The Journal of
Mathematics and Computer Science

Available online at
http://www.TJMCS.com

The Journal of Mathematics and Computer Science Vol .3 No.2 (2011) 236- 245

# Estimating the Average Worth of a Subset Selected from Binomial Populations

**Riyadh Al-Mosawi**

*Department of Mathematics, Thiqar University, Thiqar, Iraq.*

Received: May 2011, Revised: June 2011
Online Publication: December 2011

**Abstract**

Suppose $\overline{X} = (\overline{X}_1, \cdots, \overline{X}_p), (p \geq 2)$, where $\overline{X}_i$ represents the mean of a random sample of size $n_i$ drawn from binomial $bin(1, \theta_i)$ population. Assume the parameters $\theta_1, \cdots, \theta_p$ are unknown and the populations $bin(1, \theta_1), \cdots, bin(1, \theta_p)$ are independent. A subset of random size is selected using Gupta's (Gupta, S. S. (1965). On some multiple decision(selection and ranking) rules. Technometrics 7,225-245) subset selection procedure. In this paper, we estimate of the average worth of the parameters for the selected subset under squared error loss and normalized squared error loss functions. First, we show that neither the unbiased estimator nor the risk-unbiased estimator of the average worth (corresponding to the normalized squared error loss function) exist based on a single-stage sample. Second, when additional observations are available from the selected populations, we derive an unbiased and risk-unbiased estimators of the average worth and also prove that the natural estimator of the average worth is positively biased. Finally, the bias and risk of the natural, unbiased and risk-unbiased estimators are computed and compared using Monti Carlo simulation method.

*Key words*: binomial populations, selected subset, average worth estimation

## 1 Introduction

The problem of estimation the parameters of the selected subset was initiated by Jeyarathnam and Panchapakesan (1984, 1986). The problem is also considered by Vellaisamy and Sharma (1990) and Vellaisamy (1992, 1996), Misra (1994), Vellaisamy and Punnen (2002), Moeschlin and Vellaisamy (2002), Gangopadhyay and Kumar (2005), Kumar, Mahapatra and Vellaisamy (2009), Vallaisamy and Al-Mosawi (2010) and Al-Mosawi, Shanubhogue and Vellaisamy (2011).
Suppose for $i = 1, 2 \cdots, p$, $X_i$ and $\overline{X}_i$ represent the sum and the mean, respectively, of a random sample of size $n_i$ drawn from binomial population $bin(1, \theta_i)$. Assume the populations are independent and the parameters $\theta_1, \ldots, \theta_p$ are unknown. It is well-known that $X_i \sim bin(n_i, \theta_i)$ while $\overline{X}_i$ is no longer binomial. Let $\theta_{[1]} = max\{\theta_1, \cdots, \theta_p\}$ and $\overline{X}_{(1)} = \max\{\overline{X}_1, \cdots, \overline{X}_p\}$ and let

$\theta = (\theta_1, \ldots, \theta_p)$, $X = (X_1, \ldots, X_p)$ and $\overline{X} = (\overline{X}_1, \ldots, \overline{X}_p)$. The population associated with $\theta_{[1]}$ is called the best population and in case of ties, we randomly tagged any of the tied populations. Using Gupta's (1965) subset selection approach, a subset of the binomial populations of random size is selected according to the following subset selection rule (Gupta and Panchpakasen (2002), page 257)

$$\text{R: Select the population } Bin(1, \theta_i) \text{ in the subset iff } \overline{X}_i \geq \overline{X}_{(1)} - d, \tag{1.1}$$

where $d = d(p, P^*, n_1, \ldots, n_p) > 0$ is the smallest nonnegative number for which

$$\min_{i=1,\ldots,p} \{ \inf_{0 \leq \lambda \leq 1} P(CS; \lambda, d, n_i)\} \geq P^*,$$

where $P(CS; \lambda, d, n_i)$ is the probability of including $Bin(1, \theta_i)$ in the selected subset when $\theta_1 = \ldots, \theta_p = \lambda$ and $P^*((1/p) < P^* \leq 1)$ is a pre-specified quantity.

In this paper, we considered the estimation problem of the average worth of the parameters for the selected populations. Namely our estimand is

$$\Phi = \sum_{i=1}^{p} \theta_i \psi_i(\overline{X}), \tag{1.2}$$

where $\psi_i(\overline{X}) = \dfrac{I(\overline{X}_i \geq \overline{X}_{(1)} - d)}{\sum_{j=1}^{p} I(\overline{X}_j \geq \overline{X}_{(1)} - d)}$. It is easy to see that $I(\overline{X}_i \geq \overline{X}_{(1)} - d) \equiv I(\overline{X}_i \geq \overline{X}_{(1)i} - d)$, where $\overline{X}_{(1)} = \max(\overline{X}_1, \ldots, \overline{X}_{i-1}, \overline{X}_{i+1}, \cdots, \overline{X}_p)$. The loss function considered here is given by

$$L(\Phi, \Phi_1) = \frac{(\Phi_1 - \Phi)^2}{\Phi^k}, \tag{1.3}$$

where $\Phi_1$ is an estimand of $\Phi$ and $k = 0, 1$. The loss (1.3) for $k = 0$ correspond to the squared error loss function while for $k = 1$ correspond to the normalized squared error loss function.

Unlike the classical estimation problem, our estimand is not fixed but it is a random parametric function which is a function of parameters and observations, as well.

In the following, we introduce some definitions. Let $\Omega$ denote the parameter space i.e. $\Omega = \{\theta : 0 < \theta_i < 1, i = 1, \cdots, p\}$.

**Definition 1.1** *(Vellaisamy (1993)) An estimand $\Phi$ is said to be U-estimable if there is an estimator $\Phi_1$ such that*

$$\mathbb{E}_\theta \Phi_1 = \mathbb{E}_\theta \Phi,$$

*for all $\theta \in \Omega$.*

The following definition is Lehmann's risk-unbiased definition (see Lehmann annd Casella (1998)) under normalized squared error loss function.

**Definition 1.2** *(Al-Mosawi, Vellaisamy and Shanubhogue (2011)) An estimand $\Phi$ is said to be RU-estimable if there is an estimator $\Phi_2$ such that*

$$\mathbb{E}_\theta \Phi_2^2 = \mathbb{E}_\theta \Phi^2,$$

*for all $\theta \in \Omega$.*

## 2 Estimation Based on Single-Stage Sample

This section is devoted to estimate the average worth of the parameters for the selected subset, using observations from single-stage sample scheme. We prove that the average worth of the parameters for the selected subset is neither U-estimable nor RU-estimable.

**Lemma 2.1** *The random parametric function $\theta_i^r g(\overline{x}), r \in \mathbb{Z}_+$, the set of nonnegative integers, is not U-esimable i.e there is no estimator $\delta$ such that*

$$\mathbb{E}_\theta(\theta_i^r g(\overline{X}) = \mathbb{E}_\theta(\delta(X)),$$

*where g is any real-valued function defined on p-fold Cratesian product of the interval [0,1].*

**Proof :** Observe that

$$\begin{aligned}
\mathbb{E}_\theta(\theta_i^r g(\overline{X}) = & \sum_{x_p=0}^{n_p} \cdots \sum_{x_1=0}^{n_1} g(\overline{x}_1, \ldots, \overline{x}_p)\theta_i^r \prod_{j=1}^{p} c(n_j, x_j)\theta_j^{x_j}(1-\theta_j)^{n_j - x_j} \\
=& g(0,\ldots,0)\theta_i^r \prod_{j=1}^{p}(1-\theta_j)^{n_j} + \cdots + g(1,\cdots,1)\theta_i^r \prod_{j\neq i}^{p} \theta_j^{n_j} \\
=& g(0,\ldots,0) \sum_{x_p=0}^{n_p} \cdots \sum_{x_1=0}^{n_1} (-1)^{\sum_{j=1}^{p} x_j} \prod_{j=1}^{p} c(n_j, x_j)\theta_i^r \prod_{j=1}^{p} \theta_j^{x_j} + \cdots + g(1,\cdots,1)\theta_i^r \prod_{j=1}^{p} \theta_j^{n_j} \\
=& \theta_i^{n_i + r} \prod_{j\neq i}^{p} \theta_j^{n_j} \left( g(1,1,\cdots,1) + \cdots + (-1)^{\sum_{j=1}^{p} n_j} g(0,0,\cdots,0) \right) + \cdots + g(0,\cdots,0)\theta_i^r,
\end{aligned}$$

where $c(a,b) = \dfrac{a!}{b!(a-b)!}$. From the last equation, we see that $\mathbb{E}_\theta(\theta_i g(\overline{X})$ is a polynomial of degree $(n_1, \ldots, n_{i-1}, n_i + r, n_{i+1}, \ldots, n_p)$. Now, since $\mathbb{E}_\theta(\delta(X))$ is a polynomial of degree $(n_1, \ldots, n_p)$ then the function $\theta_i g(\overline{X})$ is not U-estimable (Lehmann and Casella (1998), page 100). ∎

**Lemma 2.2** *The estimand $\Phi$ given in (1.2) is not U-estimable.*

**Proof :** Since $\theta_i\psi_i(\overline{x}), i = 1, \cdots, p$ are not U-estimable using Lemma 2.1, then $\Phi$ is not U-estimable. ∎

Similarly, we can prove the following lemma.

**Lemma 2.3** *The estimand $\Phi$ given in (1.2) is not RU-estimable.*

## 3 Estimation Based on Two-Stage Sample

Assume additional observations are available through second stage sample. We find the natural estimators for the average worth of the selected subset and prove these estimators are positively biased. Also we find an unbiased and risk-unbiased estimators for the average worth of the selected subset in case of additional observations from the selected populations are available through second stage. In the following lemma, we obtain an unbiased estimator for the one-dimensional random function $\lambda^r g(W)$.

**Lemma 3.1** *Let $W$ and $Y_r$, $r \geq 1$ be independent random variables , where $W \sim bin(n, \lambda)$ and $Y_r \sim bin(r, \lambda)$. If $g(w)$ is a real-valued function defined on $\mathbb{Z}_+$, the set of non-negative integers, such that $E_\theta(|g(W)|) < \infty$, then*

$$E_\lambda\big(\lambda^r g(W)\big) = E_\lambda\bigg(\frac{(W+Y_r)^{(r)}}{(n+r)^{(r)}}\, g(W+Y_r-r)\bigg).$$

**Proof :** It is clear that $T = W + Y_r \sim bin(n+r, \lambda)$. Observe that

$$\begin{aligned}
E_\lambda\bigg(\frac{T^{(r)}}{(n+r)^{(r)}}\, g(T-r)\bigg) &= \sum_{t=0}^{n+r} \frac{t^{(r)}}{(n+r)^{(r)}}\, g(t-r)c(n+r,t)\lambda^t(1-\lambda)^{n+r-t} \\
&= \sum_{t=r}^{n+r} g(t-r)c(n,t-r)\lambda^t(1-\lambda)^{n+r-t} \\
&= \lambda^r \sum_{w=0}^{n} g(w)c(n,w)\lambda^w(1-\lambda)^{n-w} \\
&= \lambda^r E_\lambda(g(W)).
\end{aligned}$$

∎

The following lemma extends Lemma 3.1 for p-dimensional case.

**Lemma 3.2** *Let $X = (X_1, \cdots, X_p)$ and $Y = (Y_{r_1,1}, \cdots, Y_{r_p,p})$ be vectors of p independent binomial random variables, where $X_i \sim bin(n_i, \theta_i)$, $Y_{r_i,i} \sim bin(r_i, \theta_i)$ and $r_i \geq 1$; $1 \leq i \leq p$. Assume Y is independent of X. Let $\overline{X} = (\overline{X}_1, \cdots, \overline{X}_p)$ where $\overline{X}_i = X_i/n_i$. If $f(\overline{x})$ is a real-valued function such that $E_\theta(|f(\overline{X})|) < \infty$, then*

$$\mathbb{E}_\theta\left(f(\overline{X})\prod_{i=1}^{p}\theta_i^{r_i}\right) = \mathbb{E}_\theta\left(f\left(\overline{X} + \sum_{i=1}^{p}\frac{(Y_{r_i,i} - r_i)}{n_i}e_i\right)\prod_{i=1}^{p}\frac{(X_i + Y_{r_i,i})^{(r_i)}}{(n_i + r_i)^{(r_i)}}\right), \tag{3.1}$$

*where $a^{(b)} = a(a-1)\ldots(a-b+1)$ and $e_i$ is a p-vector whose i-th component one and the rest are zero.*

**Proof :** Write for simplicity $X = (X_i, X^{(i)}), Y = (Y_{r_i,i}, Y^{(i)})$ and $\theta = (\theta_i, \theta^{(i)})$, where $X^{(i)} = (X_1, \cdots, X_{i-1}, X_{i+1}, \cdots, X_p)$, $Y^{(i)} = (Y_{r_1,1}, \cdots, Y_{r_{i-1},i-1}, Y_{r_{i+1},i+1}, \cdots, Y_{r_p,p})$ and $\theta^{(i)} = (\theta_1, \cdots, \theta_{i-1}, \theta_{i+1}, \cdots, \theta_p)$. By the independence of the $X_i$'s and $Y_{r_i,i}$,

$$\begin{aligned}
&E_\theta\left(f\left(\overline{X} + \sum_{j=1}^{p}\frac{Y_{r_j,j} - r_j}{n_j}e_j\right)\prod_{i=1}^{p}\frac{(X_i + Y_{r_i,i})^{(r_i)}}{(n_i + r_i)^{(r_i)}}\right)\\
=&E_{\theta^{(1)}}\prod_{i=2}^{p}\frac{(X_i + Y_{r_i,i})^{(r_i)}}{(n_i + r_i)^{(r_i)}}\left(E_{\theta_1}\left(\frac{(X_1 + Y_{r_1,1})^{(r_1)}}{(n_1 + r_1)^{(r_1)}}f\left(\overline{X} + \sum_{j=1}^{p}\frac{Y_{r_j,j} - r_j}{n_j}e_j\right)\big|X^{(1)}, Y^{(1)}\right)\right)\\
=&E_{\theta^{(1)}}\prod_{i=2}^{p}\frac{(X_i + Y_{r_i,i})^{(r_i)}}{(n_i + r_i)^{(r_i)}}\left(E_{\theta_1}\theta_1^{r_1}f\left(\overline{X} + \sum_{j=2}^{p}\frac{Y_{r_j,j} - r_j}{n_j}e_j\right)\big|X^{(1)}, Y^{(1)}\right)\Big)(\text{ using Lemma 3.1})\\
=&E_{\theta^{(2)}}\prod_{i=3}^{p}\frac{(X_i + Y_{r_i,i})^{(r_i)}}{(n_i + r_i)^{(r_i)}}\left(E_{\theta_2}\theta_1^{r_1}\theta_2^{r_2}f\left(\overline{X} + \sum_{j=3}^{p}\frac{Y_{r_j,j} - r_j}{n_j}e_j\right)\big|X^{(2)}, Y^{(2)}\right)\Big),
\end{aligned}$$

and continuing this process completes the proof. ∎

From classical estimation theory, it is known that $\dfrac{X_i + Y_{1,i}}{n_i + 1}$ is the UMVUE of $\theta_i, i = 1, \cdots, p$, so that the natural estimator of $\Phi$ based on single additional observation from each of the selected populations is given by

$$\Phi_N = \sum_{i=1}^{p}\frac{X_i + Y_{1,i}}{n_i + 1}\psi_i(\overline{X}). \tag{3.2}$$

In the following theorem, we show that the natural estimator $\Phi_N$ is a biased estimator.

**Theorem 3.1** *The natural estimator (3.2) is positively biased i.e. $\mathbb{E}_\theta\Phi_N \geq \mathbb{E}_\theta\Phi$, for all $\theta \in \Omega$.*

**Proof :** Without loss of generality, we consider the case $p = 2$. For $p = 2$, (1.2) and (3.2) are, respectively, reduced to

$$\Phi = \begin{cases} \theta_1, & \text{if } \overline{x}_2 < \overline{x}_1 - d; \\ \theta_2, & \text{if } \overline{x}_1 < \overline{x}_2 - d; \\ \frac{1}{2}(\theta_1 + \theta_2), & \text{if } \overline{x}_2 - d \leq \overline{x}_1 \leq \overline{x}_2 + d. \end{cases}$$

and

$$\Phi_N = \begin{cases} \frac{X_1+Y_{1,1}}{n_1+1}, & \text{if } \overline{x}_2 < \overline{x}_1 - d; \\ \frac{X_2+Y_{1,2}}{n_2+1}, & \text{if } \overline{x}_1 < \overline{x}_2 - d; \\ \frac{X_1+Y_{1,1}}{2(n_1+1)} + \frac{X_2+Y_{1,2}}{2(n_2+1)}, & \text{if } \overline{x}_2 - d \leq \overline{x}_1 \leq \overline{x}_2 + d. \end{cases}$$

Now

$$\begin{aligned} \mathbb{E}_\theta(\Phi_N - \Phi) =& \mathbb{E}_\theta\left(\frac{X_1 + Y_{1,1}}{n_1 + 1} - \theta_1\right) I(\overline{X}_2 < \overline{X}_1 - d) + \mathbb{E}_\theta\left(\frac{X_2 + Y_{1,2}}{n_2 + 1} - \theta_2\right) I(\overline{X}_1 < \overline{X}_2 - d) \\ &+ \frac{1}{2}\mathbb{E}_\theta\left(\frac{X_1 + Y_{1,1}}{n_1 + 1} + \frac{X_2 + Y_{1,2}}{n_2 + 1} - \theta_1 - \theta_2\right) I(\overline{X}_2 - d \leq \overline{X}_1 \leq \overline{X}_2 + d)) \\ =& \frac{1}{2}\mathbb{E}_\theta\left(\frac{X_1 + Y_{1,1}}{n_1 + 1} - \theta_1\right) I(\overline{X}_2 < \overline{X}_1 - d) + \frac{1}{2}\mathbb{E}_\theta\left(\frac{X_2 + Y_{1,2}}{n_2 + 1} - \theta_2\right) I(\overline{X}_1 < \overline{X}_2 - d) \\ =& \frac{1}{2}\eta_1 + \frac{1}{2}\eta_2 (say), \end{aligned}$$

since $I(\overline{X}_2 - d \leq \overline{X}_1 \leq \overline{X}_2 + d) = 1 - I(\overline{X}_1 < \overline{X}_2 - d) - I(\overline{X}_2 < \overline{X}_1 - d)$ and $(X_i + Y_{1,i})/(n_i + 1)$ is an unbiased estimator of $\theta_i, i = 1, 2$. Now, using Lemma 3.2 in $\eta_1$, we obtain

$$\begin{aligned} \eta_1 =& \mathbb{E}_\theta\left(\frac{X_1 + Y_{1,1}}{n_1 + 1} - \theta_1\right) I(\overline{X}_2 < \overline{X}_1 - d) \\ =& \mathbb{E}_\theta\left(\frac{X_1 + Y_{1,1}}{n_1 + 1} I(\overline{X}_2 < \overline{X}_1 - d) - \frac{X_1 + Y_{1,1}}{n_1 + 1} I\left(\overline{X}_2 < \overline{X}_1 + \frac{Y_{1,1} - 1}{n_1} - d\right)\right) \\ =& \mathbb{E}_\theta\left(\frac{X_1 + Y_{1,1}}{n_1 + 1}\left(I(\overline{X}_2 < \overline{X}_1 - d) - I\left(\overline{X}_2 < \overline{X}_1 + \frac{Y_{1,1} - 1}{n_1} - d\right)\right)\right) \\ =& \mathbb{E}_\theta\left(\frac{X_1 + Y_{1,1}}{n_1 + 1} I\left(\overline{X}_2 + d < \overline{X}_1 \leq \overline{X}_2 + d - \frac{Y_{1,1} - 1}{n_1}\right)\right) \\ >& 0, \end{aligned}$$

since $P\left(\overline{X}_2 + d < \overline{X}_1 \leq \overline{X}_2 + d - \frac{Y_{1,1}-1}{n_1}\right) > 0$. Similarly, we can prove $\eta_2 > 0$ and this completes the proof. ∎

Using Lemma 3.2, we find an unbiased estimator for the average worth of the selected subset.

**Theorem 3.2** *The estimator*

$$\Phi_U = \sum_{i=1}^{p} \frac{X_i + Y_{1,i}}{n_i + 1} \psi_i \Big( \overline{X} + \Big( \frac{Y_{1,i} - 1}{n_i} \Big) e_i \Big) \tag{3.3}$$

*is an unbiased estimator of* $\Phi$.

It is of interest to write an explicit form of (3.3) for the special case $p = 2$. When $p = 2$, (3.3) reduces to

$$\Phi_U = \begin{cases} \frac{X_2 + Y_{1,2}}{n_2 + 1}, & \text{if } \overline{X}_1 < \overline{X}_2 - d - \frac{1 - Y_{1,2}}{n_2}; \\ \frac{X_2 + Y_{1,2}}{2(n_2 + 1)}, & \text{if } \overline{X}_2 - d - \frac{1 - Y_{1,2}}{n_2} \le \overline{X}_1 < \overline{X}_2 - d - \frac{1 - Y_{1,1}}{n_1}; \\ \frac{X_1 + Y_{1,1}}{2(n_1 + 1)} + \frac{X_2 + Y_{1,2}}{2(n_2 + 1)}, & \text{if } \overline{X}_2 - d + \frac{1 - Y_{1,1}}{n_1} \le \overline{X}_1 \le \overline{X}_2 + d - \frac{1 - Y_{1,2}}{n_2}; \\ \frac{X_1 + Y_{1,1}}{2(n_1 + 1)}, & \text{if } \overline{X}_2 + d - \frac{1 - Y_{1,2}}{n_2} < \overline{X}_1 < \overline{X}_2 + d + \frac{1 - Y_{1,1}}{n_1}; \\ \frac{X_1 + Y_{1,1}}{n_1 + 1}, & \text{if } \overline{X}_1 > \overline{X}_2 + d + \frac{1 - Y_{1,1}}{n_1}. \end{cases}$$

Now, we find a risk-unbiased estimator with respect normalized squared error loss function of the average worth for the selected subset using Definition 1.2.

**Theorem 3.3** *The estimator* $\Phi_{RU}$ *such that*

$$\Phi_{RU}^2 = \sum_{i=1}^{p} \frac{(X_i + Y_{2,i})^{(2)}}{(n_i + 2)^{(2)}} \psi_i^2 \Big( \overline{X} + \frac{Y_{2,i} - 2}{n_i} e_i \Big)$$
$$+ 2 \sum_{i=j+1}^{p} \sum_{j=1}^{p-1} \frac{(X_i + Y_{1,i})(X_j + Y_{1,j})}{(n_i + 1)(n_j + 1)} \psi_i \Big( \overline{X} + \frac{Y_{1,i} - 1}{n_i} e_i + \frac{Y_{1,j} - 1}{n_j} e_j \Big) \psi_j \Big( \overline{X} + \frac{Y_{1,i} - 1}{n_i} e_i + \frac{Y_{1,j} - 1}{n_j} e_j \Big) \tag{3.4}$$

*is a risk-unbiased estimator of* $\Phi$.

**Proof :** It is easy to see

$$\Phi^2 = \sum_{i=1}^{p} \theta_i^2 \psi_i^2(\overline{X}) + 2 \sum_{i=j+1}^{p} \sum_{j=1}^{p-1} \theta_i \theta_j \psi_i(\overline{X}) \psi_j(\overline{X}).$$

So that

$$\begin{aligned} \mathbb{E}_\theta \Phi^2 &= \sum_{i=1}^{p} \mathbb{E}_\theta \theta_i^2 \psi_i^2(\overline{X}) + 2 \sum_{i=j+1}^{p} \sum_{j=1}^{p-1} \mathbb{E}_\theta \theta_i \theta_j \psi_i(\overline{X}) \psi_j(\overline{X}) \\ &= \sum_{i=1}^{p} \mathbb{E}_\theta \frac{(X_i + Y_{2,i})^{(2)}}{(n_i + 2)^{(2)}} \psi_i^2 (\overline{X} + \frac{Y_{2,i} - 2}{n_i} e_i) \\ &\quad + 2 \sum_{i=j+1}^{p} \sum_{j=1}^{p-1} \mathbb{E}_\theta \frac{(X_i + Y_{1,i})(X_j + Y_{1,j})}{(n_i + 1)(n_j + 1)} \psi_i (\overline{X} + \frac{Y_{1,i} - 1}{n_i} e_i + \frac{Y_{1,j} - 1}{n_j} e_j) \psi_j (\overline{X} + \frac{Y_{1,i} - 1}{n_i} e_i + \frac{Y_{1,j} - 1}{n_j} e_j) \end{aligned}$$

using Lemma 3.2. This completes the proof. ∎

# 4 Monti-Carlo simulation

In this section, a comparison of the performance of the natural, unbiased and risk-unbiased estimators, using Monti-carlo simulation technique is performed. The simulation is done using Matlab 7.6 (R2008a) for the case and $n_i = n, i = 1, 2, \cdots, p$ and the values of $d$ are selected from Gibbons, Olkin and Sobel (1999), Table Q.2, page 503. We follow the simulation procedure used by Vellaisamy and Al-Mosawi (2011). First the value of $p$ and $n$ are chosen and then a set of $\{\theta_1, \cdots, \theta_p\}$ of parameter values are chosen at random within the range $(0, 1)$. In the second step, an observation $X_i$ is randomly chosen from the distribution $bin(n, \theta_i), 1 \leq i \leq p$. In step 3, the selection rule $R$, defined in (1.1), is used to select the subset. To estimate the average worth of the parameters associated with the selected populations, we compute the bias and risk of the natural $\Phi_N$ and unbiased $\Phi_U$ estimators and the risk-bias and risk of the natural $\Phi_{N2}$ and risk-unbiased $\Phi_{RU}$ estimators. The above procedure is repeated 1500 times and the averages of the risks are calculated. The above procedure is repeated a number of times with different sets of parameters in $(0, 1)$ and then the averages are also calculated and presented in the Table 1.

We observe the following facts from the simulation results. The bias of the unbiased and risk-unbiased estimators are clearly close to zero. The bias of the natural estimator is positive and it increases with $p$,number of populations, increases and decreases with $n$ increases for almost all the cases. The risk of the natural estimator is apparently less than that of the unbiased and risk-unbiased estimators, for all the cases.

Table 1: The bias and risk of the natural, unbiased and risk-unbiased estimators

| $n$ | $p$ | $Bias(N)$ | $Bias(US)$ | $Risk - Bias(RS)$ | $Risk(N)$ | $Risk(US)$ | $Risk(RS)$ |
|---|---|---|---|---|---|---|---|
| 5 | 2 | 0.0324 | 0.0005 | -0.0007 | 0.0197 | 0.0258 | 0.0269 |
| | 5 | 0.0338 | 0.0002 | 0 | 0.0084 | 0.0112 | 0.0097 |
| | 10 | 0.0473 | -0.0004 | -0.0007 | 0.0066 | 0.0072 | 0.0057 |
| | 20 | 0.0226 | -0.0002 | -0.0002 | 0.0022 | 0.0021 | 0.0018 |
| 10 | 2 | 0.0221 | -0.0002 | -0.0008 | 0.0116 | 0.0172 | 0.0189 |
| | 5 | 0.0402 | -0.0001 | -0.0015 | 0.008 | 0.014 | 0.013 |
| | 10 | 0.0535 | 0.0004 | -0.0014 | 0.0063 | 0.0113 | 0.0088 |
| | 20 | 0.0457 | 0.0001 | 0 | 0.0036 | 0.0045 | 0.0032 |
| 20 | 2 | 0.0205 | 0 | -0.0009 | 0.0092 | 0.0139 | 0.015 |
| | 5 | 0.0303 | -0.0005 | -0.0023 | 0.0059 | 0.0138 | 0.0147 |
| | 10 | 0.0444 | 0.0007 | -0.0009 | 0.0049 | 0.0127 | 0.0105 |
| | 20 | 0.0443 | 0.0009 | 0.0005 | 0.0034 | 0.0064 | 0.0041 |

# References


[1] Al-Mosawi, R., Vellaisamy P. and Shanubhogue, A. (2011). Risk-Unbiased estimation of the selected subset of Poisson populations. To appear in *Journal of Indian Statistical Association*, **49**, 2011.

[2] Al-Mosawi, R. R.,Shanubhogue A. and Vellaisamy, P. (2011). Average worth estimation of the selected subset of Poisson populations. *Statistitcs*,,pp:1-19.

[3] Gibbons, J. D., Olkin, I. and Sobel, M. (1999). Selecting and ordering populations: a new statistical methodology.*Society for Industrial and Applied Mathematics (SIAM),Philadelphia, PA, USA.*

[4] Gupta, S. S. (1965). On some multiple decision (selection and ranking) rules. *Technometrics,* **7**,225-245.

[5] Gupta, S. S. and Panchapakesan, S. (2002). Multiple decision procedures: theory and methodology of selection and ranking populations. *Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, USA.*

[6] 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), Marcel-Dekker, New York.

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

[8] Kumar, S., Mahapatra, A. K. and Vellaisamy, P.(2009). Relaibility estimation of the selected exponential populations. *Statistics and Probability Letters.*,**52**,B,Pt.3,pp:305-318.

[9] Lehmann, E. L. and Casella, G. (1998). Theory of point estimation. *Second Edition, Springer-Verlag, New York.*

[10] Vellaisamy, P. (1992). Average worth and simulatneous estimation of the selected subset. *Ann. Inst. Statist. Math.,* , **44**, 551-562.

[11] Vellaisamy, P. (1993). On UMVUE estimation following selection. *Comm. Statist.-Theory Methods*, **22**(4),1031-1043.

[12] Vellaisamy, P. (1996). Simultaneous estimation of the selected subset of uniform populations. *J. of Appl. Statist.Sci.*, **5**, pp:39-46.

[13] Vellaisamy, P., Al-Mosawi, R. R. (2010). Simultaneous estimation of Poisson means of the selected subset. *J. of Statist. Plann. Infer.*,**140**,pp:3355-3364.