Shrinkage Estimators Dominating Some Naive Estimators of the Selected Entropy

Masihuddin® and Neeraj Misra®

Abstract—Consider two populations characterized by independent random variables $X_1$ and $X_2$ such that $X_i$, $i = 1, 2$, follows a gamma distribution with an unknown scale parameter $\theta_i > 0$, and known shape parameter $\alpha > 0$ (the same shape parameter for both the populations), here $(X_1, X_2)$ may be an appropriate minimal sufficient statistic based on independent random samples from the two populations. The population associated with the larger (smaller) Shannon entropy is referred to as the “worse” (“better”) population. For the goal of selecting the worse (better) population, a natural selection rule is the one that selects the population corresponding to $\max\{X_1, X_2\} \{\min\{X_1, X_2\}\}$ as the worse (better) population. This natural selection rule is known to possess several optimum properties. We consider the problem of estimating the Shannon entropy of the population selected using the natural selection rule (to be referred to as the selected entropy) under the mean squared error criterion. In order to improve upon various naive estimators of the selected entropy, we derive a class of shrinkage estimators that shrink various naive estimators towards the central entropy. For this purpose, we first consider a class of naive estimators comprising linear, scale and permutation equivariant estimators and identify optimum estimators within this class. The class of naive estimators considered by us contains three natural plug-in estimators. To further improve upon the optimum naive estimators, we consider a general class of equivariant estimators and obtain dominating shrinkage estimators. We also present a simulation study on the performances of various competing estimators. Illustrative examples have been given using real data sets.

Index Terms—Admissible estimator, generalized Bayes estimator, inadmissible estimator, linear, scale and permutation equivariant estimators, mean squared error, naive estimators, natural selection rule, plug-in estimators, selected better entropy, selected worse entropy, Shannon’s entropy, shrinkage estimator.

I. INTRODUCTION

In numerous real life situations, it may be of interest to select a target population (e.g., the one having the largest mean or the smallest variance) from the $k$ ($\geq 2$) available populations. Based on sample observations from the $k$ populations, one of the $k$ populations is chosen as the target population using a predetermined decision/selection rule. For example, if the target population is the one having the largest mean (or the smallest variance) then an experimenter may use the decision rule that selects the population yielding the largest sample mean (or the smallest sample variance). Such selection problems fall under the general framework of so called “Ranking and Selection Problems”. The indifference-zone approach of [1] and the subset selection approach of [2] are the two well-known classical approaches that have been used in the formulation of ranking and selection problems. After the selection of the target population, the experimenter may also be interested in estimating some population characteristic(s) (e.g., mean, variance, entropy etc.) of the selected population. In the statistical literature, such estimation problems are popularly known as “Estimation After Selection Problems”. Because of their applications in various fields of science, engineering, economics and medicine, these estimation problems have attracted the attention of many eminent researchers and there exists a vast literature on these problems. Some of the key contributions are due to [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], and [14]. For the majority of univariate problems, the problem of ordering and selection are based on the associated location or scale parameters of the underlying populations. But in many real life situations, ranking and selection problems may be of interest in terms of suitably defined functions of location/scale parameters. For example, in statistical information theory, one example could be ranking and selection procedures from several populations based on some suitable measure of information or uncertainty. In this context, [15] and [16] have discussed the problem of ranking and selection of populations in terms of entropy and divergences. In this article, we have considered the problem of estimating Shannon entropy of the selected gamma population.

In 1948, Claude Shannon in his seminal paper ([17]) introduced the notion of entropy which paved the way for a separate field of research, called Information Theory. The notion of entropy has been used in various fields of science, engineering, economics and medicine to describe uncertainty. In the field of Statistics, entropy measures the uncertainty associated with a random variable (rv). Let $X_1$ and $X_2$ be two rvs with the Lebesgue probability density functions (pdfs) $f_{\theta_1}(\cdot)$ and $f_{\theta_2}(\cdot)$, respectively, where $\theta_i \in \Omega \subset \mathbb{R}$, $i = 1, 2$; here $\mathbb{R}$ denotes the real line and $\Omega$ denotes the common parameter space of $\theta_i$s. The Shannon entropy of $X_i$ (or of the associated probability distribution) is defined by:

$$H(\theta_i) = \mathbb{E}_{\theta_i} \left( - \ln f_{\theta_i}(X_i) \right), \quad \theta_i \in \Omega, \quad i = 1, 2.$$
For two populations $\Pi_1$ and $\Pi_2$, characterized by random variables $X_1$ and $X_2$, the one corresponding to the larger (smaller) entropy $\max\{H(\theta_1), H(\theta_2)\}$ ($\min\{H(\theta_1), H(\theta_2)\}$) is considered to be more (less) volatile (chaotic) than the other. For example, if $\Pi_1$ and $\Pi_2$ are populations of daily stock prices of two stocks, then the stock corresponding to $\max\{H(\theta_1), H(\theta_2)\}$ ($\min\{H(\theta_1), H(\theta_2)\}$) is considered to be more riskier (safer) than the other. We call the population corresponding to riskier (safer) stock the “worse” (“better”) population. In such a situation, it may be of interest to select the worse (better) population and also to have an estimate of the entropy (volatility index) of the selected population. Since $H(\theta_i)$s are unknown, they may be estimated by appropriate estimators $\hat{H}_1(X)$ and $\hat{H}_2(X)$ based on independent random samples from the two populations, and the population corresponding to $\max\{\hat{H}_1(X), \hat{H}_2(X)\}$ ($\min\{\hat{H}_1(X), \hat{H}_2(X)\}$) may be identified as the worse (better) population. We call such a selection rule the “natural selection rule”. Under the assumption that the underlying pdfs $\{f_\theta : \theta \in \Omega\}$ have the monotone likelihood ratio property, the natural selection rule is known to possess several optimum properties (e.g., the selection rule is best invariant, minimax, etc. For more details, one may refer [18], [19]). After the target population (riskier or safer stock) has been selected, it may be desired to have an estimate of the volatility index (Shannon entropy) of the selected population.

Under the set up of $X_1$ and $X_2$ following gamma distributions with unknown scale parameters and having a common known shape parameter, we study this problem and obtain shrinkage estimators that shrink some naive estimators of the selected entropy towards the central entropy. Let $\Pi_1$ and $\Pi_2$ be two populations such that the observations from the population $\Pi_i$, $i = 1, 2$, follow a gamma distribution with an unknown scale parameter $\theta_i$ ($> 0$) and known shape parameter $\beta$ ($> 0$). Here the gamma distributions corresponding to both the populations have the same shape parameter $\beta$. Let $X_{1i}, X_{2i}, \ldots, X_{ni}$ be a random sample of size $n$ drawn from the population $\Pi_i$ and $X_i = \sum_{j=1}^{n} X_{ij}$, $i = 1, 2$. Assume that the two random samples are mutually independent. Then $X = (X_1, X_2)$ is a complete-sufficient (and, hence minimal sufficient) statistic for $(\theta_1, \theta_2) \in \Theta = (0, \infty) \times (0, \infty)$ (see [13]), $X_1$ and $X_2$ are independently distributed, and $X_i$, $i = 1, 2$, follows a gamma distribution with an unknown scale parameter $\theta_i$ and known shape parameter $\alpha = n\beta$. The pdf of $X_i$ is given by

$$f(x|\theta_i, \alpha) = \frac{1}{\Gamma(\alpha)\theta_i^\alpha} x^{\alpha-1} e^{-\frac{x}{\theta_i}} ; \quad x > 0, \alpha > 0, \theta_i > 0,$$

for $i = 1, 2$, where, $\Gamma(\alpha)$ denotes the usual gamma function. The Shannon entropy of the population $\Pi_i$ ($i = 1, 2$) is given by

$$H(\theta_i) = \mathbb{E}_{\theta_i}(-\ln f(X_i|\theta_i, \alpha)) = \ln \theta_i + \alpha + \ln(\Gamma(\alpha)) + (1 - \alpha)\psi(\alpha), \quad i = 1, 2,$$

where, $\psi(\alpha) = \Gamma(\alpha)/\Gamma(\alpha - 1), \alpha > 0$, denotes the digamma function; here $\Gamma'(\alpha)$ denotes the derivative of the gamma function $\Gamma(\alpha)$. Call the population corresponding to $\max\{H(\theta_1), H(\theta_2)\}$ ($\min\{H(\theta_1), H(\theta_2)\}$) the “worse” (“better”) population. Equivalently the population corresponding to $\max\{\theta_1, \theta_2\}$ ($\min\{\theta_1, \theta_2\}$) is called the “worse” (“better”) population. Consider the natural selection rule that selects the population corresponding to $Z_2 = \max\{X_1, X_2\}$ ($Z_1 = \min\{X_1, X_2\}$) as the worse (better) population. The natural selection rule is known to possess several desirable optimum properties (see [18], [19], [20], [21]). We first discuss estimation of the selected worse entropy (entropy of the population identified as the worse population). Let $S = S(X)$ denote the index of the selected population, i.e., $S = i$, if $X_i = Z_2$, $i = 1, 2$; here $X = (X_1, X_2)$. Our goal is to estimate the Shannon entropy of the selected population, which is equivalent to the estimation of

$$H_S(\theta) = \left\{ \begin{array}{ll} \ln \theta_1, & \text{if } X_1 \geq X_2 \\ \ln \theta_2, & \text{if } X_1 < X_2 \end{array} \right.$$  

where, $\theta = (\theta_1, \theta_2)$ and, for any event $A$, $I(A)$ denotes its indicator function. We consider the squared error loss function, given by

$$L(\theta, a) = (a - H_S(\theta))^2, \quad a \in \Theta, \quad a \in A,$$

where, $\Theta = (0, \infty) \times (0, \infty)$ is the parameter space and $A = \mathbb{R}$ is the action space.

A natural estimator of $H_S(\theta)$ can be obtained by plugging in suitable estimators of $\theta_1$ and $\theta_2$ (or $\ln \theta_1$ and $\ln \theta_2$) in the expression of $H_S(\theta)$, for $\theta_1$ and $\theta_2$ (or $\ln \theta_1$ and $\ln \theta_2$). Note that, $X_1/\alpha$ and $X_2/\alpha$, respectively, are the maximum likelihood estimator (mle) and the best scale equivariant estimators of $\theta_i, i = 1, 2$ (see [13]). Thus two natural estimators of $H_S(\theta)$ are $\hat{H}_N^{(1)}(X) = H_S(X_1/\alpha, X_2/\alpha) = \ln Z_2 - \ln(\alpha)$ and $\hat{H}_N^{(2)}(X) = H_S(X_1/\alpha, X_2/\alpha) = \ln Z_2 - \ln(\alpha + 1)$. Also, note that, $\ln(X_i) - \psi(\alpha)$ is the best scale equivariant estimator (also the uniformly minimum variance unbiased estimator) of $\ln \theta_i, i = 1, 2$, under the squared error loss function. Thus, another natural estimator of $H_S(\theta)$ is $\hat{H}_N(\alpha)(X) = \ln Z_2 - c, c \in \mathbb{R}$. We call the class $K_1$, the class of naive estimators of $H_S(\theta)$. Interestingly, the class of naive estimators $K_1$ also contains a class of generalized Bayes estimators of $H_S(\theta)$ under the squared error loss function (1.2), with respect to a class of improper prior distributions (that also contains the Jeffreys non-informative prior distribution). It will be meaningful to identify optimum estimators within class $K_1$ of naive estimators. To deal with this issue, we consider characterizing estimators that are admissible (or inadmissible) within class $K_1$ of naive estimators. We will also find naive estimators dominating over the inadmissible naive estimators in class $K_1$.

The given estimation problem is invariant under the scale group of transformations $G_a = \{g_a : g_a(x) = ax, a > 0\}$ and also under the group of permutations $G_h = \{h_1, h_2\}$, where $x = (x_1, x_2), ax = (ax_1, ax_2), h_1(x_1, x_2) = (x_1, x_2)$ and $h_2(x_1, x_2) = (x_2, x_1), (x_1, x_2) \in (0, \infty) \times (0, \infty)$. Under the
group of transformations $G_a$, $X \rightarrow aX$, $\theta \rightarrow a\theta$, so that
$\ln \theta_1 \rightarrow \ln \theta_i + \ln a, i = 1, 2,$ and $H_S(\theta) \rightarrow H_S(\theta) + \ln a.$
Under the transformation $(x_1, x_2) \rightarrow (x_2, x_1)$, $(\theta_1, \theta_2) \rightarrow
(\theta_2, \theta_1)$ and $H_S(\theta) \rightarrow H_S(\theta)$. Therefore, it is natural to
consider estimators $\delta$ satisfying $\delta(aX_1, aX_2) = \delta(X_1, X_2) + \ln a, \forall a > 0,$ and $\delta(X_1, X_2) = \delta(X_2, X_1).$ Any such estimator will have the following form
$$\delta_S(X) = \ln Z_2 - \Phi(T), \quad (I.3)$$
for some function $\Phi(\cdot)$ defined on $(0,1]$, $T = \frac{Z_2}{Z_1}$. An estimator
of the type (I.3) will be called a scale and permutation equivariant estimator of $H_S(\theta)$. Let $K_2$ denote the class
of all scale and permutation equivariant estimators of the type
(I.3). Clearly, $K_1 \subseteq K_2$. We observe that, for an equivariant
estimator $\delta_S \in K_2$, the risk function (mean squared error)
$R(\theta, \delta_S) = E_{\theta}(\delta_S(X) - H_S(\theta))^2$ depends on $\theta$ through
$\mu = \frac{Z_2}{Z_1}$, where, $\mu_1 = \min\{\theta_1, \theta_2\}$, $\mu_2 = \max\{\theta_1, \theta_2\}$ and
$\mu \geq 1$. Therefore, for notational simplicity we denote $R(\theta, \delta_S)$
in $R_{\mu}(\delta), \mu \geq 1$.

In the literature on decision theoretic estimation problems,
notion of shrinkage has been widely used to find improvements
over naive estimators under different settings. A shrinkage
estimate of a naive estimate is an estimate obtained by shrinking
a raw extreme estimate towards a central value. In many
situations developing shrinkage estimators of naive estimators
result in better estimators. Following seminal works of [22],
[23], [24], [25], and [26], several researchers have obtained
shrinkage estimators dominating naive estimators in different
settings. Since $K_1 \subseteq K_2$, in order to find shrinkage estimators
dominating naive estimators in class $K_1$, we explore the class
$K_2$ of scale and permutation equivariant estimators. Note that
a naive estimator $\delta_S \in K_1$ is a plug-in estimator with plug-ins
for $\ln \theta_1$ and $\ln \theta_2$ as $\ln X_1 - c$ and $\ln X_2 - c$, respectively. If the
data pretends that $\ln \theta_1$ and $\ln \theta_2$, or equivalently $\theta_1$ and $\theta_2$,
are close (evidenced through large observed value of $\frac{Z_2}{Z_1}$), then,
derunder the notion of shrinkage, it may be appropriate to shrink
the naive estimator $\delta_S(X_1, X_2)$ towards a plug-in estimate
corresponding to common scale parameter $\theta (\theta_1 = \theta_2 = \theta)$. This
leads to considering shrinkage estimators of the type
$$\delta_{c,d}(X) = \begin{cases} 
\ln Z_2 - c, & \text{if } \frac{Z_1}{Z_2} < d, \\
\ln(X_1 + X_2) - c - \ln(1 + d), & \text{if } \frac{Z_1}{Z_2} \geq d,
\end{cases} \quad c \in \mathbb{R}, \ d > 0.$$ 
Here the choice $c - \ln(1 + d)$ attached to
$\ln(X_1 + X_2)$ ensures continuity (smoothness) of $\delta_{c,d}$. Following
the ideas of [23] and [26], we derive shrinkage estimators of
$H_S(\theta)$ dominating over some optimum naive estimators
(including some generalized Bayes estimators) belonging to
class $K_2$. As a consequence of this general result, we obtain
shrinkage estimators improving upon several optimum naive
estimators belonging to class $K_1$. In Section III, analogous
results are obtained for the problem of estimating the selected
better entropy. In Section IV, we report a numerical study
assessing performances of various competing estimators. Finally,
to illustrate applicability of various optimum estimators, analysis
of two real data sets is provided in Section V of the paper.

II. ESTIMATION ENTROPY OF THE WORSE SELECTED
POPULATION

The following lemma will be useful in obtaining various
findings of the paper.

Lemma 2.1: Let $X \sim Gamma(\alpha, \beta)$ (gamma distribution
with scale parameter $\beta > 0$ and shape parameter $\alpha > 0$),
so that the pdf of $X$ is,
$$f(x|\alpha, \beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} e^{-\frac{x}{\beta}} \frac{x^{\alpha-1}}{\Gamma(\alpha)}; \ x > 0, \alpha > 0, \beta > 0.$$

Let
$$c_1(\alpha) = \psi(\alpha), \ c_2(\alpha) = \frac{1}{\beta^\alpha \Gamma(\alpha)}.$$

Proof: The proof of the first assertion above is straightforward and the second assertion follows using Jensen’s
inequality on $\psi(\alpha) = E (\ln \frac{X}{\beta}).$ (iii) For proving the third assertion, let $Z_{1,\alpha}$ and $Z_{2,\alpha}$ be
independent and identically distributed (iid) as $Gamma(\alpha, 1).$ Further let $Z_{2,\alpha} \sim Gamma(2\alpha, 1),$ so that $Z_{1,\alpha} + Z_{2,\alpha} \overset{d}{=} Z_{2,\alpha} \ (\text{where} \ d \ \text{means equality in distribution}),$ and
$$c_2(\alpha) = \frac{2}{\beta^\alpha \Gamma(\alpha)}.$$

Also
$$c_2(\alpha) - \psi(2\alpha) = E (\ln (\max \{Z_{1,\alpha}, Z_{2,\alpha}\})) - \ln (Z_{1,\alpha} + Z_{2,\alpha}) \geq \frac{1}{2} \ln \left(\frac{Z_{1,\alpha} + Z_{2,\alpha}}{2}\right), \text{w.p.1}$$
$$\implies c_2(\alpha) > \psi(2\alpha) - \ln(2).$$
Finally
\[ \psi(2\alpha) - c_1(\alpha) = E\left[\ln \left(\frac{Z_{1,a} + Z_{2,a}}{Z_{1,a} + Z_{2,a}}\right)\right] \]
\[ = E \left[ \ln \left( \frac{Z_{1,a} + Z_{2,a}}{Z_{1,a} + Z_{2,a}} \right) \right] > - \ln \left( \frac{Z_{1,a} + Z_{2,a}}{Z_{1,a} + Z_{2,a}} \right) \]
\[ = - \ln \left( \frac{1}{2} \right) = \ln 2, \]
where \( B_{\alpha,\alpha} = \frac{Z_{1,a} + Z_{2,a}}{Z_{1,a} + Z_{2,a}} \) has the Beta\((\alpha, \alpha)\) distribution.

We first obtain a class of generalized Bayes estimators of \( H_S(\theta) \), under the squared error loss function (I.2). For this, we consider following class of improper prior densities for \( \theta = (\theta_1, \theta_2) \):
\[ \Pi_\beta(\theta) = \begin{cases} \frac{1}{\theta_1 \theta_2}, & \text{if } \theta \in \Theta, \\ 0, & \text{otherwise}. \end{cases} \quad (I.1) \]

Note that \( \Pi_0(\theta) = \frac{1}{\theta_1 \theta_2}, \) \( \theta \in \Theta, \) is the Jeffreys non-informative prior density. The posterior density of \( \theta = (\theta_1, \theta_2) \) given \( Y = (x_1, x_2) \in (0, \infty) \times (0, \infty) \) is obtained as
\[ \Pi_{\theta_2}(\theta) = \left\{ \prod_{i=1}^{2} \left\{ \frac{\phi^n(x_i \theta_i)}{\Gamma(\alpha) \theta_i^{\alpha+\beta}} e^{-\frac{x_i^2}{\theta_i}} \right\} \right\} \quad (I.2) \]
where \( \beta > -\alpha, \) i.e., the posterior distribution of \( \theta \), given \( Y = (x_1, x_2) \in (0, \infty) \times (0, \infty) \), is such that \( \theta_1^{-1} \) and \( \theta_2^{-1} \) are independent and \( \theta_1^{-1} \sim \text{Gamma}(\alpha + \beta, x_i^2), i = 1, 2. \) Also, the posterior risk of an estimator \( \delta \) is given by
\[ r_\beta(\delta, Y) = E_{\Pi_{\theta_2}}(\delta - H_S(\theta))^2, x \in (0, \infty) \times (0, \infty), \beta > -\alpha. \]

The generalized Bayes estimator, \( \delta_{GB}(X) \), which minimizes the posterior risk (II.2), is obtained as
\[ \delta_{GB}(X) = E_{\Pi_{\theta_2}}[H_S(\theta)] \]
\[ = \begin{cases} E_{\Pi_{\theta_2}}(\ln \theta_1), & \text{if } X_1 \geq X_2 \\ E_{\Pi_{\theta_2}}(\ln \theta_2), & \text{if } X_1 < X_2 \\ \ln X_1 - \psi(\alpha + \beta), & \text{if } X_1 \geq X_2 \\ \ln X_2 - \psi(\alpha + \beta), & \text{if } X_1 < X_2 \end{cases} \]
\[ = \delta_{\psi(\alpha + \beta)}(X). \]
Clearly the class \( K_{GB} = \{ \delta_{GB} : \beta > -\alpha \} \) of generalized Bayes estimators is contained in the class \( K_1 \) of naive estimators. In particular, the natural estimator \( \bar{H}_{K_1}^{GB}(X) = \delta_{GB}(X) = \ln Z_2 - \psi(\alpha) \) (analogue of the best scale equivariant estimator of \( \ln(\theta_1) \) and \( \ln(\theta_2) \)) is the generalized Bayes estimator with respect to Jeffrey’s non-informative prior density \( \Pi_0(\theta) = \frac{1}{\theta_1 \theta_2}, \theta \in \Theta. \)

Now we will attempt to find optimum estimators within the class \( K_1 \) of naive estimators under the squared error loss function (I.2). The following lemma will be useful in obtaining admissible estimators within the subclass \( K_1. \)

**Lemma 2.2:** Let \( U = \ln Z_2 - H_S(\theta) \). Then, for any \( \theta \in \Theta, \)
\[ E_{\theta}(U) = \int_{0}^{\infty} \ln(z) G_{\alpha} \left( \frac{z}{\mu} \right) g_{\alpha}(z) dz \]
\[ + \int_{0}^{\infty} \ln(z) G_{\alpha}(\mu z) g_{\alpha}(z) dz, \mu \geq 1, \quad (I.3) \]
where \( G_{\alpha}(\cdot) \) and \( g_{\alpha}(\cdot) \), respectively, denote the df and the pdf of Gamma\((\alpha, 1)\) distribution.

**Proof:** Let \( Y_i = \frac{X_i}{\theta_i}, i = 1, 2 \). Then \( Y_1 \) and \( Y_2 \) are iid Gamma\((\alpha, 1)\) random variables. Note that, the pdf of \( U \) is a permutation symmetric function of \((\theta_1, \theta_2)\). Thus, without loss of generality, we may take \( \theta_i = \mu_i, i = 1, 2. \) Then, for any \( \theta \in \Theta, \)
\[ E_{\theta}(U) = E_{\theta}[\ln Z_2 - H_S(\theta)] \]
\[ = E_{\theta}(\ln(\frac{X_1}{\theta_1})) I(X_1 > X_2) \]
\[ + E_{\theta}(\ln(\frac{X_2}{\theta_2})) I(X_2 > X_1) \]
\[ = E(\ln(Y_1)) I(Y_1 > \frac{\theta_1}{\theta_2} Y_2) \]
\[ + E(\ln(Y_2)) I(Y_2 > \frac{\theta_1}{\theta_2} Y_1) \]
\[ = \int_{0}^{\infty} \ln(z) G_{\alpha}(\frac{z}{\mu}) g_{\alpha}(z) dz + \int_{0}^{\infty} \ln(z) G_{\alpha}(\mu z) g_{\alpha}(z) dz. \]

The risk (mean squared error) function of an estimator \( \delta_\epsilon \in K_1 \), is given by
\[ R_\psi(\delta_\epsilon) = E[\ln Z_2 - \frac{1}{\theta_2^2} H_S(\theta)]^2, \mu \geq 1, c \in \mathbb{R}. \quad (II.4) \]
For any fixed \( \mu \geq 1, \) the risk function in (II.4) is minimized at
\[ c = c^*(\mu) \]
\[ = E_{\theta}[\ln Z_2 - H_S(\theta)] = E_{\theta}(U) \]
\[ = \int_{0}^{\infty} \ln(z) G_{\alpha} \left( \frac{z}{\mu} \right) g_{\alpha}(z) dz \]
\[ + \int_{0}^{\infty} \ln(z) G_{\alpha}(\mu z) g_{\alpha}(z) dz, \]
using (I.3). Clearly,
\[ \frac{d}{d\mu} c^*(\mu) = - \frac{1}{\mu^2} \int_{0}^{\infty} z \ln(z) g_{\alpha}(\frac{z}{\mu}) g_{\alpha}(z) dz \]
\[ + \int_{0}^{\infty} z \ln(z) g_{\alpha}(\mu z) g_{\alpha}(z) dz \]
\[ = - \frac{\Gamma(2\alpha) \mu^{\alpha-1}}{(\Gamma(\alpha))^2(1 + \mu^2)} \ln \mu \leq 0, \forall \mu \geq 1. \]
Consequently,
\[ \inf_{\mu \geq 1} c^*(\mu) = \psi(\alpha) = c_1(\alpha), \text{ say,} \quad (I.5) \]
and
\[ \sup_{\mu \geq 1} c^*(\mu) = 2 \int_{0}^{\infty} \ln(z) G_{\alpha}(z) g_{\alpha}(z) dz = c_2(\alpha), \text{ say.} \quad (I.6) \]
In the following theorem, we will characterize estimators that are admissible/inadmissible within the class \( \mathcal{K}_1 \) of naive estimators.

**Theorem 2.1:** Let \( c_1(\alpha) = \psi(\alpha) \) and \( c_2(\alpha) = 2\int_0^\infty \ln(z) G_\alpha(z) g_\alpha(z) dz \), \( \alpha > 0 \). Then, for estimating \( H_S(\theta) \), under the mean squared error criterion, the estimators in the class \( \mathcal{K}_1, M = \{ \delta_\epsilon \in \mathcal{K}_1 : c \in [c_1(\alpha), c_2(\alpha)] \} \) are admissible within the class of estimators \( \mathcal{K}_1 \) of naive estimators. The estimators in the class \( \mathcal{K}_1, I = \{ \delta_\epsilon \in \mathcal{K}_2 : c \in (-\infty, c_1(\alpha)) \cup (c_2(\alpha), \infty) \} \) are inadmissible. Furthermore, for any \(-\infty < d < c \leq c_1(\alpha) \) or \( c_2(\alpha) \leq c < d < \infty\),

\[
R_\mu(\delta_\epsilon) < R_\mu(\delta_\epsilon), \quad \forall \mu \geq 1.
\]

**Proof:** For any fixed \( \mu \geq 1 \), the risk function \( R_\mu(\delta_\epsilon) \), defined by (II.4), is a strictly increasing function of \( c \) on \( [c^*(\mu), \infty) \), it is a decreasing function of \( c \) on \( (-\infty, c^*(\mu)] \), and it achieves its minimum at \( c = c^*(\mu) \). Since \( c^*(\mu) \) is a continuous function of \( \mu \), using (II.5-II.6), it follows that \( c^*(\mu) \) takes all values in the interval \( [c_1(\alpha), c_2(\alpha)] \). Thus, we conclude that each \( c \in (c_1(\alpha), c_2(\alpha)) \) minimizes the risk function \( R_\mu(\delta_\epsilon) \) at some \( \mu \in [1, \infty) \). This establishes that the estimators \( \delta_\epsilon(\cdot) \), for \( c \in (c_1(\alpha), c_2(\alpha)) \) are admissible within the subclass \( \mathcal{K}_2 \). Further, continuity of the risk function ensures the admissibility of estimator \( \delta_{c_2(\alpha)}(\cdot) \) within the subclass \( \mathcal{K}_1 \) (see proof of Theorem 3.1 of [23], [27]). This proves the first assertion. Also, since \( c_1(\alpha) = \psi(\alpha) \leq c^*(\mu) < c_2(\alpha), \forall \mu \geq 1 \), it follows for any \( \mu \geq 1 \) the risk function \( R_\mu(\delta_\epsilon) \) is a strictly decreasing function of \( c \) on \( (-\infty, c_1(\alpha)] \) and it is a strictly increasing function of \( c \) on \( [c_2(\alpha), \infty) \). This proves the second assertion. Hence the result follows.

In the sequel we discuss optimality of natural estimators \( \hat{H}_{N_1}(X) \), \( i = 1, 2, 3 \). It directly follows from Theorem 2.1 that within the class \( \mathcal{K}_1 \) of naive estimators of \( H_S(\theta) \), the natural estimator \( \hat{H}_{N_1}(X) = \delta_{c_1(\alpha)}(\hat{X}) = \ln Z_2 - \psi(\alpha) \) is an admissible estimator of \( H_S(\theta) \). Using lemma 2.1, we have \( \ln(\alpha + 1) > \ln(\alpha) - \psi(\alpha) = c_1(\alpha), \forall \alpha > 0 \). It can be numerically verified (see Table I and Figure 1-2) that \( c_2(\alpha) > (\ln(\alpha) - \ln(\alpha + 0.63) \end{IEEEeqnarray*} for \( \alpha > 0 \)). The above discussion along with Theorem 2.1, yields the following corollary.

**Corollary 1.1:** (a) Within the class \( \mathcal{K}_1 \) of naive estimators of \( H_S(\theta) \), the natural estimator \( \hat{H}_{N_1}(X) = \delta_{c_1(\alpha)}(\hat{X}) = \ln Z_2 - \psi(\alpha) \) is an admissible estimator for estimating \( H_S(\theta) \).

(b) For \( \alpha \in (0, 0.63) \), the natural estimator \( \hat{H}_{N_1}(X) = \hat{\delta}_{\alpha}(\hat{X}) = \ln Z_2 - \ln(\alpha + 1) \) is an inadmissible estimator for estimating \( H_S(\theta) \) and it is dominated by the naive estimator \( \hat{\delta}_{c_2(\alpha)}(\hat{X}) = \ln Z_2 - c_2(\alpha) \). However, for \( \alpha \in (0.63, \infty) \), the natural estimator \( \hat{H}_{N_1}(X) \) is an admissible estimator within the class \( \mathcal{K}_1 \) of naive estimators.

(c) For \( \alpha \in (0, 0.65) \), the natural estimator \( \hat{H}_{N_2}(X) = \hat{\delta}_{\alpha}(\hat{X}) \) is an inadmissible estimator of \( \hat{H}_{N_1}(X) = \hat{\delta}_{\alpha}(\hat{X}) \) and it is dominated by the naive estimator \( \hat{\delta}_{c_1(\alpha)}(\hat{X}) = \ln Z_2 - c_1(\alpha) \). However, for \( \alpha \in (0.65, \infty) \), the natural estimator \( \hat{H}_{N_2}(X) = \hat{\delta}_{\alpha}(\hat{X}) \) is an admissible estimator for estimating \( H_S(\theta) \) and it is dominated by the naive estimator \( \hat{\delta}_{c_1(\alpha)}(\hat{X}) = \ln Z_2 - c_1(\alpha) \). However, for \( \alpha \in
TABLE I
TABLE 2.1: VALUES OF $c_1(\alpha)$, $c_2(\alpha)$, $\delta_3(\alpha) = \psi^{-1}(c_2(\alpha))$, and $\psi(2\alpha) - \ln 2$ for Various Values of $\alpha$

| $\alpha$ | $c_1(\alpha)$ | $c_2(\alpha)$ | $\ln \alpha$ | $\ln(\alpha + 1)$ | $\delta_3(\alpha)$ | $\psi(2\alpha) - \ln 2$ |
|----------|----------------|----------------|--------------|------------------|------------------|------------------------|
| 0.2      | -5.289         | -2.682         | -1.609       | 0.182            | 0.183            | -3.254                 |
| 0.4      | -2.561         | -1.158         | -0.916       | 0.336            | 0.322            | -1.658                 |
| 0.6      | -1.54          | -0.51          | -0.47        | 0.47             | 0.428            | -0.982                 |
| 0.8      | -0.965         | -0.135         | -0.197       | 0.514            | 0.514            | -0.587                 |
| 1        | -0.577         | 0.115          | 0            | 0.693            | 0.588            | -0.270                 |
| 1.5      | 0.036          | 0.566          | 0.045        | 0.738            | 0.229            |                        |
| 2        | 0.422          | 0.685          | 0.593        | 0.86             | 0.562            |                        |
| 2.5      | 0.703          | 0.911          | 0.716        | 0.904            | 0.512            |                        |
| 3        | 0.922          | 1.272          | 1.098        | 1.07             | 0.30             |                        |
| 3.5      | 1.103          | 1.423          | 1.252        | 1.141            | 0.17             |                        |
| 4        | 1.256          | 1.553          | 1.366        | 1.218            | 0.12             |                        |
| 4.5      | 1.388          | 1.667          | 1.504        | 1.291            | 0.11             |                        |
| 5        | 1.506          | 1.769          | 1.609        | 1.359            | 0.15             |                        |
| 5.5      | 1.611          | 1.861          | 1.704        | 1.424            | 0.18             |                        |
| 6        | 1.706          | 1.944          | 1.791        | 1.486            | 0.19             |                        |
| 6.05     | 1.716          | 1.953          | 1.8         | 1.54             | 0.21             |                        |
| 6.5      | 1.792          | 2.021          | 1.871        | 2.014            | 1.35             |                        |
| 7        | 1.872          | 2.072          | 1.945        | 2.161            | 1.38             |                        |
| 8        | 2.015          | 2.22           | 2.079        | 2.197            | 2.28             |                        |
| 9        | 2.14           | 2.333          | 2.197        | 2.302            | 2.55             |                        |
| 10       | 2.351          | 2.431          | 2.302        | 2.397            | 2.77             |                        |
| 12       | 2.442          | 2.608          | 2.484        | 2.564            | 3.91             |                        |
| 15       | 2.674          | 2.832          | 2.708        | 2.772            | 5.93             |                        |
| 16       | 2.741          | 2.833          | 2.772        | 3.782            | 2.76             |                        |
| 18       | 2.862          | 2.996          | 2.890        | 3.944            | 2.76             |                        |
| 20       | 2.970          | 3.098          | 2.995        | 3.044            | 2.79             |                        |

Fig. 2. Plot of $\Delta_2(\alpha) = 2\int_0^\infty \ln(z)G_\alpha(z) \, g_\alpha(z) \, dz - \ln(\alpha + 1)$.

equivariant estimator. As a consequence of this general result, we obtain shrinkage estimators dominating various naive estimators of $H_S(\theta)$ belonging to the class $K_1$.

A. Shrinkage Type Improvements Over Scale and Permutation Equivariant Estimators

In this section we will attempt to derive conditions under which shrinkage type improvements over an arbitrary scale and permutation equivariant estimator can be found. For this purpose, we will consider orbit by orbit improvement of the risk function, as proposed by [26]. Recall that a typical estimator in the class $K_2$ of scale and permutation equivariant estimator of $H_S(\theta)$ is of the form $\delta_\Phi(X) = \ln Z_2 - \Phi(T)$, for some function function $\Phi(\cdot) : (0,1] \rightarrow \mathbb{R}$, where $T = \frac{Z_2}{Z_1}$.

We first provide two supporting lemmas that will be useful in proving the main result of this section.

**Lemma 2.3:** Let $U = \ln Z_2 - H_S(\theta)$. For any fixed $t \in (0,1]$ and $\mu \geq 1$, the conditional pdf of $U$, given $T = t$, is given by

$$f_{1,\theta}(u|t) = \frac{1}{\Gamma(2\alpha)} \left[ \frac{1}{\mu (1 + \frac{u}{\mu})^{\alpha / 2}} + \frac{\mu^{\alpha / 2}}{(1 + t \mu)^{\alpha / 2}} \right] ^{\frac{1}{2}} - \infty < u < \infty.$$

**Proof:** Let $t \in (0,1]$ and $\mu \in [1,\infty)$ be fixed. Since the pdf of $U$, given $T = t$, is a permutation symmetric function of $(\theta_1, \theta_2)$, without loss of generality, we may assume that $\theta_1 = \theta_i$, $i = 1,2$. Let $h_\theta(\cdot)$ denote the pdf of $T$. Then, for any fixed $t \in (0,1)$, the df of $U$, given $T = t$, is

$$F_{1,\theta}(u|t) = P_{\theta}(U \leq u|T = t) = \frac{1}{h_\theta(t)} \lim_{h \downarrow 0} N_1(h(u|t, \theta)), \quad (II.7)$$

where, for $-\infty < u < \infty$ and $h > 0$ (sufficiently small),

$$N_1(h(u|t, \theta) = P_{\theta}(U \leq u|T = t),$$

$$= P_{\theta}(\ln Z_2 - H_S(\theta) \leq u, t - h < T \leq t)$$

$$= P_{\theta}(X_1 \geq X_2, \ln(X_1) - \ln(\theta_1) \leq u, t - h < X_1 < X_2 \leq t)$$

$$= P_{\theta}(X_2 > X_1, \ln(X_2) - \ln(\theta_2) \leq u, t - h < X_1 < X_2 \leq t)$$

$$= P_{\theta}(\ln(Y_1) \leq u, \frac{(t - h)}{\mu} < Y_1 < \frac{t}{\mu})$$

$$+ P_{\mu}(\ln(Y_2) \leq u, \frac{(t - h)}{\mu} < Y_2 \leq \frac{t \mu}{Y_2})$$

$$= P_{\theta}(Y_1 \leq e^u, \frac{(t - h)}{\mu} Y_1 \mu < Y_2 \leq \frac{t \mu}{Y_2})$$

$$+ P_{\mu}(Y_2 \leq e^u, \frac{(t - h)}{\mu} Y_2 \mu < Y_2 \leq \frac{t \mu}{Y_2}),$$

where, $Y_i = \frac{X_i}{\theta_i}$, $i = 1,2$, so that $Y_1$ and $Y_2$ are iid Gamma($\alpha, 1$). Let $G_\alpha$ and $g_\alpha$ denote the df and pdf, respectively, of $Y_1$. Then,

$$N_1(h(u|t, \theta)$$

$$= \int_0^e e^u \left[ G_\alpha \left( \frac{t \mu}{Y_2} \right) - G_\alpha \left( \frac{(t - h)\mu}{Y_2} \right) \right] g_\alpha(y) \, dy + \int_0^e g_\alpha \left( \frac{t \mu}{Y_2} \right) \, dy,$$

$$- \infty < u < \infty, h > 0.$$

$$\Rightarrow \lim_{h \downarrow 0} N_1(h(u|t, \theta)$$

$$= \int_0^e \frac{\mu}{\mu} g_\alpha \left( \frac{t \mu}{Y_2} \right) g_\alpha(y) \, dy + \int_0^e \frac{\mu}{\mu} g_\alpha \left( \frac{t \mu}{Y_2} \right) \, dy,$$

$$- \infty < u < \infty, \quad (II.8)$$
Using (II.6) and (II.7), for any fixed $t \in (0, 1]$, the conditional pdf of $U$, given $T = t$, is given by

$$f_{1,\theta}(u|t) = \frac{\frac{\alpha}{\mu} g_0 \left( \frac{t^\alpha}{\mu} \right) g_0 \left( e^u \right) + \mu e^{2\alpha u} g_0 \left( \mu e^u \right) g_0 \left( e^u \right)}{\Gamma(2\alpha) \left( \frac{1}{\mu^\alpha + \frac{\mu^\alpha}{(1+\mu)^{2\alpha}}} \right)},$$

where $\Phi(\cdot)$ is the standard normal CDF.

Using Lemma 2.3, where

$$N = \frac{1}{\mu^\alpha} \int_{-\infty}^{\infty} u e^{2\alpha u} e^{-\left(1 + \frac{\alpha}{\mu} \right) u} du \quad + \quad \frac{\mu^\alpha}{\mu^\alpha} \int_{-\infty}^{\infty} u e^{2\alpha u} e^{-\left(1 + \mu \right) u} du$$

Therefore, for any fixed $t \in (0, 1]$, the risk function of any scale and permutation equivariant estimator $\Phi_\mu(t) = \psi(2\alpha) - k_1(\mu)$, $\mu \geq 1$,

where, $k_1(\mu)$ is defined by (II.9). Further, using Lemma 2.4, we get

$$\Phi(t) = \sup_{\mu \geq 1} R_1(\mu, \Phi(t)) = 0 < t \leq 1.$$
Since, \( P_\alpha \left\{ T : \Phi(T) > \Phi_\alpha(T) \right\} > 0 \), for some \( \alpha \in \Theta \), we conclude that \( R(\bar{\theta}, \delta_{\alpha}) = R(\bar{\theta}, \delta_{\alpha_1}) \geq 0, \forall \theta \in \Theta \), with strictly inequality for some \( \theta \in \Theta \). Hence, the result follows.

Now we will discuss applications of the above theorem in finding shrinkage type improvements over various naive estimators belonging to class \( K_1 \). Let \( c > \psi(2\alpha) - \varphi \). Then

\[
P_\alpha (c > \Phi_\alpha(T)) = P_\alpha (\ln(1 + T) > \psi(2\alpha) - c) > 0,
\]

\( \forall \theta \in \Theta \). Consequently, any naive estimator \( \delta_{\alpha}(X) = \ln Z_2 - c \), with \( c > \psi(2\alpha) - \varphi \), is inadmissible for estimating \( H_S(\theta) \) and is dominated by the shrinkage estimator

\[
\delta^{(S)}_{\alpha}(X) = \begin{cases} 
\ln Z_2 - c, & \text{if } Z_2^{1/2} < e^{\psi(2\alpha) - c} - 1; \\
\ln(X_1 + X_2) - \psi(2\alpha), & \text{if } Z_2^{1/2} \geq e^{\psi(2\alpha) - c} - 1.
\end{cases}
\]

(II.16)

From Lemma 2.1 (ii), we have, for any \( \alpha > 0 \), \( \psi(2\alpha) < \ln(2\alpha) \), i.e., \( \ln \alpha + 1 > \alpha > \psi(2\alpha) - \varphi \). Now, from the above discussion, it follows that natural estimators \( \hat{H}_{N_1}^{(1)}(X) = \delta_{\ln \alpha}(X) = \ln Z_2 - \ln \alpha \) and \( \hat{H}_{N_2}^{(1)}(X) = \delta_{\ln(\alpha + 1)}(X) = \ln Z_2 - \ln(\alpha + 1) \) are inadmissible for estimating \( H_S(\theta) \) and are dominated by shrinkage estimators

\[
\delta^{(S)}_{\ln \alpha}(X) = \begin{cases} 
\ln Z_2 - \ln \alpha, & \text{if } Z_2^{1/2} < e^{\psi(2\alpha) - c} - 1; \\
\ln(X_1 + X_2) - \psi(2\alpha), & \text{if } Z_2^{1/2} \geq e^{\psi(2\alpha) - c} - 1.
\end{cases}
\]

(II.17)

and

\[
\delta^{(S)}_{\ln(\alpha + 1)}(X) = \begin{cases} 
\ln Z_2 - \ln(\alpha + 1), & \text{if } Z_2^{1/2} < e^{\psi(2\alpha) - c} - 1; \\
\ln(X_1 + X_2) - \psi(2\alpha), & \text{if } Z_2^{1/2} \geq e^{\psi(2\alpha) - c} - 1.
\end{cases}
\]

(II.18)

respectively. A summary of the above discussion, in conjunction with Theorem 2.1, Lemma 2.1 (iii) and Theorem 2.2, is provided in the form of following theorem.

**Theorem 2.3:** (a) The naive estimators \( \{ \delta_{\psi} : c \in (-\infty, c_1(\alpha)) \cup (\psi(2\alpha) - \ln 2, \infty) \} \) are inadmissible for estimating \( H_S(\theta) \) under the squared error loss function (I.2). For any \( c \in (-\infty, c_1(\alpha)) \) the naive estimator \( \delta_{\psi}(X) = \ln Z_2 - c \) is dominated by the natural estimator \( \delta_{c_1(\alpha)}(X) = \ln Z_2 - c_1(\alpha) \); for any \( c \in (\psi(2\alpha) - \ln 2, c_2(\alpha)) \), the naive estimator \( \delta_{c}(X) \) is dominated by the shrinkage estimator, defined by (2.15) and for any \( c \in [c_2(\alpha), \infty) \) the naive estimator \( \delta_{\psi} \) is dominated by the shrinkage estimator \( \delta_{c_2(\alpha)}(X) \), as defined in (II.16).

(b) The natural estimator \( \hat{H}_{N_1}^{(1)}(X) = \delta_{\ln \alpha}(X) = \ln Z_2 - \ln \alpha \) is inadmissible for estimating \( H_S(\theta) \) and is dominated by the shrinkage estimator \( \delta^{(S)}_{\ln \alpha}(X) \), defined by (II.17).

(c) The natural estimator \( \hat{H}_{N_2}^{(1)}(X) = \delta_{\ln(\alpha + 1)}(X) = \ln Z_2 - \ln(\alpha + 1) \) is inadmissible for estimating \( H_S(\theta) \) and is dominated by the shrinkage estimator \( \delta^{(S)}_{\ln(\alpha + 1)}(X) \), defined by (II.18).

Remark 3.1: (i) The global admissibility of naive estimators \( \{ \delta_{\psi} : c \in [c_1(\alpha), \psi(2\alpha) - \ln 2] \} \) is unresolved. We believe that these estimators are globally admissible but we have not been able to prove this. This seems to be an interesting problem for future research.

(ii) A natural question that arises is whether an unbiased estimator of \( H_S(\theta) \) exists? We have tried to address this question and could not succeed in resolving it. Through our experience with analysis carried out to resolve the question, we conjecture that an unbiased estimator for the selected entropy \( H_S(\theta) \) does not exist. However, we have not been able to settle this question and it remains an open problem, that may also be considered in our future research.

### III. ESTIMATION OF THE ENTROPY OF THE BETTER SELECTED POPULATION

We call the population associated with \( \min \{ H(\theta_1), H(\theta_2) \} \), as the “better” population. A natural selection rule for selecting the better population is to choose the population corresponding to \( Z_1 = \min(X_1, X_2) \). Let \( M = M(X) \) denotes the index of the better selected population, i.e., \( M = i, \) if \( X_i = Z_1, i = 1, 2 \). Following selection of the better population, our goal is to estimate the Shannon entropy of the selected better population, which is equivalent to estimation of

\[
H_M(\theta) = \begin{cases} 
\ln \theta_1, & \text{if } X_1 \leq X_2; \\
\ln \theta_2, & \text{if } X_1 > X_2
\end{cases}
\]

(III.1)

In this section, we consider estimation of \( H_M(\theta) \) under the squared error loss function,

\[
L(\theta, a) = (a - H_M(\theta))^2, \quad \theta \in \Theta = (0, \infty) \times (0, \infty), (III.2)
\]

where \( a \in A = \mathbb{R} \). For the goal of estimating \( H_M(\theta) \), any scale and permutation equivariant estimator is of the form,

\[
d_\phi(X) = \ln Z_1 - \phi(V),
\]

(III.3)

with \( V = \frac{Z_1}{Z_2} \) and, for some real valued function \( \phi(\cdot) \) defined on \([1, \infty)\). We define \( M_2 \) the class of all scale and permutation equivariant estimators of \( H_M(\theta) \), given by (III.3). We observe that for any estimator \( d_\phi \in M_2 \), the risk function \( R(\theta, \phi) = \mathbb{E}_\theta ((\delta - H_M(\theta))^2) \) depends on \( \theta \) through \( \theta = \frac{\mu_1}{\mu_2} \), where \( \mu_1 = \min\{\mu_1, \mu_2\}, \mu_2 = \max\{\mu_1, \mu_2\} \). Clearly, \( \theta \in [0, 1] \). Therefore, for notational simplicity, we denote \( R(\theta, \phi) \) by \( R_\phi(\delta) \).

As in Section I, three naive estimators of \( H_M(\theta) \) based on MLEs and best equivariant estimators of \( \theta_1 \) and \( \theta_2 \) (or \( \ln \theta_1 \) and \( \ln \theta_2 \)) are

\[
d_{N_1}(X) = \ln Z_1 - \ln(\alpha), \quad d_{N_2}(X) = \ln Z_1 - \ln(\alpha + 1), \quad d_{N_3}(X) = \ln Z_1 - \psi(\alpha).
\]

Motivated by form of the estimators \( d_{N_1}(X), d_{N_2}(X) \) and \( d_{N_3}(X) \), we consider a subclass \( M_1 \) \( = \{ d_{\psi}(\cdot) : c \in \mathbb{R} \} \) of estimators, where \( d_{\psi}(\cdot) = \ln Z_1 - c, c \in \mathbb{R} \). We call class \( M_1 \), the class of linear, scale and permutation equivariant estimators. Clearly \( M_1 \subseteq M_2 \).
We will first provide a class of generalized Bayes estimators of $H_M(\theta)$ that is contained in class $M_1$. We obtain a result characterizing admissible and inadmissible estimators within the subclass $M_1 = \{d_\theta(\cdot) : d_\theta(X) = \ln Z_1 - c, c \in \mathbb{R}\}$ under the mean squared error criterion. Further, we also derive a sufficient condition for inadmissibility of an arbitrary scale and permutation equivariant estimator of $H_M(\theta)$ under the criterion of mean squared error. As a consequence of this general result, we obtain shrinkage estimators improving upon various naive estimators belonging to the class $M_1$.

Utilizing the arguments used in obtaining the class of generalized Bayes estimators $K_{GB}(H_S(\theta))$, we obtain the class $M_{GB} = \{d^{GB}_\beta : \beta > -\alpha\}$, with $d^{GB}_\beta(X) = \ln Z_1 - \psi(\alpha + \beta)$ as the class of generalized Bayes estimators of $H_M(\theta)$ under the squared error loss function (III.2), and improper priors (II.1). Remarkably, the naive estimator $d_{\psi(\alpha)}(X) = d^0_0(X) = \ln Z_1 - \psi(\alpha)$ (analogue of the best scale equivariant estimators of $\ln \theta_1$ and $\ln \theta_2$) is the generalized Bayes estimator with respect to Jeffrey’s non-informative prior density $P_0(\theta) = \frac{1}{\pi \Gamma(\alpha)}$, $\theta \in \Theta$.

The following lemma is useful in proving Theorem 3.1, reported in the sequel.

**Lemma 3.1:** Let $U_1 = \ln Z_1 - H_M(\theta)$. Then, for any $\theta \in \Theta$,
\[
E_\theta(U_1) = \int_0^\infty \ln(z)[1 - G_\alpha(\theta z)]g_\alpha(z)dz + \int_0^\infty \ln(z)[1 - G_\alpha(\frac{z}{\theta})]g_\alpha(z)dz, 0 < \theta \leq 1.
\]
where $G_\alpha(\cdot)$ and $g_\alpha(\cdot)$ denote the df and pdf, respectively, of $\Gamma(\alpha, 1)$ distribution.

**Proof:** Similar to the proof of Lemma 2.2.

The risk function of an estimator $\delta_c \in M_1$, under the squared error loss function (III.2), is given by
\[
R_\theta(\delta_c) = E \left[(\ln Z_1 - c - H_M(\theta))^2\right], 0 < \theta \leq 1, c \in \mathbb{R}.
\]
For any fixed $0 < \theta \leq 1$, the risk function in (III.5) is minimized at
\[
c \equiv c^*(\theta) = \inf_{0 < \theta \leq 1} c^*(\theta) = \inf_{0 < \theta \leq 1} E_\theta[(\ln Z_1 - H_M(\theta))] = E_\theta(U_1)
\]
\[
= \int_0^\infty \ln(z)[1 - G_\alpha(\theta z)]g_\alpha(z)dz + \int_0^\infty \ln(z)[1 - G_\alpha(\frac{z}{\theta})]g_\alpha(z)dz.
\]
Since $c^*(\theta)$ is a decreasing function of $\theta$,
\[
\left(\frac{d}{d\theta}c^*(\theta) = \frac{\Gamma(2\alpha)}{(1 + \theta)^{2\alpha + 1}}\ln \theta \leq 0, \forall \theta \in (0, 1]\right),
\]
we have,
\[
\inf_{0 < \theta \leq 1} c^*(\theta) = 2\int_0^\infty \ln(z)[1 - G_\alpha(\theta z)]g_\alpha(z)dz = c_3(\alpha),
\]
where $c_3(\alpha)$ is the generalized Bayes estimator $d_{\psi(\alpha)}(X) = d^0_0(X) = \ln Z_1 - \psi(\alpha)$.

**Theorem 3.1:** Let $c_3(\alpha)$ minimize the risk function of an estimator $\delta_c \in M_1$, under the mean squared error criterion, the estimators in the class $M_{1,M} = \{d_\theta \in M_1 : c \in [c_3(\alpha), c_1(\alpha)]\}$ are admissible within the class $M_1$, whereas the estimators in the class $M_{1,1} = \{d_\theta \in M_1 : c \in (-\infty, c_3(\alpha)) \cup (c_1(\alpha), \infty)\}$ are inadmissible for estimating $H_M(\theta)$. Moreover, for any $-\infty < b < c \leq c_3(\alpha)$ or $c_1(\alpha) \leq c < b < \infty$,
\[
R_\theta(d_c) < R_\theta(d_b), \forall 0 < \theta \leq 1.
\]
As a consequence of Theorem 3.1, we have the following corollary, addressing admissibility and inadmissibility of some naive estimators, including generalized Bayes estimators of $H_M(\theta)$, belonging to the class $M_{GB}$.

**Corollary 1.1:** (i) For estimating $H_M(\theta)$ under the mean squared error criterion, the naive estimator $d_{\psi(\alpha)}(X) = d^0_0(X) = \ln Z_1 - \psi(\alpha)$ is admissible within class $M_1$, whereas the estimators $d_{\ln(\alpha)}(X) = \ln Z_1 - \ln \alpha$ and $d_{\ln(\alpha+1)}(X) = \ln Z_1 - \ln(\alpha + 1)$ are inadmissible and are dominated by $d_{\psi(\alpha)}(X) = \ln Z_1 - \psi(\alpha)$.

(ii) For any $\alpha > 0$, define $\beta_1(\alpha) = \psi^{-1}(c_3(\alpha)) - \alpha$. Then
(a) generalized Bayes estimators $\{d_{\psi(\alpha + \beta)} : \beta_1(\alpha) \leq \beta \leq 0\}$ are admissible within the class $M_1$ of naive estimators.

(b) generalized Bayes estimators $\{d_{\psi(\alpha + \beta)} : \beta \in (-\alpha, \beta_1(\alpha) \cup (0, \infty)\}$ are inadmissible for estimating $H_M(\theta)$.

For any $\beta \in (-\alpha, \beta_1(\alpha))$, the generalized Bayes estimator $d_{\psi(\alpha + \beta)}(X) = \ln Z_1 - \psi(\alpha + \beta)$ is dominated by the naive estimator $d_{\ln(\alpha+1)}(X) = \ln Z_1 - \ln(\alpha + 1)$, and for any $\beta \in (0, \infty)$, the generalized Bayes estimator $d_{\psi(\alpha + \beta)}(X) = \ln Z_1 - \psi(\alpha)$.

The following Lemma will be useful in deriving the result stated in Theorem 3.2.

**Lemma 3.2:** (i) For fixed $v \in [1, \infty)$, the conditional pdf of $U_1 = \ln Z_1 - H_M(\theta)$, given $V = \frac{Z_1}{Z} = v$, is given by
\[
f_{2,\theta}(u|v) = \frac{\Gamma(2\alpha)}{(1 + \theta)^{2\alpha + 1}}\theta^\alpha e^{2\alpha u e^{-\theta} - (1 + v) e^u - \theta^\alpha u e^{\theta(1 + v)}},
\]
\[
\Gamma(2\alpha)\left[\frac{1}{(1 + \theta)^{2\alpha + 1}}\theta^\alpha e^{2\alpha u e^{-\theta} - (1 + v) e^u - \theta^\alpha u e^{\theta(1 + v)}}\right]^{-\alpha} < u < \infty, 0 < \theta \leq 1.
\]

(ii) For any fixed $\alpha > 0$,
\[
k_v(\theta) = \frac{(1 + v \theta)^{2\alpha} \ln (1 + \theta^\alpha) + (\theta + v)^{2\alpha} \ln (1 + v \theta)}{(1 + v \theta)^{2\alpha} + (\theta + v)^{2\alpha}}, \quad v \in [1, \infty), \quad 0 < \theta \leq 1.
\]
Then, for
\[
\min\left\{1 + \frac{1}{2\alpha}, 1 + \sqrt{3}\right\}, \quad \inf_{0 < \theta \leq 1} k_v(\theta) = \ln(1 + v).
\]
**Proof:** (i) Similar to the proof of Lemma 2.3.
(ii) Since \( \lim_{\theta \to 0} k_v(\theta) = \infty \), we have, \( \sup_{0 < \theta \leq 1} k_v(\theta) = \infty \). Note that 
\( k_v(1) = \ln(1+v), v \geq 1 \). Also, for \( 0 < \theta \leq 1 \) and \( v \geq 1 \), 
we have \( k_v(\theta) \geq k_v(1) \), if, and only if,

\[
(1+v\theta)^{2\alpha} \ln \left( \frac{1+v/\theta}{1+v} \right) \geq (\theta + v)^{2\alpha} \ln \left( \frac{1+v/\theta}{1+v} \right)
\]

\[
\iff \quad \frac{1+v\theta}{\theta + v} \geq \frac{\ln \left( \frac{1+v/\theta}{1+v} \right)}{\ln \left( \frac{1+v/\theta}{1+v} \right)}.
\]

Since, for \( 2\alpha \leq \frac{1}{1-v}, \frac{1+v/\theta}{\theta + v} \geq \frac{1}{1-v} \), to show that 
\( \inf_{0 < \theta \leq 1} k_v(\theta) = \ln(1+v) \), for \( 1 \leq v \leq \min \left\{ 1 + \frac{1}{2\alpha}, 1 + \sqrt{3} \right\} \),
it suffices to show that,

\[
\frac{1+v\theta}{\theta + v} \geq \frac{\ln \left( \frac{1+v/\theta}{1+v} \right)}{\ln \left( \frac{1+v/\theta}{1+v} \right)}, \tag{III.9}
\]

provided \( 1 \leq v \leq \min \left\{ 1 + \frac{1}{2\alpha}, 1 + \sqrt{3} \right\} \). Let \( \frac{1+v/\theta}{\theta + v} = x \) and \( \frac{1}{1-v} = \beta \), so that \( \beta > 0 \) and \( x \in \left( \frac{\beta}{\beta+1}, 1 \right) \). Proving inequality (3.8) is equivalent to showing

\[
\Psi(x) = -x^\beta \left[ \ln(\beta+1) + \ln \left( x - \frac{\beta}{\beta+1} \right) \right] - \beta \ln - \ln \left( \frac{\beta+1}{\beta} - x \right) + \ln x \geq 0, \tag{III.10}
\]

for all \( x \in \left( \frac{\beta}{\beta+1}, 1 \right) \) and \( \beta \geq \max \{2\alpha, \frac{1}{\sqrt{3}}\} \).

Note that, \( \lim_{x \to \frac{\beta}{\beta+1}} \Psi(x) = \infty \) and \( \lim_{x \to 1} \Psi(x) = 0 \). Thus, to prove (III.10), it is sufficient to show that \( \Psi'(x) \leq 0 \), 
\( \forall x \in \left( \frac{\beta}{\beta+1}, 1 \right) \), \( \beta \geq \max \{2\alpha, \frac{1}{\sqrt{3}}\} \). We have

\[
\Psi'(x) = x^{\beta-1}k_1(x), \quad x \in \left( \frac{\beta}{\beta+1}, 1 \right), \quad \beta \geq \max \left\{ 2\alpha, \frac{1}{\sqrt{3}} \right\},
\]

where

\[
k_1(x) = -x \ln(\beta+1) + \ln \left( x - \frac{\beta}{\beta+1} \right) - \ln \left( \frac{\beta+1}{\beta} - x \right) + \ln x \geq 0
\]

We have \( \lim_{x \to \frac{\beta}{\beta+1}} k_1(x) = -\infty \) and \( \lim_{x \to 1} k_1(x) = 0 \).

Thus, to show that \( \Psi'(x) \leq 0 \), \( \forall x \in \left( \frac{\beta}{\beta+1}, 1 \right) \), \( \beta \geq \max \{2\alpha, \frac{1}{\sqrt{3}}\} \), it suffices to show that \( k_1'(x) \geq 0 \), \( \forall x \in \left( \frac{\beta}{\beta+1}, 1 \right) \), \( \beta \geq \max \{2\alpha, \frac{1}{\sqrt{3}}\} \). We have

\[
k_1'(x) = (1-x) \left[ \frac{\beta}{x - \frac{\beta}{\beta+1}} - \frac{(\beta+1)^2}{\beta x^{\beta+1} \left( \frac{\beta+1}{\beta} - x \right)} \right],
\]

\( x \in \left( \frac{\beta}{\beta+1}, 1 \right), \beta \geq \max \left\{ 2\alpha, \frac{1}{\sqrt{3}} \right\} \).

To show that \( k_1'(x) \geq 0 \), \( \forall x \in \left( \frac{\beta}{\beta+1}, 1 \right) \), \( \beta \geq \max \{2\alpha, \frac{1}{\sqrt{3}}\} \), we will show that

\[
k_2(x) = (\beta+1) x^{\beta+1} \left( \frac{\beta+1}{\beta} - x \right) - (\beta+1) \left( x - \frac{\beta}{\beta+1} \right)
\]

\[
= (\beta+1) x^{\beta+1} - \beta x^{\beta+1} - (\beta+1) x + \beta \geq 0,
\]

\( \forall x \in \left( \frac{\beta}{\beta+1}, 1 \right), \beta \geq \max \{2\alpha, \frac{1}{\sqrt{3}}\} \).

We have, for \( x \in \left( \frac{\beta}{\beta+1}, 1 \right) \), \( \beta \geq \max \{2\alpha, \frac{1}{\sqrt{3}}\} \),

\[
k_2' \leq \lim_{x \to \frac{\beta}{\beta+1}} k_2'(x)
\]

\[
= \frac{\beta^{\beta+1}}{(2+\beta+1)^{\beta+1} \left( \beta+1 \right)} \left[ 3\beta + 1 - 2\beta^2 \left( 1 + \frac{1}{\beta} \right)^{\beta+3} \right]
\]

\[
\geq \frac{\beta^{\beta+1}}{(2+\beta+1)^{\beta+1} \left( \beta+1 \right)} \left[ 3\beta + 1 - 2\beta^2 \left( 1 + \frac{\beta+3}{2\beta} \right) \right]
\]

\[
= \frac{\beta^{\beta+1}}{(2+\beta+1)^{\beta+1} \left( \beta+1 \right)} \left[ 1 - 3\beta^2 \right] \leq 0,
\]

\( \Rightarrow \quad k_2(x) \geq \lim_{x \to \frac{\beta}{\beta+1}} k_2(x) = 0 \).

Hence, the assertion follows.

Let \( d_\Psi(V) = \ln Z_1 - \phi(V) \) be a scale and permutation equivariant estimator of \( H_M(\theta) \). For any fixed \( v \in [1, \infty) \), the conditional risk of \( d_\Psi(V) \), given \( V = v \), is obtained as

\[
R_1(\theta, \phi(v)) = \mathbb{E}_d \left[ (\ln Z_1 - \phi(V) - H_M(\theta))^2 \right] \big| V = v \big] = 0, \quad 0 < \theta \leq 1.
\]

For any fixed \( v \in [1, \infty) \) and \( \theta \in (0, 1) \), the choice of \( \phi(\cdot) \) that minimizes the conditional risk (3.10) is obtained as

\[
\phi_\theta(v) = \mathbb{E}_d \left[ (\ln Z_1 - H_M(\theta))^2 \right] \big| V = v \big) = \mathbb{E}_d \left[ U_1 \big| V = v \big).
\]

Using Lemma 3.2(i), we obtain \( \phi_\theta(v) = \psi(2\alpha) - k_\theta(\theta), 0 < \theta < 1, v \in [1, \infty) \), where, \( k_\theta(\theta) \) is given by (III.8).

Further, using Lemma 3.2(ii), we have, for \( 1 \leq v \leq \min \{1 + 1/(2\alpha), 1 + \sqrt{3} \} \),

\[
\sup_{0 < \theta \leq 1} \phi_\theta(v) = \psi(2\alpha) - \ln(1+v) = \phi_\theta(v), \quad \text{(say)}.
\]
The following theorem is an analogue of Theorem 2.2 and provides a sufficient condition for inadmissibility of an arbitrary scale and permutation equivariant estimator of $H_M(\theta)$, under the mean squared error criterion. Since the proof of the following theorem is similar to that of Theorem 2.2, it is omitted here.

**Theorem 3.2:** For estimating $H_M(\theta)$ under the mean squared error criterion, consider a scale and permutation equivariant estimator, $d_{\phi}(X) = \ln Z_1 - \phi(V)$, where, $V = \frac{Z_2}{Z_1}$ and $\phi(\cdot)$ is a real valued function defined on $[1, \infty)$. Let
\[
P_{\sum} \left[ V : V \leq \min \left\{ 1 + \frac{1}{2\alpha}, 1 + \sqrt{3} \right\} \text{ and } \phi(V) > \phi_*(V) \right] > 0, \quad \text{for some } \theta \in \Theta,
\]
where, $\phi_*(V)$ is defined by (III.12). Then, the estimator $d_{\phi}(\cdot)$ is dominated by $d_{\phi}^I(X) = \ln Z_1 - \phi^I(V)$, where,
\[
\phi^I(v) = \begin{cases} 
\phi_*(v), & \text{if } \phi(v) > \phi_*(v) \\
\text{and } v \leq \min \left\{ 1 + \frac{1}{2\alpha}, 1 + \sqrt{3} \right\}, & (\text{III.13})
\end{cases}
\]
otherwise.

For estimating $H_M(\theta)$ under the mean squared error criterion, we obtain the following result on inadmissibility of any naive estimator $d_{c}(X)$. As a consequence of Theorem 3.2, we also provide the shrinkage estimators dominating on $d_{c}(X)$.

**Corollary 2.1:** (i) For estimating $H_M(\theta)$ under the mean squared error criterion, any natural estimator $d_{c}(X) = \ln Z_1 - c$, with $c > \psi(2\alpha) - \ln(1 + \lambda)$; $\lambda = \min \left\{ 1 + \frac{1}{2\alpha}, 1 + \sqrt{3} \right\}$, is inadmissible and is dominated by the shrinkage estimator
\[
d_{c}^{S}(X) = \begin{cases} 
\ln(Z_1 + Z_2) - \psi(2\alpha), & \text{if } 1 \leq \frac{Z_2}{Z_1} \leq \min \left\{ 1 + \frac{1}{2\alpha}, 1 + \sqrt{3} \right\} \\
\ln Z_1 - c, & \text{otherwise.}
\end{cases} \quad (\text{III.14})
\]

(ii) For estimating $H_M(\theta)$ under the mean squared error criterion, the naive estimators $d_{ln}\alpha(X) = \ln Z_1 - \ln \alpha$ and $d_{ln(\alpha+1)}(X) = \ln Z_1 - \ln(\alpha + 1)$ are inadmissible and are dominated by the shrinkage estimators
\[
d_{ln}\alpha(\cdot) = \begin{cases} 
\ln(Z_1 + Z_2) - \psi(2\alpha), & \text{if } 1 \leq \frac{Z_2}{Z_1} \leq \min \left\{ 1 + \frac{1}{2\alpha}, 1 + \sqrt{3} \right\} \\
\ln Z_1 - \ln \alpha, & \text{otherwise,}
\end{cases} \quad (\text{III.15})
\]
and
\[
d_{ln(\alpha+1)}(\cdot) = \begin{cases} 
\ln(Z_1 + Z_2) - \psi(2\alpha), & \text{if } 1 \leq \frac{Z_2}{Z_1} \leq \min \left\{ 1 + \frac{1}{2\alpha}, 1 + \sqrt{3} \right\} \\
\ln Z_1 - \ln(\alpha + 1), & \text{otherwise.}
\end{cases} \quad (\text{III.16})
\]

Fig. 3. Mean squared error plots of the three natural estimators ($d_{ln}\alpha$, $d_{ln(\alpha+1)}$) and $d_{\psi(\alpha)}$ of $H_S(\theta)$.

**IV. SIMULATION**

In this section, we present results of a simulation study carried out to numerically assess the performances of various estimators of $H_S(\theta)$ and $H_M(\theta)$ in terms of the mean squared error (mse) and the absolute bias. For the numerical study,
we have taken random samples of size \( n = 3, 5, 10, 15, 20 \) etc. from relevant gamma distributions for different configurations of \( \alpha \). The risk (mse) and the absolute bias values of various proposed estimators are simulated based on 60,000 samples of size \( n \). In Figure 3–4 (Figure 7–8), we have plotted simulated

mises of the naive estimators \( \delta_{\ln(\alpha)} \), \( \delta_{\ln(\alpha+1)} \) and \( \delta_{\psi(\alpha)} \) of \( H_S(\theta) \).

In Figures 9–10 and 11–12, we have plotted simulated absolute bias values of the three naive estimators of \( H_S(\theta) \) and \( H_M(\theta) \), respectively. In Figure 5–6, we have plotted the mse values of the estimators \( \delta_{\ln(\alpha)} \) and \( \delta_{\ln(\alpha+1)} \) and their improved versions \( \delta_{\ln(\alpha)}^{(S)} \) and \( \delta_{\ln(\alpha+1)}^{(S)} \) respectively. In Figure 13–14, we have
plotted mse values of estimators $d_{\ln\alpha}$ and $d_{\ln(\alpha+1)}$ and their improved versions $d_{\ln(\alpha)}^{(S)}$ and $d_{\ln(\alpha+1)}^{(S)}$, respectively. Based on these plotted graphs, the following conclusions are obvious:

(i) For estimating $H_S(\theta)$, under the mean squared error criterion, among the three naive estimators ($d_{\ln\alpha}$, $d_{\ln(\alpha+1)}$ and $d_{\psi(\alpha)}$), the estimator $d_{\psi(\alpha)}$ performs better except for smaller values of $\mu$ (see Figure 3-4).

(ii) For estimating $H_S(\theta)$, mean squared errors (mses) of the estimators $d_{\ln\alpha}^{(S)}$ and $d_{\ln(\alpha+1)}^{(S)}$ are nowhere larger than mses of $d_{\ln\alpha}$ and $d_{\ln(\alpha+1)}$ respectively (see Figure 5-6).

(iii) For estimating $H_M(\theta)$, under the mean squared error criterion, the estimator $d_{\psi(\alpha)}$ uniformly performs better.
among all the three naive estimators \( d_{\ln \alpha}, d_{\ln(\alpha+1)} \) and \( d_{\psi(\alpha)} \) of \( H_M(\theta) \).

(iv) For estimating \( H_S(\theta) \), the estimator \( d_{\ln(\alpha+1)} \) has smaller absolute bias than the other two naive estimators (\( d_{\ln(\alpha)} \) and \( d_{\psi(\alpha)} \)) for smaller values of \( \mu \). For large values of \( \mu \), \( d_{\psi(\alpha)} \) seems to be a good choice, in terms of the absolute bias (see Figure 9-10.).

(v) For estimating \( H_M(\theta) \) in terms of absolute bias, the naive estimators \( d_{\ln \alpha} \) and \( d_{\ln(\alpha+1)} \) are uniformly dominated by the naive estimator \( d_{\psi(\alpha)} \). Moreover, \( d_{\ln \alpha} \) dominates \( d_{\ln(\alpha+1)} \) in terms of the absolute bias (see Figure 11-12.).

(vi) For estimating \( H_M(\theta) \), the mean squared error of the estimators \( d_{\ln \alpha}^{(S)} \) and \( d_{\ln(\alpha+1)}^{(S)} \) are nowhere
larger than mses of $d_{ln\alpha}$ and $d_{ln(\alpha+1)}$ respectively (see Figure 13-14).

(vii) Based on the simulation study, we observe that the estimator $\delta_{\psi(\alpha)}$ performs better (in terms of the mse and the absolute bias), for estimating the selected Shannon entropy $H_S(\theta)$.

V. REAL DATA EXAMPLE

Example 1: Here we consider a real data set provided in [29]. The data set reflects the failure times (in hours) of the air conditioning systems of Boeing 720 jet planes “7913” and “7914”. The data set contains sample
observations of equal size $n = 24$ and is given as below:

**Plane 7913**: 97, 51, 11, 4, 141, 18, 142, 68, 77, 80, 1, 16, 106, 206, 82, 54, 31, 216, 46, 111, 39, 63, 18, 191.

**Plane 7914**: 50, 44, 102, 72, 22, 39, 3, 15, 197, 188, 79, 88, 46, 5, 5, 36, 22, 139, 210, 97, 30, 23, 13, 14.

For testing the goodness of fit of gamma distribution, we analysed the given data in R software using the Kolmogorov-Smirnov goodness of fit test. We observe that at 5% level of significance, one cannot reject the hypothesis that the data for Planes 7913 and 7914 are from $\text{Gamma}(1, 68)$ and $\text{Gamma}(1, 61)$ distributions, respectively. Based on the...
For estimating $H_M(\theta)$ (the Shannon entropy of the selected plane “7914”) under the squared error loss function, various estimates are provided in the following table:

**Table 5.2: Various estimates of $H_M(\theta)$**

| $d_{ln(1)}$  | $d_{ln(1)+1}$ | $d_{\psi(1)}$ | $d_{\phi(1)}$ | $d_{X(1)}$ | $d_{X(1)+1}$ |
|-------------|---------------|---------------|---------------|-------------|--------------|
| 4.160834    | 4.120012      | 4.181812      | 4.27313       | 4.27313     |

It is recommended to use $\delta_\psi$ ($d_\psi$) for estimating the selected Shannon entropy $H_S(\theta)(H_M(\theta))$ of the failure times of the air conditioning systems of Boeing jet planes.

**Example 2:** Consider the data given in Table 5.3, which has been taken from [30] [page-208,4.19]. The data given in the following table represent failure times, in minutes, for two types of electrical insulations (Type A and Type B) in an experiment in which the insulation was subjected to a continuously increasing voltage stress.

**Table 5.3: Failure times (in minutes) for two types of electric insulation**

| Type   | $X_1$ | $X_2$ |
|--------|-------|-------|
| Type A | 219.3 | 79.4  |
|        | 121.9 | 40.5  |
| Type B | 21.8  | 70.7  |
|        | 12.3  | 95.5  |

We analyzed the given data set in R software. Using the Kolmogorov-Smirnov goodness of fit test, it has been verified that the data is coming from two-parameter gamma distribution with the same shape parameter. It is observed, that at 5% level of significance, one can’t reject the hypothesis that data for the two types of electrical insulations are from $Gamma(2,42)$ and $Gamma(2,32)$ distributions, respectively.

For the given data set, we computed, $X_1 = 1010.7$, $X_2 = 807.1$. Therefore, Type A insulation is selected as the “worse” one and Type B as the “better” insulation. Our goal is to estimate the Shannon entropy of the selected insulation. For this purpose, the values of the statistics $Z_1$, $Z_2$, $T$ and $V$ are computed as $Z_1 = 807.1$, $Z_2 = 1010.7$, $T = 0.7985555$ and $V = 1.252261$. For estimating $H_S(\theta)$ (the Shannon entropy of the selected insulation “Type A”) under the squared error loss function, various estimates are provided in the following table:

**Table 5.4: Estimates of $H_S(\theta)$**

| $\delta_{ln(1)}$ | $\delta_{ln(1)+1}$ | $\delta_{\psi}$ | $\delta_{\phi}$ | $\delta_{X(1)}$ | $\delta_{X(1)+1}$ |
|-----------------|--------------------|-----------------|-----------------|-----------------|------------------|
| 3.740           | 3.699              | 3.761           | 4.348           | 4.348           |

For estimating $H_M(\theta)$ (the Shannon entropy of the selected insulation “Type B”) under the squared error loss function, various estimates are provided in the following table:

**Table 5.5: Estimates of $H_M(\theta)$**

| $d_{ln(1)}$ | $d_{ln(1)+1}$ | $d_{\psi}$ | $d_{\phi}$ | $d_{X(1)}$ | $d_{X(1)+1}$ |
|-------------|---------------|-------------|-------------|-------------|--------------|
| 3.515       | 3.474         | 3.536       | 3.515       | 3.474       |

**VI. Final Remarks**

The Shannon entropy is a useful measure of uncertainty that has applications in a variety of fields including wireless
communication, weather forecasting, economic modelling, molecular biology, and so on. Estimation of the selected Shannon entropy is an important practical problem in these areas. In the present article, we have focussed on efficient estimation of the Shannon entropy of the selected gamma population and proposed several estimators. We have derived various admissibility results for a class of naive estimators of the selected entropy and also obtained shrinkage estimators dominating upon various naive estimators. Although, our simulation study suggests that the generalized Bayes estimator with respect to the Jeffreys non informative prior is admissible within the class of scale and permutation equivariant estimators of \( H_S(\theta) \) (or \( H_M(\theta) \)) under the mean squared error criterion, we have not been able to prove it. From our analysis, we believe that there does not exist any unbiased estimator of the selected Shannon entropy \( H_S(\theta) \) (or \( H_M(\theta) \)), but, unfortunately, we have not been able to prove this result too. It will also be interesting to investigate whether the results obtained in this paper can be adapted to \( k \) (\( \geq 2 \)) populations. These are some interesting questions for future research.

**ACKNOWLEDGMENT**

The authors would like to thank the two anonymous referees and the Editor of the journal for their comments and suggestions that have greatly improved this manuscript.

**REFERENCES**

[1] R. E. Bechhofer, “A single-sample multiple decision procedure for ranking means of normal populations with known variances,” *Ann. Math. Statist.*, vol. 25, no. 1, pp. 16–39, Mar. 1954.

[2] S. Gupta, “On a decision rule for a problem in ranking means,” Ph.D. thesis, Inst. Statist., Univ. North Carolina Chapel Hill, Chapel Hill, NC, USA, 1956.

[3] K. Sarkadi, “Estimation after selection,” *Studia Scientarium Mathematicarum Hungarica*, vol. 2, no. 1, pp. 341–350, 1967.

[4] J. Putter and D. Rubinstein, “On estimating the mean of a selected population,” Dept. Statist., Univ. Wisconsin Madison, Madison, WI, USA, Tech. Rep. 165, 1968.

[5] H.-K. Hsieh, “On estimating the mean of the selected population with unknown variance,” *Commun. Statist., Theory Methods*, vol. 10, no. 18, pp. 1869–1878, Jan. 1981.

[6] A. Cohen and H. B. Sackrowitz “Estimating the mean of the selected population,” in *Statistical Decision Theory & Related Topics-III*, vol. 1, S. S. Gupta and J. O. Berger, Eds. New York, NY, USA: Academic, pp. 243–270.

[7] H. Sackrowitz and E. Samuel-Cahn, “Evaluating the chosen population: A Bayes and minimax approach,” *Lect. Notes-Monograph Ser.*, vol. 8, pp. 386–399, Jan. 1986.

[8] P. Vellaisamy, “Inadmissibility results for the selected scale parameters,” *Ann. Statist.*, vol. 20, no. 4, pp. 2183–2191, Dec. 1992.

[9] J. T. Hwang, “Empirical Bayes estimation for the means of the selected populations,” *Sankhya, Indian J. Statist.*, vol. 55, no. 2, pp. 285–304, 1993.

[10] P. Vellaisamy, “On UMVU estimation following selection,” *Commun. Statist., Theory Methods*, vol. 25, no. 4, pp. 1031–1043, 1993.

[11] N. Misra, “Estimation of the average worth of the selected subset of gamma populations,” *Sankhya, Indian J. Statist.*, vol. 56, no. 3, pp. 344–355, 1994.

[12] M. N. Qomi, N. Nematollahi, and A. Parsian, “On admissibility and inadmissibility of estimators after selection under reflected gamma loss function,” *Hacetette J. Math. Statist.*, vol. 44, no. 5, pp. 1109–1124, 2015.

[13] M. Arshad, N. Misra, and P. Vellaisamy, “Estimation after selection from gamma populations with unequal known shape parameters,” *J. Stat. Theory Pract.*, vol. 9, no. 2, pp. 395–418, Apr. 2015.

[14] T. Routtenberg and L. Tong, “Estimation after parameter selection: Performance analysis and estimation methods,” *IEEE Trans. Signal Process.*, vol. 64, no. 20, pp. 5268–5281, Oct. 2016.

[15] M. L. Menéndez, L. Pardo, C. Tsairidis, and K. Zografos, “Selection of the best population: An information theoretic approach,” *Metrika*, vol. 58, no. 2, pp. 117–147, Sep. 2003.

[16] M. L. Menéndez, L. Pardo, and K. Zografos, “Ordering and selecting extreme populations by means of entropies and divergences,” *J. Comput. Appl. Math.*, vol. 232, no. 2, pp. 335–350, Oct. 2009.

[17] C. E. Shannon, “A mathematical theory of communication,” *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, Jul. 1948.

[18] M. L. Eaton, “Some optimum properties of ranking procedures,” *Ann. Math. Statist.*, vol. 38, no. 1, pp. 124–137, Feb. 1967.

[19] N. Misra and I. D. Dhariyal, “Non-minimality of natural decision rules under heteroscedasticity,” *Statist. Risk Model.*, vol. 12, no. 1, pp. 79–98, Jan. 1994.

[20] R. R. Bahadur and L. A. Goodman, “Impartial decision rules and sufficient statistics,” *Ann. Math. Statist.*, vol. 23, no. 4, pp. 553–562, Dec. 1952.

[21] N. Misra and M. Arshad, “Selecting the best of two gamma populations having unequal shape parameters,” *Stat. Methodol.*, vol. 18, pp. 41–63, May 2014.

[22] C. Stein, “Inadmissibility of the usual estimator for the mean of a multivariate normal distribution,” in *Proc. 3rd Berkeley Symp. Math. Statist. Probab.*, vol. I, Berkeley, CA, USA: Univ. of California Press, Berkeley, 1956, pp. 197–206.

[23] C. Stein, “Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean,” *Ann. Inst. Stat. Math.*, vol. 16, no. 1, pp. 155–160, Dec. 1964.

[24] C. Stein, “Estimation of the mean of a multivariate normal distribution,” in *Proc. Prague Symp. Asymptotic Statist.*, vol. II, Prague, Czech Republic: Charles Univ., 1974, pp. 34–381.

[25] W. James and C. Stein, “Estimation with quadratic loss,” in *Proc. 4th Berkeley Symp. Math. Statist.*, vol. I, Berkeley, CA, USA: Univ. of California Press, Berkeley, 1961, pp. 361–379.

[26] J. F. Bremer and J. V. Zidek, “Improving on equivariant estimators,” *Ann. Statist.*, vol. 2, no. 1, pp. 21–38, Jan. 1974.

[27] Masihuddin and N. Misra, “Equivariant estimation following selection from two normal populations having common unknown variance,” *Statistics*, vol. 55, no. 6, pp. 1407–1438, Nov. 2021.

[28] M. Abramovitz and I. A. Stegun, *Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables*, vol. 55, Washington, DC, USA: U.S. Government Printing Office, 1970.

[29] F. Proschan, “Theoretical explanation of observed decreasing failure rate,” *Technometrics*, vol. 5, no. 3, pp. 375–383, Aug. 1963.

[30] J. F. Lawless, *Statistical Models and Methods for Lifetime Data*, Hoboken, NJ, USA: Wiley, 2011.