Brazilian Journal of Probability and Statistics
2020, Vol. 34, No. 3, 580–593
https://doi.org/10.1214/19-BJPS450

Improved estimators of the entropy in scale mixture of exponential distributions

Constantinos Petropoulos[a], **Lakshmi Kanta Patra**[b] and **Somesh Kumar**[c]

[a]*University of Patras*
[b]*Indian Institute of Petroleum and Energy Visakhapatnam*
[c]*Indian Institute of Technology Kharagpur*

Abstract. In the present communication, the problem of estimating entropy of a scale mixture of exponential distributions is considered under the squared error loss. Inadmissibility of the best affine equivariant estimator(BAEE) is established by deriving an improved estimator which is not smooth. Using the integral expression of risk difference (IERD) approach of Kubokawa (*The Annals of Statistics* **22** (1994) 290–299), classes of estimators are obtained which improve upon the BAEE. The boundary estimator of this class is the Brewster and Zidek-type estimator and this estimator is smooth. We have shown that the Brewster and Zidek-type estimator is a generalized Bayes estimator. As an application of these results, we have obtained improved estimators for the entropy of a multivariate Lomax distribution. Finally, percentage risk reduction of the improved estimators for the entropy of a multivariate Lomax distribution is plotted to compare the risk performance of the improved estimators.

1 Introduction

It is well known that entropy and information can be considered as a measure of uncertainty of probability distribution. In the literature, several examples of entropies are proposed. Among all these entropies the most famous is the Shannon entropy. The Shannon's entropy of a random variable X with density function $f(x;\theta)$ is defined by

$$Q(\theta) = E\bigl(-\ln f(x;\theta)\bigr). \tag{1}$$

For a detailed description on the entropy, one may refer to Robinson (2008), Broadbridge and Guttmann (2009) and Cover and Thomas (2012).

Shannon's Entropy is widely used in various areas of science and technology such as ecology, hydrology and water resources, social studies, economics, biology etc. In molecular sciences, estimation of the entropy of molecules plays an important role to understand various chemical and biological processes (see Nalewajski (2002)). The concept of entropy is used in software reliability to measure uncertainty (see Kamavaram and Goseva-Popstojanova (2002)). In estimating uncertainty of a system with several independent components (connected in parallel/series or both) we need to estimate uncertainty in individual components. In economics, entropy estimation (see Golan, Judge and Miller (1996)) often allows the researchers to use data for the improvement of the assumptions on the parameters in econometric models.

Similar to mean, standard deviation, variance and quantile, entropy is also an important characteristic of a parametric family of distributions. Recently, Devi, Kumar and Kour (2017) derived the entropy of Lomax distribution which is used in business, economics etc. In the

Key words and phrases. Decision theory, squared error loss, best affine equivariant estimator, multivariate Lomax distribution.

Received February 2018; accepted May 2019.

present article, we are interested to estimate it for further insights into the nature of the distribution.

In the last decade, several researchers studied the estimation of Shannon's entropy of various continuous probability distributions. Misra, Singh and Demchuk (2005) considered the problem of estimating Shannon's entropy of a multivariate normal distribution under the quadratic loss function. They obtained improvements over the BAEE using the techniques of Stein (1964) and Brewster and Zidek (1974). The problem of estimating the entropy of an exponential distribution under a linex loss function was considered by Kayal and Kumar (2011). They proved the minimaxity and admissibility of the best scale equivariant estimator of the entropy of the negative exponential model. It was shown that for the shifted exponential distribution the BAEE of the entropy is inadmissible. Kayal and Kumar (2013) considered the estimation of the entropy of several shifted exponential populations with different locations, but a common scale parameter with respect to squared error loss function. They proved a general inadmissibility result for the scale equivariant estimator. Recently Kayal et al. (2015) have considered the problem of estimating the Renyi entropy of k exponential populations with a common location but different scale parameters. They derived the uniformly minimum variance unbiased estimator of the Renyi entropy. They also obtained the sufficient conditions for improvement over affine and scale equivariant estimators. Recently, Kayal and Kumar (2017) studied the problem of estimating entropy of several exponential distributions under linex loss function. They have proved that the best affine equivariant estimator is inadmissible by deriving an improved estimator. For some recent work on the estimation of entropy of logistic, half-logistic and generalized half-logistic distributions one may also refer to Kang et al. (2012), Seo and Kang (2014, 2015) and Seo and Kim (2017).

In this paper, we consider the problem of estimating Shannon's entropy of a scale mixture of exponential distributions. Consider the random vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$, $n \geq 1$, where for a given $\tau > 0$, $X_1, X_2, \ldots, X_n$ are i.i.d. with common exponential distribution $\exp(\mu, \sigma/\tau)$. So the unconditional density function of $X_1, X_2, \ldots, X_n$ is

$$\begin{aligned} &f(x_1, x_2, \ldots, x_n; \mu, \sigma) \\ &\quad = \int_0^\infty \frac{\tau^n}{\sigma^n} \exp\left\{-\frac{\tau}{\sigma}\sum_{i=1}^{n}(x_i - \mu)\right\} I_{(\mu,\infty)}(x_{(1)})\, dG(\tau), \end{aligned} \tag{2}$$

where $x_{(1)} = \min\{x_1, \ldots, x_n\}$ and $I_A(\cdot)$ denotes the indicator function of the set A. The mixing parameter τ is assumed to have a distribution function $G(\cdot)$ defined on the positive real line. Here we use the symbol τ to denote either a random variable having distribution function $G(\cdot)$ and density function $\nu(\cdot)$ or a fixed value of this random variable.

The mixture model (2) was proposed by Lindley and Singpurwalla (1986) in connection with certain reliability problems. For the estimation problem of the parameters of the model (2), one may refer to Petropoulos and Kourouklis (2005) and Petropoulos (2006, 2010).

Denote by $X_{(1)} = \min\{X_1, \ldots, X_n\}$. Let $X = nX_{(1)}$ and $S = \sum_{i=1}^{n}(X_i - X_{(1)})$. A minimal sufficient statistics for this model is (X, S). For a given $\tau > 0$, $X|\tau \sim \exp(n\mu, \sigma/\tau)$ and $S|\tau \sim \Gamma(n-1, \sigma/\tau)$. Let $V = \tau S/\sigma$ and $U = \tau X/\sigma$, then V and U have respective densities

$$\begin{aligned} g(v) &= \frac{1}{\Gamma(n-1)} v^{n-2} e^{-v}, \quad v > 0 \quad \text{and} \\ h(u, \rho) &= e^{-(u - n\tau\mu/\sigma)}, \quad u > n\tau\mu/\sigma. \end{aligned} \tag{3}$$

Clearly if $G(\cdot)$ is degenerate at $\tau = 1$, we get the exponential distribution with location parameter μ and scale parameter σ. We denote $\rho = n\tau\mu/\sigma$. In specific situations for the mixing distribution $G(\cdot)$, we can provide known distributions in the literature such as multivariate Lomax distribution and Exponential Inverse Gaussian (EIG) model (see, for instance, Bhattacharya and Kumar (1986)).

If $G(\cdot)$ is a gamma distribution with density function $\tau^{b-1}e^{-\tau}/\Gamma(b)$, $\tau > 0$, then (2) becomes the multivariate Lomax distribution with the density

$$f(x_1, x_2, \dots, x_n; \mu, \sigma) = \frac{\Gamma(b+n)}{\Gamma(b)\sigma^n}\left(1 + \frac{1}{\sigma}\sum_{i=1}^{n}(x_i - \mu)\right)^{-(b+n)} I_{(\mu,\infty)}(x_{(1)}). \tag{4}$$

This distribution is also referred as multivariate Pareto type II distribution. For details, see Johnson, Kotz and Balakrishnan (2002) and Arnold (2015). Nayak (1987) reported several properties of the multivariate Lomax distribution and he also mentioned their usefulness in reliability theory. Multivariate Lomax distribution is a generalization of the univariate Lomax distribution. Lomax (1954) used the univariate Lomax distribution in the analysis of income data and business failure data. It is seen from Ahmed and Gokhale (1989), that the estimation of the entropy in model (4) is equivalent to the estimation of $\ln \sigma$ for scale mixture of exponential distributions.

Here we consider the estimation of $\theta = \ln \sigma$ under the squared error loss function

$$L(\theta, \delta) = (\delta - \theta)^2. \tag{5}$$

The problem under study, is invariant under the group of affine transformations, given by $G_{a,b} = \{g_{a,b} : g_{a,b}(x) = ax + b; a > 0, b \in \mathbb{R}\}$. Under the transformation $g_{a,b}$, $\mu \to a\mu + b$, $\sigma \to a\sigma$ and consequently $\ln \sigma \to \ln \sigma + \ln a$. The form of an affine equivariant estimator is

$$\delta_c(X, S) = \ln S + c, \tag{6}$$

where c is a real number.

Lemma 1.1. *Under the squared error loss function* (5), *the BAEE of* θ *is* δ_{c_0} *with* $c_0 = E(\ln \tau) - \psi(n-1)$, *where* $\psi(t)$ *denotes the Euler psi* (*digamma*) *function*, *defined as* $\psi(t) = \frac{d}{dt}\ln \Gamma(t)$.

Proof. For any real constant the risk of the estimators δ_c of the form (6) is

$$R(\delta_c, \mu, \sigma) = E(\ln S + c - \ln \sigma)^2. \tag{7}$$

Differentiating $R(\delta_c, \mu, \sigma)$ with respect to c and equating to zero, we have the minimizing choice of c is

$$c_0 = -E^{\tau}E\big[\ln(S/\sigma)|\tau\big] = -E^{\tau}E\big[\ln(\tau S/\sigma) - \ln \tau|\tau\big] = E(\ln \tau) - E(\ln V), \tag{8}$$

where $V = \frac{\tau S}{\sigma} \sim \Gamma(n-1, 1)$.

But, $E(\ln V) = \int_0^\infty \ln v \frac{1}{\Gamma(n-1)} v^{n-2}e^{-v}\,dv = \frac{\Gamma'(n-1)}{\Gamma(n-1)} = \psi(n-1)$, because $\Gamma(z) = \int_0^\infty x^{z-1}e^{-x}\,dx$ and $\Gamma'(z) = \int_0^\infty \ln x x^{z-1}e^{-x}\,dx$. So, from Equation (8) we have

$$c_0 = E(\ln \tau) - \psi(n-1). \tag{9}$$

This proves the lemma. □

The rest of the paper organized as follows. In Section 2, we have established that the BAEE of θ is inadmissible by deriving improved estimators. To obtain the improved estimator we adopt the approach of Stein (1964). In Section 3, we propose another class of estimators and using the IERD approach of Kubokawa (1994) we derive sufficient conditions under which the proposed estimators improve upon the BAEE of θ. We have also proved that a generalized Bayes estimator coincides with the Brewster and Zidek-type estimator. In particular we have obtained the explicit estimators of the entropy of a multivariate Lomax distribution in Section 4. We also provide plots percentage risk reduction of proposed estimators for illustration purpose.

The following lemma due to Bobotas and Kourouklis (2009) has been used in this paper. For the sake of completeness, we state it below.

Lemma 1.2. *Let $a_1(x)$ and $a_2(x)$ be densities supported on the intervals Ω_1 and Ω_2 respectively, where $\Omega_1 \subseteq \Omega_2$ and $a_1(x)/a_2(x)$ is increasing in $x \in \Omega_1$. If X is a random variable having density $a_1(x)$ or $a_2(x)$ and $g(x)$, $x \in \Omega_1$, is increasing (decreasing) then $E_{a_1} g(X) \geq (\leq) E_{a_2} g(X)$. Moreover, if $h(x)$ and $a_1(x)/a_2(x)$ are strictly monotone, then $E_{a_1} g(X) > (<) E_{a_2} g(X)$.*

2 Inadmissibility of the best affine equivariant estimator

In this section, we will derive an estimator that dominates the BAEE δ_{c_0}. To do so, we explore a larger class than the affine equivariant estimators. We consider one such class of estimators as

$$\delta_\phi(X, S) = \ln S + \phi(W), \tag{10}$$

where $W = X/S$. Notice that the BAEE belongs in the above class, corresponding to $\phi_0(W) = c_0$. Define

$$\phi_1(W) = \ln(1+W) + \frac{E(\tau^n \ln \tau)}{E\tau^n} - \psi(n) \quad \text{for } W > 0.$$

Theorem 2.1. *Let $\phi_{ST1}(W) = \min\{\phi_1(W), c_0\}$. Then, under the squared error loss function* (5) *the best affine equivariant estimator δ_{c_0} is inadmissible and is dominated by the estimator*

$$\delta_{ST1}(X, S) = \begin{cases} \ln S + \phi_{ST1}(W), & \textit{if } W > 0, \\ \ln S + c_0, & \textit{otherwise} \end{cases}$$

provided $\phi_1(w) < c_0$ on a set of positive probability.

Proof. The risk of the estimator $\delta_\phi(W)$ given by (10) depends on (μ, σ) only through μ/σ, so without loss of generality we take $\sigma = 1$ and the risk function is

$$R(\delta, \mu) = E^W \{ E [(\ln S + \phi(W) - \ln \sigma)^2 | W = w] \}.$$

The conditional expectation $E[(\ln S + \phi(W) - \ln\sigma)^2 | W = w]$ in minimized at

$$\phi_{\text{op}}(\mu, w) = -E[\ln S | W = w]. \tag{11}$$

In the process, we have to find the upper bound of $\phi_{\text{op}}(\mu, w)$ as a function of μ for each $w > 0$. The joint density of S and W is

$$f(s, w) = \frac{1}{\Gamma(n-1)} s^{n-1} \tau^n e^{n\mu\tau} e^{-\tau s(1+w)}, \quad w > \frac{n\mu}{s}, s > 0.$$

Now we find the conditional distribution of S given $W = w$.

Case I: $\mu \geq 0$, $w > 0$.

In this case $\frac{n\mu}{w} < s < \infty$. For a given $\tau > 0$ the conditional distribution of S given $W = w$ is

$$f(s|w) = \frac{e^{n\mu\tau} \tau^n s^{n-1} e^{-\tau s(1+w)}}{\int_0^\infty \int_{n\mu/w}^\infty e^{n\mu\tau} \tau^n s^{n-1} e^{-\tau s(1+w)} \, ds \, dG(\tau)}.$$

So we have

$$\phi_{\text{op}}(\mu, w) = -\frac{\int_0^\infty \int_{n\mu/w}^\infty \ln s e^{n\mu\tau} \tau^n s^{n-1} e^{-\tau s(1+w)} \, ds \, dG(\tau)}{\int_0^\infty \int_{n\mu/w}^\infty e^{n\mu\tau} \tau^n s^{n-1} e^{-\tau s(1+w)} \, ds \, dG(\tau)}.$$

By the transformation $z = \tau s(1+w)$ we get

$$\phi_{\text{op}}(\mu, w) = \ln(1+w) + H_1(w) - H_2(w), \tag{12}$$

where

$$H_1(w) = \frac{\int_0^\infty \int_\xi^\infty \ln \tau e^{n\mu\tau} e^{-z} z^{n-1}\, dz\, dG(\tau)}{\int_0^\infty \int_\xi^\infty e^{n\mu\tau} e^{-z} z^{n-1}\, dz\, dG(\tau)}, \quad \text{and}$$

$$H_2(w) = \frac{\int_0^\infty \int_\xi^\infty \ln z e^{n\mu\tau} e^{-z} z^{n-1}\, dz\, dG(\tau)}{\int_0^\infty \int_\xi^\infty e^{n\mu\tau} e^{-z} z^{n-1}\, dz\, dG(\tau)}$$

with $\xi = \frac{\tau n\mu(1+w)}{w}$. Now for a given $\tau > 0$

$$\frac{\int_\xi^\infty \ln z e^{-z} z^{n-1}\, dz}{\int_\xi^\infty e^{-z} z^{n-1}\, dz} = E_\xi \ln Z,$$

where Z has density $g(z, \xi) \propto e^{-z} z^{n-1} I_{(\xi,\infty)}(z)$. By Lemma (1.2), for $\xi > 0$ we have $E_\xi \ln Z \geq E_0 \ln Z = \psi(n)$ and hence

$$\int_\xi^\infty \ln z e^{-z} z^{n-1}\, dz \geq \psi(n) \int_\xi^\infty e^{-z} z^{n-1}\, dz \quad \text{which implies that } H_2(W) \geq \psi(n). \tag{13}$$

Again by the transformation $\tau p = z$ and for $\xi_2 = \frac{n\mu(1+w)}{w}$, $H_1(w)$ becomes

$$\begin{aligned} H_1(w) &= \frac{\int_0^\infty \int_{\xi_2}^\infty \ln \tau e^{n\mu\tau} e^{-\tau p} \tau^n p^{n-1}\, dp\, dG(\tau)}{\int_0^\infty \int_{\xi_2}^\infty e^{n\mu\tau} e^{-\tau p} \tau^n p^{n-1}\, dp\, dG(\tau)} \\ &= \frac{\int_{\xi_2}^\infty p^{n-1} \int_0^\infty \ln \tau \tau^n e^{-(p-n\mu)\tau}\, dG(\tau)\, dp}{\int_{\xi_2}^\infty p^{n-1} \int_0^\infty \tau^n e^{-(p-n\mu)\tau}\, dG(\tau)\, dp}. \end{aligned} \tag{14}$$

Now for $\lambda = p - n\mu > 0$

$$\frac{\int_0^\infty \ln \tau e^{-(p-n\mu)\tau} \tau^n\, dG(\tau)}{\int_0^\infty \tau^n e^{-(p-n\mu)\tau}\, dG(\tau)} = E_\lambda \ln \tau,$$

where τ have the density $f(\tau; \lambda) \propto e^{-\lambda\tau} \tau^n \nu(\tau)$. But, $\frac{f(\tau;\lambda)}{f(\tau;0)}$ is nonincreasing in τ, so using Lemma (1.2) we have

$$E_\mu \ln \tau \leq E_0 \ln \tau = \frac{E(\tau^n \ln \tau)}{E\tau^n}. \tag{15}$$

Using (13) and (15) from (12), we get

$$\phi_{\text{op}}(w) \leq \ln(1+w) + \frac{E(\tau^n \ln \tau)}{E(\tau^n)} - \psi(n). \tag{16}$$

Case II: $\mu < 0$, $w > 0$.

In this case $0 < s < \infty$. For a given $\tau > 0$, the conditional distribution of S given $W = w$ is

$$f(s|w) = \frac{e^{n\mu\tau} \tau^n s^{n-1} e^{-\tau s(1+w)}}{\int_0^\infty \int_0^\infty e^{n\mu\tau} \tau^n s^{n-1} e^{-\tau s(1+w)}\, ds\, dG(\tau)}.$$

By the transformation $z = \tau s(1+w)$, we get

$$\phi_{\text{op}}(\mu, w) = \ln(1+w) + K_1(w) - K_2(w), \tag{17}$$

where

$$K_1(w) = \frac{\int_0^\infty \int_0^\infty \ln \tau e^{n\mu\tau} e^{-z} z^{n-1}\, dz\, dG(\tau)}{\int_0^\infty \int_0^\infty e^{n\mu\tau} e^{-z} z^{n-1}\, dz\, dG(\tau)}, \quad \text{and}$$

$$K_2(w) = \frac{\int_0^\infty \int_0^\infty \ln z e^{n\mu\tau} e^{-z} z^{n-1}\, dz\, dG(\tau)}{\int_0^\infty \int_0^\infty e^{n\mu\tau} e^{-z} z^{n-1}\, dz\, dG(\tau)}.$$

Proceeding in a similar way as in Case I, we have $K_2(w) = \psi(n)$ and $K_1(w) \le \frac{E(\tau^n \ln \tau)}{E\tau^n}$, so

$$\phi_{\text{op}}(w) \le \ln(1+w) + \frac{E(\tau^n \ln \tau)}{E\tau^n} - \psi(n). \tag{18}$$

Hence for any μ and $w > 0$, we have

$$\phi_{\text{op}}(w) \le \ln(1+w) + \frac{E(\tau^n \ln \tau)}{E\tau^n} - \psi(n) = \phi_1(w). \tag{19}$$

Since the risk function $R^*(c, w) = E[(\ln S + c - \ln \sigma)^2 | W = w]$ is convex in c, then $R^*(c, w)$ is strictly increasing in c for $c > \phi_{\text{op}}(\mu, w)$. Now if $\phi_1(w) \le c_0$ on a set of positive probability, then we have $E[(\ln S + \phi_1(W) - \ln \sigma)^2 | W = w] \le E[(\ln S + c_0 - \ln \sigma)^2 | W = w]$. This proves the theorem. □

Remark 2.1. When $\tau = 1$ with probability (w.p.) 1, Theorem 2.1 provides the improved estimator upon the BAEE derived in Kayal and Kumar (2013, Corollary 1) in the case of $k = 1$.

3 A class of improved estimators

In this section, we consider a class of estimators of the form

$$\delta_\phi(X, S) = \begin{cases} \ln S + \phi(W), & \text{if } W > 0, \\ \delta_{c_0}, & \text{otherwise}, \end{cases} \tag{20}$$

where ϕ is an absolutely continuous function. Employing IERD approach of Kubokawa (1994) we prove that the class of estimators defined in (20) dominate the BAEE δ_{c_0} for estimating θ under the squared error loss function (5). Define

$$H(u, \rho) = \int_0^u h(x; \rho) I(x > \rho)\, dx \quad \text{and}$$

$$H(u) = \int_0^u h(x)\, dx, \quad \text{where } h(x) = h(x; 0).$$

Theorem 3.1. *The estimator δ_ϕ improves upon the BAEE δ_{c_0} for estimating θ under the squared error loss function (5), if the function ϕ satisfies the following conditions*

(i) $\phi(z)$ *is nondecreasing with* $\lim_{z\to\infty} \phi(z) = E(\ln \tau) - \psi(n-1)$
(ii) $\phi(z) \ge \phi^*(z) = E \ln \tau - \psi(n-1) - \frac{\ln(1+z)}{(1+z)^{n-1} - 1}$

Proof. For $t > 0$, the risk difference of δ_{c_0} and δ_ϕ can be written as

$$\begin{aligned} \Delta(\delta_{c_0}, \delta_\phi) &= 2E\left\{\int_1^\infty \left(\ln\left(\frac{S}{\sigma}\right) + \phi(Wt)\right)\phi'(Wt) W\, dt\right\} I(W > 0) \\ &= 2\int_0^\infty \int_0^\infty \int_0^\infty \int_1^\infty (\ln y + \phi(wt))\phi'(wt) w g(y\tau) h(w\tau y; \rho) \\ &\quad \times I(w\tau y > \rho)\tau^2 y\, dt\, dy\, dw\, dG(\tau). \end{aligned}$$

Substituting $v = \tau y$ the risk difference reduces to

$$\Delta(\delta_{c_0}, \delta_\phi) = 2\int_0^\infty \int_0^\infty \int_0^\infty \int_1^\infty \{\ln(v/\tau) + \phi(wt)\}\phi'(wt)wvg(v)h(vw;\rho) \times I(vw > \rho)\,dt\,dv\,dw\,dG(\tau).$$

Making the transformation $z = tw$ and then $x = z/t$, the risk difference becomes

$$\begin{aligned}\Delta(\delta_{c_0}, \delta_\phi) &= 2\int_0^\infty \int_0^\infty \int_0^\infty \int_0^z \{\ln v - \ln\tau + \phi(z)\}v\phi'(z)g(v)h(xv;\rho) \\ &\quad \times I(xv > \rho)\,dx\,dv\,dz\,dG(\tau) \\ &= 2\int_0^\infty \phi'(z)\left[\int_0^\infty \int_0^\infty \{\ln v - \ln\tau + \phi(z)\}g(v)H(zv;\rho)\,dv\,dG(\tau)\right]dz.\end{aligned}$$

Under the condition (i), that is, $\phi'(z) \geq 0$ the risk difference is nonnegative provided

$$\phi(z) \geq \frac{\int_0^\infty \int_0^\infty \ln\tau g(v)H(zv;\rho)\,dv\,dG(\tau)}{\int_0^\infty \int_0^\infty g(v)H(zv;\rho)\,dv\,dG(\tau)} - \frac{\int_0^\infty \int_0^\infty \ln v g(v)H(zv;\rho)\,dv\,dG(\tau)}{\int_0^\infty \int_0^\infty g(v)H(zv;\rho)\,dv\,dG(\tau)}. \tag{21}$$

Now we can write

$$\frac{\int_0^\infty \int_0^\infty \ln v g(v)H(zv;\rho)\,dv\,dG(\tau)}{\int_0^\infty \int_0^\infty g(v)H(zv;\rho)\,dv\,dG(\tau)} = E_\rho \ln V,$$

where the density of v is $f_\rho(v) \propto g(v)H(zv;\rho)$.

In the case $\mu \leq 0$, that means $\rho \leq 0$, $I(zv > \rho) = 1$, so $H(zv;\rho) = e^\rho H(zv)$ and

$$\begin{aligned}E_\rho \ln V &= \frac{\int_0^\infty \int_0^\infty \ln v g(v)e^\rho H(zv)\,dv\,dG(\tau)}{\int_0^\infty \int_0^\infty g(v)e^\rho H(zv)\,dv\,dG(\tau)} \\ &= \frac{\int_0^\infty \ln v g(v)H(zv)\,dv \int_0^\infty e^\rho\,dG(\tau)}{\int_0^\infty g(v)H(zv)\,dv \int_0^\infty e^\rho\,dG(\tau)}.\end{aligned}$$

Or else,

$$E_\rho \ln V = \psi(n-1) + \frac{\ln(1+z)}{(1+z)^{n-1} - 1}. \tag{22}$$

For $\rho > 0$, $f_\rho(v)/f_0(u)$ is nondecreasing in v, then from Lemma 1.2 we have

$$E_\rho \ln V \geq E_0 \ln V = \psi(n-1) + \frac{\ln(1+z)}{(1+z)^{n-1} - 1}. \tag{23}$$

Combining (22) and (23), we conclude that for every $\rho \in \mathbb{R}$,

$$E_\rho \ln V \geq \psi(n-1) + \frac{\ln(1+z)}{(1+z)^{n-1} - 1}. \tag{24}$$

Using (24), the inequality (21) holds provided

$$\phi(z) \geq \frac{\int_0^\infty \int_0^\infty \ln\tau g(v)H(zv;\rho)\,dv\,dG(\tau)}{\int_0^\infty \int_0^\infty g(v)H(zv;\rho)\,dv\,dG(\tau)} - \psi(n-1) - \frac{\ln(1+z)}{(1+z)^{n-1} - 1}. \tag{25}$$

Now we find an upper bound for

$$J = \frac{\int_0^\infty \int_0^\infty \ln \tau g(v) H(zv; \rho)\, dv\, dG(\tau)}{\int_0^\infty \int_0^\infty g(v) H(zv; \rho)\, dv\, dG(\tau)}.$$

Case I: $\mu < 0$

In this case, $\rho < 0$ and $I(zv > \rho) = 1$, so $H(zv; \rho) = e^{\rho} H(zv)$ and we have

$$J = \frac{\int_0^\infty e^{\rho} \ln \tau\, dG(\tau)}{\int_0^\infty e^{\rho}\, dG(\tau)} = \frac{\int_0^\infty e^{n\mu\tau/\sigma} \ln \tau\, dG(\tau)}{\int_0^\infty e^{n\mu\tau/\sigma}\, dG(\tau)} = E_\mu \ln \tau,$$

the random variable τ having density $p_\mu(\tau) \propto e^{n\mu\tau/\sigma} \nu(\tau)$. Since $\mu < 0$, $p_\rho(\tau)/p_0(\tau)$ is nonincreasing in τ and so by Lemma 1.2 we have

$$E_\mu \ln \tau \le E_0 \ln \tau = E \ln \tau.$$

Case II: $\mu > 0$

In this case, we have

$$J = \frac{\int_0^\infty (1 - e^{\rho}(1+z)^{1-n}) \ln \tau\, dG(\tau)}{\int_0^\infty (1 - e^{\rho}(1+z)^{1-n})\, dG(\tau)} = E_{r_\mu} \ln \tau,$$

where the density of τ is $r_\mu(\tau) \propto (1 - e^{n\mu\tau/\sigma}(1+z)^{1-n})\nu(\tau)$. But, for $\mu > 0$, $r_\mu(\tau)/r_0(\tau)$ is nonincreasing in τ and so by Lemma 1.2 we have

$$E_\mu \ln \tau < E_0 \ln \tau = E \ln \tau.$$

Obviously, for $\mu = 0$

$$E_\mu \ln \tau = E_0 \ln \tau = E \ln \tau. \tag{26}$$

Hence, for every $\mu \in R$ the inequality (25) holds provided

$$\phi(z) \ge E \ln \tau - \psi(n-1) - \frac{\ln(1+z)}{(1+z)^{n-1} - 1}. \tag{27}$$

This proves the theorem. □

In the following remarks, we give estimators for θ that belong to the class of estimators (20) and improve upon δ_{c_0}, using Theorem 3.1.

Remark 3.1. The estimator δ_{ϕ^*} is a Brewster and Zidek (1974)-type estimator for $\theta = \ln \sigma$ and the conditions of Theorem 3.1 can be easily verified. Actually, for $W > 0$, δ_{ϕ^*} coincides with the generalized Bayes estimator of θ with respect to the prior $\pi(\mu, \sigma) = 1/\sigma$, $\mu > 0$, $\sigma > 0$. The posterior density of (μ, σ) given $S = s$, $X = x > 0$ and τ, is

$$\pi(\mu, \sigma | s, x, \tau) \propto \frac{\tau^n}{\sigma^{n+1}} \exp\{-\tau(s+x)/\sigma + n\tau\mu/\sigma\}, \quad 0 < \mu < x/n, \sigma > 0.$$

The generalized Bayes estimator of θ under the quadratic loss (5) is given by

$$\delta_{\mathrm{GB}} = E \ln \sigma,$$

where the expectations are taken under $\pi(\mu, \sigma | s, x) = \int_0^\infty \pi(\mu, \sigma | s, x, \tau)\, dG(\tau)$. By direct computations, it is easy to verify that $\delta_{\mathrm{GB}} = \delta_{\phi^*}$.

Remark 3.2. Let $\phi_{ST2}(W) = \min\{\phi_2(W), \phi_0(W)\}$, where $\phi_2(w) = \ln(1+w) + E(\ln\tau) - \psi(n)$, then we can choose as an estimator for θ

$$\delta_{ST2}(X,S) = \begin{cases} \ln S + \phi_{ST2}(W), & \text{if } W > 0, \\ \ln S + c_0, & \text{otherwise} \end{cases}$$

$$= \begin{cases} \ln S + \ln(1+W) + E\ln\tau - \psi(n), & \text{if } 0 < W < e^{1/(n-1)} - 1, \\ \ln S + E\ln\tau - \psi(n-1), & \text{otherwise.} \end{cases}$$

We notice that $\phi_{ST2}(W)$ satisfies the conditions of Theorem 3.1, so δ_{ST2} improves upon δ_{c_0} for estimating θ under the squared error loss. It is mentioned that the Stein-type improvement $\delta_{ST1}(W)$ belongs to the class of estimators (20) but we cannot provide this improvement through Theorem 3.1. When $\tau = 1$ w.p.1, then δ_{ST2} coincides with δ_{ST1} in Theorem 2.1.

When $\tau = 1$, w.p.1, Theorem 3.1 reduces to the following result.

Theorem 3.2. *When $\tau = 1$, w.p.1, we define the estimators of θ as*

$$\delta^1_{c_0} = \ln S + \psi(n-1) \quad \textit{and} \quad \delta^1_\phi = \begin{cases} \ln S + \phi(W), & \textit{if } W > 0, \\ \delta^1_{c_0}, & \textit{otherwise}. \end{cases} \tag{28}$$

Then the estimator δ^1_ϕ improves upon the BAEE $\delta^1_{c_0}$ for estimating θ under the squared error loss function (5), if the function ϕ satisfies the following conditions

(i′) *$\phi(z)$ is nondecreasing with $\lim_{z\to\infty}\phi(z) = -\psi(n-1)$*
(ii′) *$\phi(z) \geq -\psi(n-1) - \frac{\ln(1+z)}{(1+z)^{n-1}-1}$.*

In the next theorem, we provide some robustness properties of the estimators mentioned in this section, we show that the improved estimators of θ can be derived from the improved estimators of θ in the case $\tau = 1$, w.p.1 (this is called the degenerate case), when the conditions of Theorem 3.2 are satisfied. Also, it is verified that the improvement of δ^1_ϕ over $\delta^1_{c_0}$, shown in Theorem 3.2, is robust in a specified neighborhood of the degenerate case.

Theorem 3.3. *Let δ^1_ϕ and $\delta^1_{c_0}$ be as in (28), where the function ϕ satisfies conditions* (i′) *and* (ii′) *of Theorem 3.2. Then we have the following.*

(a) *The estimator $\delta_\phi = E\ln\tau + \delta^1_\phi$ improves upon δ_{c_0} for estimating θ under the loss (5)*
(b) *The estimator δ^1_ϕ improves upon $\delta^1_{c_0}$ for estimating θ under the loss (5), uniformly for all $G(\cdot)$ such that $E\ln\tau \leq 0$.*

Proof. (a) It is an immediate consequence of Theorems 3.1 and 3.2.

(b) From Theorem 3.1, it is easily verified that δ^1_ϕ improves upon $\delta^1_{c_0}$, for all $G(\cdot)$, provided that condition (i′) of Theorem 3.2 and condition (ii) of Theorem 3.1 hold. The later is true because of condition (ii′) of Theorem 3.2 and the fact that $E\ln\tau \leq 0$. □

4 Multivariate Lomax distribution

In this section, we obtain the improved estimators of entropy of multivariate Lomax distribution. Consider the mixing density of τ in the model (2) to be Gamma distribution $\Gamma(b,1)$, then the resulting distribution becomes a multivariate Lomax distribution. In this case, $E(\ln\tau) = \psi(b)$, $E(\tau^n\ln\tau) = \frac{\Gamma(n+b)}{\Gamma(b)}\psi(n+b)$ and $E(\tau^n) = \frac{\Gamma(n+b)}{\Gamma(b)}$. So the BAEE is

$$\delta_{c_0} = \ln S + \psi(b) - \psi(n-1). \tag{29}$$

Now we have

$$\phi_1(w) = \ln(1+w) + \psi(n+b) - \psi(n) \quad \text{for } w > 0. \tag{30}$$

The following theorem gives the Stein type improved estimator for θ. It is mentioned that $c_0 = \psi(b) - \psi(n-1)$.

Theorem 4.1. *The estimator*

$$\delta^{\mathrm{ML}}_{ST1}(X,S) = \begin{cases} \ln S + \min\{\phi_1(W), c_0\}, & \text{if } W > 0, \\ \ln S + c_0, & \text{otherwise} \end{cases} \tag{31}$$

improve upon the best affine equivariant estimator δ_{c_0} under the squared error loss function (5).

The following theorem gives a class of improved estimator for estimating $\theta = \ln\sigma$ under the squared error loss function.

Theorem 4.2. *The estimator*

$$\delta^{\mathrm{ML}}_{\phi}(X,S) = \begin{cases} \ln S + \phi(W), & \text{if } W > 0, \\ \delta_{c_0}, & \text{otherwise} \end{cases}$$

improves upon the BAEE δ_{c_0} for estimating θ under the loss function (5), *if the function ϕ satisfies the following conditions*

(i) $\phi(w)$ *is nondecreasing with* $\lim_{w\to\infty}\phi(w) = \psi(b) - \psi(n-1)$
(ii) $\phi(w) \geq \psi(b) - \psi(n-1) - \frac{\ln(1+w)}{(1+w)^{n-1}-1}$.

From Theorem 4.2 we can derive, at least, two estimators which improve upon δ_{c_0} for estimating θ, namely the Brewster and Zidek (1974)-type estimator, $\delta^{\mathrm{ML}}_{\mathrm{BZ}}(X,S)$ and a Stein (1964)-type estimator which is different from the estimator derived in Theorem 4.1, $\delta^{\mathrm{ML}}_{ST2}(X,S)$, where

$$\delta^{\mathrm{ML}}_{\mathrm{BZ}}(X,S) = \begin{cases} \ln S + \psi(b) - \psi(n-1) - \dfrac{\ln(1+W)}{(1+W)^{n-1}-1}, & \text{if } W > 0, \\ \delta_{c_0}, & \text{otherwise} \end{cases}$$

and

$$\delta^{\mathrm{ML}}_{ST2}(X,S) = \begin{cases} \ln S + \min\{\phi_2(W), c_0\}, & \text{if } W > 0, \\ \ln S + c_0, & \text{otherwise,} \end{cases} \tag{32}$$

where $\phi_2(w) = \ln(1+w) + \psi(b) - \psi(n)$. For illustrative purposes, we have computed numerically, using Mathematica v.9, the relative quadratic risk improvement (RI $= \frac{R(\delta_{c_0};\mu,\sigma) - R(\delta;\mu,\sigma)}{R(\delta_{c_0};\mu,\sigma)} 100\%$) of these two estimators. Without loss of generality we have taken $\sigma = 1$. In Figures 1 and 2, we have computed the risk improvement of the estimator $\delta^{\mathrm{ML}}_{ST2}$ upon δ_{c_0} for estimating θ and for different values of n and b. It is noticed that the best improvement is achieved for values of μ near 0 (this is typical for Stein-type estimators). As n increases, this improvement is getting less in value. Also, as b increases the range of the values of μ, that this improvement exists, is getting more narrow. In Figures 3 and 4, we have computed the risk improvement of the estimator $\delta^{\mathrm{ML}}_{\mathrm{BZ}}$ upon δ_{c_0} for estimating θ and for different values of n and b. Of course, in this situation we have no improvement at $\mu = 0$ (this is typical for Brewster and Zidek-type estimators) and our remarks are similar with those made for Figures 1 and 2.

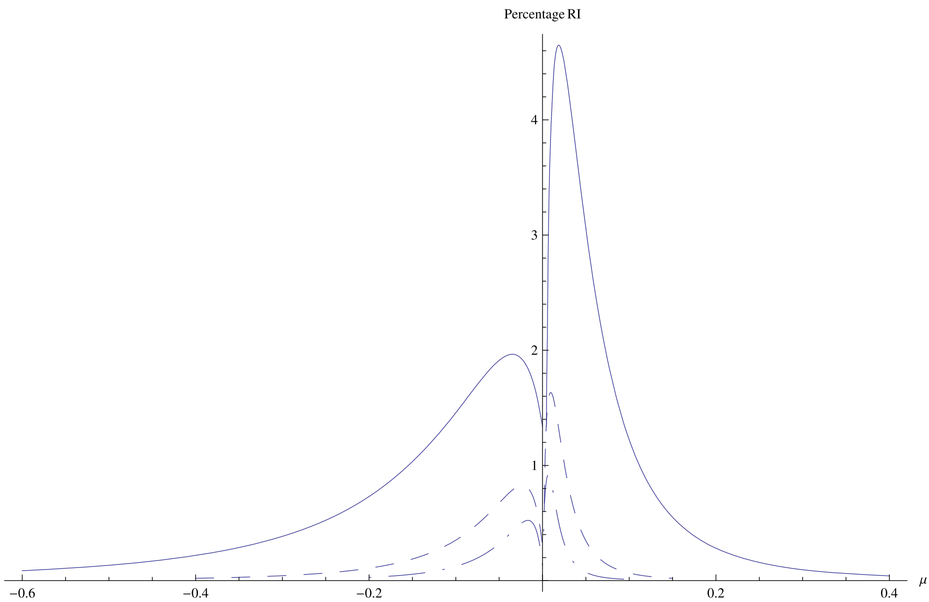

Figure 1 *Percentage risk improvement* (*RI*) *of* $\delta^{\mathrm{ML}}_{ST2}$ *over* δ_{c_0} *as a function of* μ ($\sigma = 1$) *for* $b = 4$, *when* $n = 5$ (—), $n = 10$ (- - -) *and* $n = 15$ (- . -).

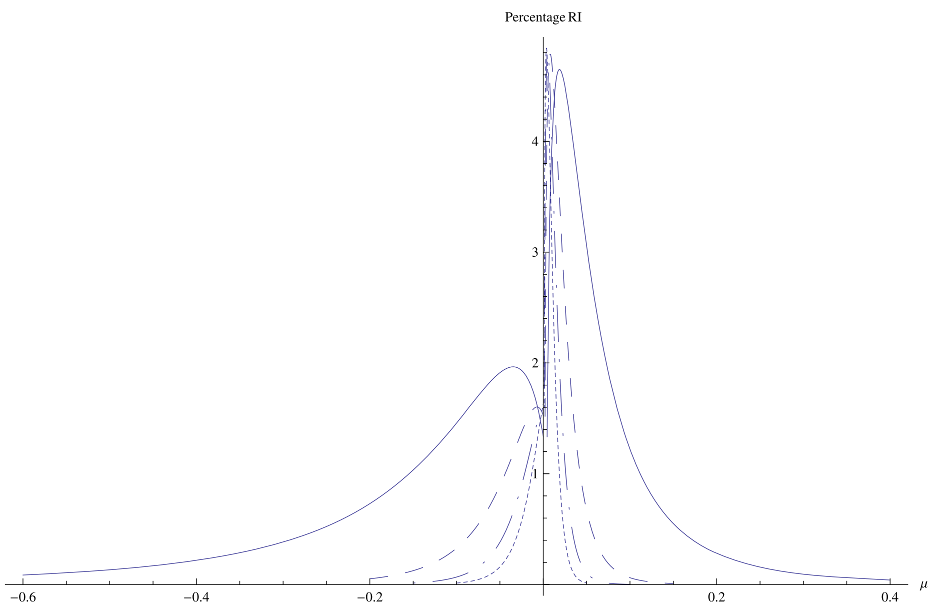

Figure 2 *Percentage risk improvement* (*RI*) *of* $\delta^{\mathrm{ML}}_{ST2}$ *over* δ_{c_0} *as a function of* μ ($\sigma = 1$) *for* $n = 5$, *when* $b = 4$ (—), $b = 8$ (- - -), $b = 12$ (- . -) *and* $b = 16$ (....).

5 Application

In this section, we provide an application of the Multivariate Lomax distribution. We consider a data set from Lawless (1982), the data are the number of million revolutions before failure for each of the 23 ball bearing in the life test and they are 17.88, 28.92, 33, 41.52, 42.12, 45.60, 48.40, 51.84, 51.96, 54.12, 55.56, 67.80, 68.64, 68.64, 68.88, 84.12, 93.12, 98.64, 105.12, 105.84, 127.92, 128.04, 173.40. By straightforward calculations and for different

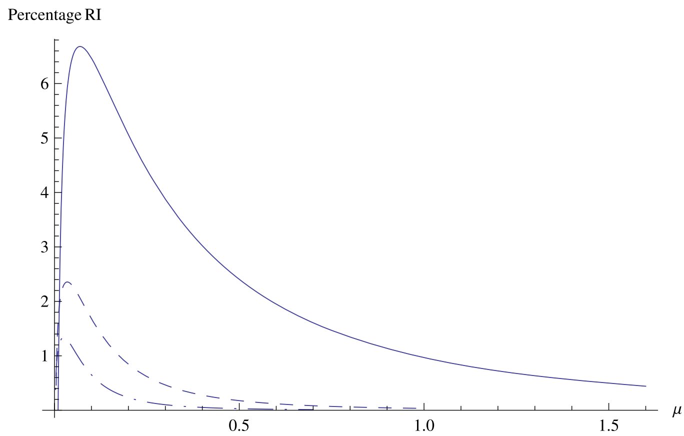

Figure 3 *Percentage risk improvement (RI) of $\delta_{\mathrm{BZ}}^{\mathrm{ML}}$ over δ_{c_0} as a function of μ ($\sigma = 1$) for $b = 4$, when $n = 5$ (—), $n = 10$ (- - -) and $n = 15$ (- . -).*

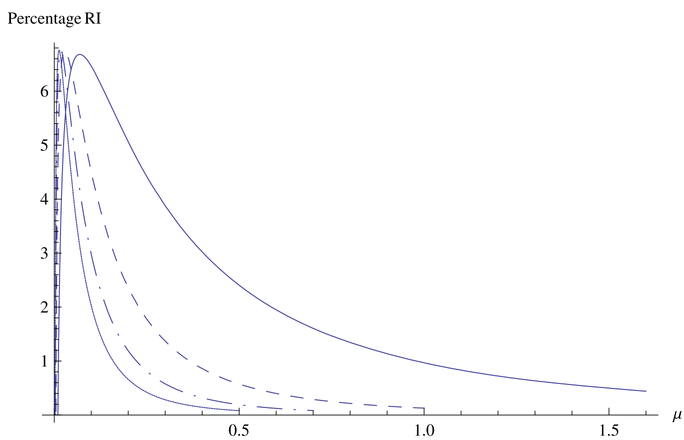

Figure 4 *Percentage risk improvement (RI) of $\delta_{\mathrm{BZ}}^{\mathrm{ML}}$ over δ_{c_0} as a function of μ ($\sigma = 1$) for $n = 5$, when $b = 4$ (—), $b = 8$ (- - -), $b = 12$ (- . -) and $b = 16$ (....).*

Table 1 *Values of the proposed estimators*

b	δ_{c_0}	$\delta_{ST1}^{\mathrm{ML}}$	$\delta_{ST2}^{\mathrm{ML}}$	$\delta_{\mathrm{BZ}}^{\mathrm{ML}}$
4	5.3187	5.3187	5.3187	5.3182
8	6.0782	6.0782	6.0782	6.0777
12	6.5052	6.5052	6.5052	6.5047
16	6.8036	6.8036	6.8036	6.8031

values of b, we compute the values of the estimators, proposed in the previous section, as in Table 1

As we can see, improvement upon δ_{c_0} can be achieved if we choose the Brewster and Zidek type estimator $\delta_{\mathrm{BZ}}^{\mathrm{ML}}$.

Remark 5.1. From our calculations, as b increases, we can achieve better improvement upon δ_{c_0} using the Stein-type estimator $\delta_{ST1}^{\mathrm{ML}}$, too.

6 Concluding remarks

In this paper, we have investigated the problem of estimating the entropy in a mixture model of exponential distributions with respect to squared error loss function. We have proved that the best affine equivariant estimator is inadmissible by deriving Stein (1964)-type improved estimators. Further, we considered a class of scale equivariant estimators and sufficient conditions are given under which this class of estimators improves upon the best affine equivatiant estimators. In this class of estimators, some robust properties are also derived. It has been seen that a generalized Bayes estimator coincides with the Brewster and Zidek (1974)-type estimator. Especially, we have considered estimation for the entropy in the multivariate Lomax distribution, plots of percentage risk reduction of the improved estimator are given for implement purpose.

References

Ahmed, N. A. and Gokhale, D. (1989). Entropy expressions and their estimators for multivariate distributions. *IEEE Transactions on Information Theory* **35**, 688–692. MR1022093 https://doi.org/10.1109/18.30996

Arnold, B. C. (2015). *Pareto Distribution*. New York: CRC Press. MR3618736

Bhattacharya, S. K. and Kumar, S. (1986). E-ig model in life testing. *Calcutta Statistical Association Bulletin* **35**, 85–90.

Bobotas, P. and Kourouklis, S. (2009). Strawderman-type estimators for a scale parameter with application to the exponential distribution. *Journal of Statistical Planning and Inference* **139**, 3001–3012. MR2535178 https://doi.org/10.1016/j.jspi.2009.02.004

Brewster, J. F. and Zidek, J. (1974). Improving on equivariant estimators. *The Annals of Statistics* **2**, 21–38. MR0381098

Broadbridge, P. and Guttmann, A. J. (2009). Concepts of entropy and their applications. *Entropy* **11**, 59–61. MR2534817 https://doi.org/10.3390/e11010059

Cover, T. M. and Thomas, J. A. (2012). *Elements of Information Theory*. New York: John Wiley and Sons. MR1122806 https://doi.org/10.1002/0471200611

Devi, B., Kumar, P. and Kour, K. (2017). Entropy of Lomax probability distribution and its order statistic. *International Journal of Statistics and Systems* **12**, 175–181.

Golan, A., Judge, G. G. and Miller, D. (1996). Maximum entropy econometrics. Technical report. MR1678448 https://doi.org/10.1007/978-94-011-5028-6_4

Johnson, N. L., Kotz, S. and Balakrishnan, N. (2002). *Continuous Multivariate Distributions: Models and Applications, Vol. 1*. New York: John Wiley and Sons. MR1788152 https://doi.org/10.1002/0471722065

Kamavaram, S. and Goseva-Popstojanova, K. (2002). Entropy as a measure of uncertainty in software reliability. In *13th Int'l Symp. Software Reliability Engineering*, 209–210.

Kang, S.-B., Cho, Y.-S., Han, J.-T. and Kim, J. (2012). An estimation of the entropy for a double exponential distribution based on multiply type-II censored samples. *Entropy* **14**, 161–173. MR2899420 https://doi.org/10.3390/e14020161

Kayal, S. and Kumar, S. (2011). Estimating the entropy of an exponential population under the linex loss function. *Journal of the Indian Statistical Association* **49**, 91–112. MR2931598

Kayal, S. and Kumar, S. (2013). Estimation of the Shannon's entropy of several shifted exponential populations. *Statistics & Probability Letters* **83**, 1127–1135. MR3041385 https://doi.org/10.1016/j.spl.2013.01.012

Kayal, S. and Kumar, S. (2017). Estimating Renyi entropy of several exponential distributions under an asymmetric loss function. *REVSTAT Statistical Journal* **15**, 501–522. MR3719213

Kayal, S., Kumar, S., Vellaisamy, P., et al (2015). Estimating the Rényi entropy of several exponential populations. *Brazilian Journal of Probability and Statistics* **29**, 94–111. MR3299109 https://doi.org/10.1214/13-BJPS230

Kubokawa, T. (1994). A unified approach to improving equivariant estimators. *The Annals of Statistics* **22**, 290–299. MR1272084 https://doi.org/10.1214/aos/1176325369

Lawless, J. (1982). *Statistical Models and Methods for Lifetime Data*. New York: Wiley. MR0640866

Lindley, D. V. and Singpurwalla, N. D. (1986). Multivariate distributions for the life lengths of components of a system sharing a common environment. *Journal of Applied Probability* **23**, 418–431. MR0839996 https://doi.org/10.2307/3214184

Lomax, K. S. (1954). Business failures: Another example of the analysis of failure data. *Journal of the American Statistical Association* **49**, 847–852.

Misra, N., Singh, H. and Demchuk, E. (2005). Estimation of the entropy of a multivariate normal distribution. *Journal of Multivariate Analysis* **92**, 324–342. MR2107880 https://doi.org/10.1016/j.jmva.2003.10.003

Nalewajski, R. F. (2002). Applications of the information theory to problems of molecular electronic structure and chemical reactivity. *International Journal of Molecular Sciences* **3**, 237–259.

Nayak, T. K. (1987). Multivariate Lomax distribution: Properties and usefulness in reliability theory. *Journal of Applied Probability* **24**, 170–177. MR0876178 https://doi.org/10.2307/3214068

Petropoulos, C. (2006). Estimation of a quantile in a mixture model of exponential distributions with unknown location and scale parameters. *Sankhya The Indian Journal of Statistics* **68**, 240–251. MR2303083

Petropoulos, C. (2010). A class of improved estimators for the scale parameter of a mixture model of exponential distribution with unknown location. *Communications in Statistics Theory and Methods* **39**, 3153–3162. MR2755430 https://doi.org/10.1080/03610920903205198

Petropoulos, C. and Kourouklis, S. (2005). Estimation of a scale parameter in mixture models with unknown location. *Journal of Statistical Planning and Inference* **128**, 191–218. MR2110184 https://doi.org/10.1016/j.jspi.2003.09.028

Robinson, D. W. (2008). Entropy and uncertainty. *Entropy* **10**, 493–506. MR2465846 https://doi.org/10.3390/e10040493

Seo, J.-I. and Kang, S.-B. (2014). Entropy estimation of generalized half-logistic distribution (ghld) based on type-II censored samples. *Entropy* **16**, 443–454. MR3360712 https://doi.org/10.1016/j.spl.2015.05.011

Seo, J.-I. and Kang, S.-B. (2015). Bayesian estimation of the entropy of the half-logistic distribution based on type-II censored samples. *International Journal of Applied Mathematics & Statistics* **53**, 58–66. MR3412655

Seo, J. I. and Kim, Y. (2017). Objective Bayesian entropy inference for two-parameter logistic distribution using upper record values. *Entropy* **19**, 208.

Stein, C. (1964). Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean. *Annals of the Institute of Statistical Mathematics* **16**, 155–160. MR0171344 https://doi.org/10.1007/BF02868569

C. Constantinos
Department of Mathematics
University of Patras
Rio, Patras 26504
Greece
E-mail: costas@math.upatras.gr

L. Kanta Patra
Department of Mathematics and Statistics
Indian Institute of Technology Kanpur
Kanpur, 208016
India
E-mail: patralakshmi@gmail.com

S. Kumar
Department of Mathematics
Indian Institute of Technology Kharagpur
Kharagpur 721302
India
E-mail: smsh@iitkgp.ac.in