A Unified Theory of Sampling from Finite Populations

By V. P. Godambe

Bureau of Economics and Statistics, Government of Bombay, Bombay

[Received August, 1954]

SUMMARY

The most general type of linear estimate is defined for a general sampling design. It is demonstrated that an unbiased linear estimate with least variance does not exist uniquely for the entire class of linear estimates. A slightly modified criterion for best estimate, with reference to certain given prior information is proposed. Some sampling designs have been analysed accordingly.

1. *Introduction*

Let x be a variate defined over a finite population of N individuals, x_λ being the value of x associated with the λ^{th} individual in the population ($\lambda = 1, \ldots, N$). If it is proposed to estimate the total

$$T = \sum_{\lambda=1}^{N} x_\lambda$$

by making n successive random drawings, we get a very general type of sampling design when we allow the probabilities of drawing different individuals at any particular draw to depend upon the out-comes of the earlier draws. Thus let p be a function with arguments (q, s_{q-1}, λ) such that

$$p(q, s_{q-1}, \lambda)$$

is the probability of drawing the λ^{th} individual ($\lambda = 1, \ldots, N$) at the q^{th} draw ($q = 1, \ldots, n$) when s_{q-1} is the sequence of individuals that turn up in the first $q - 1$ draws, s_0 denoting the absence of earlier drawings.

Then we call p a sampling design. It is important to note that p is given *a priori*, i.e. before any drawing is made.

For instance if $n = 2$ and p is such that $p(1, s_0, \lambda) = 0$ for all $\lambda > N'$ and $p(2, s_1, \lambda) = 0$ for all $\lambda \leqslant N'$ and all s_1, then p represents a sampling design of stratification, the first N' individuals forming one stratum and the remaining ones the other. Obviously almost all the known sampling designs could be expressed in a similar way.

We denote by s_n the sequence of individuals that turn up in n successive drawings, i.e. a sample, or by s simply. Once p is given, the probability of s turning up, P_s say, is also uniquely determined for all s. The total number of such s that could turn up logically (though perhaps not practically because of P_s being zero for some s) is N^n.

Conversely, however, if P_s is given for all s, p may or may not exist. Yet for a theoretical discussion it would sometimes be convenient to ignore the latter possibility and concentrate on P_s for all s.

For a given sampling design p there usually exists a wide class of linear, unbiased estimates of T. Only some of these estimates are practically serviceable. In the present paper a new criterion for the *bestness* of an estimate is suggested and some sampling designs have been investigated accordingly.

2. *The Most General Type of Linear Estimate*

Many of the usually employed linear estimates e_s based on s, i.e. based on the values of x associated with the different individuals that turn up in n successive drawings, constituting s, can be put into three classes (Horwitz and Thomson, 1952):

(i)
$$e_s = \sum_{\lambda \in s} \beta_\lambda x_\lambda$$

where $\sum_{\lambda \in s}$ stands for the summation over all the different individuals in s (these may be, in number, $\leqslant n$), β_λ being defined in advance for all $\lambda = 1, \quad . \ . \ . \quad , N$.

(ii)
$$e_s = \sum_{q=1}^{n} \beta_q x_q.$$

Here x_q is the value of x associated with the individual that turns up at the q^{th} draw, β_q being defined in advance for $q = 1, \quad . \ . \ . \quad , n$.

(iii)
$$e_s = \beta_s \sum_{\lambda \in s} x_\lambda.$$

In this case β_s is defined in advance for all s.

This classification can be illustrated as follows: When the population is stratified and in each stratum sampling is carried out with equal probabilities and without replacement the usual estimate of T,

$$e_s = \sum_{\kappa} \frac{N(\kappa)}{n(\kappa)} \sum_{\lambda \in s(\kappa)} x_\lambda$$

where $s(\kappa)$ denotes the sample from the κ^{th} stratum, $N(\kappa)$ and $n(\kappa)$ being the sizes of the population and sample respectively in that stratum, clearly belongs to class (i). As a special case for a simple random sampling (from an unstratified population) without replacement and with equal probabilities of selection,

$$e_s = \frac{N}{n} \sum_{\lambda \in s} x_\lambda$$

also belongs to class (i). Again, if the sampling is carried out with equal probabilities but with replacement one can form an unbiased estimate of T which belongs to this class as

$$e_s = \sum_{\lambda \in s} x_\lambda \Big/ \left[1 - \left(1 - \frac{1}{N}\right)^n\right]$$

It is interesting to note that when sampling is carried out with equal probabilities and with replacement the usual estimate of T

$$e_s = \frac{N}{n} \sum_{q=1}^{n} x_q$$

belongs to class (ii) and not to class (i). Yet it is difficult to imagine that any practically important estimate could belong to class (ii) strictly. Even the estimate just now referred to is a special case of a more general type of estimate, which, as will be seen later, in its general form does not belong to class (ii).

To the class (iii) belongs the well known ratio estimate. If y is a correlated variable the corresponding estimate of T is

$$e_s = \left(\sum_{\lambda=1}^{N} y_\lambda\right) \frac{\sum_{\lambda \in s} x_\lambda}{\sum_{\lambda \in s} y_\lambda}.$$

In this case

$$\beta_s = \sum_{\lambda=1}^{N} y_\lambda \Big/ \sum_{\lambda \in s} y_\lambda*.$$

* The β_s in this case are not known in advance as required by the definition of class (iii), because $\sum_{\lambda \in s} y_\lambda$ are not listed before sampling. However, the logical possibility of their being known in advance is enough for the theoretical discussion.

That the above classification is not exhaustive can be easily demonstrated. Let sampling be carried out with replacement and the probability of selection for the λ^{th} individual be a_λ, i.e. fixed for all the successive drawings. Then an unbiased estimate of T is

$$e_s = \frac{1}{n} \sum_{q=1}^{n} x_q/a_q.$$

This does not belong to either of the classes (i), (ii), or (iii), only as a special case when all a_λ are equal it belongs to class (ii), as stated already. The coefficient of x_λ in e_s here is partly determined *a priori* inasmuch as the a_λ are given in advance and partly it is determined by the number of times the λ^{th} individual is selected in n successive drawings, which constitute s. Again, for sampling without replacement, let $n = 2$ and the first drawing be made with probability of selection a_λ for the λ^{th} individual ($\lambda = 1, \ldots, N$). Suppose the λ^{th} individual is selected at the first draw, then the second drawing is made from the remaining individuals with the probability $a_{\lambda}'/1 - a_\lambda$ of selection for the λ'^{th} individual, for all individuals $\lambda' \neq \lambda$. Then an unbiased estimate of T is given by (Das, 1951)

$$e_s = \frac{1}{2} \left\{ \frac{x_\lambda}{a_\lambda} + \frac{1}{N-1} \cdot \frac{1 - a_\lambda}{a_\lambda} \cdot \frac{x_\lambda'}{a_\lambda'} \right\}.$$

The coefficient of an individual x-value in e_s here is partly determined *a priori* so far as the a_λ are given in advance, partly by the order in which the individual has been drawn and also by the individual, if any, that precedes it.

The above discussion suggests that the most general type of linear estimate of T may be defined as

$$e_s = \sum_{\lambda \in s} \beta_{s\lambda} x_\lambda \qquad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (1)$$

where $\beta_{s\lambda}$ is defined in advance for all the logically possible s which in all are N^n, and for all $\lambda \in s$. It is easy to see that estimates in classes (i), (ii) and (iii) are particular cases of this estimate and the same remark holds for all the known linear estimates.

3. *Unbiasedness and Least Variance*

It has already been observed that once a sampling design p is given, the probability P_s of outcome, of every logically possible s is uniquely determined. Then in a given p, e_s is said to be an unbiased estimate of T if

$$E(e_s) = T,$$

i.e.

$$\sum_s e_s P_s = T$$

for whatever x, $\sum_s$ standing for summation over all s. Hence the necessary and sufficient condition for the unbiasedness of e_s in (1) is

$$\sum_{s \supset \lambda} \beta_{s\lambda} P_s = 1 \qquad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (2)$$

for $\lambda = 1, \ldots, N$; $\sum_{s \supset \lambda}$ standing for summation over all s which include the λ^{th} individual. Now let B_p denote the class of all β for which the corresponding estimates $e_s(x, \beta)$ are unbiased, i.e. every member of B_p satisfies (2). Then one would usually define $e_s(x, \bar{\beta})$ as the best linear estimate of T provided that $\bar{\beta} \in B_p$ and

$$\text{Variance } e_s(x, \bar{\beta}) \leqslant \text{Variance } e_s(x, \beta)$$

for all $\beta \in B_p$ whatever x may be. That such a $\bar{\beta}$ does not exist is, however, intuitively evident. A formal proof of its non-existence is as follows:

When e_s is unbiased its variance is given by

$$V(e_s) = \sum_s e_s^2 P_s - T^2. \qquad . \quad . \quad . \quad . \quad . \quad . \qquad (3)$$

Now suppose $\bar{\beta}$ exists. Then for a given x and $\lambda \varepsilon s$ we have

$$\left\{\frac{\partial}{\partial \beta_{s\lambda}} V(e_s) - \mu_\lambda \frac{\partial}{\partial \beta_{s\lambda s}} \sum_{\supset \lambda} \beta_{s\lambda} P_s\right\} \bar{\beta} = 0$$

where μ_λ is a Lagrangian Multiplier. It follows that

$$2\, x_\lambda \{e_s\}_{\bar{\beta}} = \mu_\lambda$$

for all $s \supset \lambda$ and where $P_s \neq 0$. Moreover this must hold for all x since $\bar{\beta}$ is supposed to give minimum for all x. That is for any s_1 and $s_2 \supset \lambda$; P_{s1}, P_{s2} being $\neq 0$

$$\{e_{s1}\}_{\bar{\beta}} = \{e_{s2}\}_{\bar{\beta}}$$

for all x. In particular, putting $x_\lambda = 1$ and $x_{\lambda}' = 0$ for all $\lambda' \neq \lambda$ we get

$$\bar{\beta}_{s\lambda} = \bar{\beta}_\lambda.$$

In addition, from (2)

$$\bar{\beta}_{s\lambda} = \bar{\beta}_\lambda = 1/\sum_{s \supset \lambda} P_s = 1/P(\lambda) \quad . \quad . \quad . \quad . \quad . \qquad (4)$$

for all s and λ, $P(\lambda)$ being the probability of λ^{th} individual being included in sample at least once. Hence if $\bar{\beta}$ exists it must necessarily be given by (4). Conversely $\bar{\beta}$ in (4) is the one minimizing $V(e_s)$ for all x if and only if it satisfies

$$\{e_{s1}\}_{\bar{\beta}} = \{e_{s2}\}_{\bar{\beta}}$$

for all s_1, $s_2 \supset \lambda$; P_{s1}, P_{s2} being $\neq 0$. Obviously $\bar{\beta}$ in (4) cannot satisfy this condition. Thus the conclusion that $\bar{\beta}$ does not exist.

This result can be illustrated. Let p be given by

$$p(q, s_{q-1}, \lambda) = p_\lambda$$

for $q = 1, \ldots, n$ and $\lambda = 1, \ldots, N$, i.e. the usual sampling design where sampling is done with replacement, p_λ; $\lambda = 1, \ldots, N$ being the probability of selection for the λ^{th} individual at different draws $q = 1, \ldots, n$. Thus in p the probability of the λ^{th} individual being selected in a sample at least once is

$$P(\lambda) = 1 - (1 - p_\lambda)^n.$$

Now we have two linear unbiased estimates of T viz.

$$e_s = \frac{1}{n} \sum_{q=1}^{n} x_q/p_q$$

and

$$e'_s = \sum_{\lambda \varepsilon s} x_\lambda/P(\lambda), \text{ i.e.} = \sum_{\lambda \varepsilon s} x_\lambda/\{1 - (1 - p_\lambda)^n\}.$$

Further according to (4) if a linear, unbiased and least variance estimate of T exists uniquely, in p it should be given by e'_s, i.e.

$$V(e'_s) \text{ should be less than } V(e_s)$$

for whatever x. This however is not the case* for in particular putting $p_\lambda = 1/N$ we have

$$V(e'_s) = \frac{(1 - 1/N)^n - (1 - 2/N)^n}{\{1 - (1 - 1/N)^n\}^2} \sum_{\lambda=1}^{N} x^2{}_\lambda - \frac{(1 - 1/N)^{2n} - (1 - 2/N)^n}{\{1 - (1 - 1/N)^n\}^2} T^2$$

and

$$V(e_s) = N^2\{\sum_{\lambda=1}^{N} x^2{}_\lambda - T^2/N\}/nN.$$

* This result is also evident since $V(e_s) = 0$ and $V(e'_s) > 0$ when $p_\lambda \propto x_\lambda$.

Now it can be seen that in fact

$$V(e'_s) > V(e_s)$$

if the coefficient of variation of x in the population is sufficiently small. This proves the non-existence of a unique linear, unbiased and least variance estimate for whatever x.

4. *Markoff Theorem and the Theory of Efficient Estimators*

The existing theory of the best (i.e. unbiased and least variance) linear estimates is generally based on Markoff's Theorem on least squares (Neyman, 1934; David, 1938). For instance to get the best linear estimate of the population mean $\bar{x}_N$ in a simple random sampling design, which consists of n drawings with equal probability and without replacement we proceed as follows. The expectation of the q^{th} draw $E(x_q)$, equals $\bar{x}_N$ and the variance of x_q is σ^2 $(= \sum_{\lambda} (x_\lambda - \bar{x}_N)^2/N,$ $\lambda = 1, \ldots, N)$. Then according to Markoff's Theorem the best linear estimate of $\bar{x}_N$ is $\hat{x}_N$ for which $\sum_{q} (x_q - \hat{x}_N)^2, q = 1, \ldots, n$ is minimum (Sukhatme, 1954).

Evidently the class of linear estimates considered above, while choosing the best estimate is one identical with the sub-class (ii) in section (2) of the *entire* class of linear estimates defined in equation (1). Incidently the best linear estimate $\hat{x}_N$ obtained above is also the best linear estimate in sub-classes (i) and (iii) of section (2) (Horwitz and Thomson, 1952). Similarly in case of stratified random sampling design the best linear estimate is obtained from a sub-class of linear estimates (Neyman, 1934). The ratio estimate is also found to be the best linear estimate in a sub-class of linear estimates (Cochran, 1953; Sukhatme, 1954). The estimate employed by Horwitz and Thomson (1952) $\sum_{\lambda \in s} x_\lambda/P(\lambda)$ is the only unbiased and hence the best linear estimate of sub-class (i).

Thus in the earlier theory simple *sub-classes* of (1) were considered to find the best linear estimate and in fact, as shown in section (3), best linear estimate does not exist uniquely for the *entire* class of linear estimates. However in section (5) we will define another criterion of *bestness* for the *entire* class of linear estimates and in later sections we shall prove that most of the frequently employed linear estimates are in this sense best estimates. This provides further justification for the existing practice.

5. *Unbiasedness and Least Expected Variance*

Now when for a given sampling design the attempts to secure a unique (i.e. for all x) linear, unbiased estimate with least variance fail, the next best thing that we can do is to search for a procedure of estimation which when employed repeatedly would secure on the average a least variance.

Intuitively some such criterion is quite frequently employed when the statistician expects some estimate to be efficient for one population and some other estimate for a different population. An exact statistical explanation for the above intuitive criterion in many of the situations usually arising may be as follows:

On the basis of past experience regarding several factors which influence the value of the variate under study (x in the present case), or because of the knowledge of the distribution of one or more correlated variates, the statistician often may have certain expectations of the values of x associated with different individuals in the population. The calculations of these expectations is a matter of statistical skill, in addition their *sharpness* (Bross, 1954) depends upon the degree of relevant knowledge on the part of the statistician.* These expectations are *a priori* expectations in the sense that they exist before any drawing is made for the present sample. Moreover these admit simple interpretations in terms of *a priori* probabilities, the corresponding interpretation in terms of frequency being mean values (Carnap, 1950).

Let us denote by

$$\varepsilon(x_1), \ldots, \varepsilon(x_N) \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

* A similar idea derived from the device of regarding the finite population (conditioned by ancillary variable) as a sample from an infinite one occurs in Yates (1950) and Cochran (1939).

the *a priori* expectations for different individuals $\lambda = 1, \ldots, N$ respectively.* It will be assumed in the subsequent discussion that these $\varepsilon(x_1), \ldots, \varepsilon(x_N)$ do not change because of drawing one or more individuals from the population and observing the values of x associated with them, for the present sample. This condition would obviously be violated if sampling is done with replacement, for then if the λ^{th} individual is selected at the first draw and the value of x viz. x_λ be noted, at the second and the subsequent draws its *a priori* expectation is no more $\varepsilon(x_\lambda)$ but is x_λ. Hence we shall confine ourselves henceforth to sampling without replacement. In addition if $\varepsilon(x_\lambda \,.\, x_{\lambda'}), = \varepsilon(x_\lambda / x_{\lambda'})\, \varepsilon(x_{\lambda'})$, denotes the *a priori* expectation of $x_\lambda \,.\, x_{\lambda'}$, $\varepsilon(x_\lambda/x_{\lambda'})$ being the conditional *a priori* expectation of x_λ given $x_{\lambda'}, (\lambda' \neq \lambda)$ then $\varepsilon(x_\lambda/x_{\lambda'}) = \varepsilon(x_\lambda)$ because of the assumption that $\varepsilon(x_1), \ldots, \varepsilon(x_N)$, or more specifically $\varepsilon(x_\lambda)$, remain unaffected during the entire sampling, i.e. we have

$$\varepsilon(x_\lambda \,.\, x_{\lambda'}) = \varepsilon(x_\lambda)\, \varepsilon(x_{\lambda'}) \qquad (6)$$

for all $\lambda, \lambda', \lambda \neq \lambda'$. Let further

$$v(x_1), \ldots, v(x_N) \qquad (7)$$

denote the *a priori* variances for individuals $\lambda = 1, \ldots, N$ respectively where

$$v(x_\lambda) = \varepsilon[x_\lambda - \varepsilon(x_\lambda)]^2$$

$\lambda = 1, \ldots, N$. Putting hereafter ε and v in general for *a priori* expectation and variance respectively while preserving symbols E and V for expectation and variance respectively over all possible samples s, we have from (3), (5), (6) and (7)

$$\begin{aligned}
\varepsilon V(e_s) &= \sum_s P_s\, \varepsilon(e_s^2) - \varepsilon(T^2) \\
&= \sum_{\lambda=1}^{N} v(x_\lambda) \sum_{s \supset \lambda} \beta_{s\lambda}^2 P_s + \sum_{\lambda, \lambda'=1}^{N} \varepsilon(x_\lambda)\, \varepsilon(x_{\lambda'}) \sum_{s \supset \lambda, \lambda'} \beta_{s\lambda}\, \beta_{s\lambda'} P_s \\
&\quad - \sum_{\lambda=1}^{N} v(x_\lambda) - [\sum_{\lambda=1}^{N} \varepsilon(x_\lambda)]^2 \qquad (8)
\end{aligned}$$

$s \supset \lambda, \lambda'$ denoting all samples which include λ and λ'^{th} individuals in the population.

Now for a given sampling design p we define $e_s(x, \bar{\beta})$ as the best linear estimate of T provided $\bar{\beta} \,\varepsilon\, B_p$ where B_p as before (Section 3) denotes the class of all β for which $e_s(x, \beta)$ are unbiased estimates of T and

$$\varepsilon V[e_s(x, \bar{\beta})] \leqslant \varepsilon V[e_s(x, \beta)]$$

for all $\beta \,\varepsilon\, B_p$.

That such $\bar{\beta}$ exists is evident from (8); however, if $\bar{\beta}$ depends upon $v(x_1), \ldots, v(x_N)$ it is of little practical use since $v(x_1), \ldots, v(x_N)$ are almost never known to the statistician. It will be shown later that for many of the sampling designs usually employed in practice $\bar{\beta}$ is independent of $v(x_1), \ldots, v(x_N)$. In fact it will be demonstrated in the subsequent sections that most of the usually employed estimates satisfy this criterion of *bestness*.

Hereafter for convenience we put, for a given sampling design p

$$e_s(x, \bar{\beta}) = e_s[p] \text{ simply}$$

and

$$\varepsilon V[e_s(x, \bar{\beta})] = \varepsilon V[p]$$

* For instance let x_λ be the yield of certain crop in the λ^{th} village of a population of N villages for which it is proposed to estimate $T = \sum_\lambda x_\lambda$. Now in case no other information excepting the acreages under the crop, $A_\lambda, \lambda = 1, \ldots, N$ in the N villages is available, it would be reasonable to put

$$\varepsilon(x_\lambda) = \text{const. } A_\lambda$$

$\lambda = 1, \ldots, N$.

6. *A Useful Form of* $\varepsilon V[p]$

To obtain $\varepsilon V[p]$ we have to minimize $\varepsilon V(e_s)$ in (8) viz.

$$\varepsilon V(e_s) = \sum_s P_s\, \varepsilon(e_s^2) - \varepsilon(T^2)$$

for variations of β subject to the condition

$$\sum_{s \supset \lambda} \beta_{s\lambda} P_s = 1 \qquad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (9)$$

for $\lambda = 1, \ldots, N$. Then since $\beta = \bar{\beta}$ is the solution,

$$\left\{ \frac{\partial}{\partial \beta_{s\lambda}} \varepsilon V(e_s) - \mu_\lambda \frac{\partial}{\partial \beta_{s\lambda}} \sum_{s \supset \lambda} \beta_{s\lambda} P_s \right\}_{\bar{\beta}} = 0 \qquad . \quad . \quad . \quad . \quad . \qquad (10)$$

for all $\lambda \in s$ and all s, μ_λ, $\lambda = 1, \ldots, N$ being the Lagrangian multipliers. Hence

$$\varepsilon\{x_\lambda (e_s)_{\bar{\beta}}\} - \frac{\mu_\lambda}{2} = 0$$

for all $\lambda \in s$ and all s having $P_s \neq 0$. Multiplying (10) by $\bar{\beta}_{s\lambda}$ and then adding such equations for all $\lambda \in s$ we have

$$\varepsilon\{(e_s^2)_{\bar{\beta}}\} - \frac{1}{2} \sum_{\lambda \in s} \mu_\lambda \bar{\beta}_{s\lambda} = 0.$$

Again multiplying this equation by P_s and summing over all s we have

$$\sum_s P_s\, \varepsilon\{(e_s^2)_{\bar{\beta}}\} - \frac{1}{2} \sum_s P_s \sum_{\lambda \in s} \mu_\lambda \bar{\beta}_{s\lambda} = 0.$$

But since from (9)

$$\sum_s P_s \sum_{\lambda \in s} \mu_\lambda \bar{\beta}_{s\lambda} = \sum_{\lambda=1}^{N} \mu_\lambda \sum_{s \supset \lambda} \bar{\beta}_{s\lambda} P_s = \sum_{\lambda=1}^{N} \mu_\lambda$$

we get

$$\sum_s P_s\, \varepsilon\{(e_s^2)_{\bar{\beta}}\} = \frac{1}{2} \sum_{\lambda=1}^{N} \mu_\lambda.$$

Thus

$$\varepsilon V[p] = \frac{1}{2} \sum_{\lambda=1}^{N} \mu_\lambda - \varepsilon(T^2). \qquad . \quad . \quad . \quad . \quad . \quad . \qquad (11)$$

7. *A Well Known Class of Sampling Designs*

It is fairly difficult to solve the equations (9) and (10) above for $\bar{\beta}$ in a general case. However their solutions given below, in particular cases, can be easily verified.

Let D^n denote the class of all the sampling designs p such that if $p \in D^n$ then in p

(i) the sample size $= n$;
(ii) sampling is done without replacement;
(iii) the probability of the λ^{th} individual being included in the sample is proportional to $\varepsilon(x_\lambda)$, i.e.

$$P(\lambda) = n\, \varepsilon(x_\lambda)/\varepsilon(T).$$

Now $\bar{\beta}$ and μ satisfying (9) and (10) for all $p \in D^n$ are

$$\bar{\beta}_{s\lambda} = \frac{\varepsilon(T)}{n} \cdot \frac{1}{\varepsilon(x_\lambda)} \qquad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (12)$$

for all $\lambda \in s$ and all s and

$$\mu_\lambda = 2\left\{ \varepsilon(T)\, \varepsilon(x_\lambda) + \frac{\varepsilon(T)}{n} \frac{v(x_\lambda)}{\varepsilon(x_\lambda)} \right\} \qquad . \quad . \quad . \quad . \quad . \quad . \qquad (13)$$

for $\lambda = 1, \ldots, N$ (to verify results (12), (13) use equation (6)). Hence for all $p \in D^n$ the best linear estimate of T is given by

$$e_s[p \in D^n] = \frac{\varepsilon(T)}{n} \sum_{\lambda \in s} \frac{x_\lambda}{\varepsilon(x_\lambda)}. \qquad (14)$$

Further from (11) and (13)

$$\varepsilon V[p \in D^n] = \frac{\varepsilon(T)}{n} \sum_{\lambda=1}^{N} \frac{v(x_\lambda)}{\varepsilon(x_\lambda)} - v(T).^* \qquad (15)$$

8. *An Optimum Property of Sampling Designs* $p \in D^n$

In any sampling design p the variance of an unbiased estimate e_s,

$$V(e_s) = \sum_s e_s^2 \cdot P_s - T^2.$$

Hence

$$\varepsilon V(e_s) = \sum_s \varepsilon(e_s^2) \cdot P_s - \varepsilon(T^2)$$

$$= \sum_s \varepsilon^2(e_s) P_s + \sum_s v(e_s) \cdot P_s - \varepsilon(T^2)$$

Further since from (6), for all $\lambda \neq \lambda'$

$$\varepsilon(x_\lambda \cdot x_{\lambda'}) = \varepsilon(x_\lambda)\, \varepsilon(x_{\lambda'})$$

we have

$$\varepsilon V(e_s) = \sum_s \varepsilon^2(e_s) \cdot P_s + \sum_{\lambda=1}^{N} v(x_\lambda) \sum_{s \supset \lambda} \beta^2_{s\lambda} P_s - \varepsilon(T^2). \qquad (16)$$

Now subject to the condition

$$\sum_s e_s P_s = T$$

$$\sum_s \varepsilon^2(e_s) P_s \geqslant \varepsilon^2(T) \qquad (17)$$

and subject to the condition

$$\sum_{s \supset \lambda} \beta_{s\lambda} P_s = 1$$

$$\sum_{s \supset \lambda} \beta^2_{s\lambda} P_s \geqslant 1 / \sum_{s \supset \lambda} P_s = 1/P(\lambda) \qquad (18)$$

$\lambda = 1, \ldots, N$. From (16), (17) and (18) it follows that for any sampling design p

$$\varepsilon V[p] \geqslant \sum_{\lambda=1}^{N} \frac{v(x_\lambda)}{P(\lambda)} - v(T) \qquad (19)$$

$P(\lambda)$ as before being the probability of including λ^{th} individual $\lambda = 1, \ldots, N$, in the sample drawn with sampling design p.

The inequality (19) can be illustrated. Let $n = 1$ and p be such that the probability of drawing λ^{th} individual is p_λ which in this case is also equal to $P(\lambda)$; the only unbiased estimate of T then is $x_\lambda / P(\lambda)$ and hence

$$\varepsilon V[p] = \sum_{\lambda=1}^{N} \left\{ \frac{\varepsilon(x_\lambda)}{P(\lambda)} - \varepsilon(T) \right\}^2 P(\lambda) + \sum_{\lambda=1}^{N} \frac{v(x_\lambda)}{P(\lambda)} - v(T)$$

$$\geqslant \sum_{\lambda=1}^{N} \frac{V(x_\lambda)}{P(\lambda)} - v(T)$$

* (15) can also be obtained from $V(\hat{T})$ given by Horwitz and Thomson ((1951) eq. (8) pp. 670). Calculate $\varepsilon V(\hat{T})$ putting therein $P(u_i) = n\, \varepsilon(x_i)/\varepsilon(T)$ and $\varepsilon(x_i \cdot x_j) = \varepsilon(x_i)\, \varepsilon(x_j)$. It is interesting that $\varepsilon V(\hat{T})$ then is independent of $P(u_i, u_j)$. (Note: $\sum_{i \neq j} P(u_i, u_j) = n(n-1)$.)

Moreover the lower bound of $\varepsilon V[p]$ establishe in equation (19) in in fact attained for all the sampling designs $p \in D^n$. This is evident from equation (15). Or otherwise from equations (9) and (10) we have in any sampling design p

$$\varepsilon\{x_\lambda (e_s)_{\bar{\beta}}\} - \frac{\mu_\lambda}{2} = 0$$

for all $\lambda \in s$ and all s having $P_s \neq 0$ which because of equation (6) is equivalent to

$$\varepsilon(x_\lambda)\, \varepsilon(e_s)_{\bar{\beta}} + v(x_\lambda)\bar{\beta}_{s\lambda} - \frac{\mu_\lambda}{2} = 0$$

Multiplying this equation by P_s and summing over all $s \supset \lambda$ and subsequently over all λ, $\lambda = 1, \ldots, N$, we have

$$\sum_{\lambda=1}^{N} \frac{\varepsilon(x_\lambda)}{P(\lambda)} \sum_{s \supset \lambda} \varepsilon(e_s)_{\bar{\beta}} P_s + \sum_{\lambda=1}^{N} \frac{v(x_\lambda)}{P(\lambda)} \sum_{s \supset \lambda} \bar{\beta}_{s\lambda} P_s - \frac{1}{2} \sum_{\lambda=1}^{N} \mu_\lambda = 0$$

and from equations (9) and (11)

$$\varepsilon V[P] = \sum_{\lambda=1}^{N} \frac{\varepsilon(x_\lambda)}{P(\lambda)} \sum_{s \supset \lambda} \varepsilon(e_s)_{\bar{\beta}} P_s + \sum_{\lambda=1}^{N} \frac{v(x_\lambda)}{P(\lambda)} - \varepsilon(T^2)$$

Now for any unbiased estimate e_s

$$\sum_{\lambda=1}^{N} \sum_{s \supset \lambda} \varepsilon(e_s) P_s = n \sum_{s} \varepsilon(e_s) P_s = n\varepsilon(T)$$

and when $p \in D^n$, $P(\lambda) = n\varepsilon(x_\lambda)/\varepsilon(T)$ which gives the necessary result

$$\varepsilon V[p \in D^n] = \frac{\varepsilon(T)}{n} \sum_{\lambda=1}^{N} \frac{v(x_\lambda)}{\varepsilon(x_\lambda)} - v(T) \quad \ldots \ldots \quad (20)$$

i.e. the lower bound of $\varepsilon V[p]$ in equation (19) is attained for all the sampling designs $p \in D^n$.

Further it can be seen by minimizing the right-hand side of the inequality (19) for all the variations of $P(\lambda)$, $\lambda = 1, \ldots, N$ subject to the condition

$$\sum_{\lambda=1}^{N} P(\lambda) = n$$

and from equation (20), that in case of all the populations fulfilling the condition

$$v(x_\lambda) \propto \varepsilon^2(x_\lambda) \quad \ldots \ldots \quad (21)$$

$\lambda = 1, \ldots, N$, *for whatever sampling design p' consisting of n drawings without replacement*

$$\varepsilon V[p \in D^n] \leqslant \varepsilon V[p'].$$

This provides a justification for employing sampling designs $p \in D^n$ in case of populations satisfying or approximately satisfying the condition (21).

Acknowledgments

I am very grateful to Professor M. G. Kendall from London, and Dr. A. R. Kamat from Fergusson College, Poona for their useful suggestions. I must also thank Miss Pednekar for carrying out the typing work.

References

BROSS, I. D. J. (1953), *Design for Decision*. New York: Macmillan, 48–53.
CARNAP (1951), *Logical Foundations of Probability*. University of Chicago.
COCHRAN, W. G. (1939), "The use of the analysis of variance in enumeration by sampling", *J. Amer. Statist. Ass.*, **34**, 492–510.
—— (1953) *Sampling Techniques*. John Wiley.
DAS, A. C. (1951), "On Sampling with varying probabilities", *Bull. Inst. int. Statist.*, **33,** II, 105–112.
DAVID, F. N. & NEYMAN, J. (1938), "Extension of the Markoff Theorem on least squares", *Statist. Res. Mem.*, II, 105–116.
HORWITZ, D. G. & THOMSON, D. J. (1952), "A generalization of sampling without replacement from a finite population", *J. Amer. Statist. Ass.*, **47**, 668–670.
NEYMAN, J. (1934), "On two aspects of representative method", *J.R. Statist. Soc.*, **97**, 593–600.
SUKHATME, P. V. (1954), "Sampling theory of surveys with applications", *Ind. Soc. Agric. Stat., New Delhi*, **23**, 194–198.
YATES, F. (1949), *Sampling Methods for Censuses and Surveys*. London: Griffin.