Ann. Inst. Statist. Math.
Vol. 44, No. 3, 551–562 (1992)

AVERAGE WORTH AND SIMULTANEOUS ESTIMATION OF THE SELECTED SUBSET

P. VELLAISAMY

Department of Mathematics, Indian Institute of Technology, Powai, Bombay - 400 076, India

(Received February 6, 1990; revised March 7, 1991)

Abstract. Suppose a subset of populations is selected from the given k gamma $G(\theta_i, p)$ $(i = 1, 2, \dots, k)$ populations, using Gupta's rule (1963, *Ann. Inst. Statist. Math.*, **14**, 199–216). The problem of estimating the average worth of the selected subset is first considered. The natural estimator is shown to be positively biased and the UMVUE is obtained using Robbins' UV method of estimation (1988, *Statistical Decision Theory and Related Topics IV, Vol.* 1 (eds. S. S. Gupta and J. O. Berger), 265–270, Springer, New York). A class of estimators that dominate the natural estimator for an arbitrary k is derived. Similar results are observed for the simultaneous estimation of the selected subset.

Key words and phrases: Gamma populations, subset selection, estimation after subset selection, average worth, natural estimator, UMVUE, inadmissibility, simultaneous estimation after selection.

1. Introduction

Let $\pi_1, \pi_2, \dots, \pi_k$ denote k independent $G(\theta_i, p)$ populations with densities

$$f_i(x \mid \theta_i, p) = \frac{\theta_i^{-p}}{\Gamma(p)} e^{-x/\theta_i} x^{p-1}, \qquad i = 1, 2, \dots, k, \tag{1.1}$$

where the θ_i's are the unknown scale parameters, and p is the common known shape parameter. Let $\theta_{[1]} \geq \theta_{[2]} \geq \cdots \geq \theta_{[k]}$ represent the ordered parameters (use arbitrary ordering if some of the θ_i's are equal). The population having the largest scale parameter $\theta_{[1]}$ is called the best population.

Suppose Y_i $(i = 1, 2, \dots, k)$ denote the mean of a random sample of (equal) size n from the i-th population. To select a nonempty subset of the given populations, containing the best population, Gupta (1963) proposed the following rule R:

Select π_i iff

$$Y_i \geq cY_{(1)}, \tag{1.2}$$

where $0 < c < 1$, and $Y_{(1)}$ is the largest of the Y_i's. A selection which contains the best population is called the correct selection. The constant c is the largest number satisfying the basic probability requirement

$$\min_{\underline{\boldsymbol{\theta}}} P_{\underline{\boldsymbol{\theta}}}(\text{CS} \mid R) = P^*, \tag{1.3}$$

where P^* is the preassigned probability, "CS" stands for the correct selection, and $\underline{\boldsymbol{\theta}} = (\theta_1, \theta_2, \ldots, \theta_k)$. This value of c is given by (with $p' = np$)

$$\int_0^\infty G_{p'}^{k-1}\left(\frac{z}{c}\right) g_{p'}(z)dz = P^*, \tag{1.4}$$

where here (and henceforth) $G_p(x)$ and $g_p(x)$ denote the distribution function and the density of a gamma $G(1, p)$ variate.

Since $Y_i \sim G(\theta_i/n, np)$, we consider without loss of generality the case $n = 1$. Let $(X_1, X_2, \ldots, X_k)$ denote a random sample, where X_i is from π_i. Also, $X_{(1)} \geq X_{(2)} \geq \cdots \geq X_{(k)}$ denote the ordered values of the X_i's. Throughout the paper, $\underline{\boldsymbol{X}}$ and $\underline{\boldsymbol{\theta}}$ represent, unless stated otherwise, $(X_1, X_2, \ldots, X_k)$ and $(\theta_1, \theta_2, \ldots, \theta_k)$ respectively.

Suppose a subset S (of random size) is selected using the aforementioned Gupta's rule. Then the two problems of interest of the selected subset are the estimation of the average worth W and the simultaneous estimation of Q, defined by

$$W = p\sum_{r=1}^{k}\left[\left(\sum_{j=1}^{r}\theta_{(j)}/r\right)I(X_{(r+1)} < cX_{(1)} \leq X_{(r)})\right] \tag{1.5}$$

and

$$Q = p\sum_{r=1}^{k}\theta^{(r)}I(X_{(r+1)} < cX_{(1)} \leq X_{(r)}), \tag{1.6}$$

where $\theta_{(i)}$ is the parameter associated with $X_{(i)}$, $\theta^{(r)} = (\theta_{(1)}, \theta_{(2)}, \ldots, \theta_{(r)})$ and I is defined in (2.1). Observe that both W and Q are random, and

$$\theta_{(i)} = \theta_j \quad \text{if} \quad X_j = X_{(i)}, \quad j = 1, 2, \ldots, k.$$

We mention here that when $c = 1$, the rule R selects the population corresponding to $X_{(1)}$ and W in this case is the mean of the selected population. This problem has, of late, been receiving a lot of attention. For some recent papers, see Cohen and Sackrowitz (1982, 1989), Sackrowitz and Samuel-Cahn (1984, 1987) and Venter (1988).

Estimation after subset selection has been initiated and studied by Jeyarathnam and Panchapakesan (1984, 1986) for normal and exponential populations. In this paper, the estimation of the selected subset from gamma populations is considered. In Section 2, the natural estimator is shown to be positively biased and the uniformly minimum variance unbiased estimator (UMVUE) is derived for an arbitrary k. Also, a class of estimators that dominate the natural estimator, for the squared error loss, is obtained. Similar results are observed for the simultaneous estimation of Q in Section 3.

2. Estimation of average worth W

2.1 *Natural estimator*

Let $I(A)$ denote the indicator function of an event A, that is,

$$I(A) = \begin{cases} 1 & \text{if } \underline{\boldsymbol{X}} \in A \\ 0 & \text{elsewhere.} \end{cases} \tag{2.1}$$

(This notation will be followed throughout the paper.)

A natural estimator of W is

$$T = \sum_{r=1}^{k} \left[\left(\sum_{j=1}^{r} X_{(j)}/r \right) \right] I(X_{(r+1)} < cX_{(1)} \le X_{(r)}). \tag{2.2}$$

When, for example, $k = 2$, W and T reduce to

$$p^{-1}W = \begin{cases} \theta_1 & \text{if } \dfrac{X_1}{X_2} > \dfrac{1}{c} \\ \dfrac{\theta_1 + \theta_2}{2} & \text{if } c \le \dfrac{X_1}{X_2} \le \dfrac{1}{c} \\ \theta_2 & \text{if } 0 < \dfrac{X_1}{X_2} < c \end{cases}$$

and

$$T = \begin{cases} X_1 & \text{if } \dfrac{X_1}{X_2} > c \\ \dfrac{X_1 + X_2}{2} & \text{if } c \le \dfrac{X_1}{X_2} \le \dfrac{1}{c} \\ X_2 & \text{if } 0 < \dfrac{X_1}{X_2} < c. \end{cases}$$

The bias $B_{\underline{\theta}}(T)$ of T as an estimator of W is defined by

$$B_{\underline{\theta}}(T) = E_{\underline{\theta}}(T - W).$$

We now have the following lemma.

LEMMA 2.1. *The bias of the natural estimator T, when $k = 2$, is*

$$B_{\underline{\theta}}(T) = \frac{(\theta_1 + \theta_2)}{2B(p,p)} \left\{ \frac{(c\theta)^p}{(1 + c\theta)^{2p}} + \frac{(c/\theta)^p}{(1 + c/\theta)^{2p}} \right\}, \tag{2.3}$$

where $\theta = \theta_1/\theta_2$, and $B(\cdot,\cdot)$ denotes the usual beta function.

PROOF. Write

$$D(\theta_1, \theta_2) = E_{\underline{\theta}}(X_1 - p\theta_1) I\left(\frac{X_1}{X_2} > \frac{1}{c}\right) + \frac{1}{2} E_{\underline{\theta}}(X_1 - p\theta_1) I\left(c \le \frac{X_1}{X_2} \le \frac{1}{c}\right).$$

Let Z_1 and Z_2 be i.i.d. $G(1,p)$ variables, and $Z = Z_1/Z_2$. Then

(2.4) $$D(\theta_1,\theta_2)$$
$$= \theta_1\left\{E(Z_1-p)I\left(Z > \frac{1}{c\theta}\right) + \frac{1}{2}E(Z_1-p)I\left(\frac{c}{\theta} \le Z \le \frac{1}{c\theta}\right)\right\}$$
$$= \frac{\theta_1}{2}\left\{E(p-Z_1)I\left(Z < \frac{1}{c\theta}\right) + E(p-Z_1)I\left(Z < \frac{c}{\theta}\right)\right\}.$$

It is easy to note that by interchanging the role of (X_1,θ_1) and (X_2,θ_2), we obtain $D(\theta_2,\theta_1)$ so that

(2.5) $$B_{\underline{\theta}}(T) = D(\theta_1,\theta_2) + D(\theta_2,\theta_1).$$

We now proceed to evaluate $D(\theta_1,\theta_2)$. For $t > 0$,

$$E(p-Z_1)I(Z<t) = \int_0^\infty\left\{\int_0^{tz_2}(p-z_1)g_p(z_1)dz_1\right\}g_p(z_2)dz_2$$
$$= p\int_0^\infty\{G_p(tz_2) - G_{p+1}(tz_2)\}g_p(z_2)dz_2.$$

Using the fact
$$G_p(x) - G_{p+1}(x) = g_{p+1}(x)$$
for $x > 0$, we obtain

(2.6) $$E(p-Z_1)I(Z<t) = \frac{1}{B(p,p)}\frac{t^p}{(1+t)^{2p}}.$$

The lemma now follows from (2.4)–(2.6).

It is now clear from (2.3) that the bias of the natural estimator, $B_{\underline{\theta}}(T) > 0$ for all $\underline{\theta} = (\theta_1,\theta_2)$. In fact, the bias becomes infinite as, for example, θ_2 tends to infinity along the line $\theta_2 = b\theta_1$, $b > 0$.

Remark 2.1. The above lemma for the case $p = n$, a positive integer, is proved in Jeyarathnam and Panchapakesan (1986). Their proof is rather involved, and uses much notation and a large number of results. In contrast, the proof of Lemma 2.1 given above, which applies to any $p > 0$, is simple and straightforward.

2.2 *The UMVUE of W*

We now derive the UMVUE of W using the UV method of estimation (Robbins (1988)).

Let X be a $G(\alpha,p)$ variate with p known. Then for any given $u(x)$ for which
$$v(x) = x^{1-p}\int_0^x u(t)t^{p-1}dt$$

exists, we have

$$E_\alpha(v(X)) = \alpha E_\alpha(u(X)) \qquad \text{for all } \alpha.$$

The following lemma is a straightforward generalization of the above fact to k independent gamma variables.

LEMMA 2.2. *Suppose $X_1, X_2, \ldots, X_k$ are k independent random variables with densities defined in (1.1). Let $u(\underline{\boldsymbol{x}})$ be any real-valued function defined on R^k such that*

(i) $E_{\underline{\boldsymbol{\theta}}}|u(\underline{\boldsymbol{X}})| < \infty$

(ii) *the indefinite integral*

$$h_i(\underline{\boldsymbol{x}}) = \int_0^{x_i} u(x_1, \ldots, x_{i-1}, t, x_{i+1}, \ldots, x_k) t^{p-1} dt$$

exists for all $x_i \in R^1$. Then

$$v(\underline{\boldsymbol{x}}) = x_i^{1-p} h_i(\underline{\boldsymbol{x}}) \tag{2.7}$$

satisfies the condition

$$E_{\underline{\boldsymbol{\theta}}}(v(\underline{\boldsymbol{X}})) = \theta_i E_{\underline{\boldsymbol{\theta}}}(u(\underline{\boldsymbol{X}})) \qquad \textit{for all } \underline{\boldsymbol{\theta}}. \tag{2.8}$$

The above lemma, which will be exploited throughout the paper, is very useful in obtaining unbiased estimators of certain functions of $\underline{\boldsymbol{X}}$ and $\underline{\boldsymbol{\theta}}$.

We first find an unbiased estimator for the random quantity $p\theta_{(1)} I(X_{(r+1)} < cX_{(1)} \leq X_{(r)})$, $1 \leq r \leq k$. Let

$$q_{1,r}(\underline{\boldsymbol{\theta}}) = pE_{\underline{\boldsymbol{\theta}}}[\theta_{(1)} I(X_{(r+1)} < cX_{(1)} \leq X_{(r)})],$$

where $1 \leq r \leq k$. Observe that

$$\begin{aligned} q_{1,r}(\underline{\boldsymbol{\theta}}) &= \sum_{i=1}^{k} p\theta_i P_{\underline{\boldsymbol{\theta}}}(X_{(r)i} < cX_i \leq X_{(r-1)i}; X_i > X_{(1)i}) \\ &= \sum_{i=1}^{k} q_{1,r:i}(\underline{\boldsymbol{\theta}}) \qquad \text{(say)}, \end{aligned}$$

where $\infty \equiv X_{(0)i} > X_{(1)i} \geq X_{(2)i} \geq \cdots \geq X_{(k-1)i} > X_{(k)i} \equiv 0$ denote the order statistics of $X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_k$.

Define now

$$u_{1,r:1}(\underline{\boldsymbol{X}}) = \begin{cases} p & \text{if } \underline{\boldsymbol{X}} \in A = \{\underline{\boldsymbol{X}} : X_{(r)1} < cX_1 < X_{(r-1)1}; X_1 > X_{(1)1}\} \\ 0 & \text{otherwise.} \end{cases}$$

Let

$$l_1(\underline{\boldsymbol{X}}) = \min\left\{X_1, \frac{1}{c} X_{(r-1)1}\right\}$$

and

$$l_2(\underline{\boldsymbol{X}}) = \max\left\{X_{(1)1}, \frac{1}{c}X_{(r)1}\right\}.$$

Then the corresponding $v_{1,r:1}(\underline{\boldsymbol{X}})$ given by (2.7) is

$$\begin{aligned} v_{1,r:1}(\underline{\boldsymbol{X}}) &= pX_1^{1-p}\left(\int_{l_2(\underline{\boldsymbol{X}})}^{l_1(\underline{\boldsymbol{X}})} t^{p-1}dt\right) I(l_1(\underline{\boldsymbol{X}}) > l_2(\underline{\boldsymbol{X}})) \\ &= X_1^{1-p}\left[\left\{\min\left(X_1, \frac{1}{c}X_{(r-1)1}\right)\right\}^p - \left\{\max\left(X_{(1)1}, \frac{1}{c}X_{(r)1}\right)\right\}^p\right] \\ &\quad \times I\left(\min\left(X_1, \frac{1}{c}X_{(r-1)1}\right) > \max\left(X_{(1)1}, \frac{1}{c}X_{(r)1}\right)\right). \end{aligned}$$

An application of Lemma 2.2 shows that $v_{1,r:1}(\underline{\boldsymbol{X}})$ is an unbiased estimator of $\theta_1 E_{\underline{\boldsymbol{\theta}}}(u_{1,r:1}(\underline{\boldsymbol{X}})) = q_{1,r:1}(\underline{\boldsymbol{\theta}})$, for all $\underline{\boldsymbol{\theta}}$.

For $2 \le j \le k$, let $v_{1,r:j}(\underline{\boldsymbol{X}})$ be defined by

$$v_{1,r:j}(\underline{\boldsymbol{X}}) = v_{1,r:1}(X_j, X_2, \ldots, X_{j-1}, X_1, X_{j+1}, \ldots, X_k),$$

that is, $v_{1,r:j}(\underline{\boldsymbol{X}})$ is obtained from $v_{1,r:1}(\underline{\boldsymbol{X}})$ by interchanging X_1 and X_j. Then it is easy to note that

$$E_{\underline{\boldsymbol{\theta}}}(v_{1,r:j}(\underline{\boldsymbol{X}})) = q_{1,r:j}(\underline{\boldsymbol{\theta}}) \qquad \text{for all } \underline{\boldsymbol{\theta}}.$$

Therefore, an unbiased estimator of $q_{1,r}(\underline{\boldsymbol{\theta}})$ is

$$\begin{aligned} (2.9)\quad h_{1,r}(\underline{\boldsymbol{X}}) &= \sum_{j=1}^{k} v_{1,r:j}(\underline{\boldsymbol{X}}) \\ &= X_{(1)}^{1-p}\left[\left\{\min\left(X_{(1)}, \frac{1}{c}X_{(r)}\right)\right\}^p - \left\{\max\left(X_{(2)}, \frac{1}{c}X_{(r+1)}\right)\right\}^p\right] \\ &\quad \times I\left(\min\left(X_{(1)}, \frac{1}{c}X_{(r)}\right) > \max\left(X_{(2)}, \frac{1}{c}X_{(r+1)}\right)\right). \end{aligned}$$

Consider next the unbiased estimation of $p\theta_{(j)} I(X_{(j+1)} < cX_{(1)} \le X_{(j)})$, where $2 \le j \le k$. Let, as before,

$$q_{j,j:1}(\underline{\boldsymbol{\theta}}) = p\theta_1 P_{\underline{\boldsymbol{\theta}}}(X_{(j)1} < cX_{(1)1} < X_1 \le X_{(j-1)1}).$$

Then an unbiased estimator of $q_{j,j:1}(\underline{\boldsymbol{\theta}})$ can be seen to be

$$\begin{aligned} h_{j,j:1}(\underline{\boldsymbol{X}}) &= X_1^{1-p}[X_1^p - (cX_{(1)1})^p] I(X_{(j)1} < cX_{(1)1} < X_1 \le X_{(j-1)1}) \\ &\quad + X_1^{1-p}[X_{(j-1)1}^p - (cX_{(1)1})^p] I(X_{(j)1} < cX_{(1)1} < X_{(j-1)1} < X_1). \end{aligned}$$

Following the arguments just preceding (2.9) and rearranging the terms, we get

$$\begin{aligned} (2.10)\quad h_{j,j}(\underline{\boldsymbol{X}}) &= X_{(1)}^{1-p}[X_{(j)}^p - (cX_{(2)})^p] I(X_{(j+1)} < cX_{(2)} \le X_{(j)}) \\ &\quad + \left(\sum_{s=2}^{j} X_{(s)}^{1-p}\right)[X_{(j)}^p - (cX_{(1)})^p] I(X_{(j+1)} < cX_{(1)} \le X_{(j)}) \end{aligned}$$

as an unbiased estimator of

$$q_{j,j}(\underline{\theta}) = E_{\underline{\theta}}[p\theta_{(j)} I(X_{(j+1)} < cX_{(1)} \le X_{(j)})].$$

Similarly, it can be shown, after some lengthy calculations, that

$$(2.11)\quad h_{j,r}(\underline{\boldsymbol{X}}) = X_{(1)}^{1-p}[X_{(j)}^p - X_{(j+1)}^p] I(X_{(r+1)} < cX_{(2)} \le X_{(r)}) + \left(\sum_{s=2}^{j} X_{(s)}^{1-p}\right)[X_{(j)}^p - X_{(j+1)}^p] I(X_{(r+1)} < cX_{(1)} \le X_{(r)})$$

is an unbiased estimator of

$$q_{j,r}(\underline{\theta}) = E_{\underline{\theta}}[p\theta_{(j)} I(X_{(r+1)} < cX_{(1)} \le X_{(r)})],$$

where $j+1 \le r \le k$, and $2 \le j \le k$.

Since $(X_1, X_2, \ldots, X_k)$ is complete sufficient, the following theorem is now established.

THEOREM 2.1. *The UMVUE of W is given by*

$$(2.12)\qquad H_k(\underline{\boldsymbol{X}}) = \sum_{r=1}^{k}\left[\sum_{j=1}^{r} h_{j,r}(\underline{\boldsymbol{X}})/r\right],$$

where $h_{j,r}(\underline{\boldsymbol{X}})$, $1 \le j \le r \le k$, *are given in* (2.9)–(2.11).

As the form of the UMVUE for a general k is rather complicated, we give below the explicit expressions of $H_k(\underline{\boldsymbol{X}})$ for some special cases of interest.

After some simple calculations, it can be shown that

$$H_2(\underline{\boldsymbol{X}}) = \begin{cases} H_{21}(\underline{\boldsymbol{X}}) & \text{if } 0 < \dfrac{X_{(2)}}{X_{(1)}} < c \\[2ex] H_{22}(\underline{\boldsymbol{X}}) & \text{if } c \le \dfrac{X_{(2)}}{X_{(1)}} < 1 \end{cases}$$

and tedious but straightforward calculations lead to

$$H_3(\underline{\boldsymbol{X}}) = \begin{cases} H_{21}(\underline{\boldsymbol{X}}) + H_{31}(\underline{\boldsymbol{X}}) & \text{if } X_{(2)} < cX_{(1)} \\ H_{22}(\underline{\boldsymbol{X}}) + H_{31}(\underline{\boldsymbol{X}}) & \text{if } X_{(3)} < cX_{(1)} \le X_{(2)} \\ H_{32}(\underline{\boldsymbol{X}}) & \text{if } 0 < cX_{(1)} \le X_{(3)}, \end{cases}$$

where

$$H_{21}(\underline{\boldsymbol{X}}) = X_{(1)} - \frac{X_{(1)}}{2}\left[\left(\frac{cX_{(2)}}{X_{(1)}}\right)^p + \left(\frac{X_{(2)}}{cX_{(1)}}\right)^p\right],$$

$$H_{22}(\underline{\boldsymbol{X}}) = \frac{X_{(1)} + X_{(2)}}{2} - \frac{c^p}{2}\left[X_{(1)}\left(\frac{X_{(2)}}{X_{(1)}}\right)^p + X_{(2)}\left(\frac{X_{(1)}}{X_{(2)}}\right)^p\right],$$

$$H_{31}(\underline{\boldsymbol{X}}) = \frac{1}{6}\left[\left(\frac{cX_{(2)}}{X_{(1)}}\right)^p - \left(\frac{X_{(3)}}{cX_{(1)}}\right)^p\right] I(X_{(3)} > cX_{(2)}) \quad \text{and}$$

$$H_{32}(\underline{\boldsymbol{X}}) = \frac{X_{(1)} + X_{(2)} + X_{(3)}}{3} - \frac{c^p}{3}\left[X_{(1)}\left(\frac{X_{(2)}}{X_{(1)}}\right)^p + X_{(2)}\left(\frac{X_{(1)}}{X_{(2)}}\right)^p + X_{(3)}\left(\frac{X_{(1)}}{X_{(3)}}\right)^p\right].$$

Remark 2.2. When $p = n$, a positive integer, $H_2(\underline{\boldsymbol{X}})$ coincides with the UMVUE of the subset selected based on a random sample of size n from two exponential populations, derived by Vellaisamy and Sharma (1990). The ad hoc approach followed there is difficult to apply and, in fact, does not extend to an arbitrary k. An advantage of the method used in this paper is that it yields the UMVUE of W for any k although its explicit form is rather complicated even for moderately large values of k.

2.3 *Inadmissibility of the natural estimator*

We now establish the inadmissibility of the natural estimator of W for squared error loss defined by

$$L(T, W) = (T - W)^2.$$

In fact, we find a class of estimators that are better than the natural estimator for any k.

Consider the class of estimators of the form

$$T_a = \begin{cases} aX_{(1)} & \text{if } 0 < X_{(2)} < cX_{(1)} \\ \sum_{j=1}^{r}(X_{(j)}/r) & \text{if } X_{(r+1)} < cX_{(1)} \le X_{(r)},\ r = 2, \dots, k, \end{cases} \tag{2.13}$$

where $0 < a < 1$ and $X_{(k+1)} \equiv 0$. Observe that the natural estimator T corresponds to the choice $a = 1$.

THEOREM 2.2. *Assume* T_a *is an estimator of the form* (2.13). *Then for* $(p-1)/(p+1) \le a < 1$, T_a *dominates* T *as an estimator of* W *with respect to squared error loss.*

PROOF. Let $\Delta(\underline{\boldsymbol{\theta}})$ denote the risk difference

$$\Delta(\underline{\boldsymbol{\theta}}) = R(T_a, \underline{\boldsymbol{\theta}}) - R(T, \underline{\boldsymbol{\theta}}).$$

It is easy to see that

$$\begin{aligned} \Delta(\underline{\boldsymbol{\theta}}) &= E_{\underline{\boldsymbol{\theta}}}\{(aX_{(1)} - p\theta_{(1)})^2 - (X_{(1)} - p\theta_{(1)})^2\} I\left(\frac{X_{(2)}}{X_{(1)}} < c\right) \\ &= (a-1)E_{\underline{\boldsymbol{\theta}}}\{(a+1)X_{(1)}^2 - 2p\theta_{(1)}X_{(1)}\} I\left(\frac{X_{(2)}}{X_{(1)}} < c\right). \end{aligned} \tag{2.14}$$

Let

$$u_i(\underline{\boldsymbol{X}}) = \begin{cases} pX_i & \text{if } cX_i > X_{(1)i} \\ 0 & \text{otherwise,} \end{cases}$$

where $X_{(1)i}$ denotes, as before, the maximum of $X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_k$. Applying Lemma 2.2, we have

$$v_i(\underline{\boldsymbol{X}}) = \frac{p}{p+1} X_i^2 \left[1 - \left(\frac{X_{(1)i}}{cX_i}\right)^{p+1}\right] I\left(X_i > \frac{1}{c} X_{(1)i}\right)$$

as the unbiased estimator of $p\theta_i E_{\underline{\theta}}[X_i I(cX_i > X_{(1)i})]$. Therefore, an unbiased estimator of

$$p\theta_{(1)}X_{(1)}I\left(\frac{X_{(2)}}{X_{(1)}} < c\right) = p\sum_{i=1}^{k}\theta_i E_{\underline{\theta}}[X_i I(cX_i > X_{(1)i})]$$

can be seen to be

$$\frac{p}{p+1}X_{(1)}^2\left[1-\left(\frac{X_{(2)}}{cX_{(1)}}\right)^{p+1}\right]I(cX_{(1)} > X_{(2)}).$$

Hence, the unbiased estimator $\delta(\underline{\boldsymbol{X}})$ of $\Delta(\underline{\theta})$ is

$$\delta(\underline{\boldsymbol{X}}) = (a-1)X_{(1)}^2\left\{(a+1) - \frac{2p}{(p+1)}\left[1-\left(\frac{X_{(2)}}{cX_{(1)}}\right)^{p+1}\right]\right\} \times I(cX_{(1)} > X_{(2)}). \tag{2.15}$$

It is immediate that, when

$$(a+1) \geq \frac{2p}{p+1} = \sup_{\underline{x}\in R^k}\frac{2p}{p+1}\left[1-\left(\frac{X_{(2)}}{cX_{(1)}}\right)^{p+1}\right]I(cX_{(1)} > X_{(2)}),$$

$\delta(\underline{\boldsymbol{X}}) \leq 0$ and $\delta(\underline{\boldsymbol{X}}) < 0$ with probability $P_{\underline{\theta}}(cX_{(1)} > X_{(2)}) > 0$ for all $\underline{\theta}$.

Hence, when $(p-1)/(p+1) \leq a < 1$,

$$E_{\underline{\theta}}(\delta(\underline{\boldsymbol{X}})) = \Delta(\underline{\theta}) < 0 \quad \text{for all } \underline{\theta}.$$

This completes the proof of the theorem.

Remark 2.3. Let $a_0 = p/(p+1)$. A consequence of the above theorem is that T_{a_0} improves upon T for squared error loss. A similar result has been established by Vellaisamy and Sharma (1990) for the selected subset from two exponential populations. The technique used there is that of Brewster and Zidek (1974), which is difficult to apply for an arbitrary k.

3. Simultaneous estimation of the selected subset

Another problem of interest is the simultaneous estimation of Q, the parameters associated with the populations that are in the selected subset. That is, we consider the estimation of

$$Q = p\theta^{(r)} \quad \text{if} \quad X_{(r+1)} < cX_{(1)} \leq X_{(r)}, \tag{3.1}$$

where $\theta^{(r)} = (\theta_{(1)}, \ldots, \theta_{(r)})$, $\theta_{(i)}$ is the parameter associated with $X_{(i)}$ and $r = 1, 2, \ldots, k$.

This problem has not received much attention and to the knowledge of the author, there is hardly any literature on this topic. The problem is of importance especially when one is more interested, as in fact is often the case, in estimating the individual performances of the populations contained in the selected subset.

A natural estimator of Q is

$$U = X^{(r)} \quad \text{if} \quad X_{(r+1)} < cX_{(1)} \leq X_{(r)}$$

with $X^{(r)} = (X_{(1)}, \ldots, X_{(r)})$ and $r = 1, 2, \ldots, k$.

Alternatively, we can write

$$Q = p \sum_{r=1}^{k} \theta^{(r)} I(X_{(r+1)} < cX_{(1)} \leq X_{(r)})$$

and

$$U = \sum_{r=1}^{k} X^{(r)} I(X_{(r+1)} < cX_{(1)} \leq X_{(r)}).$$

Define now

$$G^{(r)}(\underline{\boldsymbol{X}}) = (h_{1,r}(\underline{\boldsymbol{X}}), \ldots, h_{r,r}(\underline{\boldsymbol{X}}))$$

and

$$q^{(r)}(\underline{\boldsymbol{\theta}}) = (q_{1,r}(\underline{\boldsymbol{\theta}}), \ldots, q_{r,r}(\underline{\boldsymbol{\theta}})).$$

The following theorem for the estimation of Q is an analogue of Theorem 2.1.

THEOREM 3.1. *For the estimation of Q, the estimator*

$$G_k(\underline{\boldsymbol{X}}) = \sum_{r=1}^{k} G^{(r)}(\underline{\boldsymbol{X}}),$$

is unbiased and is more concentrated about $q_k(\underline{\boldsymbol{\theta}}) = \sum_{r=1}^{k} q^{(r)}(\underline{\boldsymbol{\theta}})$ than any other unbiased estimator.

The proof of the above theorem is immediate by noting that $G^{(r)}(\underline{\boldsymbol{X}})$, $1 \leq r \leq k$, is unbiased and is more concentrated about $q^{(r)}(\underline{\boldsymbol{\theta}})$ than any other unbiased estimator of $q^{(r)}(\underline{\boldsymbol{\theta}})$ (cf. Lehmann ((1983), p. 291)).

Consider next the estimation of Q with squared error loss, i.e., for example, the loss involved in estimating Q by U is

$$L(U, Q) = \sum_{r=1}^{k} \left[\sum_{j=1}^{r} (X_{(j)} - p\theta_{(j)})^2 \right] I(X_{(r+1)} < cX_{(1)} \leq X_{(r)}).$$

Considering, as in Subsection 2.3, the estimator

$$U_a = aX_{(1)} I(X_{(2)} < cX_{(1)}) + \sum_{r=2}^{k} X^{(r)} I(X_{(r+1)} < cX_{(1)} \leq X_{(r)}),$$

$0 < a < 1$, the difference $\Delta^*(\underline{\theta})$ of the risk of U_a and that of U is

$$\begin{aligned}\Delta^*(\underline{\theta}) &= R(U_a, \underline{\theta}) - R(U, \underline{\theta}) \\ &= E_{\underline{\theta}}\{(aX_{(1)} - p\theta_{(1)})^2 - (X_{(1)} - p\theta_{(1)})^2\} I(X_{(2)} < cX_{(1)}),\end{aligned}$$

which is the same as $\Delta(\underline{\theta})$ defined in (2.14). Thus, we have the following theorem for simultaneous estimation after selection.

THEOREM 3.2. *The natural estimator U of Q is inadmissible and U_a, for $(p-1)/(p+1) \leq a < 1$, improves upon U for squared error loss.*

It follows that, as a special case, U_{a_0} dominates U.

Remark 3.1. The above theorem asserts that when $X_{(2)} < cX_{(1)}$, $(p/(p+1))X_{(1)}$ dominates $X_{(1)}$ as an estimator of $\theta_{(1)}$. However, it can be checked, by the same procedure adopted above, that the same thing is not true when, for example, $X_{(3)} < cX_{(1)} \leq X_{(2)}$. Similar remarks apply for the estimation of $\theta_{(2)}$ and etc.

4. Concluding remarks

We have seen in particular that T_{a_0} and U_{a_0} respectively dominate the natural estimators T and U (of W and Q) for squared error loss. Suppose instead the loss function is scale invariant, i.e., for example,

$$L(T, W) = \left(\frac{T}{W} - 1\right)^2.$$

Then the method followed in this paper does not go through. However, we believe that the above result is still true for scale invariant loss also.

It is well-known that in the usual estimation problem of $p\theta_1$ with $k = 1$, the best scale invariant estimator $pX_1/(p+1)$ is admissible when the loss is squared error. But, for the simultaneous estimation of $\underline{\theta} = (\theta_1, \ldots, \theta_k)$, where no selection is involved, the estimator $\delta(\underline{X}) = (\delta_1(\underline{X}), \ldots, \delta_k(\underline{X}))$ with $\delta_i(\underline{X}) = pX_i/(p+1)$ is inadmissible for $k \geq 2$. For an improved estimator, see for example, Berger ((1980), p. 560). It would be interesting in this context, and moreover in general, to investigate the admissibility of the estimators T_{a_0} and U_{a_0}. Our feeling is that these estimators could further be improved.

Acknowledgements

The author is grateful to the Indian Statistical Institute, New Delhi, where most of the work was carried out, and to the referee for several suggestions which led to the improvement in the presentation of the paper.

References

Berger, J. O. (1980). Improving on inadmissible estimators in continuous exponential families with applications to simultaneous estimation of gamma scale parameters, *Ann. Statist.*, **8**, 545–571.

Brewster, J. F. and Zidek, J. V. (1974). Improving on equivariant estimators, *Ann. Statist.*, **2**, 21–38.

Cohen, A. and Sackrowitz, H. (1982). Estimating the mean of the selected population, *Statistical Decision Theory and Related topics-III*, *Vol.* 1 (eds. S. S. Gupta and J. O. Berger), 243–270, Academic Press, New York.

Cohen, A. and Sackrowitz, H. (1989). Two stage conditionally unbiased estimators of the selected mean, *Statist. Probab. Lett.*, **8**, 273–278.

Gupta, S. S. (1963). On a selection and ranking procedure for gamma population, *Ann. Inst. Statist. Math.*, **14**, 199–216.

Jeyarathnam, S. and Panchapakesan, S. (1984). An estimation problem relating to subset selection for normal populations, *Design of Experiments*: *Ranking and Selection* (eds. T. J. Santner and A. C. Tamhane), Dekker, New York.

Jeyarathnam, S. and Panchapakesan, S. (1986). Estimation after subset selection from exponential populations, *Comm. Statist. Theory Methods*, **15**, 3459–3473.

Lehmann, E. L. (1983). *Theory of Point Estimation*, Wiley, New York.

Robbins, H. E. (1988). The UV method of estimation, *Statistical Decision Theory and Related Topics IV*, *Vol.* 1 (eds. S. S. Gupta and J. O. Berger), 265–270, Springer, New York.

Sackrowitz, H. and Samuel-Cahn, E. (1984). Estimation of the mean of a selected negative exponential population, *J. Roy. Statist. Soc. Ser. B*, **46**, 242–249.

Sackrowitz, H. and Samuel-Cahn, E. (1987). Evaluating the chosen population: a Bayes and minimax approach, *Adaptive Statistical Procedures and Related Topics*, IMS Lecture Notes—Monograph Series, Vol. 8, 386–399, Hayward, California.

Vellaisamy, P. and Sharma, D. (1990). On estimation after subset selection from exponential populations, *Sankhyā Ser. B*, **52**, 305–318.

Venter, J. H. (1988). Estimation of the mean of the selected population, *Comm. Statist. Theory Methods*, **17**, 797–805.