SUFFICIENCY IN SAMPLING THEORY

By P. K. Pathak[1]

University of Illinois

0. Summary. The present paper is an attempt to define sufficiency in simple terms in the theory of sampling. This definition is a suitable version of the existing notion of sufficiency as defined by Fisher, Halmos and Savage, Bahadur and others. The paper gives justification for the use of sufficient statistics in sampling theory. Applications to interpenetrating subsampling and two-stage sampling are given.

In interpenetrating subsampling, it is proved that for estimating the population mean when the subsamples are drawn by simple random sampling without replacement, an estimator better than the usual overall average of subsample means is given by the average of distinct sample units. An improved estimator of the population variance is derived. In two-stage sampling where the first-stage units are drawn with unequal probabilities and second-stage units by simple random sampling (without replacement), two estimators of the population mean which are better than the estimator in current use are given.

1. Introduction. The notion of sufficiency was first introduced by Fisher (1922) and was made rigorous in abstract terms by Halmos and Savage (1949), Lehmann and Scheffé (1950) and later by Bahadur (1954). Recently it has been noted by Basu (1958), Hájek (1959) and the author (1961) that the theory of sufficiency can be used in a great many problems in sampling theory. The reason why the theory of sufficient statistic has not yet been fully used in sampling theory is probably because most results concerning sufficient statistics and their use are not available in the literature in a language understandable to most research workers in sampling. The notion of sufficiency as it exists today in abstract terms is neither necessary nor is it important from the view-point of sampling theory and actually obscures the main problems of estimation therein. This paper is being written in simple terms to make precise as to what is meant by sufficiency in sampling theory and to justify the use of sufficient statistics in sampling and to show that the existing theorems like Rao-Blackwell's on sufficiency apply here with almost no modification. Section 2 of this paper is, therefore, mainly expository.

2. A general sampling scheme. Consider a population π of N elements of arbitrary nature and let $U_1, U_2, \cdots, U_N$ be the values of a certain vector-valued variable under study associated with the N elements of the population. U_j is said to be the U-value of the jth population element and it is assumed that in U_j is incorporated its unit-index j ($j = 1, \cdots, N$). It is considered worthwhile to define the following terms.

Received 22 October 1962; revised 27 November 1963.

[1] On leave from Indian Statistical Institute.

DEFINITION 2.1. A sample s from π is defined to be a finite sequence of U's selected from $(U_1, U_2, \cdots, U_N)$ with or without replacement.

DEFINITION 2.2. A sampling scheme on π is defined to be an arbitrary collection $S = \{s\}$ of samples s from π, to be called the sample space, on which a probability measure has been defined. Thus a sampling scheme is expressed as $[\{s, p(s)\}, s \varepsilon S]$ where $p(s)$ denotes the probability of selection of the sample s and $\sum_{s\varepsilon S} p(s) = 1$. Sometimes a sampling scheme would be denoted by $\{S, P\}$ or just by S. It may be noted that S has at most a denumerable number of samples.

The above definition of a sampling scheme is quite general as any method of element-by-element sampling from a population can be expressed in the above form. For example simple random sampling without replacement of size n from a population of N consists of $\binom{N}{n}$ sequences corresponding to $\binom{N}{n}$ combinations of n chosen out of N elements and $p(s) = 1/\binom{N}{n}$ for each sample, and simple random sampling (with replacement) consists of N^n sequences of n elements drawn with replacement from N elements of the population and $p(s) = 1/N^n$ for each sample.

DEFINITION 2.3. Two samples s_1 and s_2 are said to be equivalent if they both contain the same elements of the population. For example $s_1 = (U_1, U_2, U_2, U_3)$ and $s_2 = (U_1, U_3, U_2)$ are equivalent as they both lead to the inclusion of the first three elements of the population.

DEFINITION 2.4. A partition of the sample space is a division of the totality of samples $S = \{s\}$ into mutually disjoint subsets of S. A typical partition will be denoted by $S_T = \{s_T\}$.

DEFINITION 2.5. A partition of S into subsets of equivalent samples is called a sufficient partition. Thus $S_T = \{s_T\}$ is a sufficient partition if each s_T of S_T contains only equivalent samples. s_T is called an element of S_T.

In some cases, it will be desirable to express a sufficient partition together with its probability measure as $S_T = \{s_T, p(s_T)\}$ where $p(s_T) = \sum_{s\varepsilon s_T} p(s)$.

A statistic $T(s)$ [a function defined from S onto another space Y] induces a partition S_T on S such that $T(s)$ is the same for all $s \varepsilon s_T$ for each $s_T \varepsilon S_T$. Further if for some $s_1, s_2 \varepsilon S$, $T(s_1) \neq T(s_2)$ then s_1 and s_2 belong to different s_T's of S_T. We, therefore, have the following definition.

DEFINITION 2.6. A statistic $T(s)$ is said to be sufficient if the partition, S_T, induced by T is sufficient.

REMARK: Suppose $S_T = \{s_t\}$ is a sufficient partition of S. Let T be the indexing set of s_t's. Then the function $T(s) = t$ for all $s \varepsilon s_t$ is a sufficient statistic as it induces the sufficient partition S_T. Thus given any sufficient partition there exists a sufficient statistic T that induces it and so any sufficient partition can be replaced by its sufficient statistic. In the sequel, whenever we say that T is a sufficient statistic we refer to the sufficient partition induced by T viz $\{s_t\}$.

The reason why partitioning of S into equivalent samples is sufficient is the following. A sample $s \varepsilon S$ provides us information about those population elements that are in the sample. If two samples contain the same population elements then they give the same information about the population. Thus a partition of S into subsets of equivalent samples is sufficient because each of the samples of a given subset is equally informative and also given one sample of a subset it is possible to work out all the equivalent samples of that subset without any information about the other samples. In other words this means that the sampling distribution of samples depends on the population elements only through equivalent samples. It is to be noted that $\mathbf{U} = (U_1, U_2, \cdots, U_N)$ the U-values of the population elements play the role of the parameter. The sufficiency here is with regard to $\mathbf{U}$.

The proof of the sufficiency of a sufficient partition in abstract terms can be given in a rigorous manner and has been omitted here for reasons of simplicity. Bahadur's paper (1954) may be referred to for the definition of sufficiency in abstract terms. It is believed that it is quite unnecessary to define sufficiency in abstract terms in sampling. Definition 2.5 is a suitable version of sufficiency in sampling theory. With this variation in the definition of sufficiency, existing theorems on sufficiency can be applied to sampling theory without any modification.

DEFINITION 2.7. A sufficient statistic T_1 (or the sufficient partition S_{T_1}) is said to be smaller than a sufficient statistic T_2 (or the sufficient partition S_{T_2}), $T_1 \subseteq T_2$ ($S_{T_1} \subseteq S_{T_2}$) in symbols, if for each $s_{T_2} \varepsilon S_{T_2}$ there is an $s_{T_1} \varepsilon S_{T_1}$ such that $s_{T_2} \subset s_{T_1}$.

If $T_1 \subseteq T_2$ (or $S_{T_1} \subseteq S_{T_2}$), then the elements of S_{T_1} can be obtained from S_{T_2} by grouping together some elements of S_{T_2}.

DEFINITION 2.8. For each $s \varepsilon S$, let

$$T(s) = [U_{(1)}, U_{(2)}, \cdots, U_{(d)}] \tag{1}$$

be the set of d U-values corresponding to different population units included in s and arranged in increasing order of their unit-indices, e.g., if $s = [U_3, U_5, U_2, U_5]$ then $T(s) = [U_2, U_3, U_5]$. $T(s)$ is called the "order-statistic"; $U_{(1)}, U_{(2)}, \cdots$, etc. are called respectively the first, second order-statistic etc. This definition of the order-statistic is slightly different from the customary definition of order-statistic where the units are arranged in order of magnitude of a certain characteristic associated with the units. This definition is used to avoid possible ambiguity in defining the "order-statistic" when a sample may contain two different population units of the same order of magnitude.

It can be easily seen that the "order-statistic" is a sufficient statistic since it induces a partition of equivalent samples of S.

DEFINITION 2.9. Let $f(s)$ be a real-valued function defined on the sample space S. Then the conditional expectation of $f(s)$ given a sufficient partition $S_T = \{s_T, p(s_T)\}$ is given by

$$E[f(s) \mid T] = \textstyle\sum_1 f(s)p(s) / \sum_1 p(s) \tag{2}$$

where the summation, $\sum_1$, is taken over the samples $s \varepsilon s_T$. Note that $E[f(s) \mid T]$ is defined on S_T.

LEMMA 2.1. *Let L be a real-valued convex function defined on an interval D of the real line. Let $\{x_n\}$ be a sequence of points in D and let $\{p_n\}$ be a sequence of non-negative numbers with $\sum p_n = 1$ and $\sum p_n|x_n| < \infty$. Then*

$$L(\sum p_n x_n) \leqq \sum p_n L(x_n) \tag{3}$$

with strict sign of inequality holding if L is strictly convex.

The definition of a convex function and the proof of the lemma can be found in Blackwell and Girshick (1954) p. 41.

The well-known Rao-Blackwell theorem is given below. The proof of this theorem in abstract terms can be found in standard texts on estimation (see Fraser (1957), p. 57). An alternative proof in terms of the definitions introduced in this paper is given here.

THEOREM 1 (Rao-Blackwell). *Let $T = \{s_T, p(s_T)\}$ be a sufficient statistic for $\{s, p(s), s \varepsilon S\}$. Let $f(s)$ be an unbiased estimator of a real-valued parameter $g(U_1, \cdots, U_N)$. Then $f_T(s) = E[f(s) \mid T]$ is also unbiased for $g(U_1, \cdots, U_N)$ and for any convex loss function $f_T(s)$ has smaller expected loss than $f(s)$ unless $f(s) = f_T(s)$ for all $s \varepsilon S$ in which case $f(s)$ and $f_T(s)$ have the same expected loss.*

Proof. By definition

$$E[f_T(s)] = \textstyle\sum_2 f_T(s)p(s_T) \tag{4}$$

where the summation, $\sum_2$, extends over all $s_T \varepsilon S_T$. Using (2),

$$\begin{aligned} E[f_T(s)] &= \textstyle\sum_2 \{\sum_1 f(s)p(s)/p(s_T)\}p(s_T) \\ &= \textstyle\sum_2 \sum_1 f(s)p(s) = \sum_{s\varepsilon S} f(s)p(s) = E[f(s)] = g(U_1, \cdots, U_N). \end{aligned}$$

Thus $f_T(s)$ is also an unbiased estimator of $g(U_1, \cdots, U_N)$. To prove that $f_T(s)$ has smaller expected loss function than $f(s)$, let $L(x)$ be a convex loss function. Then

$$L[f_T(s)] = L[\textstyle\sum_1 p(s)f(s)/p(s_T)] \leqq \sum_1 \{p(s)/p(s_T)\}L[f(s)] \tag{5}$$

by (3), the strict inequality holding when $L(x)$ is strictly convex. On taking the expectation on both sides of (5), we get

$$E\{L[f_T(s)]\} \leqq E\{L[f(s)]\}. \tag{6}$$

This completes the proof of the theorem.

COROLLARY 1.1. *If the loss function is the squared error, then $f_T(s)$ has smaller variance than $f(s)$. The decrease in variance is equal to $E\{f(s) - f_T(s)\}^2$.*

DEFINITION 2.10. Of two estimators $f_1(s)$ and $f_2(s)$ of $g(U_1, \cdots, U_N)$, $f_1(s)$ will be said to be uniformly better than $f_2(s)$ if for any convex loss function $f_1(s)$ does not have greater expected loss than $f_2(s)$ for all $(U_1, \cdots, U_N)$. Using a method of proof similar to that of Theorem 1, one can prove the following theorem.

THEOREM 2. *Let $f(s)$ be an unbiased estimator of $g(U_1, \cdots, U_N)$. Let T_1 and T_2 be two sufficient statistics such that $T_1 \subseteq T_2$. Then $f_{T_1}(s) = E[f(s) \mid T_1]$ and*

$f_{T_2}(s) = E[f(s) \mid T_2]$ *are uniformly better than* $f(s)$. *Further* $f_{T_1}(s)$ *is uniformly better than* $f_{T_2}(s)$ *and* $f_{T_1}(s) = E[f_{T_2}(s) \mid T_1]$.

Applications of these results to interpenetrating subsampling and two-stage sampling are given below.

3. Interpenetrating subsampling. Let us consider a population of N units. Let U_j be the jth population unit and let Y_j be the value of some real-valued Y-characteristic associated with U_j, $j = 1, \cdots, N$. In conformity with the notations commonly in use capital letters refer to the population and small letters refer to the sample. The following sampling scheme is considered herein: k simple random (without replacement) subsamples each of size n are drawn from the above population independently of each other, i.e., each subsample is drawn by simple random sampling (without replacement) and is replaced to the population for subsequent selection of subsamples.

The usual estimator of the population mean $\bar{Y} = N^{-1} \sum Y_j$, based on the ith subsample is given by the sample mean

$$\bar{y}_i = n^{-1} \sum_{r=1}^{n} y_{ir} \qquad i = 1, \cdots, k \tag{7}$$

where the summation is taken over all units of the ith subsample. Obviously any linear function $\sum_{i=1}^{k} c_i \bar{y}_i$ is also an unbiased estimator of $\bar{Y}$ provided $\sum_{i=1}^{k} c_i = 1$. Of all the nk units observed in these subsamples, not all will be distinct. Let $u_{(1)}, u_{(2)}, \cdots, u_{(m)}$ be the m distinct units observed. Then the order-statistic $T = [u_{(1)}, u_{(2)}, \cdots, u_{(m)}]$ is sufficient. Therefore, if any estimator of $\bar{Y}$ does not depend on T it can be reduced by conditioning through T to yield a better estimator of $\bar{Y}$. For doing so and for other interest, we will need the probability distribution of T. The following lemma will be found useful for this purpose.

LEMMA 3.1. *Let* $A_1, A_2, \cdots, A_m$ *be* m *events defined on a probability space. Let* $A = \bigcup_{i=1}^{m} A_i$ *and* $B_i = (A - A_i)$, $i = 1, \cdots, m$. *Then*

$$P\left[\bigcap_{i=1}^{m} A_i\right] = P(A) - \sum{}' P(B_1) + \sum{}' P(B_1 \cap B_2) - \cdots \tag{8}$$

where the summation $\sum'$ *is taken over all combinations of* B's *chosen out of* $B_1, \cdots, B_m$.

The proof is omitted.

It is easy to show on letting $A_i = \{u_{(i)} \text{ is included in the sample}\}$ in Lemma 3.1, that the probability of selecting any preassigned m distinct units, in this sampling scheme, is given by

$$P(T) = \binom{m}{n}^k - \binom{m}{1}\binom{m-1}{n}^k + \cdots (-1)^{m-n}\binom{m}{m-n}\binom{n}{n}^k \Big/ \binom{N}{n}^k \tag{9}$$

A consequence of Lemma 3.1 and Theorem 1 is the following theorem.

THEOREM 3. *An estimator uniformly better than* $\bar{y}_1$ *is given by*

$$(10) \qquad \bar{y}_m = E[\bar{y}_1 \mid T]$$

where $\bar{y}_m$ denotes the average of the m distinct units observed in the sample.

Proof. Obviously

$$(11) \quad E(\bar{y}_1 \mid T) = E\left[n^{-1}\sum_{r=1}^{n} y_{1r} \mid T\right] = E\,(y_{11} \mid T) = \sum_{i=1}^{m} y_{(i)}\, P[u_{11} = u_{(i)} \mid T]$$

where u_{11} stands for the first sample unit of the first subsample, y_{11} being its Y-characteristic value, and $u_{(i)}$ is the ith order-statistic. Further

$$(12) \qquad P[u_{11} = u_{(i)} \mid T] = P[u_{11} = u_{(i)} \cap T]/P(T).$$

Now letting $A_j = [u_{11} = u_{(1)}] \cap [u_{(j)}$ is included in the sample$]$, $j = 1, \cdots, m$, in Lemma 3.1, we get

$$(13) \quad \begin{aligned} &P\,[u_{11} = u_{(i)} \cap T] \\ &= \binom{m-1}{n-1}\binom{m}{n}^{k-1} - \binom{m-1}{1}\binom{m-2}{n-1}\binom{m-1}{n}^{k-1} \\ &\qquad \cdots (-1)^{m-n}\binom{m-1}{m-n}\binom{n-1}{n-1}\binom{n}{n}^{k-1} \Big/ N\binom{N-1}{n-1}\binom{N}{n}^{k-1} \\ &= \frac{1}{m}\left[\binom{m}{n}^{k} - \binom{m}{1}\binom{m-1}{n}^{k} + \cdots (-1)^{m-n}\binom{m}{m-n}\binom{n}{n}^{k}\right] \Big/ \binom{N}{n}^{k}. \end{aligned}$$

From equations (9), (10), (11), (12) and (13), it follows $E(\bar{y}_1 \mid T] = \bar{y}_m$ which completes the proof of the theorem.

Corollary 3.1. *For estimating the population mean, $\bar{Y}$, an estimator uniformly better than any linear function $\sum_{i=1}^{k} c_i\bar{y}_i$, $\sum c_i = 1$, of the subsample means is given by $\bar{y}_m$.*

It is observed on taking $c_1 = c_2 = \cdots = c_k = k^{-1}$ that $\bar{y}_m$ is uniformly better than the overall average of the subsample means, the estimator in current use.

If $n = 1$, then the above sampling scheme reduces to simple random sampling with replacement. This theorem then shows that the average of distinct sample units is a uniformly better estimator of the population mean than the overall average of all sample units. This result has been proved by Basu (1958) and Hájek (1959). Des Raj and Khamis (1958) have shown that the former estimator has smaller variance than the latter.

An alternative proof of the above theorem based on intuitive argument which would not require the evaluation of $P[u_{11} = u_{(i)} \cap T]$ can be given. However, the important point in favor of this proof is that it applies on similar lines to the case when the subsample sizes are different. For this case the intuitive approach for the derivation of (12) will be rather vague and not too obvious and therefore has not been adopted here.

4. Variance of $\bar{y}_m$. Obviously

$$(14) \quad V(\bar{y}_m) = E[V(\bar{y}_m \mid m)] + V[E(\bar{y}_m \mid m)] = E[V(\bar{y}_m \mid m)] = E[(m^{-1} - N^{-1})S^2]$$

where $S^2 = (N-1)^{-1}\sum_{j=1}^{N}(Y_j - \bar{Y})^2$.

The author (1961a) has proved that if under a given sampling scheme m denotes the number of distinct units in a sample then

$$E(m^{-1}) = \frac{1}{N}\left[1 + \frac{1}{(N-1)}\sum{}' q_1 + \frac{1\cdot 2}{(N-1)(N-2)}\sum{}' q_{12} + \cdots \cdots + \frac{1\cdot 2\cdot \cdots \cdot (N-1)}{(N-1)\cdot \cdots 2\cdot 1}\sum{}' q_{12\cdots N-1}\right] \tag{15}$$

where N is the population size, $q_{12\cdots r}$ is the probability of non-inclusion of first r population units in the sample and the summation $\sum'$ is taken over all possible combinations.

Since in the sampling scheme under consideration

$$q_{12\cdots r} = \binom{N-r}{n}^k \Big/ \binom{N}{n}^k, \qquad r = 1, \cdots, N-n \tag{16}$$

we have

$$E\left(\frac{1}{m}\right) = \binom{N}{n}^{-k}\left[\binom{N}{n}^k \Big/ N + \binom{N-1}{n}^k \Big/ (N-1) + \cdots + \binom{n}{n}^k \Big/ n\right] \tag{17}$$

An asymptotic expression for $V(\bar{y}_m)$ valid for terms up to order N^{-1} is given by

$$V(\bar{y}_m) = [(nk)^{-1} - (2N)^{-1}\{1 + (n-1)/(nk-1)\}]S^2. \tag{18}$$

Thus the approximate reduction in variance by using $\bar{y}_m$ over that of the customary estimator, $k^{-1}\sum_{i=1}^k \bar{y}_i$, is given by

$$V\left(\frac{1}{k}\sum_{i=1}^k \bar{y}_i\right) - V(\bar{y}_m) = \left(\frac{1}{nk} - \frac{1}{Nk}\right)S^2 - \left[\frac{1}{nk} - \frac{1}{2N}\left\{1 + \frac{(n-1)}{(nk-1)}\right\}\right]S^2 = \frac{(nk-2)(k-1)}{2k(nk-1)N}S^2. \tag{19}$$

5. Estimation of variance. Considering now the problem of estimating the population variance

$$S^2 = (N-1)^{-1}\sum_{j=1}^N (Y_j - \bar{Y})^2 \tag{20}$$

from interpenetrating subsamples, it is easily seen that an unbiased estimator of S^2 is given by

$$s^2 = [k(n-1)]^{-1}\sum_{i=1}^k \sum_{r=1}^n (y_{ir} - \bar{y}_i)^2. \tag{21}$$

The theorem given below gives an estimator uniformly better than s^2.

THEOREM 4. *An estimator uniformly better than s^2 is given by*

$$s_m^2 = (m-1)^{-1} \sum_{i=1}^{m} (y_{(i)} - \bar{y}_m)^2. \tag{22}$$

(It is assumed that $n \geqq 2$. For $n = 1$ refer to Pathak (1962).)

PROOF. Evidently, an estimator uniformly better than s^2 is given by

$$\begin{aligned} E[s^2 \mid T] &= E\left[(n-1)^{-1} \sum_{r=1}^{n} (y_{1r} - \bar{y}_1)^2 \mid T\right] \\ &= E\left[[2n(n-1)]^{-1} \sum_{r \neq r'=1}^{n} (y_{1r} - y_{1r'})^2 \mid T\right] \\ &= E\left[\tfrac{1}{2}(y_{11} - y_{12})^2 \mid T\right] \\ &= \sum_{i \neq i'=1}^{m} \tfrac{1}{2}(y_{(i)} - y_{(i')})^2 P[u_{11} = u_{(i)}, u_{12} = u_{(i')} \mid T] \end{aligned} \tag{23}$$

where u_{11}, u_{12} are the first two sample units of the first subsample, y_{11} and y_{12} being their Y-characteristic values respectively and $u_{(i)}$ and $u_{(i')}$ are the ith and i'th order-statistics respectively. Now

$$P[u_{11} = u_{(i)}, u_{12} = u_{(i')} \mid T] = P[u_{11} = u_{(i)}, u_{12} = u_{(i')} \cap T]/P(T). \tag{24}$$

An application of Lemma 3.1 will show that

$$\begin{aligned} &P[u_{11} = u_{(i)}, u_{12} = u_{(i')} \cap T] \\ &= \frac{\left[\binom{m-2}{n-2}\binom{m}{n}^{k-1} - \binom{m-2}{1}\binom{m-3}{n-2}\binom{m-1}{n}^{k-1} + \cdots (-1)^{m-n}\binom{m-2}{m-n}\binom{n-2}{n-2}\binom{n}{n}^{k-1}\right]}{N(N-1)\binom{N-2}{n-2}\binom{N}{n}^{k-1}} \\ &= \frac{\left[\binom{m}{n}^k - \binom{m}{1}\binom{m-1}{n}^k + \cdots (-1)^{m-n}\binom{m}{m-n}\binom{n}{n}^k\right]}{m(m-1)\binom{N}{n}^k}. \end{aligned} \tag{25}$$

From the above two equations it immediately follows that

$$E[s^2 \mid T] = [2m(m-1)]^{-1} \sum_{i \neq i'=1}^{m} (y_{(i)} - y_{(i')})^2 = s_m^2.$$

This completes the proof.

6. Estimation of $V(\bar{y}_m)$. For the purpose of estimating the variance of $\bar{y}_m$, the following estimator is suggested.

$$v(\bar{y}_m) = (m^{-1} - N^{-1})s_m^2. \tag{26}$$

7. Two-stage sampling. Let $U_1, U_2, \cdots, U_N$ be the N first-stage units of a population. Suppose that U_j consists of M_j second-stage units and let U_{jh} be the hth second-stage unit of U_j ($h = 1, \cdots, M_j$; $j = 1, \cdots, N$). Let Y_{jh} be some real-valued Y-characteristic of U_{jh}. In this section also capital letters refer to the population and small letters refer to the sample, e.g., $u_1, u_2, \cdots, u_n$ stand for the n first-stage sample units (in order of draw), by u_{ir} we mean the rth second-stage unit (in order of draw) in the ith first-stage sample unit. It is assumed that all relevant information about the units such as their unit-indices, probabilities of selection etc. are incorporated in the symbols u_i and u_{ir}.

Let us now consider a two-stage sampling scheme where the first-stage units are selected with unequal probabilities (with replacement) and each subsample of second-stage units is selected by simple random sampling (without replacement). In this sampling scheme if the jth first-stage unit, U_j, is included λ_j times in the sample, λ_j subsamples of m_j units each are drawn therefrom independently of each other by simple random sampling (without replacement), i.e., each subsample is drawn by simple random sampling (without replacement) and is replaced to the population for subsequent selection of the subsamples. Let P_j be the probability of selection of U_j ($\sum P_j = 1$) and call

$$Z_{jh} = (Y_{jh}/P_j)(M_j/\sum M_j) \tag{27}$$

the z-value of U_{jh}. Unless otherwise stated, j runs from 1 to N, h from 1 to M_j, i from 1 to n, r from 1 to M_{u_i}, (i) from (1) to (d) and (ir) from $(i1)$ to $(id_{(i)})$.

In this sampling scheme the usual unbiased estimator of the population mean, $\bar{Y} = (\sum M_j)^{-1} \sum \sum Y_{jh}$, is given by (Sukhatme 1953)

$$\bar{\bar{z}}_n = n^{-1} \sum_i \bar{z}_i \tag{28}$$

where $\bar{z}_i = m_{u_i}^{-1} \sum_r z_{ir}$, z_{ir} being the z-value of u_{ir} and m_{u_i} denotes the number of second stage units selected from u_i. The summations $\sum_i$ and $\sum_r$ are taken over the first-stage sample units and the second-stage sample units of a given first-stage sample unit respectively.

In the sequel $\bar{\bar{z}}_n$ is shown to be inefficient and two estimators uniformly better than $\bar{\bar{z}}_n$ are derived. The first estimator suggests the immediate necessity of employing it in practice as it is simple to compute and has smaller expected loss. The second estimator though even uniformly better than the first, is difficult to compute and is not of much use in practice.

Since the first-stage units are drawn with replacement, let $u_{(1)}, u_{(2)}, \cdots, u_{(d)}$ be the d ($\leqq n$) distinct first-stage units in the sample arranged in an increasing order of their unit-indices. Let $\lambda_{(i)}$ be the number of times $u_{(i)}$ is included in the sample ($\sum \lambda_{(i)} = n$). Finally, let $u_{(i1)}, \cdots, u_{(id_{(i)})}$ be $d_{(i)}$ ($\leqq \lambda_{(i)} m_{u_{(i)}}$) distinct second-stage units of $u_{(i)}$ arranged in an increasing order of their unit-indices. Now the statistic

$$T^* = [\{u_{(i)}, \lambda_{(i)}; u_{(i1)}, \cdots, u_{(id_{(i)})}\}\ i = 1, \cdots, d] \tag{29}$$

is sufficient as it partitions the sample space into disjoint subsets of equivalent samples. It can be seen that the probability of getting a sample with a given T^* is

$$P(T^*) = \frac{n!}{\lambda_{(1)}! \cdots \lambda_{(d)}!} \prod_{i=1}^{d} [p_{(i)}]^{\lambda_{(i)}} \Big/ \left[\binom{M_{(i)}}{m_{(i)}}\right]^{\lambda_{(i)}}$$

$$(30) \qquad \cdot \left\{ \left[\binom{d_{(i)}}{m_{(i)}}\right]^{\lambda_{(i)}} - \binom{d_{(i)}}{1}\left[\binom{d_{(i)}-1}{m_{(i)}}\right]^{\lambda_{(i)}} + \cdots (-1)^{d_{(i)}-m_{(i)}} \binom{d_{(i)}}{d_{(i)}-m_{(i)}}\left[\binom{m_{(i)}}{m_{(i)}}\right]^{\lambda_{(i)}} \right\}.$$

Notice the similarity of the second term on the right side of (30) with the right side of (9).

Therefore, by Rao-Blackwell theorem an estimator uniformly better than $\bar{\bar{z}}_n$ can be obtained by taking the conditional expectation of $\bar{\bar{z}}_n$ given T^*. We, thus, have

THEOREM 5. *An estimator uniformly better than $\bar{\bar{z}}_n$ is given by*

$$(31) \qquad \bar{\bar{z}}_d^* = n^{-1} \sum \lambda_{(i)} \bar{z}_{d(i)}$$

where $\bar{z}_{d(i)} = [d_{(i)}]^{-1} \sum z_{(ir)}$, $z_{(ir)}$ *being the z-value of* $u_{(ir)}$.

PROOF. Clearly an estimator uniformly better than $\bar{\bar{z}}_n$ is given by

$$(32) \qquad E[\bar{\bar{z}}_n \mid T^*] = E[z_{11} \mid T^*] = \sum_{i=1}^{d} \sum_{r=1}^{d(i)} z_{(ir)} P[u_{11} = u_{(ir)} \mid T^*].$$

It can be seen from (30) and (13) that

$$(33) \qquad P[u_{11} = u_{(ir)} \mid T^*] = n^{-1}\lambda_{(i)}/d_{(i)}$$

and therefore,

$$E[\bar{\bar{z}}_n \mid T^*] = n^{-1} \sum_{i=1}^{d} (\lambda_{(i)}/d_{(i)}) \sum z_{(ir)} = n^{-1} \sum \lambda_{(i)} \bar{z}_{d(i)} .$$

Hence the theorem is proved.

The theorem thus proves the inefficiency of $\bar{\bar{z}}_n$. It may be noted that $\bar{\bar{z}}_d^*$ can be gotten from $\bar{\bar{z}}_n$ by replacing $\bar{z}_i$ by the average of distinct second-stage units sampled from u_i. The author recommends the use of this estimator in practice as it is simple to compute and in addition will have smaller variance than $\bar{\bar{z}}_n$, the estimator in current use.

8. Variance of $\bar{\bar{z}}_d^*$. We have

$$(34) \qquad \begin{aligned} V(\bar{\bar{z}}_d^*) &= E[V\{n^{-1} \sum \lambda_{(i)} \bar{z}_{d(i)} \mid \lambda_{(1)}, \cdots, \lambda_{(d)}\}] \\ &\quad + V[E\{n^{-1} \sum \lambda_{(i)} \bar{z}_{d(i)} \mid \lambda_{(1)}, \cdots, \lambda_{(d)}\}] \\ &= n^{-2} E[\sum \lambda_{(i)}^2 V(\bar{z}_{d(i)} \mid \lambda_{(i)})] + V[n^{-1} \sum \lambda_{(i)} \bar{Z}_{(i)}] \end{aligned}$$

where $\bar{Z}_{(i)} = [M_{(i)}]^{-1} \sum_{h=1}^{M_{(i)}} Z_{(ih)}$ and $M_{(i)}$ is the number of second-stage units in $u_{(i)}$.

It is evident from (14) and (17) that

$$V[\bar{z}_{d_{(i)}} \mid \lambda_{(i)}] = E\left[\left(\frac{1}{d_{(i)}} - \frac{1}{M_{(i)}}\right) S^2_{(i)}(z) \mid \lambda_{(i)}\right]$$

$$\begin{aligned} &= \left[\binom{M_{(i)}}{m_{(i)}}^{-\lambda_{(i)}} \left\{\binom{M_{(i)} - 1}{m_{(i)}}^{\lambda_{(i)}} \Big/ (M_{(i)} - 1) \right.\right. \\ (35) \qquad &+ \binom{M_{(i)} - 2}{m_{(i)}}^{\lambda_{(i)}} \Big/ (M_{(i)} - 2) + \cdots \\ &\left.\left. + \binom{m_{(i)}}{m_{(i)}}^{\lambda_{(i)}} \Big/ m_{(i)}\right\}\right] S^2_{(i)}(z) = \left\{\sum_{r=1}^{M_{(i)}-m_{(i)}} \frac{\phi_{(i)}^{\lambda_{(i)}}(r)}{(M_{(i)} - r)}\right\} S^2_{(i)}(z) \end{aligned}$$

where

$$\phi_{(i)}(r) = \binom{M_{(i)} - r}{m_{(i)}} \Big/ \binom{M_{(i)}}{m_{(i)}}$$

and

$$S^2_i(z) = (M_{(i)} - 1)^{-1} \sum_h (Z_{(i)h} - [M_{(i)}]^{-1} \sum Z_{(i)h})^2.$$

Further it is easy to see that

$$V[n^{-1} \sum \lambda_{(i)} \bar{Z}_{(i)}] = \sigma^2_{bz}/n \tag{36}$$

where $\sigma^2_{bz} = \sum_j P_j(\bar{Z}_j - \sum P_j\bar{Z}_j)^2$. Thus (34) can be written as

$$V(\bar{\bar{z}}^*_d) = E\left[\frac{1}{n^2} \sum_{j=1}^{N} \lambda_j^2 \left(\sum_{r=1}^{M_j-m_j} \frac{\phi_j^{\lambda_j}(r)}{(M_j - r_j)}\right) S^2_j(z)\right] + \frac{\sigma^2_{bz}}{n} \tag{37}$$

where λ_j is the number of times U_j is included in the sample.

Since λ_j is a binomial random variable with parameters n and P_j, we have

$$E\{\lambda_j^2\phi_j^{\lambda_j}(r)\} = nP_j\phi_j(r)\{1 - P_j + n\phi_j(r)P_j\}\{1 - P_j + \phi_j(r)P_j\}^{n-2} \tag{38}$$
$$(r = 1 \cdots, M_j - m_j ; j = 1, \cdots, N)$$

so that

$$\begin{aligned} V(\bar{\bar{z}}^*_d) &= n^{-1} \sum_{j=1}^{N} P_j S^2_j(z) \sum_{r=1}^{M_j-m_j} \phi_j(r)/(M_j - r_j) \\ (39) \qquad &\times \{1 - P_j + n\phi_j(r)P_j\} \{1 - P_j + \phi_j(r)P_j\}^{n-2} + \sigma^2_{bz}/n. \end{aligned}$$

9. An estimator better than $\bar{\bar{z}}^*_d$. If from the statistic T^*, we take away $\lambda_{(1)}, \cdots, \lambda_{(d)}$, we get another sufficient statistic

$$T = [\{u_{(i)} ; u_{(i1)}, \cdots, u_{(id_{(i)})}\} i = 1, \cdots, d]. \tag{40}$$

The lemma given below has been used to derive the probability distribution of T.

LEMMA 9.1.

$$\sum{}' \frac{n!}{\lambda_1! \cdots \lambda_d!} \prod_{i=1} x_i^{\lambda_i} \left\{ \binom{d_i}{m_i}^{\lambda_i} - \binom{d_i}{1}\binom{d_i - 1}{m_i}^{\lambda_i} + \cdots \right\}$$

$$(41) \qquad = \sum_{r=0}^{\Sigma d_i - 1} (-1)^r \sum{}_1 \binom{d_1}{\alpha_1} \cdots \cdots \binom{d_d}{\alpha_d} \left[x_1 \binom{d_1 - \alpha_1}{m_1} + \cdots + x_d \binom{d_d - \alpha_d}{m_d} \right]^n$$

where $\sum'$ *stands for the summation over all positive integral* λ_i*'s such that* $\sum \lambda_i = n$, *and* $\sum_1$ *stands for the summation over non-negative* α_i*'s such that* $\alpha_1 + \alpha_2 + \cdots + \alpha_d = r$. $\binom{d_i}{r}$ *is to be regarded as zero for* $r > d_i$.

PROOF. The proof is obtained on observing that

$$\prod_{i=1}^{d} x_i^{\lambda_i} \left\{ \binom{d_i}{m_i}^{\lambda_i} - \binom{d_i}{1}\binom{d_i - 1}{m_i}^{\lambda_i} + \cdots \right\}$$

$$(42) \qquad = \sum_{r=0}^{\Sigma d_i - 1} (-1)^r \sum{}_1 \binom{d_1}{\alpha_1} \cdots \binom{d_d}{\alpha_d} \prod_{i=1}^{d} \left[x_i^{\lambda_i} \binom{d_i - \alpha_i}{m_i}^{\lambda_i} \right]$$

and taking the sum of the left side of (41) for all λ_i's such that $\sum \lambda_i = n$ and observing by direct expansion of the right side of (41) that the terms of the type (42) vanish if some $\lambda_i = 0$.

Clearly from (30), the probability of getting a sample with a given T is

$$P(T) = \sum{}' \frac{n!}{\lambda_{(1)}! \cdots \lambda_{(d)}!} \prod_{i=1}^{d} [p_{(i)}]^{\lambda_{(i)}} \Big/ \left[\binom{M_{(i)}}{m_{(i)}} \right]^{\lambda_{(i)}}$$

$$(43) \qquad \cdot \left\{ \left[\binom{d_{(i)}}{m_{(i)}} \right]^{\lambda_{(i)}} - \binom{d_{(i)}}{1} \left[\binom{d_{(i)} - 1}{m_{(i)}} \right]^{\lambda_{(i)}} + \cdots \right\}$$

where the summation $\sum'$ has been defined in (41).

It follows from Lemma 9.1 that

$$P(T) = \sum_{r=0}^{\Sigma d_{(i)} - 1} (-1)^r \sum{}_1 \binom{d_{(1)}}{\alpha_1} \cdots \binom{d_{(d)}}{\alpha_d}$$

$$(44) \qquad \cdot \left[p_{(1)} \binom{d_{(1)} - \alpha_1}{m_{(1)}} \Big/ \binom{M_{(1)}}{m_{(1)}} + \cdots + p_{(d)} \binom{d_{(d)} - \alpha_d}{m_{(d)}} \Big/ \binom{M_{(d)}}{m_{(d)}} \right]^n$$

where the summation $\sum_1$ has been defined in (41).

The statistic T is smaller than T^* (Definition 2.7). Therefore, any estimator which depends on T^* can again be uniformly improved by the use of Rao-Blackwell theorem. We, therefore, have the following theorem.

THEOREM 6. *An estimator uniformly better than* $\bar{\bar{z}}_d^*$ (*hence better than* $\bar{z}_n$ *Theorem* 2) *is given by*

$$(45) \qquad \bar{\bar{z}}_d = \sum_{i=1}^{d} c_{(i)} \bar{z}_{d_{(i)}}$$

where

$$c_{(i)} = p_{(i)} \left\{ \sum_{r=0}^{\Sigma d_{(i)}-1} (-1)^r \textstyle\sum_1 \binom{d_{(1)}}{\alpha_1} \cdots \binom{d_{(d)}}{\alpha_d} \binom{d_{(i)} - \alpha_i}{m_{(i)}} \Big/ \binom{M_{(i)}}{m_{(i)}} \right.$$

$$\left. \cdot \left[p_{(1)} \binom{d_{(1)} - \alpha_1}{m_1} \Big/ \binom{M_{(1)}}{m_{(1)}} + \cdots + p_{(d)} \binom{d_{(d)} - \alpha_d}{m_{(d)}} \Big/ \binom{M_{(d)}}{m_{(d)}} \right]^{n-1} \right\} \Big/ P(T)$$

and $P(T)$ *is given by* (44).

PROOF. It is obvious that an estimator uniformly better than $\bar{\bar{z}}_d^*$ is given by

$$E[\bar{\bar{z}}_d^* \mid T] = \sum E([\lambda_{(i)}/n] \mid T) \cdot \bar{z}_{d_{(i)}} . \tag{46}$$

Moreover

$$E\left[\frac{\lambda_{(i)}}{n} \middle| T\right] = \sum{}' \frac{\lambda_{(i)}}{n} \cdot \frac{n!}{\lambda_{(1)}! \cdots \lambda_{(d)}!} \prod_{i=1}^{d} [p_{(i)}]^{\lambda_{(i)}} \Big/ \left[\binom{M_{(i)}}{m_{(i)}}\right]^{\lambda_{(i)}}$$
$$\cdot \left\{ \left[\binom{d_{(i)}}{m_{(i)}}\right]^{\lambda_{(i)}} - \binom{d_{(i)}}{1} \left[\binom{d_{(i)} - 1}{m_{(i)}}\right]^{\lambda_{(i)}} + \cdots \right\} \Big/ P(T). \tag{47}$$

An argument similar to that of Lemma 9.1 will show that

$$E[(\lambda_{(i)}/n) \mid T] = c_{(i)} \qquad (i = 1, \cdots, d). \tag{48}$$

This completes the proof.

COROLLARY 6.1. *When* $P_j = M_j / \sum M_j$ *and* $m_j = 1$, *the above estimator takes the simple form*

$$\bar{\bar{z}}_d = [\textstyle\sum d_{(i)}]^{-1} [\sum d_{(i)} \bar{z}_{d_{(i)}}]. \tag{49}$$

Though the estimator $\bar{\bar{z}}_d$ is better than $\bar{\bar{z}}_d^*$, it cannot be of much use in practice except in the trivial case when $P_j = M_j/\sum M_j$ and $m_j = 1$. It is better to rely on $\bar{\bar{z}}_d^*$ which, though less efficient than $\bar{\bar{z}}_d$, has the merit of simplicity.

10. Estimation of variance. The problem of estimating the variance of $\bar{\bar{z}}_d^*$ (or of $\bar{\bar{z}}_d$) is even more complicated and is not considered worthwhile to discuss it here. However, in practice, it is recommended to use

$$[n(n-1)]^{-1} \sum_{i=1}^{n} (\bar{z}_i - \bar{\bar{z}}_n)^2$$

as an estimator of $V(\bar{\bar{z}}_d^*)$. As it over-estimates $V(\bar{\bar{z}}_d^*)$, we will always be on the safe side.

Acknowledgment. The author wishes to express his thanks to Professor D. Basu for his valuable help and guidance, to Professor Chapman and to the referee for their helpful remarks.

REFERENCES

BAHADUR, R. R. (1954). Sufficiency and statistical decision functions. *Ann. Math. Statist.* **25** 423–462.

BASU, D. (1958). On sampling with and without replacement *Sankhyā* **20** 287–294.

BLACKWELL, DAVID and GIRSHICK, M. A. (1954). *Theory of Games and Statistical Decisions*. Wiley, New York.

DES RAJ and KHAMIS, S. H. (1958). Some remarks on sampling with replacement. *Ann. Math. Statist.* **29** 550–557.

FISHER, R. A. (1922). On mathematical foundations of theoretical statistics. *Philos. Trans. Roy. Soc. London Ser. A* **222** 309–368.

FRASER, D. A. S. (1957). *Nonparametric Methods in Statistics*. Wiley, New York.

HÁJEK, JAROSLAV (1959). Optimum strategy and other problems in probability sampling, *Časopis Pěst. Mat.* **84** 387–423.

HALMOS, P. R. and SAVAGE, L. J. (1949). Applications of the Radon-Nikodym theorem to the theory of sufficient statistics. *Ann. Math. Statist.* **20** 225–241.

HANUMANTHA RAO, T. V. (1962). An existence theorem in sampling theory. *Sankhyā* **24** 327–330.

LEHMANN, E. L. and SCHEFFÉ H., (1950). Completeness, similar regions, and unbiased estimation. *Sankhyā* **10** 305–340.

MAHALANOBIS, P. C. (1946). Recent experiments in statistical sampling in Indian Statistical Institute. *J. Roy. Statist. Soc.* **109** 325–370.

PATHAK, P. K. (1961a). On the evaluation of moments of distinct units in a sample. *Sankhyā Ser. A* **23** 415–420.

PATHAK, P. K. (1961). Use of 'order-statistic' in sampling without replacement. *Sankhyā Ser. A* **23** 409–414.

PATHAK, P. K. (1962). On simple random sampling with replacement. *Sankhyā Ser. A* **24** 287–302.

SUKHATME, P. V. (1953). *Sampling Theory of Surveys with Applications*. Indian Society of Agricultural Statistics, New Delhi.