BOUNDED RISK ESTIMATION OF A FINITE POPULATION MEAN: OPTIMAL STRATEGIES

by

Nitis Mukhopadhyay
Pranab Kumar Sen
Bikas Kumar Sinha

Institute of Statitics Mimeo Series No. 1812

December 1986

BOUNDED RISK ESTIMATION OF A FINITE POPULATION MEAN:

OPTIMAL STRATEGIES

NITIS MUKHOPADHYAY, PRANAB KUMAR SEN AND BIKAS KUMAR SINHA*

ABSTRACT

For the mean of a finite population, a bounded risk estimation problem is considered for both the situations where the population variance may or may not be known. In this context, three popular (equal probability) sampling strategies are considered. These are the analogues of (i) simple random sampling with replacement, mean per unit estimation, (ii) simple random sampling with replacement, mean per distinct unit estimation, and (iii) simple random sampling without replacement, mean per unit estimation. It is well known that in the conventional fixed-sample size scheme, (iii) fares better than (ii) and (ii) better than (i). However, in the current context, the sample sizes are dictated by (possibly, degenerate) stopping times, and visualizing the cost (due to measurements/recording, etc.) as a function of the number of distinct units in the sample (as pertinent to schemes (i) and (ii)) and identifying that in scheme (iii), the number of distinct units is equal to the sample size itself, we are able to show that the second strategy still fares better than the first, although the third strategy may *not* perform better than the second one. Actually, in the case of known population variance, it is shown that in the light of the number of distinct units, the ASN (average sample number) for the second strategy can never be greater than two plus the ASN for the third strategy and can never be less than the latter minus one. A similar relationship also holds in the case of unknown population variance when we define the stopping rules in a coherent manner. Interestingly enough, this is quite contrary to our age-old belief that simple random sampling with replacement can never perform better than simple random sampling without replacement. Our theoretical results are backed up by numerical examples, too. Also, dominance of Strategy (ii) over (i) in a general sequential setup constitutes an important task of the current study. Finally, to reconcile Strategies (ii) and (iii) in a general sequential setup, the coherence of the associated stopping times has also been discussed thoroughly.

KEY WORDS: ASN; Distinct units; Minimum risk; Sequential estimation; Simple random sampling; With replacement; Without replacement.

*Nitis Mukhopadhyay is Professor, Department of Statistics, University of Connecticut, Storrs, CT 06268; Pranab Kumar Sen is Cary C. Boshamer Professor of Biostatistics and Adjunct Professor of Statistics, University of North Carolina, Chapel Hill, NC 27514; and Bikas Kumar Sinha is Professor, Indian Statistical Institute, Calcutta 700 035, India. He is currently visiting the University of Illinois at Chicago, IL 60680. This work was partially supported by (i) Office of Naval Research, Contract N00014-85-K-0548, and (ii) Office of Naval Research, Contract N00014-83-K-387.

1. INTRODUCTION

We consider a finite (labelled) population of N units, serially numbered 1,...,N. Denote by Y the study-variate which assumes values $Y_1,\ldots,Y_N$ on the units 1,...,N, respectively. The finite population mean ($\bar{Y}$) and variance (σ^2) are defined by

$$\bar{Y} = N^{-1}\Sigma_{i=1}^{N} Y_i \quad \text{and} \quad \sigma^2 = N^{-1}\Sigma_{i=1}^{N}(Y_i-\bar{Y})^2. \tag{1.1}$$

Also, for later use, we write

$$S^2 = (N-1)^{-1}\Sigma_{i=1}^{N}(Y_i-\bar{Y})^2 \quad (= N(N-1)^{-1}\sigma^2). \tag{1.2}$$

We are primarily interested in the estimation of $\bar{Y}$ (i.e., the finite population mean) with a bounded risk. In this problem, σ^2 may or may not be known. Also, for this problem, we may consider the following sampling strategies (as extended to the sequential case, whenever needed):

(i) Simple random sampling with replacement (SRSWR), mean per unit estimation;

(ii) SRSWR, mean per distinct unit estimation;

(iii) Simple random sampling without replacement (SRSWOR), mean per unit estimation.

In the conventional fixed-sample case, a relative comparison of the above strategies is well known [viz., Basu (1958), Raj and Khamis (1958) and Asok (1980), among others]. The strategies are known to be progressively better. However, in the current context, the results seem to indicate that while the ordering between the analogues of the first and second strategies remains the same, the ordering between the analogues of the second and third strategies may change in some cases. This is contrary to the popular belief that SRSWOR *always* performs at least as good as, or better than, the SRSWR.

To set our analysis in the proper perspective, in a SRSWR, we denote the random variables and indexes associated with the successive drawings by (y_k, r_k), $k \geq 1$, so that for each $k(\geq 1)$, r_k takes on the values $1,\ldots,N$ with equal probability N^{-1} and $y_k = Y_{r_k}$, $k \geq 1$. Then, the simple mean per unit estimate of $\bar{Y}$ (based on a sample of size n) is given by

$$\bar{y}_n = n^{-1}\Sigma_{i=1}^{n} y_i \ . \tag{1.3}$$

It is well known that $\bar{y}_n$ is an unbiased estimator of $\bar{Y}$ and

$$E(\bar{y}_n - \bar{Y})^2 = n^{-1}\sigma^2, \quad \forall\ n \geq 1. \tag{1.4}$$

Let us then consider a sequence $\{I_k;\ k \geq 1\}$ of indicator variables, where

$$I_k = \begin{cases} 1, & r_k \notin \{r_1,\ldots,r_{k-1}\}, \\ 0, & \text{otherwise}; \end{cases} \qquad k \geq 1 \tag{1.5}$$

with $I_1 = 1$.

Then, for every $n \geq 1$,

$\nu_n = \Sigma_{k=1}^{n} I_k$ denotes the number of distinct units in the sample of size n. (1.6)

Note that $\nu_1 = 1$, ν_n is ↗ in n, and

$$E(\nu_n) = N\{1 - (1 - 1/N)^n\}, \quad \forall\ n \geq 1. \tag{1.7}$$

The mean per distinct unit (in the sample of size n) is given by

$$\bar{y}_{(\nu_n)} = \nu_n^{-1}\Sigma_{k=1}^n I_k y_k \ ; \ n \geq 1. \tag{1.8}$$

Note that $\bar{y}_{(\nu_n)}$ is also unbiased for $\bar{Y}$ and

$$E(\bar{y}_{(\nu_n)} - \bar{Y})^2 = S^2\{E(\nu_n^{-1}) - N^{-1}\}$$

$$= \sigma^2(\frac{N}{N-1})\{E(\frac{1}{\nu_n}) - \frac{1}{N}\}. \tag{1.9}$$

In SRSWOR, for n sample units, the indices are denoted by R_1, $\ldots, R_n$, so that $\underset{\sim}{R}_n = (R_1, \ldots, R_n)$ takes on any (unordered) subset of n out of N numbers $\{1, \ldots, N\}$ with the same probability $N^{-[n]} = \{N \ldots (N - n + 1)\}^{-1}$, and the sample random variables $y_1', \ldots, y_n'$ are given by $y_k' = Y_{R_k}$, $k = 1, \ldots, n$. The mean per unit estimator is

$$\bar{y}_n' = n^{-1}\Sigma_{i=1}^n y_i' \ , \ n \geq 1. \tag{1.10}$$

Like the other two estimators, $\bar{y}_n'$ is also unbiased for $\bar{Y}$ and

$$E(\bar{y}_n' - \bar{Y})^2 = S^2(\frac{1}{n} - \frac{1}{N}). \tag{1.11}$$

Note that in the SRSWR, whenever a unit is chosen in the sample for the second time (or later), we do not incur any cost for its measurement (or recording), and, hence, it seems quite plausible to have a cost function $c(n) = c\nu_n$, $n \geq 1$, where $c(> 0)$ is a constant. Thus, $c(n)$ is stochastic in nature and

$$Ec(n) = cE\nu_n = cN\{1-(1-N^{-1})^n\},\ n \geq 1. \tag{1.12}$$

In SRSWOR, $\nu_n = n$, and hence, $c(n) = cn$ is non-stochastic. This difference in the nature of the cost function plays a basic role in the sequential schemes to be considered here.

We may consider, for an arbitrary estimator T_n of $\bar{Y}$, the usual squared error loss function $A\,(T_n - \bar{Y})^2$, where $A(> 0)$ is a given positive constant. Then corresponding to a given upper bound $W(> 0)$ to the risk of any estimator of $\bar{Y}$, we may define

$$n_0 = \min\{n \geq 1 : AE(T_n - \bar{Y})^2 \leq W\}. \tag{1.13}$$

We may then compare the $Ec(n_0)$ for different sampling strategies. This is what we may call the *bounded risk* approach for the comparison of the different sampling strategies. It may be noted that generally $E(T_n - \bar{Y})^2$ involves the unknown σ^2 (or S^2), and, hence, we may need to consider suitably modified *stopping rules* which, of course, would generally make the analysis more complicated. This aspect will be studied in detail in Section 3. An alternative approach to (1.13) would be to consider the *risk function*

$$\rho_n(T_n,\bar{Y}) = AE(T_n - \bar{Y})^2 + Ec(n) \tag{1.14}$$

and to determine n in such a way that (1.14) is a minimum. Then, it seems quite plausible to compare these "minimum risks" for the different strategies. Here also a "stopping rule" approach is needed when σ^2 (or S^2) is not known. We shall mainly confine ourselves to the first (i.e., bounded risk) approach, and indicate how parallel results hold for the "minimum risk" approach.

2. BOUNDED RISK ESTIMATION OF $\bar{Y}$: σ^2 KNOWN

Consider the following strategies:

Strategy I: Keep in mind (1.3) - (1.4). Adopt (SRSWR, n_0) sampling and use $\bar{y}_{n_0}$ as an estimator of $\bar{Y}$, where n_0 is so chosen that

$$AW^{-1}\sigma^2 \leq n_0 < 1 + AW^{-1}\sigma^2. \tag{2.1}$$

Note that by virtue of (1.7), for this strategy, we have

$$Ec(n_0) = cE(\nu_{n_0}) = cN\{1 - (1 - N^{-1})^{n_0}\}. \tag{2.2}$$

Strategy II: Adopt (SRSWR, n^*) sampling and use $\bar{y}_{(\nu_{n^*})}$ as an estimator of $\bar{Y}$, where $\bar{y}_{(\nu_n)}$ is defined by (1.8) and n^* is so chosen that

$$E\{\frac{1}{\nu_{n^*}}\} \leq \frac{1}{N} + \frac{W}{AS^2} < E\{\frac{1}{\nu_{n^*-1}}\} \; ; \tag{2.3}$$

the motivation for this choice of n^* is derived from (1.9). For this strategy, we have

$$Ec(n^*) = cE\nu_{n^*} = cN\{1 - (1 - N^{-1})^{n^*}\}. \tag{2.4}$$

Strategy III: Keep in mind (1.10) - (1.11). Adopt (SRSWOR, n^{**}) sampling and use $\bar{y}'_{n^{**}}$ as an estimator of $\bar{Y}$, where n^{**} is so chosen that

$$1/n^{**} \leq \frac{1}{N} + \frac{W}{AS^2} < 1/(n^{**}-1). \tag{2.5}$$

In this case, we have

$$Ec(n^{**}) = cn^{**}. \qquad (2.6)$$

Clearly, for known σ^2 (or S^2), each of the strategies leads to an unbiased estimator of $\bar{Y}$ with "bounded risk", and hence a comparison of (2.2), (2.4), and (2.6) would reveal the relative efficiencies of these strategies. We term (2.2), (2.4), and (2.6) as the *Average Cost Function* of Strategies I, II, and III, respectively, and denote them by ACF (I), ACF(II), and ACF(III) in that order. Note that these are all functions of A, W, σ^2 and N.

Theorem 2.1. Uniformly in A, W, N, *and* σ^2,

$$\text{ACF(II)} \leq \text{ACF(I)}. \qquad (2.7)$$

Proof. First, we may note that [c.f., Asok (1980)]

$$E\left(\frac{1}{\nu_n}\right) - \frac{1}{N} \leq \frac{N-1}{Nn}, \ \forall\ n \geq 1, \qquad (2.8)$$

so that

$$E\left\{\frac{1}{\nu_{n^*-1}}\right\} \leq \frac{N+n^*-2}{N(n^*-1)}. \qquad (2.9)$$

Writing $B = N^{-1} + W(AS^2)^{-1}$, we have from (2.3) and (2.9),

$$B < (N+n^*-2)/\{N(n^*-1)\}. \qquad (2.10)$$

Since $S^2 = N(N-1)^{-1}\sigma^2$, (2.10) and some routine steps lead us to

$$n^* < 1 + AW^{-1}\sigma^2. \qquad (2.11)$$

Thus, by (2.1) and (2.11), we have

$$n^* < n_0 + 1, \qquad (2.12)$$

and as n^* and n_0 are both positive integers, we have, therefore, $n^* \le n_0$. Consequently, by (2.2) and (2.4), we have $E\nu_{n^*} \le E\nu_{n_0}$ and this implies (2.7). Q.E.D.

Theorem 2.2. Uniformly in A, W, N *and* σ^2,

$$- c < ACF(II) - ACF(III) < 2c. \tag{2.13}$$

Before we proceed to prove this theorem, we may note that (2.13) actually relates to the inequality:

$$-1 < E(\nu_{n^*}) - n^{**} < 2, \text{ uniformly in A, W, N and } \sigma^2. \tag{2.14}$$

Or, in other words, $E(\nu_{n^*})$ cannot be smaller than $n^{**} - 1$ and also it cannot exceed $n^{**} + 2$. We shall show by some numerical examples that $E\nu_{n^*}$ may be sometimes less than n^{**}, while it may also be greater than n^{**}. The major implication of this theorem is that Strategy III may not always perform better than Strategy II; they are generally very "close" in their performance characteristics. In this context, we need the following.

Lemma 2.3. For SRSWR (N,n), *for every* $n \ge 2$

$$(E\nu_n)^{-1} \le E(\nu_n^{-1}) < (E\nu_{n-1})^{-1}. \tag{2.15}$$

Proof. Note that [viz., Chakrabarti (1965), Korwar and Serfling (1970), Pathak (1961)]

$$\begin{aligned} E(\nu_n^{-1}) &= N^{-n}\Sigma_{j=1}^{N} j^{n-1} = N^{-1}\Sigma_{j=1}^{N}(j/N)^{n-1} \\ &= N^{-1}\{1 + \Sigma_{j=1}^{N-1}(j/N)^{n-1}\} \\ &= N^{-1}\{1 + \Sigma_{k=1}^{N-1}(1-k/N)^{n-1}\}. \end{aligned} \tag{2.16}$$

On the other hand, by (1.7),

$$E\nu_{n-1} = N\{1 - (1 - 1/N)^{n-1}\}, \ n \geq 2. \tag{2.17}$$

Therefore,

$$(E\nu_{n-1})^{-1} = N^{-1}\{1 - (1 - 1/N)^{n-1}\}^{-1}$$

$$= N^{-1}\{1 + \Sigma_{k=1}^{\infty}(1 - N^{-1})^{k(n-1)}\} \tag{2.18}$$

$$> N^{-1}\{1 + \Sigma_{k=1}^{N-1}(1 - N^{-1})^{k(N-1)}\}. \tag{2.19}$$

Again, it follows easily (by induction on $k \geq 1$) that

$$(1 - N^{-1})^{k(n-1)} \geq (1 - kN^{-1})^{(n-1)}, \ \forall \ n \geq 1, \ k \geq 1. \tag{2.20}$$

By (2.16), (2.18), and (2.20), we immediately get that

$$(E\nu_{n-1})^{-1} > E\nu_n^{-1}. \tag{2.21}$$

On the other hand, $E(\nu_n)E(\nu_n^{-1}) \geq 1$, so that

$$E(\nu_n^{-1}) \geq (E\nu_n)^{-1}. \tag{2.22}$$

Thus, (2.15) follows from (2.21) and (2.22).

Proof of Theorem 2.2. By (2.3) and (2.5),

$$E(\nu_{n^*-1}^{-1}) > N^{-1} + W(AS^2)^{-1} \geq 1/n^{**}, \tag{2.23}$$

while by (2.15) and (2.23),

$$(E\nu_{n^*-2})^{-1} > E(\nu_{n^*-1}^{-1}) \geq 1/n^{**}. \tag{2.24}$$

Note that (2.24) ensures that

$$E\nu_{n^*-2} < n^{**}. \tag{2.25}$$

On the other hand, by (1.7)

$$E\nu_n = 1 + (1 - \frac{1}{N}) + (1 - \frac{1}{N})^2 E\nu_{n-2}, \tag{2.26}$$

so that by (2.25) and (2.26), we have

$$E\nu_{n*} < 1 + (1 - \frac{1}{N}) + (1 - \frac{1}{N})^2 n^{**}$$

$$= n^{**} + (1 - N^{-1}n^{**})(2 - N^{-1}). \tag{2.27}$$

Further, by (2.3), (2.5), and (2.22),

$$\frac{1}{n^{**}-1} > N^{-1} + W(AS^2)^{-1} \geq E(\nu_{n*}^{-1}) \geq (E\nu_{n*})^{-1}, \tag{2.28}$$

so that

$$E\nu_{n*} > n^{**} - 1. \tag{2.29}$$

Combining (2.27) and (2.29), we have

$$n^{**} - 1 < E\nu_{n*} < n^{**} + (1-N^{-1}n^{**})(2-N^{-1}) < n^{**} + 2 \tag{2.30}$$

for every N, W, A, and S^2. This completes the proof of Theorem 2.2. Q.E.D.

Remark 1: It may be noted that in the above analysis, choice of W is quite arbitrary and it is generally left to the experimenter. Two particular choices based on cost considerations may be suggested.

(a) Choose n^{**} beforehand and set $W_1 = AS^2\{(n^{**})^{-1} - N^{-1}\}$ = risk attained by the use of $\{\text{SRSWOR}(N,n^{**}), \bar{y}'_{n^{**}}\}$ strategy. For the competing strategy $\{\text{SRSWR}(N,n^*), \bar{y}_{(\nu_{n*})}\}$ with the *same* bound W_1 to the risk, the expected sample size $E(\nu_{n*})$ satisfies the inequality

$$n^{**} < E(\nu_{n*}) < n^{**} + (1 - \frac{n^{**}}{N})(2 - N^{-1}).$$

(b) Choose n^* beforehand and set $W_2 = AS^2\{E((\nu_{n^*})^{-1}) - N^{-1}\}$ = risk attained by the use of $\{SRSWR(N,n^*), \bar{y}_{(\nu_{n^*})}\}$ strategy. Then determine n^{**} such that the use of $\{SRSWOR(N,n^{**}), \bar{y}'_{n^{**}}\}$ strategy yields the same bound W_2 to the risk. This time we can prove a slightly improved version of (2.30), viz.,

$$n^{**} - 1 < E(\nu_{n^*}) < 1 + (1 - N^{-1})n^{**} < n^{**} + 1.$$

Next, recall that $B = \frac{1}{N} + \frac{W}{AS^2}$ so that if B^{-1} is an integer, then, of course, $n^{**} = B^{-1}$, and hence $E(\nu_{n^*}) \geq n^{**}$.

On the other hand, when B^{-1} is not an integer, let us set $B^{-1} = \alpha + \beta$, $0 < \beta < 1$, $\alpha = [B^{-1}]$, the integral part of B^{-1}. Then, $n^{**} = 1 + \alpha$ while $E(\nu_{n^*}) \geq B^{-1} = \alpha + \beta$. Thus, at least for small values of β, there is a possibility of $E(\nu_{n^*})$ being smaller than n^{**}. This is indeed true in some cases as evidenced by Table 1.

[Table 1 goes approximately here.]

Remark 2: If $W \to 0$ so that $B^{-1} \to N$, α becomes large and, hence, we can expect that for a wider range of β-values, $E(\nu_{n^*})$ would be smaller than n^{**}.

We conclude this section with some comments on the "minimum risk" approach. Note that by (1.4), (1.9), and (2.8), $E(\bar{y}_{(\nu_n)} - \bar{Y})^2 \leq E(\bar{y}_n - \bar{Y})^2$, $\forall\ n \geq 1$, and, hence, noting that $Ec(n)$ is the same for both Strategies I and II, we conclude that

$$\inf_n \rho_n(\bar{y}_{(\nu_n)}, \bar{Y}) \leq \inf_n \rho_n(\bar{y}_n, \bar{Y}), \tag{2.A}$$

so that Strategy II fares better than I. To compare Strategies II and III, we note that if n** is the specific value of n for which $\rho_n(\bar{y}_n', \bar{Y})$ is a minimum, we have

$$AS^2\left(\frac{1}{n^{**}(n^{**}+1)}\right) \leq c \leq AS^2\left(\frac{1}{n^{**}(n^{**}-1)}\right). \tag{2.B}$$

On the other hand, $\inf_n \rho_n(\bar{y}_n', \bar{Y}) = AS^2\left(\frac{1}{n^{**}} - \frac{1}{N}\right) + cn^{**}$.

Therefore, we have

$$c(2n^{**}-1) - N^{-1}AS^2 < \inf_n \rho_n(\bar{y}_n', \bar{Y}) \leq c(2n^{**}+1) - N^{-1}AS^2. \tag{2.C}$$

On the other hand, suppose that $\rho_{n^*}(\bar{y}_{(\nu_{n^*})}, \bar{Y}) = \inf_n \rho_n(\bar{y}_{(\nu_n)}, \bar{Y})$.

Then

$$\begin{aligned} \rho_{n^*}(\bar{y}_{(\nu_{n^*})}, \bar{Y}) &= AS^2\{E(\nu_{n^*}^{-1}) - N^{-1}\} + cE\nu_{n^*} \\ &\geq AS^2\{(E\nu_{n^*})^{-1} - N^{-1}\} + cE\nu_{n^*}, \end{aligned} \tag{2.D}$$

where $E\nu_{n^*}$ need not be a (positive) integer, while n** is so. Thus, whenever $n^{**} - 1 < E\nu_{n^*} < n^{**} + 1$, but $(A/c)^{\frac{1}{2}}S$ is not an integer, the right-hand side of (2.D) may actually be smaller than $\rho_{n^{**}}(\bar{y}_{n^{**}}', \bar{Y})$. However, if $n^{**} = (A/c)^{\frac{1}{2}}S$, then (2.D) cannot be smaller than $\rho_{n^{**}}(\bar{y}_{n^{**}}', \bar{Y})$, the true minimum of $AS^2\left(\frac{1}{u} - \frac{1}{N}\right) + cu$, over $u > 0$. In other words, a lower bound for $\rho_{n^*}(\bar{y}_{(\nu_{n^*})}, \bar{Y})$ may not necessarily be greater than or equal to $\rho_{n^{**}}(\bar{y}_{n^{**}}', \bar{Y})$. On the other hand, $E\nu_{n^*} = 1 + (1-N^{-1})E\nu_{n^*-1}$ and by Lemma 2.3, $E\nu_{n^*}^{-1} \leq (E\nu_{n^*-1})^{-1}$. Thus, we have

$$\begin{aligned} \rho_{n^*}(\bar{y}_{(\nu_{n^*})}, \bar{Y}) &\leq AS^2\{(E\nu_{n^*-1})^{-1} - N^{-1}\} \\ &\quad + c\{E(\nu_{n^*-1}) + (1 - N^{-1}E(\nu_{n^*-1}))\}. \end{aligned}$$

In fact, since $\rho_{n^*}(\bar{y}_{(\nu_{n^*})},\bar{Y}) \leq \rho_m(\bar{y}'_{\nu_m},\bar{Y})$, for all m, we may even use a crude upper bound:

$$\rho_{n^*}(\bar{y}_{(\nu_{n^*})},\bar{Y}) \leq AS^2\{(E\nu_m)^{-1}-N^{-1}\} + c\{E\nu_m+(1-N^{-1}E\nu_m)\} \qquad (2.E)$$

for an arbitrary m. We choose m such that $n^{**} \leq E\nu_m < n^{**} + 1$. Then from (2.E), we have

$$\rho_{n^*}(\bar{y}_{(\nu_{n^*})},\bar{Y}) \leq AS^2\{\frac{1}{n^{**}} - \frac{1}{N}\} + c\{n^{**} + 1 + 1 - \frac{n^{**}}{N}\}$$

$$= \rho_{n^{**}}(\bar{y}'_{n^{**}},\bar{Y}) + c(2 - \frac{n^{**}}{N}). \qquad (2.F)$$

(2.F) is comparable to (2.30). Note that by (2.B), $n^{**} = O_e(c^{-\frac{1}{2}})$ as $c \downarrow 0$, while $\rho_{n^{**}}(\bar{y}'_{n^{**}},\bar{Y}) = O_e(c^{\frac{1}{2}})$. Thus, by (2.F), we conclude that as $c \downarrow 0$,

$$\inf_n \rho_n(\bar{y}_{(\nu_n)},\bar{Y}) \leq \inf_n \rho_n(\bar{y}'_n,\bar{Y}) + O(c). \qquad (2.G)$$

This clearly depicts the "closeness" of the two minimum risks for the Strategies II and III.

3. BOUNDED RISK ESTIMATION OF $\bar{Y}$: σ^2 UNKNOWN

For the case of infinite population, sequential procedures for this problem were considered by Robbins (1959), Chow and Robbins (1965), Ghosh and Mukhopadhyay (1979) among others. Along their lines, we may consider the following (modified) strategies.

Strategy I': Sample units one by one at random and with replacement, governed by the stopping rule:

$$\tau_0 = \min\{n \geq 2: n \geq W^{-1}A(s_n^2 + n^{-\gamma})\}. \qquad (3.1)$$

Here, $s_n^2 = (n-1)^{-1} \Sigma_{i=1}^n (y_i - \bar{y}_n)^2$ for every $n \geq 2$ and γ is an arbitrary positive constant. Then, $\bar{y}_{\tau_0}$ is the desired (sequential) estimator of $\bar{Y}$.

Strategy II': Sample units one by one at random and with replacement, governed by the stopping rule:

$$\tau^* = \min\{\nu_n \geq 2 : \nu_n \geq W^{-1}A(\frac{N-\nu_n}{N} s_{(\nu_n)}^2 + \nu_n^{-\gamma})\}. \qquad (3.2)$$

Here, $s_{(\nu_n)}^2$ is the sample variance based on ν_n distinct units (with division ν_n-1) and γ is an arbitrary positive constant. Consider $\bar{y}_{(\tau^*)}$, the mean per distinct units, as the (sequential) estimator of $\bar{Y}$.

Strategy III': Sample units one by one at random and without replacement along the stopping rule:

$$\tau^{**} = \min\{n \geq 2 : n \geq W^{-1}A(\frac{N-n}{N} s_n^2 + n^{-\gamma})\}. \qquad (3.3)$$

Here, s_n^2 and γ are, respectively, the same as they were in (3.1). Consider $\bar{y}'_{(\tau^{**})}$ as the (sequential) estimator of $\bar{Y}$.

In the case of σ^2 unknown, whichever strategy is adopted, the sample size is a random variable, and, consequently, the properties of the estimators of $\bar{Y}$ may change. As a matter of fact, it is not difficult to verify that none of the above estimators remain unbiased in a general setup. Thus, it may be more pertinent to compare $Ec(\tau_0)$, $cE(\tau^*)$ and $cE(\tau^{**})$ along with the mean squared errors (MSE) of the sequential estimators $\bar{y}_{\tau_0}$, $\bar{y}_{(\tau*)}$ and $\bar{y}'_{(\tau**)}$.

Regarding Strategies I' and II', the former involves the sample variance based on all the observations, while the latter is based on $s^2_{(\nu_n)}$ (i.e., on the distinct units only), although both are adapted to the SRSWR. Bahadur (1954) has pointed out that in sequential decision problems, attention can be confined to sequential decision (including stopping/action) rules which depend at each stage, n, on a *transitive sufficient* sub-sigma field B^*_n whenever the latter exists. We show in the Appendix that under SRSWR (in a sequential setup) there exists a *minimal sufficient transitive sequence* $\{B^*_n, n \geq 1\}$ based on the distinct units in the sample. This means that in the definition of the stopping rule, attention can be concentrated on the use of ν_n and $s^2_{(\nu_n)}$, instead of the s^2_n. Consequently, Strategy II' is more relevant. Thus, we would advocate the use of II' instead of I'.

Next, we come to the comparison of Strategy II' and III'. Let us examine the two stopping rules $\tau*$ and $\tau**$ in (3.2) and (3.3). Recall that ν_n ($\leq n$) is a positive integer valued random variable (in a SRSWR), while n in a SRSWOR is non-random. However, using Hájek's (1964) *rejective sampling* (equal probability) scheme, we may equivalently reduce the SRSWOR to SRSWR with distinct units only if we consider a sequence $\{M_n, n \geq 1\}$ of integer valued random variables, defined by

$$M_n = \min\{k \geq 1 : \nu_k = n\}, \quad n \geq 1. \tag{3.4}$$

As such, we may write $\bar{y}_n' \underline{\underline{D}} \bar{y}_{(\nu_{M_n})}$ for all $n \geq 1$. Here, we write $U \underline{\underline{D}} V$ to mean that the random variables U and V have identical distributions. A similar distributional identity holds for s_n^2 (in SRSWOR) and $s^2_{(\nu_{M_n})}$. Consequently, by (3.2) and (3.3), we may conclude readily that

$$\tau^* \underline{\underline{D}} \ \tau^{**} \text{ and } \ \bar{y}_{(\tau^*)} \underline{\underline{D}} \bar{y}'_{(\tau^{**})}, \tag{3.5}$$

so that Strategy II' and III' share the common properties. This feature is not surprising, as in (3.2) and (3.3) we have used essentially the same stopping rule. In the case of σ^2 known, the situation was slightly different [as ν_n was random while $\nu_{M_n} = n$ was not]. Looking at (2.3) [and (2.16)], we may as well consider a modified stopping rule

$$\tau_0^* = \min\{\nu_n \geq 2 : W/A \geq (s^2_{(\nu_n)} + \nu_n^{-\gamma})(E\nu_n^{-1} - N^{-1})\} \tag{3.6}$$

and propose the distinct mean estimator $\bar{y}_{(\tau_0^*)}$ for $\bar{Y}$. The stopping rule in (3.6) may still be motivated by the transitive sufficiency of $\{B_n^*, n \geq 1\}$ based on the distinct units (in SRSWR) and, apparently, this is more in line with (2.3) [than (3.2)]. With this stopping rule in (3.6), the distributional equivalence results in (3.5) do not hold (when τ^* is replaced by τ_0^*). However, as in the case of σ^2 known, here also τ^{**} (or $\bar{y}'_{\tau^{**}}$) may not *always* dominate τ_0^* (or $\bar{y}_{\tau_0^*}$). Towards this, we consider the following numerical example which shows that we can simultaneously realize (i) $E(\nu_{\tau_0^*}) < E(\tau^{**})$ and (ii) $MSE(\bar{y}_{\tau_0^*}) < MSE(\bar{y}'_{\tau^{**}})$. In other words, we demonstrate that τ_0^* may indeed be better than τ^{**} in some cases.

Example: We take N = 5 and choose A = W, $\gamma = 1$. Let the variate values be $Y_1 = 0$, $Y_2 = 1$, $Y_3 = 1.2$, $Y_4 = 1.4$, $Y_5 = 2.5$ so that $\bar{Y} = 1.22$.

Strategy III': Stopping Rule τ^* in (3.3). Samples: $\{(i\ j) | 1 \leq i \neq j \leq 5$ except (15) and (51)$\}$, (152), (153), (154), (512), (513) and (514) where i,j, etc., refer to the labels of the sampled units. Now $E(\tau^{**}) = 2.10$, $E(\bar{y}'_{\tau^{**}}) = 1.2183$ and $E(\bar{y}'_{\tau^{**}} - \bar{Y})^2 = .240814$.

Strategy II': Stopping Rule τ_0^* in (3.6). Samples: $\{(i\ j),(iij),(iiij),\ldots$ for all (ij), $i \neq j$, except (5 1) and (1 5)$\}$, (1 5 1), (1 5 2), (1 5 3), (1 5 4), (1 5 5), (5 1 1), (5 1 2), (5 1 3), (5 1 4) and (5 1 5). Now, $E(\tau_0^*) = 2.048$, $E(\bar{y}_{\tau_0^*}) = 1.2192$ and $E(\bar{y}_{\tau_0^*} - \bar{Y})^2 = .240697$.

We hereby conclude that τ_0^* provides smaller Bias, ASN and MSE compared to τ^{**}.

APPENDIX: MINIMAL SUFFICIENCY AND TRANSITIVITY IN SRSWR

Let $\mathcal{N} = \{1,\ldots,N\}$, $\mathcal{Y} = \{Y_1,\ldots,Y_N\}$, $\bar{Y} = N^{-1}\Sigma_{i\varepsilon\mathcal{N}}Y_i$

$\bar{y}_n = n^{-1}\Sigma_{i\varepsilon\mathcal{N}}f_{ni}Y_i$, f_{ni} = # of times the index i appears in the sample $S_n = (s_1,\ldots,s_n)$

$n = \Sigma_{i\varepsilon\mathcal{N}}f_{ni}$, $\quad \mathcal{A}_n = \mathcal{A}(f_{n1},\ldots,f_{nN})$, $n \geq 1$ (increasing).

Let

$$g_{ni} = \begin{cases} 1, & \text{if } f_{ni} \geq 1 \\ 0, & \text{if } f_{ni} = 0 \end{cases}, \quad 1 \leq i \leq N$$

$\nu_n = \Sigma_{i\varepsilon\mathcal{N}}g_{ni}$ $(\leq n)$, $\mathcal{B}_n = \mathcal{B}(g_{n1},\ldots,g_{nN})$, $n \geq 1$ ($\nearrow$)

$\bar{y}_{(\nu_n)} = \nu_n^{-1}\Sigma_{i\varepsilon\mathcal{N}}g_{ni}Y_i$, $\mathcal{B}_n \subset \mathcal{A}_n$, $\forall\, n \geq 1$.

$\mathcal{N} = \mathcal{N}_{n0} \cup \mathcal{N}_{n1}$, $\quad (\Sigma_{i\varepsilon\mathcal{N}_{n1}}f_{ni} = n)$

$\mathcal{N}_{n0} = \{i \;\varepsilon\; \mathcal{N} : g_{ni} = 0\}$,

$\mathcal{N}_{n1} = \{i \;\varepsilon\; \mathcal{N} : g_{ni} = 1\}$, $\quad \mathcal{N}_{n0} \cap \mathcal{N}_{n1} = \phi$.

Cardinality of $\mathcal{N}_{n1} = \nu_n$, Cardinality of $\mathcal{N}_{n0} = N - \nu_n$.

$\mathcal{B}_n^* = \mathcal{B}(\nu_n, \mathcal{N}_{n1}, \mathcal{N}_{n0})$ $\quad (\subset \mathcal{B}_n \subset \mathcal{A}_n)$.

Note that conditional on $\mathcal{B}_n^*$, the joint probability function of $f_n = (f_{n1},\ldots,f_{nN})$ remains invariant under any permutation of the ν_n indices $\{i \;\varepsilon\; \mathcal{N}_{n1}\}$ among themselves and $N - \nu_n$ indices $\{i \;\varepsilon\; \mathcal{N}_{n0}\}$ among themselves. Therefore,

$$
\begin{aligned}
E\{\bar{y}_n|B_n^*\} &= n^{-1}\Sigma_{i\epsilon N}Y_iE(f_{ni}|B_n^*) \\
&= n^{-1}\Sigma_{i\epsilon N_{n1}}Y_iE(f_{ni}|B_n^*) + n^{-1}\Sigma_{i\epsilon N_{n0}}Y_iE(f_{ni}|B_n^*) \\
&= n^{-1}\Sigma_{i\epsilon N_{n1}}Y_iE(f_{ni}|B_n^*) + 0 \\
&= n^{-1}\Sigma_{i\epsilon N_{n1}}Y_i\{\nu_n^{-1}\Sigma_{j\epsilon N_{n1}}f_{nj}\} \\
&= n^{-1}\Sigma_{i\epsilon N_{n1}}Y_i\{\nu_n^{-1}n\} \\
&= \nu_n^{-1}\Sigma_{i\epsilon N_{n1}}Y_i \\
&= \nu_n^{-1}\Sigma_{i\epsilon N}Y_ig_{ni} \\
&= \bar{y}_{(\nu_n)}, \qquad \forall\ n \geq 1.
\end{aligned}
$$

Therefore, by the Rao-Blackwell theorem, for any convex loss $L(a,b)$,

$$EL(\bar{y}_n, \bar{Y}) \geq EL(\bar{y}_{(\nu_n)}, \bar{Y}).$$

In particular,

$$E(\bar{y}_n - \bar{Y})^2 \geq E(\bar{y}_{(\nu_n)} - \bar{Y})^2, \quad \forall\ n \geq 1.$$

We write

$$\underset{\sim}{f}_n' = (f_{n1},\ldots,f_{nN}), \text{ so that } \underset{\sim}{1}_N' \underset{\sim}{f}_n = n, \quad \forall\ n \geq 1.$$

Let $\underset{\sim}{f}_{n+1} = \underset{\sim}{f}_n + \underset{\sim}{u}_{n+1}$, $\underset{\sim}{u}_{n+1}$ independent of $\underset{\sim}{f}_n$ and $\underset{\sim}{1}_N'\underset{\sim}{u}_{n+1} = 1$, $\forall\ n \geq 0$.

$$\underset{\sim}{g}_n' = (g_{n1},\ldots,g_{nN}), \ \underset{\sim}{g}_n'\underset{\sim}{1}_N = \nu_n, \ n \geq 1.$$

$\underset{\sim}{g}_{n+1} = \underset{\sim}{g}_n + \underset{\sim}{v}_{n+1}$, where the distribution of $\underset{\sim}{v}_{n+1}$ depends only on ν_n and $\underset{\sim}{g}_n$. Thus, given B_n^*, the distribution of $\underset{\sim}{f}_n$ is generated by the $\nu_n!(N - \nu_n)!$ possible equally likely permutations, while $\underset{\sim}{v}_{n+1}$ can be a null vector with probability $N^{-1}\nu_n$ and a non-null vector with probability $(1 - N^{-1}\nu_n)$, there being $N - \nu_n$ equally likely realizations: $\underset{\sim}{v} = (v_1,\ldots,v_N) : v_i = 0$ for all but one i and $\underset{\sim}{g}_n' \underset{\sim}{v} = 0$. Thus, given B_n^*, $\underset{\sim}{f}_n$ and $\underset{\sim}{v}_{n+1}$ (i.e., A_n and B_{n+1}^*) are conditionally independent. Therefore, $\{B_n^*, n \geq 1\}$ is a transitive sufficient sequence.

Let $B_n^{o*} = B_1^* \vee \ldots \vee B_n^*$ be the smallest sigma field containing $B_1^*,\ldots,B_n^*$ for $n \geq 1$. Then the events $[\tau^* = t]$ (or $[\tau = t]$) are B_n^{o*}-measurable. Hence, if we are able to show that

$$E[\bar{y}_n | B_n^{o*}] = E(\bar{y}_n | B_n^*) = \bar{y}_{(\nu_n)}, \quad \forall\, n \geq 1,$$

then we would have for a B^{o*}-measurable stopping time M

$$EL(\bar{y}_M, \bar{Y}) = \Sigma_{m \geq 1} E\{L(\bar{y}_m, \bar{Y}) | M = m\} P\{M = m\}$$

where

$$E[L(\bar{y}_m, \bar{Y}) | M = m] \geq E[L(\bar{y}_{(\nu_m)}, \bar{Y}) | M = m], \quad \forall\, m \geq 1,$$

so that

$$EL(\bar{y}_m, \bar{Y}) \geq EL(\bar{y}_{(\nu_m)}, \bar{Y}).$$

Towards this note that

$$E[\bar{y}_n | B_n^{o*}] = n^{-1} \Sigma_{i \in N_{n1}} Y_i E(f_{ni} | B_n^{o*})$$

$$= n^{-1} \Sigma_{i \in N_{n1}} Y_i \Sigma_{k=1}^{n} E(u_{ki} | B_n^{o*}).$$

As $\underset{\sim}{u}_k$ is independent of B^*_{k-j}, $j \geq 1$, we have

$$E(u_{ki} | B^{o*}_n) = E(u_{ki} | B^*_k \vee \ldots \vee B^*_n).$$

Note that ν_k as well as N_{k1} are nondecreasing in k, and, hence, $N_{k1} \subseteq \cdots \subseteq N_{n1}$, $\forall\ k \leq n$. Thus, given that $i \in N_{n1}$, under $B^*_k \vee \ldots \vee B^*_n$, i will also belong to N_{k1} with conditional probability ν_k/ν_n, $\forall\ k \leq n$. On the other hand, for every $i \in N_{k1}$,

$E(u_{ki} | B^*_k \vee \ldots \vee B^*_n) = \nu_k^{-1} \Sigma_{i \epsilon N_{k1}} E(u_{ki} | \ldots)$

$= \nu_k^{-1} E(\Sigma_{i \epsilon N_{k1}} u_{ki} | B^*_k \vee \ldots \vee B^*_n) = \nu_k^{-1} E(1 | \ldots) = \nu_k^{-1}$. Therefore,

for every $i \in N_{n1}$, $k \leq n$,

$$E(u_{ki} | B^{o*}_n) = (\nu_k/\nu_n)\nu_k^{-1} = \nu_n^{-1}, \text{ and}$$

$$\begin{aligned} E(\bar{y}_n | B^{o*}_n) &= n^{-1} \Sigma_{i \epsilon N_{n1}} Y_i \Sigma^n_{k=1} \nu_n^{-1} \\ &= \nu_n^{-1} \Sigma_{i \epsilon N_{n1}} Y_i \\ &= \bar{y}_{(\nu_n)}, \qquad \forall\ n \geq 1. \end{aligned}$$

This characterizes the minimum risk property of the sequential $\bar{y}_{(\nu_M)}$ for B^{o*}-measurable M.

REFERENCES

Asok, C. (1980), "A Note on the Comparison Between Simple Mean and Mean Based on Distinct Units in Sampling with Replacement," *The American Statistician, 34,* 158.

Bahadur, R. R. (1954), "Sufficiency and Statistical Decision Functions," *Annals of Mathematical Statistics*, *25*, 423-462.

Basu, D. (1958), "On Sampling With and Without Replacement," *Sankhya, Series A*, *20*, 287-294.

Chakrabarti, M. C. (1965), "Query and Solutions," *Journal of Indian Statistical Association*, *3*, 223-232.

Chow, Y. S. and Robbins, H. (1965), "On the Asymptotic Theory of Fixed-Width Sequential Confidence Intervals for the Mean," *Annals of Mathematical Statistics*, *36*, 457-462.

Ghosh, M. and Mukhopadhyay, N. (1979), "Sequential Point Estimation of the Mean When the Distribution Is Unspecified," *Communications In Statistics, Series A*, *8*, 637-652.

Hájek, J. (1964), "Asymptotic Theory of Rejective Sampling with Varying Probabilities from a Finite Population," *Annals of Mathematical Statistics*, *35*, 1491-1523.

Korwar, R. M. and Serfling, R. J. (1970), "On Averaging Over Distinct Units in Sampling with Replacement," *Annals of Mathematical Statistics*, *41*, 2132-2134.

Pathak, P. K. (1961), "On the Evaluation of Moments of Distinct Units in a Sample," *Sankhya, Series A*, *23*, 415-420.

Raj, D. and Khamis, S. H. (1958), "Some Remarks on Sampling with Replacement," *Annals of Mathematical Statistics*, *29*, 550-557.

Robbins, H. (1959), "Sequential Estimation of the Mean of a Normal Population," *Probability and Statistics (H. Cramêr Volume), Ulf Grenander edited,* 235-245. Almquist and Wiksell, Uppsala, Sweden.

Table 1. Comparisons of n** and $E(\nu_{n*})$

N	B^{-1}	n**	n*	$E(\nu_{n*})$
5	2.1	3	3	2.44
	2.5	3	4	2.95
	2.6	3	4	2.95
	2.8	3	5	3.36
	3.1	4	5	3.36
	3.5	4	6	3.69
	3.6	4	7	3.95
	3.7	4	7	3.95
	3.8	4	7	3.95
	3.85	4	8	4.16
6	3.2	4	5	3.59
	3.4	4	5	3.59
	3.6	4	6	3.99
	3.8	4	6	3.99
	4.2	5	8	4.60
	4.4	5	8	4.60
	4.6	5	9	4.84
	4.8	5	10	5.03