ESTIMATING THE MEAN OF A FINITE POPULATION

By J. Roy and I. M. Chakravarti

Indian Statistical Institute, Calcutta

0. Introduction and summary. In sampling from a finite population, the non-existence of a uniformly minimum variance unbiased estimator for the mean μ has been demonstrated by Godambe [3], and the inadmissibility of the sample mean as an estimator for μ, when sampling is with replacement and equal probabilities, has been proved by Des Raj and Khamis [2] and by Basu [1].

In this paper, the problem of unbiased linear estimation of μ with minimum variance is considered for a very general scheme of sampling. An admissible estimator is obtained, together with a complete class of estimators. It is shown further that, for a somewhat restricted sampling scheme, amongst estimators with variance proportional to σ^2, there does exist a best estimator which, in the case of sampling with replacement and equal probabilities, is the same as that considered in [1] and [2].

1. Sampling scheme and method of estimation. Consider a population consisting of a finite number N of distinguishable elementary units u_i with associated real numbers (variate-values) y_i, $i = 1, 2, \cdots, N$. The mean and the variance of the population will be denoted respectively by

$$\mu = N^{-1}\sum_{i=1}^{N} y_i \quad \text{and} \quad \sigma^2 = N^{-1}\sum_{i=1}^{N} (y_i - \mu)^2. \tag{1.1}$$

Let $\{U\}$ denote a countable collection of finite or infinite sequences $U(x)$, $x = 1, 2, \cdots$, of the elementary units, repetitions being allowed. We shall call each $U(x)$ a "sampling unit". Let $n_i(x)$ denote the number of times u_i occurs in $U(x)$ and let

$$\nu_i(x) = 0(1) \quad \text{if} \quad n_i(x) = 0(>0). \tag{1.2}$$

To avoid triviality, it will be assumed that there is no sampling unit which contains all the N different elementary units.

The sampling scheme to be considered is as follows: Only one of the sampling units is to be selected, the probability of selecting $U(x)$ being $p(x)$ so that $\sum p(x) = 1$ (summation being over all sampling units), and the variate-values for all the elementary units in the selected sampling unit are to be determined. The total number of elementary units in $U(x)$, counting repetitions, is thus $n(x) = \sum_{i=1}^{N} n_i(x)$, and the number of *distinct* elementary units in $U(x)$ is

$$\nu(x) = \sum_{i=1}^{N} \nu_i(x). \tag{1.3}$$

The serial number of the selected sampling unit is thus a random variable X with probability distribution given by

Received March 12, 1959; revised October 6, 1959.

$$\text{Prob}\ (X = x) = p(x), \qquad x = 1, 2, \cdots \cdots \tag{1.4}$$

It is to be noted that the sampling scheme considered is of a very general type; $n(x)$ and $\nu(x)$ need not be independent of x, and $n(x)$ may not even be finite. This kind of formulation is useful because it covers cases of sequential sampling. Consider, for instance, the following scheme of sampling: Draw elementary units, one by one, with replacement, until two different elementary units are obtained, where the probability of getting a particular unit may vary from draw to draw. The sample size counting repetitions may be infinite, though the effective sample size is only two. Our complete class Theorem 2.2 shows that in this case, to estimate the mean, one may disregard the multiplicities and the order of drawing of the two elementary units.

If $U(x)$ happens to be selected, a linear function, call it $t(x)$, of the variate-values for all the elementary units in $U(x)$, will be taken as the estimate for μ. In general, $t(x)$ can be written as $t(x) = \sum_{i=1}^{N} y_i a_i(x)$, where $a_i(x)$ ($i = 1, 2, \cdots \cdots, N$; $x = 1, 2, \cdots \cdots$) are pre-determined real numbers with the restriction that $a_i(x) = 0$ whenever $n_i(x) = 0$. The estimator is thus the random variable

$$T = t(X) = \sum_{i=1}^{N} y_i a_i(X). \tag{1.5}$$

In order that the expectation of T may be equal to μ for all values of $y = (y_1, y_2, \cdots, Y_N)$ a necessary and sufficient condition is that

$$E[a_i(X)] = N^{-1}, \qquad i = 1, 2, \cdots, N. \tag{1.6}$$

The further restriction $E[a_i(X)]^2 < \infty$, $i = 1, 2, \cdots, N$, is imposed so that the variance of T may be finite for all finite values of y. A random variable T satisfying these conditions will be called a linear unbiased estimator of μ.

Obviously (1.6) cannot hold unless for every i ($i = 1, 2, \cdots, N$) there exists at least one x for which both $n_i(x) > 0$ and $p(x) > 0$; henceforth this will be tacitly assumed. (Any u_i's for which $n_i(x)p(x) = 0$ for all x are effectively outside of the sampled population.)

The variance of a linear unbiased estimator T of μ is obviously given by

$$V(T) = \sum_{i=1}^{N} \sum_{j=1}^{N} y_i y_j \delta_{ij}, \tag{1.7}$$

where $\delta_{ij} = \text{Cov}\ [a_i(X), a_j(X)] = E[a_i(X)a_j(X)] - N^{-2}$.

2. An admissible estimator and a complete class of estimators. Of two different linear unbiased estimators T and T' of μ, T will be said to be *at least as good as* T' if

$$V(T) \leqq V(T') \tag{2.1}$$

holds for all y; T will be said to be *better* than T' if (2.1) holds for all values of y with strict inequality for at least one value of y. In a given class of linear unbiased estimators of μ, T will be said to be *best* if it belongs to the class and is better than any other member of the class; it will be said to be *admissible* if the class does not contain a better member. A class $\mathcal{C}$ of linear unbiased estimators of μ will be called *complete,* if given any linear unbiased estimator of μ not belonging to the class $\mathcal{C}$ it is possible to find a member of $\mathcal{C}$ which is better.

It has been shown [3] that a best estimator in the class of all linear unbiased estimators does not exist for any sampling scheme. An admissible estimator and a complete class of estimators are obtained in this section.

Let $a_i^*(x)$ be defined by

$$a_i^*(x) = \frac{\nu_i(x)}{Nq_i}, \tag{2.2}$$

where $\nu_i(x)$ is defined by (1.2) and q_i stands for the probability that the elementary unit u_i occurs in the selected sampling unit, that is $q_i = E[\nu_i(X)]$.

Consider

$$T^* = \sum_{i=1}^{N} y_i a_i^*(X), \tag{2.3}$$

which is easily verified to be a linear unbiased estimator of μ. The variance of T^* is given by

$$V(T^*) = \sum_{i=1}^{N} \sum_{j=1}^{N} y_i y_j \delta_{ij}^* \tag{2.4}$$

where $N^2\delta_{ij}^* = (q_{ij}/q_iq_j) - 1$ where q_{ij} stands for the probability that both u_i and u_j occur in the selected sampling unit; that is, $q_{ij} = E[\nu_i(X)\nu_j(X)]$, $q_{ii} = q_i$.

THEOREM 2.1 *T^* defined by* (2.3) *is admissible in the class of all linear unbiased estimators of* μ.

PROOF. If not, there exists a better linear unbiased estimator of μ, say T given by (1.5). Then, from (1.7) and (2.4) one gets

$$V(T^*) - V(T) = \sum_{i=1}^{N} \sum_{j=1}^{N} y_i y_j (\delta_{ij}^* - \delta_{ij}) \tag{2.5}$$

which must be at least positive-semidefinite. But it is easy to verify that $\delta_{ii}^* - \delta_{ii} = -E[a_i^*(X) - a_i(X)]^2$ is not positive: this contradicts the assumption that T is better.

To obtain a complete class of estimates proceed as follows. Let $J = (j_1, j_2, \cdots, j_m)$ denote a non-empty proper subset of the set of integers $(1, 2, \cdots \cdots, N)$. There are thus $2^N - 2$ such subsets. Let S_J stand for the set of the serial numbers x of those sampling units $U(x)$ which contain the elementary units $u_{j_1}, u_{j_2}, \cdots, u_{j_m}$ and these only, thus

$$S_J = \{x : \nu_i(x) = 1 \ (0) \quad \text{for} \quad i \in J (i \notin J)\}. \tag{2.6}$$

Let $\mathcal{C}_0$ denote the class of linear unbiased estimators T of μ for which the coefficients $a_i(x)$ are equal for all $x \varepsilon S_J$ and for every subset J, that is, they are of the form: $a_i(x) = b_{iJ}$; for all $x \varepsilon S_J$ and for every subset J. Thus, $\mathcal{C}_0$ is the class of linear unbiased estimators of μ, whose coefficients depend only on which elementary units are in the sampling unit, and not on their multiplicities or ordering.

We then have the following:

THEOREM 2.2 *The class $\mathcal{C}_0$ is complete.*

PROOF. Let $T = \sum_{i=1}^{N} y_i a_i(X)$ be a linear unbiased estimate of μ. Let $\pi_J =$ Prob $(X \varepsilon S_J)$ and define

$$b_{iJ} = \sum_{x \varepsilon S_J} a_i(x)p(x)/\pi_J \quad \text{if} \quad \pi_J > 0, \qquad b_{iJ} = 0 \text{ otherwise} \tag{2.7}$$

and further let $\bar{a}_i(x) = b_{iJ}$ for all $x \varepsilon S_J$ and for every subset J. It is easy to see that the estimator $\bar{T} = \sum_{i=1}^{N} y_i \bar{a}_i(X)$ belongs to the class $\mathcal{C}_0$. Also,

$$V(T) = V(\bar{T}) + \sum_{i=1}^{N} \sum_{j=1}^{N} y_i y_j \lambda_{ij}, \tag{2.8}$$

where $\lambda_{ij} = E[\{a_i(X) - \bar{a}_i(X)\}\{a_j(X) - \bar{a}_j(X)\}]$. Since the matrix $((\lambda_{ij}))$ is at least positive-semidefinite, $\bar{T}$ is better than T unless T itself belongs to $\mathcal{C}_0$. This completes the proof.

3. Best estimator in a restricted class. Since there does not exist a best member in the class of all unbiased linear estimators we proceed to examine whether a best estimator exists if the class is suitably restricted.

A linear unbiased estimator $T = t(X)$ will be called *linearly invariant* if the transformation $y_i^* = \alpha y_i + \beta$ $(i = 1, 2, \cdots, N)$ of the variate values transforms $t(x)$ to $t^*(x)$ where $t^*(x) = \alpha t(x) + \beta$ for all x for which $p(x) > 0$. Obviously, a necessary and sufficient condition for T to be linearly invariant is that

$$\sum_{i=1}^{N} a_i(x) = 1, \qquad \text{for all } x \text{ for which } p(x) > 0. \tag{3.1}$$

We now show by a counter-example that, even in the class of linearly invariant unbiased estimators, in general there does not exist a best estimator.

Consider a population of $N = 4$ elementary units u_i with variate-values y_i $(i = 1, 2, 3, 4)$. Let the sampling units be $U(1) = [u_1, u_2, u_3]$, $U(2) = [u_1, u_2, u_4]$, $U(3) = [u_1, u_3, u_4]$ and $U(4) = [u_2, u_3, u_4]$, and let the probability of selection be the same, viz. $\frac{1}{4}$ for all the sampling units. This corresponds to taking 3 units with equal probabilities without replacement from a population of 4 units. It follows from Theorem 2.1 that the sample mean T^* is an admissible estimator in this case. Obviously, T^* is linearly invariant and its variance is $\sigma^2/9$ where $\sigma^2 = \sum_{i=1}^{4} (y_i - \mu)^2/4$. Consider now an alternative estimate, $T = \sum_{i=1}^{4} y_i a_i(X)$, whose coefficient-matrix $\{a_i(x)\}$ $(i, x = 1, 2, 3, 4)$ is given on the following page.

i \ x	1	2	3	4
1	$\frac{1}{3}+\theta$	$\frac{1}{3}-\theta$	$\frac{1}{3}$	0
2	$\frac{1}{3}-\theta$	$\frac{1}{3}+\theta$	0	$\frac{1}{3}$
3	$\frac{1}{3}$	0	$\frac{1}{3}$	$\frac{1}{3}$
4	0	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

It is easy to verify that T is linearly invariant and unbiased, and that its variance, $\frac{1}{9}\sigma^2 + \frac{1}{2}\theta^2(y_1 - y_2)^2 + \frac{1}{6}\theta(y_1 - y_2)(y_3 - y_4)$, can be made smaller than the variance of T^* by a proper choice of θ if $y_1 \neq y_2$. Therefore a best invariant estimator does not exist in this case.

It will now be shown that, if consideration is limited to a still smaller class of what we propose to call regular estimators, there does exist a best estimator, provided that the sampling scheme is somewhat restricted.

A linear unbiased estimator T of μ will be called a regular estimator if its variance is of the form

$$V(T) = k\sigma^2, \tag{3.2}$$

where k is a constant independent of y. Suppose that T is of the form (1.5) so that its variance is given by (1.7). Since (3.2) can be written as

$$V(T) = k(N-1)N^{-2}\sum_{i=1}^{N} y_i^2 - kN^{-2}\sum_{i \neq j=1}^{N}\sum y_i y_j, \tag{3.3}$$

by equating coefficients in (1.7) and (3.3) we get

$$E[a_i(X)a_j(X)] = \begin{cases} kN^{-1} - (k-1)N^{-2} & \text{if } i = j \\ -(k-1)N^{-2} & \text{if } i \neq j. \end{cases} \tag{3.4}$$

Consequently, writing $a(X) = \sum_{i=1}^{N} a_i(X)$, one gets from (3.4), $V[a(X)] = 0$. Therefore

$$\sum_{i=1}^{N} a_i(x) = 1 \qquad \text{for all } x \text{ for which } p(x) > 0. \tag{3.5}$$

We thus have

THEOREM 3.1. *A regular estimator is linearly invariant.*

Let us now compute $M = E\sum_{i=1}^{N}[a_i(X) - \nu_i(X)/\nu(X)]^2$. By virtue of (3.5), we get $\sum_{i=1}^{N}[a_i(X) - \nu_i(X)/\nu(X)]^2 = \sum_{i=1}^{N}[a_i(X)]^2 - 1/\nu(X)$. Using (3.4) we then have

$$M = k(N-1)N^{-1} + N^{-1} - E[1/\nu(X)]. \tag{3.6}$$

Since M is non-negative, we obtain the following:

LEMMA. *For the variance of any regular estimator T of μ, there exists a lower bound $V(T) \geqq K\sigma^2$ where*

$$K = (N/(N-1))E[1/\nu(X)] - (1/(N-1)). \tag{3.7}$$

This lower bound can be attained if and only if $M = 0$. But this requires that, for every x for which $p(x) > 0$, $a_i(x) = \lambda_i(X)$, where

$$\lambda_i(x) = \nu_i(x)/\nu(x). \tag{3.8}$$

In order that the linear statistic

$$L = \sum_{i=1}^{N} y_i \lambda_i(X) \tag{3.9}$$

may be an unbiased estimator of μ, a necessary and sufficient condition is that

$$E[\nu_i(X)/\nu(X)] = N^{-1} \qquad \text{for } i = 1, 2, \cdots, N. \tag{3.10}$$

A sampling scheme will be called *balanced* if (3.10) holds.

We thus have proved the following:

THEOREM 3.2. *In order that the lower bound for the variance of a regular estimator of μ may be attained, a necessary and sufficient condition is that the sampling scheme should be balanced. If the sampling scheme is balanced, the estimator L defined by* (3.9) *is best in the class of all regular estimators and its variance is given by $V(L) = K\sigma^2$ where K is defined by* (3.7).

4. Application to specific sampling schemes. The usefulness of the theorems derived in Sections 2 and 3 will be demonstrated by considering several well known sampling schemes.

4.1 *Simple Random Sampling*: In this case, a sample of n elementary units is drawn one by one with equal probabilities and with replacement. There are thus N^n sampling units, each consisting of n of the N elementary units, repetitions being allowed. The probability of selecting any one sampling unit is N^{-n}. That the sample mean $T = \sum_{i=1}^{N} y_i n_i(X)/n$ is an inadmissible estimator follows from the complete class Theorem 2.2. This result was obtained earlier in [1] and [2] by proving that the estimator $T_0 = \sum_{i=1}^{N} y_i \nu_i(X)/\nu(X)$ is better. From Theorem 3.2 we have the stronger result that T_0 is the best regulator estimator. An admissible estimator in this case is

$$T^* = \frac{\sum_{i=1}^{N} y_i \nu_i(X)}{N[1 - (1 - (1/N))^n]}$$

as obtained from Theorem 2.1. This estimator was used in [3] to prove that the sample mean T_0 is not better than T^*. However, T^* is not even a linearly invariant estimator.

4.2 *Random Sampling Without Replacement.* In this case a sample of n elementary units is drawn one by one with equal probabilities but without replacement. There are thus $N!/(N - n)!$ sampling units, each consisting of a combination of n of the N elementary units, and each such sampling unit has the probability $(N - n)!/N!$ of selection. It is easily seen from Theorems 2.1 and 3.2 that the sample mean in this case is admissible and best in the regular class.

The counter-example in Section 3 however demonstrates that a best invariant estimator does not exist in general.

4.3 *Sampling for ν Distinct Units.* In this case elementary units are drawn one by one with equal probabilities and with replacement, until ν distinct elementary units are drawn, the total sample size being thus a random variable. It is seen from Theorem 2.2 that the sample mean $T = \sum_{i=1}^{N} y_i n_i(X)/n(X)$ is inadmissible. This was proved in [1] by showing that the estimator $T^* = \sum_{i=1}^{N} y_i \nu_i(X)/\nu$ is better. It follows from Theorems 2.1 and 3.1 that T^* is admissible and best in the regular class.

REFERENCES

[1] D. Basu, "On sampling with and without replacement," *Sankhyā*, Vol. 20 (1958), pp. 287–294.

[2] Des Raj and H. S. Khamis, "Some remarks on sampling with replacement," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 550–557.

[3] V. P. Godambe, "A unified theory of sampling from finite populations," *Jour. Roy. Stat. Soc. Ser. B*, Vol. 17 (1955), pp. 269–277.