)
0.3	(2.114, 0.1496)	(24658.7, > 10 ⁵)	(48.4, 4394.4)	(18.2, 42.2)	(14.8, 11.7)	(14.2, 9.4)	(13.6, 7.9)
0.5	(1.416, 0.2487)	(19.3, 2775.4)	(7.9, 15.3)	(6.9, 5.3)	(6.5, 4.0)	(6.4, 3.6)	(6.3, 3.4)
0.7	(1.044, 0.3474)	(5.0, 20.6)	(4.1, 3.3)	(3.9, 2.4)	(3.8, 2.1)	(3.7, 2.0)	(3.7, 1.9)
1.0	(0.721, 0.4976)	(2.4, 2.0)	(2.2, 1.3)	(2.2, 1.2)	(2.2, 1.1)	(2.1, 1.1)	(2.1, 1.1)
1.5	(0.399, 0.7493)	(1.3, 0.6)	(1.2, 0.5)	(1.2, 0.5)	(1.2, 0.5)	(1.2, 0.5)	(1.2, 0.5)
2.0	(0.140, 0.9999)	(1.0, 0.2)	(1.0, 0.2)	(1.0, 0.1)	(1.0, 0.1)	(1.0, 0.1)	(1.0, 0.1)

- Moreover, the practical result referred in Quesenberry (1993), that an overall sample of size $n \times m = 400$ enables to design control charts with estimated control limits with a similar performance to the corresponding chart with true limits does not hold in the case of CUSUM- \tilde{X} charts (see, for instance, the $ARL_{1,m}$ and $SDRL_{1,m}$ values for small values of δ and, in particular, for $\delta = 0.1$).
- However, as the number of samples m increases the $ARL_{1,m}$ and $SDRL_{1,m}$ values converge to the values of the known parameters case, for each shift, although very slowly. In particular, when δ becomes large, the difference between the $ARL_{1,m}$ and $SDRL_{1,m}$ values in the known and estimated parameters cases tends to be non-significant. But the CUSUM charts are more attractive and efficient than the Shewhart charts for

detecting small changes, and thus, it is important to determine optimal parameters H' and K' in order to guarantee the desired performance even for m or δ small.

For completeness, in Table 3 we also present the in-control $ARL_{0,m}$ and $SDRL_{0,m}$ values for the same pairs (H, K) considered in Table 2. As in Table 2, we observe again that, as m increases, the $ARL_{0,m}$ and $SDRL_{0,m}$ values converge very slowly to the known parameters case values. As we can observe more than $m = 100$ samples are often needed to implement charts with known and estimated parameters with similar performance.

Table 3: $ARL_{0,m}$, $SDRL_{0,m}$, H , K values for several pairs of n and m when the process is in-control.

(H, K)	$n = 3$				
	$m = 5$	$m = 10$	$m = 20$	$m = 50$	$m = 100$
(8.003, 0.0501)	($> 10^5, > 10^5$)	($> 10^5, > 10^5$)	($> 10^5, > 10^5$)	(15540.8, $> 10^5$)	(1429.9, 16918.4)
(6.003, 0.0999)	($> 10^5, > 10^5$)	($> 10^5, > 10^5$)	($> 10^5, > 10^5$)	(2989.3, $> 10^5$)	(865.0, 3768.4)
(4.813, 0.1497)	($> 10^5, > 10^5$)	($> 10^5, > 10^5$)	($> 10^5, > 10^5$)	(70529.8, $> 10^5$)	(682.9, 1905.6)
(3.444, 0.2489)	($> 10^5, > 10^5$)	($> 10^5, > 10^5$)	(5501.3, $> 10^5$)	(853.9, 3474.9)	(544.3, 1032.7)
(2.666, 0.3478)	($> 10^5, > 10^5$)	(48309.8, $> 10^5$)	(2109.7, 49864.4)	(663.5, 1754.4)	(488.1, 777.8)
(1.965, 0.4951)	($> 10^5, > 10^5$)	(5812.4, $> 10^5$)	(1114.4, 7639.5)	(548.4, 1066.1)	(447.5, 622.0)
(1.319, 0.7432)	(20921.5, $> 10^5$)	(1593.5, 24654.9)	(704.3, 2093.0)	(470.9, 717.4)	(416.6, 515.1)
(0.934, 0.9963)	(3608.3, $> 10^5$)	(952.1, 4648.1)	(572.8, 1184.7)	(437.6, 587.0)	(402.1, 466.6)
(H, K)	$n = 5$				
	$m = 5$	$m = 10$	$m = 20$	$m = 50$	$m = 100$
(5.903, 0.0500)	($> 10^5, > 10^5$)	($> 10^5, > 10^5$)	($> 10^5, > 10^5$)	(7577.1, $> 10^5$)	(1192.1, 9102.0)
(4.269, 0.0999)	($> 10^5, > 10^5$)	($> 10^5, > 10^5$)	($> 10^5, > 10^5$)	(1883.9, 25530.6)	(747.4, 2380.3)
(3.349, 0.1496)	($> 10^5, > 10^5$)	($> 10^5, > 10^5$)	($> 10^5, > 10^5$)	(1092.2, 6198.7)	(605.9, 1346.7)
(2.329, 0.2487)	($> 10^5, > 10^5$)	(33318.5, $> 10^5$)	(2239.1, 38019.5)	(691.1, 1873.1)	(498.4, 812.6)
(1.767, 0.3473)	($> 10^5, > 10^5$)	(5584.4, $> 10^5$)	(1183.7, 7401.6)	(565.8, 1128.0)	(454.9, 643.8)
(1.270, 0.4949)	(17364.4, $> 10^5$)	(1774.2, 22295.1)	(756.6, 2351.7)	(485.6, 767.9)	(423.1, 534.6)
(0.812, 0.7467)	(2555.9, 62815.4)	(872.2, 3340.4)	(555.7, 1078.3)	(433.5, 571.3)	(400.4, 461.0)
(0.511, 0.9990)	(1497.0, 14155.3)	(696.7, 1890.8)	(500.9, 824.3)	(416.6, 511.7)	(392.6, 435.7)
(H, K)	$n = 7$				
	$m = 5$	$m = 10$	$m = 20$	$m = 50$	$m = 100$
(4.749, 0.0500)	($> 10^5, > 10^5$)	($> 10^5, > 10^5$)	($> 10^5, > 10^5$)	(5134.7, $> 10^5$)	(1060.8, 6361.6)
(3.348, 0.0999)	($> 10^5, > 10^5$)	($> 10^5, > 10^5$)	(33886.4, $> 10^5$)	(1474.5, 12982.1)	(683.9, 1848.5)
(2.586, 0.1496)	($> 10^5, > 10^5$)	($> 10^5, > 10^5$)	(5777.0, $> 10^5$)	(918.3, 3825.0)	(564.8, 1115.2)
(1.762, 0.2487)	($> 10^5, > 10^5$)	(10278.2, $> 10^5$)	(1538.4, 13243.4)	(619.3, 1401.7)	(474.6, 715.3)
(1.316, 0.3472)	(48309.2, $> 10^5$)	(2766.0, 57099.5)	(921.1, 3705.8)	(521.7, 913.9)	(438.0, 583.3)
(0.926, 0.4963)	(4897.4, $> 10^5$)	(1170.3, 6580.6)	(638.6, 1505.7)	(457.4, 656.4)	(411.1, 495.0)
(0.557, 0.7484)	(1671.5, 17793.4)	(734.8, 2135.8)	(514.2, 880.8)	(421.0, 526.9)	(394.7, 442.4)
(0.286, 0.9998)	(1282.1, 9130.3)	(652.2, 1599.9)	(485.7, 760.6)	(411.7, 494.9)	(390.3, 428.3)
(H, K)	$n = 9$				
	$m = 5$	$m = 10$	$m = 20$	$m = 50$	$m = 100$
(4.007, 0.0500)	($> 10^5, > 10^5$)	($> 10^5, > 10^5$)	($> 10^5, > 10^5$)	(3935.5, $> 10^5$)	(974.1, 4940.8)
(2.770, 0.0999)	($> 10^5, > 10^5$)	($> 10^5, > 10^5$)	(17455.0, $> 10^5$)	(1253.9, 8409.7)	(642.6, 1558.6)
(2.114, 0.1496)	($> 10^5, > 10^5$)	($> 10^5, > 10^5$)	(3795.7, $> 10^5$)	(819.6, 2826.8)	(538.3, 983.3)
(1.416, 0.2487)	($> 10^5, > 10^5$)	(5549.9, $> 10^5$)	(1230.8, 7364.1)	(576.9, 1169.2)	(459.4, 657.7)
(1.044, 0.3474)	(14888.5, $> 10^5$)	(1880.1, 19722.5)	(790.3, 2495.0)	(494.9, 799.0)	(427.2, 546.5)
(0.721, 0.4976)	(2879.4, 59731.6)	(946.2, 3818.3)	(580.8, 1186.6)	(441.5, 598.3)	(404.1, 472.5)
(0.399, 0.7493)	(1428.2, 11824.2)	(685.3, 1800.5)	(497.4, 808.7)	(415.6, 508.2)	(392.2, 434.3)
(0.140, 0.9999)	(1233.9, 8245.2)	(641.1, 1533.9)	(481.7, 744.4)	(410.4, 490.3)	(389.7, 426.3)

Since the out-of-control $ARL_{1,m}$ values are clearly different in the known and in the estimated parameters case, it is important to determine the number m of Phase I samples we should consider to get approximately the same out-of-control $ARL_{1,m}$ values in both cases, using the same optimal control chart parameters H and K displayed in Table 2. Thus, with

$$\Delta = \frac{|ARL_{1,m} - ARL_{1,\infty}|}{ARL_{1,\infty}}$$

denoting the relative difference between the out-of-control $ARL_{1,m}$ (estimated parameter case) and $ARL_{1,\infty}$ (known parameter case) values, we computed the minimum value of m satisfying $\Delta \leq 0.05$ or $\Delta \leq 0.01$. The obtained minimum number m of Phase I samples is given in Table 4 for some values of n and δ . From this table we observe that:

- The value of m satisfying $\Delta < 0.05$ or $\Delta < 0.01$ can be very large, for instance, $m > 100$ if $\delta \leq 0.3$. In particular, for $\delta = 0.1$ and $n = 9$, to have $\Delta < 0.05$ we must consider at least 417 samples.
- For the most common subgroup sample size $n = 5$, the number of subgroups must be 569 for very small shifts (say, for $\delta = 0.1$), and consequently $n \times m$ will be 2845, much larger than 400, as suggested by Quesenberry (1993).
- The number of the required samples decreases with the increase shift size δ . We also observe that the number of subgroups m that are needed decreases with the sample size n , but the number of observations of the overall sample needed for the estimation, $n \times m$, also increases.

Table 4: Minimum number m of Phase I samples required to satisfy $\Delta = 0.05$ (left value) and $\Delta = 0.01$ (right value) when the process is out of control.

δ	$\Delta = (0.05, 0.01)$			
	$n = 3$	$n = 5$	$n = 7$	$n = 9$
0.1	(711, 3339)	(569, 2669)	(479, 2241)	(417, 1945)
0.2	(319, 1485)	(237, 1101)	(191, 881)	(159, 737)
0.3	(181, 835)	(129, 597)	(101, 467)	(83, 385)
0.5	(81, 371)	(55, 257)	(43, 197)	(35, 159)
0.7	(45, 209)	(31, 141)	(23, 107)	(19, 87)
1.0	(25, 109)	(17, 73)	(13, 55)	(9, 45)
1.5	(11, 51)	(7, 35)	(5, 27)	(5, 21)
2.0	(7, 33)	(5, 21)	(3, 11)	(3, 5)

As a conclusion, we observe that in most of the cases a very large number m of Phase I samples is needed so that the charts with known and estimated parameters have the same ARL performance. But this requirement is in general

very hard to handle in practice for economical and logistic reasons. Therefore, for fixed values of m and n , the determination of adequate control chart parameters, taking into consideration the variability introduced by the parameters estimation is very challenging.

Table 5: Optimal values for H' , K' , ARL_1 and $SDRL_1$ subject to the constraint $ARL_0 = 370.4$.

δ	$n = 3$				
	$m = 5$	$m = 10$	$m = 20$	$m = 50$	$m = 100$
0.1	(1.747, 0.01) (121.6, $> 10^5$)	(2.739, 0.01) (101.3, 32840.0)	(3.976, 0.01) (88.0, 3910.3)	(5.886, 0.01) (81.9, 444.6)	(7.379, 0.01) (84.6, 151.3)
0.2	(1.747, 0.01) (45.2, 55017.6)	(2.739, 0.01) (35.7, 4544.5)	(3.976, 0.01) (32.7, 441.4)	(5.886, 0.01) (35.6, 55.1)	(7.024, 0.02) (40.5, 32.8)
0.3	(1.747, 0.01) (19.9, 11098.1)	(2.739, 0.01) (17.1, 682.8)	(3.976, 0.01) (18.0, 63.5)	(5.679, 0.02) (22.4, 17.2)	(5.197, 0.09) (25.0, 17.5)
0.5	(1.747, 0.01) (6.9, 493.1)	(2.739, 0.01) (7.7, 24.1)	(3.976, 0.01) (9.7, 6.8)	(3.678, 0.16) (11.9, 7.6)	(3.517, 0.21) (12.6, 7.8)
0.7	(1.747, 0.01) (4.0, 28.2)	(2.739, 0.01) (5.2, 3.8)	(2.960, 0.15) (6.5, 4.0)	(2.805, 0.27) (7.4, 4.3)	(2.671, 0.32) (7.7, 4.5)
1.0	(1.747, 0.01) (2.7, 2.0)	(2.197, 0.16) (3.5, 2.0)	(2.050, 0.35) (4.1, 2.3)	(1.997, 0.44) (4.4, 2.4)	(1.938, 0.48) (4.5, 2.4)
1.5	(1.476, 0.18) (1.8, 0.9)	(1.406, 0.49) (2.1, 1.1)	(1.331, 0.64) (2.3, 1.2)	(1.329, 0.70) (2.4, 1.2)	(1.346, 0.71) (2.4, 1.2)
2.0	(1.122, 0.46) (1.3, 0.6)	(1.040, 0.74) (1.5, 0.7)	(0.977, 0.88) (1.5, 0.7)	(0.952, 0.95) (1.6, 0.7)	(0.947, 0.97) (1.6, 0.8)

δ	$n = 5$				
	$m = 5$	$m = 10$	$m = 20$	$m = 50$	$m = 100$
0.1	(1.536, 0.01) (104.2, 46570.9)	(2.294, 0.01) (82.1, 7398.7)	(3.230, 0.01) (68.8, 1439.2)	(4.684, 0.01) (64.1, 232.1)	(5.827, 0.01) (67.4, 92.9)
0.2	(1.536, 0.01) (34.4, 8053.3)	(2.294, 0.01) (26.1, 916.8)	(3.230, 0.01) (24.3, 136.6)	(4.684, 0.01) (27.7, 27.7)	(4.642, 0.05) (31.4, 23.8)
0.3	(1.536, 0.01) (14.4, 1446.5)	(2.294, 0.01) (12.6, 128.4)	(3.230, 0.01) (13.8, 20.3)	(3.791, 0.06) (17.3, 12.0)	(3.511, 0.11) (18.8, 12.4)
0.5	(1.536, 0.01) (5.3, 57.0)	(2.294, 0.01) (6.1, 6.3)	(2.671, 0.08) (7.6, 4.8)	(2.429, 0.19) (8.8, 5.3)	(2.378, 0.22) (9.1, 5.3)
0.7	(1.536, 0.01) (3.3, 4.3)	(2.044, 0.07) (4.2, 2.5)	(1.898, 0.22) (4.9, 2.9)	(1.811, 0.30) (5.4, 3.0)	(1.806, 0.32) (5.5, 3.0)
1.0	(1.457, 0.05) (2.3, 1.2)	(1.328, 0.31) (2.8, 1.5)	(1.314, 0.40) (3.0, 1.6)	(1.282, 0.46) (3.2, 1.6)	(1.295, 0.47) (3.2, 1.6)
1.5	(0.948, 0.38) (1.5, 0.7)	(0.858, 0.59) (1.6, 0.8)	(0.835, 0.67) (1.7, 0.8)	(0.816, 0.72) (1.7, 0.8)	(0.805, 0.74) (1.7, 0.9)
2.0	(0.680, 0.62) (1.1, 0.4)	(0.613, 0.80) (1.2, 0.4)	(0.563, 0.90) (1.2, 0.4)	(0.553, 0.94) (1.2, 0.4)	(0.510, 0.99) (1.2, 0.4)

Table 6: (Cont'd) Optimal values for H' , K' , ARL_1 and $SDRL_1$ subject to the constraint $ARL_0 = 370.4$.

δ	$n = 7$				
	$m = 5$	$m = 10$	$m = 20$	$m = 50$	$m = 100$
0.1	(1.355, 0.01) (88.9, 19453.2)	(1.988, 0.01) (67.5, 3754.2)	(2.771, 0.01) (56.1, 844.7)	(3.987, 0.01) (53.4, 147.8)	(4.940, 0.01) (57.2, 65.7)
0.2	(1.355, 0.01) (26.4, 2921.9)	(1.988, 0.01) (20.2, 389.8)	(2.771, 0.01) (19.6, 65.3)	(3.987, 0.01) (23.3, 18.3)	(3.597, 0.06) (26.0, 18.5)
0.3	(1.355, 0.01) (10.8, 464.4)	(1.988, 0.01) (10.1, 48.2)	(2.771, 0.01) (11.6, 10.9)	(2.862, 0.08) (14.2, 9.4)	(2.671, 0.12) (15.2, 9.7)
0.5	(1.355, 0.01) (4.3, 16.0)	(1.988, 0.01) (5.2, 3.5)	(1.917, 0.13) (6.3, 3.9)	(1.842, 0.20) (7.0, 4.1)	(1.829, 0.22) (7.3, 4.1)
0.7	(1.355, 0.01) (2.9, 2.0)	(1.488, 0.14) (3.6, 2.0)	(1.397, 0.25) (4.0, 2.2)	(1.346, 0.31) (4.3, 2.3)	(1.326, 0.33) (4.4, 2.3)
1.0	(1.038, 0.18) (2.0, 1.1)	(0.997, 0.34) (2.3, 1.2)	(0.943, 0.43) (2.4, 1.3)	(0.933, 0.47) (2.5, 1.3)	(0.936, 0.48) (2.5, 1.3)
1.5	(0.668, 0.46) (1.3, 0.5)	(0.600, 0.62) (1.3, 0.6)	(0.573, 0.69) (1.4, 0.6)	(0.581, 0.71) (1.4, 0.6)	(0.568, 0.73) (1.4, 0.6)
2.0	(0.429, 0.69) (1.0, 0.2)	(0.386, 0.82) (1.1, 0.2)	(0.316, 0.93) (1.1, 0.2)	(0.320, 0.95) (1.1, 0.2)	(0.318, 0.96) (1.1, 0.2)

δ	$n = 9$				
	$m = 5$	$m = 10$	$m = 20$	$m = 50$	$m = 100$
0.1	(1.220, 0.01) (76.9, 11291.0)	(1.774, 0.01) (57.1, 2426.2)	(2.459, 0.01) (47.5, 568.6)	(3.521, 0.01) (46.4, 103.4)	(4.350, 0.01) (50.3, 50.5)
0.2	(1.220, 0.01) (21.0, 1482.2)	(1.774, 0.01) (16.5, 210.2)	(2.459, 0.01) (16.7, 37.5)	(3.143, 0.03) (20.3, 15.3)	(2.889, 0.07) (22.4, 15.5)
0.3	(1.220, 0.01) (8.6, 208.6)	(1.774, 0.01) (8.6, 23.5)	(2.459, 0.01) (10.1, 7.4)	(2.226, 0.10) (12.2, 8.0)	(2.224, 0.12) (12.9, 7.8)
0.5	(1.220, 0.01) (3.8, 6.7)	(1.685, 0.03) (4.6, 2.7)	(1.532, 0.15) (5.4, 3.2)	(1.465, 0.21) (5.9, 3.3)	(1.438, 0.23) (6.1, 3.4)
0.7	(1.220, 0.01) (2.6, 1.4)	(1.164, 0.18) (3.1, 1.7)	(1.095, 0.27) (3.4, 1.8)	(1.056, 0.32) (3.6, 1.9)	(1.061, 0.33) (3.6, 1.9)
1.0	(0.825, 0.23) (1.8, 0.9)	(0.758, 0.38) (1.9, 1.0)	(0.724, 0.45) (2.0, 1.1)	(0.736, 0.47) (2.1, 1.1)	(0.735, 0.48) (2.1, 1.1)
1.5	(0.519, 0.48) (1.1, 0.4)	(0.446, 0.63) (1.2, 0.4)	(0.401, 0.71) (1.2, 0.5)	(0.382, 0.75) (1.2, 0.5)	(0.379, 0.76) (1.2, 0.5)
2.0	(0.315, 0.68) (1.0, 0.1)	(0.250, 0.82) (1.0, 0.1)	(0.205, 0.90) (1.0, 0.1)	(0.206, 0.92) (1.0, 0.1)	(0.213, 0.92) (1.0, 0.1)

In this paper we computed, for fixed values of m and n , new chart parameters denoted as (H', K') , in order to achieve the desired in-control performance, i.e. such that, for fixed values of m and n , we have $ARL(m, n, H', K', \delta = 0) = 370.4$ and, for a fixed value of δ , $ARL(m, n, H', K', \delta)$ is the smallest out-of-control $ARL_{1,m}$. These new pairs of constants are given in Tables 5 and 6 for various

combinations of n , m and δ , and might be used as the chart parameters in the CUSUM- \tilde{X} charts defined in (3.8) and (3.7), i.e., we might choose $h_g^- = h_g^+ = H'$ and $k_g^- = k_g^+ = K'$. In each cell of Tables 5 and 6, the two numbers in the first row are new chart parameters (H', K') , and the two numbers in the second row are the $ARL_{1,m}$ and $SDRL_{1,m}$ values. As we can observe, with these constants (H', K') determined with the unconditional run length distribution, we can guarantee the same performance of the corresponding chart implemented with known process parameters, or even a better performance, except in the cases of $m = 5, 10$ and $\delta = 0.1$. For the shift size δ that must be quickly detected, the values presented in Tables 5 and 6 allow the practitioners to easily implement the most efficient median CUSUM control chart. For instance, if $n = 5$ and $m = 20$, the optimal CUSUM- \tilde{X} chart to detect a shift of size $\delta = 1$ must be designed with the constants $H' = 1.314$ and $K' = 0.40$. With these chart parameters we get the values $ARL_{1,m} = 3$ and $SDRL_{1,m} = 1.6$.

5. AN ILLUSTRATIVE EXAMPLE

In order to illustrate the use of the CUSUM- \tilde{X} chart when the parameters are estimated, let us consider the same example as the one in Castagliola and Figueiredo (2013), i.e. a 125g yogurt cup filling process for which the quality characteristic Y is the weight of each yogurt cup. The Phase I dataset used in this example consists of $m = 10$ subgroups of size $n = 5$ plotted in the left part of Figure 1 with “o”. From this Phase I dataset, using (3.3) and (3.4), we obtain $\hat{\mu}'_0 = 125.02$ and $\hat{\sigma}'_0 = 0.864$. According to the quality practitioner in charge of this process, a shift of $0.5\sigma_0$ (i.e., $\delta = 0.5$) in the process position should be interpreted as a signal that something is going wrong in the production. For $m = 10$, $n = 5$ and $\delta = 0.5$, Table 5 suggests to use $K' = 2.294$ and $H' = 0.01$.

The Phase II dataset used in this example consists of $m = 30$ subgroups of size $n = 5$ plotted in the right part of Figure 1 with “•”. The first 15 subgroups are supposed to be in-control while the last 15 subgroups are supposed to have a smaller yogurt weight, and thus, to be out-of-control. In Figure 2, we plotted the statistics G_i^- and G_i^+ corresponding to (3.5) and (3.6). This figure shows that the 7th first subgroups are in-control but, from subgroups #8 to #15, the process experiences a light out-of-control situation (increase) as the points “•” corresponding to the G_i^+ 's are above the upper limit $K' = 2.294$. During subgroups #16 and #17, the process returns to the in-control state but, as expected, suddenly experiences a new strong out-of-control situation (decrease) as the points “o” corresponding to the G_i^- 's are now below the lower limit $-K' = -2.294$.

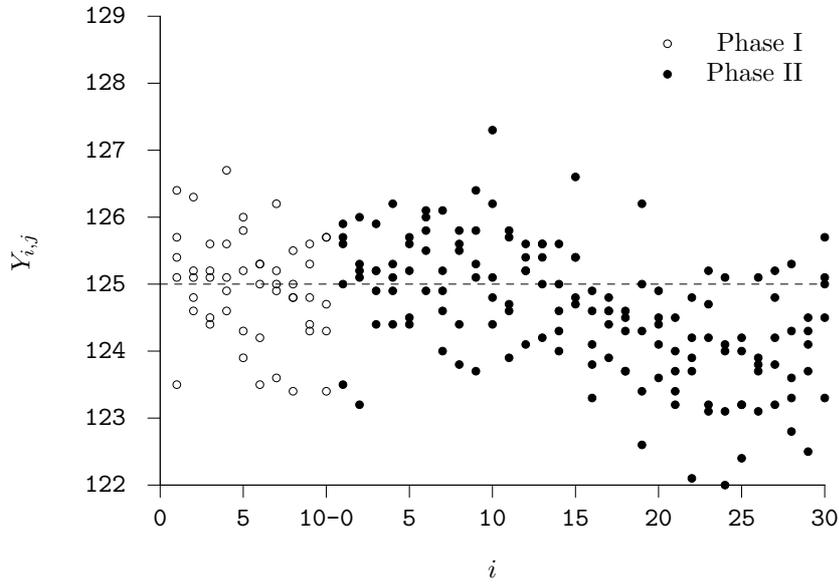


Figure 1: Phase I and Phase II samples corresponding to the 125g yogurt cup filling process.

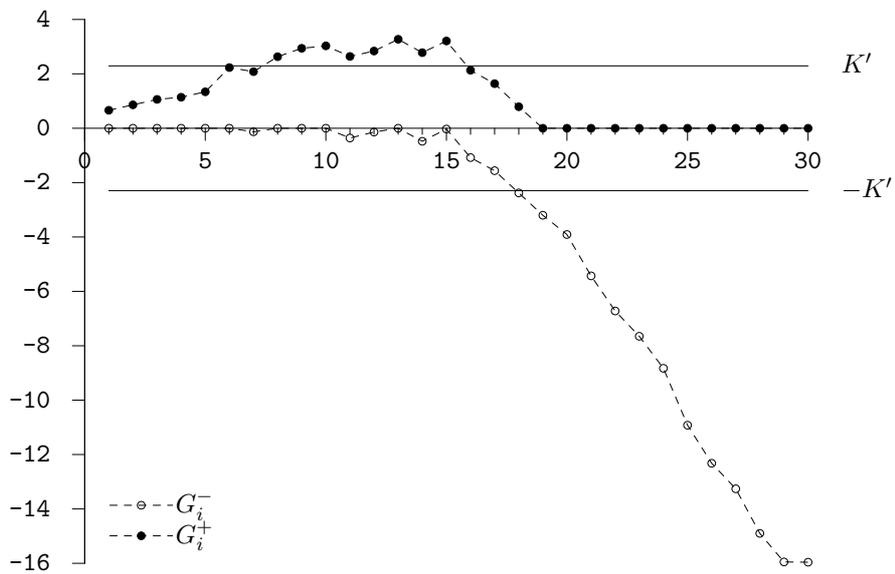


Figure 2: CUSUM- \tilde{X} chart corresponding to the Phase II sample of Figure 1.

6. CONCLUSIONS

Although the CUSUM- \tilde{X} chart has already been proposed by Yang *et al.* (2010) and Nazir *et al.* (2013a), in both of these papers the authors have only investigated the performance of this chart through simulations and compared its performance with other charts in terms of robustness. Moreover, in the implementation of the chart Yang *et al.* (2010) assumed the process parameters known, and Nazir *et al.* (2013a) also considered them fixed and known, after a prior estimation of such parameters through the use of different location and scale estimators. But they really did not analyze the effect of the parameters estimation in the performance of the chart in comparison with the performance of the corresponding chart implemented with true parameters, the main objective of our paper. We used a Markov chain methodology to compute the run length distribution and the moments of the CUSUM- \tilde{X} chart in order to study its performance when the parameters are known and estimated. In this paper we present several tables that allow us to observe that the chart implemented with estimated parameters exhibits a completely different performance in comparison to the one of the chart implemented with known parameters. We also provide modified chart parameters that allow the practitioners to implement the CUSUM- \tilde{X} chart with estimated control limits with a given desired in-control performance. More specifically, the main conclusions are: a) if the shift size δ or the number of samples m used in the estimation is small, there is a large difference between the $ARL_{1,m}$ and the $SDRL_{1,m}$ values obtained in the known and estimated parameters cases, b) for δ small, even if m is relatively large, the $ARL_{1,m}$ values are larger than the ones obtained in the case of known parameters, c) the $ARL_{1,m}$ and $SDRL_{1,m}$ values converge to the values of the known parameters case as the number of samples m increases, d) the number of subgroups m to have a relative difference between the out-of-control ARL values in the known and estimated parameters cases less than 5% or 1% can be very large, and depends on the value of δ , e) it is possible to obtain new chart parameters in order to achieve a desired in-control performance. As a general conclusion, the CUSUM- \tilde{X} chart can be a valuable alternative chart for practitioners since it is simpler than the CUSUM- \bar{X} chart. The fact that it is robust against outliers, contamination or small deviations from normality is another advantage.

ACKNOWLEDGMENTS

Research partially supported by National Funds through **FCT** — Fundação para a Ciência e a Tecnologia, project UID/MAT/00006/2013 (CEA/UL). We also acknowledge the valuable suggestions from a referee.

REFERENCES

- [1] ABUJIYA, M.R.; LEE, M.H. and RIAZ, M. (2015). Increasing the sensitivity of cumulative sum charts for location, *Quality Reliability and Engineering International*, **31**, 1035–1051.
- [2] AL-SABAH, W.S. (2010). Cumulative sum statistical control charts using ranked set sampling data, *Pakistan Journal of Statistics*, **26**(2), 365–378.
- [3] BROOK, D. and EVANS, D.A. (1972). An approach to the probability distribution of CUSUM run length, *Biometrika*, **59**(3), 539–549.
- [4] CASTAGLIOLA, P. and VANNMAN, K. (2008). Average run length when monitoring capability indices using EWMA, *Quality and Reliability European Engineering International*, **24**(8), 941–955.
- [5] CASTAGLIOLA, P. and FIGUEIREDO, F.O. (2013). The median chart with estimated parameters, *European Journal of Industrial Engineering*, **7**(5), 594–614.
- [6] CASTAGLIOLA, P.; MARAVELAKIS, P.; PSARAKIS, S. and VANNMAN, K. (2009). Monitoring capability indices using run rules, *Journal of Quality in Maintenance Engineering*, **15**(4), 358–370.
- [7] CHATTERJEE, S. and QIU, P. (2009). Distribution-free cumulative sum control charts using bootstrap-based control limits, *Annals of Applied Statistics*, **3**, 349–369.
- [8] CHOWDHURY, S.; MUKHERJEE A. and CHAKRABORTI, S. (2015). Distribution free phase II CUSUM control chart for joint monitoring of location and scale, *Quality and Reliability Engineering International*, **31**(1), 135–151.
- [9] GANDY, A. and KVALOY, J.T. (2013). Guaranteed conditional performance of control charts via bootstrap methods, *Scandinavian Journal of Statistics*, **40**, 647–668.
- [10] GRAHAM, M.A.; CHACKRABORTI, S. and MUKHERJEE, A. (2014). Design and implementation of CUSUM exceedance control charts for unknown location, *International Journal of Production Research*, **52**, 5546–5564.
- [11] HAWKINS, D.M. and OLWELL, D.H. (1999). *Cumulative Sum Charts and Charting for Quality Improvement*, Springer-Verlag New York, Inc..
- [12] HE, F.; SHU, L. and TSUI, K-L. (2014). Adaptive CUSUM charts for monitoring linear drifts in Poisson rates, *International Journal of Production Economics*, **148**, 14–20.

- [13] HOGG, R.V. (1960). Certain uncorrelated statistics, *Journal of the American Statistical Association*, **55**(290), 265–267.
- [14] JENSEN, W.A.; JONES-FARMER, L.A.; CHAMP, C.W. and WOODALL, W.H. (2006). Effects of parameter estimation on control chart properties: A literature review, *Journal of Quality Technology*, **38**, 349–364.
- [15] LATOUCHE, G. and RAMASWAMI, V. (1999). *Introduction to Matrix Analytic Methods in Stochastic Modelling*, ASA SIAM.
- [16] LI, Z. and WANG, Z. (2010). Adaptive CUSUM of the Q chart, *International Journal of Production Research*, **48**(5), 1287–1301.
- [17] LI, Z.; LUO, Y. and WANG, Z. (2010a). CUSUM of Q chart with variable sampling intervals for monitoring the process mean, *International Journal of Production Research*, **48**(16), 4861–4876.
- [18] LI, S.-Y.; TANG, L.C. and NG S.-H. (2010b). Nonparametric CUSUM and EWMA control charts for detecting mean shifts, *Journal of Quality Technology*, **42**(2), 209–226.
- [19] LIU, L.; TSUNG, F. and ZHANG, J. (2014). adaptive nonparametric CUSUM scheme for detecting unknown shifts in location, *International Journal of Production Research*, **52**(6), 1592–1606.
- [20] LOMBARD, F.; HAWKINS, D.M. and POTGIETER, C.J. (2017). Sequential rank CUSUM charts for angular data, *Computational Statistics and Data Analysis*, **105**, 268–279.
- [21] MONTGOMERY, D.C. (2009). *Introduction to Statistical Quality Control, 6th edition*, John Wiley & Sons, Inc..
- [22] MUKHERJEE, A.; GRAHAM, M.A. and CHAKRABORTI, S. (2013). Distribution free exceedance CUSUM control charts for location, *Communications in Statistics – Simulation and Computation*, **42**(5), 1153–1187.
- [23] NAZIR, H.Z.; RIAZ, M.; DOES, R.J.M.M. and ABBAS, N. (2013a). Robust CUSUM control charting, *Quality Engineering*, **25**(3), 211–224.
- [24] NAZIR, H.Z.; RIAZ, M. and DOES, R.J.M.M. (2013b). Robust CUSUM control charting for process dispersion, *Quality Reliability and Engineering International*, **31**, 369–379.
- [25] NENES, G. and TAGARAS, G. (2006). The economically designed CUSUM chart for monitoring short production runs, *International Journal of Production Research*, **44**(8), 1569–1587.
- [26] NEUTS, M.F. (1981). *Matrix-Geometric Solutions in Stochastic Models: an Algorithmic Approach*, Dover Publications Inc..

- [27] OU, Y.; WU, Z. and TSUNG, F. (2011). A comparison study of effectiveness and robustness of control charts for monitoring process mean, *International Journal of Production Economics*, **135**, 479–490.
- [28] OU, Y.; WEN, D.; WU, Z. and KHOO, M.B.C. (2012). A comparison study on effectiveness and robustness of control charts for monitoring process mean and variance, *Quality Reliability and Engineering International*, **28**, 3–17.
- [29] OU, Y.; WU, Z.; LEE, K.M. and WU, K. (2013). An adaptive CUSUM chart with single sample size for monitoring process mean and variance, *Quality Reliability and Engineering International*, **29**, 1027–1039.
- [30] PAGE, E.S. (1954). Continuous inspection schemes, *Biometrika*, **41**(1–2), 100–115.
- [31] PSARAKIS, S.; VYNIYOU, A.K. and CASTAGLIOLA, P. (2014). Some recent developments on the effects of parameter estimation on control charts, *Quality and Reliability Engineering International*, **30**(8), 1113–1129.
- [32] QIU, P. and ZHANG, J. (2015). On phase II SPC in cases when normality is invalid, *Quality Reliability and Engineering International*, **31**, 27–35.
- [33] QUESENBERRY, C.P. (1993). The effect of sample size on estimated limits for \bar{X} and x control charts, *Journal of Quality Technology*, **25**(4), 237–247.
- [34] RAKITZIS, A.; MARAVELAKIS, P. and CASTAGLIOLA, P. (2016). CUSUM control charts for the monitoring of zero-Inflated Binomial Processes, *Quality and Reliability Engineering International*, **32**(2), 465–483.
- [35] RYU, J.-H.; WAN, H. and KIM, S. (2010). Optimal design of a CUSUM chart for a mean shift of unknown size, *Journal of Quality Technology*, **42**(3), 1–16.
- [36] SAGHIR, A. and LIN, Z. (2014). Cumulative sum charts for monitoring COM-Poisson processes, *Computers and Industrial Engineering*, **68**, 65–77.
- [37] SALEH, N.A.; ZWETSLOOT, I.M.; MAHMOUD, M.A. and WOODALL, W.H. (2016). CUSUM charts with controlled conditional performance under estimated parameters, *Quality Engineering*, **28**(4), 402–415.
- [38] SANIGA, E.M.; MCWILLIAMS, T.; DAVIS, D.J. and LUCAS, J.M. (2006). Economic control chart policies for monitoring variables, *International Journal of Productivity and Quality Management*, **1**(1), 116–138.
- [39] SANIGA, E.; LUCAS, J.; DAVIS, D. and MCWILLIAMS, T. (2012). Economic control chart policies for monitoring variables when there are two components of variance. In *Frontiers in Statistical Quality* (Lenz, Wilrich and Schmid, Eds.), **10**(1), 85–95.

- [40] TEOH, W.L.; KHOO, M.B.C.; CASTAGLIOLA, P. and CHAKRABORTI, S. (2014). Optimal design of the double sampling \bar{X} chart with estimated parameters based on median run length, *Computers & Industrial Engineering*, **67**, 104–115.
- [41] WANG, T. and HUANG, S. (2016). An adaptive multivariate CUSUM control chart for signaling a range of location shifts, *Communications in Statistics – Theory and Methods*, **45**(16), 4673–4691.
- [42] WANG, D.; ZHANG, L. and XIONG, Q. (2017). A non parametric CUSUM control chart based on the Mann–Whitney statistic, *Communications in Statistics – Theory and Methods*, **46**(10), 4713–4725.
- [43] WU, Z.; JIAO, J.; YANG, M.; LIU, Y. and WANG, Z. (2009). An enhanced adaptive CUSUM control chart, *IIE Transactions*, **41**, 642–653.
- [44] YANG, L.; PAI, S. and WANG, Y.R. (2010). A novel CUSUM median control chart. In *Proceedings of the International MultiConference of Engineers and Computer Scientists (IMECS)*, **3**, 1707–1710.

AP-OPTIMUM DESIGNS FOR MINIMIZING THE AVERAGE VARIANCE AND PROBABILITY-BASED OPTIMALITY

Authors: N.M. KILANY
– Faculty of Science, Menoufia University,
Menoufia, Egypt
neveenkilany@hotmail.com

W.A. HASSANEIN
– Faculty of Science, Tanta University,
Tanta, Egypt
wfanwar@science.tanta.edu.eg

Received: December 2016

Revised: February 2017

Accepted: April 2017

Abstract:

- The purpose of this paper is to introduce a new class of compound criteria and optimum designs that provide a specified balance between minimizing the average variance and high probability of a desired outcome. The proposed criterion called AP- optimality that combines A-optimality and P-optimality and address this issue for generalized linear models. An equivalence theorem for this criterion is provided and two numerical examples are presented for different GLMs to illustrate the achieved dual properties.

Key-Words:

- *optimum design; A-optimality; P-optimality; compound criteria.*

1. INTRODUCTION

The type of any design is always an option regardless of the type of model we wish to fit (for example, first order, first order plus some interactions, full quadratic, cubic, etc.) or the objective specified for the experiment. The design of experiments for generalized linear models (GLMs) has received considerable attention in recent years, for example the research by Woods *et al.* [9]. To some extent, this has been in response to design issues raised by researchers in experimental sciences, such as new technologies (for example genomics and areas of modern biology), where the inherent characteristics of data in these fields lead to the consideration of GLMs for analysis and consequently design. GLMs are non-linear models and, as such, pose substantial challenges in terms of design, in particular in the need to have information on the model parameters prior to designing an experiment to estimate these parameters. Much of the research into design for GLMs has concentrated on quite small models: one or two variables and ‘simple’ optimality criteria, such as D-optimality, which is concerned solely with parameter estimation. However, the paper by Woods *et al.* [9] investigated complex models for binary data with several variables over a number of models in the form of a compound criterion called product design optimality. Historically, most optimal design criteria have been concerned with parameter estimation, and more recently some have combined the notions of parameter estimation and model discrimination (for example, DT-optimality, Atkinson [1]). Examples of other compound criteria can be found in Waterhouse [8] where criteria are described that also yield designs that offer efficient parameter estimation and model discrimination.

A-optimality criterion corresponds to minimize the variance of the asymptotic distribution of the maximum likelihood estimate of that parameter, employed that criterion of optimality is the one that involves the use of Fisher’s information matrix. For linear models with one discrete factor and additive general regression term the problem of characterizing A-optimal design measures for inference on treatment effects, the regression parameters and all parameters will be considered. While, P-optimal design maximizes the average probability of success of a given design.

The aim of this paper is to derive method for designing experiments from which minimizing average variance of the parameter estimates can be obtained, while at the same time maximizing the probability of a particular event that is of importance to experimenter. This paper is organized as follows: Section 2 is devoted to represent the optimum design background. In Section 3, a simple review for A — and P — optimum designs is introduced. In Section 4, the AP-optimum design is proposed to achieve the dual goals of minimizing the average variance and maximizing the average of the probability of observing an outcome.

Moreover, the equivalence theorem is derived. Two numerical examples are given in Section 5 to illustrate the method and the value of the proposed criterion in meeting the dual aims.

2. OPTIMUM DESIGN PRELIMINARIES

Consider the generalized linear models GLMs

$$(2.1) \quad \begin{aligned} E(Y) &= \mu = g^{-1}(X\beta) \\ \eta &= g(X\beta) \end{aligned}$$

which is defined by the distribution of the response Y , a matrix of independent variables (predictors) X , a vector of unknown parameters β and a linear predictor η and two functions:

1. A link function $g(\cdot)$ that describes how the mean, $E(Y_i) = \mu_i$ depends on the linear predictor $g(\mu_i) = Y_i$.
2. A variance function that describes how the variance, $Var(Y_i)$ depends on the mean

$$(2.2) \quad Var(Y_i) = \phi(V(\mu))$$

where the dispersion parameter ϕ is a constant.

In GLMs, the errors or noise ϵ_i have relaxed assumptions where it may or may not have normal distribution. GLMs are commonly used to model binary or count data. Some common link functions are used such that the identity, logit, log and probit link to induce the traditional linear regression, logistic regression, Poisson regression models.

An approximate (continuous) design is represented by the probability measure ξ over the design space δ . If the design has trials at n distinct points in δ , it can be written as

$$(2.3) \quad \xi = \left\{ \begin{array}{cccc} x_1 & x_2 & \dots & x_n \\ w_1 & w_2 & \dots & w_n \end{array} \right\}$$

A design ξ defines, for $i = 1, \dots, n$, the vector of support-point $x_i \in \chi$ related to y_i , where χ is a compact experimental domain and the experimental weights w_i corresponding to each x_i , where $\sum_{i=1}^n w_i = 1$. The design space can be then expressed as $\delta = \{\xi_i \in X^n \times [0, 1]^n : \sum_{i=1}^n w_i = 1\}$. Such designs are called approximate or continuous designs.

3. A- AND P-OPTIMUM DESIGNS

3.1. A-optimum design

A-optimality criterion introduced by Chernoff [2]; who showed that the employed criterion of optimality is the one that involves the use of Fisher's information matrix. For the case where it is desired to estimate one of the p parameters in the information matrix, this criterion corresponds to minimize the variance of the asymptotic distribution of the maximum likelihood estimate of that parameter.

A-optimality minimizes the average variance of the parameter estimates. Alternatively, it can be expressed as the following form;

$$(3.1) \quad \Phi_A(\xi) = \min_{x_i, i=1, \dots, n} \text{tr}(X^T X)$$

For a discussion on an A-optimal designs for binary models, see Sitter and Wu [6], Zhu and Wong [11]. Yang [10] introduced A-optimal designs for generalized linear models with two parameters which are logistic, probit and double exponential models.

The equivalence theorem states that, the derivative function

$$(3.2) \quad f^T(x) M^{-2}(\boldsymbol{\theta}, \xi) f(x) \leq \text{tr}[M^{-1}(\boldsymbol{\theta}, \xi)], \quad x \in \chi$$

where M is the information matrix and the equality holds only if $\xi = \xi_A^*$, $x \in \xi_A^*$.

A-efficiency of a design ξ is defined as:

$$(3.3) \quad \text{Eff}_A(\xi) = \frac{\text{tr}[M^{-1}(\boldsymbol{\theta}, \xi_A^*)]}{\text{tr}[M^{-1}(\boldsymbol{\theta}, \xi)]}$$

where ξ_A^* is A-optimal.

3.2. P-optimum designs

McGree and Eccleston [5] have offered a P-optimality criterion, which is defined as a criterion that maximizes a function of the probability of observing a particular outcome. One of the forms of P-optimality which defined is concerned with the maximization of a weighted sum of the probabilities of success.

The form of this criterion is

$$(3.4) \quad \Phi_P(\xi) = \sum_{i=1}^n \pi_i(\boldsymbol{\theta}, \xi_i) w_i$$

where, $\pi_i(\boldsymbol{\theta}, \xi_i)$ is the i -th probability of success given by ξ_i and w_i is the experimental effort relating to the i -th support point. In this criterion, design weights have been included and will play a role in maximizing the probabilities.

Let ξ_P^* be the design maximizing (3.4). Under some regularity conditions, McGree and Eccleston [5] proved an equivalence theorem for P-optimum designs, in which the derivative function $\psi_P(x, \xi_P^*) \leq 0$, $x \in \chi$, where

$$(3.5) \quad \psi_{PA}(x, \xi_P^*) = \frac{\Phi_P(x) - \Phi_P(\xi_P^*)}{\Phi_P(\xi_P^*)}$$

is the directional derivative of $\Phi_P(\xi)$. The P -efficiency of a design ξ relative to the optimum design ξ_P^* is

$$(3.6) \quad Eff_{PA}(\xi) = \frac{\sum_{i=1}^n \pi_i(\boldsymbol{\theta}, \xi_i) w_i}{\sum_{i=1}^n \pi_i(\boldsymbol{\theta}, \xi_{PA}^*) w_i}.$$

4. AP-OPTIMUM DESIGN

There is a situation when an experimenter may be interested to achieve multiple objectives. For this aim, we will construct a design that combine A-optimality with P-optimality. The new criterion will be called AP-optimality. This criterion offers a method of achieving minimizing the average variance and a high probability of a desired outcome.

The AP-optimality criterion is given by the following weighted geometric mean of efficiencies:

$$(4.1) \quad \{Eff_A(\xi)\}^\alpha \{Eff_P(\xi)\}^{1-\alpha} = \left(\frac{tr[M^{-1}(\boldsymbol{\theta}, \xi_A^*)]}{tr[M^{-1}(\boldsymbol{\theta}, \xi)]} \right)^\alpha \left(\frac{\sum_{i=1}^n \pi_i(\boldsymbol{\theta}, \xi_i) w_i}{\sum_{i=1}^n \pi_i(\boldsymbol{\theta}, \xi_P^*) w_i} \right)^{1-\alpha}$$

where the coefficients $0 \leq \alpha \leq 1$. When $\alpha = 0$, we obtain P -optimality and when $\alpha = 1$, we obtain A -optimality. To clarify the structure of the design criterion, take log in (4.1) yields

$$(4.2) \quad \begin{aligned} & \alpha \log(tr [M^{-1}(\boldsymbol{\theta}, \xi_A^*)]) - \alpha \log(tr [M^{-1}(\boldsymbol{\theta}, \xi)]) + \\ & + (1 - \alpha) \log \sum_{i=1}^n \pi_i(\boldsymbol{\theta}, \xi_i) w_i - (1 - \alpha) \log \sum_{i=1}^n \pi_i(\boldsymbol{\theta}, \xi_P^*) w_i. \end{aligned}$$

The terms involving ξ_A^* and ξ_P^* are constants when a maximum is found over ξ . Many bibliographical references presented the concept of this maximization method such as Dette [4]; Atkinson [1]; Tommasi [7]; and McGree and Eccleston [5]. So that the criterion to be maximized is

$$(4.3) \quad \Phi_{AP}(\xi) = -\alpha \log(\text{tr} [M^{-1}(\boldsymbol{\theta}, \xi)]) + (1 - \alpha) \log \sum_{i=1}^n \pi_i(\boldsymbol{\theta}, \xi_i) w_i.$$

The negative sign for the first term on the right hand side of (4.3) arises because the average variance is minimized. Designs maximizing (4.3) are called AP-optimum and denoted ξ_{AP}^* .

The equivalence theorem is stated as follows:

Theorem 4.1. *For AP-optimal design, ξ_{AP}^* , the following three statements are equivalent.*

1. *A necessary and sufficient condition for a design ξ_{AP}^* to be AP-optimum is fulfillment of the inequality $\psi_{AP}(x, \xi_{AP}^*) \leq 1$, $x \in \chi$, where the derivative function of (4.3) is given by*

$$(4.4) \quad \psi_{AP}(x, \xi_{AP}^*) = \alpha \left(\frac{f^T(x) M^{-2}(\boldsymbol{\theta}, \xi_{AP}^*) f(x)}{\Phi_A(\xi_{AP}^*)} \right) + (1 - \alpha) \left(\frac{\Phi_P(x) - \Phi_P(\xi_{AP}^*)}{\Phi_P(\xi_{AP}^*)} \right).$$

2. *The upper bound of $\psi_{AP}(x, \xi_{AP}^*)$ is achieved at the points of the optimum design.*
3. *For any non optimum design ξ , that is a design for which $\Phi_{AP}(\xi) < \Phi_{AP}(\xi_{AP}^*)$, $\sup_{x \in \chi} \psi_{AP}(x, \xi_{AP}^*) > 1$.*

Proof: Since $0 \leq \alpha \leq 1$, ψ_{AP} is a convex combination of logarithm of two design criteria. Therefore, the AP-criterion satisfies the conditions of convex optimum design theory and an equivalence theorem applies. Because of the way the terms in (4.4) have been scaled, the upper bound of ψ_{AP} over $x \in \chi$ is one, achieved at the points of the optimum design. Furthermore, ψ_{AP} is the linear combination of the directional derivatives given by A-optimality and P-optimality. Thus, the theorem has been proved. □

5. APPLICATIONS TO GENERALIZED LINEAR MODELS

In this Section, the AP-optimality criterion is applied to two types of generalized linear models, Logit and probit models, for binary data. The data were based on the work given by in Corana *et al.* [3]. The A-, P-, and the proposed compound AP-efficiencies are calculated and the optimal designs are obtained to illustrate the main objective of the compound criterion that allow both minimizing the average variance of the parameter estimates plus increasing the probability of the desired outcome.

Example 5.1. Logit Model

The considering logit model has two main factor effects besides the interaction with initial parameter estimates $\theta = [1, -2, 1, -1]^T$ with $x_j \in [-1, 1]$ as follows:

$$(5.1) \quad \text{Log} \left(\frac{\pi}{1 - \pi} \right) = 1 - 2x_1 + x_2 - x_1x_2.$$

AP-optimal designs and their A- and P-efficiencies for $\alpha = 0, 0.25, 0.5, 0.75, 1$ are obtained and presented in Table 1.

Table 1: AP-optimum design and their A- and P-efficiencies for the Logit model at different values of α .

α	x_1	x_2	w_i	π_i	A_{eff}	P_{eff}
0	-1.000	1.000	1.000	0.9933	—	1
0.25	1.0000	-1.000	0.0835	0.2689	0.822183	0.8060
	0.8020	1.000	0.0999	0.3999		
	-1.000	-1.000	0.1983	0.7311		
	-0.3980	1.000	0.6182	0.9596		
0.5	1.000	-1.000	0.1570	0.2689	1	0.6644
	1.000	1.000	0.1600	0.2889		
	-1.000	-1.000	0.2802	0.7311		
	-0.1059	1.000	0.4028	0.9103		
0.75	1.000	1.000	0.2121	0.2689	0.741011	0.5826
	1.000	-1.000	0.2121	0.2689		
	-1.000	-1.000	0.2740	0.7311		
	0.0148	1.000	0.3017	0.8761		
1	1.000	-1.000	0.2500	0.2689	0.864115	0.5352
	1.000	1.000	0.2500	0.2689		
	-1.000	-1.000	0.2500	0.7311		
	0.0680	1.000	0.2500	0.8577		

Table 1 shows the designs that maximize the AP-criterion. It can be noticed that there is little changes in the design points with high variation in design weights. That is, the P_{eff} 's are increased through the given designs as well as the probability of success is increased. Figure 1 illustrates the A- and P-efficiencies for $\alpha = 0, 0.25, 0.5, 0.75$ and 1. The dot-dashed line represents the A-efficiency of the designs, and the solid line shows their P-efficiencies. The following A-optimal design has a P-efficiency of 0.6644.

$$\xi_A^* = \begin{pmatrix} 1.0000 & -1.000 & 0.1570 \\ 1.0000 & 1.000 & 0.1600 \\ -1.000 & -1.000 & 0.2802 \\ -0.1059 & 1.000 & 0.4028 \end{pmatrix}.$$

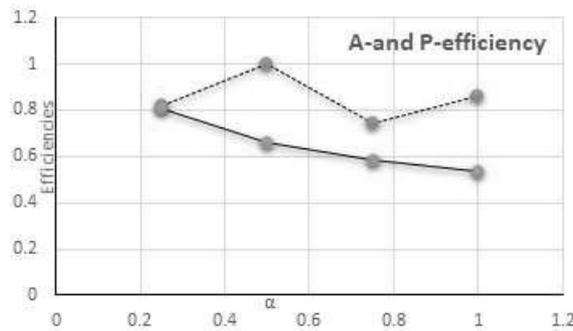


Figure 1: A- and P-efficiencies of AP-optimal designs for different values of α .

By using the AP-criterion and choosing $\alpha = 0.25$, we are able to increase the P-efficiency to 0.806, while achieving a A-efficiency of 0.822183. The AP-optimal design is

$$\xi_{AP}^* = \begin{pmatrix} 1.000 & -1.000 & 0.0835 \\ 0.802 & 1.000 & 0.0999 \\ -1.000 & -1.000 & 0.1983 \\ -0.398 & 1.000 & 0.6182 \end{pmatrix}.$$

Example 5.2. Probit Model

In the following Example, the AP-optimality criterion is applied to the probit model. The response variable is modelled via three main factor effects with initial parameters $\beta = [1, -0.5, 1, -1]$, with $x_j \in [-1, 1]$:

$$(5.2) \quad \Phi^{-1}(\pi) = 1 - 0.5 x_1 + x_2 - x_3.$$

Table 2 include the main results of the designs and their A- and P-efficiencies for $\alpha = 0, 0.2, 0.35, 0.5, 0.75, 1$. Figure 2 illustrates the A- and P-efficiencies

for $\alpha = 0, 0.2, 0.35, 0.5, 0.75, 1$. Using the compound criteria AP-criterion, at $\alpha = 0.5$, we can see that the A-efficiency and P-efficiencies have very close high efficiencies, 0.982127, 0.983084, respectively.

Table 2: AP-optimum design and their A- and P-efficiencies for the Probit model at different values of α .

α	x_1	x_2	x_3	w_i	π_i	A_{eff}	P_{eff}
0	-1.0000	-1.0000	1.0000	0.0708	0.3085	1	0.98427
	-0.6296	1.0000	-1.0000	0.1105	0.9988		
	-0.5423	1.0000	1.0000	0.1600	0.8980		
	-0.5423	-1.0000	-1.0000	0.1600	0.8980		
	-0.0186	-1.0000	1.0000	0.0658	0.1611		
	0.0186	1.0000	1.0000	0.0658	0.8389		
	0.0186	-1.0000	-1.0000	0.0658	0.8389		
	0.5423	-1.0000	1.0000	0.1600	0.1020		
	1.0000	-1.0000	-1.0000	0.0708	0.6915		
	1.0000	1.0000	1.0000	0.0708	0.6915		
0.2	-1.0000	-1.0000	1.0000	0.0608	0.3085	0.964075	1
	-0.5368	-1.0000	-1.0000	0.2315	0.8980		
	-0.5368	1.0000	1.0000	0.2315	0.8980		
	-0.5244	1.0000	-1.0000	0.1232	0.9987		
	0.5368	-1.0000	1.0000	0.2315	0.1020		
	1.0000	-1.0000	-1.0000	0.0608	0.6915		
	1.0000	1.0000	1.0000	0.0608	0.6915		
0.35	-1.0000	-1.0000	1.0000	0.0630	0.3085	0.979698	0.996276
	0.5027	1.0000	-1.0000	0.1213	0.9987		
	-0.4894	-1.0000	-1.0000	0.2299	0.8925		
	-0.4894	1.0000	1.0000	0.2299	0.8925		
	0.4894	-1.0000	1.0000	0.2299	0.1075		
	1.0000	-1.0000	-1.0000	0.0630	0.6915		
	1.0000	1.0000	1.0000	0.0630	0.6915		
0.5	-1.0000	-1.0000	1.0000	0.0618	0.3085	0.982127	0.983084
	-0.5140	1.0000	-1.0000	0.1085	0.9987		
	-0.4709	1.0000	1.0000	0.2144	0.8925		
	-0.4395	-1.0000	-1.0000	0.2241	0.8888		
	-0.0373	-1.0000	1.0000	0.0245	0.1635		
	0.0372	1.0000	1.0000	0.0245	0.8365		
	0.4709	-1.0000	1.0000	0.2144	0.1075		
	1.0000	-1.0000	-1.0000	0.0662	0.6915		
	1.0000	1.0000	1.0000	0.0618	0.6915		
0.75	-1.0000	-1.0000	1.0000	0.0504	0.3085	0.778262	0.988816
	-1.0000	-1.0000	-1.0000	0.0143	0.9332		
	-1.0000	1.0000	1.0000	0.0143	0.9332		
	-1.0000	1.0000	-1.0000	0.1306	0.0089		
	-0.0101	-1.0000	-1.0000	0.2251	0.8413		
	-0.0101	1.0000	1.0000	0.2251	0.8413		
	0.0101	-1.0000	1.0000	0.2251	0.1587		
	1.0000	-1.0000	1.0000	0.0143	0.0668		
	1.0000	-1.0000	-1.0000	0.0504	0.6915		
	1.0000	1.0000	1.0000	0.0504	0.6915		

(continues)

(continued)

α	x_1	x_2	x_3	w_i	π_i	A_{eff}	P_{eff}
0.9	-1.0000	-1.0000	1.0000	0.0533	0.3085	0.820936	0.983858
	-0.9942	1.0000	-1.0000	0.1197	0.9987		
	-0.8203	-1.0000	-1.0000	0.0212	0.9207		
	-0.8203	1.0000	1.0000	0.0212	0.9207		
	-0.0348	-1.0000	-1.0000	0.2189	0.8461		
	-0.0348	1.0000	1.0000	0.2189	0.8461		
	0.0348	-1.0000	1.0000	0.2189	0.1539		
	0.8203	-1.0000	1.0000	0.0212	0.0793		
	1.0000	1.0000	1.0000	0.0533	0.6915		
	1.0000	-1.0000	-1.0000	0.0533	0.6915		
1	-1.0000	-1.0000	1.0000	0.0550	0.3085	0.899469	0.970774
	-0.9344	1.0000	-1.0000	0.0924	0.9989		
	-0.6464	-1.0000	-1.0000	0.0515	0.9066		
	-0.6464	1.0000	1.0000	0.0515	0.9066		
	-0.0612	-1.0000	-1.0000	0.1960	0.8485		
	-0.0612	1.0000	1.0000	0.1960	0.8485		
	0.0612	-1.0000	1.0000	0.1960	0.1515		
	0.6464	-1.0000	1.0000	0.0515	0.0934		
	1.0000	-1.0000	-1.0000	0.0550	0.6915		
	1.0000	1.0000	1.0000	0.0550	0.6915		

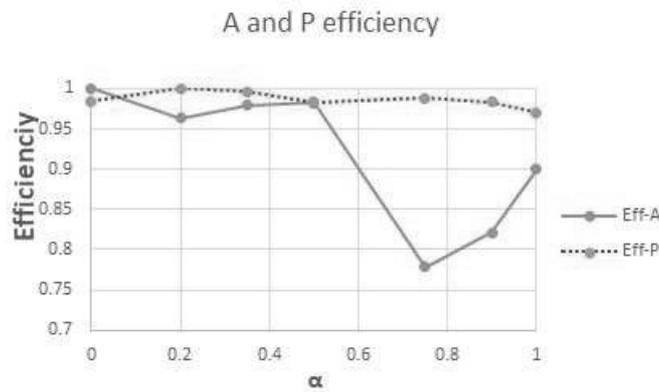


Figure 2: A- and P-efficiencies of AP-optimal designs for different values of α .

Hence, the AP-optimal design which satisfy the dual problem is obtained as:

$$\zeta_{AP}^* = \left\{ \begin{array}{ccc} -1.0000 & -1.0000 & 1.0000 \\ -0.5140 & 1.0000 & -1.0000 \\ -0.4709 & 1.0000 & 1.0000 \\ -0.4395 & -1.0000 & -1.0000 \\ -0.0373 & -1.0000 & 1.0000 \\ 0.0372 & 1.0000 & 1.0000 \\ 0.4709 & -1.0000 & 1.0000 \\ 1.0000 & -1.0000 & 1.0000 \\ 1.0000 & 1.0000 & 1.0000 \end{array} \right\}.$$

6. CONCLUSION

The criterion AP-optimum design introduced here provides a new compound criterion that yield minimum of the average variance of the parameter estimates plus a high probability of observing a particular outcome. The equivalence theorem is stated and proved for AP-optimum design. Two illustrated examples are presented for logit and probit models. The results indicate the potentiality of using the proposed AP-optimality criterion.

REFERENCES

- [1] ATKINSON, A.C. (2008). DT-optimum designs for model discrimination and parameter estimation, *Journal of Statistical Planning and Inference*, **138**, 56–64.
- [2] CHERNOFF, H. (1953). Locally optimal designs for estimating parameters, *Annals of Mathematical Statistics*, **24**, 586–602.
- [3] CORANA, A.; MARCHESI, M.; MARTINI, C. and RIDELLA, S. (1987). Minimizing multimodal functions of continuous variables with the ‘Simulated Annealing’ algorithm, *ACM Trans. Math. Software*, **13**, 262–280.
- [4] DETTE, H. (1993). On a mixture of the D- and D_1 -optimality criterion in polynomial regression, *Journal of Statistical Planning and Inference*, **35**, 233–249.
- [5] MCGREE, J.M. and ECCLESTON, J.A. (2008). Probability-based optimal design, *Australian and New Zealand Journal of Statistics*, **50**(1), 13–28.
- [6] SITTE, R. and WU, C. (1993). Optimal designs for binary response experiments: Fieller, D, and A criteria, *Scandinavian Journal of Statistics*, **20**, 329–342.
- [7] TOMMASI (2009). Optimal designs for both model discrimination and parameter estimation, *Journal of Statistical Planning and Inference*, **139**, 4123–4132.
- [8] WATERHOUSE, T.H. (2005). *Optimal Experimental Design for Nonlinear and Generalized Linear Models*, (PhD Thesis), University of Queensland, Australia.
- [9] WOODS, D.C.; LEWIS, S.M.; ECCLESTON, J.A. and RUSSELL, K.G. (2006). Designs for generalized linear models with several variables and model uncertainty, *Technometrics*, **48**, 284–292.
- [10] YANG, M. (2008). A-optimal designs for generalized linear models with two parameters, *Journal of Statistical Planning and Inference*, **138**, 624–641.
- [11] ZHU, W. and WONG, W. (2000). Multiple-objective designs in a dose-response experiment, *Journal of Biopharmaceutical Statistics*, **10**(1), 1–14.

AN INFORMATION THEORETICAL METHOD FOR ANALYZING UNREPLICATED DESIGNS WITH BINARY RESPONSE

Authors: KRYSTALLENIA DROSOU
– Department of Mathematics, National Technical University of Athens,
Athens, Greece
`drosou.kr@gmail.com`

CHRISTOS KOUKOUVINOS
– Department of Mathematics, National Technical University of Athens,
Athens, Greece
`ckoukouv@math.ntua.gr`

Received: February 2017

Revised: March 2017

Accepted: April 2017

Abstract:

- The analysis of unreplicated factorial designs constitutes a challenging but difficult issue since there are no degrees of freedom so as to estimate the error variance. In the present paper we propose a method for screening active effects in such designs, assuming Bernoulli distributed data rather than linear; something that hasn't received much attention yet. Specifically, we develop an innovating algorithm based on an information theoretical measure, the well-known symmetrical uncertainty, so that it can measure the relation between the response variable and each factor separately. The powerfulness of the proposed method is revealed via both, a thorough simulation study and a real data set analysis.

Key-Words:

- *two-level factorial designs; unreplicated experiments; generalized linear models; symmetrical uncertainty.*

AMS Subject Classification:

- 62-07, 62K15, 62J12.

1. INTRODUCTION

Factorial designs constitute a powerful tool especially in screening experiments where the goal is to identify the factors with a significant impact on the response of interest. Although two-level factorial designs are commonly used as experimental plans, the number of runs grows exponentially as the number of factors increases; thus, in case when the replication of the experiment is prohibitive due to economical or technical issues, unreplicated designs constitute an appropriate choice. Such designs are saturated; means that the number of examined factors d equals to $n - 1$, where n is the number of runs. As a result, the experimenter can estimate all the d main and interaction effects, but there are no degrees of freedom to estimate the error; therefore, the conventional analysis of variance (ANOVA) techniques cannot be applied.

Many methods, either theoretical or graphical ones, have been proposed to overcome the aforementioned problem. The standard method for identifying active effects in unreplicated designs is the probability plot of the effects, proposed by Daniel [7]. This approach consists of plotting the factor estimates on a normal or half-normal probability plot, where the inactive effects fall along a straight line while the active ones tend to fall off the line. The subjective nature of that method motivated many authors to provide more objective procedures. For a detailed review article, we refer the interested reader to Hamada and Balakrishnan [10]. Some important works include: Box and Meyer [5], Lenth [11], Dong [8], Chen and Kunert [6], Aboukalam [1], Miller [14], Voss and Wang [22], Angelopoulos and Koukouvinos [2], and Angelopoulos *et al.* [3,4].

Although many methods have been proposed for analyzing unreplicated designs for a normal response, it is evident the lack of research papers for non-normally distributed responses. This fact prompted us to develop a methodology for screening out the important effects assuming that the response of interest is a binary one; therefore, we developed a generalized linear model, say a logistic model. Our approach for analyzing unreplicated designs constitutes a statistical method inspired by some information theoretical measures, the main of which was the symmetrical uncertainty (SU). To the best of our knowledge, this is the first time such an algorithm is modified and appropriately used for variable selection in unreplicated designs. The merits of our study is encouraging enough.

The rest of the paper is organized as follows. In Section 2, we briefly discuss the basic concepts of the information theoretical measures, the formulation of the problem as well as our new SU algorithm. In Section 3, we carry out an empirical study comparing our method with two well-known feature selection algorithms, the CMIM and the mRMR. Finally, in the last Section 4, we summarize the merits of our study providing some concluding remarks.

2. A METHOD FOR SEARCHING ACTIVE EFFECTS IN UN-REPLICATED DESIGNS WITH BINARY RESPONSE

Generalized linear models (Nelder and Wedderburn [17], McCullagh and Nelder [13] and Myers *et al.* [16]) were developed to allow the fit of regression models for response data that follow a distribution belonging to the exponential family. This family includes not only the exponential but also the normal, binomial, Poisson, geometric, negative binomial, gamma and the inverse normal distributions. All these models have a common property: the mean (or expected) response at each data point and the variance of the response are related.

Consider a two-level full factorial unreplicated design where one wants to estimate the main and interaction effects in d factors with n runs. Let X be the corresponding $n \times d$ design matrix where at the i_{th} data point, $i = 1, \dots, n$ the response is a Bernoulli random variable y_i , that takes only two possible values, 0 and 1, representing “failure” or “success”, respectively. It is well known that $\mu_i = E(y_i) = P_i = P(\mathbf{x}_i)$, where P_i is the probability of success in a Bernoulli process, \mathbf{x}_i is a d -dimensional vector of the predictor variables and $Var(y_i) = P_i(1 - P_i)$ is the variance of the response. It is obvious that the variance is a function of the mean. The probability of success, $P(\mathbf{x}_i)$, in case of the logistic regression model is given as follows

$$(2.1) \quad P(\mathbf{x}_i) = \frac{1}{1 + e^{-\mathbf{x}_i^T \boldsymbol{\beta}}},$$

where the term $\mathbf{x}_i^T \boldsymbol{\beta}$ is said to be the linear predictor. For more details on logistic regression model, we refer the interested reader to Montgomery *et al.* (2006). In accordance with this scenario, we perform our simulation study by generating logistic models that has the form

$$(2.2) \quad y_i = P(\mathbf{x}_i) + \varepsilon,$$

where ε has a distribution with zero mean and variance $P(x_i)[1 - P(x_i)]$. More precisely, ε takes two possible values: $\varepsilon = 1 - P(\mathbf{x}_i)$ with probability $P(\mathbf{x}_i)$ if $y = 1$, and $\varepsilon = -P(\mathbf{x}_i)$ with probability $1 - P(\mathbf{x}_i)$ if $y = 0$. Consequently, the conditional distribution of the outcome variable has a Bernoulli distribution with success probability $P(\mathbf{x}_i)$.

2.1. Information measures

Information theory provides useful tools to quantify the uncertainty of random variables. Our method is inspired from the information theory field with the aim of identifying those effects that carry as much information as possible. This section provides some information measures which constitutes the theoretical basis of our methodology.

Let U and V be two discrete random variables. One of the most fundamental concept in information theory is that of entropy measure which was introduced by Shannon [21] and it is defined as

$$(2.3) \quad H(U) = - \sum_{u \in \mathcal{U}} p(u) \log_2(p(u)).$$

The entropy quantifies the uncertainty of U , where $p(u)$ is the prior probability for all values of U . It is a measure of the amount of information required on average to describe the random variable. The information entropy of a Bernoulli trial used in our study is defined as

$$(2.4) \quad H(Y) = -p(y)\log_2 p(y) - (1 - p(y))\log_2(1 - p(y)),$$

where $p(y)$ is the prior probability for all values of Y .

In case of two variables we could define the mutual information (MI) which is a quantity that measures the mutual dependence of these variables. It is also called information gain (Quinlan [18]) and it is defined as

$$(2.5) \quad \begin{aligned} I(U|V) &= H(X) - H(X|Y) = H(Y) - H(Y|X) \\ &= H(X) + H(Y) - H(X, Y). \end{aligned}$$

Note that the MI of a random variable with itself, is its entropy. MI can be used for feature selection with the aim to select a small subset of features that carries as much information as possible (Fleuret [9], Peng *et al.* [19]). Information gain is a symmetrical measure for two random variables. Symmetry is an appealing property for a measure of correlations between factors, but information gain is biased in favor of factors with more values. Symmetrical uncertainty (Press *et al.* [20]) counterbalances the bias of information gain towards factors with more values, and normalizes its value to the range $[0, 1]$. The definition of Symmetrical Uncertainty is given as

$$(2.6) \quad SU(U, V) = 2 \times \left[\frac{I(U|V)}{H(U) + H(V)} \right].$$

2.2. Symmetrical uncertainty algorithm

The proposed method is a modification of a feature selection algorithm, known as Fast Correlation Based Filter (FCBF, Yu and Liu [23]). More precisely, it actually performs a typical variable selection using the SU coefficient so as to determine the significant effects. The algorithm can be described as follows:

Algorithm

- a) Given a $n \times d$ unreplicated design matrix $X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_d]$, where \mathbf{x}_l , $l = 1, 2, \dots, d$, is the l_{th} column of the matrix, as well as a $n \times 1$ Bernoulli distributed vector \mathbf{y} , which is the response vector, compute the entropy and the conditional entropy with respect to the response variable.
- b) Compute the vector entropy values and the conditional entropy values for each variable as: $H(X) = (H(\mathbf{x}_1), H(\mathbf{x}_2), \dots, H(\mathbf{x}_d))$ and $H(\mathbf{X}|\mathbf{Y}) = (H(\mathbf{x}_1|\mathbf{y}), H(\mathbf{x}_2|\mathbf{y}), \dots, H(\mathbf{x}_d|\mathbf{y}))$, where $H(\mathbf{x}_j)$ is the corresponding value of the entropy measure and $H(\mathbf{x}_j|\mathbf{y})$ is the corresponding value of the conditional entropy for the j -th, $j = 1, \dots, d$ variable, respectively.
- c) Compute the vector of information gain values as: $I(X|Y) = (I(\mathbf{x}_1|\mathbf{y}), I(\mathbf{x}_2|\mathbf{y}), \dots, I(\mathbf{x}_d|\mathbf{y}))$, where $I(\mathbf{x}_j|\mathbf{y})$ is the information gain value for each variable with respect to the response variable.
- d) Compute the symmetrical uncertainty measure, $SU = (su_1, su_2, \dots, su_d)$, where

$$su_j = 2 \times \left[\frac{I(\mathbf{x}_j|\mathbf{y})}{H(\mathbf{x}_j) + H(\mathbf{y})} \right],$$
 for $j = 1, \dots, d$, represents the value of SU for the j -th variable with respect to the response variable.
- e) The last step is to identify and maintain the significant effects by retaining only those with scores greater than the predefined threshold value of the SU vector values.

2.3. Performance Criteria

The performance of the proposed methodology is evaluated using the two most known criteria, the Type I and Type II error rates. In screening designs, there are two, the probability of declaring an inactive factor to be active (Type I error), and the probability of declaring an active factor to be inactive (Type II error). Type II errors are troublesome, as addressed in Lin [12], as well as Type I errors, since they can result in unnecessary cost in follow-up experiments. Type I errors are very likely in situations of effect sparsity. Undoubtedly, Type II error rates are of highly importance and we have considered that importance during the creation and implementation of our algorithm.

3. EXPERIMENTAL RESULTS

This section presents a simulation study examining the performance of our algorithm. To assess the performance of the proposed method, we applied simulations for a wide range of underlying models. Our information-theoretic method is compared with two feature selection algorithms which are widely used in many fields of science: the Conditional Mutual Information Maximization (CMIM) algorithm proposed by Fleuret [9] and the minimal-redundancy–maximal-relevance feature selection (mRMR) algorithm proposed by Peng *et al.* [19]. These algorithms were selected to be compared with SU-algorithm since they were made based on information measures. More precisely, CMIM constitutes a feature selection technique based on conditional mutual information and it iteratively picks features which maximize their mutual information with the class to predict, conditional to any feature has already picked. MRMR algorithm performs feature selection by maximizing the mutual information between the selected features and the desired output (relevance), as well as by minimizing the mutual information between the selected features (redundancy).

3.1. Simulation scheme

Two unreplicated factorial designs served as the design matrices in our simulations experiments: a 2^4 and a 2^5 full factorial design. We used these designs since they are commonly used in a wide range of problems; thus, our results can be comparable to other existing methods and problems. For the examined designs, the true active variables were selected using two different scenarios. For each design and each number of the active factors, we randomly generated 1000 Bernoulli distributed response vectors $\mathbf{y} \sim \text{Bernoulli}(P(\mathbf{X}^T\boldsymbol{\beta}))$, where $P(u) = \frac{1}{1+e^{-u}}$. All simulations were conducted using MATLAB codes.

Scenario A: We developed logistic models with coefficients taking predefined values. The coefficients of inactive effects are set equal to zero. However, in order to examine the sensitivity of the results in terms of the selection and the number of active factors, we changed the order of columns of the active factors, using different values of β as well as different number of active factors for each unreplicated design. As a result, we considered several models that were different in this regard. We considered the cases for $p = 1, 2, 3, 4, 5, 6, 7, 8$ active effects involved in a 2^4 factorial design and for $p = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12$ active effects involved in a 2^5 factorial design.

Scenario B: We developed logistic models with coefficients taking randomly selected values from the range -5 to 5 . When a generated coefficient was “almost zero”, it was replaced by 50% of the maximum coefficient. Concerning

the true active variables, they were also selected randomly, according to the uniform distribution, using the set of $\{1, \dots, d\}$ potentially active factors and with respect to the number of active factors of the design matrix. The coefficients of the non-active variables in the true model, were set equal to zero. The number of true active variables was set at most $d/2$, based on the sparsity of effects principle (Box and Meyer [5]). This principle states that, in contrast with the initial large number of potentially active factors, only few of them are dominant, meaning that their multitude hardly exceeds $1/2$ of the total number of factors.

3.2. Simulation results

The simulation results listed in the following Tables and Figures, contain the application of the SU method along with that of CMIM and mRMR. Before performing the simulation experiments, we should set the threshold value which determines whether a factor is significant or not. Several different threshold values (0.001, 0.01, 0.05, 0.1, 0.15, 0.2, median (SU)) were examined in order to find the optimal one for the proposed method. We finally selected the median(SU) as a threshold value, since it acquires the best results. Not to mention the fact that median(SU) is based on the estimated values of the SU vector and it seems to be a reasonable choice. The following Tables summarize the results concerning scenario A of simulation study. Specifically, in Tables 1 and 3, we present the examined models for designs with four and five factors, respectively. The first column represents the number corresponding to each model with predefined values for the coefficients depicted in the second column.

Table 1: Models considered in the simulation study for a 2^4 unreplicated design (Scenario A).

Model	Predefined values of coefficients
1	$[0,0,0,0,3,0,0,0,0,0,0,0,0,0]^T$
2	$[0,0,0,0,0,0,0,0,0,0,2,0,0,3]^T$
3	$[0,0,-7,0,0,0,0,-8,0,0,0,0,0,-6]^T$
4	$[0,0,-9,0,4,0,0,0,-2,0,0,0,10]^T$
5	$[6,0,0,0,7,0,0,-5,-5,0,-7,0,0,0]^T$
6	$[0,7,0,-2,0,5,2,0,4,0,0,0,-8,0]^T$
7	$[0,0,-9,2,0,0,4,5,8,0,0,-5,-7,0]^T$
8	$[5,0,-6,8,0,-5,6,0,0,7,0,0,-7,0,1]^T$

The obtained results are summarized in Tables 2 and 4 for four and five factors, respectively. More precisely, the first column in both Tables contains

the number that corresponds to each model. The remaining columns present the results for Type I and Type II error rates correspond to each method separately.

Table 2: 2^4 unreplicated design: Performance of the proposed method for models 1–8, using 1000 simulations (Scenario A).

Model	Type I Error			Type II Error		
	SU	CMIM	mRMR	SU	CMIM	mRMR
1	0.00	0.00	0.03	0.00	0.00	0.00
2	0.15	0.04	0.08	0.17	0.31	0.45
3	0.08	0.08	0.11	0.00	0.33	0.33
4	0.09	0.10	0.11	0.24	0.27	0.28
5	0.09	0.10	0.13	0.08	0.23	0.23
6	0.00	0.11	0.22	0.33	0.33	0.33
7	0.12	0.34	0.35	0.28	0.39	0.40
8	0.00	0.13	0.18	0.12	0.12	0.16
Average	0.07	0.11	0.15	0.15	0.25	0.27

Table 2 clearly shows that SU algorithm outperforms all the others in terms of both Type I and Type II error rates. Especially, the average values of Type II error is comparatively smaller; with SU equals to 0.15 compared to 0.25 and 0.27 of CMIM and mRMR, respectively. This fact is extremely important in factorial designs since low Type II means low probability of declaring an active factor to be inactive.

Table 3: Models considered in the simulation study for a 2^5 unreplicated design (Scenario A).

Model	Predefined values of coefficients
1	$[0, 5, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, -4, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 6, 0]^T$
2	$[20, 0, -17, 0, 0, 0, 0, 0, 0, 0, 0, 0, 12, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]^T$
3	$[7, 0, 5, 0, 0, 3, 0, 0, 0, 0, 0, 0, -5, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, -7, 0, 0, 0]^T$
4	$[0, 0, 0, 17, 0, 0, 0, 0, 0, 0, 0, 8, 0, 0, 0, -7, 0, 12, 0, 0, 0, 0, 0, 0, 3, 0, 0, 0, 0, -8, 0]^T$
5	$[0, 0, -9, -5, 0, -9, 0, 0, 0, 0, 0, 0, 0, 0, -2, 0, 0, 5, 0, 0, 0, 0, 4, 0, 0, 0, 0, 0, 0, 8]^T$
6	$[0, 0, 0, 0, 0, 5, 0, 0, 7, 0, 0, 0, 7, 0, 0, 0, 5, 0, 0, 0, 0, 0, 0, 5, 9, 9, 0, 0]^T$
7	$[0, 2, 4, 0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 3, 0, 0, -2, 0, 2, 0, 0, 0, 0, 0, 3, 2, 0, 0, 0]^T$
8	$[5, 0, 4, 5, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 9, 5, 0, 4, 5, 0, 0, 0, 0, 0, 0, 0, 0, 9, 6]^T$
9	$[0, 0, 2, -4, -3, 0, 0, 0, 0, 0, -4, 3, 0, 0, 0, 0, -4, 0, 0, 0, 0, 0, 0, 0, 0, 4, 3, 0, 2, -1]^T$
10	$[0, -5, 0, 0, 0, 0, -9, 0, -7, 0, 0, 0, -4, 0, 0, 0, -5, -7, 0, 0, 0, -2, -9, 0, 0, 0, -3, -8, 0, -5]^T$
11	$[0, 7, 9, 9, 0, 0, 0, 0, 17, 0, 0, 0, 10, 7, 19, 0, 0, 0, 0, 10, 0, 0, 14, 13, 0, 0, 0, 3, -10, 0]^T$
12	$[0, 3, 0, -2, 0, 0, 0, 0, 1, 0, 0, -4, -3, 2, 0, 0, 0, -5, 1, 4, 0, -3, 0, 2, 0, 3, 0, 2, 2, 4, 0]^T$

Table 4 shows that SU algorithm achieves the lowest error rates outperforming the other two methods. More precisely, SU gathers extremely low average value of Type II while keeping low values of Type I error. Figure 1 illustrates the performance of the proposed methods considering Type II errors for scenario A and scenario B at the left and right panel, respectively, considering the design with the four factors. As depicted in this Figure, SU reveals extremely better results compared to CMIM and mRMR in all the considered cases, establishing its effectiveness.

Table 4: 2^5 unreplicated design: Performance of the proposed method for models 1–12 (Scenario A).

Model	Type I Error			Type II Error		
	SU	CMIM	mRMR	SU	CMIM	mRMR
1	0.04	0.04	0.03	0.00	0.33	0.26
2	0.04	0.04	0.04	0.00	0.33	0.33
3	0.19	0.08	0.04	0.00	0.39	0.21
4	0.15	0.06	0.10	0.00	0.26	0.42
5	0.17	0.08	0.08	0.14	0.29	0.28
6	0.08	0.07	0.09	0.14	0.24	0.32
7	0.13	0.09	0.10	0.00	0.26	0.30
8	0.22	0.18	0.15	0.00	0.43	0.36
9	0.14	0.08	0.13	0.09	0.17	0.27
10	0.18	0.20	0.21	0.27	0.36	0.39
11	0.21	0.19	0.22	0.08	0.29	0.34
12	0.04	0.24	0.17	0.13	0.23	0.18
Average	0.13	0.11	0.12	0.07	0.30	0.31

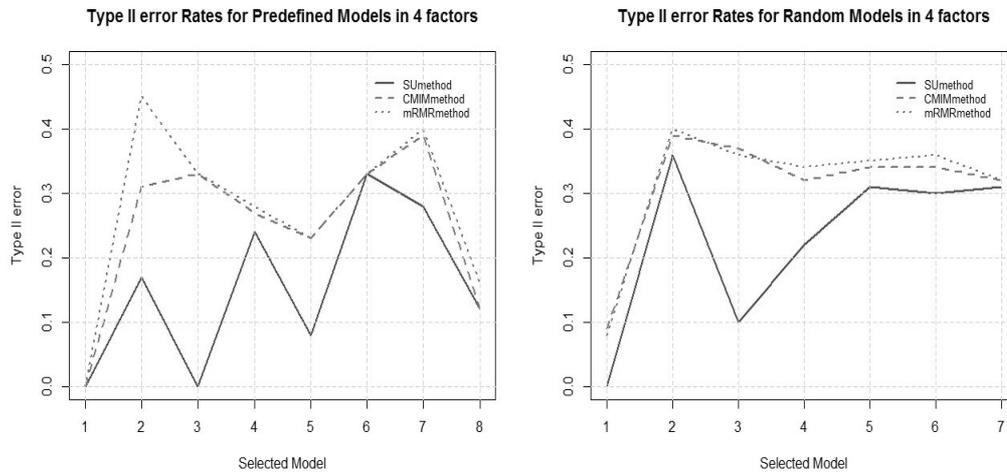


Figure 1: Comparisons of Type II error rates for Scenario A(left panel) and Scenario B (right panel) in case of four factor model. SU algorithm vs CMIM and mRMR (Scenario A).

Tables 5 and 6 are referred to scenario B. According to the simulation scheme, first column shows the number of true active effects in the simulated models which were selected randomly, and the next columns are referred to the average values of the Type I and Type II error rates for the examined approaches.

Table 5: 2^4 unreplicated design: Performance of the examined methods for random model coefficients (Scenario B).

Active effects	Type I Error			Type II Error		
	SU	CMIM	mRMR	SU	CMIM	mRMR
1	0.00	0.01	0.01	0.00	0.09	0.08
2	0.02	0.06	0.15	0.36	0.39	0.40
3	0.12	0.08	0.10	0.10	0.37	0.36
4	0.13	0.12	0.12	0.22	0.32	0.34
5	0.10	0.17	0.17	0.31	0.34	0.35
6	0.09	0.21	0.23	0.30	0.34	0.36
7	0.08	0.27	0.28	0.31	0.32	0.32
Average	0.08	0.13	0.15	0.23	0.31	0.32

A four factor unreplicated design is considered and seven different active factors from 1 to 7 were taken. Observing Table 5 we could confirm that the SU algorithm achieves an excellent performance since it has the lowest percentages of both Type I and Type II errors, say 0.08 and 0.23, while CMIM and mRMR achieve almost similar results with average values equal to 0.13 and 0.32 for Type I and Type II errors.

Table 6: 2^5 unreplicated design: Performance of the examined methods for random model coefficients (Scenario B).

Active effects	Type I Error			Type II Error		
	SU	CMIM	mRMR	SU	CMIM	mRMR
1	0.00	0.02	0.01	0.00	0.02	0.04
2	0.01	0.03	0.03	0.30	0.38	0.39
3	0.05	0.04	0.04	0.10	0.34	0.30
4	0.16	0.05	0.05	0.07	0.30	0.29
5	0.17	0.09	0.06	0.09	0.32	0.32
6	0.18	0.07	0.07	0.12	0.31	0.31
7	0.17	0.11	0.10	0.18	0.31	0.32
8	0.17	0.12	0.11	0.21	0.32	0.32
9	0.16	0.15	0.14	0.25	0.32	0.32
10	0.15	0.16	0.16	0.28	0.33	0.33
11	0.15	0.20	0.19	0.31	0.33	0.33
12	0.14	0.21	0.21	0.37	0.33	0.34
13	0.14	0.23	0.25	0.38	0.32	0.33
14	0.12	0.26	0.26	0.41	0.32	0.32
15	0.13	0.29	0.29	0.44	0.31	0.31
Average	0.13	0.14	0.13	0.23	0.30	0.30

Lastly, Table 6 aggregates the results for a five-factor unreplicated design considering different number of active factors, varying from 1 to 15. The average values of Type I error, show that SU overall outperforms the other algorithms. It is obvious that in terms of Type II error rates, SU revealed much better performance. Figure 2 illustrates a comparison of Type II error rates for scenario A (left panel) and scenario B (right panel), considering the design with the five factors. The horizontal axes show the active factors that we examined each time while the vertical axes the percentage of the Type II error. It should be noted that in cases of 12 active factors and above, the performance is relatively smaller. This fact justified by the assumption of effect sparsity which holds in the present experiment.

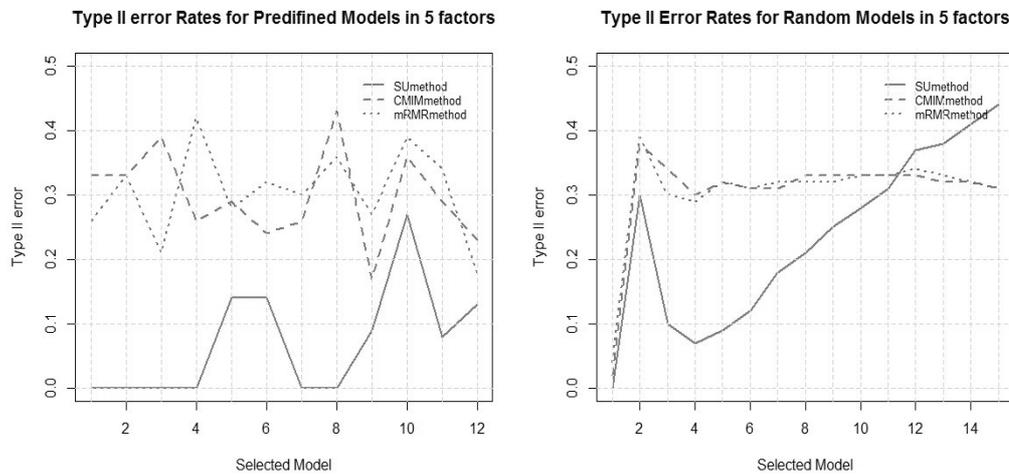


Figure 2: Comparisons of Type II error rates for Scenario A(left panel) and Scenario B (right panel) in case of five factor model. SU algorithm vs CMIM and mRMR.

3.3. Real experiment

In this subsection, we examine how the proposed screening methodology performs in the presence of real data. More precisely, we examined a real medical dataset that was collected in an annual registry conducted during the period 01/01/2005–31/12/2005 by the Hellenic Trauma and Emergency Surgery Society and which involves 30 General Hospitals in Greece. Each week, there was selected two data sets, each forms a factorial design with four and five factors, respectively, according to medical advice. There was the necessity of finding significant factors and their interactions without using an extremely large number of patients. For each patient a corresponding response variable, y , was reported which takes only two possible outcomes, denoted as 0 for survival and 1 for death. Taking all the interactions among factors a factorial design without replicates was formed.

This experiment helped us to confirm the effectiveness of our method to identify the significant factors in real life problems. This case study is of particular interest since one can identify the most significant variables and their interactions with respect to a certain effect (survival or death). The main purpose of the present real case study is to validate the practical use of our approach and to give some insights into how the proposed screening procedure contributes in real life scenarios. First of all, we present the analysis of the real data in the presence of four factors. Table 7 gives a description of the variables used in our study.

Table 7: Description of variables for a 2^4 experiment.

Variable	Description
x_{55}	immobility of limbs (0 = no, 1 = yes)
x_{56}	fluids (0 = no, 1 = yes)
x_{64}	Radiograph E.R. (0 = no, 1 = yes)
x_{72}	surgical intervention (0 = no, 1 = yes)

Table 8 presents the merits of this experiment. We denote variable x_{55} as factor A, variable x_{56} as factor B, variable x_{64} as factor C and x_{72} as the D factor. According to this notation we present the second order interactions of variables x_{55} and x_{56} as AB, of variables x_{55} and x_{64} as AC and so on. In this way we acquired the third and fourth order interactions of factors presented in Table 8. As we can conclude, all the applied methods recognize exactly the same significant variables something that confirms the efficiency of our algorithm to correctly identify significant factors.

Table 8: Significant variables for the real medical dataset using a 2^4 experiment.

Method	A	B	AB	C	AC	BC	ABC	D	AD	BD	ABD	CD	ACD	BCD	ABCD
SU	•	•	○	○	•	○	○	○	•	○	○	○	○	○	•
CMIM	•	•	○	○	•	○	○	○	•	○	○	○	○	○	•
mRMR	•	•	○	○	•	○	○	○	•	○	○	○	○	○	•

The second stage of this real experiment regards to the full factorial unreplicated design with five factors. Table 9 summarizes the description of the five variables used for this case study. In the same way as that of the four-factor case, variable x_{36} is denoted as factor A, variable x_{55} as factor B, variable x_{56} as factor C, variable x_{64} as factor D and x_{72} as the E factor. According to this notation, we present the second order interactions of variables x_{36} and x_{55} as AB, of variables x_{36} and x_{56} as AC and so on.

4. CONCLUDING REMARKS

Unreplicated experiments can be conducted in various improvement processes due to their economic run size and structure. However, the analysis of unreplicated designs doesn't constitute an easy issue since there are no degrees of freedom to estimate the experimental error. This fact makes the analysis of variance of such designs infeasible. An additional hindrance is that of dealing with a non-normal response, for instance a binary one. In this work, we propose a method for selecting the active effects in unreplicated designs, assuming a logistic regression model. We take advantage of the simplicity and the effectiveness of the SU measure so as to introduce a new method for analyzing unreplicated factorials. The novelty of the proposed method is contained on the usage of information gain and symmetrical uncertainty for analyzing unreplicated designs with a binary response. The simulation study of section 3 shows that the proposed method tends to declare at the highest rate inactive effects to be active and at the lowest rate active effects to be inactive. Compared with CMIM and mRMR, our approach has an almost similar performance concerning Type I error rates; however, Type II error is notably higher for both CMIM and mRMR leading to an unstable performance compared to SU. This fact, simultaneously leads to a very satisfactory power, that is $1 - (\text{Type II error rate})$, of the algorithm, something that constitutes an extremely characteristic for a screening procedure, such as the analysis of unreplicated designs. In conclusion, SU achieves a general stable performance and yields significantly low Type II errors, while it keeps Type I at a low level as well. It should be highlighted that there are problems, especially in real life, where one needs to perform an economic experiment with the smallest possible error. SU method gave the best average results in case of a real medical analysis not only by identifying the significant factors but also by keeping low Type I error rates. The empirical performance of the proposed algorithm reveals that this new approach constitutes a very efficient way of tackling the problem of unreplicated factorial designs while it opens new research opportunities for the application of information-theoretic methods in experimental designs where there are no degrees of freedom to estimate the experimental error.

ACKNOWLEDGMENTS

The research of the first author (K.D.) was financially supported by a scholarship awarded by the Secretariat of the Research Committee of National Technical University of Athens.

REFERENCES

- [1] ABOUKALAM, M.A.F. (2005). Quick, easy and powerful analysis of unreplicated factorial designs, *Communications in Statistics – Theory and Methods*, **34**, 1169–1175.
- [2] ANGELOPOULOS, P. and KOUKOUVINOS, C. (2008). Detecting active effects in unreplicated designs, *Journal of Applied Statistics*, **35**, 277–281.
- [3] ANGELOPOULOS, P.; EVANGELARAS, H. and KOUKOUVINOS, C. (2010). Analyzing unreplicated 2^k factorial designs by examining their projections into $k - 1$ factors, *Quality and Reliability Engineering International*, **26**, 223–233.
- [4] ANGELOPOULOS, P.; KOUKOUVINOS, C. and SKOUNTZOU, A. (2013). Clustering effects in unreplicated factorial experiments, *Communications in Statistics – Simulation and Computation*, **42**, 1998–2007.
- [5] BOX, G.E.P. and MEYER, R.D. (1986). An analysis for unreplicated fractional factorials, *Technometrics*, **28**, 11–18.
- [6] CHEN, Y. and KUNERT, J. (2004). A new quantitative method for analysing unreplicated factorial designs, *Biometrical Journal*, **46**, 125–140.
- [7] DANIEL, C. (1959). Use of half-normal plots in interpreting factorial two-level experiments, *Technometrics*, **1**, 311–341.
- [8] DONG, F. (1993). On the identification of active contrasts in unreplicated fractional factorials, *Statistica Sinica*, **3**, 209–217.
- [9] FLEURET, F. (2004). Fast binary feature selection with conditional mutual information, *Journal of Machine Learning Research*, **5**, 1531–1555.
- [10] HAMADA, M. and BALAKRISHNAN, N. (1998). Analysing unreplicated factorial experiments: A review with some new proposals, *Statistica Sinica*, **8**, 1–41.
- [11] LENTH, R.V. (1989). Quick and easy analysis of unreplicated factorial, *Technometrics*, **31**, 469–473.
- [12] LIN, D.K.J. (1995). Generating systematic supersaturated designs, *Technometrics*, **37**, 213–225.
- [13] MCCULLAGH, P. and NELDER, J.A. (1989). *Generalized Linear Models*, Chapman and Hall, London.
- [14] MILLER, A. (2005). The analysis of unreplicated factorial experiments using all possible comparisons, *Technometrics*, **47**, 51–63.
- [15] MONTGOMERY, D.C.; PECK, E.A. and VINING, G.G. (2006). *Introduction to Linear Regression Analysis*, 4th ed., New York: Wiley.
- [16] MYERS, R.H; MONTGOMERY, D.C. and VINING G.G. (2002). *Generalized Linear Models. With applications in Engineering and the Sciences*, John Wiley and Sons, New York.
- [17] NELDER, J.A. and WEDDERBURN, R.W.M. (1972). Generalized linear models, *Journal of the Royal Statistical Society Series A*, **135**, 370–384.
- [18] QUINLAN, J.R. (1986). Induction of decision trees, *Machine Learning*, **1**, 81–106.

- [19] PENG, H.; LONG, F. and DING, C. (2005). Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **27**, 1226–1238.
- [20] PRESS, W.H.; FLANNERY, B.P.; TEUKOLSKY, S.A. and VETTERLING, W.T. (1988). *Numerical Recipes*, Cambridge University Press, Cambridge.
- [21] SHANNON, C.E. (1948). A mathematical theory of communication, *Bell System Technical Journal*, **27**, 379–423 and 623–656.
- [22] VOSS, D.T. and WANG, W. (2006). On adaptive testing in orthogonal saturated designs, *Statistica Sinica*, **16**, 227–234.
- [23] YU, L. and LIU, H. (2003). Feature selection for high-dimensional data: A fast correlation-based filter solution, *Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003)*, Washington DC, pp. 856–863.

PREDICTION INTERVALS OF THE RECORD-VALUES PROCESS

Authors: AMANY E. ALY

– Department of Mathematics, Faculty of Science, Helwan University,
Ain Helwan, Cairo, Egypt
aabusabh@hotmail.com

H.M. BARAKAT

– Department of Mathematics, Faculty of Science, Zagazig University,
Zagazig, Egypt
hbarakat2@hotmail.com

MAGDY E. EL-ADLL

– Department of Mathematics, Faculty of Science, Helwan University,
Ain Helwan, Cairo, Egypt
melad112@hotmail.com

Received: March 2017

Revised: May 2017

Accepted: June 2017

Abstract:

- In this paper, exact prediction intervals of the record-values process are constructed. The record-values process model, may be considered as the collection of record-values with integer or non-integer indices. It includes both usual k -th record-values and fractional k -th record-values models. For constructing the prediction intervals, two predictive pivotal quantities are developed. The distributions of the predictive pivotal quantities are derived and it is revealed that the distribution functions of the predictive pivotal quantities are similar for the upper and lower fractional record-values. More results are obtained for the exponential upper record-values process, including two point predictors and their exact mean square errors. Some efficient algorithms are given and Monte Carlo simulation studies are conducted for comparing pivotal quantities. Finally, three real data sets are analyzed.

Key-Words:

- *record-values process; pivotal quantity; prediction interval; coverage probability; Monte Carlo simulation.*

AMS Subject Classification:

- 62E15, 62G30, 62G32, 62M20, 62F25.

1. INTRODUCTION

Record values arise naturally in many practical problems and there are several situations pertaining to meteorology, hydrology, sporting and athletic events where only record-values may be recorded. Outcomes of competitions, e.g. in athletics, arise in ascending order. In particular, sport events attract many spectators since records and best results appeal to people. Being most popular in sports, lists of best results and records are of particular interest in many other areas of real life as well. For an elaborate treatment on records and their applications see: Arnold *et al.* [6], Nevzorov [40], Gulati and Padget [27], and Ahsanullah ([1], [3]). The first result for record-values involving independent and identically observations was reported by Chandler [17]. Dziubdziela and Kopocinški [21] generalized the concept of record-values of [17] to a more generalized nature and called them k -th record-values. Since the k -th member of the sequence of the classical record-values is also known as the k -th record-value, the record-values defined in [21] is also called generalized record-values. Some properties and applications for current records are given in Barakat *et al.* [11]. Stigler [46] introduced the concept of order statistics process, which may be considered as fractional order statistics for non-integer index. Jones [31] gave an alternative construction of Stigler's uniform fractional order statistics. Namely, ordinary order statistics of a sample from uniform distribution are used to construct random variables (rv's) with the same joint distribution as Stigler's order statistics. Some applications of fractional order statistics are given in Hutson [30]. Bieniek and Szynal [14] follows a similar method of fractional order statistics to introduce the fractional record-values or the record-values process, which can be considered as a family of k -th record-values with n replaced by a positive number t .

One of the most important problems in statistics, is to predict future events based on past or current events. A predictor may be either a point or an interval predictor. Point predictor of future records was studied by Kaminsky and Nelson [32], Ahsanullah [2], Nagaraja [37], and Doganakso and Balakrishnan [19]. Prediction intervals of future records were given in Dunsmore [20], Balakrishnan *et al.* [7], Berred [13], AL-Hussaini and Ahmad [4], and Raqab and Balakrishnan [42]. Bayesian and non-Bayesian approaches have been extensively studied by many authors, e.g. Lawless [36], Kaminsky and Rhodin [34], Geisser [25], Nagaraja [39], and Kaminsky and Nelson [33]. Recent works of prediction using pivotal quantities include, Barakat *et al.* ([8], [9], [10], [12]), El-Adll [22], Aly [5], and El-Adll and Aly [23].

1.1. Motivation of the study

According to Theorem 6.3.1 page 339 of Galambos [24], the ordinary record-values are very rare to be observed. For example, one may wait too many years

before observing the next upper (lower) record of the amount of water added to a given river. Although the fractional k -th record-values cannot be observed in practice, prediction of future fractional k -th record-values is prominent in applications. As an example, the prediction of fractional k -th record-values can be applied in reliability and survival data analysis since the fractional k -th record-values may be considered as an estimator of the inverse cumulative hazard function and therefore the quantiles of the population cdf. On the other hand, the employment of the k -th fractional record-values provides an interval estimate with accurate significant level, while the use of the k -th ordinary record-values gives an interval estimate with approximate significant level (e.g., [14]).

Furthermore, our study is carried in a general framework which include prediction of the usual record-values, as well as k -th ordinary record-values, as special cases. Thereby, all the obtained new results not only have theoretical importance but also have practical importance. Thus the results of this paper, which are given in the present general framework, are beneficial when it is necessary to predict the quantiles of a distribution for which, the type of the hazard function may be changed in future.

In the next section, we give a comprehensive survey for the main results of the record-values process, that will be needed in this paper, most of these results are due to [14].

In Sections 3 and 4 of this paper, two prediction intervals for future fractional upper (lower) records are constructed based on two general predictive pivotal quantities. More details for the exponential distribution including, two point predictors and their exact mean square errors are considered for the upper record-value process. In Section 5, two simulation studies are carried out to explain the efficiency of the proposed results. In one of them, the distribution parameters are assumed to be unknown. Some applications to real data are given in Section 6. Two basic algorithms for generation ordinary record-values and fractional upper record-values, as well as an algorithm to implement these prediction intervals, are given in an Appendix.

2. PRELIMINARY RESULTS

In this section, some important preliminary and auxiliary results for the basic distribution theory of ordinary and fractional records are presented. Let $\{X_n, n \geq 1\}$ be a sequence of independent and identically distributed (iid) random variables (rv's) having a continuous cumulative distribution function (cdf) $F(x)$ and probability density function (pdf) $f(x)$. Furthermore, suppose that $X_{1:n}, X_{2:n}, \dots, X_{n:n}$ denote the order statistics of the random sample X_1, X_2, \dots, X_n .

2.1. Ordinary records

An observation X_j is called an upper record-value if its value exceeds that of all previous observations. Thus, X_j is an upper record-value if $X_j > X_i$ for every $i < j$. In other words, the upper record-value R_n of a random sample of size n can be expressed as $R_n = X_{n:n} = \max\{X_1, X_2, \dots, X_n\}$. Dziubdziela and Kopociński [21] extended the concept of upper record to the k -th upper record, for $k \geq 1$, which is formulated in the following definition.

Definition 2.1 (cf. Dziubdziela and Kopociński [21]). The k -th upper record times, $T_k(n)$, $n \geq 1$, of the sequence $\{X_i, i \geq 1\}$ is defined for fixed $k \geq 1$, as $T_k(1) = 1$ and

$$T_k(n + 1) = \min\{j > T_k(n) : X_{j:j+k-1} > X_{T_k(n):T_k(n)+k-1}\}, \quad n > 1,$$

and the k -th upper record-values as $R_n^{(k)} = X_{T_k(n):T_k(n)+k-1}$, $n \geq 1$.

The k -th lower record-values is defined similarly. Clearly, $R_1^{(k)} = X_{1:k} = \min\{X_1, \dots, X_k\}$. For $k = 1$ we have $R_n^{(1)} = R_n = X_{n:n}$. In other words, the k -th upper record-sequence is the sequence of the k -th largest yet seen. Although the term “record times” is used in all definitions related to records in statistical literature, it does not mean the time in its verbal sense. The pdf of the k -th upper record-value is

$$(2.1) \quad f_{R_n^{(k)}}(r) = \frac{k^n}{\Gamma(n)} [H(r)]^{n-1} [\bar{F}(r)]^{k-1} f(r), \quad -\infty < r < \infty,$$

where $H(r) = -\log[1 - F(r)]$ is the cumulative hazard function, $h(r) = H'(r) = f(r)/\bar{F}(r)$ denotes the hazard (failure rate) function and $\bar{F} = 1 - F$. The joint pdf of $R_m^{(k)}$ and $R_n^{(k)}$, $m < n$ for $-\infty < r_m < r_n < \infty$, can be written in the form

$$(2.2) \quad \begin{aligned} & f_{R_m^{(k)}, R_n^{(k)}}(r_m, r_n) \\ &= \frac{k^n}{\Gamma(m)\Gamma(n - m)} [H(r_m)]^{m-1} [H(r_n) - H(r_m)]^{n-m-1} [\bar{F}(r_n)]^k h(r_m)h(r_n). \end{aligned}$$

Furthermore, the joint pdf of the random vector $(R_1^{(k)}, R_2^{(k)}, \dots, R_n^{(k)})$ is given by

$$(2.3) \quad f_{R_1^{(k)}, R_2^{(k)}, \dots, R_n^{(k)}}(r_1, r_2, \dots, r_n) = k^n [\bar{F}(r_n)]^k \prod_{i=1}^n h(r_i), \quad -\infty < r_1 < r_2 < \dots < r_n < \infty.$$

For more details of the previous three relations, see [21], [26], and [38].

2.2. Upper record-values process

Let $\{W_n^{(k)}, n \geq 1\}$ denote the k -th upper record-values from the standard exponential distribution (EXP(1)). The following two facts, which are due to Ahsanullah [1], characterize the exponential distribution.

Fact 1. For any positive integers m and n , with $m < n$, the rv's $W_m^{(k)}$ and $W_n^{(k)} - W_m^{(k)}$ are independent.

Fact 2. The spacings $W_n^{(k)} - W_m^{(k)}$ follow gamma distribution with parameters $n - m$ and k , respectively.

The following definition, which is due to [14], is necessary to construct the record-value process and fractional k -th record-values.

Definition 2.2. Let $k \in \mathbb{N}$ be fixed and $W^{(k)} = \{W^{(k)}(t), t \geq 0\}$ be a stochastic process such that:

- (i) $W^{(k)}(0) = 0$ almost sure;
- (ii) $W^{(k)}(t)$ has independent increments;
- (iii) For every $t > s \geq 0$, $W^{(k)}(t) - W^{(k)}(s)$ has gamma distribution with parameters $t - s$ and k , respectively.

Then $\{W^{(k)}(t), t \geq 0\}$ is called the exponential k -th upper record-values process. Moreover, the rv's, $W^{(k)}(t), t > 0$, are said to be exponential fractional k -th upper record-values.

Remark 2.1.

1. By fractional k -th record-values, we mean k -th record-values with fractional indices.
2. We shall assume that the cdf F is continuous with pdf f and quantile function

$$F^{-1}(q) = \inf\{v : F(v) \geq q\}, \quad 0 \leq q < 1.$$

The k -th record-values process and the fractional k -th record-values based on F are formulated in the following definition:

Definition 2.3 (Bieniek and Szynal [14]). The stochastic process $Y^{(k)} = \{Y^{(k)}(t), t \geq 0\}$, where

$$Y^{(k)}(t) = F^{-1}(1 - \exp[-W^{(k)}(t)]), \quad t \geq 0,$$

is called the k -th upper record-values process based on F and the rv's $Y^{(k)}(t), t > 0$, are said to be fractional k -th upper record-values from F .

As in the ordinary record-values, the pdf $f_{Y^{(k)}(t)}(y)$ of the fractional k -th upper record-value $Y^{(k)}(t)$ is

$$(2.4) \quad f_{Y^{(k)}(t)}(y) = \frac{k^t}{\Gamma(t)} [H(y)]^{t-1} [\bar{F}(y)]^{k-1} f(y), \quad -\infty < y < \infty, \quad t > 0,$$

and the joint pdf of $Y^{(k)}(t_r)$ and $Y^{(k)}(t_s)$, $t_s > t_r \geq 0$, can be written for $-\infty < y_r < y_s < \infty$, as

$$(2.5) \quad \begin{aligned} & f_{Y^{(k)}(t_r), Y^{(k)}(t_s)}(y_r, y_s) \\ &= \frac{k^{t_s}}{\Gamma(t_r)\Gamma(t_s - t_r)} [H(y_r)]^{t_r-1} [H(y_s) - H(y_r)]^{t_s-t_r-1} [\bar{F}(y_s)]^k h(y_r)h(y_s). \end{aligned}$$

Moreover, if $0 = t_0 < t_1 < \dots < t_n$, then the joint pdf of the random vector $\mathbf{Y} = (Y^{(k)}(t_1), Y^{(k)}(t_2), \dots, Y^{(k)}(t_n))$ is given by (c.f. [14])

$$(2.6) \quad f_{\mathbf{Y}}(y_{t_1}, y_{t_2}, \dots, y_{t_n}) = k^{t_n} [\bar{F}(y_{t_n})]^k \prod_{i=1}^n \frac{(H(y_{t_i}) - H(y_{t_{i-1}}))^{t_i-t_{i-1}-1} h(y_{t_i})}{\Gamma(t_i - t_{i-1})},$$

for $-\infty < y_{t_1} < y_{t_2} < \dots < y_{t_n} < \infty$.

2.3. Lower record-values process

Let $\{Z_n^{(k)}, n \geq 1\}$ denote the k -th lower record-values from the standard negative exponential distribution (NEXP(1)), with cdf $G^*(x) = e^x, x \leq 0$.

Definition 2.4. Let $k \in \mathbb{N}$ be fixed and $Z^{(k)} = \{Z^{(k)}(t), t \geq 0\}$ be a stochastic process such that:

- (i) $Z^{(k)}(0) = 0$ almost sure;
- (ii) $Z^{(k)}(t)$ has independent increments;
- (iii) For any $t > s \geq 0$, $Z^{(k)}(t) - Z^{(k)}(s)$ has a reverse gamma distribution with parameters $t - s$ and k (the reverse gamma pdf is $f(x) = \frac{k^{t-s}}{\Gamma(t-s)} |x|^{t-s-1} e^{x/k}, x \leq 0$).

Then $\{Z^{(k)}(t), t \geq 0\}$ is called the negative exponential k -th lower record-values process. Moreover, the rv's $Z^{(k)}(t), t > 0$, are said to be negative exponential fractional k -th lower record-values.

Definition 2.5 (Bieniek and Szynal [14]). The stochastic process $X^{(k)} = \{X^{(k)}(t), t \geq 0\}$, where

$$X^{(k)}(t) = F^{-1}(\exp[Z^{(k)}(t)]), \quad t \geq 0,$$

is called the k -th lower record-values process based on the cdf F and the rv's $X^{(k)}(t), t > 0$, are said to be fractional k -th lower record-values from F .

The pdf $f_{X^{(k)}(t)}(x)$ of the fractional k -th lower record-values $X^{(k)}(t)$ is

$$(2.7) \quad f_{X^{(k)}(t)}(x) = \frac{k^t}{\Gamma(t)} [-\log F(x)]^{t-1} [F(x)]^{k-1} f(x), \quad -\infty < x < \infty, \quad t > 0,$$

and the joint pdf of $X^{(k)}(t_r)$ and $X^{(k)}(t_s)$, $t_s > t_r \geq 0$ can be written for $-\infty < x_s < x_r < \infty$, as

$$(2.8) \quad f_{X^{(k)}(t_r), X^{(k)}(t_s)}(x_r, x_s) \\ = \frac{k^{t_s}}{\Gamma(t_r)\Gamma(t_s - t_r)} [-\log F(x_r)]^{t_r-1} \left[\log \frac{F(x_r)}{F(x_s)} \right]^{t_s-t_r-1} [F(x_s)]^k \frac{f(x_r)}{F(x_r)} \frac{f(x_s)}{F(x_s)}.$$

3. PREDICTION OF FUTURE UPPER RECORD-VALUES PROCESS

In this section, two predictive pivotal quantities (the pivotal quantity is a function of the sample X_1, X_2, \dots , and on the distribution parameters, but its distribution does not depend on the distribution parameters) are developed to construct prediction intervals of future fractional upper record-values from a continuous distribution. The following theorem is formulated for the first pivotal quantity, which enables us to predict any future fractional k -th upper record-value $Y^{(k)}(t_s)$ based on one fractional k -th upper record-value $Y^{(k)}(t_r)$ with $r < s$.

Theorem 3.1. *Let $0 = t_0 < t_1 < t_2 < \dots < t_n$ be positive real numbers and $Y^{(k)}(t_r)$ be the r -th fractional k -th upper record-values from a continuous distribution with cdf F and pdf f . Then the pdf and the cdf of the pivotal quantity $P_1 = (W^{(k)}(t_s) - W^{(k)}(t_r)) / W^{(k)}(t_r)$, with $s > r$, respectively, are*

$$(3.1) \quad f_{P_1}(p_1) = \frac{1}{B(t_s - t_r, t_r)} p_1^{t_s-t_r-1} (1+p_1)^{-t_s}, \quad p_1 > 0,$$

and

$$(3.2) \quad F_{P_1}(p_1) = I_{\frac{p_1}{1+p_1}}(t_s - t_r, t_r), \quad p_1 \geq 0,$$

where $I_z(a, b) = \frac{1}{B(a, b)} \int_0^z u^{a-1} (1-u)^{b-1} du$, $0 < z < 1$, is the incomplete beta function, $B(a, b) = \int_0^1 u^{a-1} (1-u)^{b-1} du$ and

$$(3.3) \quad W^{(k)}(t_i) = -\log \bar{F}(Y^{(k)}(t_i)), \quad i = 1, 2, \dots, n.$$

A $100(1 - \delta)\%$ predictive confidence interval (PCI) for the future fractional k -th upper record-value $Y^{(k)}(t_s)$, (L, U_{P_1}) , is

$$L = Y^{(k)}(t_r) \quad \text{and} \quad U_{P_1} = F^{-1} \left(1 - \left(\bar{F}(Y^{(k)}(t_r)) \right)^{1+p_1(\delta)} \right),$$

where $p_1(\delta)$ can be obtained by solving the non linear equation $F_{P_1}(p_1(\delta)) = 1 - \delta$. Moreover, when $Y^{(k)}(t_i) = W^{(k)}(t_i), i = 1, 2, \dots, n$ (i.e., F is $EXP(1)$), the expected interval width of the PCI is $\frac{p_1(\delta)t_r}{k}$. Furthermore, $\tilde{W}^{(k)}(t_s) = \frac{t_s}{t_r} W^{(k)}(t_r)$ is an unbiased point predictor based on P_1 .

Proof: Since the transformation $w = -\log \bar{F}(y)$ is one to one and onto (monotone increasing function), the pdf of the rv $W^{(k)}(t), f_{W^{(k)}(t)}(w)$, is given by

$$f_{W^{(k)}(t)}(w) = |J|f_{Y^{(k)}(t)}(y(w)), \quad \text{where} \quad |J| = \frac{dy}{dw} = \frac{1}{f(y)}e^{-w}.$$

Therefore, by (2.4) we have $f_{W^{(k)}(t)}(w) = \frac{k^t}{\Gamma(t)}w^{t-1}e^{-kw}, w > 0$, which is the pdf of fractional k -th upper record-values based on $EXP(1)$. Thus, the joint pdf of $W^{(k)}(t_r)$ and $W^{(k)}(t_s), t_s > t_r \geq 0$ based on $EXP(1)$ can be written by (2.5) as

$$(3.4) \quad f_{W^{(k)}(t_r), W^{(k)}(t_s)}(w_r, w_s) = \frac{k^{t_s}}{\Gamma(t_r)\Gamma(t_s - t_r)}w_r^{t_r-1}(w_s - w_r)^{t_s-t_r-1}e^{-kw_s},$$

with $0 < w_r < w_s < \infty$. By a standard method of transformation of rvs, the joint pdf $f_{P_1, W_r^{(k)}}(p_1, w_r)$ of P_1 and $W^{(k)}(t_r)$ can be written in the form

$$(3.5) \quad f_{P_1, W_r^{(k)}}(p_1, w_r) = \frac{k^{t_s}}{\Gamma(t_r)\Gamma(t_s - t_r)}w_r^{t_s-1}p_1^{t_s-t_r-1}e^{-k(1+p_1)w_r}, \quad p_1 > 0, w_r > 0.$$

Thus, we have

$$\begin{aligned} f_{P_1}(p_1) &= \int_0^\infty f_{P_1, W_r^{(k)}}(p_1, w_r)dw_r \\ &= \frac{k^{t_s}}{\Gamma(t_r)\Gamma(t_s - t_r)} \int_0^\infty w_r^{t_s-1}p_1^{t_s-t_r-1}e^{-k(1+p_1)w_r}dw_r. \end{aligned}$$

By the definition of gamma function, the above integration can be simplified in the form (3.1). Moreover, the cdf of the pivotal quantity P_1 is given by

$$\begin{aligned} F_{P_1}(p_1) &= \int_0^{p_1} f_{P_1}(z)dz = \int_0^{p_1} \frac{1}{B(t_r, t_s - t_r)} z^{t_s-t_r-1}(1+z)^{-t_s}dz \\ &= \frac{1}{B(t_s - t_r, t_r)} \int_0^{p_1} \left(\frac{z}{1+z}\right)^{t_s} \left(\frac{1}{z}\right)^{t_r+1} dz. \end{aligned}$$

If we set $w = \frac{z}{1+z}$ in the above integration, it yields (3.2). If δ is such that $F_{P_1}(p_1(\delta)) = P(P_1 \leq p_1(\delta)) = 1 - \delta$, we can write

$$\begin{aligned} 1 - \delta &= P\left(0 < \frac{W^{(k)}(t_s) - W^{(k)}(t_r)}{W^{(k)}(t_r)} \leq p_\delta\right) = P\left(0 < \frac{W^{(k)}(t_s)}{W^{(k)}(t_r)} - 1 \leq p_1(\delta)\right) \\ &= P\left(W^{(k)}(t_r) < W^{(k)}(t_s) \leq (1 + p_1(\delta))W^{(k)}(t_r)\right) = P\left(L < Y^{(k)}(t_s) \leq U_{P_1}\right). \end{aligned}$$

The expected interval width of the PCI for $W^{(k)}(t_s)$ is given by

$$E \left[(1 + p_1(\delta))W^{(k)}(t_r) - W^{(k)}(t_r) \right] = E \left[p_1(\delta)W^{(k)}(t_r) \right] = \frac{p_1(\delta)t_r}{k}.$$

Finally, a point predictor based on P_1 can be obtained from the relation $\tilde{W}^{(k)}(t_s) = L + c_1(U_{P_1} - L)$, where the constant c_1 is such that $E[\tilde{W}^{(k)}(t_s)] = E[W^{(k)}(t_s)] = t_s/k$. Hence the theorem. \square

Theorem 3.2. Assume that $0 = t_0 < t_1 < t_2 < \dots < t_n$ are positive real numbers. Furthermore, let $Y^{(k)}(t_1)$ and $Y^{(k)}(t_r)$ be the first and the r -th fractional k -th upper record-values from a continuous distribution with cdf F and pdf f . Then the pdf and the cdf of the pivotal quantity

$$(3.6) \quad P_2 = \frac{W^{(k)}(t_s) - W^{(k)}(t_r)}{W^{(k)}(t_r) - W^{(k)}(t_1)}, \quad s > r > 1,$$

are given by

$$(3.7) \quad f_{P_2}(p_2) = \frac{1}{B(t_s - t_r, t_r - t_1)} (1 + p_2)^{-(t_s - t_1)} p_2^{t_s - t_r - 1}, \quad p_2 > 0,$$

and

$$(3.8) \quad F_{P_2}(p_2) = I_{\frac{p_2}{1+p_2}}(t_s - t_r, t_r - t_1), \quad p_2 \geq 0,$$

respectively, with $W^{(k)}(t_i) = -\log \bar{F}(Y^{(k)}(t_i))$, $i = 1, 2, \dots, n$. A $100(1 - \delta)\%$ PCI for the future k -th upper record-value $Y^{(k)}(t_s)$ is (L, U_{P_2}) , with

$$U_{P_2} = F^{-1} \left(1 - \bar{F}(Y^{(k)}(t_1)) \left(\frac{\bar{F}(Y^{(k)}(t_r))}{\bar{F}(Y^{(k)}(t_1))} \right)^{1+p_2(\delta)} \right),$$

where $p_2(\delta)$ can be obtained by solving the non linear equation $F_{P_2}(p_2(\delta)) = 1 - \delta$. Moreover, an unbiased point predictor based on P_2 is given by

$$\hat{W}^{(k)}(t_s) = W^{(k)}(t_r) + \left(\frac{t_s - t_r}{t_r - t_1} \right) \left(W^{(k)}(t_r) - W^{(k)}(t_1) \right), \quad s > r > 1,$$

which is the best linear unbiased predictor (BLUP) for $W^{(k)}(t_s)$.

Proof: We see from the proof of Theorem 3.1 that the rv $W^{(k)}(t_i)$, $i = 1, 2, \dots, n$, can be expressed as fractional k -th upper record-values based on EXP(1). Therefore, the joint pdf of $W^{(k)}(t_1)$, $W^{(k)}(t_r)$ and $W^{(k)}(t_s)$ is given by

$$\begin{aligned} & f_{1,r,s}(w_1, w_r, w_s) \\ &= \frac{k^{t_s}}{\Gamma(t_1)\Gamma(t_r - t_1)\Gamma(t_s - t_r)} w_1^{t_1 - 1} (w_r - w_1)^{t_r - t_1 - 1} (w_s - w_r)^{t_s - t_r - 1} e^{-kw_s}, \end{aligned}$$

for $0 < w_1 < w_r < w_s < \infty$, where for simplicity we write W_i instead of $W^{(k)}(t_i)$. On the other hand, by using the linear transformations $U = W_1, V = W_r - W_1$ and $W = W_s - W_r$, the joint pdf of the rv's U, V and W is

$$(3.9) \quad \begin{aligned} & f_{U,V,W}(u, v, w) \\ &= \frac{k^{t_s}}{\Gamma(t_1)\Gamma(t_r - t_1)\Gamma(t_s - t_r)} u^{t_1-1} v^{t_r-t_1-1} w^{t_s-t_r-1} \exp(-k(u + v + w)), \end{aligned}$$

with $u > 0, v > 0, w > 0$. The joint pdf of U, V and $P_2 = W/V$ can be written as

$$\begin{aligned} & f_{U,V,P_2}(u, v, p_2) \\ &= \frac{k^{t_s}}{\Gamma(t_1)\Gamma(t_r - t_1)\Gamma(t_s - t_r)} u^{t_1-1} v^{t_s-t_1-1} p_2^{t_s-t_r-1} \exp(-k[u + (1 + p_2)v]), \end{aligned}$$

for $u > 0, v > 0, p_2 > 0$. Thus, the pdf of the pivotal quantity P_2 is

$$f_{p_2}(p_2) = \int_0^\infty \int_0^\infty f_{U,V,P_2}(u, v, p_2) du dv.$$

By evaluating the above integration, we get (3.7) and (3.8). Moreover, we have

$$\begin{aligned} 1 - \delta &= F_{P_2}(p_2) = P(P_2 \leq p_2(\delta)) = P\left(0 < \frac{W_s - W_r}{W_r - W_1} \leq p_2(\delta)\right) \\ &= P(W_r < W_s \leq W_r + p_2(\delta)(W_r - W_1)) = P(L < Y^{(k)}(t_s) \leq U_{P_2}). \end{aligned}$$

Furthermore, the expected interval width for the PCI of $W^{(k)}(t_s)$ is given by

$$E[p_2(\delta)(W_r - W_1)] = \frac{p_2(\delta)}{k}(t_r - t_1).$$

Finally, we can obtain the point predictor, $\hat{W}^{(k)}(t_s)$, as in Theorem 3.1. By the same method of [2] (with a suitable modifications), it is not difficult to verify that $\hat{W}^{(k)}(t_s)$ is the BLIP. Hence the theorem. \square

4. PREDICTION OF FUTURE LOWER RECORD-VALUES PROCESS

In this section, the predictive pivotal quantities presented in Section 3, will be modified to construct prediction intervals of future fractional lower record-values from continuous distributions.

Theorem 4.1. *Let $0 = t_0 < t_1 < t_2 < \dots < t_n$ be positive real numbers and $X^{(k)}(t_1), X^{(k)}(t_2), \dots, X^{(k)}(t_r)$ be the first r fractional k -th lower record-values from a continuous distribution whose pdf f and cdf F . Then the pdf and*

the cdf of the pivotal quantity $P_1^* = (Z^{(k)}(t_s) - Z^{(k)}(t_r)) / Z^{(k)}(t_r)$ are given by (3.1) and (3.2), respectively, where

$$(4.1) \quad Z^{(k)}(t_i) = \log F(X^{(k)}(t_i)), \quad i = 1, 2, \dots, n.$$

A $100(1 - \delta)\%$ PCI for the future fractional k -th lower record-value $X^{(k)}(t_s)$ is $(L_{P_1^*}, U)$ where

$$L_{P_1^*} = F^{-1} \left(\left(F(X^{(k)}(t_r)) \right)^{1+p_1^*(\delta)} \right), \quad U = X^{(k)}(t_r),$$

and p_1^* can be obtained by solving the non linear equation $F_{P_1^*}(p_1^*) = 1 - \delta$. Moreover, when $X^{(k)}(t_i) = Z^{(k)}(t_i), i = 1, 2, \dots, n$ (i.e., F is NEXP(1)), the expected interval width of the PCI is $\frac{p_1^*(\delta)t_r}{k}$.

Proof: Since the transformation $Z = \log F(y)$ is one to one and onto (monotone increasing function), the pdf $f_{Z^{(k)}(t)}(z)$ of the rv $Z^{(k)}(t)$ is given by $f_{Z^{(k)}(t)}(z) = |J|f_{X^{(k)}(t)}(x(z))$, where $|J| = \frac{dx}{dz} = \frac{1}{f(x)}e^z$. Therefore, by (2.7) we have $f_{Z^{(k)}(t)}(z) = \frac{k^t}{\Gamma(t)}(-z)^{t-1}e^{kz}, z < 0$, which is the pdf of fractional k -th lower record-values based on NEXP(1). The remaining part of the proof is similar to the corresponding part of the proof of Theorem 3.1, with only obvious changes. \square

Theorem 4.2. Let $0 = t_0 < t_1 < t_2 < \dots < t_n$ be positive real numbers and $X^{(k)}(t_1)$ and $X^{(k)}(t_r)$ be the first and the r -th fractional k -th lower record-values from a continuous distribution whose pdf f and cdf F . Then the pdf and the cdf of the pivotal quantity

$$(4.2) \quad P_2^* = \frac{Z^{(k)}(t_r) - Z^{(k)}(t_s)}{Z^{(k)}(t_1) - Z^{(k)}(t_r)}, \quad s > r > 1,$$

are given by (3.7) and (3.8) respectively, with $Z^{(k)}(t_i) = \log F(X^{(k)}(t_i)), i = 1, 2, \dots, n$. A $100(1 - \delta)\%$ PCI for the future k -th lower record-value $X^{(k)}(t_s)$ is $(L_{P_2^*}, U)$ where

$$L_{P_2^*} = F^{-1} \left(F(X^{(k)}(t_r)) \left(\frac{F(X^{(k)}(t_r))}{F(X^{(k)}(t_1))} \right)^{p_2^*(\delta)} \right), \quad U = X^{(k)}(t_r),$$

and $p_2^*(\delta)$ can be obtained by solving the non linear equation $F_{P_2^*}(p_2^*(\delta)) = 1 - \delta$.

Proof: The joint pdf of the fractional k -th lower record-values $Z^{(k)}(t_1), Z^{(k)}(t_r)$ and $Z^{(k)}(t_s)$ based on NEXP(1) is given by

$$\begin{aligned} & f_{1,r,s}(z_1, z_r, z_s) \\ &= \frac{k^{t_s}}{\Gamma(t_1)\Gamma(t_r - t_1)\Gamma(t_s - t_r)} (-z_1)^{t_1-1} (z_1 - z_r)^{t_r-t_1-1} (z_r - z_s)^{t_s-t_r-1} e^{kz_s}, \end{aligned}$$

$-\infty < z_s < z_r < z_1 \leq 0$. Now, consider the linear transformations $U^* = -Z_1$, $V^* = Z_1 - Z_r$ and $W^* = Z_r - Z_s$, the joint pdf of the rv's U^* , V^* and W^* is given by relation (3.9). Therefore, the rest of the proof is similar, with only obvious changes, to the corresponding part of the proof of Theorem 3.2. \square

Remark 4.1.

1. The preceding results can be proved by the independence between the components in each of the vectors $(W^{(k)}(t_r), W^{(k)}(t_s) - W^{(k)}(t_r))$, $(W^{(k)}(t_1), W^{(k)}(t_r) - W^{(k)}(t_1), W^{(k)}(t_s) - W^{(k)}(t_r))$, $(Z^{(k)}(t_r), Z^{(k)}(t_s) - Z^{(k)}(t_r))$ and $(Z^{(k)}(t_1), Z^{(k)}(t_r) - Z^{(k)}(t_1), Z^{(k)}(t_s) - Z^{(k)}(t_r))$.
2. The lower and the upper limits of the PCI for future fractional k -th upper (lower) record-values depend on the population cdf F .
3. The ordinary upper (lower) record-values are obtained as special cases from the presented methods by setting $t_i = i$, for all $i = 1, 2, \dots, n$.
4. All the preceding results remain valid if we replace $t_i = i$, $i = 1, 2, \dots, r$, that is, fractional k -th upper (lower) record-values can be predicted via ordinary k -th upper (lower) record-values.

5. SIMULATION STUDIES

In this section, simulation studies are conducted to demonstrate the efficiency of the presented results. For this purpose, three algorithms are established in Appendix A.

Let us first check the validity of the first two algorithms, by generating ten fractional upper records $Y^{(1)}(t_i), i = 1, 2, \dots, 10$, (see Table 1) based on Weibull distribution with shape and scale parameters $\alpha = 3$ and $\beta = 30$, respectively. It is easy to compute the theoretical expectation of each of these records, namely,

$$(5.1) \quad E[Y^{(k)}(t_i)] = \beta k^{-\frac{1}{\alpha}} \frac{\Gamma(t_i + 1/\alpha)}{\Gamma(t_i)}, \quad i = 1, 2, \dots, 10.$$

The idea of this simple test is to compare the theoretical value $E[Y^{(k)}(t_i)]$ with the estimated value resulted from application of the algorithms, i.e., the average value $\bar{Y}^{(k)}(t_i)$. In order to compute the average value of each of these records, we repeat the generation processes of these ten records, different values of times, $M = 10^3, 10^4, 10^5, 10^6$, and for each of these replicates M , we compute the average $\bar{Y}^{(k)}(t_i)$, for each i . Table 1 summarizes these computations and shows that the theoretical expectations for all records are close to the estimated values, which are resulted via the application of the two algorithms.

All Computations are performed by using Mathematica version 10 with processor: Intel(R) Core(TM) i7-2640 cpu @ 2.80GHz 2.80GHz, RAM 4.00GB, and system type 64-bit operating system.

Table 1: A comparison between $\bar{Y}^{(1)}(t_i)$ and $E[Y^{(1)}(t_i)]$.

i	1	2	3	4	5	6	7	8	9	10	Run Time(s)
t_i	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	5.5	
$\bar{Y}^{(1)}(t_i),$ $M=10^3$	26.7135	31.7547	35.6315	38.8247	41.5164	43.8532	46.0244	48.2064	50.0735	51.8080	0.515
$\bar{Y}^{(1)}(t_i),$ $M=10^4$	26.8238	31.8600	35.7033	38.8794	41.6359	44.0795	46.3186	48.3856	50.2302	51.9270	2.840
$\bar{Y}^{(1)}(t_i),$ $M=10^5$	26.7416	31.8409	35.7152	38.9153	41.6562	44.1004	46.2950	48.3030	50.1555	51.8929	25.896
$\bar{Y}^{(1)}(t_i),$ $M=10^6$	26.7875	31.8445	35.7116	38.9043	41.6630	44.0974	46.2920	48.2962	50.1461	51.8737	267.182
$E[Y^{(1)}(t_i)]$	26.7894	31.8425	35.7192	38.9186	41.6724	44.1078	46.3026	48.3085	50.1612	51.8869	0.265

5.1. Exact and numerical computations

The exact expected values of the upper limits for the future fractional k -th upper record-value, $Y^{(k)}(t_s)$, from the exponential distribution with mean $1/\lambda$, based on the pivotal quantities P_1 and P_2 , respectively, are given by

$$E[U_{P_1}] = \frac{(1 + p_1(\delta))t_r}{\lambda k} \quad \text{and} \quad E[U_{P_2}] = \frac{1}{\lambda k} [t_r + p_2(\delta)(t_r - t_1)].$$

Moreover, the exact mean square errors of U_{P_1} and U_{P_2} , respectively, are given by

$$\begin{aligned} MSE_{U_{P_1}} &= E \left[U_{P_1} - Y^{(k)}(t_{s+1}) \right]^2 \\ &= \frac{1}{(\lambda k)^2} \left[(1 + p_1(\delta))^2 t_r (1 + t_r) + t_{s+1} (1 + t_{s+1}) - 2 t_r (1 + t_{s+1}) (1 + p_1(\delta)) \right] \end{aligned}$$

and

$$\begin{aligned} MSE_{U_{P_2}} &= E \left[U_{P_2} - Y^{(k)}(t_{s+1}) \right]^2 \\ &= \frac{1}{(\lambda k)^2} \left\{ p_2(\delta)(t_r - t_1) [p_2(\delta)(t_r - t_1 + 1) - 2(t_{s+1} - t_r)] + (t_{s+1} - t_r)(t_{s+1} - t_r + 1) \right\}. \end{aligned}$$

The mean square predictive errors, based on P_1 and P_2 respectively, are

$$MSE_{P_1} = E \left[\tilde{Y}^{(k)}(t_s) - Y^{(k)}(t_s) \right]^2 = \frac{t_s(t_s - t_r)}{(\lambda k)^2 t_r}, \quad r \geq 1,$$

and

$$MSE_{P_2} = E \left[\hat{Y}^{(k)}(t_s) - Y^{(k)}(t_s) \right]^2 = \frac{(t_s - t_r)(t_s - t_1)}{(\lambda k)^2 (t_r - t_1)}, \quad s > r > 1,$$

where $\tilde{Y}^{(k)}(t_s)$ and $\hat{Y}^{(k)}(t_s)$ denote the point predictors of $Y^{(k)}(t_s)$ based on the pivotal quantities P_1 and P_2 , respectively.

Remark 5.1. Clearly,

$$MSE_{P_2} - MSE_{P_1} = \frac{t_1(t_s - t_r)^2}{(\lambda k)^2 t_r(t_r - t_1)} > 0.$$

That is, $MSE_{P_2} > MSE_{P_1}$, for all $s > r > 1$.

The estimated root mean square errors for the upper limits of the PCI, respectively are defined by

$$RM\hat{S}E_{P_j} = \left[\frac{1}{M-1} \sum_{i=1}^M \left(U_{P_j}(i) - Y^{(k)}(t_{i+1}) \right)^2 \right]^{1/2}, \quad j = 1, 2.$$

Throughout this paper the following abbreviations are used:

- $\bar{Y}^{(k)}(t_s)$: The mean of fractional k -th record-value, $Y^{(k)}(t_s)$, is defined by $\bar{Y}^{(k)}(t_s) = \frac{1}{M} \sum_{i=1}^M Y_i^{(k)}(t_s)$, where M denote the number of replicants.
- PCI*: The predictive confidence interval of future fractional upper record.
- $CP_i\%$: The percent of coverage probability based on P_i , $i = 1, 2$, at $\delta = 0.10$.
- (\bar{L}, \bar{U}_{P_i}) : The average lower (upper) limits for the PCI of future fractional upper record.
- $(\bar{L}_{P_i^*}, U)$: The average lower (upper) limits for the PCI of future fractional lower record.
- BLUP*: Best linear unbiased predictor for future fractional upper record.
- $E[U_{P_i}]$: The expected value of the upper limit of the PCI based on P_i , $i = 1, 2$.
- $RMSE_{P_i}$: The exact root mean square error for the upper limit of the PCI based on P_i , $i = 1, 2$.
- $RM\hat{S}E_{P_i}$: The estimated root mean square error for the upper limit of the PCI based on P_i , $i = 1, 2$.

The rest of this section contains illustrations of the purposed methods through two simulation studies. The first study for EXP(0.1) is based on $M = 10^5$ replicates of $n = 25$ k -th upper records (including 13 ordinary records and 12 fractional records) corresponding to $t_i = 1, 1.5, 2, 2.5, \dots, 13$, $k = 2$. In this study the

first $r = 15$ upper records $Y^{(2)}(1), Y^{(2)}(1.5), \dots, Y^{(2)}(8)$ are assumed to be known and the next future 9 upper records, $Y^{(2)}(8.5), Y^{(2)}(9), \dots, Y^{(2)}(12.5)$ are to be predicted. The results which are shown in Table 2, include 90% coverage probability, two point predictors as well as two prediction intervals and the expected values of the upper limits. Moreover the exact root mean square errors for the point predictors, exact and estimated root mean square errors for the upper limits are given between parentheses. It is worth to mention here that the PCI's as well as the point predictors does not depend on the scale parameter β .

Table 2: Prediction of future ordinary and fractional upper records from EXP(0.1) based on $M = 10^5$ replicates.

t_r	t_s	$CP_1\%$	$CP_2\%$	\bar{L}	$\bar{Y}^{(2)}(t_s)$	$\tilde{Y}^{(2)}(t_s)$	$\hat{Y}^{(2)}(t_s)$	\bar{U}_{P_1}	$E[U_{P_1}]$	\bar{U}_{P_2}	$E[U_{P_2}]$
8.0	8.5	89.914	89.958	40.026	42.525	42.527 (3.644)	42.531 (3.660)	47.651 (6.283)	47.620 (6.255)	47.798 (6.449)	47.756 (6.418)
	9.0	89.998	89.993	40.026	45.031	45.029 (5.303)	45.037 (5.345)	53.375 (9.698)	53.341 (9.688)	53.688 (10.108)	53.632 (10.082)
	9.5	90.064	90.092	40.026	47.527	47.531 (6.673)	47.542 (6.748)	58.501 (12.820)	58.464 (12.816)	58.983 (13.467)	58.917 (13.440)
	10.0	90.014	90.032	40.026	50.037	50.032 (7.906)	50.047 (8.018)	63.368 (15.751)	63.327 (15.740)	64.025 (16.638)	63.947 (16.597)
	10.5	89.989	90.022	40.026	52.531	52.534 (9.057)	52.553 (9.210)	68.091 (18.544)	68.047 (18.535)	68.925 (19.676)	68.837 (19.628)
	11.0	89.967	90.024	40.026	55.040	55.035 (10.155)	55.058 (10.351)	72.722 (21.246)	72.675 (21.242)	73.737 (22.626)	73.638 (22.574)
	11.5	90.036	90.097	40.026	57.541	57.537 (11.215)	57.564 (11.456)	77.290 (23.887)	77.240 (23.887)	78.489 (25.518)	78.380 (25.461)
	12.0	89.998	90.132	40.026	60.055	60.039 (12.247)	60.069 (12.536)	81.812 (26.499)	81.760 (26.484)	83.197 (28.380)	83.078 (28.303)
	12.5	89.983	90.063	40.026	62.550	62.540 (13.258)	62.575 (13.595)	86.301 (29.053)	86.245 (29.046)	87.873 (31.193)	87.744 (31.112)
9.5	10.0	89.993	90.020	47.527	50.037	50.028 (3.627)	50.031 (3.638)	55.006 (6.100)	54.975 (6.083)	55.106 (6.211)	55.066 (6.190)
	10.5	90.003	90.144	47.527	52.531	52.529 (5.257)	52.536 (5.286)	60.562 (9.275)	60.528 (9.270)	60.774 (9.547)	60.723 (9.530)
	11.0	90.041	90.123	47.527	55.040	55.031 (6.589)	55.040 (6.642)	65.514 (12.145)	65.478 (12.152)	65.840 (12.579)	65.781 (12.565)
	11.5	90.034	90.145	47.527	57.541	57.532 (7.780)	57.545 (7.859)	70.202 (14.814)	70.163 (14.827)	70.646 (15.410)	70.577 (15.396)
	12.0	90.069	90.078	47.527	60.055	60.034 (8.885)	60.049 (8.993)	74.741 (17.366)	74.699 (17.370)	75.305 (18.124)	75.228 (18.096)
	12.5	89.965	90.104	47.527	62.550	62.535 (9.934)	62.554 (10.073)	79.185 (19.810)	79.140 (19.820)	79.871 (20.739)	79.786 (20.706)
11.0	11.5	89.975	89.967	55.040	57.541	57.542 (3.615)	57.545 (3.623)	62.417 (5.981)	62.371 (5.964)	62.490 (6.062)	62.437 (6.039)
	12.0	89.942	90.001	55.040	60.055	60.044 (5.222)	60.049 (5.244)	67.856 (8.990)	67.807 (8.976)	68.010 (9.187)	67.946 (9.161)
	12.5	90.016	90.042	55.040	62.550	62.546 (6.528)	62.554 (6.567)	72.687 (11.707)	72.634 (11.685)	72.923 (12.022)	72.851 (11.979)

In the second simulation study we assume that the first five ordinary k -th upper record-values, $Y^{(3)}(1), \dots, Y^{(3)}(5)$, have been observed from Weibull distribution with cdf,

$$F(y) = 1 - \exp \left[- \left(\frac{y}{\beta} \right)^\alpha \right], \quad y > 0, \quad \alpha > 0, \quad \beta > 0,$$

for $\alpha = 2.5$, $\beta = 40$ and we have to predict the next three ordinary k -th upper record-values and three fractional k -th upper record-values:

$$Y^{(3)}(5.5), \quad Y^{(3)}(6), \quad Y^{(3)}(6.5), \quad Y^{(3)}(7), \quad Y^{(3)}(7.5), \quad Y^{(3)}(8).$$

The prediction results are obtained in the following two situations:

- (a) The parameters are assumed to be known.
- (b) The parameters are unknown and should be estimated.

The maximum likelihood estimators (MLE's) of the parameters based on the first observed $r < n$ ordinary k -th upper record-values can be obtained by maximizing (2.6) (after replacing n with r). Namely,

$$(5.2) \quad \hat{\alpha} = \frac{r}{\sum_{i=1}^{r-1} \ln(Y^{(k)}(r)/Y^{(k)}(i))} \quad \text{and} \quad \hat{\beta} = \left(\frac{k}{r} \right)^{\frac{1}{\hat{\alpha}}} Y^{(k)}(r).$$

But the MLE's are biased and Wang and Ye [47] obtained the corrected unbiased estimators, which are

$$(5.3) \quad \tilde{\alpha} = \frac{r-2}{\sum_{i=1}^{r-1} \ln(Y^{(k)}(r)/Y^{(k)}(i))} \quad \text{and} \quad \tilde{\beta} = \frac{\Gamma(r)}{\Gamma(r+1/\tilde{\alpha})} \left(1 + \frac{\ln r}{r\tilde{\alpha}} \right)^{r-1} \hat{\beta}.$$

Moreover, an unbiased point predictor, $\tilde{Y}^{(k)}(t_s)$ based on P_1 is obtained, and is given by

$$(5.4) \quad \tilde{Y}^{(k)}(t_s) = \frac{\Gamma(t_r)\Gamma(t_s+1/\alpha)}{\Gamma(t_s)\Gamma(t_r+1/\alpha)} Y^{(k)}(t_r).$$

For each value of t_r and t_s in Table 3, the prediction results obtained based on the exact values of parameters are given in the first two lines, while when the parameters are estimated from (5.3), the prediction results are shown in the last two lines of the same value of t_r and t_s .

Table 3: Prediction of future ordinary and fractional k -th upper record-values with $k = 3$ from Weibull(2.5, 40) based on $M = 10^5$ replicates. The root mean square errors are between parentheses.

$t_r = r$	t_s	$CP_1\%$	$CP_2\%$	L	$\bar{Y}^{(3)}(t_s)$	$\tilde{Y}^{(3)}(t_s)$	U_{P_1}	U_{P_2}	$E[U_{P_1}]$	
5.0	5.5	89.969	89.894	47.9290	49.9014	49.9012 (2.7719)	53.6961 (4.4844)	53.9554 (4.7718)	53.6622	
		85.578	87.335	47.9290	49.9014	49.8351 (2.9267)	53.6967 (5.4088)	54.2708 (5.9271)		
	6.0	90.045	89.978	47.9290	51.7529	51.7633 (3.8825)	57.6233 (6.5626)	58.1218 (7.1876)	57.5869	
		83.096	85.544	47.9290	51.7529	51.6495 (4.3132)	57.6239 (8.5196)	58.7394 (9.6642)		
	6.5	90.026	89.940	47.9290	53.5099	53.5303 (4.7093)	60.8786 (8.2831)	61.5855 (9.1853)	60.8402	
		81.141	84.075	47.9290	53.5099	53.3848 (5.4889)	60.8792 (11.2884)	62.5114 (12.9732)		
	7.0	89.959	89.991	47.9290	55.2068	55.2142 (5.4164)	63.7723 (9.7591)	64.6657 (10.9042)	63.7320	
		79.384	82.639	47.9290	55.2068	55.0509 (6.5917)	63.7729 (13.8610)	65.9118 (16.0448)		
	7.5	89.954	89.986	47.9290	56.8109	56.8245 (6.0246)	66.4209 (11.0487)	67.4835 (12.4157)	66.3789	
		78.043	81.511	47.9290	56.8109	56.6559 (7.6409)	66.4216 (16.2993)	69.0622 (18.9581)		
	8.0	90.013	90.081	47.9290	58.3474	58.3693 (6.5493)	68.8847 (12.1988)	70.1027 (13.7668)	68.8412	
		76.884	80.553	47.9290	58.3474	58.2064 (8.6456)	68.8854 (18.6427)	72.0256 (21.7588)		
	6.0	6.5	90.137	90.106	51.7529	53.5099	53.5196 (2.4750)	56.8502 (3.9648)	57.0023 (4.1292)	56.8256
			86.360	87.745	51.7529	53.5099	53.4529 (2.5871)	56.8507 (4.5650)	57.1463 (4.8598)	
7.0		90.007	89.930	51.7529	55.2068	55.2031 (3.4957)	60.3395 (5.7578)	60.6362 (6.1244)	60.3135	
		84.132	86.076	51.7529	55.2068	55.0810 (3.8025)	60.3401 (7.0049)	60.9186 (7.6738)		
7.5		90.001	89.976	51.7529	56.8109	56.8131 (4.2573)	63.2519 (7.2441)	63.6779 (7.7866)	63.2247	
		82.368	84.637	51.7529	56.8109	56.6457 (4.8119)	63.2526 (9.1503)	64.1030 (10.1386)		
8.0		90.016	89.958	51.7529	58.3474	58.3576 (4.8783)	65.8570 (8.5096)	66.4008 (9.2079)	65.8286	
		80.930	83.479	51.7529	58.3474	58.1539 (5.7231)	65.8577 (11.1011)	66.9755 (12.3802)		
7.0	7.5	90.086	90.102	55.2068	56.8109	56.8169 (2.2653)	59.8084 (3.5675)	59.9066 (3.6735)	59.7786	
		86.856	87.946	55.2068	56.8109	56.7554 (2.3490)	59.8090 (3.9979)	59.9720 (4.1891)		
	8.0	90.164	90.231	55.2068	58.3474	58.3615 (3.1695)	62.9703 (5.1441)	63.1637 (5.3853)	62.9389	
		84.952	86.511	55.2068	58.3474	58.2460 (3.4058)	62.9709 (6.0285)	63.2955 (6.4710)		

6. DATA ANALYSIS

In the inverse sampling plan, one takes observations until a fixed number r of records is reached (cf. [28]). According to Hofmann and Nagaraja [28], the amount of Fisher information (FI) for both fixed sample and inverse sampling plan based on all upper records and their record times is greater than the amount of FI based on only upper records. Therefore, in this section, we shall estimate the parameters from the likelihood function for record-breaking (record-values and their inter-record times).

For the inverse sampling plan, the joint likelihood of the upper record-values Y_1, Y_2, \dots, Y_m and the inter-record times $\tau_1, \tau_2, \dots, \tau_m$ is given by

$$(6.1) \quad L(\mathbf{y}, \tau; \Theta) = f(y_1, \dots, y_m, \tau_1, \dots, \tau_m; \Theta) = \prod_{i=1}^m f(y_i; \Theta)(F(y_i; \Theta))^{\tau_i - 1},$$

where \mathbf{y} is the vector of observed upper records, $\tau_i, i = 1, 2, \dots, m - 1$, are the number of trials following the observation y_i that are needed to obtain the next upper record-value y_{i+1} with $\tau_m = 1$ and Θ is an unknown vector of parameters (e.g. [40], [28] and [35]). Similar result for lower record-breaking is given in [45], [29] and [27], that is

$$(6.2) \quad L^*(\mathbf{x}, \tau^*; \Theta) = f(x_1, \dots, x_m, \tau_1^*, \dots, \tau_m^*; \Theta) = \prod_{i=1}^m f(x_i; \Theta)(1 - F(x_i; \Theta))^{\tau_i^* - 1},$$

where \mathbf{x} is the vector of observed lower records, $\tau_i^*, i = 1, 2, \dots, m - 1$, is the number of trials needed, following x_i to obtain the next lower record x_{i+1} , $\tau_m^* = 1$ and $F(\cdot)$ is cdf of the population from which the sample is drawn. In the rest of this section, three examples to real data are analyzed.

Example 6.1 (Maximum annual temperature). The following data from Long Beach, California, represents the maximum annual temperature in Fahrenheit from 1990 to 2012:

86.7, 81.7, 84.3, 86.4, 84.9, 85.1, 89.7, 82.3, 84.2, 85.8, 81.5, 82.4,
84.3, 84.1, 90.5, 89.4, 87.5, 88.4, 90.3, 84.1, 88.4, 83.0, 86.6.

The upper records and inter-record times for the above data are $y_1 = 86.7, y_2 = 89.7, y_3 = 90.5$ and $\tau_1 = 6, \tau_2 = 8, \tau_3 = 1$. First we fit the complete data to some probability distributions. The preliminary fitting indicates that Weibull, extreme value, Frechet distributions are appropriate models for this data. Moreover, the maximum likelihood estimates (MLE's) of parameters are obtained based on (6.1) and then a comparison is performed according to Akaike information criterion (AIC) to select the best model. The results are summarized in Table 4.

Table 4: Comparison between three different distributions via the log Likelihood and AIC.

Model	Parameters	$\mathcal{L} = \text{Log } L$	AIC
Frechet(α, β)	$\hat{\alpha} = 52.7598, \hat{\beta} = 85.2217$	-11.347	26.694
EVD(α, β)	$\hat{\alpha} = 85.1787, \hat{\beta} = 1.67541$	-11.331	26.662
Weibull(α, β)	$\hat{\alpha} = 26.6179, \hat{\beta} = 86.4405$	-11.052	26.103

According to AIC Weibull distribution is better than extreme value distribution (EVD) and Frechet distributions. Based on the first three records the upper limits for the next two records and the next two half fractional records are obtained in Table 5.

Table 5: Point predictor and 95% PCI for the next two half record-values and the two record-values for annual maximum temperatures based on the first 3 records.

t_r	t_s	L	$\tilde{Y}^{(1)}(t_s)$	U_{P_1}	$E[U_{P_1}]$	U_{P_2}	$E[U_{P_2}]$
3	3.5	90.500	91.111	92.884	91.873	93.395	91.946
	4	90.500	91.633	93.960	92.936	94.720	93.533
	4.5	90.500	92.089	94.735	93.704	95.652	94.594
	5	90.500	92.494	95.358	94.320	96.385	95.406

Example 6.2 (Maximum annual earthquakes). The data consists of 151 magnitude of the annual maximum earthquakes in the United States during the period from 1769 to 1989 (some data are missing). The data are from Mathematica Documentation Center. The upper records and inter-record times for the annual maximum earthquakes are:

$$x_i = 6.0, \quad 6.5, \quad 7.2, \quad 7.4, \quad 7.6, \quad 7.9, \quad 8.0, \quad 8.3, \quad 8.4$$

$$\tau_i = 3, \quad 3, \quad 1, \quad 15, \quad 10, \quad 28, \quad 39, \quad 26, \quad 1$$

We proceed as in Example 6.1. According to AIC, and the Log likelihood function, Gumbel distribution is more suitable than several other distributions (including Weibull, EVD, Frechet distributions) for modeling the previous data. The cdf of Gumbel distribution is of the form

$$F(y) = 1 - \exp \left[-e^{(y-\alpha)/\beta} \right], \quad -\infty < y < \infty, \quad -\infty < \alpha < \infty, \quad \beta > 0.$$

The data are analyzed in the following two cases:

1. In the first case, we suppose that the first 8 record-values have been observed. The prediction results in the first 3 rows of Table 6 are obtained via the MLE's $\hat{\alpha} = 6.59296$ and $\hat{\beta} = 1.00387$, which are computed from (6.1).
2. In the second case, all the first 9 record-values are assumed to be observed. An application to (6.1) again yields, $\hat{\alpha} = 6.58459$ and $\hat{\beta} = 1.07983$, which are very close to the MLE's computed from the complete data. The prediction results according to these estimates are shown in the last two rows of Table 6.

In such cases a point predictor based on P_1 is given by

$$\tilde{Y}^{(1)}(t_s) = Y^{(1)}(t_r) + E \left[Y^{(1)}(t_s) \right] - E \left[Y^{(1)}(t_r) \right], \quad t_s > t_r,$$

where $E \left[Y^{(1)}(t_s) \right]$ and $E \left[Y^{(1)}(t_r) \right]$ are computed numerically.

Table 6: Point predictor and 95% upper limits for the next half record-value and the next record-value for the annual maximum earthquakes.

t_r	t_s	L	$\tilde{Y}^{(1)}(t_s)$	U_{P_1}	U_{P_2}
8	9	8.300	8.425	8.676	8.694
	9.5	8.300	8.537	8.776	8.799
	10	8.300	8.637	8.863	8.890
9	9.5	8.400	8.520	8.637	8.641
	10	8.400	8.628	8.759	8.767

Example 6.3 (One hour mean concentration of sulphur dioxide). The following data represents the monthly maxima of 1 h mean concentration of sulphur dioxide in parts per hundred million (pphm) from Long Beach, California, during 1956 to 1974 for the month of October:

26, 14, 27, 15, 16, 16, 11, 10, 14, 12,
15, 40, 29, 13, 20, 41, 31, 28, 11.

Roberts[44] shows that the Weibull model is a reasonably good for fitting this data. An application of extreme value $Q - Q$ plot by [16] supports Weibull model. The upper records and inter-record times for the above data are: $x_1 = 26$, $x_2 = 27$, $x_3 = 40$, $x_4 = 41$, and $\tau_1 = 2$, $\tau_2 = 9$, $\tau_3 = 4$, $\tau_4 = 1$. The MLE's of Weibull parameters based on the likelihood function (6.1) are $\hat{\alpha} = 2.3596$ and $\hat{\beta} = 24.5108$.

Based on the pivotal quantities P_1 and P_2 , 90% PCI's for the next two records, respectively, are (41, 52.328), (41, 52.103) and (41, 59.452), (41, 59.382), which are shorter than the intervals obtained by [47] ((41.1590, 60.2449) and (41.9011, 75.5765)). Moreover, an unbiased point predictors for the next two record-values are obtained from (5.4), that is, $\tilde{Y}^{(1)}(5) = 45.344$ and $\tilde{Y}^{(1)}(6) = 49.187$.

7. CONCLUSION

In this article, we have proposed two predictive pivotal quantities for constructing prediction intervals of future ordinary (fractional) upper (lower) records from any continuous distribution. More details have been given for the exponential distribution. Prediction intervals constructed using this approach have been demonstrated, by using a simulation study and by applying it to real data. Example 6.3 shows that this method gives a shorter intervals than that given by Wang and Ye [47]. Moreover, the second case in the simulation study as well as the three real data examples show that, when the cdf of the data is unknown as always in practice, the given method is applicable with acceptable degree of accuracy. Also, it is noted that the coverage probability is closed to theoretical value $1 - \delta = 0.90$ and average upper (lower) limits of PCI are closed to expected values of upper (lower) limits based on both P_1 and P_2 . Comparisons based on exact and estimated root mean square errors, indicate that the pivotal quantity P_1 is relatively better than P_2 . Moreover, the root mean square errors, increase with increasing of the difference $t_s - t_r$. Finally, three real data sets have been completely analyzed.

A. ALGORITHMS

Based on the results of Rider [43], Rahman [41], Cramer [18] and Burkschat *et al.* [15], we can generate ordinary k -th upper (lower) record-values from any continuous cdf F with pdf f , by the following algorithm.

A.1. Algorithm 1

- Step 1.** Choose the values of n , k and determine the cdf F ;
- Step 2.** Generate a random sample of size n from beta distribution, $Beta(k, 1)$, say $\mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_n$;

Step 3. Compute the k -th upper record-value $Y_r^{(k)}$, based on F by the formula

$$Y_r^{(k)} = F^{-1} \left(1 - \prod_{j=1}^r \mathcal{B}_j \right), \quad r = 1, 2, \dots, n;$$

Step 4. Compute the k -th lower record-value $X_r^{(k)}$, based on F from the relation

$$X_r^{(k)} = F^{-1} \left(\prod_{j=1}^r \mathcal{B}_j \right), \quad r = 1, 2, \dots, n.$$

The second algorithm relies on Theorem 1 and Definition 2 of Bieniek and Szynal [14]. The algorithm is formulated in a special case, whenever there is only a single fractional k -th upper record-value between two successive ordinary k -th upper record-values.

A.2. Algorithm 2

Step 1. Determine n , k and use Algorithm 1 to generate n ordinary k -th upper record-values, $W_i^{(k)}$, $i = 1, 2, \dots, n$, based on EXP(1);

Step 2. Choose the real numbers $0 = t_0 < t_1 < \dots < t_n$, such that, $i - 1 < t_i < i$, $\forall i = 1, \dots, n$;

Step 3. Compute the fractional k -th upper record-values based on EXP(1) by Theorem 1 of Bieniek and Szynal (2004), that is,

$$(A.1) \quad W^{(k)}(t_i) = (1 - B_i^*)W_{[t_i]}^{(k)} + B_i^*W_{[t_i+1]}^{(k)}, \quad i = 1, 2, \dots, n,$$

where $[t_i]$ denotes the greatest integer part of t_i , B_i^* is a random observation from beta distribution $Beta(t_i^*, 1 - t_i^*)$, independent of $W_{[t_i]}^{(k)}$, $i = 1, 2, \dots, n$, and t_i^* denotes the fractional part of the numerical value of t_i ;

Step 4. The fractional k -th upper record-values based on F , are then given by

$$(A.2) \quad Y^{(k)}(t_i) = F^{-1} \left(1 - e^{-W^{(k)}(t_i)} \right), \quad i = 1, 2, \dots, n.$$

The general case can be accomplished by Theorems 2 and 3 of [14].

A.3. Algorithm 3

- Step 1.** Determine the number n of fractional upper records to be generated, the number of repetitions M , the real numbers $0 = t_0 < t_1 < \dots < t_n$ with $i - 1 < t_i < i$, $\forall i = 1, \dots, n$ and the distribution with its parameter(s);
- Step 2.** Generate and store M arrays, each array include n of fractional k -th upper record-values;
- Step 3.** Determine the number of observed ordinary (fractional) k -th upper record-values r and the number of future ordinary (fractional) k -th upper record-value s , to be predicted;
- Step 4.** Find the numerical values of $p_i(\delta)$ by solving the nonlinear equations $F_{P_i}(p_i) = 1 - \delta$, $i = 1, 2$;
- Step 5.** Find the MLE's of the parameters based on the first r ordinary (fractional) k -th upper record-values;
- Step 6.** Compute the upper and lower limits for the *PCI* based on the pivotal quantities P_1 and P_2 by Theorems 3.1 and 3.2, and the point predictor(s) with
- (i) the true values of parameters, and
 - (ii) the MLE's of parameters;
- Step 7.** Check whether, the observed value of $Y^{(k)}(t_s)$ did belong to the *PCI*;
- Step 8.** Repeat Steps 5, 6 and 7, M times;
- Step 9.** Compute the percentage of coverage probability, that is the percent that the true value of the future fractional record lies inside the *PCI*, the average of the lower and upper limits;
- Step 10.** Compute the root mean square errors and expected values of upper limits based on P_1 and P_2 .

ACKNOWLEDGMENTS

The authors acknowledge the valuable suggestions from the referee and the associated editor for careful reading of the first version.

REFERENCES

- [1] AHSANULLAH, M. (1978). Record values and the exponential distribution, *Annals of the Institute of Statistical Mathematics*, **30**(1), 429–433.
- [2] AHSANULLAH, M. (1980). Linear prediction of record values for the two parameter exponential distribution, *Annals of the Institute of Statistical Mathematics*, **32**(1), 363–368.
- [3] AHSANULLAH, M. (2004). *Record values – theory and applications*, University Press of America.
- [4] AL-HUSSAINI, E.K. and AHMAD, A.A. (2003). On Bayesian interval prediction of future records, *Test*, **12**(1), 79–99.
- [5] ALY, A.E. (2015). Prediction and reconstruction of future and missing unobservable modified Weibull lifetime based on generalized order statistics, *Journal of the Egyptian Mathematical Society*, **24**(2), 309–318.
- [6] ARNOLD, B.C.; BALAKRISHNAN, N. and NAGARAJA, H.N. (1998). *Records*, John Wiley & Sons, New York.
- [7] BALAKRISHNAN, N.; AHSANULLAH, M. and CHAN, P.S. (1995). On the logistic record values and associated inference, *Journal of Applied Statistical Science*, **2**(3), 233–248.
- [8] BARAKAT, H.M.; EL-ADLL, M.E. and AMANY, E.ALY. (2011). Exact prediction intervals for future exponential lifetime based on random generalized order statistics, *Computers & Mathematics with Applications*, **61**(5), 1366–1378.
- [9] BARAKAT, H.M.; EL-ADLL, M.E. and AMANY, E.ALY. (2014). Prediction intervals of future observations for a sample of random size from any continuous distribution, *Mathematics and Computers in Simulation*, **97**, 1–13.
- [10] BARAKAT, H.M.; NIGM, E.M. and ALDALLAL, R.A. (2014a). Current records and record range with some applications, *Journal of the Korean Statistical Society*, **43**, 263–273.
- [11] BARAKAT, H.M.; NIGM, E.M. and ALDALLAL, R.A. (2014b). Exact prediction intervals for future current records and record range from any continuous distribution, *SORT*, **38**(2), 251–270.
- [12] BARAKAT, H.M.; NIGM, E.M.; EL-ADLL, M.E. and YUSUF, M. (2018). Prediction of future generalized order statistics based on exponential distribution with random sample size, *Statistical Papers*, **59**(2), 605–631.
- [13] BERRED, A.M. (1998). Prediction of record values, *Communications in Statistics – Theory and Methods*, **27**(9), 2221–2240.
- [14] BIENIEK, M. and SZYNAL, D. (2004). On the fractional record values, *Probability and Mathematical Statistics*, **24**(1), 27–46.
- [15] BURKSCHAT, M.; CRAMER, E. and KAMPS, U. (2003). Dual generalized order statistics, *Metron*, **61**(1), 13–26.
- [16] CHAN, P.S. (1993). *A statistical study of log-gamma distribution* (Ph.D. dissertation), McMaster University, Canada.
- [17] CHANDLER, K.N. (1952). The distribution and frequency of record values, *Journal of the Royal Statistical Society: Series B*, **14**(2), 220–228.

- [18] CRAMER, E. (2003). *Contributions to Generalized Order Statistics*, Habilitationsschrift, Reprint, University of Oldenburg.
- [19] DOGANAKSOY, N. and BALAKRISHNAN, N. (1997). A useful property of best linear unbiased predictors with applications to life-testing, *The American Statistician*, **51**(1), 22–28.
- [20] DUNSMORE, I.R. (1983). The future occurrence of records, *Annals of the Institute of Statistical Mathematics*, **35**(1), 267–277.
- [21] DZIUBDZIELA, W. and KOPOCIŃSKI, B. (1976). Limiting properties of the k -th record value, *Zastosowania Matematyki*, **15**, 187–190.
- [22] EL-ADLL, M.E. (2011). Predicting future lifetime based on random number of three parameters Weibull distribution, *Mathematics and Computers in Simulation*, **81**, 1842–1854.
- [23] EL-ADLL, M.E. and ALY, A.E. (2016). Prediction intervals of future generalized order statistics from Pareto distribution, *Journal of Applied Statistical Science*, **22**(1–2), 111–125.
- [24] GALAMBOS, J. (1987). *The Asymptotic Theory of Extreme Order Statistics*, Second edition, Krieger, Malabar, Florida.
- [25] GEISSER, S. (1993). *Predictive Inference: An Introduction*, Chapman and Hall, London.
- [26] GRUDZIEFI, Z. and SZYNAL, D. (1983). *On the expected values of k -th record values and associated characterizations of distributions*. In “Probability and Statistical Decision Theory”, Vol. A, 1985, (F. Konecny et al., Eds.), Reidel, Dordrecht, pp. 119–127.
- [27] GULATI, S. and PADGETT, W.J. (2003). *Parametric and nonparametric inference from record-breaking data*, Springer-Verlag, New York.
- [28] HOFMANN, G. and NAGARAJA, H.N. (2003). Fisher information in record data, *Metrika*, **57**, 177–193.
- [29] HOINKES, L.A. and PADGETT, W. (1994). Maximum likelihood estimation from record-breaking data for the Weibull distribution, *Quality and Reliability Engineering International*, **10**(1), 5–13.
- [30] HUTSON, A.D. (1999). Calculating nonparametric confidence intervals for quantiles using fractional order statistics, *Journal Journal of Applied Statistic*, **26**(3), 343–353.
- [31] JONES, M.C. (2002). On fractional uniform order statistics, *Statistics & Probability Letters*, **58**(1), 93–96.
- [32] KAMINSKY, K.S. and NELSON, P.I. (1975). Best Linear Unbiased Prediction of Order Statistics in Location and Scale Families, *Journal of the American Statistical Association*, **70**(349), 145–150.
- [33] KAMINSKY, K.S. and NELSON, P.I. (1998). *Prediction on order statistics*. In “Handbook of Statistics” (N. Balakrishnan and C.R. Rao, Eds.), 17, North Holland, Amsterdam, pp. 431–450.
- [34] KAMINSKY, K.S. and RHODIN, L.S. (1985). Maximum likelihood prediction, *Annals of the Institute of Statistical Mathematics*, **37**(1), 507–517.
- [35] KYZYLASLAN, F. and NADAR, M. (2015). Estimation with the generalized exponential distribution based on record values and inter-record times, *Journal of Statistical Computation and Simulation*, **85**(5), 978–999.

- [36] LAWLESS, J.F. (1971). A prediction problem concerning samples from the exponential distribution with applications in life testing, *Technometrics*, **13**(4), 725–730.
- [37] NAGARAJA, H.N. (1984). Asymptotic linear prediction of extreme order statistics, *Annals of the Institute of Statistical Mathematics*, **36**, 289–299.
- [38] NAGARAJA, H.N. (1988). Record values and related statistics a review, *Communications in Statistics – Theory and Methods*, **17**(1), 2223–2238.
- [39] NAGARAJA, H.N. (1995). *Prediction problems*. In “The Exponential Distribution: Theory and Appl.” (N. Balakrishnan and A.P. Basu, Eds.), pp. 139–163, Gordon and Breach, New York.
- [40] NEVZOROV, V.B. (2001). *Records: Mathematical Theory*, Translations of Mathematical Monographs, American Math. Soc., Providence, R.I.
- [41] RAHMAN, N.A. (1964). Some generalisations of the distributions of product statistics arising from rectangular populations, *Journal of the American Statistical Association*, **59**(306), 557–563.
- [42] RAQAB, M. and BALAKRISHNAN, N. (2008). Prediction intervals for future records, *Statistics & Probability Letters*, **78**(13), 1955–1963.
- [43] RIDER, P.R. (1955). The distribution of the product of maximum values in samples from a rectangular distribution, *Journal of the American Statistical Association*, **50**(272), 1142–1143.
- [44] ROBERTS, E. (1979). Review of statistics of extreme values with applications to air quality data: part II. Applications, *Journal of the Air Pollution Control Association*, **29**(7), 733–740.
- [45] SAMANIEGO, F.J. and WHITAKER, L.R. (1988). On estimating population characteristics from record-breaking observations. II: Nonparametric Results, *Naval Research Logistics Quarterly*, **35**(2), 221–236.
- [46] STIGLER, M.S. (1977). Fractional order statistics, with applications, *Journal of the American Statistical Association*, **72**(359), 544–550.
- [47] WANG, B.X. and YE, Z. (2015). Inference on the Weibull distribution based on record values, *Computational Statistics & Data Analysis*, **83**, 26–36.

REVSTAT – STATISTICAL JOURNAL

Background

Statistics Portugal (INE, I.P.), well aware of how vital a statistical culture is in understanding most phenomena in the present-day world, and of its responsibility in disseminating statistical knowledge, started the publication of the scientific statistical journal *Revista de Estatística*, in Portuguese, publishing three times a year papers containing original research results, and application studies, namely in the economic, social and demographic fields.

In 1998 it was decided to publish papers also 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, being now composed by national and international 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 three volumes special issue containing extended abstracts of the invited contributed papers presented at the 23rd European Meeting of Statisticians.

The name of the Journal has been 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

— The editorial policy will focus on publishing research articles at the highest level in the domains of Probability and Statistics with emphasis on the originality and importance of the research.

— All research articles will be refereed by at least two persons, one from the Editorial Board and another external.

— The only working language allowed will be English. — Four volumes are scheduled for publication, one in January, one in April, one in July and the other in October.

— The only working language allowed will be English. — Four volumes are scheduled for publication, one in January, one in April, one in July and the other in October.

Aims and Scope

The aim of REVSTAT is to publish articles of high scientific content in English, developing innovative statistical scientific methods and introducing original research, grounded in substantive problems.

REVSTAT covers all branches of Probability and Statistics.

Surveys of important areas of research in the field are also welcome.

REVSTAT is published, in English, by Statistics Portugal.

REVSTAT - Statistical Journal is an open access peer reviewed journal.

Abstract and Indexing Services

The REVSTAT is covered by the following abstracting/indexing services:

- Current Index to Statistics
- Google Scholar
- Mathematical Reviews® (MathSciNet®)
- Science Citation Index Expanded
- Zentralblatt für Mathematic
- Scimago Journal & Country Rank
- Scopus

Instructions to Authors, special-issue editors and publishers

The articles should be written in English and submitted in PDF format to the Editor-in chief (revstat@fc.ul.pt) and to the executive editor (revstat@ine.pt).

Submission of a paper means that it contains original work that has not been nor is about to be published elsewhere in any form.

Manuscripts (text, tables and figures) should be typed only in black on one side, in double-spacing, with a left margin of at least 3 cm and with less than 30 pages. The first page should include the name, institution and address of the author(s) and a summary of less than one hundred words, followed by a maximum of six key words and the AMS 2000 subject classification.

Authors are obliged to write the final version of accepted papers using LaTeX, in the REVSTAT style. This style (REVSTAT.sty), and examples file (REVSTAT.tex), which may be download to PC Windows System (Zip format), Macintosh, Linux and Solaris Systems (StuffIt format), and Mackintosh System (BinHex Format), are available in the REVSTAT link of the Statistics Portugal Website: <http://www.ine.pt/revstat/inicio.html>

Additional information for the authors, published papers and forthcoming papers are available at REVSTAT website.

Accepted papers

Authors of accepted papers are requested to provide the LaTeX files and also a postscript (PS) or an acrobat (PDF) file of the paper to the Secretary of REVSTAT: revstat@ine.pt.

Such e-mail message should include the author(s)'s name, mentioning that it has been accepted by REVSTAT.

The authors should also mention if encapsulated postscript figure files were included, and submit electronics figures separately in .tiff, .gif, .eps or .ps format. Figures must be a minimum of 300 dpi.

Copyright

Upon acceptance of an article, the author(s) will be asked to transfer copyright of the article to the publisher, Statistics Portugal, in order to ensure the widest possible dissemination of information, namely through the Statistics Portugal website (<http://www.ine.pt>).

After assigning copyright, authors may use their own material in other publications provided that REVSTAT is acknowledged as the original place of publication. The Executive Editor of the Journal must be notified in writing in advance.

Editorial Board

Editor-in-Chief

Isabel Fraga Alves, University of Lisbon, Portugal

Co-Editor

Giovani L. Silva, University of Lisbon, Portugal

Associate Editors

Marília Antunes, University of Lisbon, Portugal

Barry Arnold (2019), University of California, USA

Narayanaswamy Balakrishnan, McMaster University, Canada

Jan Beirlant, Katholieke Universiteit Leuven, Belgium

Graciela Boente (2019-2020), University of Buenos Aires, Argentina

Paula Brito, University of Porto, Portugal

Vanda Inácio de Carvalho, University of Edinburgh, UK

Arthur Charpentier, Université du Québec à Montréal, Canada

Valérie Chavez-Demoulin, University of Lausanne, Switzerland

David Conesa, University of Valencia, Spain

Charmaine Dean, University of Waterloo, Canada

Jorge Milhazes Freitas, University of Porto, Portugal

Alan Gelfand, Duke University, USA

Stéphane Girard, Inria Grenoble Rhône-Alpes, France

Wenceslao Gonzalez-Manteiga, University of Santiago de Compostela, Spain

Marie Kratz, ESSEC Business School, France

Victor Leiva, Pontificia Universidad Católica de Valparaíso, Chile

Maria Nazaré Mendes-Lopes, University of Coimbra, Portugal

Fernando Moura, Federal University of Rio de Janeiro, Brazil

John Nolan, American University, USA

Paulo Eduardo Oliveira, University of Coimbra, Portugal

Pedro Oliveira, University of Porto, Portugal

Carlos Daniel Paulino (2019-2021), University of Lisbon, Portugal

Arthur Pewsey, University of Extremadura, Spain

Gilbert Saporta, Conservatoire National des Arts et Métiers, France

Alexandra M. Schmidt, McGill University, Canada

Julio Singer, University of Sao Paulo, Brazil

Manuel Scotto, University of Lisbon, Portugal

Lisete Sousa, University of Lisbon, Portugal

Milan Stehlík, University of Valparaíso, Chile and LIT-JK University Linz, Austria

María Dolores Ugarte, Public University of Navarre, Spain

Executive Editor

José A. Pinto Martins, Statistics Portugal

Former Executive Editors

Maria José Carrilho, Statistics Portugal (2005-2015)

Ferreira da Cunha, Statistics Portugal (2003-2005)

Secretary

Liliana Martins, Statistics Portugal