Brazilian Journal of Probability and Statistics
2015, Vol. 29, No. 1, 94–111
DOI: 10.1214/13-BJPS230

Estimating the Renyi entropy of several exponential populations

Suchandan Kayal[a], Somesh Kumar[b] and P. Vellaisamy[c]

[a]*Birla Institute of Technology Mesra*
[b]*Indian Institute of Technology Kharagpur*
[c]*Indian Institute of Technology Bombay*

Abstract. Suppose independent random samples are drawn from k shifted exponential populations with a common location but unequal scale parameters. The problem of estimating the Renyi entropy is considered. The uniformly minimum variance unbiased estimator (UMVUE) is derived. Sufficient conditions for improvement over affine and scale equivariant estimators are obtained. As a consequence, improved estimators over the UMVUE and the maximum likelihood estimator (MLE) are obtained. Further, for the case $k = 1$, an estimator that dominates the best affine equivariant estimator is derived. Cases when the location parameter is constrained are also investigated in detail.

1 Introduction

Let $\Pi_1, \dots, \Pi_k$ $(k \geq 1)$ be k shifted exponential populations with a common location parameter μ and unknown scale parameters $\sigma_1, \dots, \sigma_k$, respectively. In this paper, we consider the problem of estimating the Renyi entropy of several exponential populations. Let X be a random variable with density $f(x|\theta)$. Then the Renyi entropy with parameter $\alpha \geq 0$ is defined as (Renyi (1961))

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

Note that as α tends to 1, $R_\alpha(\theta)$ tends to the Shannon entropy $H(\theta) = E_\theta(-\ln f(X|\theta))$ (Shannon (1948)). In recent years, the concept of entropy has found applications in diverse areas such as ecology, hydrology and water resources, social studies, economics, biology etc. In ecology, it is used to measure the diversity indices of different species whereas in social science, particularly in model building of urban and regional systems, the concept of entropy is widely used. The entropy is also used in earthquake forecasting (see Harte and Jones (2005)). For a detailed account, one may refer to Cover and Thomas (2006).

The problem of estimating the Shannon entropy of various continuous probability distributions has been addressed in recent years. The asymptotic distribution

Key words and phrases. Entropy estimation, equivariance, improved estimators, MLE, shifted exponential, UMVUE.

Received March 2013; accepted September 2013.

of the UMVUE of the entropy of a multivariate normal distribution is derived by Ahmed and Gokhale (1989). Misra et al. (2005) considered the problem of estimating the Shannon entropy of a multivariate normal distribution under the squared error loss. They established that the best affine equivariant estimator (BAEE) is admissible in a larger class of estimators. Further, they obtained improvements over the BAEE. Recently, Kayal and Kumar (2013) have considered estimation of entropy of several shifted exponential populations with different locations, but a common scale parameter.

In this paper, we consider the problem of estimating the Renyi entropy of k exponential populations with a common location but different scale parameters, which has not been addressed, so far, in the literature. Note that this problem is qualitatively different from the one considered in the paper by Kayal and Kumar (2013). In that paper, several exponential populations are assumed to have a common scale parameter but different location parameters. As a consequence the complete and sufficient statistics in the two models are distinct.

Rest of the paper is organized as follows. The UMVUE of the Renyi entropy of k exponential populations is derived in Section 2. In Section 3, inadmissibility results for scale and affine equivariant estimators are obtained. Consequently, estimators improving over the MLE and the UMVUE are derived. In Section 4, the case $k = 1$ is considered and the improvement over the BAEE is obtained. Special attention is paid to the cases when the location parameter is constrained. In Section 5, the risk performance of various estimators is compared numerically.

2 Derivation of the UMVUE

Assume that the ith population Π_i is shifted exponential with the probability density function (p.d.f.)

$$f_i(x) = \begin{cases} \dfrac{1}{\sigma_i}\exp\left\{-\left(\dfrac{x-\mu}{\sigma_i}\right)\right\}, & \text{if } x > \mu, \\ 0, & \text{otherwise,} \end{cases} \tag{2.1}$$

$i = 1, \dots, k$. Then the Renyi entropy of Π_i is $R_\alpha(\sigma_i) = \ln\sigma_i - \ln\alpha/(1-\alpha)$ and so the Renyi entropy based on $\Pi_1, \dots, \Pi_k$ is $R_\alpha(\sigma) = \sum_{i=1}^k \ln\sigma_i - k\ln\alpha/(1-\alpha)$, where $\sigma = (\sigma_1, \dots, \sigma_k)$. Note that the Shannon entropy in this case is $H(\sigma) = \sum_{i=1}^k \ln\sigma_i + k$. Since α is assumed to be known, the problems of estimating $R_\alpha(\sigma)$ and $H(\sigma)$ are the same as that of estimating $Q(\sigma) = \sum_{i=1}^k \ln\sigma_i$.

We first derive the UMVUE of $Q(\sigma)$. Suppose independent random observations $\{X_{i1}, \dots, X_{in}\}$ are available from the ith population Π_i. We consider the squared error loss defined by

$$L(\sigma, \delta) = \big(\delta - Q(\sigma)\big)^2, \tag{2.2}$$

where δ is an estimator of Q. Let $X_i = \min\{X_{i1}, \dots, X_{in}\}$, $Y_i = \sum_{j=1}^n X_{ij}$ and $Z_i = Y_i - nX_i$, $i = 1, \dots, k$. For ith population, (X_i, Z_i) is a complete and sufficient statistic for (μ, σ_i). It is noted that Z_i and X_i are independently distributed, where $2\sigma_i^{-1} Z_i$ follows a chi-square distribution with $2(n-1)$ degrees of freedom and X_i follows an exponential distribution with the location parameter μ and the scale parameter σ_i/n. Further, we define $X = \min\{X_1, \ldots, X_k\}$ and $T_i = Y_i - nX$. It can be shown that (X, T) is a complete and sufficient statistic for (μ, σ), where $T = (T_1, \ldots, T_k)$ (Ghosh and Razmpour (1984)). The MLE is $\delta_{\mathrm{ML}} = \sum_{i=1}^{k} \ln T_i - k \ln n$. Also, X and $T = (T_1, \ldots, T_k)$ are independently distributed with respective p.d.f.'s

$$f_X(x) = n\tau \exp\{-n\tau(x-\mu)\}, \qquad x > \mu \tag{2.3}$$

and

$$f_T(t) = (n-1) p \eta l \tau^{-1} \left(\prod_{i=1}^{k} t_i^{n-1} \right) \exp\left\{ -\left(\sum_{i=1}^{k} t_i \sigma_i^{-1} \right) \right\}, \qquad t_i > 0, \tag{2.4}$$

where $\eta = (\prod_{i=1}^{k} \sigma_i)^{-n}$, $\tau = \sum_{i=1}^{k} \sigma_i^{-1}$, $l = (\Gamma(n))^{-k}$, $p = \sum_{i=1}^{k} t_i^{-1}$ and $t = (t_1, \ldots, t_k)$. The joint density of X_i and Y_i is given by

$$g_i(x_i, y_i) = n \sigma_i^{-n} \big(\Gamma(n-1)\big)^{-1} (y_i - n x_i)^{n-2} \exp\left\{ -\left(\frac{y_i - n\mu}{\sigma_i} \right) \right\}, \tag{2.5}$$

where $x_i > \mu, y_i > n x_i$, $i = 1, \ldots, k$. Also, the joint density of X and $Y = (Y_1, \ldots, Y_k)$ can be derived as

$$\begin{aligned} g(x, y) = n(n-1)\eta l \sum_{i=1}^{k} (y_i - nx)^{-1} &\left(\prod_{i=1}^{k} (y_i - nx) \right)^{n-1} \\ \times \exp&\left\{ -\left(\sum_{i=1}^{k} (y_i - n\mu)\sigma_i^{-1} \right) \right\}, \end{aligned} \tag{2.6}$$

where $y = (y_1, \ldots, y_n)$, $y_i > nx, x > \mu$. Note that based on the ith sample, the UMVUE of $\ln \sigma_i$ is $d_i = \ln Z_i - \psi(n-1)$, where $\psi(x) = \Gamma'(x)/\Gamma(x)$ is the digamma function. The estimator d_i is no longer the UMVUE of $\ln \sigma_i$ when we consider all k populations together, though it is an unbiased estimator. To obtain the UMVUE, we apply Rao–Blackwellization to get

$$\begin{aligned} d_i &= E\big[\{\ln Z_i - \psi(n-1)\} | X = x, T = t\big] \\ &= E\big[\{\ln(Y_i - nX_i)\} | X = x, Y = y\big] - \psi(n-1), \end{aligned}$$

where $y_i = t_i + nx$. For $i = 1$, we get

$$d_1 = E\big[\{\ln(Y_1 - nX_1)\} | X = x, Y = y\big] - \psi(n-1). \tag{2.7}$$

The first term on the right-hand side of the equation (2.7) can be written as

$$E\big[\{\ln(Y_1 - nX_1)\} | X = x, Y = y\big] = \sum_{r=1}^{k} S_r \qquad \text{(say)}, \tag{2.8}$$

where

$$g(x,y)S_1 = \ln(y_1 - nx)g_1(x,y_1) \times \int_x^{y_2/n} \cdots \int_x^{y_k/n} \left(\prod_{i=2}^{k} g_i(x_i,y_i) \right) dx_k \cdots dx_2 \tag{2.9}$$

and

$$g(x,y)S_r = \int_x^{y_1/n} \ln(y_1 - nz) g_1(z,y_1) g_r(x,y_r) \int_x^{y_2/n} \cdots \int_x^{y_{r-1}/n} \int_x^{y_{r+1}/n} \cdots \times \int_x^{y_k/n} \prod_{i=2, i\neq r}^{k} g_i(x_i,y_i)\, dx_k \cdots dx_{r+1}\, dx_{r-1} \cdots dx_2\, dz, \tag{2.10}$$

where $r = 2, \ldots, k$. After some simplification, we obtain

$$S_1 = \frac{\ln(y_1 - nx)}{(y_1 - nx)\sum_{i=1}^{k}(y_i - nx)^{-1}} \tag{2.11}$$

and

$$S_r = \frac{\ln(y_1 - nx) - (n-1)^{-1}}{(y_r - nx)\sum_{i=1}^{k}(y_i - nx)^{-1}}, \qquad r = 2, \ldots, k. \tag{2.12}$$

Putting the values of $S_r, r = 1, \ldots, k$ in (2.8), we get the UMVUE of $\ln \sigma_1$ as

$$d_1 = \ln T_1 - (n-1)^{-1}\left[1 - (JT_1)^{-1}\right] - \psi(n-1),$$

where $J = \sum_{i=1}^{k} T_i^{-1}$. Similarly, the UMVUE of $\ln \sigma_i$ is obtained as

$$d_i = \ln T_i - (n-1)^{-1}\left[1 - (JT_i)^{-1}\right] - \psi(n-1), \qquad i = 2, \ldots, k.$$

Consequently, the UMVUE of $Q(\sigma)$ is

$$\delta_{\mathrm{MV}} = \sum_{i=1}^{k} \ln T_i - \frac{k-1}{n-1} - k\psi(n-1). \tag{2.13}$$

3 Inadmissibility results for equivariant estimators

In this section, we introduce the invariance considerations for the entropy estimation problem. Indeed, invariance plays an important role in information theory (Khinchin (1957)). Some general inadmissibility results are obtained for both affine as well as scale equivariant estimators. It is noted that the results presented in this section are true only for $k \geq 2$, because the complete sufficient statistic is different from the case $k = 1$. First, we consider the affine equivariant estimators.

3.1 Affine equivariant estimators

The estimation problem under study is invariant under $G_{a,b}$, a group of affine transformations, where $G_{a,b} = \{g_{a,b} : g_{a,b}(x) = ax + b, a > 0, b \in \mathbb{R}\}$. Under this transformation $X_{ij} \to aX_{ij} + b$, $X_i \to aX_i + b$, $Y_i \to aY_i + b$, $Z_i \to aZ_i$, $X \to aX + b$, $T_i \to aT_i$ and $Q(\sigma) \to Q(\sigma) + k \ln a$. Also, the loss function (2.2) is invariant under $G_{a,b}$ if $\delta \to \delta + k \ln a$. Therefore, the form of an affine equivariant estimator is

$$\begin{aligned} \delta_\phi(X, T) &= k \ln T_1 + \phi(W_1, \dots, W_{k-1}) \\ &= k \ln T_1 + \phi(W), \end{aligned} \tag{3.1}$$

where $W = (W_1, \dots, W_{k-1})$ and $W_i = (T_{i+1}/T_1)$, $i = 1, \dots, k-1$. The following result provides a general inadmissibility result for affine equivariant estimators.

Theorem 3.1. *Let δ_ϕ be an affine equivariant estimator of the form (3.1), $w = (w_1, \dots, w_{k-1})$ and $\phi_0(w) = \ln[k^k(\prod_{i=1}^{k-1} w_i)] - k\psi(kn-1)$. Further, define the estimator δ^*_ϕ by*

$$\delta^*_\phi = \begin{cases} \delta_\phi, & \text{if } \phi(w) \geq \phi_0(w), \\ \delta_{\phi_0}, & \text{if } \phi(w) < \phi_0(w). \end{cases}$$

Under the squared error loss, δ^*_ϕ improves δ_ϕ if $P_{(\mu,\sigma)}(\phi(W) < \phi_0(W)) > 0$, for some (μ, σ).

Proof. The risk function of δ_ϕ can be written as

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

where

$$R_1(\mu, \sigma, W, \delta_\phi) = E\big[\big(k \ln T_1 + \phi(W) - Q(\sigma)\big)^2 | W = w\big],$$

the conditional risk of δ_ϕ given $W = w$. Note that the conditional risk R_1 is a convex function of ϕ with minimum attained at

$$\widehat{\phi}(w, \sigma) = -kE(\ln T_1 | W = w) + Q(\sigma). \tag{3.2}$$

To evaluate $\widehat{\phi}(w, \sigma)$ in (3.2) we need to find out the conditional expectation of $T_1 | W = w$. Applying the transformations $W_i = (T_{i+1}/T_1)$, $i = 1, \dots, k-1$ and $T_1 = T_1$ to the joint p.d.f. of $T = (T_1, \dots, T_k)$ in (2.4), we get the joint p.d.f. of (T_1, W) as

$$f_{T_1,W}(t_1, w) = (n-1) l \eta \tau^{-1} \left(\prod_{i=1}^{k-1} w_i\right)^{n-1} \left(1 + \sum_{i=1}^{k-1} w_i^{-1}\right) e^{-st_1} t_1^{kn-2},$$

where $t_1 > 0$, $w_i > 0$ and $s = \sigma_1^{-1} + \sum_{i=1}^{k-1} w_i \sigma_{i+1}^{-1}$. Integrating $f_{T_1,W}(t_1, w)$ with respect to t_1, we get the marginal density of W as

$$f_W(w) = (n-1)\Gamma(kn-1) l \eta \tau^{-1} \left(\prod_{i=1}^{k-1} w_i \right)^{n-1} \left(1 + \sum_{i=1}^{k-1} w_i^{-1} \right) s^{-(kn-1)},$$

$w_i > 0$. Therefore, the conditional density of $T_1 | W = w$ is

$$f_{T_1|W}(t_1|w) = \big(\Gamma(kn-1)\big)^{-1} s^{kn-1} e^{-st_1} t_1^{kn-2}, \qquad t_1 > 0, w_i > 0, \tag{3.3}$$

which leads to

$$E(\ln T_1 | W = w) = \psi(kn-1) - \ln s. \tag{3.4}$$

Therefore, from (3.2) we get

$$\widehat{\phi}(w, \sigma) = \ln\big(s^k \eta^{-n}\big) - k\psi(kn-1).$$

To apply the orbit-by-orbit improvement technique of Brewster and Zidek (1974), it is required to find out the infimum and supremum of $\widehat{\phi}(w, \sigma)$ over σ for fixed values of w. To obtain the infimum of $(s^k \eta^{-n})$, we apply geometric mean–harmonic mean (GM-HM) inequality to the variables $\sigma_1, (\sigma_2/w_1), \ldots, (\sigma_k/w_{k-1})$. The equality sign holds if $\sigma_1 = (\sigma_2/w_1) = \cdots = (\sigma_k/w_{k-1})$. Thus, we get

$$\inf_\sigma \widehat{\phi}(w, \sigma) = \phi_0(w) \quad \text{and} \quad \sup_\sigma \widehat{\phi}(w, \sigma) = +\infty. \tag{3.5}$$

Infimum of $\widehat{\phi}(w, \sigma)$ is attained if $w_1 = \cdots = w_{k-1}$. An application of the Brewster and Zidek (1974) technique completes the proof of the theorem. □

A consequence of Theorem 3.1 is the following corollary.

Corollary 3.1. *The MLE δ_{ML} is inadmissible.*

Proof. The MLE δ_{ML} can be written as

$$\delta_{\mathrm{ML}} = k \ln T_1 + \phi(w),$$

where $\phi(w) = \ln(\frac{T_2}{T_1} \cdots \frac{T_k}{T_1}) - k \ln n$. We see that the MLE is an affine equivariant estimator of the form (3.1). The MLE is inadmissible since $\phi(w) < \phi_0(w)$, for $n \geq 2$ and $k \geq 2$. □

Remark 3.1. Note that the UMVUE can be written as

$$\delta_{\mathrm{MV}} = k \ln T_1 + \phi(w),$$

where $\phi(w) = \ln(\prod_{i=1}^{k-1} w_i) - k\psi(n-1) - \frac{k-1}{n-1}$. The UMVUE is inadmissible if $\phi(w) < \phi_0(w)$, which is not possible. Thus, Theorem 3.1 does not lead to an improvement over the UMVUE. However, we show in the next section that the UMVUE can be improved in a larger class of estimators when μ is known a priori to be negative.

Remark 3.2. Note from (3.2), the BAEE does not exist, for $k > 1$. However, when $k = 1$, $Q(\sigma) = \ln(\sigma_1)$, and the BAEE exists. It is given by

$$\delta_{\text{BA}} = \ln T_1 - \psi(n-1), \tag{3.6}$$

which is also a generalized Bayes estimator with respect to the improper prior $\pi(\sigma_1) = \sigma_1^{-1}, \sigma_1 > 0$. It is noted that the BAEE is also the UMVUE when $k = 1$.

Remark 3.3. Using the arguments in the proof of the Theorem 2.4 of Misra et al. (2005), it can be shown that the BAEE is admissible in the class of estimators depending on T_1 alone, when $k = 1$.

3.2 Scale equivariant estimators

In this section, we will look for the improvement over the UMVUE by looking at the larger class of scale-equivariant estimators with respect to the scale group of transformations $G_a = \{g_a : g_a(x) = ax, a > 0\}$. The form of the scale equivariant estimator is obtained as

$$\begin{aligned}\delta_\xi(X, T_1, T_2, \ldots, T_k) &= k \ln T_1 + \xi\left(\frac{X}{T_1}, \frac{T_2}{T_1}, \frac{T_3}{T_1}, \ldots, \frac{T_k}{T_1}\right) \\ &= k \ln T_1 + \xi(V) \qquad \text{(say)},\end{aligned} \tag{3.7}$$

where $V = (V_1, V_2, \ldots, V_k)$, $V_1 = X/T_1$, and $V_i = T_i/T_1, i = 2, 3, \ldots, k$. The risk function of δ_ξ as given in (3.7) can be written as

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

where

$$R_1(\mu, \sigma, V, \delta_\xi) = E\big[\big(k \ln T_1 + \xi(V) - Q(\sigma)\big)^2 | V = v\big] \tag{3.8}$$

which is minimized with respect to ξ at

$$\widehat{\xi}(\mu, \sigma, v) = -E(k \ln T_1 | V = v) + Q(\sigma). \tag{3.9}$$

Applying the transformations $V_1 = X/T_1, V_2 = T_2/T_1, V_3 = T_3/T_1, \ldots, V_k = T_k/T_1, T_1 = T_1$, we get the joint p.d.f. of T_1 and V as

$$\begin{aligned}f(t_1, v) = C&\left(1 + \sum_{i=2}^{k} v_i^{-1}\right) e^{n\mu\tau} \exp\left\{-\left(n v_1 \tau + \left(\sigma_1^{-1} + \sum_{i=2}^{k} v_i \sigma_i^{-1}\right)\right) t_1\right\} \\ &\times t_1^{nk-1},\end{aligned} \tag{3.10}$$

where $C = n(n-1)\eta l (v_2 v_3 \ldots v_k)^{n-1}, t_1 > 0, t_1 v_1 > \mu, v_2 > 0, v_3 > 0, v_k > 0$. To derive the conditional density of T_1 given V, the marginal density of V is required, which can be obtained by integrating $f(t_1, v)$ given in (3.10) with respect to t_1. Note that the range of t_1 is different for different values of μ and v_1. The cases are described below:

Case (i). When $\mu > 0$ and $v_1 > 0$, we have $\mu/w_1 < t_1 < \infty$. Therefore, the marginal density of V is

$$f(v) = B\int_{\mu/v_1}^{\infty} e^{-At_1} t_1^{nk-1}\, dt_1, \tag{3.11}$$

where $B = C(1 + \sum_{i=2}^{k} v_i^{-1})e^{n\mu\tau}$ and $A = nv_1\tau + (\sigma_1^{-1} + \sum_{i=2}^{k} v_i\sigma_i^{-1})$.

Also, the conditional density of T_1 given V is

$$f(t_1|v) = \frac{e^{-At_1} t_1^{nk-1}}{\int_{\mu/v_1}^{\infty} e^{-At_1} t_1^{nk-1}\, dt_1}, \qquad t_1 > \frac{\mu}{v_1}. \tag{3.12}$$

Thus, we have

$$E(\ln T_1|V = v) = \frac{\int_{A\mu/v_1}^{\infty} \ln z e^{-z} z^{nk-1}\, dz}{\int_{A\mu/v_1}^{\infty} e^{-z} z^{nk-1}\, dz} - \ln A. \tag{3.13}$$

Substituting the above expression in (3.9), we get

$$\widehat{\xi}(\mu, \sigma, v) = \ln\big(A^k \eta^{-n}\big) - k\frac{\int_{A\mu/v_1}^{\infty} \ln z e^{-z} z^{nk-1}\, dz}{\int_{A\mu/v_1}^{\infty} e^{-z} z^{nk-1}\, dz}. \tag{3.14}$$

Using the MLR property, it can be shown that the supremum and infimum of $\widehat{\xi}(\mu, \sigma, v)$ are $+\infty$ and $-\infty$, respectively.

Case (ii). When $\mu < 0$ and $v_1 > 0$, t varies from 0 and ∞. Therefore, the conditional distribution of T_1 given V can be obtained as

$$f(t_1|v) = \frac{A^{nk}}{\Gamma(nk)} e^{-At_1} t_1^{nk-1}, \qquad t_1 > 0. \tag{3.15}$$

Hence,

$$E(\ln T_1|V = v) = \psi(nk) - \ln A.$$

Thus from (3.9), we have

$$\widehat{\xi}(\mu, \sigma, v) = \ln\big(A^k \eta^{-n}\big) - k\psi(nk). \tag{3.16}$$

Note that supremum of $\widehat{\xi}(\mu, \sigma, v)$ is $+\infty$. The infimum of $\widehat{\xi}(\mu, \sigma, v)$ can be obtained by applying the GM-HM inequality on the variables $(\sigma_1/nv_1 + 1)$, $(\sigma_2/nv_1 + v_2), \ldots, (\sigma_k/nv_1 + v_k)$ and is given by

$$\inf_{\mu,\sigma} \widehat{\xi}(\mu, \sigma, v) = \ln\left(k^k (nv_1 + 1)\prod_{i=2}^{k}(nv_1 + v_i)\right) - k\psi(nk).$$

Case (iii). When $\mu < 0$ and $v_1 < 0$, we have $0 < t_1 < \mu/v_1$. Under this assumption, the conditional distribution of T_1 given V can be obtained as

$$f(t_1|v) = \frac{e^{-At_1} t_1^{nk-1}}{\int_0^{\mu/v_1} e^{-At_1} t_1^{nk-1}\, dt_1}, \qquad 0 < t_1 < \frac{\mu}{v_1}. \tag{3.17}$$

Note that the value of A can be positive or negative. For $A > 0$, we get

$$\widehat{\xi}(\mu, \sigma, v) = \ln\bigl(A^k \eta^{-n}\bigr) - k \frac{\int_0^{A\mu/v_1} \ln z e^{-z} z^{nk-1}\, dz}{\int_0^{A\mu/v_1} e^{-z} z^{nk-1}\, dz}, \tag{3.18}$$

and for $A < 0$, we have

$$\widehat{\xi}(\mu, \sigma, v) = \ln\bigl((A')^k \eta^{-n}\bigr) - k \frac{\int_0^{A'\mu/v_1} \ln z e^{-z} z^{nk-1}\, dz}{\int_0^{A'\mu/v_1} e^{-z} z^{nk-1}\, dz}, \tag{3.19}$$

where $A' = -A$. It can be shown that for both the cases when $A > 0$ or $A < 0$, the supremum and infimum of $\widehat{\xi}(\mu, \sigma, v)$ are $+\infty$ and $-\infty$, respectively.

For the function $\xi(v)$ defined in (3.7), let

$$\xi_0(v) = \begin{cases} \ln(v^*) - k\psi(nk), & \text{if } v_1 > 0 \text{ and } v^* > \exp\{\xi(v) + k\psi(nk)\}, \\ \xi(v), & \text{otherwise,} \end{cases} \tag{3.20}$$

where $v^* = k^k (nv_1 + 1) \prod_{i=2}^k (nv_1 + v_i)$. Using Brewster and Zidek (1974) technique, we get the following result for $\mu < 0$.

Theorem 3.2. *Let δ_ξ be a scale equivariant estimator of the form (3.7) and $\xi_0(v)$ be as defined in (3.20). If there exists a (μ, σ) such that $P_{(\mu,\sigma)}(\xi_0(V) \neq \xi(V)) > 0$, then the estimator δ_{ξ_0} dominates δ_ξ, for the squared error loss when $\mu < 0$.*

As a consequence of the Theorem 3.2, we get the following corollary.

Corollary 3.2. *When $\mu < 0$, the* UMVUE *and the MLE are inadmissible and are respectively dominated by the estimators given by*

$$\delta_{\mathrm{IMV}} = \begin{cases} \ln\bigl(T_1^k V^*\bigr) - k\psi(nk), \\ \qquad \text{if } V_1 > 0 \text{ and} \\ \qquad V^* > \exp\left\{ \sum_{i=2}^k \ln V_i + k\bigl(\psi(nk) - \psi(n-1)\bigr) - \dfrac{k-1}{n-1} \right\}, \\ \delta_{\mathrm{MV}}, \qquad \text{otherwise} \end{cases}$$

and

$$\delta_{\mathrm{IML}} = \begin{cases} \ln\bigl(T_1^k V^*\bigr) - k\psi(nk), \\ \qquad \text{if } V_1 > 0 \text{ and } V^* > \exp\left\{ \sum_{i=2}^k \ln V_i + k\bigl(\psi(nk) - \ln n\bigr) \right\}, \\ \delta_{\mathrm{ML}}, \qquad \text{otherwise,} \end{cases}$$

where $V^* = k^k (nV_1 + 1) \prod_{i=2}^k (nV_1 + V_i)$.

Remark 3.4. Since the complete sufficient statistics are different for the cases when $k \geq 2$ and $k = 1$, all the results for k population case do not follow directly for the one population case, as seen in the next section. We propose some estimators which improve the BAEE for the case $k = 1$.

4 Improving upon the BAEE when $k = 1$

When $k = 1$, the Renyi entropy of an exponential population is $R_\alpha(\sigma_1) = \ln \sigma_1 - \frac{\ln \alpha}{1-\alpha}$, so that the problem reduces to the estimation of $\ln \sigma_1$ with respect to the squared error loss. The form of a scale equivariant estimator is given by

$$\delta_\xi(X_1, T_1) = \ln T_1 + \xi(U), \tag{4.1}$$

where $U = X_1/T_1$. A general inadmissibility result is derived for the scale equivariant estimators of the form (4.1) in the following theorem. For the function $\xi(v)$ in (4.1), we define

$$\xi_0(u) = \begin{cases} \ln(nu+1) - \psi(n), & \text{if } 0 < u < d^* \text{ or, } m^* < u < 0, \\ \xi(u), & \text{otherwise,} \end{cases} \tag{4.2}$$

where $d^* = (\exp\{(\phi(u) + \psi(n)\} - 1)/n$ and $m^* = \max\{-1/n, d^*\}$.

Theorem 4.1. *Let δ_ξ be a scale equivariant estimator of the form (4.1) and $\xi_0(u)$ be as defined in (4.2). If there exist (μ, σ_1) such that $P_{(\mu,\sigma_1)}(\xi_0(U) \neq \xi(U)) > 0$, then with respect to the squared error loss, the estimator δ_ξ is inadmissible and is improved by δ_{ξ_0}.*

Proof. We write the risk function of δ_ξ given in (4.1) as

$$R(\mu, \sigma_1, \delta_\xi) = E^U R_1(\mu, \sigma_1, U, \delta_\xi),$$

where $R_1(\mu, \sigma_1, u, \delta_\xi)$ denotes the conditional risk of δ_ξ given $U = u$ given by

$$\begin{aligned} R_1(\mu, \sigma_1, u, \delta_\xi) &= E\big[(\delta_\xi - \ln \sigma_1)^2 | U = u\big] \\ &= E\big[\big(\ln T_1 + \xi(U) - \ln \sigma_1\big)^2 | U = u\big]. \end{aligned}$$

It can be noted that the conditional risk R_1 is a function of μ/σ_1. Therefore, without loss of generality, we take $\sigma_1 = 1$. We also see that the conditional risk R_1 is a convex function of ξ and the choice of ξ minimizing it is

$$\widehat{\xi}(\mu, u) = -E(\ln T_1 | U = u). \tag{4.3}$$

Also, from (3.10), the joint density of T_1 and U is

$$f_{T_1,U}(t_1, u) = \frac{n}{\Gamma(n-1)} e^{-(nt_1u - n\mu + t_1)} {t_1}^{n-1}, \qquad t_1 u > \mu, t_1 > 0.$$

In order to determine the marginal density of U, the density $f_{T_1,U}(t_1, u)$ needs to be integrated with respect to t_1. For different values of μ and u, we get different ranges of t_1 and these are described below.

(i) $\mu > 0$ and $u > 0$. In this case, t_1 varies from μ/u to ∞.
(ii) $\mu < 0$ and $u > 0$. In this case, t_1 varies from 0 to ∞.
(iii) $\mu < 0$ and $u < 0$. Here, t_1 varies from 0 to μ/u.

Case (i). When $\mu > 0$ and $u > 0$, putting $k = 1$ and making changes accordingly, we get from (3.14)

$$\begin{aligned}\widehat{\xi}(\mu, u) &= \ln(nu+1) - \frac{\int_{\eta_1}^{\infty} \ln p e^{-p} p^{n-1}\, dp}{\int_{\eta_1}^{\infty} e^{-p} p^{n-1}\, dp} \\ &= \ln(nu+1) - h_1(\eta_1) \qquad \text{(say)},\end{aligned} \tag{4.4}$$

where $\eta_1 = \mu(nu+1)/u$. Using the MLR property and then applying Lemma 3.4.2 of Lehmann and Romano (2005, p. 70), it can be shown that the function $h_1(\eta_1)$ given in (4.4) is nondecreasing for $\eta_1 > 0$. Thus, we get

$$\sup_{\eta_1} h_1(\eta_1) = +\infty \quad \text{and} \quad \inf_{\eta_1} h_1(\eta_1) = \psi(n)$$

and consequently

$$\sup_{\mu} \widehat{\xi}(\mu, u) = \ln(nu+1) - \psi(n) \quad \text{and} \quad \inf_{\mu} \widehat{\xi}(\mu, u) = -\infty. \tag{4.5}$$

Case (ii). When $\mu < 0$ and $u > 0$, we get from (3.16) for $k = 1$

$$\widehat{\xi}(\mu, u) = \ln(nu+1) - \psi(n). \tag{4.6}$$

Case (iii). When $\mu < 0$ and $u < 0$, we get from (3.18)

$$\begin{aligned}\widehat{\xi}(\mu, u) &= \ln(nu+1) - \frac{\int_{0}^{\eta_1} \ln p e^{-p} p^{n-1}\, dp}{\int_{0}^{\eta_1} e^{-p} p^{n-1}\, dp} \\ &= \ln(nu+1) - h_2(\eta_1) \qquad \text{(say)},\end{aligned} \tag{4.7}$$

for $nu + 1 > 0$. Using the MLR property as in Case (i), we can show that $h_2(\eta_1)$ is a nondecreasing function in $\eta_1, 0 < \eta_1 < \infty$. Hence,

$$\sup_{\eta_1} h_2(\eta_1) = \psi(n) \quad \text{and} \quad \inf_{\eta_1} h_2(\eta_1) = -\infty.$$

Thus, we get from (4.7)

$$\sup_{\mu} \widehat{\xi}(\mu, u) = +\infty \quad \text{and} \quad \inf_{\mu} \widehat{\xi}(\mu, u) = \ln(nu+1) - \psi(n). \tag{4.8}$$

Further, when $(nv + 1) < 0$, we get from (3.19)

$$\begin{aligned}\widehat{\xi}(\mu, u) &= \ln q - \frac{\int_{0}^{\eta_2} \ln p e^{p} p^{n-1}\, dp}{\int_{0}^{\eta_2} e^{p} p^{n-1}\, dp} \\ &= \ln q - h_3(\eta_2) \qquad \text{(say)},\end{aligned} \tag{4.9}$$

where $\eta_2 = (\mu q/u)$ and $q = -(nu+1)$. It can be shown that the function $h_3(\eta_2)$ is nondecreasing for $0 < \eta_2$. Therefore,

$$\sup_{\eta_2} h_3(\eta_2) = +\infty \quad \text{and} \quad \inf_{\eta_2} h_3(\eta_2) = -\infty,$$

which leads to

$$\sup_{\mu} \widehat{\xi}(\mu, u) = +\infty \quad \text{and} \quad \inf_{\mu} \widehat{\xi}(\mu, u) = -\infty. \tag{4.10}$$

An application of the Brewster–Zidek technique on the function $R_1(\mu, \sigma_1, u, \delta_\xi)$ completes the proof of the theorem. □

The following corollary is an immediate consequence of Theorem 4.1.

Corollary 4.1. *The* BAEE δ_{BA} *of* $\ln \sigma_1$ *is improved by*

$$\delta_{\mathrm{IB}} = \begin{cases} \ln T_1 + \ln(nU+1) - \psi(n), & \text{if } 0 < U < d^{**} \text{ or, } m^{**} < U < 0, \\ \ln T_1 - \psi(n-1), & \text{otherwise,} \end{cases}$$

where $d^{**} = (e^{(\psi(n)-\psi(n-1))} - 1)/n$ *and* $m^{**} = \max\{-1/n, d^{**}\}$.

In some applications to reliability and life testing problems, the minimum guarantee time may be bounded below or above. In such a case, we may consider the parameter μ to be nonnegative or negative.

4.1 The Case $k = 1$ and $\mu \geq 0$

We study in this section the problem of estimating $\ln \sigma_1$, under squared error loss, when $\mu \geq 0$. The MLE of $\ln \sigma_1$ remains unchanged as for the case of unrestricted parameter space and is $\delta_M = \ln T_1 - \ln n$. The following theorem establishes that the BAEE δ_{BA} of $\ln \sigma_1$ is inadmissible.

Theorem 4.2. *Let* d^{**} *be defined in the Corollary 4.1 and* $\mu \geq 0$. *Then the estimator defined by*

$$\delta_{\mathrm{IB}}^{+} = \begin{cases} \ln T_1 + \ln(nU+1) - \psi(n), & \text{if } U < d^{**}, \\ \ln T_1 - \psi(n-1), & \text{otherwise,} \end{cases}$$

dominates the BAEE δ_{BA} *of* $\ln \sigma_1$.

The theorem follows using (4.5) and the Brewster and Zidek technique.

We also observe that the generalized Bayes estimator of $\ln \sigma_1$ with respect to a noninformative prior $\pi(\mu, \sigma_1) = 1/\sigma_1$, $\mu > 0$, $\sigma_1 > 0$ is given by

$$\delta_{\mathrm{GB}}^{+} = \frac{(T_1 + nX_1)^{n-1} \ln T_1 - \ln(T_1 + nX_1) {T_1}^{n-1}}{(T_1 + nX_1)^{n-1} - {T_1}^{n-1}} - \psi(n-1). \tag{4.11}$$

4.2 The Case $k = 1$ and $\mu < 0$

Assume now μ is bounded above and hence, without loss of generality, we may take it to be negative. This type of situation may arise in cases where the minimum guarantee time may be known to be less than a pre-specified constant. The restricted maximum likelihood estimator (RMLE) of $\ln \sigma_1$ is given by

$$\delta_{\mathrm{RM}} = \begin{cases} \ln T_1 - \ln n, & \text{if } X_1 < 0, \\ \ln(T_1 + nX_1) - \ln n, & \text{if } X_1 \geq 0. \end{cases} \tag{4.12}$$

The following theorem proves that the BAEE of $\ln \sigma_1$ is inadmissible when $\mu < 0$.

Theorem 4.3. *Let $\mu < 0$ and m^{**} be defined as in Corollary 4.1. Then the estimator defined by*

$$\delta_{\mathrm{IB}}^{-} = \begin{cases} \ln T_1 + \ln(nU + 1) - \psi(n), & \textit{if } U > 0 \textit{ or, } m^{**} < U < 0, \\ \ln T_1 - \psi(n-1), & \textit{otherwise,} \end{cases}$$

dominates the BAEE δ_{BA} *of* $\ln \sigma_1$.

Proof. The proof follows using (4.6), (4.8) and (4.10) and the Brewster and Zidek technique.

Finally, we mention that the generalized Bayes estimator of $\ln \sigma_1$ with respect to the noninformative prior $\pi(\mu, \sigma_1) = 1/\sigma_1, \mu < 0, \sigma_1 > 0$ can be seen to be

$$\delta_{\mathrm{GB}}^{-} = \begin{cases} \ln T_1 - \psi(n-1), & \text{if } X_1 < 0, \\ \ln(T_1 + nX_1) - \psi(n-1), & \text{if } X_1 \geq 0. \end{cases} \tag{4.13}$$

□

Example 4.1. As an application of the results of this section we consider the data set given in Grubbs (1971). The data can be shown to follow a two-parameter exponential distribution. The data gives mileage at failure for nineteen military personnel carriers. Various estimators considered in this section are computed and given here (the values reported are accurate up to 2 decimal places): $\delta_{\mathrm{ML}} = 6.7277$, $\delta_{\mathrm{BA}} = 6.8098$, $\delta_{\mathrm{IB}} = 5.1217$. As δ_{IB} has the best risk performance, the estimate 5.1217 is recommended for $Q(\sigma)$.

5 Numerical comparisons

In this section, the risk performance of various estimators derived in the Sections 2, 3 and 4 is compared numerically. In Section 2, the UMVUE was derived. In Section 3, the improving estimators over the MLE in a class of affine equivariant estimators was obtained and further improvements over δ_{ML} and δ_{MV} were obtained in a class of scale equivariant estimators when $\mu < 0$. In Table 1, The relative percentage risk improvement over the estimator δ_{ML} by δ_{IML}, δ_{MV} and δ_{IMV} is presented for $k = 2$. The risk values of the estimators were calculated using simulations based on 10,000 samples of size $n = 10$ and different values of μ, σ_1

Table 1 *The relative percentage risk improvement over δ_{ML} by δ_{IML}, δ_{MV} and δ_{IMV} for $k = 2$*

μ	σ_1	σ_2	δ_{IML}	δ_{MV}	δ_{IMV}	μ	σ_1	σ_2	δ_{IML}	δ_{MV}	δ_{IMV}
−0.1	0.2	0.5	0.04	17.76	17.77	−0.02	0.2	0.5	5.29	17.77	18.67
		1.0	0.06	17.69	17.72			1.0	6.60	17.69	18.67
		2.0	0.07	17.62	17.64			2.0	7.27	17.62	18.68
	0.5	0.4	0.29	17.92	17.98		0.5	0.4	9.04	17.92	19.30
		2.0	1.72	17.72	18.05			2.0	13.24	17.72	19.78
		2.5	1.87	17.69	18.01			2.5	13.47	17.69	19.80
	1.0	0.5	1.14	17.94	18.10		1.0	0.5	11.94	17.94	17.67
		1.5	4.14	17.83	18.63			1.5	15.69	17.83	20.28
		2.0	4.84	17.79	18.40			2.0	16.19	17.79	20.30
	1.5	0.5	1.42	17.98	18.14		1.5	0.5	12.55	17.98	19.67
		1.0	4.22	17.94	18.62			1.0	15.80	17.94	20.39
		2.0	6.84	17.85	18.82			2.0	17.30	17.85	20.56

and σ_2. We have chosen two values of μ as -0.1 and -0.02. We observe marginal improvement over the δ_{ML} by δ_{IML} but substantial improvement by δ_{MV} and δ_{IMV}. The amount of improvement is more for values of μ close to 0. For the sake of space we have presented very few values, however, similar observations are made for various other values of n, μ, σ_1 and σ_2.

In Section 4, the improvements were obtained over the best affine equivariant estimator in a class of scale equivariant estimators. The cases of restrictions on μ, that is, $\mu \geq 0$ and $\mu < 0$ were also considered. The risk values of the proposed estimators are calculated using simulations based on 10,000 samples of sizes $n = 10, 20, 30$. Since the risk functions of the estimators are independent of σ_1, we take $\sigma_1 = 1$ for simulations. The plots of the risk functions of the proposed estimators for different sample sizes are given in Figures 1(a)–(i) below. On the basis of the simulation results, we observe the following:

(i) Figures 1(a), (b), (c) represent the risk function of the estimator δ_{IB} for $n = 10, 20$ and 30, respectively when μ is not restricted. The risk of the BAEE δ_{BA} is 0.1179, 0.0538 and 0.0353, respectively. We observe that the region of improvement of δ_{IB} over δ_{BA} becomes smaller for larger sample sizes. For the present case, the region of improvement is approximately within $|\mu| \leq 0.05$. Outside this region δ_{IB} takes constant value, equal to the risk of δ_{BA}. It is also observed that when $|\mu|$ becomes smaller then the margin of improvement increases and the maximum improvement occurs when $|\mu|$ is close to zero.

(ii) When μ is restricted on the positive real line, the risk functions of the estimators δ^+_{IB} and δ^+_{GB} are plotted in Figures 1(d), (e), (f). It is observed that δ^+_{IB} improves δ_{BA} when μ is close to zero. The region of improvement is approximately $\mu \leq 0.1$. We observe also that the estimator δ^+_{GB} improves δ_{BA}. δ^+_{GB} performs better than δ^+_{IB} approximately, when $0.06 \leq \mu \leq 1$, and δ^+_{IB} performs better than δ^+_{GB} when $0 \leq \mu \leq 0.06$.

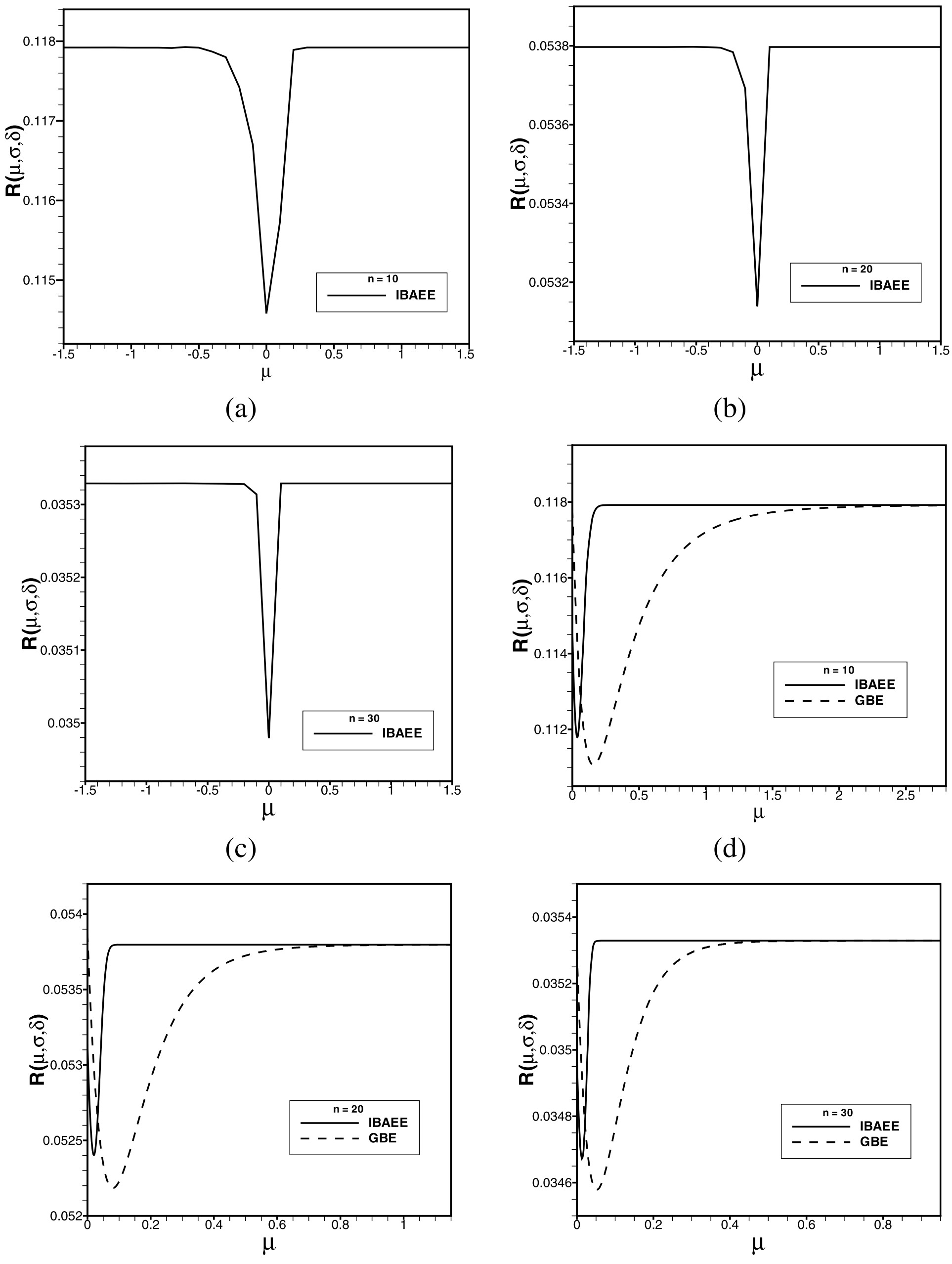
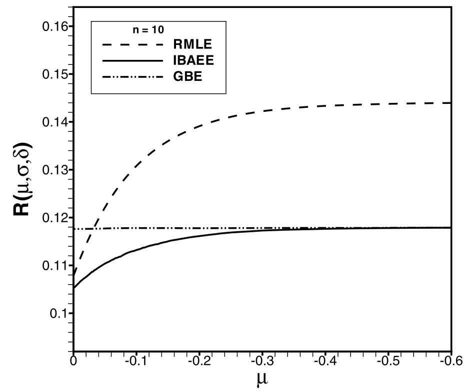
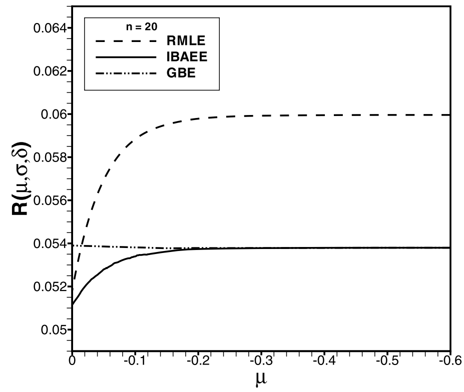


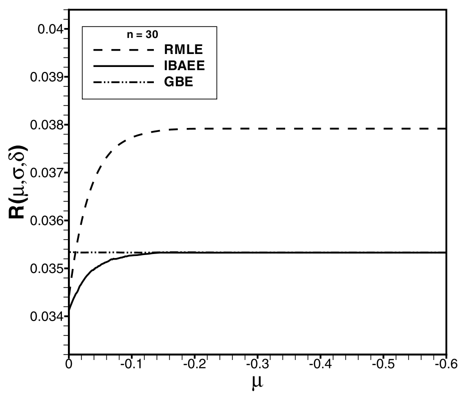
Figure 1 *Risk plots of* δ_{IB}, δ_{IB}^{+}, δ_{GB}^{+}, δ_{IB}^{-}, δ_{GB}^{-} *and* δ_{RM}.



(g)



(h)



(i)

Figure 1 *Continued.*

(iii) When μ is negative δ_{RM} performs better than the δ_M for $\mu \geq 0.1$ and for other values of μ, δ_M performs better than δ_{RM}. We observe that δ_{GB}^{-} is a better estimator than δ_{RM} when $\mu \leq -0.04$ and reverse happens for $\mu \geq -0.01$. It is also noticed that δ_{IB}^{-} always perform well than δ_{RM}. The region of improvement of the estimator δ_{IB}^{-} over δ_{BA} is $\mu \geq -0.1$. Figures 1(g), (h), (i) represent the risk functions of the estimators δ_{GB}^{-}, δ_{RM} and δ_{IB}^{-}.

6 Conclusion

The concept of entropy was introduced in thermodynamics but has later found applications in diverse areas such as communication, biological systems and social sciences (see Robinson (2008)). In this paper, the problem of estimating the Renyi entropy of k (≥ 1) exponential populations with a common location but different scale parameters has been explored. This model arises frequently in reliability and life testing studies. Consider, for example, k manufacturing processes for a product. Assume that the life of the product manufactured using the ith process has an exponential distribution with location parameter μ_i and scale parameter σ_i, $i = 1, \ldots, k$. Here due to difference in the manufacturing processes, the expected lives $(\mu_i + \sigma_i)$'s may be different, but commercial considerations may force the producers to keep the minimum guarantee times μ_i's to be the same.

In this paper, we derive the UMVUE using Rao–Blackwellization. The MLE is shown to be inadmissible in a class of affine equivariant estimators. Further estimators improving upon the MLE and the UMVUE are derived in a class of scale equivariant estimators when $\mu < 0$. As the best affine equivariant (BAEE) does not exist for k (≥ 2), we consider $k = 1$ as a special case and obtain an estimator improving the BAEE. The cases when prior considerations force μ to be non-negative or negative are also discussed. The improved performance of new estimators is demonstrated through a numerical comparison of risk functions for $k = 1, 2$ using Monte Carlo simulations. An application to the actual data set is also given.

Acknowledgments

The authors thank the reviewers for their suggestions which have improved the content and the presentation of the paper.

References

Ahmed, N. A. and Gokhale, D. V. (1989). Entropy expressions and their estimators for multivariate distributions. *IEEE Transactions on Information Theory* **35**, 688–692. MR1022093

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

Cover, T. M. and Thomas, J. A. (2006). *Elements of Information Theory*. New York: Wiley. MR2239987

Ghosh, M. and Razmpour, A. (1984). Estimation of the common location parameter of several exponentials. *Sankhyā, Ser. A* **46**, 383–394. MR0798044

Grubbs, F. E. (1971). Approximate fiducial bounds on reliability for the two parameter negative exponential distribution. *Technometrics* **13**, 873–876.

Harte, D. and Jones, D. V. (2005). The entropy and its uses in earthquake forecasting. *Pure and Applied Geophysics* **162**, 1229–1253.

Kayal, S. and Kumar, S. (2013). Estimation of the Shannon's entropy of several shifted exponential populations. *Statistics & Probability Letters* **83**, 1127–1135. MR3041385

Khinchin, A. I. (1957). *Mathematical Foundations of Information Theory*. New York: Dover. MR0092709

Lehmann, E. L. and Romano, J. P. (2005). *Testing Statistical Hypotheses*. New York: Springer. MR2135927

Misra, N., Singh, H. and Demchuk, E. (2005). Estimation of the entropy of multivariate normal distribution. *Journal of Multivariate Analysis* **92**, 324–342. MR2107880

Renyi, A. (1961). On measures of entropy and information. In *Proceedings of Fourth Berkeley Symposium on Mathematics, Statistics and Probability, Vol. 1* 547–461. Berkeley: Univ. California Press. MR0132570

Robinson, D. W. (2008). Entropy and uncertainty. *Entropy* **10**, 493–506. MR2465846

Shannon, C. (1948). A mathematical theory of communication. *The Bell System Technical Journal* **27**, 379–423. MR0026286

S. Kayal
Department of Mathematics
National Institute of Technology Rourkela
Rourkela-769008
India
E-mail: suchandan.kayal@gmail.com

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

P. Vellaisamy
Department of Mathematics
Indian Institute of Technology Bombay
Powai-400076
India
E-mail: pv@math.iitb.ac.in