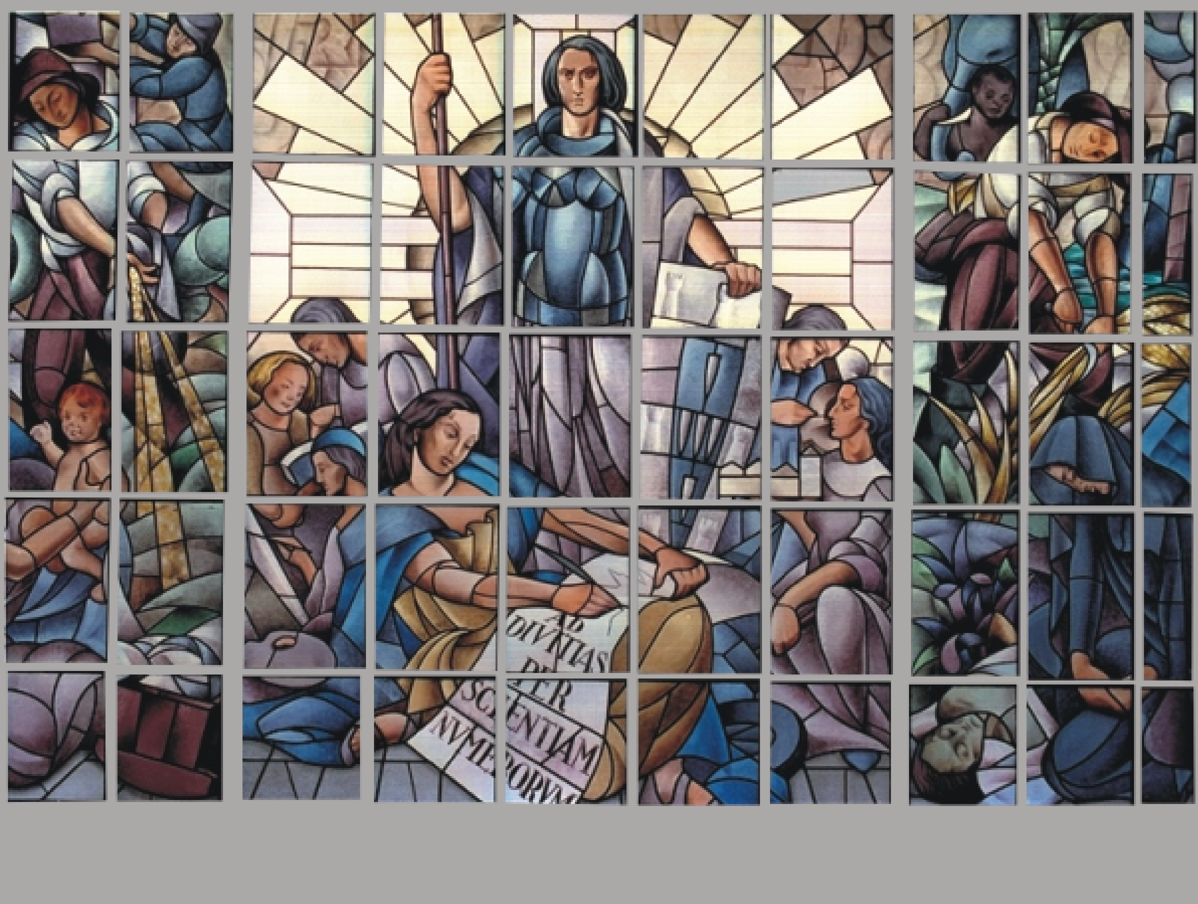
ISSN 1645-6726

Instituto Nacional de Estatística
Statistics Portugal

REVSTAT
Statistical Journal



Volume 15, No.4
October 2017

Catalogação Recomendada

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

CREDITS

- **EDITOR-IN-CHIEF**
  - *M. Ivette Gomes*
- **CO-EDITOR**
  - *M. Antónia Amaral Turkman*
- **ASSOCIATE EDITORS**
  - *Barry Arnold*
  - *Jan Beirlant*
  - *Graciela Boente*
  - *João Branco*
  - *Carlos Agra Coelho (2017-2018)*
  - *David Cox*
  - *Isabel Fraga Alves*
  - *Dani Gammerman (2014-2016)*
  - *Wenceslao Gonzalez-Manteiga*
  - *Juerg Huesler*
  - *Marie Husková*
  - *Victor Leiva*
  - *Isaac Meilijson*
  - *M. Nazaré Mendes- Lopes*
  - *Stephen Morghenthaler*
  - *António Pacheco*
  - *Carlos Daniel Paulino*
  - *Dinis Pestana*
  - *Arthur Pewsey*
  - *Vladas Pipiras*
  - *Gilbert Saporta*
  - *Julio Singer*
  - *Jef Teugels*
  - *Feridun Turkman*
- **EXECUTIVE EDITOR**
  - *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
    (National network) : 808 201 808
    Other networks: + 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**
  - *150 copies*
- **LEGAL DEPOSIT REGISTRATION**
  - *N.º 191915/03*
- **PRICE [VAT included]**
  - *€ 9,00*

© INE, Lisbon. Portugal, 2017* Reproduction authorised except for commercial purposes by indicating the source.

INDEX

Gamma-Admissibility in a Non-Regular Family with Squared-log Loss Function

Shirin Moradi Zahraie and *Hojatollah Zakerzadeh* 473

Estimation through Array-Based Group Tests

João Paulo Martins, Miguel Felgueiras and *Rui Santos* 487

Estimating Renyi Entropy of Several Exponential Distributions under an Asymmetric Loss Function

Suchandan Kayal and *Somesh Kumar* 501

Discriminating between Normal and Gumbel Distributions

Abdelaziz Qaffou and *Abdelhak Zoglat* 523

Confidence Intervals for Exceedance Probabilities with Application to Extreme Ship Motions

Dylan Glotzer, Vladas Pipiras, Vadim Belenky, Bradley Campbell and *Timothy Smith* 537

Reliability Estimation in Multistage Ranked Set Sampling

M. Mahdizadeh and *Ehsan Zamanzade* 565

Non Parametric ROC Summary Statistics

M.C. Pardo and *A.M. Franco-Pereira* 583

The Transmuted Birnbaum–Saunders Distribution

Marcelo Bourguignon, Jeremias Leão, Víctor Leiva and *Manoel Santos-Neto* 601

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.*

REVSTAT – Statistical Journal
Volume 15, Number 4, October 2017, 473–486

GAMMA-ADMISSIBILITY IN A NON-REGULAR FAMILY WITH SQUARED-LOG LOSS FUNCTION

Authors: Shirin Moradi Zahraie
– Department of Statistics, Yazd University, Yazd, Iran
sh_moradi_zahraie@stu.yazd.ac.ir

Hojatollah Zakerzadeh
– Department of Statistics, Yazd University, Yazd, Iran
hzaker@yazd.ac.ir

Received: July 2014 Revised: April 2015 Accepted: November 2015

Abstract:

- Review the admissibility of estimators under a vague prior information leads to the concept of gamma-admissibility. In this paper, the problem of estimation in a non-regular family of distributions under a squared log error loss function is considered. We find sufficient condition for a generalized Bayes estimator of a parametric function to be gamma-admissible. Some examples are given.

Key-Words:

- *Gamma-admissibility; generalized Bayes estimator; non-regular distribution; squared-log error loss function.*

AMS Subject Classification:

- 62C15, 62F15.

1. INTRODUCTION

Admissibility of estimator is an important problem in statistical decision theory; Consequently, this problem has been considered by many authors under various type of loss functions both in an exponential and in a non-regular family of distributions. For example under squared error loss function (Karlin (1958), Ghosh & Meeden (1977), Ralescu & Ralescu (1981), Sinha & Gupta (1984), Hoffmann (1985), Pulskamp & Ralescu (1991), Kim (1994) and Kim & Meeden (1994)), under entropy loss function (Sanjari Farsipour (2003,2007)) and under LINEX loss function (Tanaka (2010,2011,2012)) and squared-log error loss function (Zakerzadeh & Moradi Zahraie (2015)).

A Bayesian approach to a statistical problem requires defining a prior distribution over the parameter space. Many Bayesians believe that just one prior can be elicited. In practice, it is more frequently the case that the prior knowledge is vague and any elicited prior distribution is only an approximation to the true one. So, we elect to restrict attention to a given flexible family of priors and we choose one member from that family, which seems to best match our personal beliefs.

A gamma-admissible (Γ-admissible) approach is used which allows to take into account vague prior information on the distribution of the unknown parameter θ. The uncertainty about a prior is assumed by introducing a class Γ of priors. If prior information is scarce, the class Γ under consideration is large and a decision is close to a admissible decision. In the extreme case when no information is available the Γ-admissible setup is equivalent to the usual admissible setup. If, on the other hand, the statistician has an exactly prior information and the class Γ contains a single prior, then the Γ-admissible decision is an usual Bayes decision. So it is a middle ground between the subjective Bayes setup and full admissible.

Eichenauer-Herrmann (1992) gained a sufficient condition for an estimator of the form $(aX + b)/(cX + d)$ to be Γ-admissible under the squared error loss in a one-parameter exponential family.

The most popular convex and symmetric loss function is the squared error loss function which is widely used in decision theory due to its simple mathematical properties. However in some cases, it does not represent the true loss structure. This loss function is symmetric in nature i.e. it gives equal weightage to both over and under estimation. In real life, we encounter many situations where over-estimation may be more serious than under-estimation or vice versa. As an example, in construction an underestimate of the peak water level is usually much more serious than an overestimation.

The squared-log error loss function was introduced by Brown (1968). For an estimator δ of estimand $h(\theta)$, it is given by

$$L(h(\theta),\delta) = L(\nabla) := (\ln(\nabla))^2, \tag{1.1}$$

where both $h(\theta)$ and δ are positive and $\nabla := \delta/h(\theta)$.

We need the following definitions to express properties of the loss (1.1).

Definition 1.1. A real function $g(x)$ is *quasi-convex*, if for any given real number r, the set of all x such that $g(x) \leq r$ is convex. Any convex function is quasi-convex, but the converse is not necessarily true.

Definition 1.2. A loss function $L(h(\theta),\delta)$ is (for any $\varepsilon > 0$):

- *downside damaging* if $L(\delta-\varepsilon,\delta) \geq L(\delta+\varepsilon,\delta)$,
- *upside damaging* if $L(\delta-\varepsilon,\delta) \leq L(\delta+\varepsilon,\delta)$,
- *symmetric* if the loss function is both downside and upside damaging.

Remark 1.1. With downside damaging loss function, under-estimation is penalized more heavily, per unit distance, than over-estimation and with upside damaging loss function it is reversed.

Remark 1.2. If a loss function be downside damaging or upside damaging, then it is called *asymmetric*. By using asymmetric loss functions one is able to deal with cases where it is more damaging to miss the target on one side than the other.

Definition 1.3. The $L(h(\theta),\delta)$ is a *precautionary loss function* if and only if

(1) $L(h(\theta),\delta)$ is downside damaging, and

(2) for each fixed θ, $L(h(\theta),\delta) \to \infty$ as $\delta \to 0$.

Definition 1.4. The $L(h(\theta),\delta)$ is a *balanced loss function*, if $L(h(\theta),\delta) \to \infty$ as $\delta \to 0$ or $\delta \to \infty$. A balanced loss function takes both error of estimation and goodness of fit into account but the unbalanced loss function only considers error of estimation.

From Figure 1, we see that the loss (1.1) has the below properties:

(i) It is asymmetric.

(ii) It is quasi-convex.

(iii) It is a balanced loss function.

(iv) It is a precautionary loss function.

(v) When $0 < \nabla < 1$, it rises rapidly to infinity at zero; it has a unique minimum at $\nabla = 1$ and when $\nabla > 1$ it increases sublinearly.

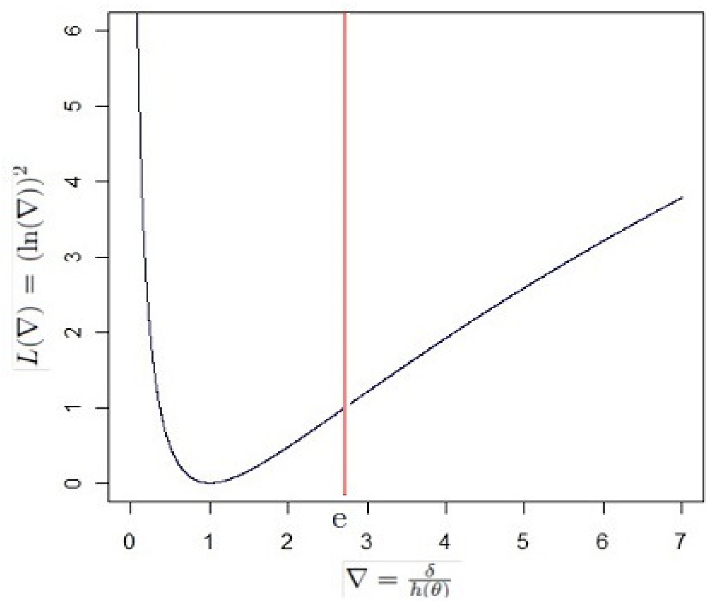


Figure 1: Plot of the squared-log error loss function.

For estimation under (1.1), see Sanjari Farsipour and Zakerzadeh (2005, 2006), Mahmoudi and Zakerzadeh (2011), Kiapour and Nematollahi (2011), Nematollahi and Jafari Tabrizi (2012) and Zakerzadeh and Moradi Zahraie (2015).

In this paper we consider the Γ-admissibility of generalized Bayes estimators in a non-regular family of distributions under the loss (1.1) where class Γ consists of all distributions which are compatible with the vague prior information. To this end, in Section 2, we state some preliminary definitions and results. In Section 3, we will obtain main theorem. Finally, in Section 4, we give an application of the Γ-admissibility in proving the Γ-minimaxity of estimators. Some examples are given.

2. PRELIMINARIES

2.1. Definition of Γ-admissibility

In the present paper it is assumed that vague prior density on the distribution of the unknown parameter θ is available. Let Π denote the set of all priors, i.e. Borel probability measures on the parameter interval Θ and Γ be a non-empty subset of Π. Suppose that the available vague prior information can be described by the set Γ, in the sense that Γ contains all prior which are compatible with the vague prior information.

Eichenauer-Herrmann (1992) defined the Γ-admissibility of an estimator as follows.

Definition 2.1. An estimator δ^* is called Γ-admissible, if

$$r(\pi,\delta) \le r(\pi,\delta^*), \qquad \pi \in \Gamma,$$

for some estimator δ implies that

$$r(\pi,\delta) = r(\pi,\delta^*), \qquad \pi \in \Gamma,$$

where $r(\pi,\delta)$ is the Bayes risk of δ.

Remark 2.1. From Definition 2.1, it is obvious that

– A Π-admissible estimator is admissible;

– A $\{\pi\}$-admissible estimator is simply a Bayes strategy with respect to the prior π;

– In general neither Γ-admissibility implies admissibility nor admissibility implies Γ-admissibility.

Hence, the available results on admissibility cannot be applied in order to prove the Γ-admissibility of an estimator. Consequently, it is necessary to study the problem of Γ-admissibility of estimators.

2.2. A non-regular family of distributions

Let X be a random variable whose probability density function with respect to some σ-finite measure μ is given by

$$f_X(x;\theta) = \begin{cases} q(\theta)r(x) & \underline{\theta} < x < \theta, \\ 0 & \text{otherwise}, \end{cases}$$

where $\theta \in \Theta := (\underline{\theta}, \bar{\theta})$ and Θ is a nondegenerate interval (possibly infinite) on the real line. Also $r(x)$ is a positive μ-measurable function of x and

$$q^{-1}(\theta) = \int_{\underline{\theta}}^{\theta} r(x) d\mu(x) < \infty$$

for $\theta \in \Theta$. This family is known as *a non-regular family of distributions.*

Suppose $\pi(\theta)$ be a prior (possibly improper) by its Lebesgue density $p_\pi(\theta)$ over Θ which is positive and continuous. Let $h(\theta)$ be a continuous function to be estimated from Θ to $\mathbb{R}$ and the loss to be (1.1). The generalized Bayes estimator of $h(\theta)$ with respect to $\pi(\theta)$ is given by $\delta_\pi(X)$, where

$$\delta_\pi(x) = \exp\left\{ \frac{\int_x^{\bar{\theta}} \{\ln(h(\theta))\} q(\theta) p_\pi(\theta) d\theta}{\int_x^{\bar{\theta}} q(\theta) p_\pi(\theta) d\theta} \right\} \tag{2.1}$$

for $\underline{\theta} < x < \bar{\theta}$, provided that the integrals in (2.1) exist and are finite.

3. MAIN RESULTS

In this section, the main results will obtain.

For some real number λ_0 let $a, b : [\lambda_0, \infty) \mapsto \Theta$ be continuously differentiable functions with $a(\lambda_0) < b(\lambda_0)$, where a and b are supposed to be strictly decreasing and strictly increasing, respectively. For $\lambda \geq \lambda_0$ a prior π_λ is defined by its Lebesgue density p_{π_λ} of the form

$$p_{\pi_\lambda}(\theta) := \left(\int_{a(\lambda)}^{b(\lambda)} p_\pi(t) dt \right)^{-1} I_{[a(\lambda), b(\lambda)]}(\theta) p_\pi(\theta).$$

Throughout this paper, we restrict estimators to the class

$$\Delta := \{\delta | \text{(A1) and (A2) are satisfied}\},$$

where

(A1) $E_\theta[\{\ln(\delta(X))\}^2] < \infty$ for all $\theta \in \Theta$;

(A2) $\int_{a(\lambda)}^{b(\lambda)} E_\theta[\{\ln(\frac{\delta(X)}{h(\theta)})\}^2] p_\pi(\theta) d\theta < \infty$ for $a(\lambda) < b(\lambda)$ and $\lambda \geq \lambda_0$.

Remark 3.1. In the statistical game (Γ, Δ, r), a Γ-admissible estimator is an admissible strategy of the second player.

The next lemma is essential to obtain our results.

Lemma 3.1. *Let $S(\theta)$ be a continuous and non-negative function over $\Theta = (\underline{\theta}, \bar{\theta})$ and $G(\lambda) := \int_{a(\lambda)}^{b(\lambda)} S(\theta)d\theta$. Suppose that there exists a positive function $R(\theta)$ such that*

$$G(\lambda) \leq 4(\min\{R(b(\lambda))b'(\lambda), -R(a(\lambda))a'(\lambda)\})^{-1/2}(G'(\lambda))^{1/2}$$

for $\lambda \geq \lambda_0$. If

$$\int_{\lambda_0}^{\infty} \min\{R(b(\lambda))b'(\lambda), -R(a(\lambda))a'(\lambda)\}d\lambda = \infty,$$

then $S(\theta) = 0$ for a.a. $\theta \in \Theta$.

Proof: See Eichenauer-Herrmann (1992). □

Theorem 3.1. *Suppose that $\delta_\pi \in \Delta$ and put*

$$K(x, \theta) := \int_x^{\theta} \{\ln(\frac{\delta_\pi(x)}{h(t)})\}q(t)p_\pi(t)dt,$$

and

$$\gamma(\theta) := \frac{1}{p_\pi(\theta)q(\theta)} \int_{\underline{\theta}}^{\theta} r(x)K^2(x, \theta)d\mu(x).$$

If $\pi_\lambda \in \Gamma$ for all $\lambda \geq \lambda_0$ and

$$\int_{\lambda_0}^{\infty} \min\{\gamma^{-1}(b(\lambda))b'(\lambda), -\gamma^{-1}(a(\lambda))a'(\lambda)\}d\lambda = \infty, \tag{3.1}$$

then $\delta_\pi(X)$ is Γ-admissible under the loss (1.1).

Proof: Let $\delta \in \Delta$ be an estimator such that $r(\pi, \delta) \leq r(\pi, \delta_\pi)$ for every prior $\pi \in \Gamma$. Since $\pi_\lambda \in \Gamma$ for $\lambda \geq \lambda_0$, we must have

$$\begin{aligned} 0 \; &\leq \; \left(\int_{a(\lambda)}^{b(\lambda)} p_\pi(t)dt\right) \{r(\pi_\lambda, \delta_\pi) - r(\pi_\lambda, \delta)\} \\ &= \int_{a(\lambda)}^{b(\lambda)} E_\theta[L(\delta_\pi, h(\theta)) - L(\delta, h(\theta))]p_\pi(\theta)d\theta \end{aligned}$$

for all $\theta \in \Theta$. From Condition (A1), we see that it is equivalent to

$$\begin{aligned} 0 \; &\leq \; \int_{a(\lambda)}^{b(\lambda)} E_\theta\left[\left\{\ln\left(\frac{\delta(X)}{\delta_\pi(X)}\right)\right\}^2\right]p_\pi(\theta)d\theta \\ &\leq \; 2\int_{a(\lambda)}^{b(\lambda)} E_\theta\left[\left\{\ln\left(\frac{\delta_\pi(X)}{h(\theta)}\right)\right\}\left\{\ln\left(\frac{\delta_\pi(X)}{\delta(X)}\right)\right\}\right]p_\pi(\theta)d\theta, \end{aligned}$$

for all $\theta \in \Theta$.

An application of the Fubini's theorem gives

$$\begin{aligned}
0 &\leq \int_{a(\lambda)}^{b(\lambda)} \int_{\underline{\theta}}^{\theta} \left\{ \ln\left(\frac{\delta(x)}{\delta_\pi(x)}\right)\right\}^2 r(x) q(\theta) p_\pi(\theta) d\mu(x) d\theta \\
&\leq 2 \int_{\underline{\theta}}^{b(\lambda)} \left[\int_x^{b(\lambda)} \left\{ \ln\left(\frac{\delta_\pi(x)}{h(\theta)}\right)\right\} p_\pi(\theta) q(\theta) d\theta \right] \left\{ \ln\left(\frac{\delta_\pi(x)}{\delta(x)}\right)\right\} r(x) d\mu(x) \\
&- 2 \int_{\underline{\theta}}^{a(\lambda)} \left[\int_x^{a(\lambda)} \left\{ \ln\left(\frac{\delta_\pi(x)}{h(\theta)}\right)\right\} p_\pi(\theta) q(\theta) d\theta \right] \left\{ \ln\left(\frac{\delta_\pi(x)}{\delta(x)}\right)\right\} r(x) d\mu(x),
\end{aligned}$$

which is guaranteed by Condition (A2).

Applying the Cauchy-Schwartz inequality, the first term of the right-hand side in the above equation, is less than

$$2 \left\{ \int_{\underline{\theta}}^{b(\lambda)} \left\{ \ln\left(\frac{\delta(x)}{\delta_\pi(x)}\right)\right\}^2 r(x) d\mu(x) \right\}^{1/2} \left\{ \int_{\underline{\theta}}^{b(\lambda)} r(x) K^2(x, b(\lambda)) d\mu(x) \right\}^{1/2}.$$

Hence, if we define

$$T(\theta) := \int_{\underline{\theta}}^{\theta} \left\{ \ln\left(\frac{\delta(x)}{\delta_\pi(x)}\right)\right\}^2 r(x) d\mu(x),$$

then we have

$$\begin{aligned}
0 &\leq \int_{a(\lambda)}^{b(\lambda)} T(\theta) q(\theta) p_\pi(\theta) d\theta \\
&\leq 2\{T(b(\lambda)) b'(\lambda) q(b(\lambda)) p_\pi(b(\lambda))\}^{1/2} \{\gamma^{-1}(b(\lambda)) b'(\lambda)\}^{-1/2} \\
&\quad +2\{-T(a(\lambda)) a'(\lambda) q(a(\lambda)) p_\pi(a(\lambda))\}^{1/2} \{-\gamma^{-1}(a(\lambda)) a'(\lambda)\}^{-1/2} \\
&\leq 4 \left(\min\{\gamma^{-1}(b(\lambda)) b'(\lambda), \gamma^{-1}(a(\lambda)) a'(\lambda)\}\right)^{-1/2} \\
&\quad \times \left(T(b(\lambda)) q(b(\lambda)) p_\pi(b(\lambda)) b'(\lambda) - T(a(\lambda)) q(a(\lambda)) p_\pi(a(\lambda)) a'(\lambda)\right)^{1/2}
\end{aligned}$$

for $\lambda \geq \lambda_0$, where the definition of the function $\gamma(\theta)$ has been used. Now a continuous, differentiable and increasing function $H : [\lambda_0, \infty] \to \mathbb{R}$ is defined by

$$H(\lambda) := \int_{a(\lambda)}^{b(\lambda)} T(\theta) q(\theta) p_\pi(\theta) d\theta.$$

So the above inequality can be written in the form

$$H(\lambda) \leq 4 \left(\min\{\gamma^{-1}(b(\lambda)) b'(\lambda), -\gamma^{-1}(a(\lambda)) a'(\lambda)\right)^{-1/2} (H'(\lambda))^{1/2}$$

for $\lambda \geq \lambda_0$. Therefore, from Lemma 3.1 we obtain $T(\theta) = 0$ for $a.a.\theta \in \Theta$, and consequently from (A1), we have $\delta(x) = \delta_\pi(x)$ $a.e.\mu$. This completes the proof. □

Remark 3.2. $K(x,\theta)$ can expressed as

$$K(x,\theta)=\frac{\int_x^{\theta}\int_{\theta}^{\bar{\theta}}\{\ln(\frac{h(s)}{h(t)})\}q(s)p_\pi(s)q(t)p_\pi(t)dsdt}{\int_x^{\bar{\theta}}q(u)p_\pi(u)du}$$

by (2.1) and the symmetry of the integrand.

Example 3.1. Suppose that X be a random variable according to an exponential distribution whose probability density function is given by

$$f_X(x,\theta)=\begin{cases} e^{x-\theta} & x<\theta,\\ 0 & x>\theta,\end{cases}$$

where $\theta(\in\mathbb{R})$ is unknown. The Generalized Bayes estimator of $h(\theta)=e^{\theta}$ with respect to the Lebesgue prior is given by $\delta_\pi(X)=\exp\{X+1\}$ which is of the form $ah(X)(a>0)$. A direct calculation gives $K(x,\theta)=e^{-\theta}(\theta-x)$ and $\gamma(\theta)=2$. Let class Γ_0 consists of all priors with mean 0, i.e., $\Gamma_0:=\{\pi\in\Pi|\int_\Theta\theta p_\pi(\theta)d\theta=0\}$. Define functions a and b by $a(\lambda)=-\lambda$ and $b(\lambda)=\lambda$ for $\lambda\geq\lambda_0>0$, i.e., the prior π_λ is the uniform distribution on the interval $[-\lambda,\lambda]$. Hence, $\pi_\lambda\in\Gamma_0$ for all $\lambda\geq\lambda_0$. Since (3.1) is satisfied, Theorem 3.1 implies that $\delta_\pi(X)$ is Γ_0-admissible under the loss (1.1).

Remark 3.3. It is difficult to express $\gamma(\theta)$ explicitly and it can have a complicated form, so to apply Theorem 3.1, we have to seek the suitable upper bound of $\gamma(\theta)$. For the case when $h(\theta)$ is bounded, we can get the next corollary.

Corollary 3.1. *Suppose that $h(\theta)$ is bounded and $\delta_\pi\in\Delta$. Put*

$$\tilde{K}(x,\theta):=\frac{\int_\theta^{\bar{\theta}}q(s)p_\pi(s)ds\int_x^{\theta}q(t)p_\pi(t)dt}{\int_x^{\bar{\theta}}q(u)p_\pi(u)du},$$

and

$$\tilde{\gamma}(\theta):=\frac{1}{p_\pi(\theta)q(\theta)}\int_{\underline{\theta}}^{\theta}r(x)\tilde{K}^2(x,\theta)d\mu(x).$$

If $\pi_\lambda\in\Gamma$ for all $\lambda\geq\lambda_0$ and

$$\int_{\lambda_0}^{\infty}\min\{\tilde{\gamma}^{-1}(b(\lambda))b'(\lambda),-\tilde{\gamma}^{-1}(a(\lambda))a'(\lambda)\}d\lambda=\infty,$$

then $\delta_\pi(X)$ is Γ-admissible under the loss (1.1).

Proof: It can be easily shown that there exists a constant C such that $|K(x,\theta)|\leq C\tilde{K}(x,\theta)$ for all $(x,\theta)\in\{(x,\theta)|\underline{\theta}<x<\theta<\bar{\theta}\}$. This completes the proof by Theorem 3.1. $\square$

Example 3.2. Suppose that $X_1, ..., X_n$ are independent and identically distributed random variables according to a uniform distribution over the interval $(0, \theta)$ where $\theta (\in \mathbb{R}^+)$ is unknown. Then the probability density function of the sufficient statistic $X = X_{(n)}$ is given by

$$f_X(x, \theta) = \begin{cases} \frac{n}{\theta^n} x^{n-1} & 0 < x < \theta, \\ 0 & \text{otherwise.} \end{cases}$$

Let $h(\theta)$ be bounded and $\pi(\theta) = 1/\theta$. We can easily obtain

$$\tilde{K}(x, \theta) = (1/(n\theta^n)) \{1 - (x/\theta)^n\},$$

and

$$\tilde{\gamma}(\theta) = \theta/(3n^2).$$

We assume that $\Gamma_m := \{\pi \in \Pi | \int_\Theta \theta p_\pi(\theta) d\theta = m\}$, i.e., Γ_m consists of all priors with mean m. Define functions a and b by $a(\lambda) = m \ln(\lambda)/(\lambda - 1)$ and $b(\lambda) = \lambda a(\lambda)$ for $\lambda \geq \lambda_0 > 1$. Since

$$\int_\Theta \theta p_{\pi_\lambda}(\theta) d(\theta) = \left(\int_{a(\lambda)}^{b(\lambda)} \frac{1}{t} dt \right)^{-1} (b(\lambda) - a(\lambda)) = m$$

for all $\lambda \geq \lambda_0$, so that $\pi_\lambda \in \Gamma_m$. A short calculation yields $a'(\lambda) = m \frac{\lambda - 1 - \lambda \ln(\lambda)}{\lambda(\lambda-1)^2} < 0$ and $b'(\lambda) = m \frac{\lambda - 1 - \ln(\lambda)}{(\lambda-1)^2} > 0$ for $\lambda \geq \lambda_0$. Because of $\lambda - 1 - \ln(\lambda) < \lambda \ln(\lambda) - \lambda + 1$ for $\lambda \geq \lambda_0$ and $\lim_{\lambda \to \infty} b(\lambda) = \infty$, one obtains

$$\begin{aligned} \int_{\lambda_0}^\infty \min\{\tilde{\gamma}^{-1}(b(\lambda)) b'(\lambda), -\tilde{\gamma}^{-1}(a(\lambda)) a'(\lambda)\} d\lambda &= (3n^2) \int_{\lambda_0}^\infty \min\{\frac{b'(\lambda)}{b(\lambda)}, \frac{a'(\lambda)}{a(\lambda)}\} d\lambda \\ &= (3n^2) \int_{\lambda_0}^\infty \frac{b'(\lambda)}{b(\lambda)} d\lambda = \infty. \end{aligned}$$

Hence, according to Corollary 3.1 the Generalized Bayes estimator of $h(\theta)$ with respect to $\pi(\theta) = 1/\theta$ is Γ_m-admissible under the loss (1.1).

Remark 3.4. Typically all the result in this paper go through with some modifications for the density

$$f_X(x, \theta) = \begin{cases} q(\theta) r(x) & \theta < x < \bar{\theta}, \\ 0 & \text{otherwise,} \end{cases}$$

where $\theta \in \Theta$ is unknown.

4. AN APPLICATION

In the presence of vague prior information frequently the Γ-minimax approach is used as underlying principle. In this section, we provide the definition of the Γ-minimaxity of an estimator and then express the relation between this concept and the Γ-admissibility. Finally, we give an example.

Definition 4.1. A Γ-minimax estimator is a minimax strategy of the second player in the statistical game (Γ, Δ, r); δ^* is called a Γ-minimax estimator, if

$$\sup_{\pi\in\Gamma} r(\pi, \delta^*) = \inf_{\delta\in\Delta} \sup_{\pi\in\Gamma} r(\pi, \delta),$$

where $r(\pi, \delta)$ is the Bayes risk of δ.

Definition 4.2. A Γ-minimax estimator δ^* is said to be unique, if

$$r(\pi, \delta) = r(\pi, \delta^*), \quad \pi \in \Gamma,$$

for any other Γ-minimax estimator δ.

Remark 4.1.

- From Definition 4.2, it is obvious that a unique Γ-minimax estimator is Γ-admissible.
- If a Γ-admissible estimator δ is an equalizer on Γ, i.e., $r(., \delta)$ is constant on Γ, then δ is a unique Γ-minimax estimator.

Example 4.1. In Example 3.1, we have $E_\theta[X] = \theta - 1$ and $E_\theta[X^2] = \theta^2 - 2\theta + 2$. Thus, from (1.1), the risk function of δ_π is equal to

$$\begin{aligned} R(e^{X+1}, e^\theta) &= E_\theta[\{\ln(e^{X+1}) - \ln(e^\theta)\}^2] \\ &= E_\theta[\{X + 1 - \theta\}^2] \\ &= Var_\theta[X] \\ &= 1. \end{aligned}$$

So, δ_π is an equalizer on Γ_0, since its risk function is constant. Hence, $\delta_\pi(X) = e^{X+1}$ is the unique Γ_0-minimax estimator for e^θ.

REFERENCES

[1] Brown, L.D. (1968). Inadmissibility of the usual estimators of scale parameters in problems with unknown location and scale parameters, *Annals of Mathematical Statistics*, **23**, 29–48.

[2] Eichenauer-Herrmann, J. (1992). Gamma-admissibility of estimators in the one-parameter exponential family, *Metrika*, **39**, 199–208.

[3] Ghosh, G. and Meeden, M. (1977). Admissibility of linear estimators in the one parameter exponential family, *Annals of Mathematical Statistics*, **5**, 772–778.

[4] Hoffmann, K. (1985). Admissibility and inadmissibility of estimators in the one-parameter exponential family, *Statistics*, **16**, 327–349.

[5] Karlin, S. (1958). Admissibility for estimation with quadratic loss, *Annals of Mathematical Statistics*, **29**, 406–436.

[6] Kiapour, A. and Nematollahi, N. (2011). Robust Baysian prediction and estimation under squared log error loss function, *Statistics Probability Letters*, **81**, 1717–1724.

[7] Kim, B.H. (1994). Admissibility of generalized Bayes estimators in an one parameter non-regular family, *Metrika*, **41**, 99–108.

[8] Kim, B.H. and Meeden, G. (1994). Admissible estimation in an one parameter nonregular family of absolutely continuous distributions, *Communications in Statistics-Theory and Methods*, **23**, 2993–3001.

[9] Mahmoudi, E. and Zakerzadeh, H. (2011). An admissible estimator of a lower-bound scale parameter under squared-log error loss function, *Kybernetika*, **47**, 595–611.

[10] Nematollahi, N. and Jafari Tabrizi, N. (2012). Minimax estimator of a lower bounded parameter of a discrete distribution under a squared log error loss function, *Journal of Sciences, Islamic Republic of Iran*, **23**, 77–84.

[11] Pulskamp, R.J. and Ralescu, D.A. (1991). A general class of nonlinear admissible estimators in the one-parameter exponential case, *Journal of Statistical Planning and Inference*, **28**, 383–390.

[12] Ralescu, D. and Ralescu, S. (1981). A class of nonlinear admissible estimators in the one-parameter exponential family, *Annals of Mathematical Statistics*, **9**, 177–183.

[13] Sanjari Farsipour, N. (2003). Admissibility of estimators in the non-regular family under entropy loss function, *Statistical Papers*, **44**, 249–256.

[14] Sanjari Farsipour, N. (2007). Admissible estimation in an one parameter nonregular family of absolutely continuous distributions, *Statistical Papers*, **48**, 337–345.

[15] Sanjari Farsipour, N. and Zakerzadeh, H. (2005). Estimation of a gamma scale parameter under asymmetric squared log error loss function, *Communications in Statistics-Theory and Methods*, **24**, 1127–1135.

[16] Sanjari Farsipour, N. and Zakerzadeh, H. (2006). Estimation of generalized variance under an asymmetric loss function "Squared log error", *Communications in Statistics-Theory and Methods*, **35**, 571–581.

[17] Sinha, B.K. and Gupta, A.D. (1984). Admissibility of generalized Bayes and Pitman estimates in the nonregular family, *Communications in Statistics-Theory and Methods*, **13**, 1709–1721.

[18] Tanaka, H. (2010). Sufficient conditions for the admissibility under LINEX loss function in regular case, *Communications in Statistics-Theory and Methods*, **39**, 1477–1489.

[19] Tanaka, H. (2011). Sufficient conditions for the admissibility under LINEX loss function in non-regular case, *Statistics*, **45**, 199–208.

[20] Tanaka, H. (2012). Admissibility under the LINEX loss function in non-regular case, *Scientiae Mathematicae Japonicae*, **75**, 351–358.

[21] Zakerzadeh, H. and Moradi Zahraie, SH. (2015). Admissibility in non-regular family under squared-log error loss, *Metrika*, **78**, 227–236.

REVSTAT – Statistical Journal
Volume 15, Number 4, October 2017, 487–500

ESTIMATION THROUGH ARRAY-BASED GROUP TESTS

Authors: João Paulo Martins
– ESTG – Polytechnic Institute of Leiria, Portugal,
CEAUL – University of Lisbon, Portugal
jpmartins@ipleiria.pt

Miguel Felgueiras
– ESTG and CIGS – Polytechnic Institute of Leiria, Portugal,
CEAUL – University of Lisbon, Portugal
mfelg@ipleiria.pt

Rui Santos
– ESTG – Polytechnic Institute of Leiria, Portugal,
CEAUL – University of Lisbon, Portugal
rui.santos@ipleiria.pt

Received: September 2015 Revised: December 2015 Accepted: December 2015

Abstract:

- Pooling individual samples for batch testing is a common procedure for reducing costs. The recent use of multidimensional array algorithms, due to the emergence of robotic pooling, is an innovative way of pooling. We show that the two-dimensional array-based group tests can provide accurate estimates for the prevalence rate even for situations in which the traditional estimators, applied to one-dimensional arrays, are not valid. Hence, a computational script was developed to determine which prevalence rate estimate minimizes the sum of the squared deviations between the number of observed and expected rows and columns whose pooled sample had a positive test result.

Key-Words:

- *estimation; prevalence rate; pooled samples; array.*

AMS Subject Classification:

- 62F10, 62P10.

1. INTRODUCTION

Evidences of using pooled samples for batch testing date as far back as 1915 (cf. [8]). However, its use in order to reduce costs started only in 1943 with Dorfman's seminal work [4] on the detection of the syphilis antigen in U.S. soldiers during World War II. Dorfman entailed pooling together biological specimens from different individuals and testing the resulting pools of specimens rather than testing each individual.

When the aim is the detection of some binary characteristic, Dorfman's process comprehends two stages. In its first stage a pool of n individuals is homogeneously mixed and a portion of the mixed sample is analyzed. A negative result in the pooled mixture indicates that none of the n individuals has that characteristic. On the other hand, a positive result implies that at least one of the n individuals possesses the characteristic under investigation. And, in this last case, a second stage takes place in which an individual test is performed to each one of the n suspected individuals. The optimal batch size n^* is the pool size which minimizes the expected number of tests since the cost of mixing samples is, in general, negligible comparing to the cost of the experimental tests, as [13] points out.

Since Dorfman's work, the research on methodologies involving pooled sample tests has been quite active. Thence, some improvements to his work have been proposed, for instance, by [6, 22, 23]. The common idea of all these algorithm improvements was to divide each positive pool into smaller subpools until eventually all specimens are individually tested. These kind of algorithms are called hierarchical algorithms with a number of stages equal to the number of times each individual may be tested. All of them are called one-dimensional since they use non-overlapping pools. More recent works, considering the experimental errors measured by the test sensitivity and test specificity, are available in [9, 11, 19, 24]. Another branch in this area has been the application to quantitative characteristics (only reliable for underlying heavy tailed distributions). More details may be found, for instance, in [5, 15, 20, 21]. Moreover, the use of pooled samples does not refer only to the classification problem (identifying all the individuals which possess some characteristic), since it may also be useful in estimating its prevalence rate p (estimation problem), as [22] stated.

As different procedures may be applied in a problem involving pooled sample tests, the expected number of experimental tests per individual is a good measure of the savings obtained with each procedure. Hence, the *relative cost* of a procedure $\mathcal{M}$ is defined as

$$\mathrm{RC}\,(\mathcal{M}) = \frac{\mathrm{E}\,(T_N)}{N}, \tag{1.1}$$

where T_N stands for the number of experimental tests performed to screen a sample with N individuals.

When the purpose is to estimate the prevalence rate, the performing of individual tests is only optional, since the goal is no longer to identify who has the characteristic under investigation. Thus, this may lead to a lower relative cost. Furthermore, the estimators obtained by applying compound tests have, under certain conditions, better performance than the traditional estimators based on individual tests, cf. [7, 14, 22]. Hence, group testing can be more efficient as well as more accurate than individual testing.

The use of more complex schemes of mixing samples, i.e., dividing the amount of sample in two or more parts and using them in different batches had not been a reasonable choice since the complexity of the process could be itself another significant source of error. However, in the beginning of this century the emergence of the robotic and automatic pooling has turned the array-based group testing into a reliable alternative to hierarchical group testing (cf. [12]).

Our purpose is to show that in the context of a prevalence rate estimation the use of two dimensional arrays may be a reasonable alternative to the traditional one dimensional array-based procedure. We also discuss strategies to obtain an estimate from each of the possible ambiguous results of a two dimensional array. A first attempt to solve those ambiguities was performed in [17]. For this goal, an improvement of an algorithm firstly proposed in [16] is provided and a small simulation study is carried out. It is also claimed that for some low sensitivity tests, such as some enzyme immunoassays to screen for *Clostridium difficile* in fecal specimens described in [1], the performance of one dimensional arrays procedures may provide prevalence rate estimates outside the interval $[0, 1]$ whereas the two dimensional arrays procedures always provide valid estimates.

The outline of this article is as follows. Section 2 describes some general assumptions and additional notation usually used in this research field. Subsequently, Section 3 describes the use of one-dimensional and two-dimensional array-based group testing in the context of the estimation problem. The core of this work is Section 4 where it is analyzed the performance of the proposed algorithm for computing an estimate to the prevalence rate based on the results of two dimensional arrays. In particular, some possible drawbacks of the algorithm are discussed. Finally, some conclusions are provided in Section 5.

2. FRAMEWORK SETTING

Let us consider a large population of individuals and let p stand for the probability of randomly choosing an individual infected with some disease. The value p is called the *prevalence rate* of the disease.

When dealing with an estimation problem, i.e., the problem of estimating the value of p, the most basic and common pooled sample methodology is to divide the individuals among the sample into groups with size n – one dimensional arrays. For simplicity, admit that n is a divisor of the sample size N (otherwise, one would have $\left[\frac{N}{n}\right]$ groups with n individuals and one group with $N - \left[\frac{N}{n}\right] \times n$ individuals, where $[x]$ stands for the highest integer lower than x). Then, it is required to perform $T_N = \frac{N}{n}$ tests. Let us also assume that the individual status (infected/not infected) within a pooled sample are independent. The probability of having an infected pooled sample is $\pi_n = 1 - (1-p)^n$. Hence, the total number of infected pooled samples is described by a binomial random variable $I \frown \mathrm{Bin}\,(T_N, \pi_n)$, where T_N is the trials number and π_n the success probability. Thus, the maximum likelihood (ML) estimator of π_n is given by

$$\widehat{\pi_n} = \frac{I}{T_N}. \tag{2.1}$$

As p is given by a simple transformation of π_n, it is straightforward to show, applying the proprieties of the ML estimators, that the ML estimator of p is given by

$$\widehat{p} = 1 - \left(1 - \frac{I}{T_N}\right)^{1/n}. \tag{2.2}$$

For $n = 1$, $\hat{p} = 1 - \left(1 - \frac{I}{T_N}\right) = \frac{I}{T_N}$ is an unbiased estimator of p. For $n > 1$, the estimator is positively biased . Expressions for the expected value and variance of the estimator can be found in [10].

As screening errors may occur, the above binomial model is, in practice, unrealistic. Let $X_{ji} = 1$ denote an infected individual and $X_{ji} = 0$ denote a non-infected individual concerning the i-th individual of the j-th pooled sample where $i = 1, ..., n$ and $j = 1, ..., T_N$. In addition, X_{ji}^{+} denotes a positive test result and X_{ji}^{-} a negative test result performed with a sample collected only from that individual. The probability $\varphi_s = \mathrm{P}\left(X_{ji}^{+}|X_{ji} = 1\right)$ is called the *test sensitivity* and $\varphi_e = P\left(X_{ji}^{-}|X_{ji} = 0\right)$ is called the *test specificity*. [19] extended the concepts of specificity and sensitivity to a specific procedure $\mathcal{M}$. These measures assess the quality of a result provided by $\mathcal{M}$. The *procedure sensitivity* is the probability of an infected individual being correctly identified by the procedure $\mathcal{M}$, that is, $\varphi_s^{\mathcal{M}} = P_{\mathcal{M}}\left(X_{ji}^{+}|X_{ji} = 1\right)$. The *procedure specificity* stands for the probability of a non-infected individual being correctly classified by the procedure $\mathcal{M}$, that is, $\varphi_e^{\mathcal{M}} = P_{\mathcal{M}}\left(X_{ji}^{-}|X_{ji} = 0\right)$.

As some interaction between the pooled specimens may occur, some general assumptions underlying our work must be settled (more details may be found in [18]).

- Assumption 1 – Any specimen X_{ji}, where $i = 1, ..., T_N$ and $i = 1, ..., n$, may be described by a Bernoulli random variable X_{ji} where $P\left(X_{ji}=1\right) = p$ (infected) and $P\left(X_{ji} = 0\right) = q = 1 - p$ (non-infected).
- Assumption 2 – The methodology sensitivity equals the test sensitivity, i.e., $\varphi_s^{\mathcal{M}} = \varphi_s$. Note that this may not always be true as individually defective specimens may generate a negative result when tested in a batch. This is called an antagonism effect.
- Assumption 3 – The methodology specificity equals the test specificity, i.e., $\varphi_e^{\mathcal{M}} = \varphi_e$. As in the last assumption, in some situations the interaction between non-infected specimens may produce a positive pooled sample test result. This phenomenon is called synergism.
- Assumption 4 – Given the true status of any pool, its test result is independent of the true status and test result of any other non-overlapping pool.

3. ARRAY-BASED GROUP TESTING

We will briefly describe how to deal with the estimation problem concerning two different procedures. In the first one, the individuals are divided into non-overlapping groups (one-dimensional arrays). In the second one, a one-stage two-dimensional array procedure, the individuals will be tested twice.

3.1. One-dimensional arrays

One-dimensional arrays are the most common arrays used for batching individuals into groups in order to perform pooled sample tests. Each individual is allocated to one and only one group and some amount of its sample is mixed with the same amount of sample from other individual(s). This procedure will be represented by $D(n)$.

Given the j-th pooled sample of size n, the probability of it being positively classified is given by

$$P\left(X_j^+ | \sum_{i=1}^n X_{ji} \geq 1\right)\left(1-(1-p)^n\right) + P\left(X_j^+ | \sum_{i=1}^n X_{ji} = 0\right)(1-p)^n \tag{3.1}$$
$$= \varphi_s + \left(1 - \varphi_s - \varphi_e\right)(1-p)^n \,.$$

Therefore, a ML estimator of p is

$$\widehat{p} = 1 - \left(\frac{\varphi_s - \frac{I}{T_N}}{\varphi_s + \varphi_e - 1} \right)^{1/n} . \tag{3.2}$$

The estimator only assumes a value in the interval $[0,1]$ if

$$1 - \varphi_e \leq \frac{I}{T_N} \leq \varphi_s . \tag{3.3}$$

Whenever this condition is not fulfilled, we are not able to provide a reasonable estimate. For instance, suppose $p = 0.01$ and that a $D(3)$ procedure is performed with 50 pools where the test verifies $\varphi_e = 0.98$ and $\varphi_s = 0.6$. There is a chance of about 15% of condition (3.3) not be fulfilled. The proportion of positive samples is not an option as it may be a very biased estimator of the prevalence rate.

3.2. Two-dimensional arrays

Two-dimensional arrays are an alternative to the one dimensional arrays which uses overlapping pools. This approach is frequently employed in genetics, cf. [11], but it is rarely applied in the infectious disease setting. In its simplest two-stage version (square array), denoted by $A\,(n)$, a sample of size n^2 is placed in a $n \times n$ matrix in the following way. Each individual is allocated to one and only one matrix position. Then, all the individuals within the same row are gathered for batch testing, and the same procedure is applied to all the individuals within the same column. Thus, the two-stage version involves $T_{n^2} \geq 2n$ tests as subsequent individual tests can be performed to the ones lying in a row and/or column which tested positively. A variant of this methodology (a three-stage procedure) consists in performing *a priori* a pooled sample test on all the n^2 individuals (master pool). If the master pool test result is negative no further testing is needed as the individuals are all classified as negative.

The expected number of tests for all these two-dimensional array group testing procedures is derived in [18] when a classification problem is dealt. In [11] are computed the operating characteristics of these procedures (with or without a master pool). An extension to higher-dimensional arrays assuming no test errors may be found in [3]. More recently, [12] introduced the possibility of misclassification.

The performance of subsequent individual tests is required to avoid ambiguities. For instance, it is possible to have a row tested positive but all columns tested negative (the number of infected individuals can be any integer from zero to n, since certainly there is at least one misclassification error) or to have two positive rows and columns (the number of infected individuals can be two, three or four even considering that there was no misclassification). However, it won't be required to identify who is infected or either if a row or column has infected elements when the next proposed methodology for an estimation problem is applied. Hence, this one-stage procedure is expected to decrease the relative cost.

4. PREVALENCE RATE ESTIMATION

Two-dimensional array-based group testing allows the inclusion of each individual into two different batches. However, as previously discussed, some ambiguities may arise due to the experimental test errors described by the test sensitivity and by the test specificity.

In this case, the use of the proportion of defective individuals (without performing any individual tests) is not advised as it may lead to an underestimation of the prevalence rate or to an increase of the relative cost as [16] points out. [16] also provides some guidelines of how to compute a ML estimate of the prevalence rate using a computational script. Moreover, it computes an exact expression for the ML estimator, assuming possible test errors, for an one stage and two-dimensional array procedure $A\,(2)$.

The inputs of that script are the test sensitivity φ_s, the test specificity φ_e and the number of arrays that have $i-1$ positive rows and $j-1$ positive columns for $i=1,2,...,r+1$ and $j=1,2,...,c+1$. These values may be recorded in a $(r+1)\times(c+1)$ matrix $\boldsymbol{O}$ which resumes the experimental results required to compute a ML estimate.

To compute the ML function value at p_0 it is also required to compute the probability of observing $i-1$ positive rows and $j-1$ positive columns, where $i=1,2,...,r+1$ and $j=1,2,...,c+1$, given p_0. Let us denote this matrix of probabilities by $\boldsymbol{P}_{p_0}$. An approximation to these probabilities can be computed by the performance of a simple simulation.

The ML function value at p_0 knowing $\boldsymbol{O}$ is given by

$$\mathrm{ML}(p_0|\boldsymbol{O}) = \prod_{i,j} \boldsymbol{P}_{p_0}(i,j)^{\boldsymbol{O}(i,j)}. \tag{4.1}$$

However, it is necessary to account some special cases which can lead to inaccurate estimates. For instance, if for some value p_0 and some values i and j, $\boldsymbol{P}_{p_0}(i,j)$ is high and $\boldsymbol{O}(i,j)$ is zero then the process could converge to a value near p_0 whereas its "likelihood" is low.

To avoid having to account for this problem, we propose using the sum of the square of the differences between the values of the matrix $\boldsymbol{O}$ and the expected values for $\boldsymbol{O}$, computed using the probabilities of the matrix $\boldsymbol{P}$, as a quality measure for comparing different estimates, i.e.,

$$Dif(p_0|\boldsymbol{O}) = \sum_{i,j} \left(\boldsymbol{O}(i,j) - s \times \boldsymbol{P}_{p_0}(i,j)\right)^2 , \tag{4.2}$$

where s is the total number of two-dimensional arrays (i.e., $s = \sum_{i,j} \boldsymbol{O}(i,j)$).

In the next subsections, we provide some guidelines of an algorithm to find the minimum of the function defined in (4.2) as well as a small simulation study.

4.1. A computational script

As only for a small number of rows and columns it is possible to compute an exact expression for the function Dif defined in (4.2), in general, it is just possible to find the minimum of that function using some computational method.

Next, we describe the script which was implemented and highlight the possible drawbacks due to working with very low values. Some of the issues are shared by the implementation of the well-known chi-square test of independence of two random variables.

Our script comprehends the following steps.

- Step 1 – Consider an increasing sequence of possible values for p, say $p_1 < p_2 < \cdots < p_m$.
  We used the golden section search optimization method, cf. [2], as it presents optimal properties in the numerical search of a maximum when no expression for the function of interest is available. This method uses $m = 4$ and two inner points given at each step by $p_2 = p_4 - \varphi * (p_4 - p_1)$ and $p_3 = p_1 + \varphi * (p_4 - p_1)$ where $\varphi \approx 0.61803$ is known as the golden ratio.
- Step 2 – Simulate a large number of individuals (say, equal to $r \times c \times rep$ with rep large) extracted from a population with a prevalence rate p_i for $i = 1, ..., m$ where r and c stand for the number of rows and columns of the array and rep is the number of arrays. These arrays will be used to obtain an estimate for the values of the matrix $\boldsymbol{P}$. We used $rep = 100$ and our simulations have shown that higher values for rep do not change significantly the final outcome of the simulation.
- Step 3 – For each p_i and for each replicate compute the (estimated)

probability of observing an array corresponding to each position of the matrix $\boldsymbol{O}$ (the matrix of the experimental results). Add that value to the position (i, j) of the matrix $\boldsymbol{P}$.
Note that regardless of the number of infected individuals in the array, it is always possible to span all possible number of positive rows and columns due to the presence of the test errors.

- Step 4 – Compute the Dif function given the matrix $\boldsymbol{O}$.
- Step 5 – Compare the Dif function at $p_1, ..., p_m$ and choose the two estimates which minimize the function, say p_{min} and p_{max} where $p_{min} < p_{max}$.
- Step 6 – Consider a new increasing sequence of possible values for p starting on p_{min} and finishing on p_{max}.
In our case, $p_2 = p_4 - \varphi * (p_{max} - p_{min})$ and $p_3 = p_1 + \varphi * (p_{max} - p_{min})$.
- Step 7 – Repeat the procedure from step two up to step six until the distance between the two estimates is lower than some tolerance tol.

4.2. Possible drawbacks

Expression (4.2) for the Dif function involves very small quantities which are not a problem for the most recent software. However, for avoiding the null estimate we advise the initial choice of p_1 to be equal or higher than tol.

One possible problematic situation that must be taken into account in order to avoid underestimation occurs when the expected number of arrays with i positive rows and j positive columns, $s \times \boldsymbol{P}(i, j)$, is higher than zero and less than one. Note that, theoretically, all values of the matrix $\boldsymbol{P}$ are higher than zero but as we are not computing the exact matrix $\boldsymbol{P}$, as there are some computational restraints, it is possible to have zero in some entry of $\boldsymbol{P}$.

In that case, we suggest adding some of those low probabilities till eventually the sum of the expected number of arrays be at least one.

Hence, suppose you inspect all the values of $\boldsymbol{P}(i, j)$ in some logic sequence and you find h values of $\boldsymbol{P}$, say $\boldsymbol{P}(r_1, c_1), ..., \boldsymbol{P}(r_h, c_h)$, for which the sum of the expected values verifies $s \sum_{i=1}^{h-1} \boldsymbol{P}(r_i, c_i) < 1$ and $s \sum_{i=1}^{h} \boldsymbol{P}(r_i, c_i) \geq 1$. Then, add all those h probability estimates $P^* = \sum_{i=1}^{h} \boldsymbol{P}(r_i, c_i)$ and do the same for the matrix $\boldsymbol{O}$, i.e., $O^* = \sum_{i=1}^{h} \boldsymbol{O}(r_i, c_i)$. Replace the position (r_h, c_h) in the matrix $\boldsymbol{P}$ and in the matrix $\boldsymbol{O}$ by P^* and O^*, respectively. All the positions $(r_1, c_1), ..., (r_{h-1}, c_{h-1})$ for both matrices should be replaced by zero. This is a process with some resemblances to the one applied to contingency tables in order to improve the chi-square test performance.

In that logic sequence of inspection of all positions it is possible to get a sequence of values of $\boldsymbol{P}$ for which the sum of the expected values does not achieve 1 due to end of the inspection process. In this case, we suggest a similar process, however, the sum P^* and O^* should be added to some value of $\boldsymbol{P}$ whose expected value is at least one. Our simulations showed that the choice of this value is not relevant.

4.3. A simulation study

A small simulation was performed to assess the algorithm performance. The chosen measure to assess the accuracy of the estimates was the absolute value of the bias.

The one-stage square array-based group procedure, $A\,(2n)$, was compared to the one-dimensional alternative $D\,(n)$. Note that both procedures present the same relative cost.

We considered four different experimental tests with the sensitivity and specificity described in Table 1.

Table 1: Test sensitivity and test specificity.

Test	φ_s	φ_e
A	0.99	0.99
B	0.80	0.98
C	0.60	0.98
D	0.99	0.80

It was considered four different prevalence rates: 0.01, 0.05, 0.10 and 0.25. For each one of them, 20 square arrays were simulated ($s = 20$) for applying $A(4)$ and $A(6)$ procedures. The one-dimensional alternative for each of these two procedures was $D\,(2)$, with 160 pools, for the first case and $D\,(3)$, with 240 pools, for the last one. In each pair of procedures (one and two-dimensional) the number of tests performed is the same. For all cases, the number of replicates was 100. The results are summarized in the Table 2.

In some cases, condition (3.3) was not fullfield leading to negative estimates. Thus, some values are not displayed (the symbol "—" is displayed instead of a numerical value) since it was observed more than 20 (in 100) negative estimates. All estimates displayed for a prevalence rate $p = 0.01$ do not use all 100 estimates since 1 to 6 of them were negative and excluded from the calculus of the mean of the absolute bias.

Table 2: Mean of the absolute bias (multiplied by 10^2) of 100 estimates obtained by $A\,(4)\,(D\,(2))|A\,(6)\,(D\,(3))$ procedures.

	$p = 0.01$	$p = 0.05$	$p = 0.10$	$p = 0.25$
A	0.48(0.34)|0.25(0.16)	0.81(0.33)|0.60(0.28)	0.96(0.49)|0.83(0.40)	1.82(0.94)|1.38(0.86)
B	0.51(—)|0.34(0.35)	0.79(0.68)|0.71(0.59)	1.12(1.03)|1.06(0.72)	2.17(1.82)|2.34(1.53)
C	0.48(—)|0.37(0.45)	0.95(1.18)|0.78(0.90)	1.24(1.41)|1.03(1.15)	2.25(2.45)|2.92(2.07)
D	0.31(0.38)|0.61(—)	1.15(1.55)|0.86(1.01)	1.46(1.61)|1.06(1.04)	1.47(1.54)|1.45(1.36)

Note that the mean of the absolute bias increases with the prevalence rate since for low prevalence rate values the value zero is a natural limit to the estimates. It is not surprising to observe $A(6)$ procedure performing better than $A(4)$ since it uses more individuals. However, when the prevalence rate increases the chance of having a very high number of positive tests greatly increases leading to a worse performance. Regardless the square array procedure, the one-dimensional pools generally outperform the square array procedures for moderate and high prevalence rates (note that the interval of possible prevalence rates is $]0, 0.50]$ since for values higher than 0.50 we can study $q = 1 - p$). For the most inaccurate test considered, test C, the behavior is similar.

5. CONCLUSION

The spreading of the possibility of robotic pooling will certainly highlight the use of arrays with dimensions higher than one as a practical alternative to the traditional one dimensional arrays for both estimation and classification problems, cf. [12].

In this work we address the problem of estimation and show that for very inaccurate tests (cheaper tests) the use of square arrays assures the experimenter a reasonable estimate (at least for low prevalence rates). However, whenever the sample size is low and φ_s and φ_e are high we have just a few square arrays and the results can be quite inaccurate. In this scenario the $D\,(n)$ methodology remains a more reasonable option as it could (almost certainly) provide a good estimate. Thence, $D\,(n)$ methodology can, in some situations, outperforms $A(2n)$ methodology whereas $A(2n)$ works well in a wider parameter support.

REFERENCES

[1] Alcalá, L.; Sánchez-Cambronero, L.; Catalán, M.P.; Sánchez-Somolinos, M.; Peláez, M.T.; Marín, M. and Bouza, E. (2008). Comparison of three commercial methods for rapid detection of *Clostridium difficile* toxins A and B from fecal specimens, *J. Clin. Microbiol.*, **46**, 3833–3835.

[2] Antia, H.M. (2002). *Numerical Methods for Scientists and Engineers*, Vol. 1, 2nd ed., Springer, New York, 356–362.

[3] Berger, T.; Mandell, J.W. and Subrahmanya, P. (2000). Maximally efficient two-stage screening, *Biometrics*, **56**, 833–840.

[4] Dorfman, R. (1943). The detection of defective members in large populations, *Ann. Math. Stat.*, **14**, 436–440.

[5] Felgueiras, M.; Martins, J.P. and Santos, R. (2014). Distributions families in counting bacteria for compound sampling, *Lecture Notes in Computer Science*, **8581**, 539–551.

[6] Finucan, H.M. (1964) The blood testing problem, *App. Stat.*, **13**, 43–50.

[7] Garner, F.C.; Stapanian, M.A.; Yfantis, E.A. and Williams, L.R. (1989). Probability estimation with sample compositing techniques, *J. Off. Stat.*, **5**, 365–374.

[8] Hughes-Oliver, J.M. (2006). *Pooling Experiments for Blood Screening and Drug Discovery.* In "Screening — Methods for Experimentation in Industry, Drug Discovery, and Genetics" (A. Dean and S. Lewis, Eds.), Springer, 46–68.

[9] Hwang, F.K. (1976). Group testing with a dilution effect, *Biometrika*, **63**, 671-673.

[10] Hung, M. and Swallow, W.H. (1999). Robustness of group testing in the estimation of proportions, *Biometrics*, **55**, 231–237.

[11] Kim, H.; Hudgens, M.; Dreyfuss, J.; Westreich, D. and Pilcher, D. (2007). Comparison of group testing algorithms for case identification in the presence of testing errors, *Biometrics*, **63**, 1152–1163.

[12] Kim, H. and Hudgens, M. (2009). Three-Dimensional Array-Based Group Testing Algorithms, *Biometrics*, **65**, 903–910.

[13] Liu, S.C.; Chiang, K.S.; Lin, C.H.; Chung, W.C.; Lin, S.H. and Yang, L.C. (2011). Cost analysis in choosing group size when group testing for Potato virus Y in the presence of classification errors, *Ann. Appl. Biol.*, **159**, 491–502.

[14] Loyer, M.W. (1983). Bad probability, good statistics, and group testing for binomial estimation, *Am. Stat.*, **37**, 57–59.

[15] Martins, J.P.; Santos, R. and Sousa, R. (2014). *Testing the maximum by the mean in quantitative group tests.* In "New Advances in Statistical Modeling and Applications, Studies in Theoretical and Applied Statistics" (A. Pacheco, R. Santos, M.R. Oliveira and C.D. Paulino, Eds.), Springer, 55–63.

[16] Martins, J.P.; Felgueiras, M. and Santos, R. (2014). Maximum Likelihood Estimation in Pooled Sample Tests, *AIP Conf. Proc.*, **1618**, 543 (online).

[17] Martins, J.P.; Santos, R. and Felgueiras, M. (2015). *A Maximum Likelihood Estimator for the Prevalence Rate Using Pooled Sample Tests*. In " Theory and Practice of Risk Assessment" (C.P. Kitsos, T.A. Oliveira, A. Rigas and S. Gulati, Eds.), Springer, 99–110.

[18] Phatarfod, C.D. and Sudbury, A. (1994). The use of a square array scheme in blood testing, *Stat. Med.*, **13**, 2337–2343.

[19] Santos, R.; Pestana, D. and Martins, J.P. (2013). *Extensions of the Dorfman's theory*. In "Studies in Theoretical and Applied Statistics, Recent Developments in Modeling and Applications in Statistics" (P.E. Oliveira, M. Graça, C. Henriques and M. Vichi, Eds.), Springer, 179–189.

[20] Santos, R.; Martins, J.P. and Felgueiras, M. (2015). Known mean, unknown maxima? Testing the maximum knowing only the mean, *Commun. Stat.-Simul. C.*, **44**, 2479–2491.

[21] Santos, R.; Martins, J.P. and Felgueiras, M. (2015). *An overview of quantitative continuous compound tests*. In "CIM Series in Mathematical Sciences, Dynamics, Games and Science III" (J.P. Bourguignon, R. Jeltsch, A. Pinto and M. Viana, Eds.), Springer, 595–607.

[22] Sobel, M. and Elashoff, R.M. (1975). Group testing with a new goal, estimation, *Biometrika*, **62**, 181–193.

[23] Sterret, A. (1957). On the detection of defective members of large populations, *Ann. Math. Stat.*, **28**, 1033–1036.

[24] Wein, L.M. and Zenios, S.A. (1996). Pooled testing for HIV screening: capturing the dilution effect, *Oper. Res.*, **44**, 543–569.

REVSTAT – Statistical Journal
Volume 15, Number 4, October 2017, 501–522

ESTIMATING RENYI ENTROPY OF SEVERAL EXPONENTIAL DISTRIBUTIONS UNDER AN ASYMMETRIC LOSS FUNCTION

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

Somesh Kumar
– Department of Mathematics, Indian Institute of Technology Kharagpur, Kharagpur-721302, India
smsh@iitkgp.ernet.ac.in

Received: December 2014 Revised: September 2015 Accepted: January 2016

Abstract:

- The present paper takes into account the estimation of the Renyi entropy of several exponential distributions under a linex loss function. The models under study are (i) several exponential distributions with a common scale parameter and unknown but unequal location parameters and (ii) several exponential distributions with a common location parameter and unknown but unequal scale parameters. Improvements over the best affine equivariant estimator are obtained for the first model considering unrestricted and restricted parameter spaces. For the second model, sufficient conditions for improvement over affine and scale equivariant estimators are obtained and consequently, improvements over the maximum likelihood estimator and the uniformly minimum variance unbiased estimator are proposed. Sections on numerical studies have been included after each model to present comparative study of the relative risk performance of the proposed improved estimators.

Key-Words:

- *Renyi entropy; linex loss function; equivariant estimator; inadmissibility; Brewster-Zidek technique.*

AMS Subject Classification:

- 62F10, 62C15.

1. INTRODUCTION

The Shannon entropy (see Shannon (1948)) is a fundamental measure of information content and has been applied in a wide variety of fields such as statistical thermodynamics, urban and regional planning, business, economics, finance, operations research, queueing theory, spectral analysis, image reconstruction, biology and manufacturing. One may refer to Kapur (1990) and Cover and Thomas (2006) for examples of various applications. Several generalized information-theoretic measures have been proposed in the literature to measure the uncertainty of a probability distribution since the seminal work of Shannon (1948). Among these, one of the most important and applicable measures is proposed by Renyi (1961). For a random variable X with probability density function $f(x|\theta), \theta \in \Theta$, the Renyi entropy is given by

$$R_\alpha(X) = \frac{1}{1-\alpha} \ln \int_{-\infty}^{\infty} f^\alpha(x|\theta)dx, \quad \alpha\ (\neq 1) > 0. \tag{1.1}$$

Note that we are using logarithm to base e in the expression given by (1.1). Here, the unit of the information measure is nat. Golshani and Pasha (2010) provide some important properties of the measure given in (1.1):

- **(i)** The Renyi entropy can be negative,
- **(ii)** It is invariant under location transformation, but not under scale transformation, and
- **(iii)** For any $\alpha_1 < \alpha_2$, $R_{\alpha_2}(X) \leq R_{\alpha_1}(X)$ and equality holds if and only if X is a uniform random variable.

Using L'Hospital rule, it can be shown that (1.1) retrieves the Shannon entropy when α tends to 1. The Renyi entropy is more or less sensitive to the shape of the probability distributions due to the parameter α. For large values of α, the measure given in (1.1) is more sensitive to events that occur often. Likewise, for small values of α, it is more sensitive to the events which happen seldom. In many instances, the Renyi entropy is seen to be more useful than the Shannon entropy (see Nilsson (2006), Maszczyk and Duch (2008) and Pharwaha and Khehra (2009)). The measure given in (1.1) has found a lot of applications in different areas of science and technology. For example, in speech recognition, different values of α determine different concepts of noisiness. Basically, small α values tend to emphasize the noise content of signal, while large α values tend to emphasize the harmonic content of a signal (see Obin and Liuni (2012)). It is also used for ultrasonic molecular imaging (see Hughes *et al.* (2009)). For properties of Renyi entropy one may refer to Song (2001), Bercher (2008), De-Gregorio and Iacus (2009), Golshani and Pasha (2010) and Renyi (2012).

Recently, the problem of estimating a common characteristic of several independent populations has received a considerable attention. There are many situations where this problem arises. For example, this situation arises when the information from several independent studies are combined or in meta-analysis. Meta-analysis is used in clinical studies. This is also seen in many statistical designs such as balanced incomplete block designs, panel models and regression models. The present paper is concerned with the problem of estimating the Renyi entropy of several exponential populations with respect to linex loss function (see Varian (1975)). The linex loss function is given as

$$(1.2)\qquad L(\triangle) = p'[\exp\{p\triangle\} - p\triangle - 1],\ \triangle = \delta - \theta,\ p \neq 0,\ p' > 0,$$

where p and p' are shape and scale parameters, respectively. Without loss of generality, we assume $p' = 1$. Note that the loss function (1.2) reduces to the squared error loss function when $|p|$ tends to 0. For more properties on linex loss function one may refer to Zellner (1986).

Let $\Pi_1, ..., \Pi_k$ be k (≥ 2) exponential populations with location and scale parameters $\underline{\mu} = (\mu_1, ..., \mu_k)$ and $\underline{\sigma} = (\sigma_1, ..., \sigma_k)$, respectively. The probability density function of the i-th population Π_i is given by

$$(1.3)\qquad f_i(x|\mu_i, \sigma_i) = \begin{cases} \sigma_i^{-1} exp\{-(x - \mu_i)/\sigma_i\}, & \text{if } x > \mu_i \\ 0, & \text{otherwise,} \end{cases}$$

where $\mu_i \in \mathbb{R},\ \sigma_i > 0$ and $i = 1, ..., k$. The expression of the Renyi entropy of k exponential distributions can be obtained as $R_\alpha(\underline{\sigma}) = \sum_{i=1}^{k} \ln \sigma_i - k \ln \alpha/(1 - \alpha)$. Several authors attempted the problem of estimating entropy of various continuous probability distributions. In this direction one may refer to Misra *et al.* (2005), Kayal and Kumar (2011a, 2011b, 2013), and Kayal *et al.* (2015). Misra *et al.* (2005) showed that the best affine equivariant estimator (BAEE) of the Shannon entropy of a multivariate normal distribution is inadmissible with respect to the squared error loss function. Under linex loss function, Kayal and Kumar (2011a) derived an estimator improving upon the BAEE of the Shannon entropy of a shifted exponential distribution. Kayal and Kumar (2011b) considered the problem of estimating the Shannon entropy of several exponential distributions with respect to both squared error and linex loss functions. Generalized Bayes estimators are showed to be admissible. Kayal and Kumar (2013) obtained improved estimator upon the BAEE in estimating the Shannon entropy of several exponential distributions with a common scale but unequal location parameters with respect to the squared error loss function. Recently, Kayal *et al.* (2015) studied the problem of estimating the Renyi entropy of several exponential distributions with a common location but unequal scale parameters with respect to squared error loss function. They derived the uniformly minimum variance unbiased estimator (UMVUE) and obtained improvements over the UMVUE and the maximum likelihood estimator (MLE). In this communication we deal with the problem of estimating the Renyi entropy in similar models considered by Kayal and Kumar (2013, 2015) with respect to the linex loss function.

The rest of the paper is organized as follows. In Section 2, the problem of estimating the Renyi entropy of several exponential distributions with a common scale and unknown but unequal location parameters is considered. The BAEE is shown to be inadmissible. Further, estimators improving upon the BAEE are obtained when the parameter space is restricted. Relative risks of the proposed estimators are compared numerically. In Section 3, the problem of estimating the Renyi entropy of several exponential distributions with a common location and unequal scale parameters is considered. Inadmissibility results for the scale and affine equivariant estimators are obtained. Further, improved estimators over the MLE and the UMVUE are derived. Some concluding remarks have been added in Section 4. Finally, relative risk performance of the proposed estimators is compared numerically.

2. COMMON SCALE BUT UNEQUAL LOCATION PARAMETERS

As mentioned earlier, in this section, we consider k independent exponential populations with unknown and possibly unequal location parameters $\underline{\mu}$ and a common but unknown scale parameter σ. This model arises in reliability engineering where location parameters can be interpreted as minimum guarantee times of several equipments, whereas the common scale parameter can be considered as unknown but possibly equal failure rate of those equipments. This model is also useful in economy where one may assume the unknown location parameters as the income levels below which the tax filling is not required in different locations. However, the average income levels may be same due to overall economic policies of the country. Let $(X_{i1}, ..., X_{in_i})$ be a random sample of size n_i drawn from the i-th $(i = 1, ..., k)$ population with the probability density function

$$f_i(x|\mu_i, \sigma) = \begin{cases} \sigma^{-1}\ exp\{-(x-\mu_i)/\sigma\}, & \text{if } x > \mu_i \\ 0, & \text{otherwise,} \end{cases} \tag{2.1}$$

where $\mu_i \in \mathbb{R}$ and $\sigma > 0$. For a population with probability density function (2.1), the Renyi entropy can be obtained as $R_\alpha(\sigma) = k \ln \sigma - k \ln \alpha/(1-\alpha)$. It should be mentioned that the problem of estimating $R_\alpha(\sigma)$ with respect to the loss function of the form $L(\theta, \delta) = W(\delta - \theta)$ is equivalent to that of estimating $Q_1(\sigma) = \ln \sigma$. Here, the loss function is given by

$$L^1(Q_1(\sigma), \delta) = exp\{p(\delta - \ln \sigma)\} - p(\delta - \ln \sigma) - 1,\ p \neq 0. \tag{2.2}$$

Note that for the i-th population, $(X_{i(1)}, Y_i)$ is a complete and sufficient statistic of (μ_i, σ), where $X_{i(1)} = min_{1\leq j\leq n_i} X_{ij}$ and $Y_i = \sum_{j=1}^{n_i} X_{ij}$. We denote $\underline{X}_{(1)} = (X_{1(1)}, ..., X_{k(1)})$, $T = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij} - X_{i(1)})$ and $n = \sum_{i=1}^{k} n_i$. Following the factorization criterion (see Lehmann and Casella, 1998, pp. 35), it can be showed that $(\underline{X}_{(1)}, T)$ is a complete and sufficient statistic of $(\underline{\mu}, \sigma)$. Further, $\underline{X}_{(1)}$ and T are independently distributed. It is seen that $X_{i(1)}$ follows an exponential distribution with location parameter μ_i and scale parameter σ/n_i, and $2T/\sigma$ follows a chi-square distribution with $2(n-k)$ degrees of freedom. The MLE and the UMVUE of $Q_1(\sigma)$ are given by $\delta^1_{ML} = \ln T - \ln n$ and $\delta^1_{MV} = \ln T - \psi(n-k)$, respectively, where ψ denotes digamma function and is given by $\psi(x) = \frac{d}{dx}(\ln \Gamma(x))$.

2.1. The best affine equivariant estimator and its improvement

In this section, we introduce invariance to the problem under study and obtain an improvement over the BAEE. Let $G_{a,b_i} = \{g_{a,b_i}(x_{ij}) : g_{a,b_i}(x_{ij}) = ax_{ij} + b_i, a > 0, b_i \in \mathbb{R}\},\ j = 1, ..., n_i,\ i = 1, ..., k$ be an affine group of transformations. Under the transformation $g_{a,b_i}(x_{ij}) = ax_{ij} + b_i$, the form of an affine equivariant estimator can be obtained as

$$\delta^1_c(\underline{X}_{(1)}, T) = \ln T - c, \tag{2.3}$$

where c is an arbitrary constant. In the following theorem we obtain the BAEE of $Q_1(\sigma)$. The proof is omitted as it is straightforward.

Theorem 2.1. *Under the linex loss function* (2.2), *the BAEE of $Q_1(\sigma)$ is $\delta^1_{c_0}(\underline{X}_{(1)}, T)$, where $c_0 = -(1/p)\ln[\Gamma(n-k)/\Gamma(n-k+p)]$.*

We consider a group of scale transformations $G_a = \{g_a(x_{ij}) = ax_{ij}, a > 0\},\ j = 1, ..., n_i,\ i = 1, ..., k$. Under the transformation $g_a(x_{ij}) = ax_{ij}$, the form of a scale equivariant estimator is

$$\delta_\phi(\underline{W}, T) = \ln T + \phi(\underline{W}), \tag{2.4}$$

where $\underline{W} = (W_1, ..., W_k),\ W_i = X_{i(1)}/T$ and ϕ is a real valued measurable function. In the following theorem, we prove a general inadmissibility result for the estimators of the form (2.4). First, define

$$\phi_0(\underline{w}) = \begin{cases} \ln u - \frac{1}{p}\ln\left(\frac{\Gamma(n+p)}{\Gamma(n)}\right), & \text{if } \underline{w} \in (B_1 \bigcap B_2) \bigcup (B_3 \bigcap B_2^c) \\ \phi(\underline{w}), & \text{otherwise,} \end{cases} \tag{2.5}$$

where $B_1 = \{\underline{w} : w_{(1)} > 0\},\ B_2 = \{\underline{w} : u < exp(\phi(\underline{w}) + (1/p)\ln(\Gamma(n+p)/\Gamma(n)))\}$, $B_3 = \{\underline{w} : w_{(k)} < 0\}$, $u = \sum_{i=1}^k n_i w_i + 1$, $w_{(1)} = min\{w_1, ..., w_k\}$, $w_{(k)} = max\{w_1, ..., w_k\}$ and $w_i = x_{i(1)}/t,\ i = 1, ..., k$.

Theorem 2.2. *Let δ_ϕ be a scale equivariant estimator of the form* (2.4) *and $\phi_0(\underline{w})$ be as defined in* (2.5). *If there exists $(\underline{\mu}, \sigma)$ such that $P_{(\underline{\mu},\sigma)}(\phi_0(\underline{W}) \neq \phi(\underline{W})) > 0$, then under linex loss function* (2.2), *the estimator δ_{ϕ_0} dominates δ_ϕ.*

Proof: The risk function of the estimators of the form (2.4) is

$$R(\underline{\mu}, \sigma, \delta_\phi) = E^{\underline{W}} R_1(\underline{\mu}, \sigma, \underline{W}, \delta_\phi),$$

where R_1 denotes the conditional risk of δ_ϕ given $\underline{W} = \underline{w}$, and is given by

$$\begin{aligned} R_1(\underline{\mu}, \sigma, \underline{w}, \delta_\phi) \;=\; & E[(exp\{p(\ln T + \phi(\underline{W}) - \ln \sigma)\} \\ & -p(\ln T + \phi(\underline{W}) - \ln \sigma) - 1)|\underline{W} = \underline{w}]. \end{aligned} \tag{2.6}$$

Note that the conditional risk function $R_1(\underline{\mu}, \sigma, \underline{w}, \delta_\phi)$ given in (2.6) is a function of the ratio $\underline{\mu}/\sigma$. Hence, without loss of generality we may assume σ to be 1. Moreover, the conditional risk is a convex function of ϕ, therefore the choice of ϕ minimizing $R_1(\underline{\mu}, \sigma, \underline{w}, \delta_\phi)$ can be obtained as

$$\hat{\phi}(\underline{\mu}, \underline{w}) = -p^{-1} \ln(E(T^p|\underline{W} = \underline{w})). \tag{2.7}$$

To get improvement over δ_ϕ, it is required to obtain the supremum and infimum of $\hat{\phi}(\underline{\mu}, \underline{w})$ given in (2.7). These can be derived along the arguments of the proof of Theorem 2 of Kayal and Kumar (2013). We omit the details here.

Case **(i):** When all μ_i's, $(i = 1, ..., k)$ are non-negative, the respective supremum and infimum of $\hat{\phi}(\underline{\mu}, \underline{w})$ can be obtained as

$$\sup_{\underline{\mu}} \hat{\phi}(\underline{\mu}, \underline{w}) = \ln u - p^{-1} \ln(\Gamma(n+p)/\Gamma(n)) \text{ and } \inf_{\underline{\mu}} \hat{\phi}(\underline{\mu}, \underline{w}) = -\infty.$$

Case **(ii):** Assume that μ_i's are negative for $i = 1, ..., k$. Under this restriction, it is required to take into account three possibilities on w_i's: (**a**) all w_i's are non-negative, (**b**) all w_i's are negative and (**c**) some of w_i's, $(i = 1, ..., k)$ are non-negative and remaining are negative. In the following we consider these three sub-cases separately and obtain supremum and infimum of $\hat{\phi}(\underline{\mu}, \underline{w})$.

(**a**) Under the assumption that w_i's are non-negative, we obtain

$$\hat{\phi}(\underline{\mu}, \underline{w}) = \ln u - p^{-1} \ln(\Gamma(n+p)/\Gamma(n)).$$

(**b**) When w_i's are negative, note that the value of u may be positive or negative. For $u > 0$,

$$\sup_{\underline{\mu}} \hat{\phi}(\underline{\mu}, \underline{w}) = +\infty \text{ and } \inf_{\underline{\mu}} \hat{\phi}(\underline{\mu}, \underline{w}) = \ln u - p^{-1} \ln(\Gamma(n+p)/\Gamma(n));$$

and for $u < 0$,

$$\sup_{\underline{\mu}} \hat{\phi}(\underline{\mu}, \underline{w}) = +\infty \text{ and } \inf_{\underline{\mu}} \hat{\phi}(\underline{\mu}, \underline{w}) = -\infty.$$

(**c**) Let some of w_i's $(i = 1, ..., k)$ assume non-negative values and the remaining w_i's assume negative values. Thus u may be positive or negative. When $u > 0$, then

$$\sup_{\underline{\mu}} \hat{\phi}(\underline{\mu}, \underline{w}) = +\infty \text{ and } \inf_{\underline{\mu}} \hat{\phi}(\underline{\mu}, \underline{w}) = \ln u - p^{-1} \ln(\Gamma(n+p)/\Gamma(n));$$

and when $u < 0$, then

$$\sup_{\underline{\mu}} \hat{\phi}(\underline{\mu}, \underline{w}) = +\infty \text{ and } \inf_{\underline{\mu}} \hat{\phi}(\underline{\mu}, \underline{w}) = -\infty.$$

Case **(iii):** Under the constraints that some of μ_i's are non-negative and remaining are negative, we consider the following sub-cases:

(**a**) For the case when $w_1, ..., w_r, ..., w_k > 0$, we obtain

$$\sup_{\underline{\mu}} \hat{\phi}(\underline{\mu}, \underline{w}) = \ln u - p^{-1}\ln(\Gamma(n+p)/\Gamma(n)) \text{ and } \inf_{\underline{\mu}} \hat{\phi}(\underline{\mu}, \underline{w}) = -\infty.$$

(**b**) Assume that $w_1, ..., w_r > 0$ and $w_{r+1}, ..., w_k < 0$. Then, for $u \neq 0$,

$$\sup_{\underline{\mu}} \hat{\phi}(\underline{\mu}, \underline{w}) = +\infty \text{ and } \inf_{\underline{\mu}} \hat{\phi}(\underline{\mu}, \underline{w}) = -\infty.$$

(**c**) Let $w_1, ..., w_r > 0$ and within $(k-r)$, some w_i's be non-negative and remaining be negative. Then, for $u \neq 0$, we obtain

$$\sup_{\underline{\mu}} \hat{\phi}(\underline{\mu}, \underline{w}) = +\infty \text{ and } \inf_{\underline{\mu}} \hat{\phi}(\underline{\mu}, \underline{w}) = -\infty.$$

An application of the Brewster and Zidek technique (see Brewster and Zidek (1974)) on the function $R_1(\underline{\mu}, \sigma, \underline{w}, \delta_\phi)$, then completes the proof of the theorem. □

As a consequence of the Theorem 2.2, we get the following corollary which shows that the BAEE obtained in Theorem 2.1 is inadmissible.

Corollary 2.1. The BAEE $\delta^1_{c_0}$ of $Q_1(\sigma)$ is dominated by the estimator

$$\delta^1_{IB} = \begin{cases} \ln(uT) - p^{-1}\ln(\Gamma(n+p)/\Gamma(n)), & \text{if } \underline{w} \in (B_1 \bigcap C_1) \bigcup (B_3 \bigcap C_1^c) \\ \ln T + p^{-1}\ln(\Gamma(n-k)/\Gamma(n-k+p)), & \text{otherwise,} \end{cases}$$

where $C_1 = \{\underline{w} : u < exp(d)\}$ and $d = p^{-1}\ln(\Gamma(n-k)\Gamma(n+p)/\Gamma(n-k+p)\Gamma(n))$.

In this part of the paper we consider the problem of estimating $Q_1(\sigma)$ in restricted parameter spaces. Here we consider the restriction on μ_i's. However, it is seen that it affects the improvement results for the estimation of $Q(\sigma)$. First assume that all μ_i's are bounded below. This arises when the minimum guarantee time of components is known to be more than a pre-specified constant. Without loss of generality, we may assume that $\mu_{(1)} \geq 0$, where $\mu_{(1)} = min\{\mu_1, ..., \mu_k\}$. In this case, δ^1_{ML} is the MLE of $Q_1(\sigma)$. Along the arguments of Case (**i**) of the proof of the Theorem 2.2, the inadmissibility of the BAEE can be established and the improved estimator is

$$\delta_{IB}^{1+} = \begin{cases} \ln(uT) - p^{-1}\ln(\Gamma(n+p)/\Gamma(n)), & \text{if } \underline{w} \in C_1, \\ \ln T + p^{-1}\ln(\Gamma(n-k)/\Gamma(n-k+p)), & \text{otherwise.} \end{cases}$$

We also consider the other case when the guarantee times of the components are known to be bounded above. Without loss of generality we assume $\mu_{(k)} < 0$, where $\mu_{(k)} = max\{\mu_1, ..., \mu_k\}$. In this case, the MLE of $Q_1(\sigma)$ is $\delta_{RM}^1 = \ln T^0 - \ln n$, where $T^0 = \sum_{i=1}^k (Y_i - n_i X_{i(1)}^0)$, $X_{i(1)}^0 = min\{0, X_{i(1)}\}$, $i = 1, ..., k$. Along the arguments of Case (**ii**) of the Theorem 2.2, the inadmissibility of the BAEE can be established. The improved estimator is given by

$$\delta_{IB}^{1-} = \begin{cases} \ln(uT) - \frac{1}{p}\ln(\frac{\Gamma(n+p)}{\Gamma(n)}), \text{ if } \underline{w} \in B_1 \bigcup (B_3 \bigcap C_1^c) \bigcup (C_2 \bigcap C_3 \bigcap C_1^c) \\ \ln T + \frac{1}{p}\ln(\frac{\Gamma(n-p)}{\Gamma(n-k+p)}), \text{ otherwise,} \end{cases}$$

where $C_2 = \{\underline{w} : w_{(r)} < 0\}$, $C_3 = \{\underline{w} : w_{(r+1)} > 0\}$.

2.2. Numerical comparisons

In this section, we present the relative risk performance of δ_{IB}^1, δ_{IB}^{1+} and δ_{IB}^{1-} over the BAEE $\delta_{c_0}^1$ through graphs for the case $k = 2$. We assume $\sigma = 1$, as the conditional risk in (2.6) is a function of $(\frac{\mu_1}{\sigma}, \frac{\mu_2}{\sigma})$. It should be mentioned that the risk values of various estimators were calculated using Monte-Carlo simulation based on $10,000$ samples of different combinations of (n_1, n_2) and different values of (μ_1, μ_2) and p. However, we present few of them considering sample sizes $(5, 5)$, $(5, 10)$, $(10, 5)$, $(10, 10)$ and $p = +0.2, -0.2$. It is worthwhile to remark that we observe similar pattern of the relative risk for other values of p and (n_1, n_2).

Based on the Fig. 1 we can conclude the following points:

(i) The margin and the region of the relative risk improvement (RRI) of δ_{IB}^1 over $\delta_{c_0}^1$ becomes small when we increase sample sizes (n_1, n_2).

(ii) We observe considerable RRI of δ_{IB}^1 over $\delta_{c_0}^1$ when both μ_1 and μ_2 approach to origin.

(iii) For fixed (n_1, n_2), the RRI of δ_{IB}^1 over $\delta_{c_0}^1$ is marginally better for negative values of p than positive values of p. For example, the RRI of δ_{IB}^1 over $\delta_{c_0}^1$ is 8.59% at $(\mu_1 = 0.18, \mu_2 = 0)$ for $(n_1 = 5, n_2 = 5)$ and $p = -2$, whereas for the same values of (μ_1, μ_2) and (n_1, n_2), the RRI of δ_{IB}^1 over $\delta_{c_0}^1$ is 8.24% when $p = 2$.

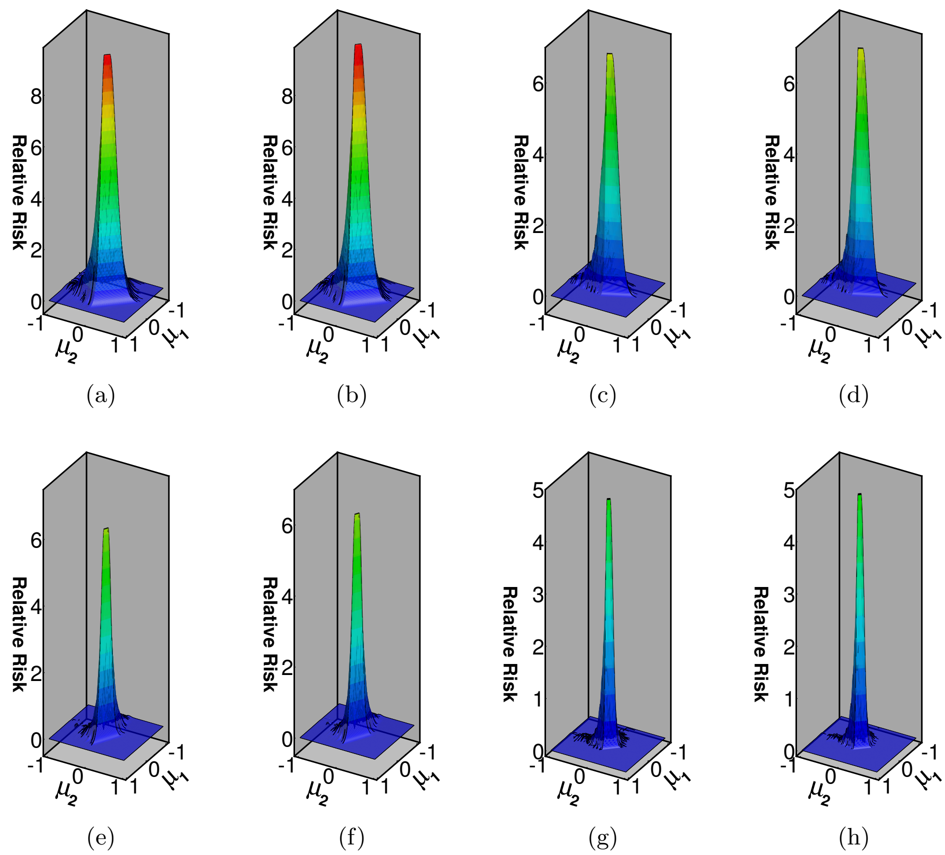


Figure 1: Fig. (a), (b), (c), (d), (e), (f), (g) and (h) represent relative percentage risk improvement plots of δ_{IB}^{1} over $\delta_{c_0}^{1}$ for (5,5,+0.2), (5,5,-0.2), (10,5,+0.2), (10,5,-0.2), (5,10,+0.2), (5,10,-0.2), (10,10,+0.2) and (10,10,-0.2), respectively when $(\mu_1, \mu_2) \in \mathbb{R}_2$. The first and second components of the triplet represent the sample sizes of the first and second population, respectively whereas the third component represents the value of p.

Based on the Fig. 2 we get the following observations.

(i) The region as well as the margin of the RRI of δ_{IB}^{1+} over $\delta_{c_0}^{1}$ become smaller for larger values of (n_1, n_2).

(ii) When μ_i tends to the zero, the RRI of δ_{IB}^{1+} over $\delta_{c_0}^{1}$ first increases and then decreases, $i = 1, 2$.

(iii) For fixed sample sizes (n_1, n_2), the RRI is marginally better for negative values of p than positive values of p. The RRI of δ_{IB}^{1+} over $\delta_{c_0}^{1}$ is 9.79% at $(\mu_1 = 0.02, \mu_2 = 0.08)$ for $(n_1 = 5, n_2 = 5)$ and $p = -2$, whereas for the same values of (μ_1, μ_2) and (n_1, n_2), the RRI of δ_{IB}^{1} over $\delta_{c_0}^{1}$ is 9.39%, when $p = 2$.

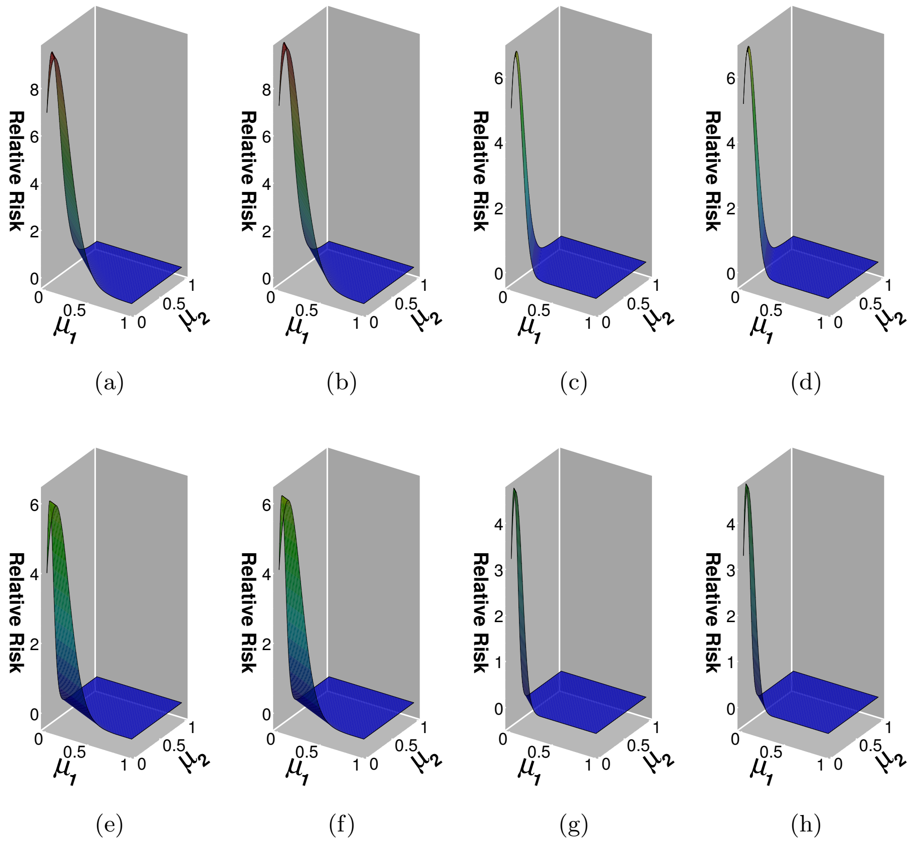
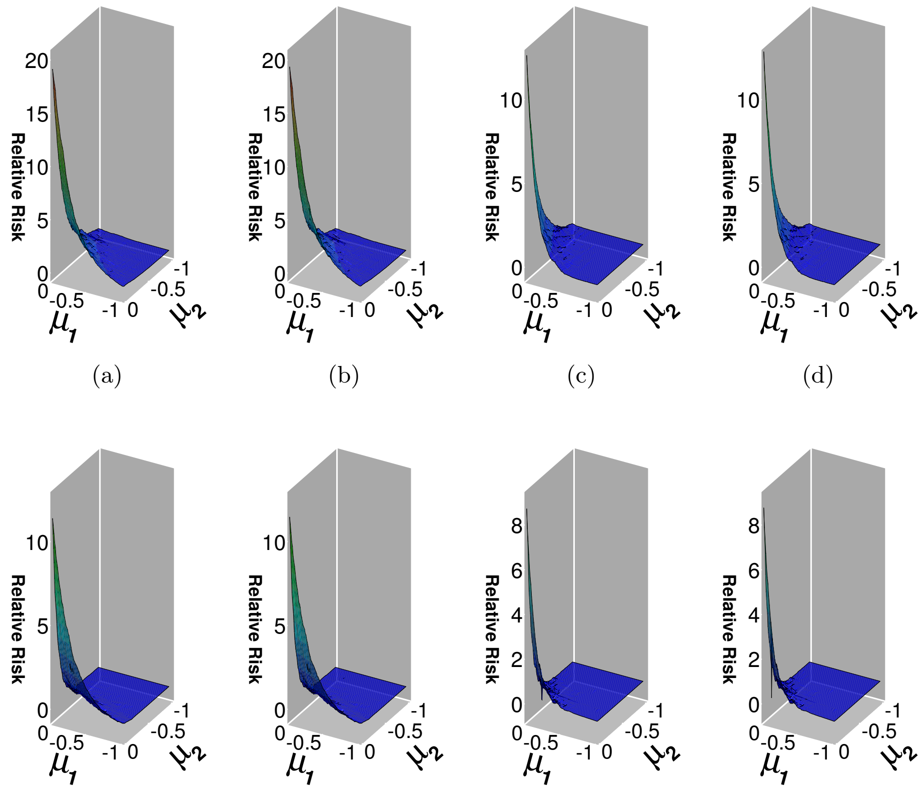


Figure 2: Fig. (a), (b), (c), (d), (e), (f), (g) and (h) represent relative percentage risk improvement plots of δ_{IB}^{1+} over $\delta_{c_0}^1$ for (5,5,+0.2), (5,5,-0.2), (10,5,+0.2), (10,5,-0.2), (5,10,+0.2), (5,10,-0.2), (10,10,+0.2) and (10,10,-0.2), respectively when $(\mu_1, \mu_2) \in \mathbb{R}_2^+$. The first and second components of the triplet represent the sample sizes of the first and second population, respectively whereas the third component represents the value of p.

Based on the Fig. 3, we notice the following points.

- **(i)** The margin and the region of the RRI of δ_{IB}^{1-} over $\delta_{c_0}^1$ become small when we increase the values of (n_1, n_2).
- **(ii)** When $(\mu_1, \mu_2) \to (0, 0)$, the RRI of δ_{IB}^{1-} over $\delta_{c_0}^1$ increases and it attains maximum at some point near origin.
- **(iii)** For fixed (n_1, n_2), the RRI of δ_{IB}^{1-} over $\delta_{c_0}^1$ is marginally better for negative values of p than positive values of p. For example, the RRI of δ_{IB}^{1-} over $\delta_{c_0}^1$ is 18.98% at $(\mu_1 = -0.01, \mu_2 = -0.01)$ for $(n_1 = 5, n_2 = 5)$ and $p = 2$, whereas for the same values of (μ_1, μ_2) and (n_1, n_2), the RRI of δ_{IB}^{1-} over $\delta_{c_0}^1$ is 19.20%, when $p = -2$.



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 3: Fig. (a), (b), (c), (d), (e), (f), (g) and (h) represent relative percentage risk improvement plots of δ_{IB}^{1-} over $\delta_{c_0}^{1}$ for (5,5,+0.2), (5,5,-0.2), (10,5,+0.2), (10,5,-0.2), (5,10,+0.2), (5,10,-0.2), (10,10,+0.2) and (10,10,-0.2), respectively when $(\mu_1, \mu_2) \in \mathbb{R}_2^-$. The first and second components of the triplet represent the sample sizes of the first and second population, respectively whereas the third component represents the value of p.

3. COMMON LOCATION BUT UNEQUAL SCALE PARAMETERS

In this section, we consider k exponential populations with a common location parameter μ and unknown but unequal scale parameters $\underline{\sigma}$. This model arises in life testing and reliability, where the common location parameter can be considered as minimum guarantee time of operation of several components and scale parameters are interpreted as unknown and possibly unequal failure rates of these components. Let the probability density function of the i-th population be

$$f_i(x|\mu,\sigma_i) = \begin{cases} \sigma_i^{-1} exp\{-(x-\mu)/\sigma_i\}, & \text{if } x > \mu \\ 0, & \text{otherwise,} \end{cases} \tag{3.1}$$

where $\mu \in \mathbb{R},\ \sigma_i > 0,\ i = 1, ..., k$. Let $(X_{i1}, ..., X_{in_i})$ be a random sample of size n_i drawn from the i-th population $(i = 1, ..., k)$ with probability density function given in (3.1). The expression of the Renyi entropy is $R_\alpha(\underline{\sigma}) = \sum_{i=1}^{k} \ln \sigma_i - k \ln \alpha/(1-\alpha)$. It is worthwhile to mention that the problem of estimating $R_\alpha(\underline{\sigma})$ with respect to the loss function of the form $L(\theta, \delta) = W(\delta - \theta)$ is equivalent to the problem of estimating $Q_2(\underline{\sigma}) = \sum_{i=1}^{k} \ln \sigma_i$. We consider the loss function as

$$L^2(Q_2(\underline{\sigma}), \delta) = \exp\Big\{p\Big(\delta - \sum_{i=1}^{k} \ln \sigma_i\Big)\Big\} - p\Big(\delta - \sum_{i=1}^{k} \ln \sigma_i\Big) - 1,\ p \neq 0. \tag{3.2}$$

Denote $Z_i = Y_i - n_i X_{i(1)},\ i = 1, ..., k$. For the i-th population, $(X_{i(1)}, Z_i)$ is a complete and sufficient statistic for (μ, σ_i). Moreover, Z_i and $X_{i(1)}$ are independently distributed, where $2\sigma_i^{-1} Z_i$ follows chi-square distribution with $2(n_i - 1)$ degrees of freedom and $X_{i(1)}$ follows an exponential distribution with location parameter μ and scale parameter σ_i/n_i. Further, define $X = min\{X_{1(1)}, ..., X_{k(1)}\}$ and $T_i = Y_i - n_i X$. It is easy to show that $(X, \underline{T})$ is a joint complete and sufficient statistic for $(\mu, \underline{\sigma})$, where $\underline{T} = (T_1, ..., T_k)$. The MLE of $Q_2(\underline{\sigma})$ is $\delta_{ML}^2 = \sum_{i=1}^{k} \ln T_i - \ln(\prod_{i=1}^{k} n_i)$. Also, X and $\underline{T}$ are independently distributed with respective probability density function

$$f_X(x) = \tau \exp\{-\tau(x-\mu)\}, \qquad x > \mu \tag{3.3}$$

and

$$f_{\underline{T}}(\underline{t}) = Nq\eta l \tau^{-1} \Big(\prod_{i=1}^{k} t_i^{n_i} - 1\Big) \ \exp\Big\{-\Big(\sum_{i=1}^{k} t_i \sigma_i^{-1}\Big)\Big\}, \quad t_i > 0, \tag{3.4}$$

where $\eta = \Big(\prod_{i=1}^{k} \sigma_i\Big)^{-n_i},\ \tau = \sum_{i=1}^{k} n_i \sigma_i^{-1},\ N = \sum_{i=1}^{k} n_i(n_i - 1),\ l = \Big(\prod_{i=1}^{k} \Gamma(n_i)\Big)^{-1},$ $q = \sum_{i=1}^{k} t_i^{-1}$ and $\underline{t} = (t_1, ..., t_k)$. Following steps analogous to Kayal *et al.* (2015), the UMVUE of $Q_2(\underline{\sigma})$ can be obtained as

$$\delta_{MV}^2 = \sum_{i=1}^{k} \ln T_i - \sum_{i=1}^{k} \frac{1 - (JT_i)^{-1}}{n_i - 1} - \sum_{i=1}^{k} \psi(n_i - 1), \tag{3.5}$$

where $J = \sum_{i=1}^{k} T_i^{-1}$.

3.1. Affine equivariant estimator

The estimation problem under study is invariant under $G_{a,b}$, a group of affine transformations, where $G_{a,b} = \{g_{a,b} : g_{a,b}(x) = ax + b,\ a > 0,\ b \in \mathbb{R}\}$. The form of an affine equivariant estimator can be obtained as (see Kayal *et al.* (2015))

$$\delta_\eta(X, \underline{T}) = k \ln T_1 + \eta(W_1, ..., W_{k-1}) = k \ln T_1 + \eta(\underline{W}), \tag{3.6}$$

where $\underline{W} = (W_1, ..., W_{k-1})$, $W_i = (T_{i+1}/T_1)$, $i = 1, ..., k-1$ and η is a real valued measurable function. The following theorem provides a general inadmissibility result for an affine equivariant estimator of the form (3.6).

Theorem 3.1. Let δ_η be the form of an affine equivariant estimator given in (3.6). Further, define the estimator δ^*_η by

$$\delta^*_\eta = \begin{cases} \delta_\eta, & \text{if} \quad \eta(\underline{w}) \geq \eta_0(\underline{w}) \\ \delta_{\eta_0}, & \text{if} \quad \eta(\underline{w}) < \eta_0(\underline{w}), \end{cases}$$

where $\underline{w} = (w_1, ..., w_{k-1})$ and $\eta_0(\underline{w}) = \ln[k^k(\prod_{i=1}^{k-1} w_i)] - \frac{1}{p} \ln\left(\frac{\Gamma(n+kp-1)}{\Gamma(n-1)}\right)$. Then under the linex loss function (3.2), δ^*_η improves δ_η if there exists $(\mu, \underline{\sigma})$ such that $P_{\mu,\underline{\sigma}}(\eta(\underline{W}) < \eta_0(\underline{W})) > 0$.

Proof: The risk function of δ_η can be written as

$$R(\mu, \underline{\sigma}, \delta_\eta) = E^{\underline{W}} R_1(\mu, \underline{\sigma}, \underline{W}, \delta_\eta),$$

where $R_1(\mu, \underline{\sigma}, \underline{w}, \delta_\eta)$ denotes the conditional risk of δ_η given $\underline{W} = \underline{w}$, and is given by

$$R_1(\mu, \underline{\sigma}, \underline{w}, \delta_\eta) = E\Big[\Big(\exp\Big\{p\Big(k \ln T_1 + \eta(\underline{W}) - \sum_{i=1}^{k} \ln \sigma_i\Big)\Big\} - p\Big(k \ln T_1 + \eta(\underline{W}) - \sum_{i=1}^{k} \ln \sigma_i\Big) - 1\Big)\Big|\underline{W} = \underline{w}\Big].$$

Note that R_1 is a convex function in η and minimized at

$$\hat{\eta}(\underline{\sigma}, \underline{w}) = \frac{1}{p} \ln\left(\frac{\Big(\prod_{i=1}^{k} \sigma_i\Big)^p}{E(T_1^{kp}|\underline{W} = \underline{w})}\right). \tag{3.7}$$

To evaluate $\hat{\eta}(\underline{\sigma}, \underline{w})$ in (3.7), we need to derive the conditional distribution of T_1 given $\underline{W} = \underline{w}$ which is given by

$$f_{T_1|\underline{W}}(t_1|\underline{w}) = \Gamma^{-1}(n-1) s^{n-1}\, t_1^{n-2}\, e^{-st_1},\ t_1 > 0,\ w_i > 0, \tag{3.8}$$

where $s = \sigma_1^{-1} + \sum_{i=1}^{k-1} w_i \sigma_{i+1}^{-1}$ and $n = \sum_{i=1}^{k} n_i$. Using (3.8) we obtain

$$E(T_1^{kp}|\underline{W} = \underline{w}) = \frac{\Gamma(n+kp-1)}{\Gamma(n-1)} \frac{1}{s^{kp}}.$$

Putting $E(T_1^{kp}|\underline{W}=\underline{w})$ in (3.7), we get

$$\hat{\eta}(\underline{\sigma},\underline{w}) = \ln\left(s^k\prod_{i=1}^{k}\sigma_i\right) - \frac{1}{p}\ln\left(\frac{\Gamma(n+kp-1)}{\Gamma(n-1)}\right). \tag{3.9}$$

For fixed values of $\underline{w}$, the supremum and infimum of $\hat{\eta}(\underline{\sigma},\underline{w})$ over $\underline{\sigma}$ can be obtained as

$$\begin{aligned} \sup_{\underline{\sigma}}\hat{\eta}(\underline{\sigma},\underline{w}) &= +\infty, \\ \inf_{\underline{\sigma}}\hat{\eta}(\underline{\sigma},\underline{w}) &= \ln\left[k^k\Big(\prod_{i=1}^{k-1}w_i\Big)\right] - \frac{1}{p}\ln\left(\frac{\Gamma(n+kp-1)}{\Gamma(n-1)}\right) \\ &= \phi_0,\ \text{say.} \end{aligned} \tag{3.10}$$

An application of the Brewster-Zidek technique on R_1, then completes the proof the theorem. □

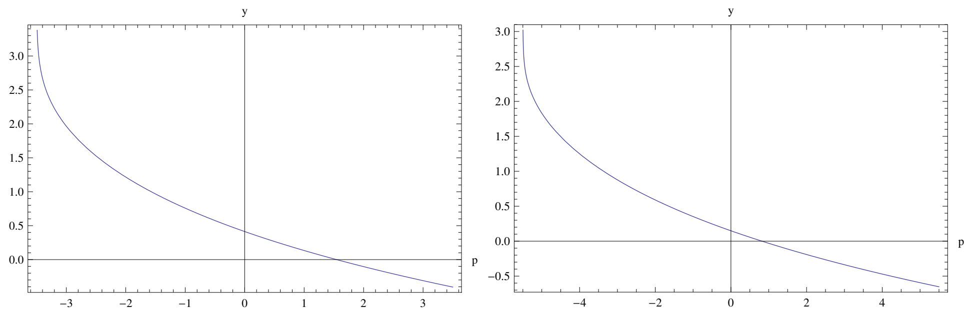
Note that δ^2_{ML} and δ^2_{MV} belong to the class of affine equivariant estimators of the form (3.6) when $\eta(\underline{w})$ is equal to $\ln\left(\frac{T_2}{T_1}\cdots\frac{T_k}{T_1}\right) - k\ln(\prod_{i=1}^k n_i)$ and $\ln\left(\frac{T_2}{T_1}\cdots\frac{T_k}{T_1}\right) - \sum_{i=1}^k\frac{1-(JT_i)^{-1}}{n_i-1} - \sum_{i=1}^k\psi(n_i-1)$, respectively. The Theorem 3.1 then leads to the following corollaries.

Corollary 3.1. The MLE δ^2_{ML} is inadmissible and dominated by

$$\delta^2_{IML} = \begin{cases} \ln[(kT_1)^k(\prod_{i=1}^{k-1}w_i)] - \frac{1}{p}\ln(\frac{\Gamma(n+kp-1)}{\Gamma(n-1)}), \\ \qquad\qquad\qquad\qquad \text{if } \ln(k^k\prod_{i=1}^k n_i) - \frac{1}{p}\ln(\frac{\Gamma(n+kp-1)}{\Gamma(n-1)}) > 0 \\ \sum_{i=1}^k \ln T_i - \ln(\prod_{i=1}^k n_i), \qquad \text{otherwise.} \end{cases}$$

Corollary 3.2. The UMVUE δ^2_{MV} is inadmissible and dominated by

$$\delta^2_{IMV} = \begin{cases} \ln[(kT_1)^k(\prod_{i=1}^{k-1}W_i)] - \frac{1}{p}\ln(\frac{\Gamma(n+kp-1)}{\Gamma(n-1)}), \\ \qquad\qquad\qquad \text{if } \ln(k^k) - \frac{1}{p}\ln(\frac{\Gamma(n+kp-1)}{\Gamma(n-1)}) \\ \qquad\qquad\qquad\qquad + \sum_{i=1}^k\psi(n_i-1) + \sum_{i=1}^k\frac{1-(JT_i)^{-1}}{n_i-1} > 0 \\ \sum_{i=1}^k \ln T_i - \sum_{i=1}^k\frac{1-(JT_i)^{-1}}{n_i-1} - \sum_{i=1}^k\psi(n_i-1), \qquad \text{otherwise.} \end{cases}$$



(a) (b)

Figure 4: Fig. (a) represents the plot of $\ln(k^k \prod_{i=1}^{k} n_i) - \frac{1}{p}\ln(\frac{\Gamma(n+kp-1)}{\Gamma(n-1)})$ for $(n_1 = 4, n_2 = 4)$ when $k = 2$; and Fig. (b) represents the plot of that for $(n_1 = 4, n_2 = 8)$, when $k = 2$;

3.2. Scale equivariant estimator

In this section, we introduce invariance to the problem under the group of scale transformations $G_a = \{g_a(x) = ax,\ a > 0\}$. The form of a scale equivariant estimator is obtained as

$$\begin{aligned}\delta_\xi(X, \underline{T}) &= k \ln T_1 + \xi\Big(\frac{X}{T_1}, \frac{T_2}{T_1}, \frac{T_3}{T_1}, ..., \frac{T_k}{T_1}\Big) \\ &= k \ln T_1 + \xi(\underline{V}), \text{(say)},\end{aligned} \tag{3.11}$$

where $\underline{V} = (V_1, V_2, ..., V_k)$, $V_1 = X/T_1$, and $V_i = T_i/T_1$, $i = 2, 3, ..., k$. The risk function of δ_ξ given in (3.11) is

$$R(\mu, \underline{\sigma}, \delta_\xi) = E^{\underline{V}} R_1(\mu, \underline{\sigma}, \underline{V}, \delta_\xi),$$

where $R_1(\mu, \underline{\sigma}, \underline{v}, \delta_\xi)$ denotes the conditional risk of δ_ξ given $\underline{V} = \underline{v}$, is given by

$$\begin{aligned}R_1(\mu, \underline{\sigma}, \underline{v}, \delta_\xi) &= E\Big[\Big(\exp\Big\{p\Big(k \ln T_1 + \xi(\underline{V}) - \sum_{i=1}^{k} \ln \sigma_i\Big)\Big\} \\ &\quad - p\Big(k \ln T_1 + \xi(\underline{V}) - \sum_{i=1}^{k} \ln \sigma_i\Big) - 1\Big)\Big|\underline{V} = \underline{v}\Big]\end{aligned}$$

which is minimized at

$$\hat{\xi}(\mu, \underline{\sigma}, \underline{v}) = \frac{1}{p} \ln\left(\frac{\Big(\prod_{i=1}^{k} \sigma_i\Big)^p}{E(T_1^{kp} | \underline{V} = \underline{v})}\right). \tag{3.12}$$

The joint probability density function of T_1 and $\underline{V}$ is

$$f(t_1, \underline{v}) \;=\; C\exp\{\mu\tau\}\Big(\prod_{i=2}^{k} v_i^{n_i-1}\Big)\Big(1+\sum_{i=2}^{k} v_i^{-1}\Big) \times\;\exp\Big\{-\Big(v_1\tau+\Big(\sigma_1^{-1}+\sum_{i=2}^{k} v_i\sigma_i^{-1}\Big)\Big)t_1\Big\}t_1^{n-1}, \tag{3.13}$$

where $C = N\eta l,\ t_1 > 0, t_1v_1 > \mu, v_2 > 0, v_3 > 0, ..., v_k > 0$. Note that to obtain the supremum and infimum of $\hat{\xi}(\mu, \underline{\sigma}, \underline{v})$, it is required to derive the conditional distribution of $T_1|\underline{V} = \underline{v}$, which can be obtained through the arguments of the cases considered in Section 3.2 of the paper by Kayal *et al.* (2015). Hence we omit the details.

Case **(i):** Under the assumptions that $\mu > 0$ and $v_1 > 0$, we obtain

$$\hat{\xi}(\mu, \underline{\sigma}, \underline{v}) = \ln\Big(\prod_{i=1}^{k}\sigma_i\Big) - \frac{1}{p}\ln(I_1^*/I_1), \tag{3.14}$$

where $I_1^* = \int_{\mu/v_1}^{\infty} exp\{-At_1\}\ t_1^{kp+n-1}dt_1,\ I_1 = \int_{\mu/v_1}^{\infty} exp\{-At_1\}\ t_1^{n-1}dt_1$ and $A = v_1\tau + (\sigma_1^{-1} + \sum_{i=2}^{k} v_i\sigma_i^{-1})$. Using the monotone likelihood ratio property it is not hard to show that

$$\sup_{\mu,\underline{\sigma}}\hat{\xi}(\mu, \underline{\sigma}, \underline{v}) = +\infty \quad\text{and}\quad \inf_{\mu,\underline{\sigma}}\hat{\xi}(\mu, \underline{\sigma}, \underline{v}) = -\infty.$$

Case **(ii):** When $\mu < 0$ and $v_1 > 0$, we get

$$\hat{\xi}(\mu, \underline{\sigma}, \underline{v}) = \ln\Big(A^k\prod_{i=1}^{k}\sigma_i\Big) - \frac{1}{p}\ln\Big(\frac{\Gamma(n+kp)}{\Gamma(n)}\Big). \tag{3.15}$$

It is easy to see that supremum of $\widehat{\xi}(\mu, \underline{\sigma}, \underline{v})$ is $+\infty$. The infimum of $\widehat{\xi}(\mu, \underline{\sigma}, \underline{v})$ can be obtained by applying geometric mean-harmonic mean inequality to the variables $(\sigma_1/n_1v_1 + 1), (\sigma_2/n_2v_1 + v_2), ..., (\sigma_k/n_kv_1 + v_k)$ and is given by

$$\inf_{\mu,\underline{\sigma}}\widehat{\xi}(\mu, \underline{\sigma}, \underline{v}) = \ln\Big(k^k(n_1v_1 + 1)\prod_{i=2}^{k}(n_iv_1 + v_i)\Big) - \frac{1}{p}\ln\Big(\frac{\Gamma(n+kp)}{\Gamma(n)}\Big).$$

Case **(iii):** Let $\mu < 0$ and $v_1 < 0$. Under these assumptions, we have

$$\hat{\xi}(\mu, \underline{\sigma}, \underline{v}) = \ln\Big(\prod_{i=1}^{k}\sigma_i\Big) - \frac{1}{p}\ln(I_2^*/I_2), \tag{3.16}$$

where $I_2^* = \int_0^{\mu/v_1} exp\{-At_1\}t_1^{kp+n-1}dt_1$ and $I_2 = \int_0^{\mu/v_1} exp\{-At_1\}t_1^{n-1}dt_1$. Note that the value of A may be positive or negative. For both $A > 0$ and $A < 0$, it can be shown that the supremum and infimum of $\hat{\xi}(\mu, \underline{\sigma}, \underline{v})$ are $+\infty$ and $-\infty$, respectively. As in the full sample space the supremum and infimum choices of $\hat{\xi}(\mu, \underline{\sigma}, \underline{v})$ are $+\infty$ and $-\infty$, respectively, therefore we do not get any improvement over the BAEE. But if we restrict the parameter space to $\mu < 0$, then an improvement over the BAEE exists, which is shown in the next theorem. Define

$$\xi_0(\underline{v}) = \begin{cases} \ln(v^*) - \frac{1}{p}\ln(\frac{\Gamma(n+kp)}{\Gamma(n)}), & \text{if } v_1 > 0 \text{ and} \\ \qquad v^* > exp\{\xi(\underline{v}) + \frac{1}{p}\ln\,(\frac{\Gamma(n+kp)}{\Gamma(n)})\} & \\ \xi(\underline{v}), & \text{otherwise,} \end{cases} \tag{3.17}$$

where $v^* = k^k(n_1 v_1 + 1)\prod_{i=2}^{k}(n_i v_1 + v_i)$.

Theorem 3.2. Let δ_ξ be a scale equivariant estimator of the form (3.12) and $\xi_0(\underline{v})$ be as defined in (3.17). If there exists a $(\mu, \underline{\sigma})$ such that $P_{(\mu,\underline{\sigma})}(\xi_0(\underline{V}) \neq \xi(\underline{V})) > 0$, then the estimator δ_{ξ_0} dominates δ_ξ, with respect to the linex loss function, when $\mu < 0$.

As a consequence of the Theorem 3.2, the following corollary immediately follows.

Corollary 3.3. When $\mu < 0$, the MLE and the UMVUE are inadmissible and dominated by

$$\delta_{IML}^{2-} = \begin{cases} \ln(T_1^k V^*) - \frac{1}{p}\ln(\frac{\Gamma(n+kp)}{\Gamma(n)}), & \text{if } V_1 > 0 \text{ and} \\ \qquad V^* > exp\{\sum_{i=2}^k \ln V_i - \ln(\prod_{i=1}^k n_i) + \frac{1}{p}\ln(\frac{\Gamma(n+kp)}{\Gamma(n)})\} & \\ \sum_{i=1}^k \ln T_i - \ln(\prod_{i=1}^k n_i), & \text{otherwise} \end{cases}$$

and

$$\delta_{IMV}^{2-} = \begin{cases} \ln(T_1^k V^*) - \frac{1}{p}\ln(\frac{\Gamma(n+kp)}{\Gamma(n)}), & \text{if } V_1 > 0 \\ \qquad \text{and } V^* > exp\{\sum_{i=2}^k \ln V_i - \sum_{i=1}^k \frac{1-(JT_i)^{-1}}{n_i - 1} & \\ \qquad\qquad - \sum_{i=1}^k \psi(n_i - 1) + \frac{1}{p}\ln(\frac{\Gamma(n+kp)}{\Gamma(n)})\} & \\ \sum_{i=1}^k \ln T_i - \sum_{i=1}^k \frac{1-(JT_i)^{-1}}{n_i - 1} - \sum_{i=1}^k \psi(n_i - 1), & \text{otherwise,} \end{cases}$$

where $V^* = k^k(n_1 V_1 + 1)\prod_{i=2}^{k}(n_i V_1 + V_i)$ and $J = \sum_{i=1}^k T_i^{-1}$.

3.3. Numerical comparisons

Here we present risk and relative risk of various estimators derived in Section 3. As in Section 2.2, the risk values were calculated using Monte-Carlo simulation based on $10,000$ samples of different combinations of sample sizes (n_1, n_2) and different values of $(\mu, \sigma_1, \sigma_2)$ and p. We present few of them in tabular form below. Table 1 is for the risk values of δ^2_{MV} and δ^2_{IMV} when $k = 2$. The RRI of the estimators δ^{2-}_{IMV}, δ^{2-}_{IML} and δ^2_{MV} over δ^2_{ML} is presented in Table 2 and Table 3 for $k = 2$. Different combinations of (n_1, n_2) and different values of p, σ_1, σ_2 have been chosen. We have considered sample sizes $(5, 5), (5, 10), (10, 5)$ and $(10, 10)$. The values of p have been chosen as $-0.5, -1, +0.5$ and $+1$. Here, we have presented very few values, however, similar observations are made for various other values of (n_1, n_2), p and μ.

Table 1: The risk values of δ^2_{MV} and δ^2_{IMV} for $k = 2$.

p	μ	(n_1,n_2)	(σ_1,σ_2)	δ^2_{MV}	δ^2_{IMV}	p	μ	(n_1,n_2)	(σ_1,σ_2)	δ^2_{MV}	δ^2_{IMV}
-0.75	0.2	(5,5)	(0.5,0.5)	0.164190	0.157732	-2	-0.2	(10,5)	(0.5,0.5)	1.157259	1.068251
			(0.5,1.0)	0.159419	0.153242				(0.5,1.0)	1.135408	1.009288
			(1.0,0.5)	0.169174	0.162430				(1.0,0.5)	1.227956	1.157454
			(1.0,1.0)	0.164190	0.157732				(1.0,1.0)	1.157259	1.068251
-1.5	0.5	(5,10)	(0.5,0.5)	0.561950	0.560949	-2.5	-0.5	(10,10)	(0.5,0.5)	0.974965	0.669685
			(0.5,1.0)	0.657281	0.653271				(0.5,1.0)	1.007035	0.680643
			(1.0,0.5)	0.622936	0.621872				(1.0,0.5)	0.931061	0.654255
			(1.0,1.0)	0.561950	0.552783				(1.0,1.0)	0.974965	0.669685

Table 2: The relative percentage risk improvement over δ^2_{ML} by δ^{2-}_{IMV}, δ^{2-}_{IML} and δ^2_{MV} for $k = 2$.

p	μ	(n_1,n_2)	(σ_1,σ_2)	δ^{2-}_{IMV}	δ^{2-}_{IML}	δ^2_{MV}	p	μ	(n_1,n_2)	(σ_1,σ_2)	δ^{2-}_{IMV}	δ^{2-}_{IML}	δ^2_{MV}
-0.5	-0.2	(5,5)	(0.5,0.5)	1.18	1.50	41.46	0.5	-0.2	(5,5)	(0.5,0.5)	0.38	0.69	17.72
			(0.5,1.0)	1.55	2.59	41.68				(0.5,1.0)	0.59	1.31	17.81
			(1.0,0.5)	3.38	3.75	41.29				(1.0,0.5)	2.23	2.41	17.85
			(1.0,1.0)	5.19	7.03	41.47				(1.0,1.0)	3.21	4.05	17.72
-0.5	-0.2	(5,10)	(0.5,0.5)	0.15	0.23	35.57	0.5	-0.2	(5,10)	(0.5,0.5)	0.12	0.16	14.02
			(0.5,1.0)	0.41	0.82	37.06				(0.5,1.0)	0.23	0.47	15.64
			(1.0,0.5)	0.08	0.16	33.86				(1.0,0.5)	0.02	0.07	13.02
			(1.0,1.0)	0.39	1.25	35.57				(1.0,1.0)	0.42	0.53	14.02
-0.5	-0.2	(10,5)	(0.5,0.5)	0.12	0.25	36.64	0.5	-0.2	(10,5)	(0.5,0.5)	0.04	0.09	15.87
			(0.5,1.0)	0.23	0.55	35.37				(0.5,1.0)	0.08	0.23	15.63
			(1.0,0.5)	0.24	1.31	37.83				(1.0,0.5)	0.10	0.45	16.71
			(1.0,1.0)	2.66	3.68	36.65				(1.0,1.0)	0.90	1.66	15.87
-0.5	-0.2	(10,10)	(0.5,0.5)	0.01	0.02	27.68	0.5	-0.2	(10,10)	(0.5,0.5)	0.02	0.05	13.95
			(0.5,1.0)	0.23	0.34	28.20				(0.5,1.0)	0.04	0.22	14.72
			(1.0,0.5)	0.42	0.48	26.89				(1.0,0.5)	0.07	0.31	14.23
			(1.0,1.0)	0.77	1.42	27.68				(1.0,1.0)	0.43	0.82	13.95

Table 3: The relative percentage risk improvement over δ^2_{ML} by δ^{2-}_{IMV}, δ^{2-}_{IML} and δ^2_{MV} for $k=2$.

p	μ	(n_1,n_2)	(σ_1,σ_2)	δ^{2-}_{IMV}	δ^{2-}_{IML}	δ^2_{MV}	p	μ	(n_1,n_2)	(σ_1,σ_2)	δ^{2-}_{IMV}	δ^{2-}_{IML}	δ^2_{MV}
-1	-0.2	(5,5)	(0.5,0.5)	1.60	2.08	50.16	1	-0.2	(5,5)	(0.5,0.5)	0.27	0.44	0.43
			(0.5,1.0)	2.25	3.21	50.37				(0.5,1.0)	0.34	0.78	0.42
			(1.0,0.5)	4.09	4.76	49.88				(1.0,0.5)	1.01	1.58	0.77
			(1.0,1.0)	6.67	8.56	50.16				(1.0,1.0)	2.55	2.98	0.44
-1	-0.2	(5,10)	(0.5,0.5)	0.15	0.23	43.88	1	-0.2	(5,10)	(0.5,0.5)	0.10	0.11	0.98
			(0.5,1.0)	0.40	1.07	45.22				(0.5,1.0)	0.21	0.37	1.03
			(1.0,0.5)	0.06	0.35	42.00				(1.0,0.5)	0.02	0.05	0.76
			(1.0,1.0)	0.51	1.81	43.88				(1.0,1.0)	0.29	0.42	1.20
-1	-0.2	(10,5)	(0.5,0.5)	0.03	0.45	44.71	1	-0.2	(10,5)	(0.5,0.5)	0.02	0.13	2.01
			(0.5,1.0)	0.59	1.12	43.14				(0.5,1.0)	0.08	0.21	2.66
			(1.0,0.5)	0.75	1.98	45.95				(1.0,0.5)	0.09	0.29	2.47
			(1.0,1.0)	4.06	5.56	44.71				(1.0,1.0)	0.45	1.15	2.36
-1	-0.2	(10,10)	(0.5,0.5)	0.12	0.16	33.43	1	-0.2	(10,10)	(0.5,0.5)	0.04	0.06	5.59
			(0.5,1.0)	0.18	0.27	33.84				(0.5,1.0)	0.05	0.09	6.49
			(1.0,0.5)	0.16	0.30	32.81				(1.0,0.5)	0.07	0.33	4.14
			(1.0,1.0)	1.24	1.73	33.43				(1.0,1.0)	0.34	0.56	5.59

The following conclusions are evident from Table 2 and Table 3:

(i) We observe marginal RRI over δ^2_{ML} by δ^{2-}_{IML}, δ^{2-}_{IMV}, but substantial improvement by δ^2_{MV}.

(ii) For fixed (n_1,n_2) and μ, the RRIs for the estimators δ^{2-}_{IMV} and δ^{2-}_{IML} over δ^2_{ML} are marginally better, whereas we observe substantial RRIs for the estimator δ^2_{MV} over δ^2_{ML} for negative values of p than positive values of p.

(iii) For fixed μ, p and (n_1,n_2) the RRI of δ^{2-}_{IML} and δ^{2-}_{IMV} over δ_{ML} approximately increases with (σ_1,σ_2), but we do not observe such behaviour for the RRI of δ^2_{MV} over δ^2_{ML}.

4. CONCLUDING REMARKS

In this paper, the problem of estimating the Renyi entropy of several exponential distributions has been investigated with respect to a linex loss function. The concept of invariance has been used to derive improved estimators over the standard ones such as MLE and UMVUE. We have considered two distinct models here. Both these models have various applications in real life experiments. In the first model, the location parameters are distinct but the scale parameter is assumed to be common. Improved estimators over the BAEE have been obtained when parameters space is restricted as well as unrestricted. In the second model, the scale parameters are distinct but the location parameter is assumed to be common. Affine and scale equivariant estimators improving over the UMVUE and MLE are obtained under restrictions on the parameter space. Finally, margins of relative risk improvements by new estimators are determined numerically using simulations.

ACKNOWLEDGMENTS

The authors wish to thank the Associate Editor and three anonymous referees for their comments which have greatly improved the initial version of this manuscript.

REFERENCES

[1] Bercher, J.F. (2008). On some entropy functionals derived from Renyi information divergence, *Information Sciences*, **178**, 2489–2506.

[2] Brewster, J.F. and Zidek, J.V. (1974). Improving on equivariant estimators, *Annals of Statistics*, **2**, 21–38.

[3] Cover, T.M. and Thomas, J.A. (2006). *Elements of Information Theory*, Wiley, New York.

[4] De-Gregorio, A. and Iacus, S.M. (2009). On Renyi information for ergodic diffusion processes, *Information Sciences*, **179**, 279–291.

[5] Golshani, L. and Pasha, E. (2010). Renyi entropy rate for Gaussian processes, *Information Sciences*, **180**, 1486–1491.

[6] Hughes, M.S.; Marsh, J.N.; Arbeit, J.M.; Neumann, R.G.; Fuhrhop, R.W.; Wallace, K.D.; Thomas, L.; Smith, J.; Agyem, K.; Lanza, G.M.; Wickline, S.A. and McCarthy, J.E. (2009). Application of Renyi entropy for ultrasonic molecular imaging, *The Journal of the Acoustical Society of America*, **125**, 3141–3145.

[7] Kapur, J.N. (1990). *Maximum Entropy Models in Science and Engineering*, Wiley, New York.

[8] Kayal, S. and Kumar, S. (2011a). Estimating the Shannon entropy of several exponential populations, *International Journal of Statistics and Economics*, **7**, 42–52.

[9] Kayal, S. and Kumar, S. (2011b). Estimating entropy of an exponential population under linex loss function, *Journal of Indian Statistical Association*, **49**, 91–112.

[10] Kayal, S. and Kumar, S. (2013). Estimation of the Shannon's entropy of several shifted exponential populations, *Statistics and Probability Letters*, **83**, 1127–1135.

[11] Kayal, S.; Kumar, S. and Vellaisamy, P. (2015). Estimating the Renyi entropy of several exponential populations, *Brazilian Journal of Probability and Statistics*, **29**, 94–111.

[12] Lehmann, E.L. and Casella, G. (1998). *Theory of Point Estimation*, Second Ed., Springer, New York.

[13] Maszczyk, T. and Duch, W. (2008). Comparison of Shannon, Renyi and Tsallis entropy used in decision trees, *In Artificial Intelligence and Soft Computing-ICAISC, Proceedings*, 5097, 643–651.

[14] Misra, N.; Singh, H. and Demchuk, E. (2005). Estimation of the entropy of a multivariate normal distribution, *Journal of Multivariate Analysis*, **92**, 324–342.

[15] Nilsson, M. (2006). *Entropy and Speech*, Ph.D. Dissertation, Royal Institute of Technology, Stockholm, KTH.

[16] Obin, N. and Liuni, M. (2012). On the generalization of Shannon entropy for speech recognition, *IEEE Workshop on Spoken Language Technology, Miami, USA*, December 2–5, 97–102.

[17] Pharwaha, A.P.S. and Khehra, B.S. (2009). Shannon and non-Shannon measures of entropy for statistical texture feature extraction in digitized mammograms, *Proceedings of the World Congress on Engineering and Computer Science, San Francisco, USA*, October 20–22, 2, 1286–1291.

[18] Renyi, A. (1961). On measures of entropy and information, *Proceedings of Fourth Berkeley Symposium on Mathematics, Statistics and Probability*, University of California Press, Berkeley, 1, 1286–1291.

[19] Renyi, A. (2012). *Probability Theory*, North-Holland, Amsterdam.

[20] Shannon, C. (1948). The mathematical theory of communication, *Bell System Technical Journal*, **27**, 379–423.

[21] Song, K.S. (2001). Renyi information, loglikelihood and an intrinsic distribution measure, *Journal of Statistical Planning and Inference*, **93**, 51–69.

[22] Varian, H.R. (1975). *A Bayesian approach to real estate assessment.* In "Studies in Bayesian Econometrics and Statistics in Honour of L.J. Savage" (S.E. Feinberg and A. Zellner, Eds.), Amsterdam North Holland, 195–208.

[23] Zellner, A. (1986). Bayesian estimation and prediction using asymmetric loss functions, *Journal of American Statistical Association*, **81**, 446–451.

REVSTAT – Statistical Journal
Volume 15, Number 4, October 2017, 523–536

DISCRIMINATING BETWEEN NORMAL AND GUMBEL DISTRIBUTIONS

Authors: Abdelaziz Qaffou
– Department of Applied Mathematics, Faculty of Sciences and Techniques, Sultan Moulay Slimane University, Beni Mellal, Morocco
aziz.qaffou@gmail.com

Abdelhak Zoglat
– Department of Mathematics, Faculty of Sciences, Mohammed V University, Rabat, Morocco
azoglat@gmail.com

Received: March 2015 Revised: December 2015 Accepted: January 2016

Abstract:

- The normal and Gumbel distributions are much alike in practical engineering, in flood frequency and they are similar in appearance, especially for small samples. So the aim of this paper is to discriminate between these two distributions. Considering the logarithm of the ratio of the maximized likelihood (RML) as test statistic, its asymptotic distribution is found under both normal and Gumbel distributions, which can be used to compute the probability of correct selection (PCS). Finally, Monte Carlo (MC) simulations are performed to examine how the asymptotic results work for finite sample sizes.

Key-Words:

- *Normal and Gumbel distributions; probability of correct selection; ratio of maximized likelihood; Monte Carlo simulation.*

AMS Subject Classification:

- 62H10, 62H12, 62H15, 65C05.

1. INTRODUCTION

In engineering practice, risk criteria and economic considerations are important parts of a project design. These criteria are crucial, for example, in the design of an urban sewer network, the sizing of a hydraulic structure, or the conception of a storage capacity system. The adequate knowledge of design events (e.g., design flood magnitudes) is often helpful for the proper sizing of a project to avoid the high initial investments associated with the oversizing of the project and the large future failure costs resulting from its undersizing.

To estimate these design events, statistical frequency analysis of hydrological data is often used; it consists of fitting a probability distribution to a set of recorded hydrological values (e.g., annual maximum flood series) and obtaining estimated results concerning the underlying population. Estimates are often needed for such quantities as the magnitude of an extreme event (quantile) x_T, corresponding to a return period T. Evidently, the reliability of the estimates depends largely on the quality of the data as well as the length of the period of record.

The aim of this paper is to discriminate between normal and Gumbel distributions. These two distributions are widely applied in engineering and often used as a model for hydrologic data sets. Some of its recent application areas include flood frequency analysis, network and software reliability engineering, nuclear engineering and epidemic modeling.

There are many practical applications where Gumbel and normal distributions are similar in appearance and the two distributions cannot be distinguished from one another. Normal and Gumbel distributions belong to the location scale family. Discriminating between any two general probability distribution functions from the location scale family was widely investigated in the literature. See, for instance, [1], [2], [4], [5], [6], [7], [15] and [9] who studied the discrimination problem in general between the two models. Besides, [16], [19] and [22] studied the discrimination problem between lognormal and gamma distributions. [3] and [10] studied the discrimination problem between Weibull and gamma distributions. Recently, Gupta and Kundu considered the discrimination problem between Weibull and generalized exponential distributions, between gamma and generalized exponential distributions and between lognormal and generalized exponential distributions (see, [12], [13], [18]).

Among the discrimination problems, the one for Weibull and lognormal distributions is particularly important and has received much attention; this is because the two distributions are the most popular ones for analyzing the lifetime of electronic products. [8] adopted the ratio of maximized likelihood (RML) in discriminating between the two distributions for complete data and provided the percentile points for some sample sizes by simulation. Recently, [17] considered the discrimination problem for complete data using the RML procedure.

In the present work, to discriminate between normal and Gumbel distributions, we consider the ratio of maximized likelihood (RML) as test statistic. Based on the result of [21], the asymptotic distribution of the logarithm of the RML is found under both normal and Gumbel distributions, which can be used to compute the probability of correct selection (PCS). For small sample size, maximum likelihood estimators (MLE) of Gumbel parameters are biased; henceforth, we will use a correction for the bias introduced by [11] and [14].

The rest of the paper is organized as follows. Section 2 is dedicated to the mathematical notations that we use in this paper. In Section 3 we describe the logarithm of RML as test statistic, their asymptotic distributions under both normal and Gumbel distributions are obtained. Monte Carlo simulations are presented in Section 4 to examine how the asymptotic results work for finite samples. Finally, we conclude the paper in Section 5.

2. NOTATION

To facilitate the analysis that follows, we use the following notations. A normal distribution with mean μ and variance σ^2, denoted by $N(\mu,\sigma^2)$, has a probability density function (pdf) given by

$$f_N(x,\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp -\frac{(x-\mu)^2}{2\sigma^2},\ \ x\in\mathbb{R}.$$

The maximum likelihood estimators of μ and σ^2 are respectively given by

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i := \bar{X} \text{ and } \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2. \tag{2.1}$$

A Gumbel distribution with location parameter α and scale parameter β, denoted by $G(\alpha,\beta)$, has a pdf given by

$$f_G(x,\alpha,\beta) = \frac{1}{\beta}\exp\left[-\frac{x-\alpha}{\beta} - \exp -\frac{x-\alpha}{\beta}\right],\ \ x\in\mathbb{R},$$

and the maximum likelihood estimators of its parameters satisfy the following equations

$$\hat{\beta} = \bar{X} - \frac{\sum_{i=1}^{n} X_i \exp -\frac{X_i}{\hat{\beta}}}{\sum_{i=1}^{n}\exp -\frac{X_i}{\hat{\beta}}} \text{ and } \hat{\alpha} = -\hat{\beta}\ln\left[\frac{1}{n}\sum_{i=1}^{n}\exp -\frac{X_i}{\hat{\beta}}\right]. \tag{2.2}$$

These estimators, obtained as numerical solutions to the above equations, are known to be biased when the sample size is small. [11] proposed a correction for that bias:

$$\hat{\beta}_c^* = \frac{\hat{\beta}}{1 - 0.8/n} \text{ and } \hat{\alpha}_c^* = -\hat{\beta}_c \ln\left[\frac{1}{n}\sum_{i=1}^{n} \exp -\frac{X_i}{\hat{\beta}_c}\right] - 0.7\,\frac{\hat{\beta}_c}{n}.$$

Using a rather theoretical analysis, [14] made more accurate corrections leading to the following estimators:

$$\hat{\beta}_c = \hat{\beta}\left(1 + \frac{0.7716}{n}\right) \text{ and } \hat{\alpha}_c = -\hat{\beta}_c \ln\left[\frac{1}{n}\sum_{i=1}^{n} \exp -\frac{X_i}{\hat{\beta}_c}\right] - 0.3698\,\frac{\hat{\beta}_c}{n}.$$

It is to be noted that in instances when a non-negative random variable is needed, it is the discrimination between the lognormal and the Weibull distributions that might be of interest, but in such a case the results of the present study remain applicable because the normal and the lognormal (also the Gumbel and the Weibull distributions) are linked by a simple logarithmic transformation. The discrimination between lognormal and Weibull has been proposed by [17].

3. THE TEST STATISTIC AND ITS ASYMPTOTIC DISTRIBUTION

Assume that the random sample $X_1,...,X_n$ is known to come from either a normal distribution, $X \sim N(\mu, \sigma^2)$, or a Gumbel distribution, $X \sim G(\alpha, \beta)$. The log-likelihood ratio statistic, T, is defined as the logarithm of the ratio of two maximized likelihood functions:

$$T = \ln\left(\frac{L_N(\hat{\mu}, \hat{\sigma}^2)}{L_G(\hat{\alpha}, \hat{\beta})}\right)$$

where $L_N(\mu, \sigma^2)$ and $L_G(\alpha, \beta)$ the likelihood functions under a normal distribution and a Gumbel distribution, respectively. The decision rule for discriminating between the normal and the Gumbel distributions is to choose the normal if $T > 0$, and to reject the normal in favor of the Gumbel, otherwise. Because both of these two distributions are of the location scale type, one important property of the T statistic is that it is independent of the parameters from both distributions (see, [8]).

Let us look at the expressions of T in terms of the corresponding MLEs.

Note that

$$\begin{aligned}
T &= \ln L_N(\hat{\mu}, \hat{\sigma}^2) - \ln L_G(\hat{\alpha}, \hat{\beta}) \\
&= \left[-n \ln \hat{\sigma} - n \ln \sqrt{2\pi} - \frac{1}{2\hat{\sigma}^2} \sum_{i=1}^{n} (X_i - \hat{\mu})^2 \right] \\
&\quad - \left[-n \ln \hat{\beta} - \sum_{i=1}^{n} \left[\frac{X_i - \hat{\alpha}}{\hat{\beta}} + \exp - \frac{X_i - \hat{\alpha}}{\hat{\beta}} \right] \right] \\
&= -n \ln \hat{\sigma} - n \ln \sqrt{2\pi} - \frac{1}{2\hat{\sigma}^2} \sum_{i=1}^{n} (X_i - \hat{\mu})^2 + n \ln \hat{\beta} \\
&\quad + \frac{1}{\hat{\beta}} \sum_{i=1}^{n} X_i - \frac{n\hat{\alpha}}{\hat{\beta}} + \sum_{i=1}^{n} \exp - \frac{X_i}{\hat{\beta}} \, \exp \frac{\hat{\alpha}}{\hat{\beta}}. \qquad (3.1)
\end{aligned}$$

Using (2.2), we get

$$n \exp \frac{-\hat{\alpha}}{\hat{\beta}} = \sum_{i=1}^{n} \exp - \frac{X_i}{\hat{\beta}}. \qquad (3.2)$$

If we replace the MLE finding in the equations (2.1) and the last equation (3.2) in (3.1), we obtain

$$T = -n \ln \frac{\hat{\sigma}}{\hat{\beta}} + n \frac{\hat{\mu} - \hat{\alpha}}{\hat{\beta}} + \frac{n}{2}(1 - \ln 2\pi).$$

We denote T_c, the new test statistic which introduces a correction for bias of maximum likelihood estimators proposed by [14]. Therefore, T_c can be written as:

$$T_c = -n \ln \frac{\hat{\sigma}}{\hat{\beta}_c} + n \frac{\hat{\mu} - \hat{\alpha}_c}{\hat{\beta}_c} + \frac{n}{2}(1 - \ln 2\pi).$$

Note that T and T_c are asymptotically equivalent, then we state the following lemma:

Lemma 3.1. *The test statistics T and T_c have the same asymptotic distribution.*

Proof: We have $\dfrac{T_c}{n} = -\ln \dfrac{\hat{\sigma}}{\hat{\beta}_c} + \dfrac{\hat{\mu} - \hat{\alpha}_c}{\hat{\beta}_c} + \dfrac{1}{2}(1 - \ln 2\pi)$ and

$$-\ln \frac{\hat{\sigma}}{\hat{\beta}_c} = -\ln \frac{\hat{\sigma}}{\hat{\beta}} + \ln\left(1 + \frac{0.7716}{n}\right) = -\ln \frac{\hat{\sigma}}{\hat{\beta}} + o(1). \qquad (3.3)$$

In addition, $\hat{\beta}_c = \hat{\beta} + o(1)$ and $\hat{\alpha}_c = \hat{\alpha} + o_p(1)$ lead to

$$\frac{\hat{\mu} - \hat{\alpha}_c}{\hat{\beta}_c} = \frac{\hat{\mu} - \hat{\alpha}}{\hat{\beta}} + o_p(1). \qquad (3.4)$$

From (3.3) and (3.4) we obtain $\frac{T_c}{n} = \frac{T}{n} + o_p(1)$. Then $\frac{T_c}{n}$ and $\frac{T}{n}$ have the same limit distribution, thus for $\epsilon > 0$ and n sufficiently large, we have

$$\left| P\left[\frac{T_c}{n} < \frac{t}{n}\right] - P\left[\frac{T}{n} < \frac{t}{n}\right] \right| < \epsilon.$$

Immediately $|P[T_c < t] - P[T < t]| < \epsilon$, for $\epsilon > 0$ and n is sufficiently large. Finally, if $\lim_{n\to+\infty} P[T < t]$ exists, then $\lim_{n\to+\infty} P[T_c < t] = \lim_{n\to+\infty} P[T < t]$. □

3.1. Asymptotic distribution of T_c under the normal distribution

Suppose data are coming from a normal distribution $N(\mu, \sigma^2)$. Based on [18], the following theorem can be stated:

Theorem 3.1. *Assume that the sample $X_1,\ldots,X_n$ follows $N(\mu, \sigma^2)$, then the test statistic T_c is asymptotically normally distributed with mean $E_N(T)$ and variance $Var_N(T)$.*

Proof: The proof of this theorem is based on the Lemma 3.1, the following Lemma 3.2 and the Central Limit Theorem (CLT). □

Lemma 3.2. *Denote $\tilde{T} = \ln\left(\frac{L_N(\mu,\sigma^2)}{L_G(\tilde{\alpha},\tilde{\beta})}\right)$, where $\tilde{\alpha}$ and $\tilde{\beta}$ are given by the following equation and may depend on μ and σ,*

$$E_N[\ln f_G(X, \tilde{\alpha}, \tilde{\beta})] = \max_{\alpha,\beta} E_N[\ln f_G(X, \alpha, \beta)],$$

then $\hat{\alpha} \to \tilde{\alpha}$ a.s, $\hat{\beta} \to \tilde{\beta}$ a.s and $\frac{T - E_N(T)}{\sqrt{n}}$ is asymptotically equivalent to $\frac{\tilde{T} - E_N(\tilde{T})}{\sqrt{n}}$.

The proof of this lemma is similar to that of Theorem 1 presented by White in [21], then the proof of Theorem 3.1 is established by proving that $\frac{\tilde{T} - E_N(\tilde{T})}{\sqrt{n}}$ is asymptotically normal based on the central limit theorem. As for the needed quantities $\widetilde{\alpha}$ and $\widetilde{\beta}$ in Lemma 3.2, $E_N(T)$ and variance $Var_N(T)$ in Theorem 3.1, they are derived by first referring to Lemma 3.2 and performing the following calculation:

$$\begin{aligned} E_N[\ln f_G(X, \alpha, \beta)] &= -\ln\beta - E_N\left(\frac{X-\alpha}{\beta}\right) - E_N\left(\exp\left(-\frac{X-\alpha}{\beta}\right)\right) \\ &= -\ln\beta - \frac{\mu-\alpha}{\beta} - \exp\left(-\frac{\mu-\alpha}{\beta} + \frac{\sigma^2}{2\beta^2}\right). \end{aligned}$$

We maximize with respect to α and β, we get $\tilde{\alpha} = \mu - \frac{\sigma}{2}$ and $\tilde{\beta} = \sigma$. By the second point of Lemma 3.2, $E_N(T)$ and $Var_N(T)$ are calculated.

$$\begin{aligned}
E_N(T) &\simeq E_N\left[\ln\left(\frac{L_N(\mu,\sigma^2)}{L_G(\tilde{\alpha},\tilde{\beta})}\right)\right] \\
&= nE_N[\ln f_N(X,\mu,\sigma^2) - \ln f_G(X,\widetilde{\alpha},\widetilde{\beta})] \\
&= nE_N[\ln f_N(X,\mu,\sigma^2)] - nE_N[\ln f_G(X,\widetilde{\alpha},\widetilde{\beta})] \\
&= nE_N\left[-\ln\sigma - \ln\sqrt{2\pi} - \frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2\right] \\
&\quad -nE_N\left[-\ln\tilde{\beta} - \frac{X-\tilde{\alpha}}{\tilde{\beta}} - \exp-\frac{X-\tilde{\alpha}}{\tilde{\beta}}\right] \\
&= n\left(-\ln\sigma - \ln\sqrt{2\pi} - \frac{1}{2} - (-\ln\sigma - \frac{3}{2})\right) \\
&= n\left(1 - \ln\sqrt{2\pi}\right),
\end{aligned}$$

for n sufficiently large, we obtain

$$\lim_{n\to+\infty} \frac{E_N(T)}{n} = 0.081016.$$

In addition, $Var_N[\ln f_N(X,\mu,\sigma^2)] = Var_N\left[-\frac{1}{2\sigma^2}\left(X-\mu\right)^2\right] = \frac{1}{2}$ and taking into account that $e^{-\frac{1}{2}}\int z^2e^{-z}\phi(z)dz = 2$ and $e^{-\frac{1}{2}}\int ze^{-z}\phi(z)dz = -1$ where $\phi(.)$ is the standard normal probability density function, then we have

$$\begin{aligned}
Var_N\left[\ln f_G(X,\widetilde{\alpha},\widetilde{\beta})\right] &= Var_N\left[-\frac{X-\widetilde{\alpha}}{\widetilde{\beta}} - \exp-\frac{X-\tilde{\alpha}}{\tilde{\beta}}\right] \\
&= Var_N\left(\frac{X-\mu}{\sigma}\right) + Var_N\left[e^{-\frac{1}{2}}\exp\left(-\frac{X-\mu}{\sigma}\right)\right] \\
&\quad +2e^{-\frac{1}{2}}Cov_N\left[\frac{X-\mu}{\sigma}; \exp-\frac{X-\mu}{\sigma}\right] \\
&= e-2
\end{aligned}$$

and

$$\begin{aligned}
&Cov_N\left[\ln f_N(X,\mu,\sigma^2), \ln f_G(X,\tilde{\alpha},\tilde{\beta})\right] \\
&= \frac{1}{2}Cov_N\left[\left(\frac{X-\mu}{\sigma}\right)^2, \frac{X-\mu}{\sigma}\right] \\
&\quad +\frac{1}{2}e^{-\frac{1}{2}}Cov_N\left[\left(\frac{X-\mu}{\sigma}\right)^2, \exp\left(-\frac{X-\mu}{\sigma}\right)\right] \\
&= \frac{1}{2},
\end{aligned}$$

thus,

$$\begin{aligned}\frac{Var_N(T)}{n} &\simeq Var_N[\ln f_N(X,\mu,\sigma^2) + Var_N[\ln f_G(X,\tilde{\alpha},\tilde{\beta})] \\ &\quad -2Cov_N[\ln f_N(X,\mu,\sigma^2); \ln f_G(X,\tilde{\alpha},\tilde{\beta})] \\ &\simeq e - \frac{5}{2}.\end{aligned}$$

Then

$$\lim_{n\to+\infty} \frac{Var_N(T)}{n} = 0.218282.$$

Finally, $\lim\limits_{n\to+\infty} \frac{E_N(T)}{n}$ and $\lim\limits_{n\to+\infty} \frac{Var_N(T)}{n}$ are independent of μ and σ, then the asymptotic distribution of T is independent of μ and σ. Then from Theorem 3.1, the test statistic T_c is asymptotically normally distributed with mean $0.081016 \times n$ and variance $0.218282 \times n$.

3.2. Asymptotic distribution of T_c under the Gumbel distribution

Now we turn to the case where the sample comes from a Gumbel distribution $G(\alpha,\beta)$. As before, based on Kundu, Gupta, and Manglick [18], the following theorem can be stated:

Theorem 3.2. *We suppose that the sample X_1,...,X_n follows $G(\alpha,\beta)$, then the test statistic T_c is asymptotically normally distributed with mean $E_G(T)$ and variance $Var_G(T)$.*

Once again, the proof of this theorem is straightforward from the central limit theorem and the following lemma.

Lemma 3.3. *Denote $\tilde{T}' = \ln\left(\frac{L_N(\tilde{\mu},\tilde{\sigma}^2)}{L_G(\alpha,\beta)}\right)$, where $\tilde{\mu}$ and $\tilde{\sigma}$ are given by the following equation and may depend on α and β:*

$$E_N[\ln f_G(X,\tilde{\mu},\tilde{\sigma}^2)] = \max_{\mu,\sigma} E_N[\ln f_G(X,\mu,\sigma^2)]$$

then $\hat{\mu} \to \tilde{\mu}$ a.s, $\hat{\sigma} \to \tilde{\sigma}$ a.s and $\frac{T - E_G(T)}{\sqrt{n}}$ is asymptotically equivalent to $\frac{\tilde{T}' - E_G(\tilde{T}')}{\sqrt{n}}$.

It is now possible to evaluate $\widetilde{\mu}$ and $\widetilde{\sigma}$ by referring to Lemma 3.3 and performing the following calculation:

$$E_G[\ln f_N(X,\mu,\sigma^2)] = E_G\left[-\frac{1}{2}\ln 2\pi - \ln\sigma - \frac{(X-\mu)^2}{2\sigma^2}\right].$$

Since X follows $G(\alpha,\beta)$, it is immediate that $E_G(X)=\alpha+\beta\gamma$ and $Var_G(X)=\dfrac{\pi^2}{6}\beta^2$, where $\gamma\simeq 0.5772...$(the Euler constant). Therefore,

$$\begin{aligned}
E_G[\ln f_N(X,\mu,\sigma^2)] &= -\frac{1}{2}\ln 2\pi - \ln\sigma - \frac{1}{2\sigma^2}E_G(X^2-2\mu X+\mu^2)\\
&= -\frac{1}{2}\ln 2\pi - \ln\sigma\\
&\quad -\frac{1}{2\sigma^2}\left[\frac{\pi^2\beta^2}{6}+(\alpha+\beta\gamma)^2-2\mu(\alpha+\beta\gamma)+\mu^2\right].
\end{aligned}$$

Maximizing with respect to μ and σ yields $\tilde{\mu}=\alpha+\beta\gamma$ and $\tilde{\sigma}=\frac{\pi}{\sqrt{6}}\beta$. The quantities $E_G(T)$ and $Var_G(T)$ can be derived using again Lemma 3.3,

$$\begin{aligned}
E_G(T) &\simeq nE_G\left[\ln f_N(X,\tilde{\mu},\tilde{\sigma}^2)-\ln f_G(X,\alpha,\beta)\right]\\
&\simeq nE_G\left[-\ln\tilde{\sigma}-\ln\sqrt{2\pi}-\frac{1}{2}\left(\frac{X-\tilde{\mu}}{\tilde{\sigma}}\right)^2\right]\\
&\quad +nE_G\left[\ln\beta+\frac{X-\alpha}{\beta}+\exp-\frac{X-\alpha}{\beta}\right]\\
&\simeq nE_G\left[-\ln\frac{\pi\beta}{\sqrt{6}}-\frac{1}{2}\ln 2\pi-\frac{1}{2}\left(\frac{X-(\alpha+\beta\gamma)}{\frac{\pi\beta}{\sqrt{6}}}\right)^2\right]\\
&\quad +nE_G\left[\ln\beta+\frac{X-\alpha}{\beta}+\exp-\frac{X-\alpha}{\beta}\right]\\
&\simeq n\left(-\frac{3}{2}\ln\pi+\frac{1}{2}\ln 3\right)\\
&\quad +nE_G\left[-\frac{3}{\pi^2}\left(\frac{X-\alpha}{\beta}-\gamma\right)^2+\frac{X-\alpha}{\beta}+e^{-\dfrac{X-\alpha}{\beta}}\right]
\end{aligned}$$

we put $Z=\dfrac{X-\alpha}{\beta}$, then Z follows $G(0,1)$ and we obtain

$$\begin{aligned}
E_G(T) &\simeq -\frac{3n}{2}\ln\pi+\frac{n}{2}\ln 3+nE_G\left[-\frac{3}{\pi^2}(Z-\gamma)^2+Z+\exp-Z\right]\\
&\simeq -\frac{3n}{2}\ln\pi+\frac{n}{2}\ln 3-\frac{3n}{\pi^2}E_G[(Z-\gamma)^2]+E_G[Z]+E_G[\exp-Z]\\
&\simeq n\left(-\frac{3}{2}\ln\pi+\frac{1}{2}\ln 3-\frac{3}{\pi^2}\frac{\pi^2}{6}+\gamma+1\right)
\end{aligned}$$

for n sufficiently large, we obtain

$$\lim_{n\to+\infty}\frac{E_G(T)}{n}=-0.090573.$$

Similarly,

$$
\begin{aligned}
\frac{Var_G(T)}{n} &\simeq Var_G\left[\ln f_N(X,\tilde{\mu},\tilde{\sigma}^2) - \ln f_G(X,\alpha,\beta)\right] \\
&\simeq Var_G\left[-\ln\tilde{\sigma} - \ln\sqrt{2\pi} - \frac{1}{2}\left(\frac{X-\tilde{\mu}}{\tilde{\sigma}}\right)^2 + \ln\beta + \frac{X-\alpha}{\beta}\right. \\
&\quad \left. + \exp-\frac{X-\alpha}{\beta}\right] \\
&\simeq Var_G\left[-\frac{3}{\pi^2}\left(\frac{X-\alpha}{\beta} - \gamma\right)^2 + \frac{X-\alpha}{\beta} + \exp-\frac{X-\alpha}{\beta}\right] \\
&\simeq Var_G\left[-\frac{3}{\pi^2}(Z-\gamma)^2 + Z + \exp-Z\right],
\end{aligned}
$$

then

$$\lim_{n\to+\infty}\frac{Var_G(T)}{n} = 0.283408.$$

Since both $\lim\limits_{n\to+\infty}\frac{E_G(T)}{n}$ and $\lim\limits_{n\to+\infty}\frac{Var_G(T)}{n}$ do not depend on α and β, the asymptotic distribution of T is independent of α and β. Then from Theorem 3.2, the test statistic T_c is asymptotically normally distributed with mean $-0.090573\times n$ and variance $0.283408\times n$.

4. PCS AND MC SIMULATION

It is assumed that the data have been generated from one of the two distributions: $N(\mu,\sigma^2)$ or $G(\alpha,\beta)$. Then the discrimination procedure based on a random sample $X = X_1, ..., X_n$ is as follows.

Choose normal distribution if $T_c > 0$ and Gumbel distribution if $T_c < 0$. If the data were originally coming from $N(\mu,\sigma^2)$, the PCS_N can be written as follows: $PCS_N = P(T_c > 0|$ data follow a normal distribution). Similarly, if the data were originally coming from $G(\alpha,\beta)$, the PCS_N can be written as follows: $PCS_G = P(T_c < 0|$ data follow Gumbel distribution). Since for normal distribution

$$
\begin{aligned}
PCS_N = P[T_c > 0] &\simeq \Phi\left(\frac{E_N(T)}{\sqrt{Var_N(T)}}\right) \\
&= \Phi\left(\frac{0.081016\times n}{\sqrt{0.218282\times n}}\right) \\
&= \Phi(0.1734\sqrt{n})
\end{aligned}
$$

where Φ is the distribution function of the standard normal distribution. In the same manner, we have for Gumbel distribution

$$
\begin{aligned}
PCS_G = P[T_c < 0] &= 1 - P[T_c > 0] \\
&\simeq 1 - \Phi\left(\frac{E_G(T)}{\sqrt{Var_G(T)}}\right) \\
&= 1 - \Phi\left(\frac{-0.090573 \times n}{\sqrt{0.283408 \times n}}\right) \\
&= \Phi\left(\frac{0.090573 \times n}{\sqrt{0.283408 \times n}}\right) \\
&= \Phi(0.1701\sqrt{n}).
\end{aligned}
$$

We use Monte-Carlo simulations to examine how the asymptotic results work for small sizes. All computations are performed using the statistical freeware R [20]. We compute the PCS based on simulations and those based on the asymptotic normality results. Since the distribution of T_c is independent of the location and scale parameters, we take the location and scale parameters to be zero and one respectively in all cases. We consider different sample sizes, namely $n = 20$, 30, 40, 50, 60 and 100. First we consider the case when the data comes from normal distribution. In this case we generate a random sample of size n from $N(0,1)$, we compute T_c and check whether T_c is positive or negative. We replicate the process 10 000 times and obtain an estimate of PCS. Similarly, we obtain the results when the data comes from Gumbel distribution. The results are reported in Table 1.

Table 1: PCS's based on Monte Carlo simulations (MC) with 10 000 replications and those based on the asymptotic results (AR) when the data come from the normal (Gumbel) distribution respectively.

Sample size (n)	**MC**	**Asymptotic results**
10	0.62 (0.70)	0.71 (0.70)
20	0.75 (0.79)	0.78 (0.77)
30	0.81 (0.84)	0.84 (0.82)
40	0.86 (0.88)	0.86 (0.85)
50	0.90 (0.91)	0.89 (0.88)
60	0.91 (0.92)	0.91 (0.90)
70	0.93 (0.94)	0.93 (0.92)
80	0.94 (0.95)	0.94 (0.94)
90	0.95 (0.96)	0.95 (0.95)
100	0.96 (0.97)	0.96 (0.95)

The comparison between the MC simulation and the asymptotic results shows that the asymptotic approximation works quite well even for small samples. Results also reveal that it is easy to discriminate between normal and Gumbel distributions even for a small sample as 20. For example, the comparison of the results of Table 1 with those of Kundu and Manglick [17] shows that the selection between the normal and Gumbel distributions gives an asymptotic approximation more accurate even for a small sample size when the data comes from Gumbel distribution. Table 1 shows that the minimum sample size needed to choose between normal and Gumbel distributions is less than 50; it is also clear that the power of the test varies between 0.62 and 0.96 as the sample size varies between 10 and 100.

5. CONCLUSION

The normal and Gumbel distributions are often considered as competing models when the variable of interest takes values from $-\infty$ to $+\infty$. In this work we consider the statistic based on the RML and obtain asymptotic distributions of the test statistics under null hypothesis. Using MC simulations we compare the probability of correct selection with these asymptotic result and it is observed that even when the sample size is as small as 20, these asymptotic results work quite well for a wide range of the parameter space. Therefore, these asymptotic results can be used to estimate the PCS. Our method can be used for discriminating between any two members of the different location and scale families.

ACKNOWLEDGMENTS

The authors would like to thank the Reviewer for his valuable comments which had improved the earlier version of the paper and one Associate Editor for some very constructive suggestions.

REFERENCES

[1] Atkinson, A. (1969). A test for discriminating between models, *Biometrika*, **56**, 337–347.

[2] Atkinson, A. (1970). A method for discriminating between models (with discussions), *Journal of Royal Statistical Society, Series B*, **32**, 323–353.

[3] Bain, L.J. and Engelhardt, M. (1980). Probability of correct selection of Weibull versus Gamma based on likelihood ratio, *Statit.-Theor. Meth*, **9**, 375–381.

[4] Chambers, E.A. and Cox, D.R. (1967). Discriminating between alternative binary response models, *Biometrika*, **54**, 573–578.

[5] Chen, W.W. (1980). On the tests of separate families of hypotheses with small sample size, *Journal of Statistical Computations and Simulations*, **2**, 183–187.

[6] Cox, D.R. (1961). Tests of separate families of hypotheses, *Proceedings of the Fourth Berkeley Symposium in Mathematical Statistics and Probability*, Berkeley, University of California Press, 105–123.

[7] Cox, D.R. (1962). Further results on tests of separate families of hypotheses, *J. Roy. Statist. Soc. Ser. B*, **24**, 406–424.

[8] Dumonceaux, R.; Antle, C.E. and Haas, G. (1973). Likelihood ratio test for discriminating between two models with unknown location and scale parameters, *Technometrics*, **15**, 19–31.

[9] Dyer, A.R. (1973). Discriminating procedure for separate families of hypotheses, *Journal of American Statistics Association*, **68**, 970–974.

[10] Fearn, D.H. and Nebenzahl, E. (1991). On the maximum likelihood ratio method of deciding between the Weibull and gamma distribution, *Communications in Statistics — Theory and Methods*, **20**, 579–593.

[11] Fiorentino, M. and Gabriele, S. (1984). A correction for the bias of maximum-likelihood estimators of Gumbel parameters, *Journal of Hydrology*, **73**, 39–49.

[12] Gupta, R.D. and Kundu, D. (2003). Discriminating between the Weibull and generalized exponential distributions, *Computational Statistics and Data Analysis*, **43**, 179–196.

[13] Gupta, R.D. and Kundu, D. (2004). Discriminating between the gamma and generalized exponential distributions, *Journal of Statistical Computations and Simulations*, **74**(1), 107–121.

[14] Hosking, J.R.M. (1985). A correction for the bias of maximum-likelihood estimators of Gumbel parameters-comment, *Journal of Hydrology*, **78**, 393–396.

[15] Jackson, O.A.Y. (1968). Some results on tests of separate families of hypotheses, *Biometrika*, **55**, 355–363.

[16] Jackson, O.A.Y. (1969). Fitting a gamma or log-normal distributions to fibrediameter measurements of wool tops, *Applied Statistics*, **18**, 70–75.

[17] Kundu, D. and Manglick, A. (2004). Discriminating between the Weibull and lognormal distributions, *Naval Research Logistics*, **51**, 893–905.

[18] Kundu, D.; Gupta, R.D. and Manglick, A. (2005). Discriminating between the lognormal and generalized exponential distributions, *Journal of Statistical Planning and Inference*, **127**, 213–227.

[19] Quesenberry, C.P. and Kent, J. (1982). Selecting among probability distributions used in reliability, *Technometric*, **24**, 59–65.

[20] R Development Core Team. (2009). *R: A language and environment for statistical computing*, ISBN 3-900051-07-0, Vienna, Austria.

[21] White, H. (1982). Regularity conditions for Cox's test of non-nested hypotheses, *Journal of Econometrics*, **19**, 301–318.

[22] Wiens, B.L. (1999). When lognormal and gamma models give different results: A case study, *American Statistics*, **53**, 89–93.

REVSTAT – Statistical Journal
Volume 15, Number 4, October 2017, 537–563

CONFIDENCE INTERVALS FOR EXCEEDANCE PROBABILITIES WITH APPLICATION TO EXTREME SHIP MOTIONS

Authors: Dylan Glotzer, Vladas Pipiras
– Dept. of Statistics and Operations Research, UNC at Chapel Hill, CB#3260, Hanes Hall, Chapel Hill, NC 27599, USA
dglotzer@email.unc.edu
pipiras@email.unc.edu

Vadim Belenky, Bradley Campbell, Timothy Smith
– Naval Surface Warfare Center Carderock Division (NSWCCD), 9500 MacArthur Blvd, W. Bethesda, MD 20817, USA
vadim.belenky@navy.mil
bradley.campbell@navy.mil
timothy.c.smith1@navy.mil

Received: May 2015 Revised: January 2016 Accepted: February 2016

Abstract:

- Statistical inference of a probability of exceeding a large critical value is studied in the peaks-over-threshold (POT) approach. The focus is on assessing the performance of the various confidence intervals for the exceedance probability, both for the generalized Pareto distribution used above a selected threshold and in the POT setting for general distributions. The developed confidence intervals perform well in an application to extreme ship motion data. Finally, several approaches to uncertainty reduction are also considered.

Key-Words:

- *exceedance probability; quantiles; confidence intervals; peaks over threshold; generalized Pareto distribution; threshold selection; uncertainty reduction.*

AMS Subject Classification:

- 62G32, 62P30, 62F25, 62G15.

1. INTRODUCTION

We describe first the real-life application which sets the directions and frames the questions pursued in this work (Section 1.1). We then outline the contributions and the structure of this work (Section 1.2).

1.1. Motivation

This work is motivated by applications to ship motions and, more specifically, their stability in irregular seas. See, for example, Lewis [25], Benford [7], Belenky and Sevastianov [4], Neves *et al.* [28] for more information on this research area. When it comes to ship stability, the focus is on several variables characterizing the ship motion including *roll* and *pitch* angles, which are, respectively, the rotational movements around longitudinal (stern-to-bow) and lateral (starboard-to-port side or right-to-left side) axes, as well as vertical and lateral *accelerations* in various locations on the ship. See Figure 1. The ship stability failures are related directly to the exceedance of certain values by these variables. For example, the exceedance of a certain roll angle can lead to a cargo shift (which then can lead to capsizing), loss or damage of cargo in containers on deck, or down-flooding internal volumes of a ship. A large enough acceleration can lead to serious injuries or even death of a crew and passengers, as well as cargo damage. Such occurrences are well known not only among the researchers working in the area but also often make it to the popular media.[1]

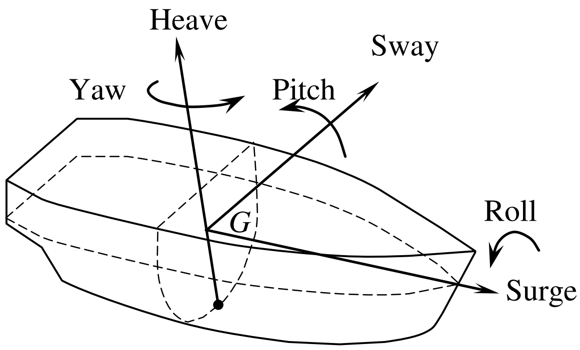


Figure 1: The motions of the ship.

[1]Recent examples of accidents related to intact stability failures include: Ro/Ro Ferry *Aratere* on 3rd March 2006 (Maritime New Zealand, 2007), Cruise ship *Pacific Sun* on 30 July 2008 (Marine Accident Investigation Branch, 2009), Ferry *Ariake* on 13 November, 2009 (Transportation Safety Board, 2011), to name but a few.

The measured variables of interest to stability are understandably affected by the *geometry* and *loading* of the ship, the *operational parameters* and the surrounding *sea*. The operational side includes the *heading* (the angle between the vector of ship speed and predominant direction of wave propagation) and the value of *speed* of the ship. The state of the sea is usually described by a spectrum of wave elevations. Note that a wide range of *conditions* (the values of the above descriptors) are possible. What can be expected under a particular condition is often suggested from the understanding of the dynamics governing the ship motion.

An appealing but also critical feature of the research area is the availability of computer programs simulating ship motions, see the recent state-of-the-art review by Reed *et al.* [30]. In this work, we use a fast volume-based ship motion simulation algorithm developed in Weems and Wundrow [37]. The developed code does not incorporate finer hydrodynamics features of ship motions such as the influence of a ship motion on wave pressure field (i.e. wave diffraction and radiation; cf. Large Amplitude Motion Program or LAMP, see Lin and Yue [26]). But it is considered qualitatively representative of ship motions and their extremes. Moreover, the code is fast enough (in fact, the only such realistic method available) to be used in validation, where very long time histories of ship motions are necessary (see Section 3 below).

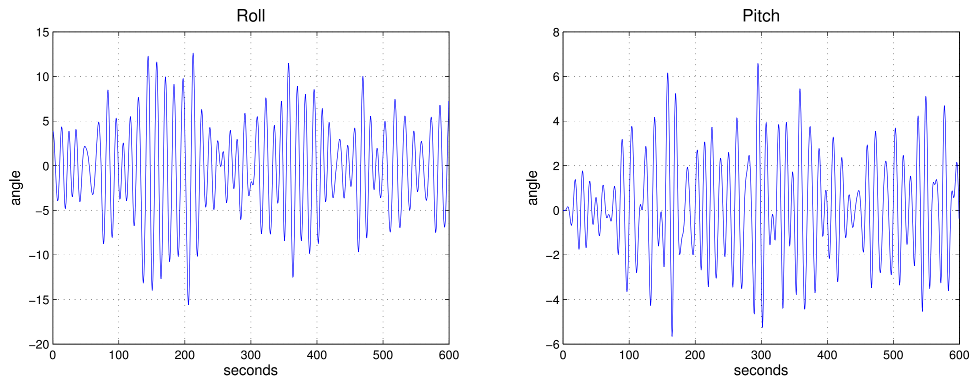


Figure 2: The roll and pitch angle series for 10 minutes.

Figure 2 depicts the time series of roll and pitch angles obtained by the above referenced code for a 10 minute time window at 0.5 second measurement intervals. The ship geometry is that of the ONR tumblehome top (Bishop *et al.* [8]). The heading is at 45 degrees, the speed is 6 knots, the waves are characterized by significant height of 9m and mean zero-crossing period of 10.65s which corresponds to 15s of the modal period, using Bretschneider spectrum in open ocean (Lewis [25]).

A basic problem is to estimate the probability of roll, pitch or other variable of interest exceeding a critical value, as well as to provide a confidence interval. For example, in the condition of Figure 2, one could be interested in the roll angle exceeding 60 degrees (in either positive or negative direction). Inference would have to be made from the roll series of, for example, 100 hours, which would typically not contain such extreme occurrences. Again, the critical angle is often suggested from real-life considerations.

A method suggested for the problem above (and, more specifically, the associated confidence intervals) can be assessed through a validation procedure. The computer code mentioned above can be used to generate millions of hours of ship motion data which would contain exceedances of the target of interest. The "true" exceedance probability can then be estimated directly from this long history of the ship motion. In the validation procedure, the performance of the suggested method could be checked against the "true" exceedance probability at hand. See Section 3 for further details and a solution to the estimation problem.

1.2. Description of work and contributions

A natural mathematical framework to address the problem of estimating exceedance probabilities described above is the peaks-over-threshold (POT) approach (see, for example, Embrechts *et al.* [18], Coles [9], Beirlant *et al.* [2], as well as de Carvalho *et al.* [11], Ferreira and de Haan [19] for more recent related work). According to this approach, the probability of exceeding a given target of interest is computed as the product of the probability of exceeding a smaller threshold and the (conditional) probability of exceeding the target above the threshold. The former probability is estimated simply as the proportion of data above the chosen threshold. The peaks over the threshold are modeled using the *generalized Pareto distribution* (GPD), whose complementary distribution function has the form

(1.1)

$$\overline{F}_{\mu,\xi,\sigma}(x) := \Big(1+\frac{\xi(x-\mu)}{\sigma}\Big)_+^{-1/\xi} := \begin{cases} \Big(1+\frac{\xi(x-\mu)}{\sigma}\Big)^{-1/\xi}, & \mu<x, & \text{if } \xi>0, \\ e^{-\frac{x-\mu}{\sigma}}, & \mu<x, & \text{if } \xi=0, \\ \Big(1+\frac{\xi(x-\mu)}{\sigma}\Big)^{-1/\xi}, & \mu<x<\mu-\frac{\sigma}{\xi}, & \text{if } \xi<0, \end{cases}$$

where ξ is the *shape* parameter, σ is the *scale* parameter and μ is a *threshold.* Note that the GPD has an upper bound $(-\sigma/\xi)$ (above the threshold) for a negative shape parameter $\xi<0$. When $\xi=0$, the GPD is the usual exponential distribution.

We are interested here in what confidence intervals should be used for an exceedance probability. As indicated above, in the POT approach, this exceedance probability is the product of two probabilities, one of them being the exceedance probability for GPD. The questions then is what confidence intervals should be used for the exceedance probability in the GPD framework. The probability of the GPD exceeding a fixed *target* c (above the threshold), and its estimator are given by:

$$p_c = p_c(\xi, \sigma) = \Big(1 + \frac{\xi c}{\sigma}\Big)^{-1/\xi}, \quad \widehat{p}_c = p_c(\widehat{\xi}, \widehat{\sigma}) = \Big(1 + \frac{\widehat{\xi} c}{\widehat{\sigma}}\Big)^{-1/\widehat{\xi}}, \tag{1.2}$$

where $\widehat{\xi}$ and $\widehat{\sigma}$ are some estimators of the shape and scale parameters, respectively. Somewhat surprisingly, the question of confidence intervals for the exceedance probability in (1.2) has apparently not been considered in much depth in the literature on extreme values. The paper by Smith [33], which laid the mathematical foundations for the ML estimators of the GPD, considers the problem of estimating the exceedance probability and provides the asymptotic normality result for the probability estimator (Section 8 of Smith [33]). This can in turn be used for confidence intervals but the normality assumption is not particularly appropriate (see Section 2 below).

Estimation of exceedance probabilities has also been considered by others but with different goals in mind. For example, Smith and Shively [34] are interested in trends for exceedance probabilities. Exceedance probabilities in the spatial context appear in Draghicescu and Ignaccolo [16]. Considerable interest in exceedance (also sometimes referred to as failure) probabilities is when working with multivariate extremes. See, for example, de Haan and Sinha [13], de Haan and de Ronde [12], Heffernan and Tawn [20], Drees and de Haan [17].

Much of the focus in the extreme value analysis, on the other hand, has been on the related inverse problem of quantile estimation (see, for example, Embrechts *et al.* [18], Coles [9], Beirlant *et al.* [2]). The quantiles have been of greater practical interest in many applications driving the extreme value analysis, including finance (VaR calculations), insurance and hydrology (1-in-T years event). A closer look at the confidence intervals for quantiles can be found in Hosking and Wallis [22], Tajvidi [36] and also Section 4.3.3 of Coles [9], Section 5.5 of Beirlant *et al.* [2].

In applications to ship motions, as indicated in Section 1.1, it is common to look at the probabilities of exceeding a particular target rather than quantiles. Though perhaps not surprisingly, the two perspectives are also complementary. In fact, one of our findings is that the confidence intervals for exceedance probabilities perform well if constructed from those for quantiles. Another reason to focus on probabilities rather than quantiles is that probabilities can be aggregated naturally into "lifetime" probabilities, when integrated over a set of conditions of interest (as discussed, for example, in Section 1 of Belenky and Sevastianov [4]).

We study a number of ways to construct confidence intervals for the exceedance probability of the GPD and, more generally, in the POT framework in Section 2. We consider both direct methods, which are based on the functional form of exceedance probability (1.2) and the sampling distribution of the underlying estimators $\widehat{\xi}, \widehat{\sigma}$, and indirect (inverse) methods, which construct confidence intervals from those for quantiles.

The application of the considered confidence intervals to ship motions can be found in Section 3. In the validation procedure, the performance of the confidence intervals is analogous to that found under the idealized GPD framework. In particular, the methods recommended under the GPD framework also perform well and best in the application to ship motions. It should also be noted that the proposed solution is the first to address satisfactorily the estimation problem of the exceedance probabilities in ship stability. Some earlier attempts include Belenky and Campbell [5] who used the Weibull distribution (instead of the GPD) to fit peaks over threshold, and McTaggart [27].

Finally, in Section 4, we discuss the issue of uncertainty (the size of confidence intervals) and its reduction. Conclusions can be found in Section 5.

2. CONFIDENCE INTERVALS FOR EXCEEDANCE PROBABILITIES

2.1. Methods for GPD

We study and assess here several ways to construct confidence intervals for the exceedance probability p_c of the GPD given in (1.2). The probability is estimated through $\widehat{p}_c$ in (1.2) where we use the ML estimators $\widehat{\xi}$ and $\widehat{\sigma}$ computed from the sample $y_1, ..., y_n$ of size n. The large sample asymptotics of the ML estimators (Smith [33]) is the bivariate normal,

$$\sqrt{n}\begin{pmatrix} \widehat{\xi} - \xi_0 \\ \widehat{\sigma} - \sigma_0 \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, W^{-1}), \tag{2.1}$$

where ξ_0, σ_0 denote the true values and

$$W^{-1} = \begin{pmatrix} 1+\xi_0 & -\sigma_0 \\ -\sigma_0 & 2\sigma_0^2 \end{pmatrix}. \tag{2.2}$$

In practice, the limiting covariance matrix can be estimated by replacing ξ_0 and σ_0 with their respective estimators $\widehat{\xi}$ and $\widehat{\sigma}$. Another common choice is to approximate nW through the observed information matrix

$$n\widehat{W} = \begin{pmatrix} -\frac{\partial^2}{\partial \xi^2} l(\xi,\sigma) & -\frac{\partial^2}{\partial \xi \partial \sigma} l(\xi,\sigma) \\ -\frac{\partial^2}{\partial \xi \partial \sigma} l(\xi,\sigma) & -\frac{\partial^2}{\partial \sigma^2} l(\xi,\sigma) \end{pmatrix}_{(\xi,\sigma)=(\widehat{\xi},\widehat{\sigma})}, \tag{2.3}$$

where $l(\xi,\sigma) = \sum_{i=1}^{n} \ln f_{\xi,\sigma}(y_i)$ is the log-likelihood and $f_{\xi,\sigma}(y)$ denotes the density of the GPD. Strictly speaking, the asymptotic result (2.1) holds for $\xi > -1/2$ only (Smith [33]). It should also be noted that other estimation methods than the MLE are possible for ξ_0 and σ_0. See, for example, a review paper by de Zea Bermudez and Kotz [14, 15] and references therein. Some of these estimators outperform the ML estimators for small samples. For the sample sizes relevant to our problem of interest, the ML estimators seem to perform quite well and, in particular, to be approximately normal as stated in (2.1), and will be used throughout this work.

The exceedance probability $p_c = p_c(\xi,\sigma)$ in (1.2) is a function of ξ and σ, and is estimated through (1.2) by replacing the two parameters ξ and σ by their ML estimates. A confidence interval for p_c can then naturally be obtained through the standard delta method, using the asymptotic result (2.1). This is the approach seemingly adopted by Smith [33], Section 8. However, we found the delta method to perform poorly, in part because p_c can be very small and the normal approximation of $\widehat{p}_c$ may be sufficiently wide to include negative values. We have also tried the delta method for $\log p_c$ but the normal approximation did not appear to provide a good fit to $\log \widehat{p}_c$. Consequently, we consider below several, potentially more accurate ways to construct confidence intervals for the exceedance probabilities: the normal and lognormal methods, the boundary method, the bootstrap method, the profile likelihood method and the quantile method. The terminology behind the normal, lognormal, boundary and quantile methods are ours.

Normal method: The idea behind the normal method is still to use (2.1), which as mentioned earlier provides a good approximation in practice, but not to linearize the function $p_c(\xi,\sigma)$ (or $\log p_c(\xi,\sigma)$) as in the unsatisfactory delta method. In fact, assuming the bivariate normal approximation for $\widehat{\xi}$ and $\widehat{\sigma}$ according to (2.1), we can derive the exact distribution of $\widehat{p}_c$ as follows. Observe that the distribution function of $\widehat{p}_c$ is: for $0 \le z \le 1$,

$$\begin{aligned} F_{\widehat{p}_c}(z) &= P\Big(\Big(1+\frac{\widehat{\xi}c}{\widehat{\sigma}}\Big)^{-1/\widehat{\xi}} \le z\Big) \\ &= P\Big(\Big(1+\frac{\widehat{\xi}c}{\widehat{\sigma}}\Big)^{-1/\widehat{\xi}} \le z, 1+\frac{\widehat{\xi}c}{\widehat{\sigma}} > 0\Big) + P\Big(1+\frac{\widehat{\xi}c}{\widehat{\sigma}} \le 0\Big), \end{aligned}$$

where we use the fact that $\widehat{p}_c = 0$ if $1+\widehat{\xi}c/\widehat{\sigma} \le 0$. This can further be expressed as

$$F_{\widehat{p}_c}(z) = P\Big(\widehat{\sigma} \le \frac{\widehat{\xi}c}{z^{-\widehat{\xi}}-1}, \widehat{\sigma} > -\widehat{\xi}c\Big) + P(\widehat{\sigma} \le -\widehat{\xi}c),$$

if we assume that $\widehat{\sigma}$ takes only positive values. (Note also that $\widehat{\xi}/(z^{-\widehat{\xi}}-1) > 0$ for both $\widehat{\xi} < 0$ and $\widehat{\xi} > 0$.) Note, however, that it is not possible to have $\widehat{\sigma} > \widehat{\xi}c/(z^{-\xi}-1)$ and $\widehat{\sigma} \le -\widehat{\xi}c$. Indeed, this is certainly not possible if $\widehat{\xi} > 0$, since then $-\widehat{\xi}c < 0$ and $\widehat{\xi}c/(z^{-\widehat{\xi}}-1) > 0$. If $\widehat{\xi} < 0$, on the other hand, this is not possible since $-\widehat{\xi}c \le \widehat{\xi}c/(z^{-\widehat{\xi}} - 1)$ or, equivalently, $z^{-\widehat{\xi}} < 1$. Hence, we also have

$$F_{\widehat{p}_c}(z) = P\Big(\widehat{\sigma} \le \frac{\widehat{\xi}c}{z^{-\widehat{\xi}} - 1}\Big) = \int_{\sigma \le \xi c/(z^{-\xi}-1)} g_{\widehat{\xi},\widehat{\sigma}}(\xi,\sigma)d\xi d\sigma, \tag{2.4}$$

where $g_{\widehat{\xi},\widehat{\sigma}}(\xi,\sigma)$ denotes the bivariate normal density of the limit law (2.1) (replacing ξ_0 and σ_0 by $\widehat{\xi}$ and $\widehat{\sigma}$). In practice, the distribution function $F_{\widehat{p}_c}(z)$ is computed numerically and the $100(1-\alpha)\%$ confidence interval is set as (z_1, z_2) where $z_j = \inf\{z : F_{\widehat{p}_c}(z) \ge \alpha_j\}$, $j = 1, 2$, where $\alpha_1 = \alpha/2$ and $\alpha_2 = 1 - \alpha/2$. We use the generalized inverse in the last expression since $F_{\widehat{p}_c}(z)$ can have a discontinuity (mass) at $z = 0$.

Lognormal method: In the normal method above, we assumed that $\widehat{\sigma}$ does not take negative values or that, from a practical perspective, the probability of $\widehat{\sigma}$ being negative according to (2.1) is negligible. This may not be the case for smaller values of σ and sample sizes n. A natural way to address this is by parameterizing the GPD through ξ and $\ln\sigma$, instead of σ. The difference is that $\ln\sigma$ now takes possibly negative values. The asymptotic normality result then becomes

$$\sqrt{n}\begin{pmatrix} \widehat{\xi} - \xi_0 \\ \widehat{\ln\sigma} - \ln\sigma_0 \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, W_1^{-1}), \tag{2.5}$$

where

$$W_1^{-1} = \mathrm{diag}\{1, \sigma_0^{-1}\} W^{-1} \mathrm{diag}\{1, \sigma_0^{-1}\}. \tag{2.6}$$

Arguing as in the normal method above, we have

$$F_{\widehat{p}_c}(z) = P\Big(\widehat{\ln\sigma} \le \ln\frac{\widehat{\xi}c}{z^{-\widehat{\xi}} - 1}\Big) = \int_{\ln\sigma \le \ln(\xi c/(z^{-\xi}-1))} g_{\widehat{\xi},\widehat{\ln\sigma}}(\xi,\ln\sigma)d\xi d\ln\sigma, \tag{2.7}$$

where $g_{\widehat{\xi},\widehat{\ln\sigma}}(\xi,\ln\sigma)$ denotes the bivariate normal density of the limit law (2.5). The confidence interval can then be computed as in the normal method above. We shall refer to this as the lognormal method. A nice feature of the normal and lognormal methods is that they provide confidence intervals even in the case when $\widehat{\xi} < 0$ and the target is beyond the estimated support bound $(-\widehat{\sigma}/\widehat{\xi})$.

Boundary method: The normal and lognormal methods described above involve a relatively intensive numerical computation of the integrals (2.4) and (2.7). An approximate confidence interval which is fast to compute and easy to implement, can be constructed through the following boundary method. That is, take the confidence interval as

$$\Big(\min_{j,k=1,2} p_c(\xi_j,\sigma_k),\ \max_{j,k=1,2} p_c(\xi_j,\sigma_k)\Big), \tag{2.8}$$

where ξ_1, ξ_2 and σ_1, σ_2 are suitable critical values of the distributions of $\widehat{\xi}$ and $\widehat{\sigma}$, respectively. If $\widehat{\xi}$ and $\widehat{\sigma}$ were asymptotically uncorrelated, it would be natural to consider $\xi_j = \widehat{\xi} \pm C_{\sqrt{\alpha}}\mathrm{se}_{\widehat{\xi}}$ and $\sigma_k = \widehat{\sigma} \pm C_{\sqrt{\alpha}}\mathrm{se}_{\widehat{\sigma}}$, where se stands for standard error, C_β denotes the $100(\beta/2)\%$ quantile of the standard normal distribution and $(1-\alpha)\%$ is the confidence level sought. To account for the correlation between $\widehat{\xi}$ and $\widehat{\sigma}$, we take

$$\begin{pmatrix} \xi_j \\ \sigma_k \end{pmatrix} = V \begin{pmatrix} \xi_{0,j} - \widehat{\xi} \\ \sigma_{0,k} - \widehat{\sigma} \end{pmatrix} + \begin{pmatrix} \widehat{\xi} \\ \widehat{\sigma} \end{pmatrix}, \tag{2.9}$$

where $n^{-1}W^{-1} = VDV'$ with a diagonal $D = \mathrm{diag}\{d_1, d_2\}$ and $\xi_{0,j} = \widehat{\xi} \pm C_{\sqrt{\alpha}}\sqrt{d_1}$ and $\sigma_{0,k} = \widehat{\sigma} \pm C_{\sqrt{\alpha}}\sqrt{d_2}$. Note that the confidence intervals obtained by the boundary method are expected to be conservative. Indeed, the region determined by the points (ξ_j, σ_k) can be thought as the $100(1-\alpha)\%$ confidence region for the parameters ξ_0 and σ_0. But since $p_c(\xi, \sigma)$ is not a one-to-one function, there are points (ξ, σ) outside the confidence region for which the value $p_c(\xi, \sigma)$ falls inside the confidence interval (2.8).

Bootstrap method: The bootstrap method is somewhat standard with the confidence interval determined by the $100(\alpha/2)\%$ and $100(1-\alpha/2)\%$ quantiles of the bootstrap distribution of the exceedance probability.

Profile (likelihood) method: The profile (likelihood) method refers to another standard method to construct confidence intervals based on the profile likelihood. This is achieved by first expressing σ as a function of ξ and the exceedance probability p_c,

$$\sigma = \frac{\xi c}{p_c^{-\xi} - 1},$$

then parameterizing the likelihood in terms of ξ and p_c (instead of σ), and finally constructing the confidence interval based on the profile likelihood in a standard way. (See Coles [9] for the same approach when estimating a return level, instead of an exceedance probability.) Since the exceedance probability is constrained to be nonnegative, the use of the profile likelihood may be questionable.

Quantile method: Finally, the quantile method actually refers to a set of methods. The basic idea is the following. Exceedance probabilities p (p_c above) are associated with respective return levels (quantiles) x_p (c above) of the GPD distribution. A return level x_p can be estimated with a confidence interval $\widehat{x}_p \pm m_p$. Any of the methods discussed above (normal, lognormal, boundary, bootstrap and profile) can be adapted to construct a confidence interval for x_p – the difference being that the function (1.2) is now the return level

$$x_p = x_p(\xi, \sigma) = \frac{\sigma}{\xi}\Big(p^{-\xi} - 1\Big). \tag{2.10}$$

Moreover, the plot of $(-\ln p, \widehat{x}_p)$ with added confidence intervals is known as a return level plot (e.g. Coles [9]). To indicate the underlying method used to set confidence intervals for return levels, we will refer to the quantile method as quantile-boundary, quantile-lognormal, etc. A natural way to set a confidence interval for the exceedance probability p_c of the level c is then

$$(p_1, p_2), \tag{2.11}$$

where $p_1 = \inf\{p : \widehat{x}_p + m_p \geq c\}$ and $p_2 = \inf\{p : \widehat{x}_p - m_p \geq c\}$ (with $\inf\{\emptyset\} = 0$). See Figure 3. For the parameter values considered below, the functions $\widehat{x}_p + m_p$ and $\widehat{x}_p - m_p$ are increasing and continuous in the argument $(-\ln p)$. The quantile approach is appealing in that it makes estimation of exceedance probabilities and return levels consistent.

In the reliability context and for a location-scale family of distributions, the quantile approach was studied in Hong *et al.* [21] (see also Section I-C therein for earlier uses of connections between confidence intervals for quantiles and exceedance probabilities).

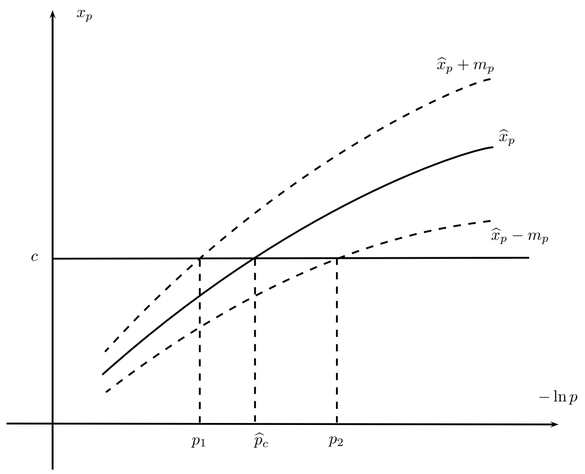


Figure 3: The quantile method to set confidence intervals for exceedance probability.

2.2. Simulation study for GPD

We examine here the confidence intervals proposed in Section 2.1 through a simulation study. The empirical coverage frequencies of the confidence intervals (based on 500 Monte Carlo replications) are reported in Tables 1 and 2 for the sample sizes $n = 100$ and $n = 50$, respectively. The sample size of approximately $n = 100$ is a typical value that we encounter in the application to ship motions described in Section 3 below. The results are also presented for the smaller sample size $n = 50$, since in practice, one does not expect many peaks over a threshold for which the GPD is used as a model.

The first four columns in the tables present the true values of the parameters ξ_0, σ_0, and also the target c and the corresponding exceedance probability p_c. The values of $\xi_0 = \pm.1$ are some of the typical values encountered in our application of interest. When $\xi_0 = .6$, the GPD has infinite variance but finite mean. σ_0 is just a scale parameter, which we fix at 1. For the other two true parameters, we fix the exceedance probability p_c and compute the respective target c.

The other columns of the tables correspond to the methods considered. The normal, lognormal and boundary methods use the limiting covariance matrix W^{-1}/n in (2.1). It is approximated by the inverse of the observed information matrix (2.3), which we found to yield better results than using, for example, the expression (2.2) (with ξ_0, σ_0 replaced by $\widehat{\xi}, \widehat{\sigma}$). The bootstrap method is based on 500 bootstrap replications. Finally, for the quantile methods, we consider three ways to construct confidence intervals for the return levels: lognormal, boundary and profile.

A number of observations can be drawn from Tables 1 and 2. The normal and lognormal methods are slightly anti-conservative, with the lognormal method preferred. The reason for the methods being anti-conservative is the estimation of the limiting covariance matrix W^{-1}/n in (2.1). The intervals have the expected coverage probability if the true covariance matrix (2.2) is taken (the exact coverage probabilities not reported here). As claimed in Section 2.1, the boundary method yields slightly conservative confidence intervals. The bootstrap and profile methods do not work well, especially for the value of ξ_0 close to zero or negative. Again, we suspect that this is due to the fact that the probability cannot be negative. Issues with bootstrap for the GPD were also reported and studied in Tajvidi [36].

Turning to the quantile methods, the quantile-lognormal method is slightly anti-conservative, as is the direct lognormal method. The quantile-boundary method is, on the other hand, slightly conservative. The quantile-profile method seems to perform best, with the coverage probabilities consistently close to the nominal level. Note that the profile-likelihood method for return levels does not have such pronounced limitation of the same method for exceedance probabilities – although it is true that a return level cannot be negative, the confidence interval would rarely reach zero. Note also that the results for $n = 100$ and $n = 50$ are comparable. One notable difference is that the quantile methods become slightly more anti-conservative when the sample size is reduced from $n = 100$ to $n = 50$.

In conclusion, the quantile method based on profile likelihood seems to perform best among the methods considered. The (log)normal and boundary methods, for both direct and indirect (quantile) approaches, can also be recommended but keeping in mind their (anti)conservative nature. Finally, we also note that the direct (log)normal and boundary methods are computationally less intensive compared to the indirect (quantile) methods.

Table 1: Coverage frequencies for confidence intervals when $n = 100$.

true values				direct methods					quantile methods		
ξ_0	σ_0	c	p_c	norm	logn	bound	boot	profl	logn	bound	profl
$-.1$	1	6.02	10^{-4}	90.4	90.8	96.2	68.2	76.2	92	97	95
		6.84	10^{-5}	95.2	95.6	97.6	65.6	78.8	88.8	96.6	94.6
		7.49	10^{-6}	94	94.6	96	74.6	80.8	91.2	96.8	94.2
.1	1	15.12	10^{-4}	92.6	93.2	98	87	98.4	89.6	97	94.2
		21.62	10^{-5}	90.6	91.2	97.2	82.8	97.6	92.4	98.6	95.2
		29.81	10^{-6}	91.2	92.6	97.6	81.2	97.8	91.2	98.4	94.6
.3	1	49.5	10^{-4}	91.8	92.4	98	89	97.2	89.4	97.4	92.2
		102.08	10^{-5}	88.8	89.2	98.4	86.6	97.6	92.2	97.8	95
		206.99	10^{-6}	93.4	94.2	99	91.8	98.6	92.6	98.4	94.4
.6	1	416.98	10^{-4}	90.8	90.8	97.8	91	93.8	92.4	98.6	94.2
		1665	10^{-5}	92.8	93.2	98	92.6	95.6	92.6	98.4	94
		6633.45	10^{-6}	94	93.8	98.6	92	95.8	93.4	99.4	95.2

Table 2: Coverage frequencies for confidence intervals when $n = 50$.

true values				direct methods					quantile methods		
ξ_0	σ_0	c	p_c	norm	logn	bound	boot	profl	logn	bound	profl
$-.1$	1	6.02	10^{-4}	90.2	91.8	96.2	65.8	78	83.6	94	92.8
		6.84	10^{-5}	96	96	94.4	69.8	71.8	89.6	95.4	93.4
		7.49	10^{-6}	95.4	95.8	93.8	74.4	75.4	88.2	94.4	92.2
.1	1	15.12	10^{-4}	91.4	92.2	97.6	75.6	98.4	89.9	97.4	92.6
		21.62	10^{-5}	87.6	89	96.6	67	98.2	88.4	96.8	94
		29.81	10^{-6}	90.8	92.2	96.2	70	98.2	88.8	96.4	93
.3	1	49.5	10^{-4}	88.2	91	96	88	99	90	97.2	95.8
		102.08	10^{-5}	85.8	87.8	94.4	82.6	98.6	90.2	96.8	93.4
		206.99	10^{-6}	88.4	90	96.8	85.4	98.2	90.8	96.6	93.8
.6	1	416.98	10^{-4}	80.4	90.6	98.2	88	96.4	89.8	98.4	93.4
		1665	10^{-5}	80.8	91.4	97	89.8	98.2	89.4	96.8	93
		6633.45	10^{-6}	83.4	90.2	97.8	89.4	98	92.6	97.4	92.8

2.3. The POT framework

Suppose now that $x_1, ..., x_N$ are i.i.d. observations of a general (i.e. non-GPD) random variable X, and that we are interested in estimating the probability $P(X > x_{cr})$ of the variable X exceeding a critical value x_{cr}. Again, in the peaks-over-threshold (POT) approach, the probability is written as

$$\begin{aligned} P(X > x_{cr}) &= P(X > u)P(X > x_{cr}|X > u) \\ &= P(X > u)P(X - u > x_{cr} - u|X > u) =: P_{nr} \cdot P_r, \end{aligned} \tag{2.12}$$

where u stands for an intermediate threshold, and the subscripts nr and r refer to the non-rare and rare problems, respectively. The non-rare probability is estimated directly from the data as the proportion of data above the threshold u, $\widehat{P}_{nr} = \sum_{j=1}^{N} 1_{\{x_j > u\}}/N$, with the respective confidence interval based on standard binomial calculations. The rare probability is estimated supposing that the peaks over threshold $Y = X - u$ follow a GPD, and setting

$$\widehat{P}_r = \widehat{p}_{x_{cr}-u},$$

where $\widehat{p}_c$ is the exceedance probability (1.2) in the GPD framework, estimated from the data $y_i = x_{i'} - u$ of the peaks exceeding the threshold. The confidence intervals for $P_r = p_{x_{cr}-u}$ are constructed by one of the methods of Section 2.1. The confidence interval for the original exceedance probability $P(X > x_{cr})$ is obtained by multiplying the respective endpoints of the confidence intervals of P_{nr} and P_r.

Threshold selection has been discussed and studied by many authors (for example, a review is given in Scarrott and MacDonald [32]) and is not the focus here. A special feature of the application to ship motions discussed in Section 3 is that the threshold selection should be automated, but with the possibility of closer examination if needed. The automatic selection is naturally sought in the ship motion application because multiple records need to be analyzed for the accuracy that is meaningful for practical applications.

In the automatic selection that we use, the threshold u is selected as the maximum of the thresholds u_{sh}, u_{ms}, u_{me} and u_{rt} chosen by the following four automatic procedures. The thresholds u_{sh}, u_{ms} and u_{me} are selected automatically from the commonly used shape parameter, modified scale parameter and mean excess plots, respectively. For example, the plot of the estimated shape parameters with confidence intervals (against thresholds) should be about constant over the range where GPD fit is appropriate. The threshold u_{sh} is chosen as the smallest threshold for which the horizontal line drawn from the corresponding estimate passes through the confidence intervals of the shape parameter for all the larger thresholds. The thresholds u_{ms} and u_{me} are chosen similarly except that the line in the mean excess plot does not need to be horizontal. The choice of the three thresholds is illustrated in Figure 4, for one of the data sets considered in Section 3 below.

The threshold u_{rt}, on the other hand, is selected following the Reiss and Thomas [31], p. 137, automatic procedure (see also Neves and Fraga Alves [29]). Let $\xi_{k,n}$ be the estimates of the shape parameter ξ based on the k largest values

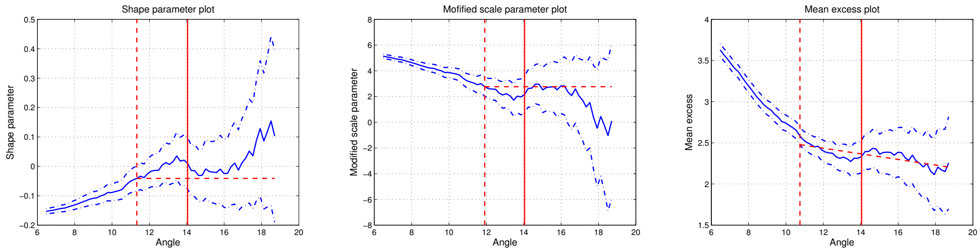


Figure 4: Shape parameter, modified scale parameter and mean excess plots. The vertical dashed line indicates the thresholds chosen with the corresponding (horizontal or arbitrary) lines passing through the confidence intervals for the larger thresholds. The vertical solid line indicates the threshold choice using the Reiss and Thomas method.

of y_i (by using the moment estimation for computational efficiency). Choose k^* as the value that minimizes

$$\frac{1}{k}\sum_{i\leq k} i^{\beta}|\xi_{k,n} - \mathrm{med}(\xi_{1,n},...,\xi_{k,n})|,$$

where $\beta = 1/2$ (though other values of $\beta < 1/2$ can be considered as well) and med denotes the median. In practice, after the suggestion of Reiss and Thomas, the function above is slightly smoothed. The threshold u_{rt} is then chosen as the k^* largest value of y_i. It is depicted as a vertical solid line in Figure 4 and probably better corresponds to a visually desired choice of threshold. In our experience, the Reiss and Thomas choice most often provides the largest (most conservative) value among the methods considered.

Table 3 presents the empirical coverage frequencies of the confidence intervals constructed through the above POT approach for several non-GPDs. The distributions considered are: the Weibull distribution with the CDF

$$F(x) = 1 - e^{-\lambda x^{\tau}}, \quad x > 0,$$

with parameters $\lambda > 0$, $\tau > 0$; the Burr distribution with the CDF

$$F(x) = 1 - \Big(\frac{\beta}{\beta + x^{\tau}}\Big)^{\lambda}, \quad x > 0,$$

with parameters $\lambda > 0$, $\tau > 0$, $\beta > 0$; and the reverse Burr distribution with the CDF

$$F(x) = 1 - \Big(\frac{\beta}{\beta + (x_+ - x)^{-\tau}}\Big)^{\lambda}, \quad x < x_+,$$

with parameters $\lambda > 0$, $\tau > 0$, $\beta > 0$. Two choices of the parameter τ are considered for the Weibull distribution, with $\tau = 1/2$ ($\tau = 2$, resp.) providing heavier (lighter, resp.) tails than exponential (but both associated with the shape parameter $\xi = 0$ in the POT framework). The Burr distribution has a power-law tail, corresponding to the shape parameter $\xi = 1/(\tau\lambda)$ in the POT framework. Similarly, the reverse Burr distribution has a finite upper bound x_+, and corresponds to the negative shape parameter $\xi = -1/(\tau\lambda)$ in the POT framework.

Table 3: Empirical coverage frequencies in the non-GPD context using the POT approach.

non-GPD						direct		quantile		
model	parameters	N	n	c	p_c	logn	bound	logn	bound	profl
Weibull	$(\lambda, \tau) = (1, 1/2)$	2000	126	132.5	10^{-5}	90.0	99.2	97.0	99.6	95.0
			123	190.9	10^{-6}	92.2	99.4	94.8	98.8	93.4
	$(\lambda, \tau) = (1, 2)$	2000	194	3.4	10^{-5}	94.6	96.4	90.0	96.4	94.4
			195	3.7	10^{-6}	94.4	97.2	86.8	94.2	93.2
Burr	$(\beta, \tau, \lambda) = (1, 2, 2)$	2000	221	17.8	10^{-5}	95.6	99.2	90.4	97.2	92.8
			210	31.6	10^{-6}	96.4	99.6	87.4	94.8	92.4
Reverse	$(\beta, x_+) = (0.1, 10)$	2000	156.5	9.8	10^{-5}	96.8	93	83.4	92.4	90.8
Burr	$(\tau, \lambda) = (2, 2)$		150	9.9	10^{-6}	98.2	92.2	80.4	90.2	89.0

Under the direct approach in Table 3, the coverage probabilities are reported only for the lognormal and boundary methods. The quantile methods use the proportion of data above the threshold to estimate P_{nr} but do not take the estimation uncertainty of P_{nr} into account. Two of the columns also give the sample size N and the average number of peaks over threshold n. As before, p_c is the exceedance probability and c is the corresponding critical target.

Our goal with Table 3 is not to provide an exhaustive study of the POT approach in the non-GPD framework, but rather to make a few general comments. First, note from the table that the approach works quite well. Second, note that the performance of the considered methods is not as uniformly good as in the GPD context. Thus, the performance of the methods for non-GPDs depends not only on the way to produce confidence intervals above a threshold but also on the non-GPD itself, as well as the (automatic) choice of the threshold.

3. APPLICATION TO EXTREME SHIP MOTIONS

We shall use the POT approach outlined in Section 2.3 to estimate the probability of roll and pitch angle exceeding a critical value. Several issues need to be addressed before we can apply the methods for constructing confidence intervals discussed in Section 2.3. An important and pressing issue is the presence of temporal dependence as clearly seen from Figure 2. A related issue is also what is meant by an exceedance probability and how it relates to time.

The issue of temporal dependence is addressed through the following envelope approach. Motivated by the periodic nature of a ship motion, the maxima and minima are first found between consecutive zero crossings of the series. These are the positive and negative peaks in the series of interest. The absolute values of the peaks are then connected by a piecewise linear function producing an *envelope* of the series. This is depicted in Figure 5. The left plot includes the original roll series for 5 minutes, with the positive and negative envelope. The right plot depicts the absolute values of the roll and the positive envelope connecting linearly the absolute values of the peaks.

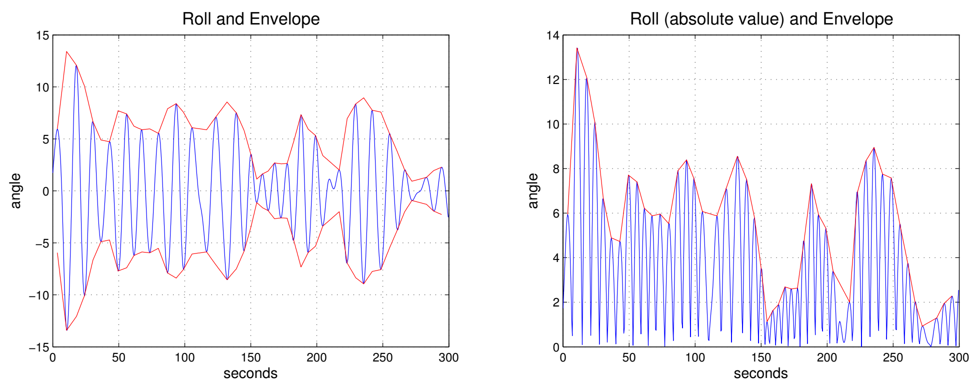


Figure 5: The roll angle series with envelope for 5 minutes. Left: original roll series. Right: roll series in absolute value.

After the envelope is found for the whole roll time series (not just the 5 minutes shown), its average value is computed. Next, the maxima and minima are found in the envelope between consecutive crossings of the average envelope value. These are the envelope peaks above/below the envelope average. This is illustrated in Figure 6, where the envelope average is plotted as a horizontal line and the envelope peaks above/below the envelope average are indicated by small black marks.

Note from Figure 6 that focusing on the envelope peaks (above the average) deals, at least qualitatively, with temporal dependence. That is, the larger values close in time are "clustered" and only the largest values in "clusters" are recorded as envelope peaks. (A closer look at the decorrelation properties of the envelope peak series can be found in a report by Belenky and Campbell [5].) In what follows, we shall work only with the envelope peaks. It is also important to note that the envelope approach is automated. This is particularly convenient when dealing with multiple conditions and many records.

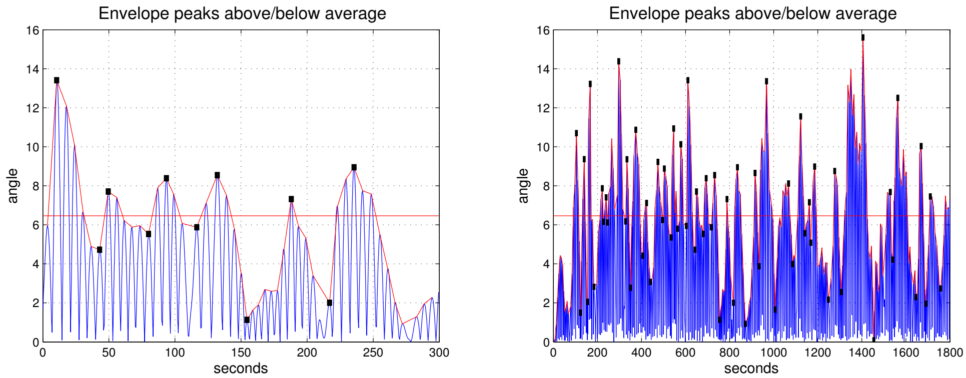


Figure 6: Envelope peaks above/below envelope average.
Left: 5 minutes. Right: 30 minutes.

Focusing on the envelope peaks also simplifies the notion of exceedance and the associated exceedance probabilities. Note that the series of interest will exceed a large target when an envelope peak will exceed the target. It is then natural to think of an exceedance probability as that for the envelope peaks. This is the perspective adopted throughout the paper.

We should also clarify what we mean by probabilities, which are now related to the envelope peaks. Suppose a series contains $1,000$ envelope peaks of which 45 exceed a given threshold. Then, the estimated probability is $45/1000 = .045$ of exceeding the threshold. This probability is not informative without a reference to time. Suppose the series is actually recorded over 15 minutes or $15 \cdot 60 = 900$ seconds. It is then more informative to consider the (probability) rate of $45/900 = .05$ envelope peaks (over the threshold) per second. Though we will continue referring to probabilities below, the results will be reported in terms of (probability) rates, rather than probabilities themselves.

If $x_1, ..., x_N$ are the envelope peaks of the series at hand, the exceedance probability is then estimated with a confidence interval as explained in Section 2.3. The performance of the confidence intervals can be assessed through a validation procedure as follows. The computer code (discussed in Section 1) can be used to generate significantly more series of ship motions, which contain rare events of interest and from which exceedance probabilities can be estimated by direct counting. More specifically, for the same condition used in Figures 2–6, the code was used to generate $115,000$ hours of the ship motion. With the target roll angle of $x_{cr} = 60$ degrees, the probability rate of exceedance obtained by direct counting based on rare events from the available records is 7.25×10^{-8} envelope peaks per second (that is, 30 envelope peaks above 60 degrees in $115,000$ hours).

This "true" rate estimate can be supplemented by the confidence interval obtained by a standard binomial argument.

A typical given series (record) to make inference from covers only 100 hours and would not contain rare events of interest. For each record, confidence intervals for exceedance probabilities can be computed as in Section 2.3. The confidence intervals can then be assessed by their coverage frequencies of the "true" exceedance probability. This could be examined graphically as in Figure 7 where the lognormal, boundary, quantile-lognormal and quantile-profile confidence intervals are presented for 100 records of the total length of 100 hours. The critical value of interest is the roll of 60 degrees as above. Note that the vertical axis for the probability rate is in the log scale, and that we truncated the confidence intervals and the probability (rate) estimates at a practically negligible probability rate of 10^{-15}. The horizontal dashed lines indicate the confidence bounds for the "true" probability. The small circles are the probability rate estimates.

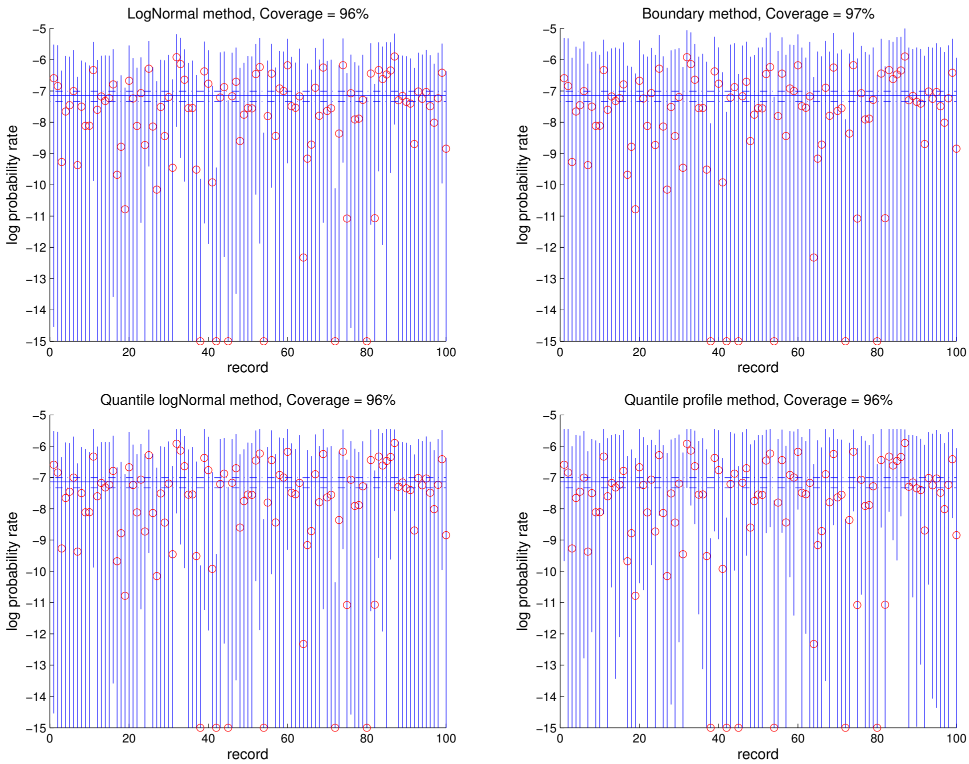


Figure 7: Confidence intervals for 100 records of the length of 100 hours. Roll series for 45° heading and with critical roll angle of 60°. Top left: lognormal method. Top right: boundary method. Bottom left: quantile-lognormal method. Bottom right: quantile-profile method.

For the roll and pitch motion at 45 and 30 degree headings, we also report the coverage frequencies for the methods of Section 2 in Table 4, based on the results in 100 records. The columns under $\widehat{\xi}$ and n provide the average estimates of the shape parameter and the number of peaks over threshold. The standard errors are given in parentheses. In the parentheses under the coverage probabilities, we provide the average of the sizes of the suggested confidence intervals above the true value (supposing it is contained), which will be discussed further in Section 4 below.

Table 4: Headings of 30 and 45 degress. Roll: target is 60 at 45° and 35 at 30°. Pitch: target is 10.

series				direct methods				quantile methods		
varble	head	$\widehat{\xi}$	n	logn	bound	boot	profl	logn	bound	profl
roll	45	0.19 (0.13)	96.06 (31.76)	96 (1.19)	97 (1.46)	92 (1.15)	100 (1.52)	96 (1.18)	97 (1.41)	96 (1.40)
	30	0.04 (0.13)	105.03 (46.96)	84 (0.82)	91 (1.13)	76 (0.87)	99 (1.37)	84 (0.82)	91 (1.12)	89 (1.07)
pitch	45	−0.06 (0.11)	107.06 (50.92)	99 (0.62)	100 (0.73)	95 (0.57)	98 (0.73)	99 (0.62)	100 (0.73)	100 (0.74)
	30	−0.08 (0.11)	107.63 (46.61)	97 (0.43)	98 (0.49)	94 (0.41)	96 (0.51)	97 (0.43)	98 (0.48)	98 (0.51)

Note from Table 4 that the performance of the confidence intervals is similar to those in Sections 2.2 and 2.3. Target values are chosen based on available rare events in the large set of records. The performance seems also satisfactory, validating the approach from a practical perspective. The point of using such validation is to show that the approach works on the ship motion data generated by a qualitatively correct computer code, before applying the methods to real or experimental data (where a large number of records are naturally not available). Or, put differently, had the methods not passed the validation, no applied researcher would be confident in using them.

The approach to estimate the exceedance probabilities certainly works in part because of the mathematical justification as outlined in Section 2.3. But this is not the whole story! Another important component to success is related to the length of the record and the physics of the ship motion. The 100–hour records are typical for Naval Architecture purposes. Our results show that these records have sufficiently enough physics to allow one to extrapolate into the tail using the POT framework.

4. UNCERTAINTY REDUCTION

An interesting but also practically important question is whether the uncertainty of the estimators or, equivalently, the size of of the confidence intervals can be reduced. For example, in Figure 7, the right (top) endpoints of the confidence intervals are about one order of magnitude above the true value. One order seems acceptable from a practical perspective. But we also encounter conditions where the uncertainty could be as high as two or three orders of magnitude.

Can the uncertainty (or the size of confidence intervals) be reduced? It surely depends on the approach and model used (that is, the POT approach with the two parameter GPD above threshold), the sample size (that is, the number of exceedances above threshold), and the efficiency of the estimation method used. Efficiency cannot be improved considerably since the ML estimators of the GPD parameters are used. But several directions could be explored when it comes to the first two points.

More specifically, in Section 4.1, we study the situation where it may be meaningful to fix a right upper bound when a negative shape parameter is expected. A substantial uncertainty reduction is achieved with this approach but it may not be promising to search for extensions to positive shape parameters, or ways of fixing a bound. Section 4.2 contains a short and, in our view, informative account of several other possibilities that we tried but which did not lead to much of the uncertainty reduction.

4.1. Fixing upper bound

When the shape parameter of a GP distribution is negative, the distribution has a finite upper bound. One direction for uncertainty reduction is to fix this upper bound before estimation based on some physical considerations, e.g. limiting angle for roll after which ship capsizes. Fixing the bound reduces the number of parameters from 2 to 1, so that the reduction of uncertainty is expected.

In applications to ship stability, the pitch motion typically yields a negative shape parameter, as can already be seen from Table 4 (3rd column). There are physical reasons for this phenomenon which, in technical terms, have to do with the form of the stiffness of the pitch motion. Moreover, again for physical reasons, an upper bound for the pitch motion may be expected at about 15°–20°, as roll stiffness of ONR Tumblehome becomes flat and does not support any resonance excitation. Details of the physics of the pitch motion go beyond the scope of this paper.

From a statistical standpoint, deriving the GPD framework with a fixed upper bound is straightforward. Suppose for notational simplicity that the threshold μ is 0, and denote a fixed upper bound by $y_{\max}$. When the shape parameter ξ of the GPD (1.1) is negative, the upper bound is given by $(-\sigma/\xi)$. Setting $y_{\max} = -\sigma/\xi$, solving for $\xi = -\sigma/y_{\max}$ and substituting this into (1.1) when $\xi < 0$, we obtain the complementary GPD function with the upper bound $y_{\max}$,

$$\overline{F}_\sigma(y) = \Big(1 - \frac{y}{y_{\max}}\Big)^{y_{\max}/\sigma}, \quad 0 < y < y_{\max}. \tag{4.1}$$

Note that the function (4.1) depends only on the scale parameter σ (with the shape parameter of the GPD being $\xi = -\sigma/y_{\max}$).

The parameter σ in (4.1) can be estimated using ML. Given observations $y_1, ..., y_n$ (all smaller than $y_{\max}$), optimizing the log-likelihood

$$\ell(\sigma) = \sum_{i=1}^{n} \log\Big(\frac{1}{\sigma}\Big(1 - \frac{y_i}{y_{\max}}\Big)^{y_{\max}/\sigma - 1}\Big)$$

leads to the ML estimator

$$\widehat{\sigma} = -\frac{y_{\max}}{n}\sum_{i=1}^{n} \log\Big(1 - \frac{y_i}{y_{\max}}\Big). \tag{4.2}$$

The inverse of the observed information matrix can easily be checked to be

$$\Big(-\frac{\partial^2 \ell}{\partial \sigma^2}\Big)^{-1}\Big|_{\sigma=\widehat{\sigma}} = \frac{\widehat{\sigma}^2}{n}. \tag{4.3}$$

A confidence interval for an exceedance probability $p_c = \overline{F}_{\sigma_0}(c)$ can then be given by the boundary method as $(\overline{F}_{\sigma_1}(c), \overline{F}_{\sigma_2}(c))$, where $\sigma_1 = \widehat{\sigma} - C_\alpha\widehat{\sigma}/\sqrt{n}$ and $\sigma_2 = \widehat{\sigma} + C_\alpha\widehat{\sigma}/\sqrt{n}$ are two critical values for the distribution of $\widehat{\sigma}$ based on (4.3) (with as before, C_α denoting the $100(\alpha/2)\%$ quantile of the standard normal distribution).

Figure 8 compares the confidence intervals for the exceedance probability of the pitch motion at the 30° heading (under the same condition as earlier) obtained through the lognormal method as in Section 3, and the boundary method with the upper bound fixed at 15° as explained above. The left plot in Figure 8 corresponds to the entry of Table 4 under "pitch", "30" degree heading and "logn" method, with the uncertainty measure of 0.43 in the parentheses. The same measure for the right-plot of Figure 8 is 0.34. The reduction of uncertainty is also evident from Figure 8 itself, with smaller variability of the estimators (red circles) and the sizes of confidence intervals in the right plot.

It should also be noted that the results with the fixed upper bound are not sensitive to the choice of the bound (suggested by physical considerations). For example, fixing the bound at 17° and 20° leads to the same coverage frequency of 99%, with the exception that the uncertainty measure above becomes slightly

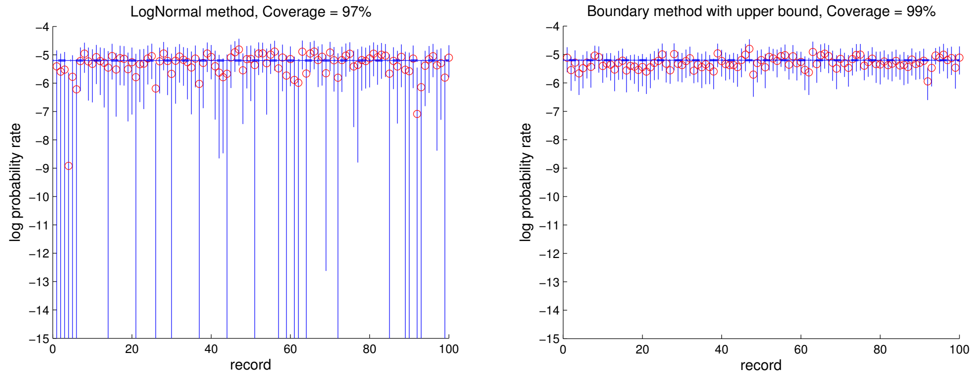


Figure 8: Confidence intervals for 100 records of the length of 100 hours. Pitch 30°. Left: lognormal method. Right: boundary method with fixed upper bound at 15°.

larger, 0.36 and 0.38, respectively. The conclusions are the same for the pitch motion at the 30° heading (not reported here).

Remark 4.1. Whether a similar approach can be developed for a positive shape parameter remains an open question. One idea we entertained was to experiment with truncated GPD models in the spirit of, for example, Aban *et al.* [1], Beirlant *et al.* [3]. (Truncation seems natural because, for example, the roll and pitch angles are bounded by 180 degrees.) But the truncated GPD models did not appear to fit the data well. In Belenky *et al.* [6], we study extreme value methods on mathematically tractable physical models mimicking ship motion dynamics, and expect to gain further insight into the above issues from this approach.

4.2. Other possibilities

We explored or thought about several other possibilities for uncertainty reduction. One natural possibility would be to view the variables describing different conditions as covariates and then pool the data across different conditions by modeling covariates to reduce uncertainty. This idea is particularly relevant in the application of interest here since naval engineers have to take measurements regularly across a range of conditions. The idea also has a sound statistical footing, as developed in Davison and Smith [10] and described, for example, in Chapter 6 of Coles [9].

Following this approach, we have modeled records across a number of headings (e.g. 15°, 22.5°, 30°, 37.5°, 45° degrees). But we generally found the reduction in uncertainty small if any. Some of this is due to a small reduction of uncertainty even under ideal situations (when the model incorporating the covariates is known). The uncertainty in the underlying model for the covariates (entering the POT framework) also plays a role.

Finally, another possibility might be to use some of the more advanced approaches in modeling dependent peaks over threshold, as in e.g. Smith *et al.* [35]. The idea here is that this would seemingly allow for a larger sample size to be considered. Even if the dependence structure is captured correctly by these approaches, we also expect them to lead to little uncertainty reduction. As with the covariates above, we view these approaches as serving different purposes and used to answer different questions.

5. CONCLUSIONS

In this work, we studied the various methods to construct confidence intervals for exceedance probabilities in the peaks-over-threshold approach. The performance of the confidence intervals was assessed through several simulation studies, pointing to the superior performance of some of the considered methods. The developed methods were applied to build confidence intervals for the probabilities of extreme ship motions, leading to satisfactory results overall. Finally, several uncertainty reduction approaches were considered, with a promising solution when a negative shape parameter is expected. Whether uncertainty reduction can be achieved in the case of a positive shape parameter remains an open question.

ACKNOWLEDGMENTS

The authors thank the anonymous Reviewer whose comments helped improve the paper considerably. The authors would also like to thank Prof. Ross Leadbetter (University of North Carolina, Chapel Hill) for valuable discussions about the paper. The work described in this paper has been funded by the Office of Naval Research (ONR) under Dr. Patrick Purtell and Dr. Ki-Han Kim as well as NSWCCD under Dr. Jack Price. Participation of the first author was facilitated by the Summer Faculty Program supported by ONR and managed by Dr. Jack Price. The hospitality of the NSWCCD and the ONR support are gratefully acknowledged.

REFERENCES

[1] Aban, I.B.; Meerschaert, M.M. and Panorska, A.K., (2006). Parameter estimation for the truncated Pareto distribution, *Journal of the American Statistical Association*, **101**(473), 270–277.

[2] Beirlant, J.; Goegebeur, Y.; Teugels, J. and Segers, Jo. (2004). *Statistics of Extremes*, John Wiley & Sons, Ltd., Chichester.

[3] Beirlant, J.; Fraga Alves, I. and Gomes, I. (2015). *Tail fitting for truncated and non-truncated Pareto-type distributions*, Preprint.

[4] Belenky, V.L. and Sevastianov, N.B. (2007). *Stability and Safety of Ships: Risk of Capsizing, 2nd edition*, Society of Naval Architects and Marine Engineers.

[5] Belenky, V. and Campbell, B.L. (2011). Valuation of the exceedance rate of a stationary stochastic process by statistical extrapolation using the envelope peaks over threshold (EPOT) method, *Naval Surface Warfare Center, Carderock Division*, Hydromechanics Department Report, NSWCCD-50-TR-2011/032.

[6] Belenky, V.; Glotzer, D.; Pipiras, V. and Sapsis, T.P. (2016). *Distribution tail structure and extreme value analysis of constrained piecewise linear oscillators*, Preprint.

[7] Benford, H. (1991). *Naval Architecture for Non-Naval Architects*, Society of Naval Architects and Marine Engineers.

[8] Bishop, R.C.; Belknap, W.; Turner, C.; Simon, B. and Kim, J.H. (2005). Parametric investigation on the influence of GM, roll damping, and above-water form on the roll response of model 5613, *David Taylor Model Basin – NSWCCD*, Report, NSWCCD-50-TR-2005/027.

[9] Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*, Springer-Verlag London Ltd., London.

[10] Davison, A.C. and Smith, R.L. (1990). Models for exceedances over high thresholds, *Journal of the Royal Statistical Society. Series B. Methodological*, **52**(3), with discussion and a reply by the authors.

[11] de Carvalho, M.; Turkman, K.F. and Rua, A. (2013). Dynamic threshold modelling and the US business cycle, *Journal of the Royal Statistical Society. Series C. Applied Statistics*, **62**(4), 535–550.

[12] de Haan, L. and de Ronde, J. (1998). Sea and wind: multivariate extremes at work, *Extremes*, **1**(1), 7–45.

[13] de Haan, L. and Sinha, A.K. (1999). Estimating the probability of a rare event, *The Annals of Statistics*, **27**(2), 732–759.

[14] de Zea Bermudez, P. and Kotz, S. (2010). Parameter estimation of the generalized Pareto distribution, I, *Journal of Statistical Planning and Inference*, **140**(6), 1353–1373.

[15] de Zea Bermudez, P. and Kotz, S. (2010). Parameter estimation of the generalized Pareto distribution, II, *Journal of Statistical Planning and Inference*, **140**(6), 1374–1388.

[16] Draghicescu, D. and Ignaccolo, R. (2009). Modeling threshold exceedance probabilities of spatially correlated time series, *Electronic Journal of Statistics*, **3**, 149–164.

[17] Drees, H. and de Haan, L. (2014). *Estimating failure probabilities*, Preprint.

[18] Embrechts, P.; Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extreme Events*, Springer-Verlag, New York.

[19] Ferreira, A. and de Haan, L. (2014). The generalized Pareto process; with a view towards application and simulation, *Bernoulli*, **20**(4), 1717–1737.

[20] Heffernan, J.E. and Tawn, J.A. (2010). A conditional approach for multivariate extreme values, *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, **66**(3), 497–546, with discussions and reply by the authors.

[21] Hong, Y.; Meeker, W.Q. and Escobar, L.A. (2008). The relationship between confidence intervals for failure probabilities and life time quantiles, *Reliability, IEEE Transactions on*, **57**(2), 260–266.

[22] Hosking, J.R.M. and Wallis, J.R. (1987). Parameter and quantile estimation for the generalized Pareto distribution, *Technometrics*, **29**(3), 339–349.

[23] Klein, J.P. and Moeschberger, M.L. (2003). *Survival Analysis: Techniques for Censored and Truncated Data*, Springer.

[24] Lawless, J.F. (1978). Confidence interval estimation for the Weibull and extreme value distributions, *Technometrics*, **20**(4, part 1), 355–365, with a discussion by R.G. Easterling.

[25] Lewis, E.V. and Society of Naval Architects and Marine Engineers (U.S.) (1990). *Principles of Naval Architecture: Volume III Motions in Waves and Controllability*, Society of Naval Architects and Marine Engineers.

[26] Lin, W.M. and Yue, D.K.P. (1991). Numerical solutions for large amplitude ship motions in the time-domain, *Proceedings of the 18th Symposium on Naval Hydrodynamics, Ann Arbor*, 41–66.

[27] McTaggart, K.A. (2000). Ship capsize risk in a seaway using fitted distributions to roll maxima, *Journal of Offshore Mechanics and Arctic Engineering*, **122**(2), 141–146.

[28] Neves, M.A.S.; Belenky, V.L.; de Kat, J.O.; Spyrou, K. and Umeda, N. (2011). *Contemporary Ideas on Ship Stability and Capsizing in Waves*, Springer.

[29] Neves, C. and Fraga Alves, M.I. (2004). Reiss and Thomas' automatic selection of the number of extremes, *Computational Statistics & Data Analysis*, **47**(4), 689–704.

[30] Reed, A.; Beck, R. and Belknap, W. (2014). Advances in predictive capability of ship dynamics in waves, *Proceedings of the 30th Symposium of Naval Hydrodynamics, Hobart.*

[31] Reiss, R.D. and Thomas, M. (2007). *Statistical Analysis of Extreme Values With Applications To Insurance, Finance, Hydrology and Other Fields*, Birkhäuser Verlag, Basel.

[32] Scarrott, C. and MacDonald, A. (2012). A review of extreme value threshold estimation and uncertainty quantification, *REVSTAT*, **10**(1), 33–60.

[33] Smith, R.L. (1987). Estimating tails of probability distributions, *The Annals of Statistics*, **15**, 1174–1207.

[34] Smith, R.L. and Shively, T.S. (1995). Point process approach to modeling trends in tropospheric ozone based on exceedances of a high threshold, *Atmospheric Environment*, **29**(23), 3489–3499.

[35] Smith, R.L.; Tawn, J.A. and Coles, S.G. (1997). Markov chain models for threshold exceedances, *Biometrika*, **84**(2), 249–268.

[36] Tajvidi, N. (2003). Confidence intervals and accuracy estimation for heavy-tailed generalized Pareto distributions, *Extremes*, **6**(2), 111–123.

[37] Weems, K. and Wundrow, D. (2013). Hybrid models for fast time-domain simulation of stability failures in irregular waves with volume-based calculations for Froude-Krylov and hydrostatic forces, *Proceedings of the 13th International Ship Stability Workshop, Brest.*

[38] Zhao, W.; Pan, R.; Aron, A. and Mettas, A. (2006). Some properties of confidence bounds on reliability estimation for parts under varying stresses, *Reliability, IEEE Transactions on*, **55**(1), 7–17.

REVSTAT – Statistical Journal
Volume 15, Number 4, October 2017, 565–581

RELIABILITY ESTIMATION IN MULTISTAGE RANKED SET SAMPLING

Authors: M. Mahdizadeh
– Department of Statistics, Hakim Sabzevari University,
P.O. Box 397, Sabzevar, Iran
mahdizadeh.m@live.com

Ehsan Zamanzade
– Department of Statistics, University of Isfahan,
Isfahan, 81746-73441, Iran
e.zamanzade@sci.ui.ac.ir
ehsanzamanzadeh@yahoo.com

Received: October 2015 Revised: March 2016 Accepted: March 2016

Abstract:

- A nonparametric reliability estimator based on multistage ranked set sampling is developed. It is shown that the estimator is unbiased and its efficiency relative to the simple random sampling rival is increasing in the number of stages. Numerical experiments are used to illustrate the theoretical findings. The suggested procedure is applied on a sport data set.

Key-Words:

- *covariate information; judgment ranking; stress-strength model.*

1. INTRODUCTION

Ranked set sampling (RSS) is a data collection technique which is advantageous in settings where precise measurement is difficult (i.e. time-consuming, expensive or destructive), but small sets of units can be accurately ranked without actual quantification. The ranking of the units is usually done by using expert opinion, concomitant variable, or a combination of them, and need not to be exact.

The RSS method was introduced by McIntyre (1952) for estimating average yields in agriculture. In this setup, precise measurement entails harvesting the crops, and thus is expensive. An expert, however, can accurately rank the yields in a small set of adjacent fields by visual inspection. There has been a surge of research on RSS in the last two decades. The RSS has been applied in a variety of areas such as forestry, environmental science and medicine. For a book-length treatment of RSS and its applications, see Chen *et al.* (2004).

The RSS design can be elucidated as follows:

1. Draw m random samples, each of size m, from the target population.

2. Apply judgement ordering, by any cheap method, on the elements of the ith $(i = 1, ..., m)$ sample and identify the ith smallest unit.

3. Actually measure the m identified units in step 2.

4. Repeat steps 1–3, p times (cycles), if necessary, to obtain a ranked set sample of size $M = p\,m$.

Let X_{ik} be the ith judgement order statistic from the kth cycle. Then, the resulting ranked set sample is denoted by $\{X_{ik} : i = 1, ..., m\,; k = 1, ..., p\}$. The design parameter m is called set size.

A ranked set sample contains more information than a simple random sample of comparable size because it contains not only information carried by quantified observations but also information provided by the judgment ranking mechanism. Thus, statistical procedures based on RSS tend to be superior to their simple random sampling (SRS) analogs.

The success of RSS hinges on accuracy of the ranking process. To reduce possible errors, the set size m should be kept small in the basic version of RSS. Al-Saleh and Al-Kadiri (2000) suggested double RSS (DRSS) that increases efficiency of the RSS mean estimator, given a fixed m. Al-Saleh and Al-Omari (2002) generalized DRSS to multistage RSS (MSRSS), and showed that further gain in efficiency can be achieved in estimating the population mean. Al-Saleh and Samuh (2008) investigated the distribution function and the median estimation based on MSRSS.

The MSRSS scheme can be summarized as follows:

1. Randomly identify m^{r+1} units from the population of interest, where r is the number of stages.

2. Allot the m^{r+1} units randomly into m^{r-1} sets of m^2 units each.

3. For each set in step 2, apply steps 1–2 of RSS procedure explained above, to get a (judgement) ranked set of size m. This step gives m^{r-1} (judgement) ranked sets, each of size m.

4. Without actual measuring of the ranked sets, apply step 3 on the m^{r-1} ranked set to gain m^{r-2} second stage (judgement) ranked sets, of size m each.

5. Repeat step 3, without any actual measurement, until an rth stage (judgement) ranked set of size m is acquired.

6. Actually measure the m identified units in step 5.

7. Repeat steps 1–6, p times (cycles), if necessary, to obtain an rth stage ranked set sample of size $M = p\,m$.

Similar to our previous notation, $\{X_{ik}^{(r)} : i = 1, ..., m\,; k = 1, ..., p\}$ denotes the rth stage ranked set sample. Clearly, the especial case of MSRSS with $r = 1$ corresponds to RSS. Also, DRSS is obtained by setting $r = 2$.

The estimation of system reliability has drawn much attention in the statistical literature. Reliability of a component with strength X which is subjected to stress Y is quantified by $\theta = P(X > Y)$. This approach is known as the stress-strength model. The estimation of θ has been extensively investigated in the literature when X and Y are independent random variables, and belong to the same family of distributions. A comprehensive account of this topic appear in Kotz *et al.* (2003). In this article, we study reliability estimation in MSRSS setup.

In Section 2, a nonparametric estimator is proposed and its properties are investigated in theory. Section 3 is given to a Monte Carlo analysis of the finite sample behavior of the estimator. A sport data set is analyzed in Section 4. The paper is concluded with a summary in Section 5.

2. ESTIMATION USING MSRSS

Let $X_1, ..., X_m$ and $Y_1, ..., Y_n$ be independent random samples from two populations with density functions f and g, respectively. The corresponding distribution functions are denoted by F and G. The standard nonparametric estimator of θ is

$$\hat{\theta} = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} I(X_i > Y_j),$$

where $I(.)$ is the indicator function.

To construct an estimator under MSRSS, one needs two ranked set samples of sizes m and n from f and g. It is assumed that the samples are drawn using a single cycle. The results in the general setup are then easily followed. If $X_i^{(r)}$, $i = 1, ..., m$, and $Y_j^{(s)}$, $j = 1, ..., n$, are the two multistage ranked set samples, then

$$\hat{\theta}_{r,s} = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} I(X_i^{(r)} > Y_j^{(s)})$$

is a natural estimator of θ. The especial case of $r = s = 1$ was treated by Sengupta and Mukhuti (2008).

Let $f_i^{(r)}$ and $F_i^{(r)}$ be the density and distribution function of $X_i^{(r)}$, respectively. The notation $g_j^{(s)}$ and $G_j^{(s)}$ will be used for similar functions associated with $Y_j^{(s)}$. Suppose the ith order statistic of an $(r-1)$th stage ranked set sample of size m from f, say $Z_1^{(r-1)}, ..., Z_m^{(r-1)}$, is denoted by $Z_{(i)}^{(r-1)}$. Under the assumption of no error in judgment ranking, we have $X_i^{(r)} \stackrel{d}{=} Z_{(i)}^{(r-1)}$.

In our mathematical development, the two identities

$$\frac{1}{m} \sum_{i=1}^{m} f_i^{(r)}(x) = f(x)$$

and

$$\frac{1}{n} \sum_{j=1}^{n} g_j^{(s)}(y) = g(y),$$

observed by Al-Saleh and Al-Omari (2002), are repeatedly used. The above identities can be expressed in terms of distribution functions, as well.

It is straightforward to see that $\hat{\theta}$ is unbiased. The unbiasedness of $\hat{\theta}_{r,s}$ is verified in the following proposition.

Proposition 2.1. *$\hat{\theta}_{r,s}$ is an unbiased estimator of θ.*

Proof:

$$
\begin{aligned}
E\left\{\sum_{i=1}^{m}\sum_{j=1}^{n} I(X_i^{(r)} > Y_j^{(s)})\right\} &= \sum_{i=1}^{m}\sum_{j=1}^{n} P(X_i^{(r)} > Y_j^{(s)}) \\
&= \sum_{i=1}^{m}\sum_{j=1}^{n} \int P(X_i^{(r)} > y) g_j^{(s)}(y)\,\mathrm{d}y \\
&= n\sum_{i=1}^{m} \int P(X_i^{(r)} > y) g(y)\,\mathrm{d}y \\
&= n\sum_{i=1}^{m} P(X_i^{(r)} > Y) \\
&= n\sum_{i=1}^{m} \int P(x > Y) f_i^{(r)}(x)\,\mathrm{d}x \\
&= mn\int P(x > Y) f(x)\,\mathrm{d}x \\
&= mnP(X > Y). \qquad \square
\end{aligned}
$$

We now derive variance expressions of the two estimators.

Proposition 2.2. *The variances of $\hat{\theta}$ and $\hat{\theta}_{r,s}$ are given by*

$$
\begin{aligned}
m^2n^2Var(\hat{\theta}) &= m(m-1)n(n-1)\theta^2 + nm(m-1)E\left\{\bar{F}(Y)\right\}^2 \\
&+ mn(n-1)E\left\{G(X)\right\}^2 + mn\theta - m^2n^2\theta^2, \qquad (2.1)
\end{aligned}
$$

and

$$
\begin{aligned}
m^2n^2Var(\hat{\theta}_{r,s}) &= E\left\{m^2\Big[\sum_{j=1}^{n}\bar{F}(Y_j^{(s)})\Big]^2 - \sum_{i=1}^{m}\Big[\sum_{j=1}^{n}\bar{F}_i^{(r)}(Y_j^{(s)})\Big]^2\right\} \\
&+ mE\left\{n^2\Big[G(X)\Big]^2 - \sum_{j=1}^{n}\Big[G_j^{(s)}(X)\Big]^2\right\} \\
&+ mn\theta - m^2n^2\theta^2. \qquad (2.2)
\end{aligned}
$$

Proof: It is easy to show that

$$
m^2n^2E(\hat{\theta}^2) = E(A_1 + A_2 + A_3 + A_4), \qquad (2.3)
$$

where

$$
\begin{aligned}
E(A_1) &= E\Bigg\{\sum_{i\neq i'=1}^{m}\sum_{j\neq j'=1}^{n} I(X_i > Y_j)I(X_{i'} > Y_{j'})\Bigg\} \\
&= m(m-1)n(n-1)\theta^2,
\end{aligned}
\tag{2.4}
$$

$$
\begin{aligned}
E(A_2) &= E\Bigg\{\sum_{j=1}^{n}\sum_{i\neq i'=1}^{m} I(X_i > Y_j)I(X_{i'} > Y_j)\Bigg\} \\
&= \sum_{j=1}^{n}\sum_{i\neq i'=1}^{m} EE\Bigg\{I(X_i > Y_j)I(X_{i'} > Y_j)\Big| Y_j\Bigg\} \\
&= \sum_{j=1}^{n}\sum_{i\neq i'=1}^{m} E\Bigg\{\bar{F}(Y)\Bigg\}^2 = nm(m-1)E\Bigg\{\bar{F}(Y)\Bigg\}^2,
\end{aligned}
\tag{2.5}
$$

$$
\begin{aligned}
E(A_3) &= E\Bigg\{\sum_{i=1}^{m}\sum_{j\neq j'=1}^{n} I(X_i > Y_j)I(X_i > Y_{j'})\Bigg\} \\
&= \sum_{i=1}^{m}\sum_{j\neq j'=1}^{n} EE\Bigg\{I(X_i > Y_j)I(X_i > Y_{j'})\Big| X_i\Bigg\} \\
&= \sum_{i=1}^{m}\sum_{j\neq j'=1}^{n} E\Bigg\{G(X)\Bigg\}^2 = mn(n-1)E\Bigg\{G(X)\Bigg\}^2,
\end{aligned}
\tag{2.6}
$$

and

$$
E(A_4) = E\Bigg\{\sum_{i=1}^{m}\sum_{j=1}^{n} I(X_i > Y_j)\Bigg\} = mn\theta. \tag{2.7}
$$

From (2.3)–(2.7) and unbiasedness of $\hat{\theta}$, the proof of the first part is complete. Similarly,

$$
m^2n^2E(\hat{\theta}^2_{r,s}) = E(B_1 + B_2 + B_3), \tag{2.8}
$$

where

$$\begin{aligned}
E(B_1) &= E\Bigg\{ \sum_{i\neq i'=1}^{m} \sum_{j\neq j'=1}^{n} I(X_i^{(r)} > Y_j^{(s)}) I(X_{i'}^{(r)} > Y_{j'}^{(s)}) \\
&+ \sum_{j=1}^{n} \sum_{i\neq i'=1}^{m} I(X_i^{(r)} > Y_j^{(s)}) I(X_{i'}^{(r)} > Y_j^{(s)}) \Bigg\} \\
&= \sum_{i\neq i'=1}^{m} \sum_{j\neq j'=1}^{n} EE\Bigg\{ I(X_i^{(r)} > Y_j^{(s)}) \Big| Y_j^{(s)} \Bigg\} EE\Bigg\{ I(X_{i'}^{(r)} > Y_{j'}^{(s)}) \Big| Y_{j'}^{(s)} \Bigg\} \\
&+ \sum_{j=1}^{n} \sum_{i\neq i'=1}^{m} EE\Bigg\{ I(X_i^{(r)} > Y_j^{(s)}) I(X_{i'}^{(r)} > Y_j^{(s)}) \Big| Y_j^{(s)} \Bigg\} \\
&= E\Bigg\{ \sum_{i\neq i'=1}^{m} \sum_{j\neq j'=1}^{n} \Big[\bar{F}_i^{(r)}(Y_j^{(s)})\Big]\Big[\bar{F}_{i'}^{(r)}(Y_{j'}^{(s)})\Big] \\
&+ \sum_{j=1}^{n} \sum_{i\neq i'=1}^{m} \Big[\bar{F}_i^{(r)}(Y_j^{(s)})\Big]\Big[\bar{F}_{i'}^{(r)}(Y_j^{(s)})\Big] \Bigg\} \\
&= E\Bigg\{ \Big[\sum_{i=1}^{m}\sum_{j=1}^{n} \bar{F}_i^{(r)}(Y_j^{(s)})\Big]^2 - \sum_{i=1}^{m}\sum_{j=1}^{n} \Big[\bar{F}_i^{(r)}(Y_j^{(s)})\Big]^2 \\
&- \sum_{i=1}^{m} \sum_{j\neq j'=1}^{n} \Big[\bar{F}_i^{(r)}(Y_j^{(s)})\Big]\Big[\bar{F}_i^{(r)}(Y_{j'}^{(s)})\Big] \Bigg\} \\
&= E\Bigg\{ m^2 \Big[\sum_{j=1}^{n} \bar{F}(Y_j^{(s)})\Big]^2 - \sum_{i=1}^{m} \Big[\sum_{j=1}^{n} \bar{F}_i^{(r)}(Y_j^{(s)})\Big]^2 \Bigg\}, \qquad (2.9)
\end{aligned}$$

$$\begin{aligned}
E(B_2) &= E\Bigg\{ \sum_{i=1}^{m} \sum_{j\neq j'=1}^{n} I(X_i^{(r)} > Y_j^{(s)}) I(X_i^{(r)} > Y_{j'}^{(s)}) \Bigg\} \\
&= m \sum_{j\neq j'=1}^{n} E\Bigg\{ I(X > Y_j^{(s)}) I(X > Y_{j'}^{(s)}) \Bigg\} \\
&= m \sum_{j\neq j'=1}^{n} EE\Bigg\{ I(X > Y_j^{(s)}) I(X > Y_{j'}^{(s)}) \Big| X \Bigg\} \\
&= m \sum_{j\neq j'=1}^{n} E\Bigg\{ \Big[G_j^{(s)}(X)\Big]\Big[G_{j'}^{(s)}(X)\Big] \Bigg\} \\
&= mE\Bigg\{ n^2 \Big[G(X)\Big]^2 - \sum_{j=1}^{n} \Big[G_j^{(s)}(X)\Big]^2 \Bigg\}, \qquad (2.10)
\end{aligned}$$

and

$$E(B_3) = E\Bigg\{ \sum_{i=1}^{m}\sum_{j=1}^{n} I(X_i^{(r)} > Y_j^{(s)}) \Bigg\} = mn\theta. \qquad (2.11)$$

Now the second part follows from (2.8)–(2.11) and unbiasedness of $\hat{\theta}_{r,s}$. $\square$

The variances of $\hat{\theta}$ and $\hat{\theta}_{r,s}$ are compared in the next proposition.

Proposition 2.3. *For any $m, n \geq 2$ and $r, s \geq 1$, $Var(\hat{\theta}_{r,s}) \leq Var(\hat{\theta})$.*

Proof: Using equations (2.1) and (2.2), it can be shown

$$m^2n^2\Big[Var(\hat{\theta}) - Var(\hat{\theta}_{r,s})\Big] = C_1 + C_2 + C_3,$$

where

$$\begin{aligned} C_1 &= E\Bigg\{\sum_{i=1}^{m}\Big[\sum_{j=1}^{n}\bar{F}_i^{(r)}(Y_j^{(s)})\Big]^2 - m\Big[\sum_{j=1}^{n}\bar{F}(Y_j^{(s)})\Big]^2\Bigg\} \\ &= E\Bigg\{\sum_{i=1}^{m}\Big(\sum_{j=1}^{n}\Big[\bar{F}_i^{(r)}(Y_j^{(s)}) - \bar{F}(Y_j^{(s)})\Big]\Big)^2\Bigg\}, \end{aligned}$$

$$\begin{aligned} C_2 &= mn(n-1)E\Bigg\{G(X)\Bigg\}^2 - mE\Bigg\{n^2\Big[G(X)\Big]^2 - \sum_{j=1}^{n}\Big[G_j^{(s)}(X)\Big]^2\Bigg\} \\ &= mE\Bigg\{\sum_{j=1}^{n}\Big[G_j^{(s)}(X)\Big]^2 - n\Big[G(X)\Big]^2\Bigg\} \\ &= mE\Bigg\{\sum_{j=1}^{n}\Big[G_j^{(s)}(X) - G(X)\Big]^2\Bigg\}, \end{aligned}$$

and

$$\begin{aligned} C_3 &= m(m-1)n(n-1)\theta^2 + nm(m-1)E\Bigg\{\bar{F}(Y)\Bigg\}^2 \\ &- m(m-1)E\Bigg\{\Big[\sum_{j=1}^{n}\bar{F}(Y_j^{(s)})\Big]^2\Bigg\} \\ &= m(m-1)\Bigg[(1-\frac{1}{n})\Bigg(\sum_{j=1}^{n}E\Bigg\{\bar{F}(Y_j^{(s)})\Bigg\}\Bigg)^2 \\ &- \sum_{j\neq j'=1}^{n}E\Bigg\{\bar{F}(Y_j^{(s)})\Bigg\}E\Bigg\{\bar{F}(Y_{j'}^{(s)})\Bigg\}\Bigg] \\ &= m(m-1)\Bigg[\sum_{j=1}^{n}E^2\Bigg\{\bar{F}(Y_j^{(s)})\Bigg\} - \frac{1}{n}\Bigg(\sum_{j=1}^{n}E\Bigg\{\bar{F}(Y_j^{(s)})\Bigg\}\Bigg)^2\Bigg] \\ &= m(m-1)\sum_{j=1}^{n}E^2\Bigg\{\bar{F}(Y_j^{(s)}) - \bar{F}(Y)\Bigg\}. \end{aligned}$$

Clearly, $C_i \geq 0$, $i = 1, 2, 3$, as was asserted. $\square$

As mentioned earlier, increasing the number of stages leads to improvement in the context of mean and distribution function estimation based on MSRSS. So, it is natural to observe similar trend in the case of reliability estimation. The next result attends to this problem.

Proposition 2.4. *For fixed m and n, $Var(\hat{\theta}_{r,s})$ is decreasing in r and s.*

Proof: It suffices to show that $Var(\hat{\theta}_{r,s}) \leq Var(\hat{\theta}_{r-1,s})$ and $Var(\hat{\theta}_{r,s}) \leq Var(\hat{\theta}_{r,s-1})$. From the beginning of proof for the second part of Proposition 2.2, one can write

$$\begin{aligned} m^2n^2E(\hat{\theta}^2_{r,s}) &= E\Bigg\{ \sum_{i\neq i'=1}^{m} \sum_{j\neq j'=1}^{n} I(X_i^{(r)} > Y_j^{(s)}) I(X_{i'}^{(r)} > Y_{j'}^{(s)}) \\ &\quad + \sum_{i=1}^{m} \sum_{j\neq j'=1}^{n} I(X_i^{(r)} > Y_j^{(s)}) I(X_i^{(r)} > Y_{j'}^{(s)}) \\ &\quad + \sum_{j=1}^{n} \sum_{i\neq i'=1}^{m} I(X_i^{(r)} > Y_j^{(s)}) I(X_{i'}^{(r)} > Y_j^{(s)}) \\ &\quad + \sum_{i=1}^{m}\sum_{j=1}^{n} I(X_i^{(r)} > Y_j^{(s)}) \Bigg\}. \end{aligned} \tag{2.12}$$

We now establish some equalities and inequalities regarding the four expectation terms on the right-hand side of the above equation. Let $W_{(i)}^{(r-1)}$ be the ith order statistic of an $(r-1)$th stage ranked set sample of size m from f. As to the first term, we have

$$\begin{aligned} &E\Big\{ I(X_i^{(r)} > Y_j^{(s)}) I(X_{i'}^{(r)} > Y_{j'}^{(s)}) \Big\} \\ &\qquad = EE\Big\{ I(X_i^{(r)} > Y_j^{(s)}) I(X_{i'}^{(r)} > Y_{j'}^{(s)}) \Big| Y_j^{(s)}, Y_{j'}^{(s)} \Big\} \\ &\qquad = E\Big[E\Big\{ I(X_i^{(r)} > Y_j^{(s)}) \Big| Y_j^{(s)}, Y_{j'}^{(s)} \Big\} \\ &\qquad \times E\Big\{ I(X_{i'}^{(r)} > Y_{j'}^{(s)}) \Big| Y_j^{(s)}, Y_{j'}^{(s)} \Big\} \Big] \\ &\qquad = E\Big[E\Big\{ I(W_{(i)}^{(r-1)} > Y_j^{(s)}) \Big| Y_j^{(s)}, Y_{j'}^{(s)} \Big\} \\ &\qquad \times E\Big\{ I(W_{(i')}^{(r-1)} > Y_{j'}^{(s)}) \Big| Y_j^{(s)}, Y_{j'}^{(s)} \Big\} \Big] \\ &\qquad \leq EE\Big\{ I(W_{(i)}^{(r-1)} > Y_j^{(s)}) I(W_{(i')}^{(r-1)} > Y_{j'}^{(s)}) \Big| Y_j^{(s)}, Y_{j'}^{(s)} \Big\} \\ &\qquad = E\Big\{ I(W_{(i)}^{(r-1)} > Y_j^{(s)}) I(W_{(i')}^{(r-1)} > Y_{j'}^{(s)}) \Big\}, \end{aligned} \tag{2.13}$$

where the inequality holds owing to the positive covariance between any pair of order statistics in a sample (see Lehmann (1966)).

Similarly, it follows that

$$
\begin{aligned}
E\Big\{I(X_i^{(r)} > Y_j^{(s)})I(X_{i'}^{(r)} > Y_j^{(s)})\Big\}
&= EE\Big\{I(X_i^{(r)} > Y_j^{(s)})I(X_{i'}^{(r)} > Y_j^{(s)})\Big|Y_j^{(s)}\Big\} \\
&= E\Big[E\Big\{I(X_i^{(r)} > Y_j^{(s)})\Big|Y_j^{(s)}\Big\} \\
&\times E\Big\{I(X_{i'}^{(r)} > Y_j^{(s)})\Big|Y_j^{(s)}\Big\}\Big] \\
&= E\Big[E\Big\{I(W_{(i)}^{(r-1)} > Y_j^{(s)})\Big|Y_j^{(s)}\Big\} \\
&\times E\Big\{I(W_{(i')}^{(r-1)} > Y_j^{(s)})\Big|Y_j^{(s)}\Big\}\Big] \\
&\le EE\Big\{I(W_{(i)}^{(r-1)} > Y_j^{(s)})I(W_{(i')}^{(r-1)} > Y_j^{(s)})\Big|Y_j^{(s)}\Big\} \\
&= E\Big\{I(W_{(i)}^{(r-1)} > Y_j^{(s)})I(W_{(i')}^{(r-1)} > Y_j^{(s)})\Big\}. \qquad (2.14)
\end{aligned}
$$

In addition,

$$
\begin{aligned}
E\Big\{I(X_i^{(r)} > Y_j^{(s)})I(X_i^{(r)} > Y_{j'}^{(s)})\Big\}
&= EE\Big\{I(X_i^{(r)} > Y_j^{(s)})I(X_i^{(r)} > Y_{j'}^{(s)})\Big|Y_j^{(s)}, Y_{j'}^{(s)}\Big\} \\
&= EE\Big\{I(W_{(i)}^{(r-1)} > Y_j^{(s)})I(W_{(i)}^{(r-1)} > Y_{j'}^{(s)})\Big|Y_j^{(s)}, Y_{j'}^{(s)}\Big\} \\
&= E\Big\{I(W_{(i)}^{(r-1)} > Y_j^{(s)})I(W_{(i)}^{(r-1)} > Y_{j'}^{(s)})\Big\}, \qquad (2.15)
\end{aligned}
$$

and

$$
\begin{aligned}
E\Big\{I(X_i^{(r)} > Y_j^{(s)})\Big\} &= EE\Big\{I(X_i^{(r)} > Y_j^{(s)})\Big|Y_j^{(s)}\Big\} \\
&= EE\Big\{I(W_{(i)}^{(r-1)} > Y_j^{(s)})\Big|Y_j^{(s)}\Big\} \\
&= E\Big\{I(W_{(i)}^{(r-1)} > Y_j^{(s)})\Big\}. \qquad (2.16)
\end{aligned}
$$

Putting (2.12)–(2.16) together, we get

$$\begin{aligned} m^2n^2E(\hat{\theta}^2_{r,s}) \leq E\Bigg\{ & \sum_{i\neq i'=1}^{m}\sum_{j\neq j'=1}^{n} I(W_{(i)}^{(r-1)} > Y_j^{(s)})I(W_{(i')}^{(r-1)} > Y_{j'}^{(s)}) \\ & + \sum_{i=1}^{m}\sum_{j\neq j'=1}^{n} I(W_{(i)}^{(r-1)} > Y_j^{(s)})I(W_{(i)}^{(r-1)} > Y_{j'}^{(s)}) \\ & + \sum_{j=1}^{n}\sum_{i\neq i'=1}^{m} I(W_{(i)}^{(r-1)} > Y_j^{(s)})I(W_{(i')}^{(r-1)} > Y_j^{(s)}) \\ & + \sum_{i=1}^{m}\sum_{j=1}^{n} I(W_{(i)}^{(r-1)} > Y_j^{(s)}) \Bigg\} = m^2n^2E(\hat{\theta}^2_{r-1,s}). \end{aligned}$$

This implies that $Var(\hat{\theta}_{r,s}) \leq Var(\hat{\theta}_{r-1,s})$ because $\hat{\theta}_{r,s}$ is unbiased for any $r, s \geq 1$. A similar argument proves the second part. □

The above theoretical development assumes perfect rankings. It is possible to obtain some results in the imperfect ranking situation. Suppose the ranking mechanism is such that

$$\frac{1}{m}\sum_{i=1}^{m}\tilde{f}_i^{(r)}(x) = f(x)$$

and

$$\frac{1}{n}\sum_{j=1}^{n}\tilde{g}_j^{(s)}(y) = g(y),$$

where $\tilde{f}_i^{(r)}$ and $\tilde{g}_j^{(s)}$ are the density functions of the multistage judgment order statistics drawn from the two populations. Then one can simply verify that Propositions 2.1 and 2.3 still hold. However, it may not be an easy job to prove Proposition 2.4 in this setup. In the next section, effect of the ranking errors is assessed using Monte Carlo simulations.

3. NUMERICAL RESULTS

This section reports results of simulation studies carried out to compare the performances of $\hat{\theta}$ and $\hat{\theta}_{r,s}$. It is assumed that both populations follow normal, exponential or uniform distribution. Suppose X and $Y - \mu$ are standard normal random variables. Then, it is simply shown that

$$\theta = \Phi\left(\frac{-\mu}{\sqrt{2}}\right),$$

where $\Phi(.)$ is the distribution function of X. Similarly, for standard exponential random variables X and Y/α, we have

$$\theta = \frac{1}{1+\alpha}.$$

Finally, let X and Y/β be uniformly distributed on the unit interval. Then, it follows that

$$\theta = \begin{cases} 1-\beta/2 & 0<\beta<1 \\ 1/(2\beta) & \beta \geq 1 \end{cases}.$$

Under each parent distribution, three values were assigned to the associated parameter so as to produce $\theta = 0.25, 0.5, 0.75$ which are referred to as cases A, B and C, respectively. The appropriate parameter values are given in Table 1. Also, sample sizes $(m,n) \in \{(3,3),(4,4),(5,5)\}$ and stage numbers $(r,s) \in \{(1,1),$ $(2,2),(2,4),(3,3),(4,4),(4,6),(5,5)\}$ were selected.

Table 1: Parameter values corresponding to case A, B and C.

Parameter	A	B	C
μ	0.95387	0	-0.95387
α	3	1	1/3
β	2	1	1/2

We assume that the ranking the variables of interest X and Y are done based on concomitant variables $\mathcal{X}$ and $\mathcal{Y}$ which are related according to equations

$$\mathcal{X} = \rho_1\left(\frac{X-\mu_x}{\sigma_x}\right) + \sqrt{1-\rho_1^2}Z_1$$

and

$$\mathcal{Y} = \rho_2\left(\frac{Y-\mu_y}{\sigma_y}\right) + \sqrt{1-\rho_2^2}Z_2,$$

where $\rho_i \in [0,1]$ $(i=1,2)$, and Z_1 (Z_2) is a standard normal random variable independent from X (Y). Moreover, Z_1 and Z_2 are independent. The quality of rankings are controlled by the parameter ρ_i's. It is easy to see that $Corr(X,\mathcal{X}) = \rho_1$ and $Corr(Y,\mathcal{Y}) = \rho_2$. The chosen values of (ρ_1,ρ_2) are $(1,1)$ for perfect rankings of X and Y, $(1,0.8)$ for perfect ranking of X and fairly accurate ranking of Y, and $(0.8,0.8)$ for fairly accurate rankings of X and Y.

For each combination of distribution, sample sizes and correlations, 5,000 pairs of samples were generated in SRS and MSRSS (with the aforesaid stage numbers). The two estimators were computed from each pair of samples, and their variances were determined. The relative efficiency (RE) is defined as the ratio of $\widehat{Var}(\hat{\theta})$ to $\widehat{Var}(\hat{\theta}_{r,s})$. The RE values larger than one indicate that $\hat{\theta}_{r,s}$ is more efficient than $\hat{\theta}$. Tables 2–4 display the results.

Table 2: Estimated REs for different sample sizes and stage numbers under normal distribution.

(m,n)	(r,s)	$(\rho_1,\rho_2)=(1,1)$			$(\rho_1,\rho_2)=(1,0.8)$			$(\rho_1,\rho_2)=(0.8,0.8)$		
		A	B	C	A	B	C	A	B	C
(3,3)	(1,1)	1.719	1.860	1.737	1.613	1.546	1.408	1.425	1.402	1.262
	(2,2)	2.281	2.547	2.255	1.963	1.778	1.482	1.582	1.469	1.311
	(2,4)	2.623	3.284	2.599	2.127	1.906	1.460	1.654	1.563	1.284
	(3,3)	2.662	3.323	2.724	2.391	1.996	1.527	1.766	1.584	1.355
	(4,4)	3.078	4.034	3.052	2.535	2.149	1.697	1.880	1.685	1.410
	(4,6)	3.295	4.325	3.155	2.468	2.223	1.664	1.778	1.692	1.470
	(5,5)	3.291	4.435	3.304	2.641	2.224	1.634	1.834	1.652	1.423
(4,4)	(1,1)	2.141	2.334	2.118	1.847	1.721	1.493	1.526	1.461	1.302
	(2,2)	3.006	3.760	2.959	2.421	2.152	1.703	1.755	1.678	1.442
	(2,4)	3.626	4.559	3.534	2.598	2.275	1.766	1.910	1.738	1.502
	(3,3)	3.911	5.064	3.952	2.757	2.401	1.787	1.934	1.790	1.517
	(4,4)	4.440	6.059	4.389	3.125	2.669	1.948	2.046	1.818	1.596
	(4,6)	4.698	6.641	4.638	3.128	2.677	1.881	2.067	1.892	1.510
	(5,5)	4.625	6.685	4.666	3.259	2.793	2.038	2.019	1.877	1.594
(5,5)	(1,1)	2.458	2.813	2.501	2.083	1.840	1.591	1.645	1.551	1.368
	(2,2)	3.904	4.994	3.948	2.674	2.325	1.879	1.860	1.749	1.530
	(2,4)	4.902	6.412	4.825	3.145	2.683	1.944	2.102	1.886	1.604
	(3,3)	5.061	6.916	5.019	3.258	2.741	1.942	2.067	1.882	1.536
	(4,4)	6.071	8.783	6.079	3.415	2.925	2.080	2.156	2.014	1.680
	(4,6)	6.435	9.502	6.405	3.379	2.942	2.054	2.111	1.978	1.589
	(5,5)	6.627	10.050	6.726	3.768	3.162	2.166	2.261	2.042	1.641

Table 3: Estimated REs for different sample sizes and stage numbers under exponential distribution.

(m,n)	(r,s)	$(\rho_1,\rho_2)=(1,1)$			$(\rho_1,\rho_2)=(1,0.8)$			$(\rho_1,\rho_2)=(0.8,0.8)$		
		A	B	C	A	B	C	A	B	C
(3,3)	(1,1)	1.699	1.894	1.727	1.570	1.559	1.372	1.247	1.312	1.264
	(2,2)	2.310	2.724	2.279	1.798	1.853	1.544	1.267	1.421	1.323
	(2,4)	2.762	3.141	2.479	1.972	1.948	1.553	1.352	1.450	1.350
	(3,3)	2.733	3.420	2.667	2.067	2.120	1.643	1.313	1.504	1.379
	(4,4)	3.074	3.996	3.138	2.331	2.179	1.673	1.344	1.471	1.380
	(4,6)	3.418	4.262	3.191	2.296	2.214	1.667	1.428	1.548	1.394
	(5,5)	3.347	4.358	3.364	2.472	2.430	1.735	1.402	1.624	1.425
(4,4)	(1,1)	2.145	2.322	2.065	1.806	1.733	1.434	1.326	1.385	1.267
	(2,2)	3.102	3.811	3.101	2.371	2.152	1.711	1.407	1.536	1.396
	(2,4)	3.973	4.597	3.494	2.644	2.485	1.750	1.471	1.661	1.442
	(3,3)	3.832	5.034	3.880	2.589	2.476	1.810	1.393	1.557	1.443
	(4,4)	4.507	6.178	4.567	3.063	2.774	1.943	1.423	1.624	1.554
	(4,6)	5.282	7.028	4.795	2.991	2.789	1.983	1.469	1.673	1.601
	(5,5)	4.903	7.039	5.043	3.185	2.764	1.869	1.490	1.640	1.471
(5,5)	(1,1)	2.486	2.756	2.454	1.974	1.871	1.532	1.303	1.416	1.338
	(2,2)	3.961	4.954	4.081	2.896	2.606	1.829	1.513	1.712	1.518
	(2,4)	5.338	6.308	4.510	2.979	2.604	1.742	1.490	1.622	1.411
	(3,3)	5.468	7.232	5.453	3.347	2.765	1.780	1.545	1.716	1.466
	(4,4)	6.069	8.652	6.104	3.604	2.848	1.881	1.600	1.722	1.520
	(4,6)	7.227	9.814	6.458	3.607	3.057	1.960	1.541	1.747	1.545
	(5,5)	7.156	10.408	7.145	4.048	3.162	1.996	1.645	1.811	1.570

Table 4: Estimated REs for different sample sizes and stage numbers under uniform distribution.

(m,n)	(r,s)	$(\rho_1,\rho_2)=(1,1)$			$(\rho_1,\rho_2)=(1,0.8)$			$(\rho_1,\rho_2)=(0.8,0.8)$		
		A	B	C	A	B	C	A	B	C
(3,3)	(1,1)	1.718	1.813	1.684	1.638	1.665	1.429	1.401	1.489	1.371
	(2,2)	2.369	2.656	2.301	2.182	2.160	1.749	1.645	1.755	1.647
	(2,4)	2.973	3.210	2.529	2.117	2.344	1.914	1.592	1.846	1.725
	(3,3)	2.927	3.370	2.866	2.429	2.341	1.848	1.725	1.795	1.707
	(4,4)	3.272	3.913	3.350	2.869	2.678	1.926	1.808	1.984	1.713
	(4,6)	3.788	4.311	3.463	2.720	2.592	1.922	1.688	1.907	1.691
	(5,5)	3.625	4.348	3.690	2.942	2.764	2.045	1.793	1.994	1.817
(4,4)	(1,1)	2.024	2.298	2.030	1.903	1.916	1.653	1.422	1.613	1.524
	(2,2)	3.142	3.726	3.283	2.763	2.593	2.004	1.766	1.915	1.796
	(2,4)	4.435	4.670	3.410	2.839	2.808	2.199	1.734	2.037	1.993
	(3,3)	4.209	5.158	4.263	3.377	2.900	2.019	1.853	2.040	1.792
	(4,4)	4.832	5.916	4.810	4.204	3.368	2.216	2.068	2.209	1.906
	(4,6)	5.515	6.499	5.035	3.824	3.462	2.383	1.935	2.220	2.055
	(5,5)	5.351	6.774	5.462	4.097	3.378	2.281	1.958	2.240	1.959
(5,5)	(1,1)	2.375	2.806	2.328	2.317	2.196	1.722	1.649	1.805	1.577
	(2,2)	4.166	5.092	4.147	3.350	2.908	2.052	1.874	2.072	1.840
	(2,4)	6.162	6.384	4.527	3.504	3.238	2.251	1.910	2.189	1.928
	(3,3)	5.593	6.732	5.547	4.096	3.395	2.235	1.995	2.244	1.894
	(4,4)	6.934	8.888	6.965	4.683	3.794	2.365	1.962	2.305	2.011
	(4,6)	8.384	9.715	7.319	4.943	3.732	2.452	2.125	2.391	2.077
	(5,5)	8.206	10.226	8.064	5.590	4.256	2.535	2.200	2.508	2.144

It is observed that that MSRSS based estimator outperforms its SRS contender in all situations considered. Moreover, for any (m,n), the RE is increasing in both r and s, when the other factors are fixed. For example, compare entries for $m = n = 3$. In general, no comparison can be made between REs in two setups that one stage number is increased, and the other one is decreased. The efficiency gain could be substantial if the set sizes and stage numbers are large, e.g. when $m = n = r = s = 5$, the parent distribution is uniform, and the rankings are perfect. It is to be mentioned that when $(\rho_1, \rho_2) = (1,1)$, the REs for cases A and C are in good agreement (and smaller than that of case B) for all distributions and sample sizes, particularly when $r = s$. As expected, the REs diminish in the presence of ranking errors. The smallest values are obtained for $(\rho_1, \rho_2) = (0.8, 0.8)$.

4. APPLICATION TO REAL DATA

The MSRSS can be very efficient if the variable of interest is highly correlated to a concomitant variable. In this case, if the second variable can be measured with negligible cost, then we may use it in judgment ranking process (see Stokes (1977) for more details). In doing so, in step 2 of the RSS procedure, the elements of the ith sample are ordered according to the concomitant variable, and then study variable is actually measured for unit ranked ith smallest. The MSRSS case is treated similarly.

In this section, we illustrate the proposed procedure using a data set collected at the Australian Institute of Sport. It is made up of thirteen measured variables on 102 male and 100 female athletes[1]. We will consider lean body mass (LBM) and body mass index (BMI) for each athlete. The LBM is a component of body composition, calculated by subtracting body fat weight from total body weight. Exact measurement of the LBM is done using various technologies such as dual energy X-ray absorptiometry (DEXA) which is costly. On the other hand, the BMI is a well-accepted measure of obesity which is easy to calculate and readily accessible. A BMI value is is simply weight (in kg) divided by square of height (in m). The correlation coefficient between the two variables is 0.71. So, the BMI can serve as a concomitant variable.

Let X and Y be the LBM variable for the male and female populations, respectively. It is of interest to estimate $\theta = P(X > Y)$. For $m = n = 4$, 50,000 samples were drawn from the two hypothetical populations based on SRS and MSRSS (with $r = s = 1, 2$) designs. The sampling is done with replacement to ensure that the measured units are independent of each other. From each sample, the corresponding estimator was computed, and its variance was finally determined. The efficiencies of $\hat{\theta}_{1,1}$ and $\hat{\theta}_{2,2}$ relative to $\hat{\theta}$ are estimated as 1.193 and 1.275, respectively. As expected, the SRS estimator is outperformed by its RSS and DRSS versions. It is to be noted that the RE values are not much bigger than unity. This may root in the relatively low correlation of 0.71 between the variable of interest and the concomitant variable.

5. CONCLUSION

The RSS design is known to be a viable alternate to the usual SRS in situations that cost-efficiency is of high importance. It employs auxiliary information to direct attention toward the actual measurement of more representative units in the population under study. The success of RSS largely depends on the quality of ranking process. Since judgment ranking on large sets of units is prone to errors, the set size is chosen small in practice. The MSRSS allows to construct more efficient procedures by increasing the number of stages rather that the set size.

This article deals with reliability estimation for the stress-strength model using MSRSS. A nonparametric estimator is presented, and shown to be unbiased

[1] The data set can be found at `http://www.statsci.org/data/oz/ais.html`

with smaller variance as compared with the usual estimator in SRS. It is further proved that the estimator becomes more efficient by increasing the number of stages for ranked set samples drawn from the two populations. Results of simulation studies support the mathematical findings. An application to a real data set clarifies how judgment ranking can be implemented using a concomitant variable.

ACKNOWLEDGMENTS

The authors are grateful to the reviewer and the Associate Editor for their comments that have largely contributed to improve the original manuscript.

REFERENCES

[1] Al-Saleh, M.F. and Al-Kadiri, M. (2000). Double-ranked set sampling, *Statistics and Probability Letters*, **48**, 205–212.

[2] Al-Saleh, M.F. and Al-Omari, A.I. (2002). Multistage ranked set sampling, *Journal of Statistical Planning and Inference*, **102**, 273–286.

[3] Al-Saleh, M.F. and Samuh, M.J. (2008). On multistage ranked set sampling for distribution and median estimation, *Computational Statistics and Data Analysis*, **52**, 2066–2078.

[4] Chen, Z.; Bai, Z. and Sinha, B.K. (2004). *Ranked set sampling: Theory and Applications*, Springer, New York.

[5] Kotz, S.; Lumelskii, Y. and Pensky, M. (2003). *The stress-strength model and its generalizations. Theory and applications*, World Scientific, Singapore.

[6] Lehmann, E.L. (1966). Some concepts of dependence, *The Annals of Mathematical Statistics*, **37**, 1137–1153.

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

[8] Sengupta, S. and Mukhuti, S. (2008). Unbiased estimation of $P(X > Y)$ using ranked set sample data, *Statistics*, **42**, 223–230.

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

REVSTAT – Statistical Journal
Volume 15, Number 4, October 2017, 583–600

NON PARAMETRIC *ROC* SUMMARY STATISTICS

Authors: M.C. Pardo
– Department of Statistics and O.R. (I), Complutense University of Madrid, 28040-Madrid, Spain

A.M. Franco-Pereira
– Department of Statistics and O.R. (I), Complutense University of Madrid, 28040-Madrid, Spain

Received: November 2015 Revised: February 2016 Accepted: April 2016

Abstract:

- Receiver operating characteristic (ROC) curves are useful statistical tools for medical diagnostic testing. It has been proved its capability to assess diagnostic marker's ability to distinguish between healthy and diseased subjects and to compare different diagnostic markers. In this paper we introduce non parametric ROC summary statistics to assess a ROC curve across the entire range of $FPFs \in (0,1)$ as well as over a restricted range of $FPFs$ and compare them with some existing ones through a simulation study and through some real data examples. We also show their capability to compare two diagnostic markers.

Key-Words:

- *receiver operating characteristic; non parametric methods; diagnostic marker.*

1. INTRODUCTION

In a diagnostic setting, the performance of any continuous diagnostic marker is primarily assessed through the receiver operating characteristic (ROC) curve and the area under the ROC curve (AUC). The ROC curve is a plot of the sensitivity (the probability that the marker will be above a given threshold for the diseased subjects) against 1−specificity (the specificity being the probability that the marker will be below the threshold for the healthy subjects) or, equivalently, of the true positive fraction (TPF) against false positive fraction (FPF). Using a threshold c,

$$ROC(\cdot) = \{(FPF(c), TPF(c)), c \in (-\infty, \infty)\}.$$

The AUC is a summary measure of the sensitivity and specificity over the range of thresholds. Because of the AUC is scale free, ranging between 0.5 and 1, this measure provides a natural common scale for comparing the different markers regardless of their measurement scale. The ROC curve essentially provides a distribution-free description of the separation between the distributions of diseased and healthy subjects. Therefore, each of the summary measures is, in a sense, a summary of the distance between these two distributions. In fact, the empirical estimator of the AUC is equivalent to the Mann-Whitney U-statistic, thus representing the probability that a subject, randomly selected among the diseased, shows a marker value higher than a subject randomly extracted from the healthy. Other summary measure is the maximum vertical distance between the ROC curve and the 45^o line, which is an indicator of how far the curve is from that of the uninformative test. It ranges from 0 for the uninformative test to 1 for an ideal test. This index is closely related to Kolmogorov-Smirnov measure of distance between two distributions ([7], [5]). Other test statistics such as Anderson-Darling, Neyman and Watson tests were studied in [16] to assess diagnostic markers. They conclude that Anderson-Darling test is more powerful than Kolmogorov-Smirnov test and it is a good alternative to AUC. However, it can not be written in terms of functionals of the empirical ROC curve and it does not have value itself. In this paper, we propose to measure the distance between the ROC curve and the 45^o line through their derivatives to assess the discriminatory ability of a biomarker. This approach is closely related to a nonparametric test for two sample problem based on an order statistic introduced in [1]. It does not have value itself since it is not bounded but it has a geometric interpretation in terms of the ROC curve.

When measurements on two diagnostic markers A and B are available, the question of interest is which marker best discriminates between healthy and diseased subjects. Various methods have been proposed for comparing the performances of two diagnostic markers. See, for example [8], [11], [19], [23] and [5]. The most commonly approach to comparing ROC curves is to test the equality of their respective $AUCs$. The nonparametric version of the area test was developed in [9] and [10] for both unpaired and paired data. The test was refined in [6]. Two permutation tests for comparing paired ROC curves were proposed in [2] and [4]. However, when there is no uniform dominance between the involved curves, we can find different curves with the same AUC. Therefore, these tests are not valid to compare the equality among the ROC curves. In [21] it was developed a fully nonparametric test to compare two ROC curves when the data are paired and continuous. Later, [22] extended it for continuous unpaired data. In [16] it was suggested that the Anderson-Darling statistic can be viably used in comparing two diagnostic markers. Recently, [12] and [13] used the analogy between the ROC curve and the cumulative distribution function to propose a general methodology which allows us to use the traditional k-sample tests to the ROC curves comparison problem on unpaired and paired designs, respectively. Therefore, we propose, following [3], the difference between the values of our approach for each marker to compare ROC curves.

Although we focus primarily on comparing a ROC curve across the entire range of $FPFs \in (0,1)$, in practice, one might also be interested in a part of the ROC curve that is of primary interest. For example, in screening studies, $FPFs$ must be kept very low and so the ROC curve over a restricted range of $FPFs$ may be of interest. If $FPFs$ in the range $(0, t_0)$ is of interest, the value of partial ROC analysis has been recognized. [14] and [15] proposed a method for comparing a portion of ROC curves when binomial is appropriate. [24] present a nonparametric method for the analysis of partial ROC curves. Recently, [18] contruct nonparametric confidence intervals for the partial AUC. However, to our knowledge neither of the above approaches used to evaluate the whole ROC curve based on two-sample tests, have been extended to evaluate the ROC curve over a specific range. In order to fill this gap, we extend our summary statistic to evaluate ROC curves over a range of $FPFs$ of interest.

This paper is organized as follows: in Section 2 a new ROC summary statistic which can be written as a nonparametric test based on spacings is provided as well as its partial counterpart. In Section 3 its statistical power is investigated in extensive simulations and compared with that of the standard test on AUC and the Anderson Darling test. Furthermore, the performance of the difference of our ROC summary statistic for each marker for comparing ROC curves is studied across the entire as well as restricted range of $FPFs$. In Section 4 the new proposed method is applied to two real data sets. Finally, in Section 5, we make some concluding remarks.

2. THE NEW *ROC* SUMMARY STATISTICS

Some *ROC* summary measures are based on evaluating geometrically the distance between the *ROC* curve and the 45^o line, which is an indicator of how far the curve is from that of an uninformative marker. For example, considering the area between them or the maximum vertical distance between them, we obtain the well-known *AUC* or the Kolmogorov-Smirnov index, respectively. However, to our knowledge, the distance between the derivatives of these two functions has not been explored as a *ROC* summary statistic. Therefore, our proposal is to take into account that the *ROC* of a noninformative marker verifies

$$\frac{dROC(t)}{dt} = \lim_{\Delta t \to 0} \frac{ROC(t+\Delta t) - ROC(t)}{\Delta t} = 1$$

and to define as a summary statistic the sum of the squared differences between an approach of the derivative of the *ROC* curve

$$\frac{ROC(t) - ROC(t - \frac{1}{N})}{\frac{1}{N}}, \text{ for } N \text{ big enough,}$$

and the derivative of $y = x$, which is 1, for a number N of equidistant points

$$\sum_{k=1}^{N} \Big(\frac{ROC(\frac{k}{N}) - ROC(\frac{k-1}{N})}{\frac{1}{N}} - 1 \Big)^2 .$$

In particular, we propose to consider $N = 1 + n_{\overline{D}}$ where $n_{\overline{D}}$ is the number of healthy subjects and to define

$$\eta = \sum_{k=1}^{n_{\overline{D}}+1} \left(\left(ROC\left(\frac{k}{n_{\overline{D}}+1}\right) - ROC\left(\frac{k-1}{n_{\overline{D}}+1}\right) \right) - \frac{1}{n_{\overline{D}}+1} \right)^2 .$$

Note that the value of this summary statistic is not worthwhile by itself but it can be used to test if a biomarker is discriminatory of healthy and diseased individuals.

Let $\left\{ Y_{\overline{D}_i}, i = 1, ..., n_{\overline{D}} \right\}$ be an i.i.d. sample of a continuous distribution F representing $n_{\overline{D}}$ measurements of healthy subjects and let $\left\{ Y_{D_j}, j = 1, ..., n_D \right\}$ be an i.i.d. sample of a continuous distribution G representing n_D measurements of diseased subjects. It is common in the *ROC* methodology to assume that diseased subjects tend to have higher measurements than healthy subjects.

The empirical estimator of the *ROC* curve simply applies the definition of the *ROC* curve to the observed data. Thus, for each possible cut-point c, the empirical true and false positive fractions are calculated as follows:

$$\widehat{TPF}(c) = \frac{\sum_{i=1}^{n_D} I\left(Y_{D_i} \geq c\right)}{n_D}$$

$$\widehat{FPF}(c) = \frac{\sum_{j=1}^{n_{\overline{D}}} I\left(Y_{\overline{D}_j} \geq c\right)}{n_{\overline{D}}}.$$

The empirical ROC curve is a plot of $\widehat{TPF}(c)$ versus $\widehat{FPF}(c)$ for all $c \in (-\infty, \infty)$. Equivalently, the empirical ROC can be written as

$$\widehat{ROC}(t) = \widehat{TPF}\left(\widehat{FPF}^{-1}(t)\right), \quad t \in (0,1).$$

Let $-\infty = Y_{\overline{D}_{(0)}} \leq Y_{\overline{D}_{(1)}} \leq Y_{\overline{D}_{(2)}} \leq ... \leq Y_{\overline{D}_{(n_{\overline{D}})}} \leq Y_{\overline{D}_{(n_{\overline{D}}+1)}} = \infty$ be the order statistics constructed from $\left\{Y_{\overline{D}_j}, j = 1, ..., n_{\overline{D}}\right\}$. Therefore, an estimator of η can be obtained replacing ROC by its empirical estimator:

$$\widehat{\eta} = \sum_{k=1}^{n_{\overline{D}}+1} \left(\left(\widehat{ROC}\left(\frac{k}{n_{\overline{D}}+1}\right) - \widehat{ROC}\left(\frac{k-1}{n_{\overline{D}}+1}\right)\right) - \frac{1}{n_{\overline{D}}+1}\right)^2.$$

Note that this index is the sum of squared errors between the jump of the ROC curve evaluated in two equidistant points and the distance between these two equidistant points. The value 0 means to be a noninformative test. Furthermore, this index, $\widehat{\eta}$, is closely related to the nonparametric test for a two sample problem based on order statistics proposed in [1]. Indeed, we see that

$$\widehat{ROC}\left(\frac{k}{n_{\overline{D}}+1}\right) = \widehat{TPF}\left(\widehat{FPF}^{-1}\left(\left(\frac{k}{n_{\overline{D}}+1}\right)\right)\right)$$

so first we look for a value v such as

$$\widehat{FPF}(v) = \frac{k}{n_{\overline{D}}+1}$$

or equivalently,

$$\sum_{j=1}^{n_{\overline{D}}+1} I\left(Y_{\overline{D}_j} \geq v\right) = k$$

so $v = Y_{\overline{D}_{(n_{\overline{D}}-k+1)}}$. Therefore,

$$\widehat{ROC}\left(\frac{k}{n_{\overline{D}}+1}\right) = \frac{\sum_{i=1}^{n_D} I\left(Y_{D_i} \geq Y_{\overline{D}_{(n_{\overline{D}}-k+1)}}\right)}{n_D}.$$

In a similar way,

$$\widehat{ROC}\left(\frac{k-1}{n_{\overline{D}}+1}\right) = \frac{\sum\limits_{i=1}^{n_D} I\left(Y_{D_i} \geq Y_{\overline{D}_{(n_{\overline{D}}-k+2)}}\right)}{n_D}.$$

Finally,

$$\widehat{\eta} = \sum_{k=1}^{n_{\overline{D}}+1}\left(\frac{\sum\limits_{i=1}^{n_D}\xi_k^i}{n_D} - \frac{1}{n_{\overline{D}}+1}\right)^2$$

where

$$\xi_k^i = \begin{cases} 1, & Y_{D_i} \in \Delta_k \\ 0, & Y_{D_i} \notin \Delta_k \end{cases} \quad \text{for } k = 1, ..., n_{\overline{D}}+1, i = 1, ..., n_D,$$

with $\Delta_k = \left[Y_{\overline{D}_{(n_{\overline{D}}-k+1)}}, Y_{\overline{D}_{(n_{\overline{D}}-k+2)}}\right)$, is the test statistic proposed in [1]. They obtained its exact distribution that can be seen in Theorem 1.

If $FPFs$ in the range $(0, t_0)$ is of interest, the partial $\widehat{\eta}$ can be similarly defined as

(2.1)

$$\widehat{\eta}_p(t_0) = \sum_{1\leq k\leq [t_0(n_{\overline{D}}+1)]}\sum_{k=1}^{n_{\overline{D}}+1}\left(\left(\widehat{ROC}\left(\frac{k}{n_{\overline{D}}+1}\right) - \widehat{ROC}\left(\frac{k-1}{n_{\overline{D}}+1}\right)\right) - \frac{1}{n_{\overline{D}}+1}\right)^2$$

where $[\cdot]$ denotes the integer part of $\cdot$.

In the following section we evaluate the performance of $\widehat{\eta}$ and compare it to the ordinary nonparametric ROC test $\widehat{AUC}$ given by

$$\widehat{AUC} = \frac{\sum_{i=1}^{n_{\overline{D}}}\sum_{j=1}^{n_D} I(Y_{\overline{D}_i} < Y_{D_j})}{n_D n_{\overline{D}}}.$$

and the Anderson-Darling test of uniformity of the distribution of the false positive fraction, proposed in [16] (AD) to assess one diagnostic marker.

On the other hand, the test statistic

$$T = \frac{\widehat{AUC}_A - \widehat{AUC}_B}{\sqrt{var(\widehat{AUC}_A) + var(\widehat{AUC}_B) - 2covar(\widehat{AUC}_A, \widehat{AUC}_B)}},$$

proposed by [6] and $\Delta Z = Z_B - Z_A$, for $Z_L = \hat{\eta}, AD$, where $L = A, B$, indicates the value of the test statistic for biomarker A or B, are compared to assess two biomarkers. Finally, the partial summary measure $\widehat{\eta}_p(t_0)$ is compared to the partial AUC, $pAUC(t_0)$, via bootstrap.

3. SIMULATION STUDIES

Firstly, simulations are conducted to assess the performance of the new *ROC* summary statistic $\widehat{\eta}$, to evaluate one marker. We have compared the power of our statistic $\widehat{\eta}$ with $\widehat{AUC}$ and AD.

Table 1 compares the power of $\widehat{\eta}$ obtained when the exact distribution studied in [1] is used to obtain the critical values and when 1,000 Monte Carlo replicates are used instead. Due to the relatively large computational time required for the implementation of the exact procedure, the comparisons presented here are limited to small samples ($n_D = n_{\overline{D}} = 15$). However, even with these small samples, there is a good agreement between the exact and simulated test. Thus, for the large sample sizes as presented in the subsequent tables, we calculate only the power of the simulated test since the results for the exact test should be essentially the same.

Table 1: Comparison of the power of $\widehat{\eta}$ obtained using the exact distribution and the one obtained via 1,000 independent Monte Carlo simulations, for $n_D = n_{\overline{D}} = 15$. Healthy subjects follow (from left to right) a $N(0,1)$, $\Gamma(1/2,1/2)$ or $LN(0,1)$ distribution while diseased subjects are sampled from G.

G	Exact	MC	G	Exact	MC	G	Exact	MC
$N(0.3,1)$	0.074	0.074	$\Gamma(2,1)$	0.987	0.986	$LN(1.275,0.5)$	0.793	0.771
$N(0.3,1.4^2)$	0.133	0.119	$\Gamma(4,1)$	1.000	1.000	$LN(0,3/2)$	0.056	0.050
$N(0.3,0.3^2)$	0.981	0.977	$\Gamma(4.3,4)$	1.000	1.000	$LN(0.7,0.2)$	0.894	0.894
$N(0,1.4^2)$	0.117	0.111	$\Gamma(1/8,1/8)$	0.844	0.830	$LN(-3/2,2)$	0.576	0.540
$N(0,0.3^2)$	0.973	0.969	$\Gamma(4,4)$	1.000	1.000	$LN(1/4,1/2)$	0.279	0.258

For 1,000 independent simulations, one-sided tests were conducted at level $\alpha = 0.05$ to compare the $\widehat{AUC}$, AD and $\widehat{\eta}$ tests. To determine appropriate critical values we have carried out Monte Carlo simulation with $M = 5,000$ replicates. The type I error values are not presented as they are all around 0.05 but they can be provided by the authors upon request. Tables 2–4 compare the proportion of rejections (power) for different pairs of distributions for diseased and healthy subjects. These three tables distinguish three different distributions for the markers: Normal, Gamma and Lognormal, respectively. The markers for the healthy subjects are generated from a N(0,1), Gamma(1/2,1/2) and LN(0,1), respectively while the markers for the diseased subjects are generated from five different alternatives each one. These alternatives have been considered taking into account all the possible combinations changing the location and shape of the distribution of the diseased subjects in relation to the healthy subjects. Some of the probability distribution functions and their corresponding ROC curves for Table 2 can be seen in Figure 1.

Table 2: Power based on 1,000 independent simulations of Normal random variables. Healthy subjects follow a $N(0,1)$ distribution while diseased subjects are sampled from G.

G	Test	$n_D = n_{\overline{D}}$			
		15	30	50	100
$N(0.3,1)$	$\widehat{AUC}$	0.118	0.216	0.319	0.539
	$\widehat{\eta}$	0.067	0.098	0.087	0.096
	AD	0.097	0.145	0.270	0.489
$N(0.3,1.4^2)$	$\widehat{AUC}$	0.104	0.156	0.202	0.438
	$\widehat{\eta}$	0.057	0.098	0.156	0.347
	AD	0.112	0.188	0.284	0.650
$N(0.3,0.3^2)$	$\widehat{AUC}$	0.210	0.332	0.520	0.801
	$\widehat{\eta}$	0.626	0.817	0.929	0.993
	AD	0.059	0.119	0.254	0.676
$N(0,1.4^2)$	$\widehat{AUC}$	0.057	0.068	0.053	0.058
	$\widehat{\eta}$	0.049	0.060	0.108	0.220
	AD	0.089	0.101	0.091	0.275
$N(0,0.3^2)$	$\widehat{AUC}$	0.076	0.065	0.060	0.048
	$\widehat{\eta}$	0.590	0.778	0.909	0.992
	AD	0.036	0.113	0.208	0.622

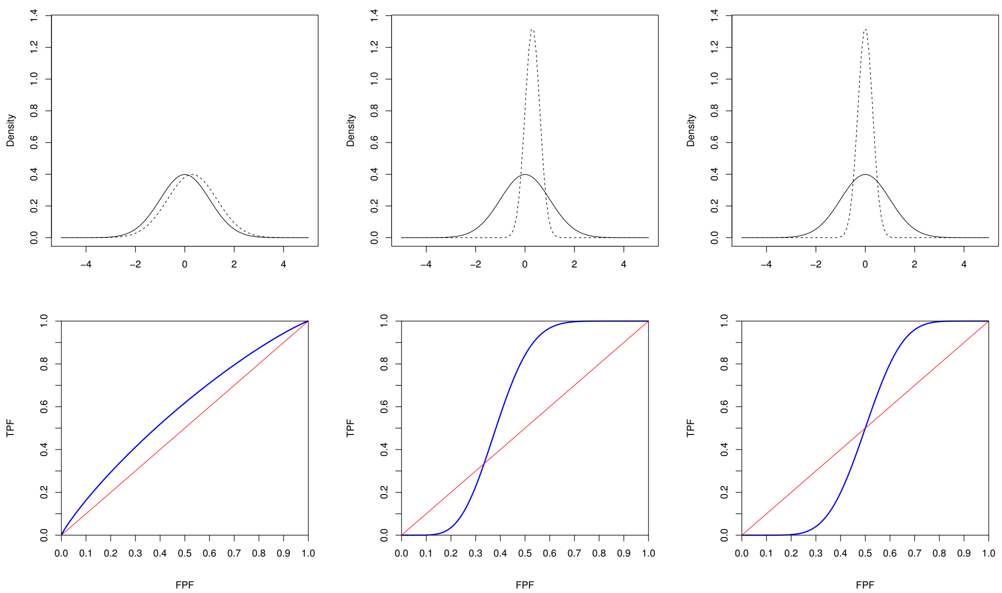


Figure 1: Probability distribution functions and their corresponding *ROC* curves ($n_D = n_{\overline{D}} = 100$) for some cases described in Table 2. From left to right: $N(0,1)$ versus $N(0.3,1)$, $N(0.3,0.3^2)$ and $N(0,0.3^2)$, respectively.

Table 3: Power based on 1,000 independent simulations of Gamma random variables. Healthy subjects follow a $\Gamma(1/2, 1/2)$ distribution while diseased subjects are sampled from G.

G	Test	$n_D = n_{\overline{D}}$			
		15	30	50	100
$\Gamma(2,1)$	$\widehat{AUC}$	0.755	0.958	0.998	1.000
	$\widehat{\eta}$	0.393	0.601	0.782	0.926
	AD	0.147	0.441	0.708	0.977
$\Gamma(4,1)$	$\widehat{AUC}$	0.999	1.000	1.000	1.000
	$\widehat{\eta}$	0.885	0.989	1.000	1.000
	AD	0.129	0.560	0.789	0.991
$\Gamma(4.3,4)$	$\widehat{AUC}$	0.363	0.635	0.817	0.981
	$\widehat{\eta}$	0.578	0.750	0.907	0.995
	AD	0.063	0.158	0.307	0.738
$\Gamma(1/8,1/8)$	$\widehat{AUC}$	0.522	0.796	0.953	1.000
	$\widehat{\eta}$	0.416	0.753	0.943	1.000
	AD	0.617	0.903	0.995	1.000
$\Gamma(4,4)$	$\widehat{AUC}$	0.329	0.507	0.754	0.964
	$\widehat{\eta}$	0.573	0.721	0.887	0.996
	AD	0.049	0.129	0.319	0.742

Table 4: Power based on 1,000 independent simulations of LogNormal random variables. Healthy subjects follow a $LN(0, 1)$ distribution while diseased subjects are sampled from G.

G	Test	$n_D = n_{\overline{D}}$			
		15	30	50	100
$LN(1.275, 0.5)$	$\widehat{AUC}$	0.976	1.000	1.000	1.000
	$\widehat{\eta}$	0.734	0.946	0.995	0.999
	AD	0.133	0.424	0.744	0.985
$LN(0, 3/2)$	$\widehat{AUC}$	0.055	0.064	0.051	0.051
	$\widehat{\eta}$	0.058	0.083	0.182	0.325
	AD	0.089	0.106	0.174	0.420
$LN(0.7, 0.2)$	$\widehat{AUC}$	0.706	0.922	0.999	1.000
	$\widehat{\eta}$	0.883	0.987	1.000	1.000
	AD	0.038	0.096	0.198	0.541
$LN(-3/2, 2)$	$\widehat{AUC}$	0.668	0.941	0.995	1.000
	$\widehat{\eta}$	0.536	0.895	0.993	1.000
	AD	0.711	0.976	0.999	1.000
$LN(1/4, 1/2)$	$\widehat{AUC}$	0.132	0.224	0.334	0.584
	$\widehat{\eta}$	0.270	0.364	0.517	0.718
	AD	0.067	0.159	0.283	0.688

Every table follows the same pattern: in the first three designs the mean of the diseased subjects is larger than the mean corresponding to the healthy ones and in the last two designs it does not change. In the first design, the standard deviation of both groups of patients is the same, in the second and forth one the standard deviation of the diseased subjects is larger, and in the third and fifth one the standard deviation of the diseased subject is smaller.

As [16] already observed, these results reveal that the $\widehat{AUC}$ test is more powerful when location differences between the distributions under consideration are primarily involved. However, in our study, although the mean increases if the standard deviation decreases (design 3) our procedure has higher power than the others. When scale differences are prominent, $\widehat{AUC}$ test is incapable of discriminating between these distributions. In particular, when the standard deviation of the distribution of the healthy subjects is larger than that for the diseased subjects, the new measure $\widehat{\eta}$ is significantly better than $\widehat{AUC}$ and AD tests. If the standard deviation of the distribution of the healthy subjects is smaller than that for the diseased subjects, the AD test is preferable to the others. Therefore, our procedure is the best when the standard deviation of the distribution of the healthy subjects is larger than that for the diseased subjects independently that the location of the distribution of the diseased subjects changes or not (designs 3 and 5). These two designs as can be seen in Figure 1 correspond to ROC curves crossing the diagonal reference line. Moreover, for the other designs it is the second best except for designs 1 and 2 in Table 2. Although $\widehat{AUC}$ test is preferable when location differences between the distributions under consideration are primarily involved, we must not use it in the other situations. Finally, AD test has only slight high power than our procedure in one of the five considered designs.

3.1. Assessment of two diagnostic markers

We compare the performance of our test statistic, $\Delta\widehat{\eta}$, to that of [6], T, and Anderson Darling approach, ΔAD, via simulation. In [16] it was concluded that the DeLong test is in general more powerful than the Anderson-Darling approach to assess two diagnostic markers, particularly when the correlation between measurements is substantial. In the simulations to obtain the distributions of AD test and our test we have used bootstrap following [3].

We perform simulations to investigate the empirical power for different underlying AUCs, correlations between the markers ($\rho = 0, 0.5$) and different simple sizes ($n_D = n_{\overline{D}} = 20, 40, 80$) at level $\alpha = 0.05$. In these simulations, the marker values of the healthy subjects were generated from a standard normal distribution and those of the diseased subjects from $N(\mu_A, \sigma_A^2 = 1)$ and $N(\mu_B, \sigma_B^2)$ for markers A and B, respectively. The uniform alternative (where one curve is uniformly above the other) occurs when $\sigma_A^2 = \sigma_B^2$ and the crossing alternative (when the two curves cross) when $4\sigma_A^2 = \sigma_B^2$. For each considered scenario, 1000 replications were used. The different scenarios are that considered in [21].

For equal $AUCs$ arising from crossing ROC curves, the power of our test is the highest as can be seen in Table 5 and the use of the T test is inappropriate. On the other hand, highly correlated biomarkers lead to increase power. For non-crossing ROC curves, the power of T test is the highest as can be seen in Table 6. The power of $\Delta\widehat{\eta}$ is higher than the power of ΔAD.

Table 5: Power against crossing alternatives.

AUC_A	AUC_B	$n_D = n_{\overline{D}}$	T		$\Delta\widehat{\eta}$		ΔAD	
			$\rho = 0$	$\rho = 0.5$	$\rho = 0$	$\rho = 0.5$	$\rho = 0$	$\rho = 0.5$
0.6	0.6	20	0.046	0.066	0.014	0.019	0.015	0.018
		40	0.063	0.048	0.088	0.075	0.072	0.073
		80	0.046	0.049	0.327	0.338	0.162	0.190
0.7	0.7	20	0.055	0.046	0.042	0.025	0.040	0.039
		40	0.057	0.045	0.132	0.148	0.095	0.127
		80	0.039	0.043	0.421	0.478	0.092	0.136
0.8	0.8	20	0.054	0.037	0.069	0.074	0.061	0.062
		40	0.061	0.051	0.193	0.209	0.110	0.155
		80	0.051	0.053	0.475	0.577	0.111	0.155
0.9	0.9	20	0.044	0.038	0.092	0.108	0.070	0.095
		40	0.049	0.041	0.195	0.218	0.147	0.157
		80	0.056	0.054	0.459	0.541	0.232	0.277

Table 6: Power against uniform alternatives.

AUC_A	AUC_B	$n_D = n_{\overline{D}}$	T		$\Delta\widehat{\eta}$		ΔAD	
			$\rho = 0$	$\rho = 0.5$	$\rho = 0$	$\rho = 0.5$	$\rho = 0$	$\rho = 0.5$
0.6	0.7	20	0.116	0.207	0.018	0.013	0.002	0.003
		40	0.209	0.340	0.022	0.021	0.043	0.043
		80	0.368	0.626	0.054	0.055	0.214	0.264
0.6	0.8	20	0.410	0.623	0.105	0.123	0.005	0.014
		40	0.689	0.925	0.249	0.242	0.183	0.214
		80	0.951	0.998	0.542	0.554	0.675	0.779
0.6	0.9	20	0.821	0.967	0.481	0.511	0.013	0.010
		40	0.987	1.000	0.835	0.854	0.323	0.324
		80	1.000	1.000	0.982	0.995	0.857	0.877
0.7	0.8	20	0.140	0.219	0.066	0.057	0.007	0.003
		40	0.282	0.419	0.106	0.108	0.058	0.059
		80	0.443	0.709	0.222	0.222	0.176	0.232
0.7	0.9	20	0.561	0.766	0.350	0.391	0.012	0.014
		40	0.837	0.984	0.608	0.702	0.108	0.110
		80	0.991	1.000	0.878	0.933	0.395	0.434
0.8	0.9	20	0.210	0.283	0.163	0.181	0.012	0.013
		40	0.354	0.605	0.263	0.304	0.032	0.031
		80	0.688	0.888	0.473	0.526	0.075	0.092

In summary, the behavior of our test is the same as AD test studied in [16] and permutation tests introduced in [21]. That is to say, they have clearly superior power in Table 5 to T test but the power of ours is the highest. However, the power of the permutation test proposed in [2] is close to the nominal significance level suggesting that a rejection of the null hypothesis is unlikely to occur. On the other hand, for non-crossing ROC curves, T test is preferable although as sample size increases the power of $\Delta\widehat{\eta}$ is closer to the power of T test. Note that in most of the cases ΔAD test has very low power (see Table 6).

Finally, suppose one is only interested in some range of specificities. For example, acceptable specificities are high for early cancer detection tests. A lower specificity for a large population leads to many more falsely classified non-diseased subjects who may have to undergo a more invasive test subsequently. It is thus desired to compare screening markers at a higher range of specificities. The partial AUC, which summarizes part of the ROC curve in the range of desired specificities, uses to be a better alternative to T test. The value of partial ROC analysis has been recognized and several methods have been developed. See [14], [15], [20] and [17]. However, the methods for analysing partial ROC presented in these papers use a parametric approach which assumes the data have an underlying normal distribution.

We perform a new simulation to compare $pAUC$ and the proposed $\widehat{\eta}_p$, defined in (2.1), for crossing ROC curves only, since in those cases T test doesn't work properly and $pAUC$ is an alternative to focus on some range of interest. We consider two different ranges $(0, 0.4)$ and $(0, 0.8)$ although by brevity we only present the results for $t_0 = 0.4$ in Table 7.

Table 7: Power of the partial measures against crossing alternatives.

AUC_A	AUC_B	$n_D = n_{\overline{D}}$	$pAUC$		$\widehat{\eta}_p$	
			$\rho = 0$	$\rho = 0.5$	$\rho = 0$	$\rho = 0.5$
0.6	0.6	20	0.054	0.051	0.031	0.031
		40	0.140	0.175	0.071	0.078
		80	0.255	0.369	0.230	0.314
0.7	0.7	20	0.042	0.029	0.038	0.051
		40	0.104	0.117	0.136	0.147
		80	0.219	0.296	0.389	0.470
0.8	0.8	20	0.022	0.024	0.059	0.084
		40	0.072	0.086	0.207	0.244
		80	0.132	0.172	0.470	0.566
0.9	0.9	20	0.007	0.002	0.071	0.085
		40	0.032	0.023	0.202	0.247
		80	0.069	0.063	0.445	0.555

As can be seen in Table 7, $pAUC$ works better that its counterpart T with higher power in most of the cases. However, our proposed summary statistic $\widehat{\eta}_p$ works much better than $pAUC$ and similarly to its counterpart $\widehat{\eta}$. That is to say, the new partial summary statistic seems to be a good alternative.

4. REAL DATA EXAMPLES

4.1. Pancreatic cancer biomarker study

The first dataset studied has been used by various statisticians to illustrate statistical techniques for diagnostic tests. First published in [23], it is a case-control study ding 90 cases with pancreatic cancer and 51 controls that did not have cancer but who had pancreatitis. Serum samples from each patient were assayed for CA-125, a cancer antigen, and CA-19-9, a carbohydrate antigen, both of which are measured on a continuous positive scale. It can be assumed that both biomarkers are independent. A natural question is to determine which of the two markers best discriminates diseased from healthy subjects. See Figure 2 (a).

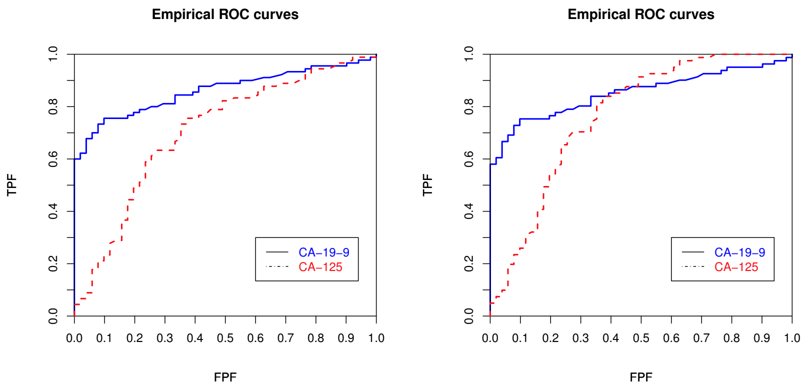


Figure 2: (a) Empirical ROC curves and (b) Empirical ROC curves once we have eliminated from the data those cases with the smallest values for the second biomarker.

The $\widehat{AUC}$ values are 0.861 and 0.706 for CA-125 (called biomarker A) and CA-19-9 (called biomarker B), respectively. The T statistic, which is based on the methodology described in [6] for paired data, is statistically significantly different from 0 ($p = 0.007$). The two differences $\Delta\widehat{\eta} = \widehat{\eta}_A - \widehat{\eta}_B$ and $\Delta AD = AD_A - AD_B$ are also statistically significantly different from 0 ($p = 0$ and $p = 0.015$, respectively). As [23], we have also focus our comparison on the range of $FPFs$ below 0.2 using the differences of the partial measures $pAUC(0.2)$ and $\widehat{\eta}_p(0.2)$.

The difference is highly significant from 0 based on the bootstrap distribution ($p = 0.002$ and $p = 0$, respectively).

In order to illustrate the behaviour of the tests in a different scenario (crossing *ROC* curves), we have eliminated from the data those cases with the smallest values for the second biomarker. Therefore, now we consider 80 cases and 51 controls. In this case, the two test statistics $\Delta\widehat{\eta} = \widehat{\eta}_A - \widehat{\eta}_B$ and $\triangle AD = AD_A - AD_B$ are also statistically significantly different from 0 ($p = 0$) but the statistic T leads us to conclude that both biomarkers are not significantly different ($p = 0.128$). See Figure 2 (b).

4.2. A method for early recognition of malignant melanoma

The second data set we have considered can be found in [21]. The dataset consists of the clinical scoring scheme without a dermoscope and a dermoscope scoring scheme on 72 suspicious lesions in order to determine whether the dermoscope contributes diagnostic information. The p-value for T for paired data, constructed following [6], is $p = 0.882$. The p-values for $\Delta\widehat{\eta}$ and ΔAD are 0.717 and 0.555, respectively. We have also compared both biomarkers through the differences of the partial measures $pAUC(0.2)$ and $\widehat{\eta}_p(0.2)$ obtaining $p = 0.716$ and $p = 0.763$, respectively. Then, we can conclude that both biomarkers are statistically significantly equal. Therefore, the dermoscope contributes no useful information in this sense.

5. DISCUSSION

There is an interesting relationship between some summary measures for ROC curves and two sample test statistics. Some of them, the Mann-Whitney U-statistic and the Kolmogorov-Smirnov statistic, can be written in terms of functionals of the empirical ROC curve. The former is the well-known AUC (area under the *ROC* curve) and the later Youden index. Other test statistics such as Anderson-Darling, Neyman and Watson tests were studied in [16] to assess diagnostic markers. However, it can not be written in terms of functionals of the empirical *ROC* curve and they do not have value themselves. In this paper, we propose the sum of squared errors between the derivative of the *ROC* curve and 1, that is the derivative of the 45^o line, as a *ROC* summary statistic. This statistic is closely related to a nonparametric test for two sample problem based on an order statistic introduced in [1]. The exact distribution of this index is known but the simulated version is used ought to computational time since it is checked that the exact test should be essentially the same. For the purpose of assesing part of a ROC curve, we also define a new partial summary statistic based on the same idea as above but ending the summation as close as possible to the specific FPF of interest.

The simulations show that our ROC summary statistics exhibit much higher power in discriminating between the diseased and healthy distributions and are thus an attractive alternative to ROC-based methodology and indeed constitute in many cases an improvement over AUC and $pAUC$, respectively. Nevertheless, the fact that our ROC summary statistic does not have value itself is a drawback. The same index value can be obtained for two absolutely different curves. Therefore, after concluding that a new marker is diagnostic, we should study in which way the diseased and healthy distributions are different.

In case of the comparison of two diagnostic markers in the whole range, the use of the difference of our individual ROC summary statistics associated with the two diagnostic markers has higher power than the conventional non-parametric test in [6], the test based on AD test statistic and the permutation test proposed in [21] for crossing ROC curves. However, if the primary interest is to detect differences in $AUCs$, then the permutation tests of [2] and [4] should be used. On the other hand, when we are interested on a specific range of specificity, $pAUC$ uses to be an alternative to AUC but we show that our partial summary statistic $\widehat{\eta}_p$ is better to discriminate between two ROC curves that cross each other when the biomarkers are not correlated as well as when they are correlated.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the constructive comments and suggestions of the Associate Editor as well as of an anonymous referee, which led to great improvements in the manuscript. The first author acknowledges support from the project MTM2013-40788-R and the second author from the project MTM2014-55966-P of the Spanish Ministry of Economy and Competitiveness.

REFERENCES

[1] Bairamov, I.G. and Özkaya, N. (2000). On the nonparametric test for two sample problem based on spacings, *J. Appl. Stat. Sci.*, **10**, 57–68.

[2] Bandos, A.I.; Howard, E.R. and Gur, D. (2005). A permutation test sensitive to differences in areas for comparing ROC curves from a paired design, *Stat. Med.*, **24**, 2873–2893.

[3] Bloch, D.A. (1997). Comparing two diagnostic tests against the same "gold standard" in the same sample, *Biometrics*, **53**, 73–85.

[4] Braun, T.M. and Alonzo, T.A. (2008). A modified sign test for comparing paired ROC curves, *Biostat.*, **9**, 364–372.

[5] Campbell, G. (1994). Advances in statistical methodology for the evaluation of diagnostic and laboratory tests, *Stat. Med.*, **13**, 499–508.

[6] DeLong, E.R.; DeLong, D.M. and Clarke-Pearson, D.L. (1988). Comparing the areas under two or more correlated receiver operating characteristics curves: A nonparametric approach, *Biometrics*, **44**, 837–845.

[7] Gail, M.H. and Green, S.B. (1976). A generalization of the one-sided two-sample Kolmogorov–Smirnov statistic for evaluating diagnostic tests, *Biometrics*, **32**, 561–570.

[8] Greenhouse, S.W. and Mantel, N. (1950). The evaluation of diagnostic tests, *Biometrics*, **6**, 399–412.

[9] Hanley, J.A. and McNeil, B.J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve, *Radiology*, **143**, 29–36.

[10] Hanley, J.A. and McNeil, B.J. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases, *Radiology*, **148**, 839–843.

[11] Linnet, K. (1987). Comparison of quantitative diagnostic tests: Type 1 error, power, and sample size, *Stat. Med.*, **6**, 147–158.

[12] Martínez-Camblor, P.; Carleos, C. and Corral, N. (2011). Powerful nonparametric statistics to compare k-independent ROC curves, *J. Appl. Stat.*, **38**, 1317–1332.

[13] Martínez-Camblor, P.; Carleos, C. and Corral, N. (2013). General nonparametric ROC curves comparison, *J. Korean Stat. Soc.*, **42**, 71–81.

[14] McClish, D.K. (1989). Analyzing a portion of the ROC curve, *Med. Decision Making*, **9**, 190–195.

[15] McClish, D.K. (1990). Determining a range of false positive rates for which ROC curves differ, *Med. Decision Making*, **10**, 283–287.

[16] Nakas, C.; Yiannoutsos, C.T.; Bosch, R.J. and Moyssiadis, C. (2003). Assessment of diagnostic markers by goodness-of-fit tests, *Stat. Med.*, **10**, 2503–2513.

[17] Obuchowski, N.A. and McClish, D.K. (1997). Sample size determination for diagnostic accuracy studies involving binormal ROC curve indices, *Stat. Med.*, **16**, 1529–1542.

[18] Qin, G.; Jin, X. and Zhou, X.H. (2011). Non-parametric interval estimation for the partial area under the ROC curve, *Can. J. of Stat.*, **39**, 17–33.

[19] Shapiro, D.E. (1999). The interpretation of diagnostic tests, *Stat. Methods Med. Res.*, **8**, 113–134.

[20] Thompson, M.L. and Zucchini, W. (1989). On the statistical analysis of ROC curves, *Stat. Med.*, **8**, 1277–1290.

[21] Venkatraman, E.S. and Begg, C.B. (1996). A distribution-free procedure for comparing receiver operating characteristic curves from a paired experiment, *Biometrika*, **83**, 835–848.

[22] VENKATRAMAN, E.S. (2000). A Permutation Test to Compare Receiver Operating Characteristic Curves, *Biometrics*, **56**, 1134–1138.

[23] WIEAND, S.; GAIL, M.H.; JAMES, B.R. and JAMES, K.L. (1989). A family of nonparametric statistics for comparing diagnostic markers with paired and unpaired data, *Biometrika*, **76**, 585–592.

[24] ZHANG, D.D.; ZHOU, X.H.; FREEMAN, D.H. and FREEMAN, J.L. (2002). A non-parametric method for the comparison of partial areas under ROC curves and its application to large health care data sets, *Stat. Med.*, **21**, 701–715.

REVSTAT – Statistical Journal
Volume 15, Number 4, October 2017, 601–628

THE TRANSMUTED BIRNBAUM–SAUNDERS DISTRIBUTION

Authors: Marcelo Bourguignon
– Department of Statistics, Universidade Federal do Rio Grande do Norte, Brazil
m.p.bourguignon@gmail.com

Jeremias Leão
– Department of Statistics, Universidade Federal do Amazonas, Brazil
leaojeremiass@gmail.com

Víctor Leiva
– School of Industrial Engineering, Pontificia Universidad Católica de Valparaíso, Chile
victorleivasanchez@gmail.com, www.victorleiva.cl

Manoel Santos-Neto
– Department of Statistics, Universidade Federal de Campina Grande, Brazil
manoel.ferreira@ufcg.edu.br, www.santosnetoce.com.br

Received: February 2016 Revised: April 2016 Accepted: April 2016

Abstract:

- The Birnbaum–Saunders distribution has been largely studied and applied because of its attractive properties. We introduce a transmuted version of this distribution. Various of its mathematical and statistical features are derived. We use the maximum likelihood method for estimating its parameters and determine the score vector and Hessian matrix for inference and diagnostic purposes. We evaluate the performance of the maximum likelihood estimators by a Monte Carlo study. We illustrate the potential applications of the new transmuted Birnbaum–Saunders distribution by means of three real-world data sets from different areas.

Key-Words:

- *data analysis; likelihood methods; Monte Carlo simulations; Ox and R softwares; transmutation map.*

AMS Subject Classification:

- 62F10; 62J20; 62P99.

1. INTRODUCTION

The Birnbaum–Saunders (BS) distribution has been widely studied and applied due to its interesting properties. Some of its more relevant characteristics are the following:

- **(i)** it is a transformation of the normal distribution, inheriting several of its properties;
- **(ii)** it has two parameters, modifying its shape and scale;
- **(iii)** it has positive skewness, doing its probability density function (PDF) to be asymmetrical to the right, but due to its flexibility, symmetric data can also be modeled by the BS distribution;
- **(iv)** its PDF and failure rate (FR) are unimodal, but also other shapes for its FR may be modeled;
- **(v)** it belongs to the scale and closed under reciprocation families of distributions; and
- **(vi)** its scale parameter is also its median, so that the BS distribution can be seen as an analogue, but in an asymmetrical setting, of the normal distribution, which has the mean as one of its parameters.

For more details of the BS distribution, see Birnbaum and Saunders (1969b), Johnson *et al.* (1995, pp. 651–663), or the recent book by Leiva (2016). The BS distribution is a direct competitor of the gamma, inverse Gaussian (IG), lognormal and Weibull distributions; see details of these last distributions in Johnson *et al.* (1994).

The BS distribution has its genesis from fatigue of materials. Then, its natural applications have been mainly focussed on engineering and reliability. However, today they range diverse fields including business, environment and medicine. For some of its more recent applications, see Villegas *et al.* (2011), Marchant *et al.* (2013, 2016a,b), Saulo *et al.* (2013, 2017), Leiva *et al.* (2014a,d, 2015a, 2016c, 2015b, 2017), Rojas *et al.* (2015), Wanke and Leiva (2015), Desousa *et al.* (2017), Garcia-Papani *et al.* (2017), Leão *et al.* (2017a,b), Lillo *et al.* (2016) and Mohammadi *et al.* (2017). These and other applications, as well as several extensions and generalizations of the BS distribution, have been conducted by an international, transdisciplinary group of researchers. The first extension of the BS distribution is attributed to Volodin and Dzhungurova (2000), which established that the BS distribution is the mixture equally weighted of an IG distribution and its convolution with the chi-squared distribution with one degree of freedom. The authors provided a physical interpretation in terms of fatigue-life models and introduced a general family of distributions, with members such as the IG, normal and BS distributions, as well as others used in reliability applications. Then, Díaz-García and Leiva (2005) introduced the generalized BS (GBS) distribution; see also Azevedo *et al.* (2012). Owen (2006) proposed a three-parameter extension of the BS distribution. Vilca and Leiva (2006) derived a BS distribution based on skew-normal models (skew-BS). Gómez *et al.* (2009) extended the BS distribution from the slash-elliptic model. Guiraud *et al.* (2009) deducted a non-central version of the BS distribution. Leiva *et al.* (2009) provided a length-biased version of the BS distribution. Ahmed *et al.* (2010) analyzed a truncated version of the BS distribution. Kotz *et al.* (2010) performed mixture models related to the BS distribution. Vilca *et al.* (2010) and Castillo *et al.* (2011) developed the epsilon-skew BS distribution. Balakrishnan *et al.* (2011) considered mixture BS distributions. Cordeiro and Lemonte (2011) defined the beta-BS distribution. Leiva *et al.* (2011) modeled wind energy flux using a shifted BS distribution. Athayde *et al.* (2012) viewed the BS distributions as part of the Johnson system, allowing location-scale BS distributions to be obtained. Ferreira *et al.* (2012) and Leiva *et al.* (2016a) proposed an extreme value version of the BS distribution and its modeling. Santos-Neto *et al.* (2012, 2014, 2016) and Leiva *et al.* (2014c) reparameterized the BS distribution obtaining interesting properties and modeling. Saulo *et al.* (2012) presented the Kumaraswamy-BS distribution. Fierro *et al.* (2013) generated the BS distribution from a non-homogeneous Poisson process. Lemonte (2013) studied the Marshall–Olkin-BS (MOBS) distribution. Bourguignon *et al.* (2014) derived the power-series BS class of distributions. Martinez *et al.* (2014) introduced an alpha-power extension of the BS distribution. Leiva *et al.* (2016c) derived a zero-adjusted BS distribution.

The above mentioned review about extensions and generalizations of the BS distribution is in agreement with the important growing that the distribution theory has had in the last decades. This is because, although the Gaussian (or normal) distribution has dominated this theory during more than 100 years, many real-world applications cannot be well modeled by this distribution. Then, non-normal distributions which must be flexible in skewness and kurtosis are needed. The interested reader can find a good collection of non-normal distributions in Johnson *et al.* (1994, 1995). Several of these distributions were constructed using methods early proposed by Pearson (1895), Edgeworth (1917), Cornish and Fisher (1937) and Johnson (1949), based on differential equations, mathematical approximations and translation techniques; see more details in Johnson *et al.* (1994, pp. 15–62). A more recent proposal on non-normal distributions is attributed to Azzalini (1985). In the line of these works and motivated from financial mathematics, where applications in the calculation of value at risk and corrections to the Black–Scholes options need flexibility in skewness and kurtosis, Shaw and Buckley (2009) introduced new parametric families of distributions based on the transmutation method. This method modifies the skewness and/or kurtosis into symmetric and asymmetric distributions and generates a new distributional family known as transmuted (or changed in its shape) distributions. The transmutation method proposed by Shaw and Buckley (2009) carries out a function composition between the cumulative distribution function (CDF) of a distribution and the quantile function (QF) of another. Aryal and Tsokos (2009) defined the transmuted extreme value distribution. Aryal and Tsokos (2011) and Khan and King (2013) presented transmuted Weibull distributions. Aryal (2013) proposed the transmuted log-logistic distribution. Ashour and Eltehiwy (2013) analyzed the transmuted Lomax distribution. Mroz (2013a,b) studied the transmuted Lindley and Rayleigh distributions, whereas Sharma *et al.* (2014) derived a transmuted inverse Rayleigh distribution. Khan and King (2014) considered the transmuted inverse Weibull distribution. Merovci and Puka (2014) deduced the transmuted Pareto distribution. Tiana *et al.* (2014) developed the transmuted linear exponential distribution. Saboor *et al.* (2015) created a transmuted exponential-Weibull distribution. Louzada and Granzotto (2016) introduced the transmuted log-logistic regression model. To our best knowledge, no transmuted versions of the BS distribution exist. Therefore, the main objective of this paper is to propose and derive the transmuted BS (TBS) distribution, as well as a comprehensive treatment of its mathematical and statistical properties.

Section 2 presents the TBS distribution and derives some of its characteristics including its PDF, CDF and QF, as well as its FR, moments and a generator of random numbers. Section 3 provides the estimation of the TBS parameters using the maximum likelihood (ML) method, including the corresponding score vector and Hessian matrix for inferential and diagnostic purposes. In this section, the performance of the ML estimators is evaluated by means of Monte Carlo (MC) simulations. In addition, diagnostic tools are derived to detect influential data in the ML estimation. Section 4 illustrates the potential applications of the TBS distribution with three real-world data sets from different areas. Section 5 discusses the conclusions of this work and future research about the topic.

2. FORMULATION AND CHARACTERISTICS

In this section, we provide a background of the BS distribution, formulate the new distribution and obtain some of its more relevant characteristics.

2.1. The BS distribution

A random variable T_1 has a BS distribution with shape ($\alpha > 0$) and scale ($\beta > 0$) parameters if it can be represented by

$$T_1 \;=\; \beta\Big(\alpha Z/2 + \big((\alpha Z/2)^2+1\big)^{1/2}\Big)^2\,,$$

where $Z \sim \mathrm{N}(0,1)$. In this case, the notation $T_1 \sim \mathrm{BS}(\alpha,\beta)$ is used. The CDF of T_1 is

$$F_{\mathrm{BS}}(t;\alpha,\beta) \;=\; \Phi\big((1/\alpha)\,\rho(t/\beta)\big)\,, \qquad t>0\,,$$

where $\rho(y) = y^{1/2} - y^{-1/2}$, for $y > 0$, and Φ denotes the standard normal CDF.

The PDF of T_1 is

$$f_{\mathrm{BS}}(t;\alpha,\beta) \,=\, \kappa(\alpha,\beta)\, t^{-3/2}(t+\beta)\, \exp\bigl(-\tau(t/\beta)/(2\alpha^2)\bigr)\,, \qquad t>0\,,$$

where $\kappa(\alpha,\beta) = \exp(1/\alpha^2)/(2\alpha\sqrt{2\pi\beta})$ and $\tau(y) = y + 1/y$, for $y > 0$. Note that the inverse function of the CDF of a random variable, also known as QF, is defined by $F^{-1}(y) = \inf_{x\in\mathbb{R}}\{F(x) \geq y\}$, for $y \in [0,1]$. Then, the QF of T_1 is

$$t_1(q;\alpha,\beta) \,=\, F_{\mathrm{BS}}^{-1}(q;\alpha,\beta) \,=\, \beta\Big(\alpha\, z(q)/2 + \big((\alpha\, z(q)/2)^2+1\big)^{1/2}\Big)^2, \qquad 0<q<1\,,$$

where $z = \Phi^{-1}$ is the inverse function of the standard normal CDF (or QF), and F_{BS}^{-1} is the inverse function of F_{BS}. As mentioned, the BS distribution holds the following scale and reciprocation properties:

(**i**) $b\,T_1 \sim \mathrm{BS}(\alpha, b\,\beta)$, for $b > 0$, and

(**ii**) $1/T_1 \sim \mathrm{BS}(\alpha, 1/\beta)$, respectively.

The rth moment of T_1 is

$$\mathrm{E}(T_1^r) \,=\, \frac{\beta^r\big(K_{r+1/2}(1/\alpha^2) + K_{r-1/2}(1/\alpha^2)\big)}{2\,K_{1/2}(1/\alpha^2)}\,,$$

with $K_\nu(u)$ denoting the modified Bessel function of the third kind of order ν and argument u given by

$$K_\nu(u) \,=\, \frac{1}{2}\left(\frac{u}{2}\right)^{\nu}\int_0^\infty w^{-\nu-1}\exp\left(-w - \frac{u^2}{4w}\right)\mathrm{d}w\,;$$

see Gradshteyn and Randzhik (2000, p. 907).

2.2. The TBS distribution

The TBS distribution that we propose is motivated by the work of Shaw and Buckley (2009). As mentioned, they introduced a class of generalized distributions based on the transmutation method, which is described next. Let F_1 and F_2 be the CDFs of two distributions with a common sample space and F_1^{-1} and F_2^{-1} be their inverse functions, that is, their QFs, respectively. The general rank transmutation map as given in Shaw and Buckley (2009) is defined by $G_{12}(u) = F_2(F_1^{-1}(u))$ and $G_{21}(u) = F_1(F_2^{-1}(u))$. The functions G_{12} and G_{21} both map the unit interval $[0,1]$ into itself. Under suitable assumptions, G_{12} and G_{21} satisfy $G_{ij}(0) = 0$ and $G_{ij}(1) = 1$, for $i,j = 1,2$, with $i \neq j$. A quadratic rank transmutation map is defined as $G_{12}(u) = u + \lambda\, u(1-u)$, for $|\lambda| \leq 1$, from which follows that the CDF satisfies the relationship $F_2(x) = (1+\lambda)\,F_1(x) - \lambda(F_1(x))^2$. Then, by differentiation, it yields $f_2(x) = f_1(x)\,(1+\lambda-2\lambda\,F_1(x))$, where f_1 and f_2 are the corresponding PDFs associated with the CDFs F_1 and F_2, respectively.

For more details about the quadratic rank transmutation map, see Shaw and Buckley (2009). By using the BS CDF and PDF, we have the TBS CDF and PDF, with parameters α, β and λ, given respectively by

$$\begin{aligned} F_{\text{TBS}}(t;\alpha,\beta,\lambda) &= (1+\lambda)\,\Phi\big((1/\alpha)\,\rho(t/\beta)\big) - \lambda\Big(\Phi\big((1/\alpha)\,\rho(t/\beta)\big)\Big)^2, \\ f_{\text{TBS}}(t;\alpha,\beta,\lambda) &= \Big(1+\lambda-2\lambda\,\Phi\big((1/\alpha)\,\rho(t/\beta)\big)\Big)\,f_{\text{BS}}(t;\alpha,\beta), \qquad t>0, \end{aligned} \tag{2.1}$$

where $|\lambda| \le 1$ is an additional skewness parameter, whose role is to introduce skewness and to vary the corresponding tail weights. Hereafter, a random variable T with CDF or PDF given as in (2.1) is denoted by $T \sim \text{TBS}(\alpha,\beta,\lambda)$. Note that, at $\lambda = 0$, we have the BS distribution. It can also be shown that

$$\lim_{t\to 0} f_{\text{TBS}}(t;\alpha,\beta,\lambda) = \lim_{t\to\infty} f_{\text{TBS}}(t;\alpha,\beta,\lambda) = 0\,.$$

Observe that the PDF of T can be expressed as a finite linear combination of $\text{BS}(\alpha,\beta)$ and skew-$\text{BS}(\alpha,\beta,1)$ PDFs; see Vilca and Leiva (2006) for details on the skew-BS distribution and its features. Thus,

$$f_{\text{TBS}}(t;\alpha,\beta,\lambda) = (1+\lambda)\,f_{\text{BS}}(t;\alpha,\beta) - \lambda f_{\text{skew-BS}}(t;\alpha,\beta,1)\,,$$

where $f_{\text{skew-BS}}(t;\alpha,\beta,\eta) = 2\,\Phi\big((\eta/\alpha)\,\rho(t/\beta)\big)\,f_{\text{BS}}(t;\alpha,\beta)$, for $\eta \in \mathbb{R}$. In addition, if $\lambda = -1$, then $T \sim$ skew-$\text{BS}(\alpha,\beta,1)$. Figure 1 (first panel/row) displays several shapes of the PDF given in (2.1) for some parameter values. These shapes reveal that the TBS distribution is very versatile and that the additional skewness parameter λ has substantial effects on its skewness. Note that the shapes of the TBS distribution are much more flexible than those of the BS distribution.

2.3. Characteristics of the TBS distribution

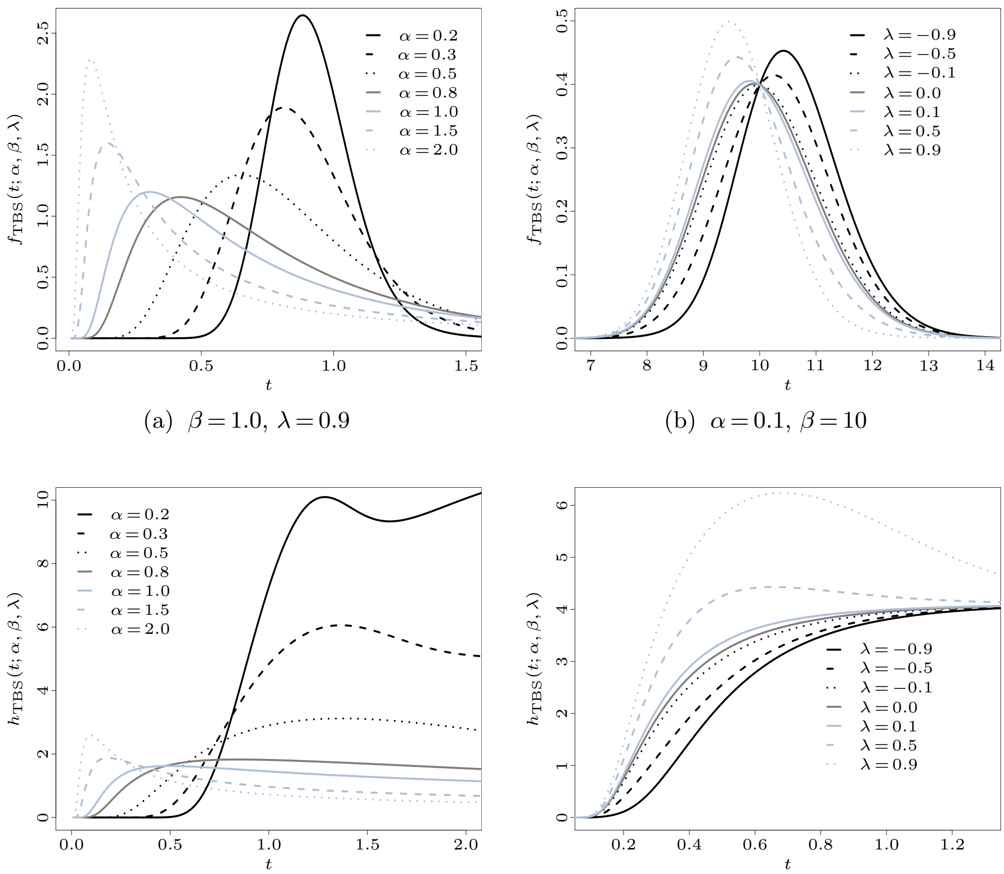
Several of the mathematical properties of the TBS distribution can be obtained directly from the BS and skew-BS distributions. For example, the ordinary moments and moment generating function of the TBS distribution follow immediately from the moments of BS and skew-BS distributions. For more details of the skew-BS distribution, see Vilca and Leiva (2006) and Saulo *et al.* (2013). Some properties of the TBS distribution are as follow. If $T \sim \text{TBS}(\alpha,\beta,\lambda)$, then:

- **(i)** $b\,T \sim \text{TBS}(\alpha, b\,\beta, \lambda)$, for $b > 0$, that is, the TBS distribution is closed under scale transformations;
- **(ii)** $1/T \sim \text{TBS}(\alpha, 1/\beta, -\lambda)$, that is, the TBS distribution is closed under reciprocation;
- **(iii)** $Y = (\alpha^2/\beta)T \sim \text{TBS}(\alpha, \alpha^2, \lambda)$, that is, Y follows a two-parameter TBS distribution.

The FR of $T \sim \mathrm{TBS}(\alpha,\beta,\lambda)$ is

$$h_{\mathrm{TBS}}(t;\alpha,\beta,\lambda) = \frac{\Big(1+\lambda-2\lambda\,\Phi\big((1/\alpha)\,\rho(t/\beta)\big)\Big)\,h_{\mathrm{BS}}(t;\alpha,\beta)}{1-\lambda\,\Phi\big((1/\alpha)\,\rho(t/\beta)\big)}\,, \qquad t>0\,, \tag{2.2}$$

where $h_{\mathrm{BS}}(t) = f_{\mathrm{BS}}(t)/(1-F_{\mathrm{BS}}(t))$ is the FR of the BS distribution. Figure 1 (second panel/row) shows the FR of the TBS distribution for some parameter values.



(a) $\beta=1.0,\ \lambda=0.9$

(b) $\alpha=0.1,\ \beta=10$

(c) $\beta=1.0,\ \lambda=0.9$

(d) $\alpha=0.5,\ \beta=0.5$

Figure 1: Plots of the TBS PDF (first panel/row) and FR (second panel/row) for the indicated value of its parameters.

We can verify that the TBS FR is upside-down. From (2.2), note that:

(i) $h_{\mathrm{TBS}}(t;\alpha,\beta,\lambda)/h_{\mathrm{BS}}(t;\alpha,\beta)$ is decreasing in t for $\lambda \geq 0$;

(ii) $h_{\mathrm{TBS}}(t;\alpha,\beta,\lambda)/h_{\mathrm{BS}}(t;\alpha,\beta)$ is increasing in t for $\lambda \leq 0$;

(iii) $h_{\mathrm{BS}}(t;\alpha,\beta) \leq h_{\mathrm{TBS}}(t;\alpha,\beta,\lambda) \leq (1+\lambda)\,h_{\mathrm{TBS}}(t;\alpha,\beta,\lambda)$ for $\lambda \geq 0$;

(iv) $(1+\lambda)\,h_{\mathrm{BS}}(t;\alpha,\beta) \leq h_{\mathrm{TBS}}(t;\alpha,\beta,\lambda) \leq h_{\mathrm{BS}}(t;\alpha,\beta)$ for $\lambda \leq 0$; and

(v) $\lim_{t\to 0} h_{\mathrm{TBS}}(t;\alpha,\beta,\lambda) = 0$ and $\lim_{t\to\infty} h_{\mathrm{TBS}}(t;\alpha,\beta,\lambda) = 1/(2\alpha^2\beta)$, that is, the limiting behaviors of the FRs of the TBS and BS distributions are the same.

Observe that the expression given in (2.2) may also be written as

$$h_{\mathrm{TBS}}(t;\alpha,\beta,\lambda) \;=\; p(t)\, h_{\mathrm{BS}}(t;\alpha,\beta) + \big(1-p(t)\big)\, h_{\text{skew-BS}}(t;\alpha,\beta,1)\,,$$

where $p(t) = \big((1+\lambda)\,\big(1-\Phi\big((1/\alpha)\,\rho(t/\beta)\big)\big)\big)/\big(1-(1+\lambda)\,\Phi\big((1/\alpha)\,\rho(t/\beta)\big) + \lambda\big(\Phi\big((1/\alpha)\,\rho(t/\beta)\big)\big)^2\big)$, whereas h_{BS} and $h_{\text{skew-BS}}$ are the FRs of the BS and skew-BS distributions, respectively.

Many important features of a distribution can be obtained through its moments. Let $T_1 \sim \mathrm{BS}(\alpha,\beta)$ and $T_2 \sim \text{skew-BS}(\alpha,\beta,1)$. Then, their rth moments are

$$\mathrm{E}(T_1^r) \;=\; \frac{\beta^r\alpha^{2\,r}}{2^{3r-1}} \sum_{k=0}^{r}\sum_{i=0}^{k} \binom{2\,r}{2\,k}\binom{k}{i}\left(\frac{\alpha^2}{4}\right)^{i-k}, \tag{2.3}$$

$$\mathrm{E}(T_2^r) \;=\; \beta^r \sum_{k=0}^{r}\sum_{i=0}^{k} \binom{r}{k}\binom{k}{i}\, 2^i \left(\frac{\alpha}{2}\right)^{k+1} w_{k+1;k-i}\,, \qquad r = 1,2,..., \tag{2.4}$$

where $w_{a,b} = \mathrm{E}\big(Z^a(\sqrt{\alpha^2 Z^2+4})^b\big)$ and $Z \sim \text{skew-normal}(0,1,1)$; see Azzalini (1985). The rth moment of $T \sim \mathrm{TBS}(\alpha,\beta,\lambda)$ can be written as $\mathrm{E}(T^r) = (1+\lambda)\,\mathrm{E}(T_1^r) - \lambda\mathrm{E}(T_2^r)$. Then, using the results presented in (2.3) and (2.4), we obtain

$$\begin{aligned}\mathrm{E}(T^r) \;=\; \beta^r \Bigg(\sum_{k=0}^{r}\sum_{j=0}^{k}\Bigg((1+\lambda)\binom{2\,r}{2\,k}\binom{k}{j}\frac{\alpha^{2(r-k+j)}}{2^{3(r-k+j)-1}} \\ -\lambda\binom{r}{k}\binom{k}{j}\, 2^j\left(\frac{\alpha}{2}\right)^{k+1} w_{k+1;k-j}\Bigg)\Bigg)\,.\end{aligned}$$

Therefore, the first four moments of $T \sim \mathrm{TBS}(\alpha,\beta,\lambda)$ are

$$\begin{aligned}
\mathrm{E}(T) \;&=\; \mu \;=\; \beta\left(1+\frac{\alpha^2}{2}\right)\left(1+\lambda-\lambda\left(1+\frac{\alpha\, w_{1,1}}{(2+\alpha^2)}\right)\right),\\
\mathrm{E}(T^2) \;&=\; \beta^2\left(1+2\alpha^2+\frac{3}{2}\alpha^4\right)\left(1+\lambda-\lambda\left(1+\frac{2\,\alpha\, w_{1,1}+\alpha^3 w_{3,1}}{2+4\alpha^2+3\alpha^4}\right)\right),\\
\mathrm{E}(T^3) \;&=\; \beta^3\left(1+\frac{9}{2}\alpha^2+9\alpha^4+\frac{15}{2}\alpha^6\right)\\
&\quad\times\left(1+\lambda-\lambda\left(1+\frac{3\,\alpha\, w_{1,1}+4\,\alpha^3 w_{3,1}+\alpha^5 w_{5,1}}{2+9\,\alpha^2+18\,\alpha^4+15\,\alpha^6}\right)\right),\\
\mathrm{E}(T^4) \;&=\; \beta^4\left(1+8\,\alpha^2+30\,\alpha^4+60\,\alpha^6+\frac{105}{2}\alpha^8\right)\\
&\quad\times\left(1+\lambda-\lambda\left(1+\frac{4\,\alpha\, w_{1,1}+10\,\alpha^3 w_{3,1}+6\,\alpha^5 w_{5,1}+\alpha^7 w_{7,1}}{2+16\,\alpha^2+60\,\alpha^4+120\,\alpha^6+105\,\alpha^8}\right)\right).
\end{aligned}$$

Thus, the rth moment of $T \sim \text{TBS}(\alpha,\beta,\lambda)$ about its mean is

$$\begin{aligned}\mathrm{E}\big((T-\mu)^r\big) \;&=\; (1+\lambda)\sum_{j=0}^{r}\binom{r}{j}(\mu_1-\mu)^{r-j}\,\mathrm{E}\big((T_1-\mu_1)^j\big)\\ &\quad -\lambda\sum_{j=0}^{r}\binom{r}{j}(\mu_2-\mu)^{r-j}\,\mathrm{E}\big((T_2-\mu_2)^j\big),\end{aligned}$$

where $\mu_1 = \beta\left(1+\alpha^2/2\right)$ and $\mu_2 = \beta\left(1+\alpha\,w_{1,1}+\alpha^2/2\right)$. Hence, the corresponding second, third and fourth moments about the mean are

$$\begin{aligned}\mathrm{E}\big((T-\mu)^2\big) \;&=\; \mathrm{Var}(T) \;=\; (1+\lambda)\left((\mu_1-\mu)^2+\sigma_1^2\right)-\lambda\big((\mu_2-\mu)^2+\sigma_2^2\big),\\ \mathrm{E}\big((T-\mu)^3\big) \;&=\; (1+\lambda)\left((\mu_1-\mu)^3+3\sigma_1^2+\mu_1^{(3)}\right)-\lambda\left((\mu_2-\mu)^3+3\sigma_2^2+\mu_2^{(3)}\right),\\ \mathrm{E}\big((T-\mu)^4\big) \;&=\; (1+\lambda)\left((\mu_1-\mu)^4+6\,(\mu_1-\mu)^2\sigma_1^2+6\,(\mu_1-\mu)\,\mu_1^{(3)}+\mu_1^{(4)}\right)\\ &\quad -\lambda\left((\mu_2-\mu)^4+6\,(\mu_2-\mu)^2\sigma_2^2+6\,(\mu_1-\mu)\,\mu_2^{(3)}+\mu_2^{(4)}\right),\end{aligned}$$

where

$$\begin{aligned}\sigma_1^2 \;&=\; \mathrm{Var}(T_1) \;=\; \alpha^2\beta^2\big(1+(5/4)\,\alpha^2\big),\\ \sigma_2^2 \;&=\; \mathrm{Var}(T_2) \;=\; (\beta^2/4)\left(4\,\alpha^2-\alpha^2\,w_{1,1}^2+2\,\alpha^3\,w_{3,1}-2\,\alpha^3\,w_{1,1}+5\,\alpha^4\right),\\ \mu_1^{(3)} \;&=\; \beta^3\alpha^4\big(3+(11/2)\,\alpha^2\big),\\ \mu_1^{(4)} \;&=\; \beta^4\alpha^4\big(3+(45/2)\,\alpha^2+(633/16)\,\alpha^4\big),\\ \mu_2^{(3)} \;&=\; \mu_1^{(3)}+(\alpha^3\beta^3/4)\Big(2\,\alpha^2\,w_{5,1}+2\,w_{3,1}-3\,\alpha^2\,w_{3,1}-3\,\alpha\,w_{1,1}\,w_{3,1}\\ &\quad +\,w_{1,1}^3+3\,\alpha\,w_{1,1}^2-6\,\alpha^2\,w_{1,1}-6\,w_{1,1}\Big),\end{aligned}$$

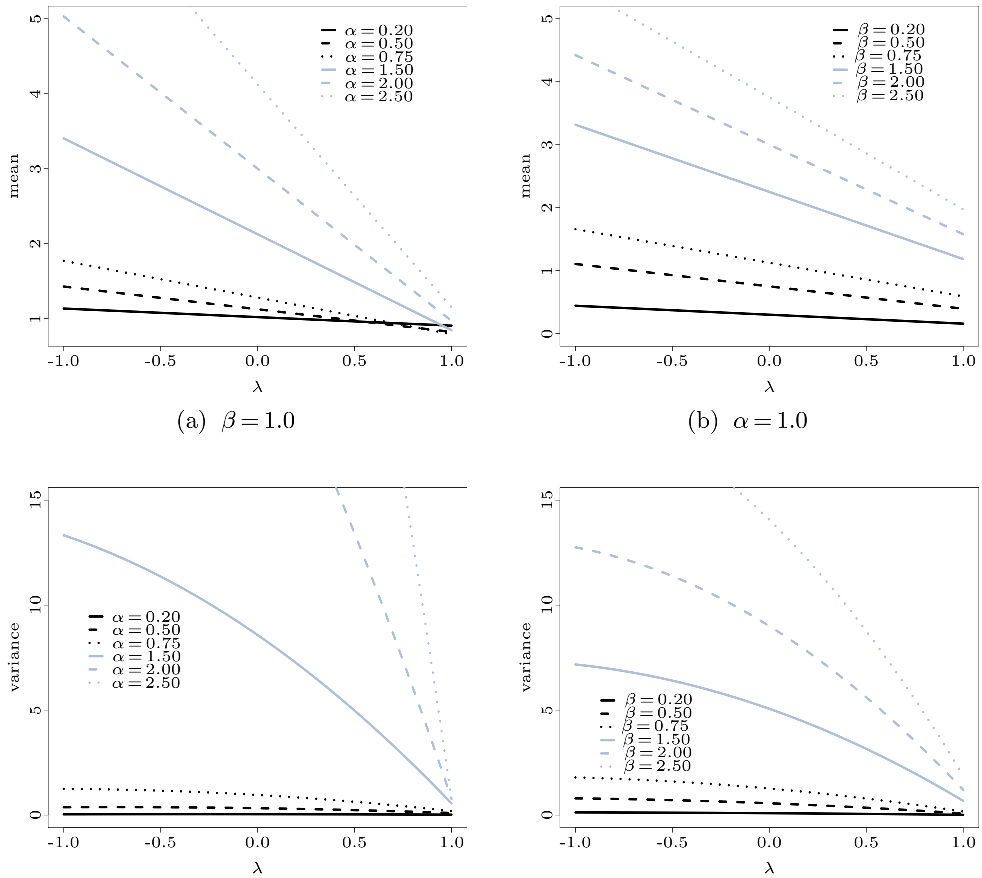
and

$$\begin{aligned}\mu_2^{(4)} \;&=\; \mu_1^{(4)}+(\alpha^4\beta^4/16)\Big(24\,\alpha^2\,w_{1,1}\,w_{2,1}+12\,w_{1,1}\,w_{3,1}^2-16\,\alpha^2\,w_{1,1}\,w_{5,1}\\ &\quad +\,18\,\alpha^2\,w_{1,1}^2-96\,\alpha\,w_{1,1}+16\,\alpha\,w_{5,1}-12\,\alpha\,w_{1,1}^3+8\,\alpha^3\,w_{7,1}\\ &\quad -\,3\,w_{1,1}^4+24\,w_{1,1}^2-16\,\alpha^3\,w_{5,1}+12\,\alpha^3\,w_{3,1}-180\,\alpha^3\,w_{1,1}\\ &\quad +\,16\,\alpha\,w_{3,1}-16\,w_{1,1}\,w_{3,1}\Big).\end{aligned}$$

Figure 2 presents graphical plots of the mean (first panel/row) and variance (second panel/row) of the TBS distribution for different values of α, β and λ. Note that the mean and variance decrease as λ increases, but the mean and variance, generally, increases as α and β increase. The QF of $T \sim \text{TBS}(\alpha,\beta,\lambda)$ is

$$t_{\mathrm{TBS}}(q;\alpha,\beta,\lambda) \;=\; \begin{cases} \beta\left(\frac{\alpha}{2}\,\Phi^{-1}(q^*)+\left(1+\frac{\alpha^2}{4}\,\Phi^{-1}(q^*)^2\right)^{1/2}\right)^2, & \lambda\neq 0;\\[2ex] \beta\left(\frac{\alpha}{2}\,\Phi^{-1}(q)+\left(1+\frac{\alpha^2}{4}\,\Phi^{-1}(q)^2\right)^{1/2}\right)^2, & \lambda = 0;\end{cases}$$

where $q^* = \left(1+\lambda-\sqrt{(1+\lambda)^2-4\,\lambda\,q}\right)/2\,\lambda$, for $q\in[0,1]$. Random numbers for the TBS distribution can be generated from the TBS QF, which is detailed by Algorithm 1.



(a) $\beta = 1.0$ (b) $\alpha = 1.0$

(c) $\beta = 1.0$ (d) $\alpha = 1.0$

Figure 2: Plots of the mean (first panel/row) and variance (second panel/row) of the TBS distribution for the indicated value of its parameters.

Algorithm 1 – Random number generator from the TBS distribution

---

1: Generate a random number u from $U \sim \mathrm{U}(0,1)$;

2: Set values for α, β and λ of $T \sim \mathrm{TBS}(\alpha, \beta, \lambda)$;

3: If $\lambda \neq 0$, then compute a random number

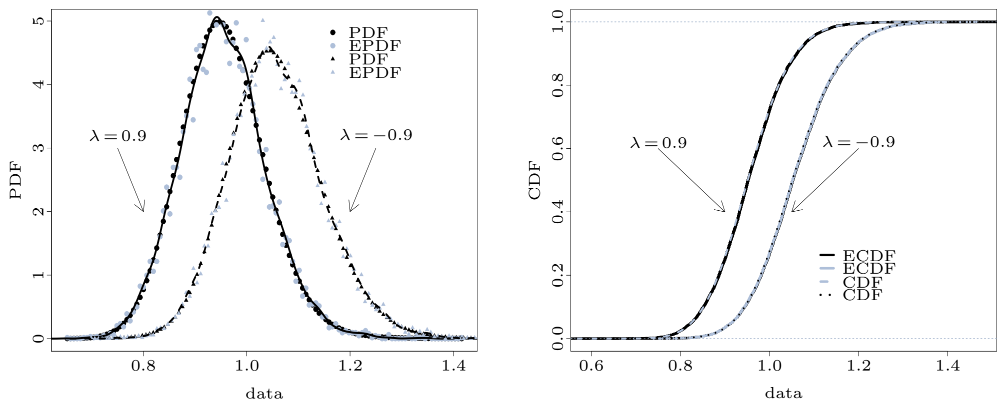
$$t \;=\; \beta\left(\frac{\alpha}{2}\,\Phi^{-1}(u^*) + \left(1 + \frac{\alpha^2}{4}\,\Phi^{-1}(u^*)^2\right)^{1/2}\right)^2$$

from $T \sim \mathrm{TBS}(\alpha, \beta, \lambda)$, with $u^* = \left(1 + \lambda - \sqrt{(1+\lambda)^2 - 4\,\lambda\, u}\right)/(2\lambda)$; otherwise

$$t \;=\; \beta\left(\frac{\alpha}{2}\,\Phi^{-1}(u) + \left(1 + \frac{\alpha^2}{4}\,\Phi^{-1}(u)^2\right)^{1/2}\right)^2 ;$$

4: Repeat steps 1 to 3 until the required amount of random numbers to be completed.

From Figure 3, note that the generator of random numbers proposed in Algorithm 1 seems to be appropriate for simulating data from a TBS distribution. We implement this algorithm in the `R` statistical software (R Core Team, 2016) and generate 10000 random numbers, considering the following values of the parameters: $\alpha = 0.1$, $\beta = 1.0$ and $\lambda \in \{-0.9, 0.9\}$. The empirical PDF (EPDF), the empirical CDF (ECDF) and the kernel density estimate (KDE) are obtained using these random numbers. Figure 3 (a) shows that the midpoints are consistent with the values obtained through the TBS PDF. Figure 3 (b) allows us to compare the ECDF and TBS CDF, which are detected to be similar.



(a) (b)

Figure 3: EPDF and TBS PDF with its KDE in solid and dashed lines (a) and ECDF and TBS CDF (b) for simulated data.

3. PARAMETER ESTIMATION, ITS PERFORMANCE AND DIAGNOSTICS

In this section, we use the ML method to estimate the TBS distribution parameters. In addition, by means of MC simulations, we study the performance of the ML estimators. Furthermore, we provide diagnostic tools to detect influential data.

3.1. ML estimation

Let $T_1, ..., T_n$ be a random sample from the TBS distribution with vector of parameters $\boldsymbol{\theta} = (\alpha, \beta, \lambda)^\top$ and $t_1, ..., t_n$ be their observations (data). The log-likelihood function for $\boldsymbol{\theta}$ is

$$\begin{aligned}\ell(\boldsymbol{\theta}) \;=\; & n\log\bigl(\kappa(\alpha,\beta)\bigr) - \frac{3}{2}\sum_{i=1}^{n}\log(t_i) + \sum_{i=1}^{n}\log(t_i+\beta) \\ & - \frac{1}{2\alpha^2}\sum_{i=1}^{n}\tau(t_i/\beta) + \sum_{i=1}^{n}\log\Bigl(1+\lambda\bigl(1-2\Phi(v_i)\bigr)\Bigr),\end{aligned} \tag{3.1}$$

where $v_i = (1/\alpha)\,\rho(t_i/\beta)$. The ML estimate $\widehat{\boldsymbol{\theta}} = (\widehat{\alpha}, \widehat{\beta}, \widehat{\lambda})^\top$ is obtained by solving the likelihood equations $U_\alpha = U_\beta = U_\lambda = 0$ simultaneously, where U_α, U_β and U_λ are the components of the score vector $\boldsymbol{U}(\boldsymbol{\theta}) = (U_\alpha, U_\beta, U_\lambda)^\top$ given by

$$\begin{aligned}
U_\alpha \;&=\; -\frac{n}{\alpha}\left(1+\frac{2}{\alpha^2}\right) + \frac{1}{\alpha^3}\sum_{i=1}^{n}\left(\frac{t_i}{\beta}+\frac{\beta}{t_i}\right) + \frac{2\lambda}{\alpha}\sum_{i=1}^{n}\frac{v_i\,\phi(v_i)}{1+\lambda-2\lambda\,\Phi(v_i)}, \\
U_\beta \;&=\; -\frac{n}{2\beta} + \sum_{i=1}^{n}\frac{1}{t_i+\beta} + \frac{1}{2\alpha^2\beta}\sum_{i=1}^{n}\left(\frac{t_i}{\beta}+\frac{\beta}{t_i}\right) - \frac{2\lambda}{\alpha\beta}\sum_{i=1}^{n}\left(\frac{\tau\bigl(\sqrt{t_i/\beta}\bigr)\,\phi(v_i)}{1+\lambda-2\lambda\,\Phi(v_i)}\right), \\
U_\lambda \;&=\; \sum_{i=1}^{n}\frac{1-2\,\Phi(v_i)}{1+\lambda\bigl(1-2\,\Phi(v_i)\bigr)},
\end{aligned}$$

with ϕ being the standard normal PDF. The equations $U_\alpha = U_\beta = U_\lambda = 0$ cannot be solved analytically, so that iterative techniques, such as bisection, Newton–Raphson and secant methods, may be used; see Lange (2001) and McNamee and Pa (2013). To obtain the ML estimates of the model parameters, we employ the subroutine `MaxBFGS` of the `Ox` software; see Doornik (2006). This subroutine uses the analytical derivatives to maximize $\ell(\boldsymbol{\theta})$; see Nocedal and Wright (1999) and Press *et al.* (2007). As starting values for the numerical procedure, we suggest to consider

$$\widetilde{\alpha} \;=\; \bigl(s/\widetilde{\beta} + \widetilde{\beta}/r - 2\bigr)^{1/2}, \qquad \widetilde{\beta} = (s\,r)^{1/2}, \qquad \widetilde{\lambda} = 0,$$

where $s = (1/n)\sum_{i=1}^{n} t_i$ and $r = 1/\bigl((1/n)\sum_{i=1}^{n}(1/t_i)\bigr)$; see Birnbaum and Saunders (1969a) and Leiva (2016, pp. 40–42).

To construct approximate confidence intervals and hypothesis tests for the parameters, we use the normal approximation of the distribution of the ML estimator of $\boldsymbol{\theta} = (\alpha, \beta, \lambda)^\top$. Specifically, assume that regularity conditions are fulfilled in the interior of the parameter space but not on the boundary; see Cox and Hinkley (1974). Then, the asymptotic distribution of $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ is $\mathrm{N}_3(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}})$, where $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ is the asymptotic variance–covariance matrix of $\widehat{\boldsymbol{\theta}}$, which can be approximated from the observed information matrix $\boldsymbol{K}(\boldsymbol{\theta}) = -\boldsymbol{J}(\boldsymbol{\theta})$, where $\boldsymbol{J}(\boldsymbol{\theta})$ is the Hessian matrix $\boldsymbol{J}(\boldsymbol{\theta}) = \partial^2\ell(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}^\top$, whose elements are

$$\begin{aligned}
J_{\alpha\alpha} &= \frac{n}{\alpha^2} + \frac{6n}{\alpha^4} - \frac{4\lambda}{\alpha^2}\sum_{i=1}^{n}\frac{v_i\,\phi(v_i)}{1+\lambda-2\lambda\,\Phi(v_i)}\left(v_i^2 + \frac{v_i\,\phi(v_i)}{1+\lambda-2\lambda\,\Phi(v_i)} - 2\right)\\
&\quad - \frac{3}{4}\sum_{i=1}^{n}\left(\frac{t_i}{\beta}+\frac{\beta}{t_i}\right),\\
J_{\alpha\beta} &= \frac{2\lambda}{\alpha^2\beta}\sum_{i=1}^{n}\frac{\tau\big(\sqrt{t_i/\beta}\big)\,\phi(v_i)}{1+\lambda\big(1-2\Phi(v_i)\big)}\left(\frac{2\lambda\,v_i\,\phi(v_i)}{1+\lambda\big(1-2\Phi(v_i)\big)} + (v_i)^2 - 1\right)\\
&\quad - \frac{1}{\alpha^3\beta}\sum_{i=1}^{n}\left(\frac{t_i}{\beta}-\frac{\beta}{x_i}\right),\\
J_{\alpha\lambda} &= -\frac{2}{\alpha}\sum_{i=1}^{n}\left(\frac{v_i\,\phi(v_i)}{\big(1+\lambda-2\lambda\,\Phi(v_i)\big)^2}\right),\\
J_{\beta\lambda} &= \frac{2}{\alpha\beta}\sum_{i=1}^{n}\left(\frac{\tau\big(\sqrt{t_i/\beta}\big)\,\phi(v_i)}{\big(1+\lambda-2\lambda\,\Phi(v_i)\big)^2}\right),\\
J_{\beta\beta} &= \frac{\lambda}{\alpha\beta^2}\sum_{i=1}^{n}\frac{\phi(v_i)}{1+\lambda-2\lambda\,\Phi(v_i)}\\
&\quad\times\left(3\sqrt{\frac{t_i}{\beta}} - \sqrt{\frac{\beta}{t_i}} - \frac{\big(\tau(t_i/\beta)\big)^2}{\alpha}\left(v_i - \frac{2\lambda\,\phi(v_i)}{1+\lambda-2\lambda\,\Phi(v_i)}\right)\right)\\
&\quad + \frac{n}{2\beta^2} - \sum_{i=1}^{n}\frac{1}{(t_i+\beta)^2} - \frac{1}{\alpha^2\beta^3}\sum_{i=1}^{n}t_i,\\
J_{\lambda\lambda} &= -\sum_{i=1}^{n}\left(\frac{1-2\Phi(v_i)}{1+\lambda\big(1-2\Phi(v_i)\big)}\right)^2.
\end{aligned}$$

Thus, this trivariate normal distribution can be used to construct approximate confidence intervals and regions for the model parameters. Note that asymptotic $100(1-\gamma/2)\%$ confidence intervals for α, β and λ are, respectively, established as

$$\widehat{\alpha} \pm z_{1-\gamma/2}\big(\widehat{\mathrm{Var}}(\widehat{\alpha})\big)^{1/2}, \qquad \widehat{\beta} \pm z_{1-\gamma/2}\big(\widehat{\mathrm{Var}}(\widehat{\beta})\big)^{1/2}, \qquad \widehat{\lambda} \pm z_{1-\gamma/2}\big(\widehat{\mathrm{Var}}(\widehat{\lambda})\big)^{1/2},$$

where $\widehat{\mathrm{Var}}(\widehat{\theta}_j)$ is the jth diagonal element of $\boldsymbol{K}^{-1}(\widehat{\boldsymbol{\theta}})$ related to each parameter θ_j, for $j=1,2,3$, with $\theta_1=\alpha$, $\theta_2=\beta$, $\theta_3=\lambda$, and $z_{\gamma/2}$ is the $100(1-\gamma/2)$th quantile of the standard normal distribution. Note that the estimated asymptotic standard errors (SEs) of the each estimator can be obtained from the square root of the diagonal element of $\boldsymbol{K}^{-1}(\widehat{\boldsymbol{\theta}})$.

3.2. Simulation study

We present a numerical experiment to evaluate the performance of the ML estimators $\widehat{\alpha}$, $\widehat{\beta}$ and $\widehat{\lambda}$. The simulation was performed using the `Ox` software. A number of 10000 MC replications were considered, sample sizes $n \in \{25, 50, 75, 100, 200, 400, 800\}$, the combination of the parameters $(\alpha, \beta) \in \{(0.10, 1.00), (0.50, 1.00), (1.50, 1.00), (2.00, 1.00)\}$ and $\lambda \in \{-0.80, -0.50, -0.20, 0.20, 0.50, 0.80\}$. Without loss of generality, we fix β at 1.00 in all experiments, because this is a scale parameter. Table 1 presents the empirical bias and square root of mean squared error of the estimators of the TBS distribution parameters. From this table, note that, generally, the bias decreases as n increases, evidencing that the ML estimators $\widehat{\alpha}$ and $\widehat{\beta}$ are asymptotically unbiased. Observe that, when varying the values of λ, the distributions of the estimators of α and β show, in general, symmetrical behaviors. In addition, when the parameter α increases, the bias of $\widehat{\beta}$ increases. Note also that the estimator $\widehat{\lambda}$ is more biased than $\widehat{\alpha}$ and $\widehat{\beta}$, considering all scenarios. Also in all of the cases, the square root of the mean square error decreases as n increases, proving that the ML estimators of the TBS distribution parameters have good precision, as known. It is important to mention that some iterations did not converge during the simulations, due possibly to the complexity of the function to be maximized or because of the difficulty to provide a good initial value from λ.

3.3. Influence diagnostics

Local influence is based on the curvature of the plane of the log-likelihood function; see Leiva *et al.* (2014b, 2016b). In the case of the TBS model given in (2.1), let $\boldsymbol{\theta} = (\alpha,\, \beta,\, \lambda)^\top$ and $\ell(\boldsymbol{\theta}|\boldsymbol{\omega})$ be the parameter vector and the log-likelihood function related to this model perturbed by $\boldsymbol{\omega}$, respectively. The perturbation vector $\boldsymbol{\omega}$ belongs to a subset $\Omega \in \mathbb{R}^n$ and $\boldsymbol{\omega}_0$ is an $n \times 1$ non-perturbation vector, such that $\ell(\boldsymbol{\theta}|\boldsymbol{\omega}_0) = \ell(\boldsymbol{\theta})$, for all $\boldsymbol{\theta}$. The corresponding likelihood distance (LD) is

$$\mathrm{LD}(\boldsymbol{\omega}) = 2\big(\ell(\widehat{\boldsymbol{\theta}}) - \ell(\widehat{\boldsymbol{\theta}}_\omega)\big)\,, \tag{3.2}$$

where $\widehat{\boldsymbol{\theta}}_\omega$ denotes the ML estimate of $\boldsymbol{\theta}$ upon the perturbed TBS model used to assess the influence of the perturbation on the ML estimate, whereas $\ell(\widehat{\boldsymbol{\theta}})$ is the usual likelihood function given in (3.1). Cook (1987) showed that the normal curvature for $\boldsymbol{\theta}$ in the direction of the vector $\boldsymbol{d}$, with $\|\boldsymbol{d}\| = 1$, is expressed as $C_d(\boldsymbol{\theta}) = 2|\boldsymbol{d}^\top \boldsymbol{\Delta}^\top \boldsymbol{J}(\boldsymbol{\theta})^{-1} \boldsymbol{\Delta}\, \boldsymbol{d}|$, where $\boldsymbol{\Delta}$ is a $3 \times n$ perturbation matrix with elements $\Delta_{ji} = \partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})/\partial\theta_j\, \partial\omega_i$ evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\omega} = \boldsymbol{\omega}_0$, for $j = 1, 2, 3$, $i = 1, ..., n$, and $\boldsymbol{J}(\boldsymbol{\theta})$ is the corresponding Hessian matrix.

Table 1: Empirical bias and square root of mean squared error (in parentheses) of the parameter estimator for the indicated value of α, β, λ, n.

n	λ	$\alpha=0.10,\ \beta=1.00$			$\alpha=0.50,\ \beta=1.00$			$\alpha=1.50,\ \beta=1.00$			$\alpha=2.00,\ \beta=1.00$		
		$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\lambda}$	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\lambda}$	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\lambda}$	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\lambda}$
25	−0.80	−0.0096 (0.0172)	0.0209 (0.0305)	0.3329 (0.3892)	−0.0509 (0.0871)	0.1198 (0.1722)	0.3472 (0.4095)	−0.1741 (0.2818)	0.3332 (0.5057)	0.3022 (0.3853)	−0.2717 (0.4122)	0.4616 (0.6787)	0.3296 (0.4115)
	−0.50	−0.0038 (0.0152)	0.0064 (0.0243)	0.0926 (0.2381)	−0.0177 (0.0763)	0.0331 (0.1222)	0.0760 (0.2328)	−0.0670 (0.2359)	0.0970 (0.3309)	0.0502 (0.2467)	−0.1000 (0.3250)	0.1457 (0.4034)	0.0898 (0.2494)
	−0.20	−0.0013 (0.0147)	−0.0087 (0.0248)	−0.1734 (0.2738)	−0.0048 (0.0745)	−0.0430 (0.1179)	−0.2053 (0.3139)	−0.0221 (0.2232)	−0.0581 (0.2662)	−0.1827 (0.2988)	−0.0306 (0.3040)	−0.0315 (0.3069)	−0.1447 (0.2613)
	0.20	−0.0012 (0.0145)	0.0093 (0.0253)	0.1731 (0.2738)	−0.0064 (0.0738)	0.0500 (0.1292)	0.1952 (0.3005)	−0.0138 (0.2236)	0.1562 (0.3613)	0.1920 (0.3042)	−0.0247 (0.3055)	0.1369 (0.3875)	0.1453 (0.2637)
	0.50	−0.0042 (0.0152)	−0.0061 (0.0236)	−0.1010 (0.2436)	−0.0196 (0.0765)	−0.0181 (0.1128)	−0.0736 (0.2416)	−0.0673 (0.2380)	−0.0104 (0.2817)	−0.0511 (0.2469)	−0.1038 (0.3267)	−0.0357 (0.3175)	−0.0896 (0.2508)
	0.80	−0.0099 (0.0175)	−0.0198 (0.0292)	−0.3239 (0.3822)	−0.0484 (0.0869)	−0.0897 (0.1338)	−0.3350 (0.3960)	−0.1726 (0.2788)	−0.1858 (0.2976)	−0.3058 (0.3889)	−0.2662 (0.4045)	−0.2336 (0.3510)	−0.3278 (0.4107)
50	−0.80	−0.0070 (0.0129)	0.0165 (0.0251)	0.2627 (0.3466)	−0.0386 (0.0667)	0.0962 (0.1406)	0.2907 (0.3771)	−0.1295 (0.2199)	0.2662 (0.4066)	0.2680 (0.3640)	−0.1916 (0.3083)	0.3345 (0.5025)	0.2754 (0.3675)
	−0.50	−0.0015 (0.0110)	0.0030 (0.0200)	0.0506 (0.2407)	−0.0062 (0.0560)	0.0160 (0.0988)	0.0420 (0.2400)	−0.0244 (0.1733)	0.0541 (0.2515)	0.0360 (0.2483)	−0.0395 (0.2321)	0.0862 (0.2943)	0.0650 (0.2415)
	−0.20	0.0008 (0.0106)	−0.0108 (0.0220)	−0.1897 (0.2940)	0.0067 (0.0549)	−0.0527 (0.1037)	−0.2051 (0.3136)	0.0179 (0.1645)	−0.0743 (0.2149)	−0.1669 (0.2879)	0.0055 (0.2139)	−0.0476 (0.2298)	−0.1282 (0.2489)
	0.20	0.0008 (0.0106)	0.0108 (0.0223)	0.1893 (0.2928)	0.0059 (0.0544)	0.0596 (0.1177)	0.1942 (0.3017)	0.0173 (0.1646)	0.1312 (0.2860)	0.1668 (0.2870)	0.0104 (0.2168)	0.1156 (0.2930)	0.1283 (0.2492)
	0.50	−0.0018 (0.0108)	−0.0033 (0.0195)	−0.0618 (0.2423)	−0.0066 (0.0550)	−0.0070 (0.0960)	−0.0432 (0.2405)	−0.0203 (0.1732)	0.0066 (0.2374)	−0.0315 (0.2462)	−0.0427 (0.2356)	−0.0235 (0.2572)	−0.0696 (0.2437)
	0.80	−0.0068 (0.0128)	−0.0159 (0.0244)	−0.2663 (0.3514)	−0.0374 (0.0652)	−0.0777 (0.1149)	−0.2848 (0.3678)	−0.1275 (0.2167)	−0.1605 (0.2584)	−0.2659 (0.3639)	−0.1932 (0.3104)	−0.1912 (0.2992)	−0.2727 (0.3694)
75	−0.80	−0.0057 (0.0110)	0.0144 (0.0230)	0.2320 (0.3297)	−0.0293 (0.0560)	0.0764 (0.1202)	0.2270 (0.3430)	−0.1118 (0.1889)	0.2359 (0.3620)	0.2490 (0.3488)	−0.1584 (0.2681)	0.2803 (0.4342)	0.2371 (0.3370)
	−0.50	−0.0005 (0.0095)	0.0018 (0.0185)	0.0342 (0.2449)	−0.0023 (0.0469)	0.0107 (0.0909)	0.0250 (0.2409)	−0.0075 (0.1480)	0.0419 (0.2244)	0.0307 (0.2446)	−0.0213 (0.1989)	0.0661 (0.2504)	0.0566 (0.2322)
	−0.20	0.0017 (0.0093)	−0.0114 (0.0211)	−0.1991 (0.3046)	0.0101 (0.0473)	−0.0504 (0.0999)	−0.2012 (0.3082)	0.0200 (0.1370)	−0.0656 (0.1905)	−0.1422 (0.2664)	0.0134 (0.1761)	−0.0500 (0.1988)	−0.1083 (0.2278)
	0.20	0.0014 (0.0091)	0.0112 (0.0213)	0.1909 (0.2965)	0.0089 (0.0467)	0.0603 (0.1115)	0.1785 (0.3028)	0.0221 (0.1384)	0.1120 (0.2508)	0.1414 (0.2652)	0.0164 (0.1781)	0.0940 (0.2485)	0.1104 (0.2307)
	0.50	−0.0009 (0.0093)	−0.0024 (0.0181)	−0.0471 (0.2400)	−0.0027 (0.0469)	−0.0046 (0.0872)	−0.0318 (0.2384)	−0.0084 (0.1491)	0.0038 (0.2200)	−0.0311 (0.2458)	−0.0207 (0.1960)	−0.0140 (0.2333)	−0.0570 (0.2332)
	0.80	−0.0059 (0.0112)	−0.0144 (0.0226)	−0.2426 (0.3408)	−0.0307 (0.0570)	−0.0681 (0.1063)	−0.2469 (0.3442)	−0.1069 (0.1887)	−0.1469 (0.2402)	−0.2408 (0.3450)	−0.1571 (0.2657)	−0.1661 (0.2691)	−0.2337 (0.3359)
100	−0.80	−0.0048 (0.0099)	0.0130 (0.0217)	0.2072 (0.3183)	−0.0286 (0.0528)	0.0758 (0.1195)	0.2349 (0.3233)	−0.1012 (0.1730)	0.2152 (0.3348)	0.2272 (0.3296)	−0.1346 (0.2385)	0.2401 (0.3852)	0.2091 (0.3107)
	−0.50	0.0000 (0.0084)	0.0008 (0.0180)	0.0204 (0.2489)	0.0005 (0.0430)	0.0072 (0.0869)	0.0176 (0.2447)	−0.0053 (0.1353)	0.0418 (0.2114)	0.0369 (0.2440)	−0.0147 (0.1790)	0.0539 (0.2296)	0.0490 (0.2259)
	−0.20	0.0020 (0.0082)	−0.0111 (0.0203)	−0.1950 (0.3011)	0.0120 (0.0430)	−0.0527 (0.0967)	−0.1909 (0.2905)	0.0217 (0.1219)	−0.0596 (0.1740)	−0.1236 (0.2492)	0.0157 (0.1562)	−0.0396 (0.1763)	−0.0885 (0.2105)
	0.20	0.0022 (0.0081)	0.0116 (0.0208)	0.1910 (0.2955)	0.0116 (0.0424)	0.0598 (0.1087)	0.1895 (0.2931)	0.0218 (0.1231)	0.1005 (0.2302)	0.1266 (0.2516)	0.0131 (0.1557)	0.0750 (0.2141)	0.0883 (0.2091)
	0.50	−0.0004 (0.0084)	−0.0015 (0.0176)	−0.0330 (0.2399)	−0.0010 (0.0418)	−0.0020 (0.0853)	−0.0257 (0.2366)	−0.0026 (0.1342)	0.0052 (0.2048)	−0.0308 (0.2406)	−0.0140 (0.1788)	−0.0028 (0.2171)	−0.0481 (0.2262)
	0.80	−0.0049 (0.0098)	−0.0126 (0.0212)	−0.2129 (0.3250)	−0.0285 (0.0510)	−0.0659 (0.1025)	−0.2401 (0.3448)	−0.0965 (0.1712)	−0.1363 (0.2235)	−0.2255 (0.3300)	−0.1342 (0.2384)	−0.1455 (0.2500)	−0.2093 (0.3116)
200	−0.80	−0.0031 (0.0079)	0.0091 (0.0187)	0.1530 (0.2852)	−0.0209 (0.0434)	0.0591 (0.1046)	0.1863 (0.3146)	−0.0744 (0.1414)	0.1624 (0.2709)	0.1780 (0.2829)	−0.0935 (0.1885)	0.1619 (0.2863)	0.1455 (0.2440)
	−0.50	0.0009 (0.0071)	−0.0007 (0.0168)	−0.0011 (0.2470)	0.0040 (0.0357)	0.0011 (0.0810)	0.0035 (0.2458)	0.0070 (0.1115)	0.0187 (0.1767)	0.0160 (0.2201)	−0.0003 (0.1376)	0.0301 (0.1785)	0.0300 (0.1906)
	−0.20	0.0024 (0.0067)	−0.0107 (0.0192)	−0.1820 (0.2963)	0.0121 (0.0352)	−0.0463 (0.0884)	−0.1712 (0.2969)	0.0138 (0.0870)	−0.0353 (0.1346)	−0.0717 (0.1941)	0.0106 (0.1097)	−0.0261 (0.1331)	−0.0525 (0.1657)
	0.20	0.0022 (0.0066)	0.0108 (0.0193)	0.1767 (0.2896)	0.0119 (0.0344)	0.0544 (0.1005)	0.1704 (0.2758)	0.0157 (0.0887)	0.0578 (0.1692)	0.0739 (0.1976)	0.0092 (0.1081)	0.0427 (0.1522)	0.0503 (0.1638)
	0.50	0.0003 (0.0068)	−0.0005 (0.0165)	−0.0202 (0.2386)	0.0024 (0.0345)	0.0005 (0.0802)	−0.0187 (0.2351)	0.0057 (0.1102)	0.0121 (0.1813)	−0.0167 (0.2213)	−0.0002 (0.1395)	0.0027 (0.1812)	−0.0276 (0.1928)
	0.80	−0.0033 (0.0080)	−0.0096 (0.0189)	−0.1635 (0.2992)	−0.0229 (0.0424)	−0.0565 (0.0943)	−0.2092 (0.3319)	−0.0729 (0.1410)	−0.1092 (0.1947)	−0.1794 (0.2867)	−0.0871 (0.1857)	−0.0977 (0.2032)	−0.1414 (0.2419)
400	−0.80	−0.0019 (0.0066)	0.0061 (0.0164)	0.1043 (0.2538)	−0.0139 (0.0369)	0.0436 (0.0923)	0.1391 (0.2816)	−0.0518 (0.1143)	0.1117 (0.2098)	0.1254 (0.2238)	−0.0547 (0.1498)	0.0962 (0.2075)	0.0886 (0.1817)
	−0.50	0.0011 (0.0060)	−0.0009 (0.0160)	−0.0048 (0.2436)	0.0053 (0.0308)	−0.0006 (0.0759)	0.0005 (0.2387)	0.0106 (0.0902)	0.0048 (0.1430)	0.0047 (0.1874)	0.0061 (0.1072)	0.0097 (0.1365)	0.0068 (0.1525)
	−0.20	0.0022 (0.0055)	−0.0097 (0.0176)	−0.1595 (0.2790)	0.0116 (0.2770)	−0.0387 (0.0813)	−0.1410 (0.0298)	0.0072 (0.0592)	−0.0142 (0.0994)	−0.0323 (0.1424)	0.0040 (0.0760)	−0.0064 (0.0980)	−0.0165 (0.1215)
	0.20	0.0019 (0.0054)	0.0097 (0.0178)	0.1567 (0.2751)	0.0109 (0.0287)	0.0439 (0.0902)	0.1348 (0.2673)	0.0070 (0.0593)	0.0269 (0.1108)	0.0317 (0.1415)	0.0043 (0.0764)	0.0178 (0.1041)	0.0188 (0.1220)
	0.50	0.0006 (0.0057)	−0.0002 (0.0153)	−0.0164 (0.2352)	0.0042 (0.0298)	0.0028 (0.0757)	−0.0135 (0.2301)	0.0098 (0.0905)	0.0177 (0.1572)	−0.0041 (0.1877)	0.0056 (0.1061)	0.0068 (0.1431)	−0.0108 (0.1529)
	0.80	−0.0021 (0.0066)	−0.0066 (0.0166)	−0.1165 (0.2671)	−0.0186 (0.0371)	−0.0497 (0.0903)	−0.1870 (0.3240)	−0.0502 (0.1138)	−0.0739 (0.1588)	−0.1213 (0.2242)	−0.0503 (0.1477)	−0.0577 (0.1659)	−0.0854 (0.1807)
800	−0.80	−0.0011 (0.0055)	0.0041 (0.0141)	0.0728 (0.2216)	−0.0095 (0.0320)	0.0322 (0.0808)	0.1045 (0.2482)	−0.0348 (0.0951)	0.0733 (0.1612)	0.0821 (0.1731)	−0.0244 (0.1233)	0.0498 (0.1568)	0.0474 (0.1400)
	−0.50	0.0010 (0.0051)	−0.0006 (0.0147)	0.0010 (0.2286)	0.0050 (0.0266)	−0.0001 (0.0707)	0.0043 (0.2266)	0.0091 (0.0713)	−0.0015 (0.1112)	−0.0037 (0.1466)	0.0061 (0.0765)	0.0008 (0.0969)	−0.0012 (0.1116)
	−0.20	0.0018 (0.0045)	−0.0074 (0.0155)	−0.1224 (0.2494)	0.0086 (0.0239)	−0.0266 (0.0681)	−0.0960 (0.2350)	0.0030 (0.0407)	−0.0037 (0.0730)	−0.0079 (0.1063)	0.0017 (0.0523)	0.0007 (0.0723)	−0.0028 (0.0918)
	0.20	0.0018 (0.0045)	0.0076 (0.0160)	0.1218 (0.2506)	0.0079 (0.0233)	0.0307 (0.0769)	0.0925 (0.2305)	0.0027 (0.0410)	0.0093 (0.0760)	0.0097 (0.1069)	0.0005 (0.0525)	0.0040 (0.0736)	0.0012 (0.0910)
	0.50	0.0004 (0.0047)	−0.0011 (0.0141)	−0.0290 (0.2227)	0.0040 (0.0261)	0.0028 (0.0708)	−0.0108 (0.2230)	0.0078 (0.0707)	0.0124 (0.1238)	−0.0005 (0.1465)	0.0033 (0.0749)	0.0057 (0.1015)	−0.0036 (0.1113)
	0.80	−0.0013 (0.0055)	−0.0043 (0.0141)	−0.0777 (0.2273)	−0.0157 (0.0331)	−0.0424 (0.0827)	−0.1607 (0.2990)	−0.0293 (0.0925)	−0.0437 (0.1311)	−0.0743 (0.1723)	−0.0229 (0.1234)	−0.0258 (0.1416)	−0.0449 (0.1384)

A local influence diagnostic is generally based on index plots. For example, the index graph of the eigenvector $\boldsymbol{d}_{\max}$ related to the maximum eigenvalue of $\boldsymbol{B}(\boldsymbol{\theta}) = -\boldsymbol{\Delta}^\top \boldsymbol{J}(\boldsymbol{\theta})^{-1}\boldsymbol{\Delta}$, $C_{d_{\max}}(\boldsymbol{\theta})$ say, evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$, can detect those cases that, under small perturbations, exercise a high influence on $\mathrm{LD}(\boldsymbol{\omega})$ given in (3.2). In addition to the direction vector of maximum normal curvature, $\boldsymbol{d}_{\max}$ say, another direction of interest is $\boldsymbol{d}_i = \boldsymbol{e}_{in}$, which corresponds to the direction of the case i, where $\boldsymbol{e}_{in}$ is an $n \times 1$ vector of zeros with a value equal to one at the ith position, that is, $\{\boldsymbol{e}_{in}, 1 \leq i \leq n\}$ is the canonical basis of $\mathbb{R}^n$. Thus, the normal curvature is $C_i(\boldsymbol{\theta}) = 2|b_{ii}|$, where b_{ii} is the ith diagonal element of $\boldsymbol{B}(\boldsymbol{\theta})$, for $i = 1, ..., n$, evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$. The case i is considered as potentially influential if $C_i(\widehat{\boldsymbol{\theta}}) > 2\overline{C}(\widehat{\boldsymbol{\theta}})$, where $\overline{C}(\widehat{\boldsymbol{\theta}}) = \sum_{i=1}^n C_i(\widehat{\boldsymbol{\theta}})/n$. This procedure is called total local influence of the case i; see Liu *et al.* (2016).

Consider the log-likelihood function given in (3.1). We obtain the respective perturbation matrix $\boldsymbol{\Delta}$, which is already evaluated at the non-perturbation vector $\boldsymbol{\omega}_0$, under the scheme of case-weight perturbation. Then, we want to evaluate whether cases with different weights in the log-likelihood function affect the ML estimate of $\boldsymbol{\theta}$. This scheme is the most used to assess local influence in a model. The log-likelihood function of the TBS model perturbed by the case-weight scheme is

$$\ell(\boldsymbol{\theta}|\boldsymbol{\omega}) = \sum_{i=1}^n \ell_i(\boldsymbol{\theta}|\omega_i) = \sum_{i=1}^n \omega_i\,\ell_i(\boldsymbol{\theta}).$$

Then, taking its derivative with respect to $\boldsymbol{\omega}^\top$, we obtain $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_\beta, \boldsymbol{\Delta}_\alpha, \boldsymbol{\Delta}_\lambda)^\top$. After evaluating at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\omega} = \boldsymbol{\omega}_0$, the elements of $\boldsymbol{\Delta}_\alpha$, $\boldsymbol{\Delta}_\beta$ and $\boldsymbol{\Delta}_\lambda$ are

$$\begin{aligned}
\Delta_\alpha^{(i)} &= -\frac{1}{\alpha}\left(1 + \frac{2}{\alpha^2}\right) + \frac{1}{\alpha^3}\left(\frac{t_i}{\beta} + \frac{\beta}{t_i}\right) + \frac{2\lambda\, v_i\, \phi(v_i)}{\alpha\big(1 + \lambda - 2\lambda\,\Phi(v_i)\big)}\,, \qquad i = 1, ..., n,\\
\Delta_\lambda^{(i)} &= \frac{1 - 2\Phi(v_i)}{1 + \lambda\big(1 - 2\Phi(v_i)\big)}\,,\\
\Delta_\beta^{(i)} &= -\frac{1}{2\beta} + \frac{1}{t_i + \beta} + \frac{1}{2\alpha^2\beta}\left(\frac{t_i}{\beta} + \frac{\beta}{t_i}\right) - \frac{2\lambda}{\alpha\beta}\left(\frac{\tau\big(\sqrt{t_i/\beta}\big)\,\phi(v_i)}{1 + \lambda - 2\lambda\,\Phi(v_i)}\right).
\end{aligned}$$

4. APPLICATIONS TO REAL-WORLD DATA

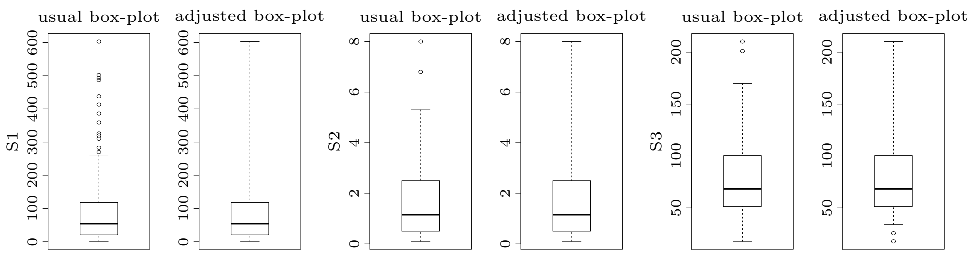
In this section, we apply the obtained results for the new model to three data sets, illustrating its potential applications. The results are compared to other competing BS distributions. All the computations were done using the `Ox` software. For each data set, we estimate the unknown parameters of the associated distribution by the ML method and evaluate its goodness of fit with suitable methods.

4.1. Exploratory analysis

The first data set (S1) corresponds to the number of successive failures for the air conditioning system of each member in a fleet of 13 Boeing 720 jet airplanes ($n = 188$); S1 can be obtained from Proschan (1963). The second data set (S2) is related to vinyl chloride concentration (in mg/L) obtained from clean upgradient monitoring wells ($n = 34$); S2 can be obtained from Bhaumik *et al.* (2009). The third data set (S3) corresponds to protein amount (in g) in the restricted diet for adult patients in a Chilean hospital ($n = 61$); S3 and more details about these data can be obtained from Leiva *et al.* (2014a). Table 2 provides some descriptive measures for the three data sets, which include central tendency statistics, the standard deviation (SD) and the coefficients of variation (CV), skewness (CS) and kurtosis (CK), among others. From these exploratory analyses, we detect asymmetrical distributions with positive skewness in all of the cases and different kurtosis levels. Figure 5 (first panel/row) shows the histograms of S1, S2 and S3, from which it is possible to observe these features.

Table 2: Descriptives statistics for the indicated data set.

Statistic	Data set		
	S1	S2	S3
n	188	34	61
Minimum	1.0	0.10	17.8
Median	54.0	1.15	68.2
Mean	92.7	1.88	80.4
Maximum	603.0	8.00	210.3
SD	107.9	1.95	42.3
CS	2.1	1.53	1.2
CK	4.9	1.72	4.0



(a) S1 (b) S2 (c) S3

Figure 4: Box-plots for the indicated data set.

Figure 4 displays the usual and adjusted box-plots, where the latter is useful in cases when the data follow a skew distribution; see Rousseeuw *et al.* (2016). From Figure 4, note that potential outliers considered by the usual box-plot are not outliers in the adjusted box-plot. This is an indication that no outliers are present at the right tail in all of the studied data sets. Figure 5 (second panel/row) confirms these facts by means of the influence index plots, which do not detect atypical cases. Therefore, the TBS distribution can be a good candidate for modeling these data sets. We compare the TBS distribution to other generalizations of the BS distribution, such as the three-parameter MOBS, exponentiated BS (EBS) and two-parameter BS distributions with the EBS and MOBS PDFs being: $f_{\mathrm{EBS}}(x;\alpha,\beta,a) = a f_{\mathrm{BS}}(x;\alpha,\beta)\, F_{\mathrm{BS}}(x;\alpha,\beta)^{a-1}$, for $x>0$, $a>0$, and $f_{\mathrm{MOBS}}(x;\eta,\alpha,\beta) = \eta f_{\mathrm{BS}}(x;\alpha,\beta)/\big(1-(1-\eta)(1-F_{\mathrm{BS}}(x;\alpha,\beta))\big)^2$, for $x>0$, $\eta>0$.

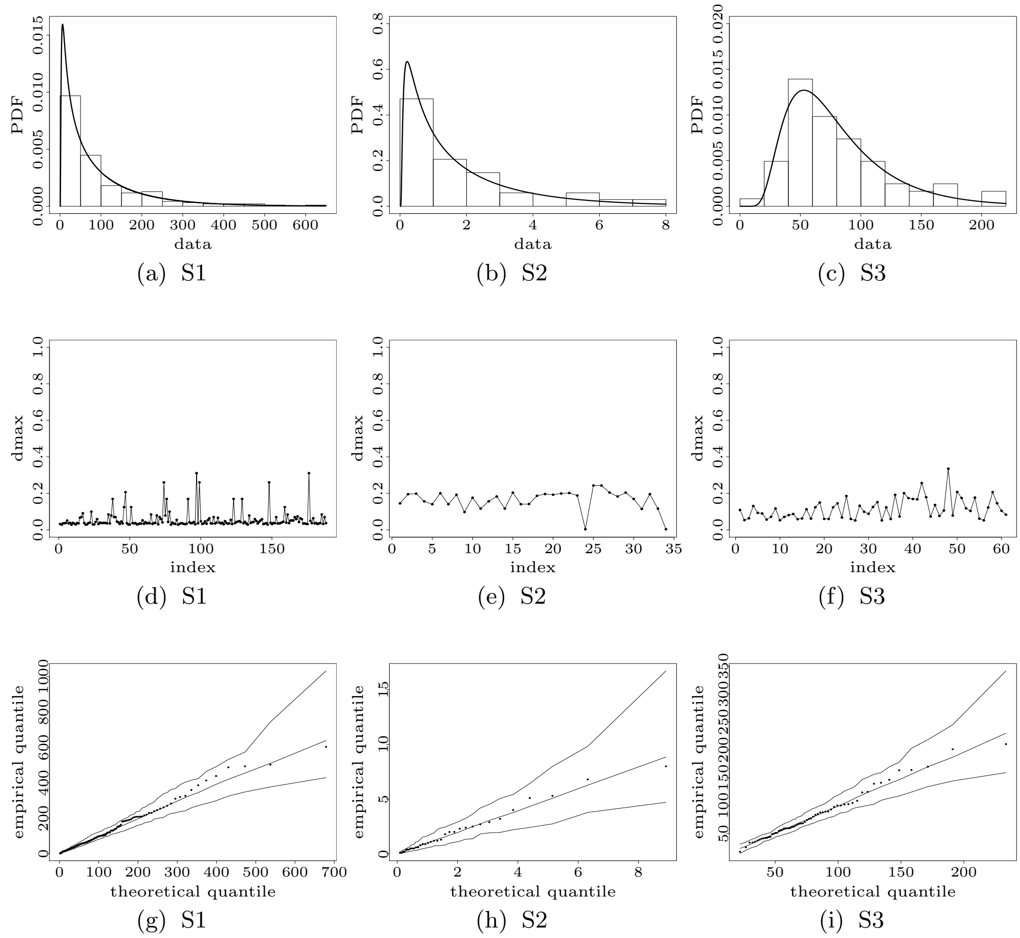


Figure 5: Histograms with estimated PDF (first panel/row), influence index (second panel/row) and plots QQ plots with envelope (third panel/row) for the TBS distribution based on the indicated data set.

4.2. Confirmatory analysis

Table 3 lists the ML estimates of the parameters and the estimated asymptotic SEs in parentheses of the corresponding estimators for the four distributions fitted to S1, S2 and S3. From this table and using the asymptotic distributions of the ML estimators proved by the simulation study of Section 3.2, we evaluate whether the additional parameters of the EBS, MOBS and TBS distributions are significatively different from zero or not for each data set. Note that, for S1, the additional parameter is always significatively different from zero at 5% for all EBS, MOBS and TBS distributions, indicating that the BS distribution should model S1 poorly. This is not the case of S2 and S3, where only the MOBS parameter is significatively different from zero at 5% in both cases.

Table 3: ML estimates (with estimated SE in parenthesis) for the indicated parameter, distribution and data set.

Distribution	S1			S2			S3		
	$\widehat{\theta}_1$	$\widehat{\theta}_2$	$\widehat{\theta}_3$	$\widehat{\theta}_1$	$\widehat{\theta}_2$	$\widehat{\theta}_3$	$\widehat{\theta}_1$	$\widehat{\theta}_2$	$\widehat{\theta}_3$
TBS(α, β, λ)	1.7432	23.407	-0.8649	1.3435	0.7525	-0.5496	0.5207	73.138	0.1133
	(0.1347)	(3.4801)	(0.1179)	(0.2526)	(0.3250)	(0.5869)	(0.0495)	(17.662)	(0.8308)
MOBS(η, α, β)	2.1975	1.5556	26.2253	1.8193	0.7451	1.2899	0.9127	0.5197	72.6927
	(0.5016)	(0.0923)	(4.3453)	(1.0968)	(0.1691)	(0.2793)	(0.7345)	(0.0471)	(17.025)
EBS(α, β, a)	2.1790	13.5871	2.5381	1.6597	0.4705	2.0973	0.4290	92.414	0.5295
	(0.2755)	(4.4206)	(0.5224)	(0.5221)	(0.3889)	(1.3836)	(0.7842)	(209.16)	(3.1384)
BS(α, β)	1.5147	41.3240	—	1.2745	1.0203	—	0.5199	70.857	—
	(0.0783)	(3.4959)	—	(0.1546)	(0.1826)	—	(0.0471)	(4.5565)	—

To confirm these facts, we apply goodness-of-fit tests detecting what distribution adjusts better each data set. We consider the Anderson–Darling (AD), Cramér–von Mises (CM) and Kolmogorov–Smirnov (KS) statistics; see Barros *et al.* (2014). Table 4 provides the p-values of the corresponding tests for S1, S2 and S3.

Table 4: p-value of the indicated statistic, model and data set.

Distribution	S1			S2			S3		
	KS	CM	AD	KS	CM	AD	KS	CM	AD
TBS	0.7919	0.5819	0.4715	0.9820	0.9332	0.9215	0.9996	0.9459	0.9202
MOBS	0.4789	0.1337	0.0903	0.9765	0.9167	0.9126	0.9995	0.9409	0.9160
EBS	0.7369	0.2015	0.1568	0.9828	0.9257	0.9212	0.9967	0.8925	0.8810
BS	0.1064	0.0501	0.0166	0.8441	0.8253	0.7129	0.9985	0.9409	0.9001

Thus, according to these tests, the TBS distribution fits the three data sets better than the other distributions, that is, such p-values indicate that all of the null hypotheses are strongly not rejected for the TBS distribution. Also, we compare the four distributions using the Akaike (AIC) and Bayesian (BIC) information criteria, as well as the Bayes factor (BF) to evaluate the magnitude of the difference between two BIC values; see Kass and Raftery (1995). Note that the BF coincides with the likelihood ratio test for nested models. We compute the AIC and BIC for the four distributions, whereas the BF is obtained to compare the distribution having a smaller BIC to the others. Decision about the best fit is made according to the interpretation of the BF presented in Table 6 of Leiva *et al.* (2015b). Table 5 provides the values of AIC, BIC and BF, indicating that the TBS distribution provides the best fit for S1 and a very competitive performance for S2 and S3.

Table 5: Value of the information criteria and BF for indicated model and data set.

Distribution	S1			S2			S3		
	AIC	BIC	BF	AIC	BIC	BF	AIC	BIC	BF
TBS	1467.1000	1476.8000	—	5.0275	9.6065	2.5891	418.7600	425.1000	4.0939
MOBS	1471.6000	1481.3000	4.4000	4.9694	9.5485	2.5311	418.7700	425.1000	4.1100
EBS	1469.0000	1478.7000	1.8000	5.0111	9.5902	2.5728	418.7800	425.1100	4.1000
BS	1481.5000	1488.0000	11.2000	3.9647	7.0174	—	416.7800	421.0000	—

These good results of the TBS distribution can be supported graphically in Figure 5, which displays the histograms with the estimated TBS PDFs (first panel/row) and the quantile versus quantile (QQ) plot with envelope (third panel/row). The QQ plot allows us to compare the empirical CDF and the estimated TBS CDF. All of these results of goodness of fit allow us to conclude the superiority of the TBS distribution in relation to the BS, EBS and MOBS distributions to model S1, S2 and S3. This shows the potential of the TBS distribution and the importance of the additional parameter. In addition, because the TBS distribution presents the best fit to the studied data sets, we analyze the influence of small perturbations in the ML estimates of its parameters. We use the scheme of case-weight perturbation. Figure 5 (second panel/row) sketches the influence index plot based on the TBS distribution for each data set. An inspection of these plots reveals that, as mentioned, none case appears with outstanding influence on the ML estimates of the TBS distribution parameters.

5. CONCLUSIONS AND FUTURE RESEARCH

We have used the transmutation method to define a new distribution that generalizes the Birnbaum–Saunders model, named the transmuted Birnbaum–Saunders distribution. Some relevant characteristics of the new distribution have been derived, such as the probabilistic functions, as well moments and a generator of random numbers. We have estimated the model parameters with the maximum likelihood method and its good performance has been evaluated by means of Monte Carlo simulations. Score vector and Hessian matrix were derived to infer about the model parameters. Diagnostic tools have been obtained to detect locally influential data in the maximum likelihood estimates. Potential applications of the new distribution have been considered by using three real-world data sets. Goodness-of-fit methods have demonstrated the suitable performance of the transmuted Birnbaum–Saunders distribution to these data in comparison to other versions of the Birnbaum–Saunders distribution. We hope that the new proposed distribution may attract wider applications in statistics. Modeling based on fixed, random and mixed effects, including semi-parametric formulations and non-parametric estimation of kernel, can be conducted with this new distribution. Multivariate versions, as well as copula methods, could also be addressed by the new transmuted Birnbaum–Saunders distribution.

ACKNOWLEDGMENTS

The authors thank the editors and reviewers for their constructive comments on an earlier version of this manuscript which resulted in this improved version. This research was supported by Capes and CNPq from Brazil, and FONDECYT 1160868 grant from Chile.

REFERENCES

[1] Ahmed, S.; Castro-Kuriss, C.; Leiva, V.; Flores, E. and Sanhueza, A. (2010). Truncated version of the Birnbaum–Saunders distribution with an application in financial risk, *Pakistan Journal of Statistics*, **26**, 293–311.

[2] Aryal, G. (2013). Transmuted log-logistic distribution, *Journal of Statistics Applications and Probability*, **2**, 11–20.

[3] Aryal, G. and Tsokos, C. (2009). On the transmuted extreme value distribution with application, *Nonlinear Analysis: Theory, Methods and Applications*, **71**, 1401–1407.

[4] Aryal, G. and Tsokos, C. (2011). Transmuted Weibull distribution: A generalization of the Weibull probability distribution, *European Journal of Pure and Applied Mathematics*, **4**, 89–102.

[5] Ashour, S. and Eltehiwy, M. (2013). Transmuted Lomax distribution, *American Journal of Applied Mathematics and Statistics*, **1**, 121–127.

[6] Athayde, E.; Azevedo, C.; Leiva, V. and Sanhueza, A. (2012). About Birnbaum–Saunders distributions based on the Johnson system, *Communications in Statistics: Theory and Methods*, **41**, 2061–2079.

[7] Azevedo, C.; Leiva, V.; Athayde, E. and Balakrishnan, N. (2012). Shape and change point analyses of the Birnbaum–Saunders-t hazard rate and associated estimation, *Computational Statistics and Data Analysis*, **56**, 3887–3897.

[8] Azzalini, A. (1985). A class of distributions which includes the normal ones, *Scandinavian Journal of Statistics*, **11**, 171–178.

[9] Balakrishnan, N.; Gupta, R.; Kundu, D.; Leiva, V. and Sanhueza, A. (2011). On some mixture models based on the Birnbaum–Saunders distribution and associated inference, *Journal of Statistical Planning and Inference*, **141**, 2175–2190.

[10] Barros, M.; Leiva, V.; Ospina, R. and Tsuyuguchi, A. (2014). Goodness-of-fit tests for the Birnbaum–Saunders distribution with censored reliability data, *IEEE Transactions on Reliability*, **63**, 543–554.

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

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

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

[14] Bourguignon, M.; Silva, R. and Cordeiro, G. (2014). A new class of fatigue life distributions, *Journal of Statistical Computation and Simulation*, **84**, 2619–2635.

[15] Castillo, N.; Gómez, H. and Bolfarine, H. (2011). Epsilon Birnbaum–Saunders distribution family: Properties and inference, *Statistical Papers*, **52**, 871–883.

[16] Cook, R.D. (1987). Influence assessment, *Journal of Applied Statistics*, **14**, 117–131.

[17] Cordeiro, M. and Lemonte, J. (2011). The β-Birnbaum–Saunders distribution: An improved distribution for fatigue life modeling, *Computational Statistics and Data Analysis*, **55**, 1445–1461.

[18] Cornish, E.A. and Fisher, R.A. (1937). Moments and cumulants in the specification of distributions, *Review of the International Statistical Institute*, **5**, 307–320.

[19] Cox, D.R. and Hinkley, D.V. (1974). *Theoretical Statistics*, Chapman and Hall, London, UK.

[20] Desousa, M.F.; Saulo, H.; Leiva, V. and Scalco, P. (2017). On a tobit-Birnbaum–Saunders model with an application to antibody response to vaccine, *Journal of Applied Statistics*, page in press available at `http://dx.doi.org/10.1080/02664763.2017.1322559`.

[21] Díaz-García, J. and Leiva, V. (2005). A new family of life distributions based on elliptically contoured distributions, *Journal of Statistical Planning and Inference*, **128**, 445–457.

[22] Doornik, J. (2006). *An Object-Oriented Matrix Language*, Timberlake Consultants Press, London, UK.

[23] Edgeworth, F.Y. (1917). On the mathematical representation of statistical data, *Journal of the Royal Statistical Society A*, **80**, 65–83; 266–288; 411–437.

[24] Ferreira, M.; Gomes, M.I. and Leiva, V. (2012). On an extreme value version of the Birnbaum–Saunders distribution, *REVSTAT Statistical Journal*, **10**, 181–210.

[25] Fierro, R.; Leiva, V.; Ruggeri, F. and Sanhueza, A. (2013). On a Birnbaum–Saunders distribution arising from a non-homogeneous Poisson process, *Statistical and Probability Letters*, **83**, 1233–1239.

[26] Garcia-Papani, F.; Uribe-Opazo, M.A.; Leiva, V. and Aykroyd, R.G. (2017). Birnbaum–Saunders spatial modelling and diagnostics applied to agricultural engineering data, *Stochastic Environmental Research and Risk Assessment*, **31**, 105–124.

[27] Gómez, H.; Olivares-Pacheco, J.F. and Bolfarine, H. (2009). An extension of the generalized Birnbaum–Saunders distribution, *Statistical and Probability Letters*, **79**, 331–338.

[28] Gradshteyn, I. and Randzhik, I. (2000). *Table of Integrals, Series, and Products*, Academic Press, New York, US.

[29] Guiraud, P.; Leiva, V. and Fierro, R. (2009). A non-central version of the Birnbaum–Saunders distribution for reliability analysis, *IEEE Transactions on Reliability*, **58**, 152–160.

[30] Johnson, N.; Kotz, S. and Balakrishnan, N. (1994). *Continuous Univariate Distributions*, volume 1, Wiley, New York, US.

[31] Johnson, N.; Kotz, S. and Balakrishnan, N. (1995). *Continuous Univariate Distributions*, volume 2, Wiley, New York, US.

[32] Johnson, N.L. (1949). Systems of frequency curves generated by methods of translation, *Biometrika*, **36**, 149–176.

[33] Kass, R. and Raftery, A. (1995). Bayes factors, *Journal of the American Statistical Association*, **90**, 773–795.

[34] Khan, M. and King, R. (2013). Transmuted modified Weibull distribution: A generalization of the modified Weibull probability distribution, *European Journal of Pure and Applied Mathematics*, **6**, 66–88.

[35] Khan, M. and King, R. (2014). A new class of transmuted inverse Weibull distribution for reliability analysis, *American Journal of Mathematical and Management*, **33**, 261–286.

[36] Kotz, S.; Leiva, V. and Sanhueza, A. (2010). Two new mixture models related to the inverse Gaussian distribution, *Methodology and Computing in Applied Probability*, **12**, 199–212.

[37] Lange, K. (2001). *Numerical Analysis for Statisticians*, Springer, New York, US.

[38] Leão, J.; Leiva, V.; Saulo, H. and Tomazella, V. (2017a). A survival model with Birnbaum–Saunders frailty for uncensored and censored cancer data, *Brazilian Journal of Probability and Statistics*, page in press.

[39] Leão, J.; Leiva, V.; Saulo, H. and Tomazella, V. (2017b). Birnbaum–Saunders frailty regression models: Diagnostics and application to medical data, *Biometrical Journal*, **59**, 291–314.

[40] Leiva, V. (2016). *The Birnbaum–Saunders Distribution*, Academic Press, New York, US.

[41] Leiva, V.; Athayde, E.; Azevedo, C. and Marchant, C. (2011). Modeling wind energy flux by a Birnbaum–Saunders distribution with unknown shift parameter, *Journal of Applied Statistics*, **38**, 2819–2838.

[42] Leiva, V.; Ferreira, M.; Gomes, M.I. and Lillo, C. (2016a). Extreme value Birnbaum–Saunders regression models applied to environmental data, *Stochastic Environmental Research and Risk Assessment*, **30**, 1045–1058.

[43] Leiva, V.; Liu, S.; Shi, L. and Cysneiros, F. (2016b). Diagnostics in elliptical regression models with stochastic restrictions applied to econometrics, *Journal of Applied Statistics*, **43**, 627–642.

[44] Leiva, V.; Marchant, C.; Ruggeri, F. and Saulo, H. (2015a). A criterion for environmental assessment using Birnbaum–Saunders attribute control charts, *Environmetrics*, **26**, 463–476.

[45] Leiva, V.; Marchant, C.; Saulo, H.; Aslam, M. and Rojas, F. (2014a). Capability indices for Birnbaum–Saunders processes applied to electronic and food industries, *Journal of Applied Statistics*, **41**, 1881–1902.

[46] Leiva, V.; Rojas, E.; Galea, M. and Sanhueza, A. (2014b). Diagnostics in Birnbaum–Saunders accelerated life models with an application to fatigue data, *Applied Stochastic Models in Business and Industry*, **30**, 115–131.

[47] Leiva, V.; Ruggeri, F.; Saulo, H. and Vivanco, J.F. (2017). A methodology based on the Birnbaum–Saunders distribution for reliability analysis applied to nano-materials, *Reliability Engineering and System Safety*, **157**, 192–201.

[48] Leiva, V.; Sanhueza, A. and Angulo, J.M. (2009). A length-biased version of the Birnbaum–Saunders distribution with application in water quality, *Stochastic Environmental Research and Risk Assessment*, **23**, 299–307.

[49] Leiva, V.; Santos-Neto, M.; Cysneiros, F.J.A. and Barros, M. (2014c). Birnbaum–Saunders statistical modelling: A new approach, *Statistical Modelling*, **14**, 21–48.

[50] Leiva, V.; Santos-Neto, M.; Cysneiros, F.J.A. and Barros, M. (2016c). A methodology for stochastic inventory models based on a zero-adjusted Birnbaum–Saunders distribution, *Applied Stochastic Models in Business and Industry*, **32**, 74–89.

[51] Leiva, V.; Saulo, H.; Leão, J. and Marchant, C. (2014d). A family of autoregressive conditional duration models applied to financial data, *Computational Statistics and Data Analysis*, **79**, 175–191.

[52] Leiva, V.; Tejo, M.; Guiraud, P.; Schmachtenberg, O.; Orio, P. and Marmolejo, F. (2015b). Modeling neural activity with cumulative damage distributions, *Biological Cybernetics*, **109**, 421–433.

[53] Lemonte, A. (2013). A new extension of the Birnbaum–Saunders distribution, *Brazilian Journal of Probability and Statistics*, **27**, 133–149.

[54] Lillo, C.; Leiva, V.; Nicolis, O. and Aykroyd, R.G. (2016). L-moments of the Birnbaum–Saunders distribution and its extreme value version: Estimation, goodness of fit and application to earthquake data, *Journal of Applied Statistics*, page in press available at `http://dx.doi.org/10.1080/02664763.2016.1269729`.

[55] Liu, S.; Leiva, V.; Ma, T. and Welsh, A. (2016). Influence diagnostic analysis in the possibly heteroskedastic linear model with exact restrictions, *Statistical Methods and Applications*, **5**, 227–249.

[56] Louzada, F. and Granzotto, D. (2016). The transmuted log-logistic regression model: A new model for time up to first calving of cows, *Statistical Papers*, **57**, 623–640.

[57] Marchant, C.; Bertin, K.; Leiva, V. and Saulo, H. (2013). Generalized Birnbaum–Saunders kernel density estimators and an analysis of financial data, *Computational Statistics and Data Analysis*, **63**, 1–15.

[58] Marchant, C.; Leiva, V. and Cysneiros, F.J.A. (2016a). A multivariate log-linear model for Birnbaum–Saunders distributions, *IEEE Transactions on Reliability*, **65**, 816–827.

[59] Marchant, C.; Leiva, V.; Cysneiros, F.J.A. and Vivanco, J.F. (2016b). Diagnostics in multivariate generalized Birnbaum–Saunders regression models, *Journal of Applied Statistics*, **43**, 2829–2849.

[60] Martinez, G.; Bolfarine, H. and Gómez, H. (2014). An alpha-power extension for the Birnbaum–Saunders distribution, *Statistics*, **48**, 896–912.

[61] McNamee, J. and Pa, V. (2013). *Numerical Methods for Roots of Polynomials, Part II*, Academic Press, New York, US.

[62] Merovci, F. and Puka, L. (2014). Transmuted Pareto distribution, *ProbStat Forum*, **7**, 1–11.

[63] Mohammadi, K.; Alavi, O. and McGowan, J.G. (2017). Use of Birnbaum–Saunders distribution for estimating wind speed and wind power probability distributions: A review, *Energy Conversion and Management*, **143**, 109–122.

[64] Mroz, T. (2013a). Transmuted Lindley distribution, *International Journal of Open Problems in Computer Science*, **6**, 63–72.

[65] Mroz, T. (2013b). Transmuted Rayleigh distribution, *Austrian Journal of Statistics*, **42**, 21–31.

[66] Nocedal, J. and Wright, S. (1999). *Numerical Optimization*, Springer, New York, US.

[67] Owen, W. (2006). A new three-parameter extension to the Birnbaum–Saunders distribution, *IEEE Transactions on Reliability*, **55**, 475–479.

[68] Pearson, K. (1895). Mathematical contribution to the theory of evolution. II. Skew variation in homogeneous material, *Philosophical Transactions of the Royal Society of London A*, **186**, 343–414.

[69] Press, W.; Teulosky, S.; Vetterling, W. and Flannery, B. (2007). *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press, Cambridge, UK.

[70] Proschan, F. (1963). Theoretical explanation of observed decreasing failure rate, *Technometrics*, **15**, 375–383.

[71] R Core Team (2016). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

[72] Rojas, F.; Leiva, V.; Wanke, P. and Marchant, C. (2015). Optimization of contribution margins in food services by modeling independent component demand, *Revista Colombiana de Estadística*, **38**, 1–30.

[73] Rousseeuw, P.; Croux, C.; Todorov, V.; Ruckstuhl, A.; Salibian-Barrera, M.; Verbeke, T.; Koller, M. and Maechler, M. (2016). *robustbase: Basic Robust Statistics*, R package version 0.92-6.

[74] Saboor, A.; Kamal, M. and Ahmad, M. (2015). The transmuted exponential-Weibull distribution with applications, *Pakistan Journal of Statistics*, **13**, 229–250.

[75] Santos-Neto, M.; Cysneiros, F.J.A.; Leiva, V. and Ahmed, S. (2012). On new parameterizations of the Birnbaum–Saunders distribution, *Pakistan Journal of Statistics*, **28**, 1–26.

[76] Santos-Neto, M.; Cysneiros, F.J.A.; Leiva, V. and Barros, M. (2014). On new parameterizations of the Birnbaum–Saunders distribution and its moments, estimation and application, *REVSTAT Statistical Journal*, **12**, 247–272.

[77] Santos-Neto, M.; Cysneiros, F.J.A.; Leiva, V. and Barros, M. (2016). Reparameterized Birnbaum–Saunders regression models with varying precision, *Electronic Journal of Statistics*, **10**, 2825–2855.

[78] Saulo, H.; Leão, J.; Leiva, V. and Aykroyd, R.G. (2017). Birnbaum–Saunders autoregressive conditional duration models applied to high-frequency financial data, *Statistical Papers*, pages in press available at `http://dx.doi.org/10.1007/s00362-017-0888-6`.

[79] Saulo, H.; Leão, V. and Bourguignon, M. (2012). The Kumaraswamy Birnbaum–Saunders distribution, *Journal of Statistical Theory and Practice*, **1**, 1–13.

[80] Saulo, H.; Leiva, V.; Ziegelmann, F.A. and Marchant, C. (2013). A nonparametric method for estimating asymmetric densities based on skewed Birnbaum–Saunders distributions applied to environmental data, *Stochastic Environmental Research and Risk Assessment*, **27**, 1479–1491.

[81] Sharma, V.; Singh, S. and Singh, U. (2014). A new upside-down bathtub shaped hazard rate model for survival data analysis, *Applied Mathematics and Computation*, **239**, 242–253.

[82] Shaw, W. and Buckley, I. (2009). The alchemy of probability distributions: Beyond Gram–Charlier expansions, and a skew-kurtotic-normal distribution from a rank transmutation map, *Working Paper*, available at `http://arxiv.org/abs/0901.0434v1`:1–8.

[83] Tiana, Y.; Tiana, M. and Zhu, Q. (2014). Transmuted linear exponential distribution: A new generalization of the linear exponential distribution, *Communications in Statistics: Simulation and Computation*, **43**, 2661–2677.

[84] Vilca, F. and Leiva, V. (2006). A new fatigue life model based on the family of skew-elliptical distributions, *Communications in Statistics: Theory and Methods*, **35**, 229–244.

[85] Vilca, F.; Sanhueza, A.; Leiva, V. and Christakos, G. (2010). An extended Birnbaum–Saunders model and its application in the study of environmental quality in Santiago, Chile, *Stochastic Environmental Research and Risk Assessment*, **24**, 771–782.

[86] Villegas, C.; Paula, G. and Leiva, V. (2011). Birnbaum–Saunders mixed models for censored reliability data analysis, *IEEE Transactions on Reliability*, **60**, 748–758.

[87] Volodin, I. and Dzhungurova, O. (2000). On limit distribution emerging in the generalized Birnbaum–Saunders model, *Journal of Mathematical Science*, **99**, 1348–1366.

[88] Wanke, P. and Leiva, V. (2015). Exploring the potential use of the Birnbaum–Saunders distribution in inventory management, *Mathematical Problems in Engineering*, Article ID 827246:1–9.

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.

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.

Abstract and Indexing Services

The REVSTAT is covered by the following abstracting/indexing services:

- Current Index to Statistics
- Google Scholar
- Mathematical Reviews
- Science Citation Index Expanded
- Zentralblatt für Mathematic

Instructions to Authors, special-issue editors and publishers

The articles should be written in English and may be submitted in two different ways:

- By sending the paper in PDF format to the Executive Editor (revstat@ine.pt) and to one of the two Editors or Associate Editors, whose opinion the author wants to be taken into account, together to the following e-mail address: revstat@fc.ul.pt

- By sending the paper in PDF format to the Executive Editor (revstat@ine.pt), together with the corresponding PDF or PostScript file to the following e-mail address: revstat@fc.ul.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 may be obtained in the above link.

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

M. Ivette Gomes, Faculdade de Ciências, Universidade de Lisboa, Portugal

Co-Editor

M. Antónia Amaral Turkman, Faculdade de Ciências, Universidade de Lisboa, Portugal

Associate Editors

Barry Arnold, University of California, Riverside, USA

Jan Beirlant, Katholieke Universiteit Leuven, Leuven, Belgium

Graciela Boente, Facultad de Ciencias Exactas and Naturales, Buenos Aires, Argentina

João Branco, Instituto Superior Técnico, Universidade de Lisboa, Portugal

Carlos Agra Coelho (2017-2018), Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Portugal

David Cox, Oxford University, United Kingdom

Isabel Fraga Alves, Faculdade de Ciências, Universidade de Lisboa, Portugal

Dani Gammerman (2014-2016), Federal University of Rio de Janeiro, Brazil

Wenceslao Gonzalez-Manteiga, University of Santiago de Compostela, Spain

Juerg Huesler, University of Bern, Switzerland

Marie Husková, Charles University of Prague, Czech Republic

Victor Leiva, Universidad Adolfo Ibáñez, Chile

Isaac Meilijson, University of Tel-Aviv, Israel

M. Nazaré Mendes- Lopes, Universidade de Coimbra, Portugal

Stephen Morghenthaler, University Laval, sainte-Foy, Canada

António Pacheco, Instituto Superior Técnico, Universidade de Lisboa, Portugal

Carlos Daniel Paulino, Instituto Superior Técnico, Universidade de Lisboa, Portugal

Dinis Pestana, Faculdade de Ciências, Universidade de Lisboa, Portugal

Arthur Pewsey, University of Extremadura, Spain

Vladas Pipiras, University of North Carolina, USA

Gilbert Saporta, Conservatoire National des Arts et Métiers (CNAM), Paris, France

Julio Singer, University of San Paulo, Brasil

Jef Teugel, Katholieke Universiteit Leuven, Belgium

Feridun Turrhman, Faculdade de Ciências, Universidade de Lisboa, Portugal

Executive Editor

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