# ON UNBIASED PRODUCT-TYPE ESTIMATORS

By
ARIJIT CHAUDHURI AND RAGHUNATH ARNAB
*Indian Statistical Institute, Calcutta*
(Received : December, 1979)

SUMMARY

**Basically following Quenouille's well-known technique of bias-reduction by splitting samples into random sub-samples Shukla's (1976) results are extended and modified to find an exactly unbiased product-type estimator equally efficient (to the second order of approximation) as a product estimator (used when a negatively correlated auxilliary variable is available).**

## 1. INTRODUCTION

Recently Shukla (1976) used Quenouille's technique of random splitting of a sample in deriving an asymptotically unbiased (to the first order of approximation) product-type estimator for a finite population total (when data on a negatively correlated auxiliary variate are available) equally efficient (to the second order of approximation) as a product estimator itself. Here we extend his method with a slightly more general set-up (allowing (*i*) sub-samples of unequal sizes (*ii*) sampling with replacement-the latter having only an academic rather than a practical interest specially because there may not be enough distinct units for effective grouping) and find an 'exactly' unbiased product type estimator with no loss of efficiency to the second order of approximation.

## 2. FORMULATION OF THE PROBLEM AND THE RESULTS

Let a finite population $U$ have units $u_i$ with values $Y_i$, $X_i$ $(i=1, ..., N)$ for two negatively correlated variables $y$ and $x$ respectively. Let $s$ be a sample of size $n$ (units may or may not be distinct) from $U$ taken using any selection-scheme. Let $s$ be divided at random (without replacement) into $k$ mutually exclusive and exhaustive groups (sub-samples) the $j$th group $s_j$ comprising $n_j$ units, not necessarily distinct. If $n_i(s)$ denotes the frequency of $i$th unit in $s$, then we have the conventional product estimator of

$$Y=\sum_1^N Y_i \text{ as } t=\frac{t_1\, t_2}{X} \quad \text{where } X=\sum_1^N X_i,$$

$$t_1=\sum_1^N b_{si}\, n_i(s)\, Y_i$$

$$t_2=\sum_1^N b_{si}\, n_i(s)\, X_i$$

with $b_{si}$'s independent of $Y_i$'s, $X_i$'s and chosen such that $t_1$ and $t_2$ are respectively unbiased for $Y$ and $X$ with $b_{si}=0$ for $i \notin s$. For the $r$th draw ($r=1, \ldots, n$) let us write

$$u_r=b_{si}Y_i, \; v_r=b_{si}\, X_i$$

in case $i$th unit ($i=1, \ldots, N$) is realised,

$$\bar{u}_j=\frac{1}{n_j}\, \Sigma_j\, u_r,$$

$$\bar{v}_j=\frac{1}{n_j}\, \Sigma_j\, v_r,\; j=1, \ldots, k$$

($\Sigma_j$ denoting the sum over the $n_j$ units falling in $s_j$, including repetitions, if any),

$$\bar{u}=\frac{1}{n}\sum_1^n u_r, \quad \bar{v}=\frac{1}{n}\sum_1^n v_r, \quad t_{3j}=n\, \bar{u}_j, \quad t_{4j}=n\, \bar{v}_j,$$

$$t_j'=\frac{t_{3j}\, t_{4j}}{X}$$

(the product estimator of type $t$ based on $s_j$). For any set of chosen weights $w_j$'s ($j=1, \ldots, k$) we will consider an estimator for $Y$ of the form

$$\bar{t}=\sum_1^k w_j\, t_j'+\left(1-\sum_1^k w_j\right) t$$

Writing $e=\dfrac{t_1-Y}{Y}$, $e'=\dfrac{t_2-X}{X}$, $e_j=\dfrac{t_{3j}-Y}{Y}$, $e_j'=\dfrac{t_{4j}-X}{X}$, and noting that $0=E(e)=E(e')=E(e_j)=E(e_j')\ \forall\ j=1, \ldots, k$, the bias of $\bar{t}$ works out as

$$B(\bar{t})=Y\Big[\sum_1^k w_j E(e_j e_j')+(1-\Sigma w_j) E(ee')\Big]$$

implying, on assuming $E(ee')\neq 0$ that

$$\sum_1^k w_j c_j+(1-\Sigma_j w_j)=0 \qquad \ldots(2.1)$$

is a necessary and sufficient condition for unbiasedness of $E$, where $c_j=E(e_j e_j')/E(ee'), j=1, \ldots, k$.

The mean square Error ($MSE$) of $\bar{t}$ is

$$\bar{M}(t)=Y^2 E[\sum_j w_j (e_j+e_j'+e_j e_j')+(1-\sum_j w_j)(e+e'+ee')]^2$$

Assuming $E e_j^r e_l'^q$ and $E e^r e'^q$ to be negligible when $0<r, q$, such that $r+q>2$ and $j, l=1, \ldots, k$, to be called assumption $A_1$ we have approximately (to be called second order of approximation)

$$M(\bar{t})\simeq \bar{M}(\bar{t})\ (\text{say})=Y^2 E\Big[\sum_1^k w_j (e_j+e_j')+\Big(1-\sum_1^k w_j\Big)(e+e')\Big]^2$$

$$=Y^2 E(z^2), \text{ with obvious notation for } z.$$

Writing $g(s)=E(z \mid s)$, the conditional (given $s$) expectation of $z$ it follows that

$$E(z)=Eg(s)=0$$

and that

$$\bar{M}(\bar{t})\geqslant Y^2 E[g(s)]^2=Y^2 E(e+e')^2=MSE \text{ of } t.$$

Thus under second order approximations with assumptions $A_1$, $\bar{t}$ cannot have $MSE$ less than that of $t$ itself.

More importantly, to the same order of approximations, $\bar{M}(\bar{t})=M(t)$ if and only if $w_i$'s are chosen such that

$$\sum_1^k w_j (e_j+e_j')=(e+e')\sum_1^k w_j \qquad \ldots(2.2)$$

with probability one.

Noting that $$(e+e') = \sum_{1}^{k} \frac{n_j}{n} (e_j+e'_j)$$

(2.2) is equivalent to

$$\sum_{1}^{k} w_j (e_j+e'_j) = \left\{ \sum_{1}^{k} \frac{n_i}{n} (e_j+e'_j) \right\} \sum_{j}^{k} w_j \quad \ldots(2.3)$$

If one chooses $$w_j = \frac{n_i}{n} \sum_{1}^{k} w_j \;\forall\; j=1, \ldots, k, \quad \ldots (2.4)$$

Then (2.3) holds leading to.

*Theorem 1*: $\bar{t}$ with $w_j$'s satisfying (2.1) and (2.4) is unbiased for $Y$ with an $MSE$ equal (to the second order of approximation under $A_1$) to that of $t$.

*Remark*. For usability $w_j$'s and hence $c_j$'s must be independent of $Y_i$'s and $X_i$'s. This need is met in particular at least, if $s$ is chosen either by SRSWOR or PPSWR method as shown below.

If $s$ is an SRSWOR, then, with a little algebra,

$$c_j = 1 + (1/n_j - 1/n)/(1/n - 1/N) \;\forall\; j=1, \ldots, k$$

leading to

$$w_j = \frac{n_j}{n} \sum_{j} w_j' \frac{1}{(1/n - 1/N)} \sum_{j} w_j (1/n_j - 1/n) + 1 = 0$$

whence $w_j'$ s free $Y_i'$ s and $X_i'$ s may be worked out. In case $n_j = n/k \;\forall\; j=1, \ldots k$, we have the choice

$$w_j = -\frac{1}{k} \frac{1}{(k-1)} \frac{N-n}{N} \;\forall\; j=1, \ldots, k.$$

For $k=2$, if $n_j = n/2$, then we have to choose $w_j = -\frac{N-n}{2N}$ (This agrees with Shukla's (1976) result).

If $k=2$, but $n_j$ are unequal, then $\bar{t}$ becomes unbiased with $\bar{M}(\bar{t})$ equalling $M(t)$ subject to $A_1$ if one takes $w_j = -n_j/n. \frac{N-n}{N}$ for $j=1, 2$.

If $s$ is chosen by PPSWR method in $n$ draws with known $p_i's$ as normed positive size-measures then taking as usual $b_{si}=1/np_i \ \forall \ i$ in $s$ we have

$$c_j = 1+n^2\left(\frac{1}{n_j}-\frac{1}{n}\right)\frac{1}{n-1}\,\frac{E\,\Sigma\,(u_r-\bar{u})\,(v_r-\bar{v})}{E\,(t_1-Y)\,(t_2-X)}$$

$$=n/n_j \text{ for } j=1,\ldots,k$$

[For a proof see appendix]

If in particular. $n_j=n/k \ \forall \ j$ we have

$$c_j=k \ \forall \ j \text{ and } w_j=-\frac{1}{k(k-1)}$$

and with this choice $\bar{t}$ will be unbiased for $Y$ with least MSE (under second order approximation with $A_1$) equalling MSE ($t$).

If $k=2$, but $n_j's$ are unequal, then the best choice of $w_j's$ is $w_j=-n_j/n \ \forall \ j$. If $k=2$ and $n_j's$ are equal, then $w_j-1/2, j=1, 2$.

ACKNOWLEDGEMENT

The authors are grateful to the referees for their comments and suggestions that led to an improvement on an earlier draft.

REFERENCE

Shukla, N.D (1976) : Almost unbiased product type estimator, *Metrika*, **23**, Fase 3, 127-133.

For PPSWR, $E\sum_{1}^{n}(u_r-\bar{u})(v_r-\bar{v})$

$$=E\sum_{1}^{n}\left(\frac{Y_r}{np_r}-\frac{1}{n}\sum_{1}^{n}\frac{Y_r}{p_r}\right)\left(\frac{X_r}{np_r}-\frac{1}{n}\sum_{1}^{n}\frac{X_r}{p_r}\right)$$

$$=\frac{1}{n}\sum_{1}^{N}\frac{Y_i X_i}{p_i}-\frac{1}{n}\text{Cov}\left(\frac{1}{n}\sum_{1}^{n}\frac{Y_r}{p_r},\frac{1}{n}\sum_{1}^{n}\frac{X_r}{p_r}\right)-\frac{1}{n}XY$$

$$=(1-1/n)\text{ Cov}\left(\frac{1}{n}\sum_{1}^{n}Y_r/p_r\, 1/n\sum_{1}^{n}X_r/p_r\right)$$

$$=\frac{n-1}{n}E(t_1-Y)(t_2-X)$$

and hence $c_j=n/n_j$ .