

CONTRIBUTIONS TO THE GENERAL THEORY OF SAMPLING FINITE POPULATIONS
WITHOUT REPLACEMENT AND WITH UNEQUAL PROBABILITIES

by

JOHN CLEMENT KOOP

Institute of Statistics
Mimeo Series No. 296
August 1957

POSTSCRIPT

This work now published in the Mimeo Series of the Institute of Statistics, partly in response to requests for copies, is an approved thesis for a doctor's degree which was submitted to the Faculty of North Carolina State College in August 1957.

J. C. Koop

Raleigh, North Carolina
September, 1961

TABLE OF CONTENTS

	Page
I. INTRODUCTION	1
1.1 General remarks	1
1.2 Statement of the problem	3
1.3 Review of literature	5
1.4 Theory treated	13
1.5 Notation	14
II. DEFINITION OF THE PROBABILITY SYSTEM	18
III. ENUMERATION AND CONSTRUCTION OF ESTIMATORS	25
3.1 The axioms of sample formation	25
3.2 The seven general linear estimators	26
3.3 The enumeration of weights	28
3.4 General comments	29
IV. UNBIASED ESTIMATORS AND THE PROBLEM OF MINIMUM VARIANCE ESTIMATORS IN EACH CLASS	31
4.1 Introductory statement	31
4.2 Class one estimators	33
4.3 Class two estimators	43
4.4 Class three estimators	50
4.5 Class four estimators	55
4.6 Class five estimators	67
4.7 Class six estimators	73
4.8 Class seven estimators	77
4.9 Summary	80

Table of Contents (Continued)

	Page
V. FURTHER NOTES ON UNBIASED ESTIMATORS	83
5.1 General notes	83
5.2 Comparison of estimators	83
5.3 Multiple comparison of efficiency	88
VI. THE PROBLEM OF NEGATIVE ESTIMATES OF VARIANCE	90
6.1 Introductory remarks	90
6.2 The case for class two estimators	91
6.3 The case for $V(T'_3)$	92
6.4 An example for $V(T'_3)$	94
6.5 A further example for $V(T'_3)$ illustrative of an unwise choice of selection probabilities	98
6.6 Some interpretations of negative estimates of variance for T'_3	99
6.7 Estimators in class four	100
6.8 Estimators in class five and class seven	101
6.9 The case of $V(T'_6)$	101
6.10 The existence of probability systems for which $V(T'_2)$, $V(T'_3)$ and $V(T''_5)$ are always positive	103
VII. A NOTE ON OPTIMUM PROBABILITIES	107
7.1 Introductory note	107
7.2 A simple example	107
7.3 A formal statement of the problem of optimum probabilities in multivariate sampling with general cost functions	110

Table of Contents (Continued)

	Page
VIII. SUMMARY AND CONCLUSIONS	112
8.1 Summary	112
8.2 Conclusions	114
8.3 Suggestions for future work	116
LITERATURE CITED	118

I. INTRODUCTION

1.1 General Remarks

The historical basis of modern sampling theory can be found in the works of Bernoulli, Poisson and Lexis. However, it was not until the appearance of Neyman's (1934) paper, based on his work in Poland, that the theory of statistical sampling, as we know it today, came into prominence. In this paper the advantages of random versus purposive sampling, the principle of unbiasedness and minimum variance (due to Gauss but popularized by Markoff) and the theory of optimum allocation ^{1/} in stratified one-stage sampling were discussed.

The next paper having a direct bearing on the subject of this thesis was by Hansen and Hurwitz (1943) ^{2/}. In what amounts to a stratified two-stage sampling scheme they selected one primary unit per stratum at the first stage with probabilities proportionate to some measure of size, and at the second stage the elements in each primary unit were selected with equal probabilities and without replacement. They showed the resulting estimator to be unbiased.

Sampling with unequal probabilities arose partly because in practice the ultimate unit of analysis and the unit of sampling are not identical. If unequal clusters of ultimate units are sampled, then obviously the sampling variance, on the basis of equal probability sampling, can be expected to be larger than in the case when they

^{1/}Subsequently Thompson (1952) pointed out that these results were anticipated by A. Tschuprow in a paper published in two parts in Metron in 1923.

^{2/}Other papers of importance to the body of sampling theory in the intervening period, i.e., 1934-1943, are mentioned in the bibliography of this paper. I have omitted direct reference to them only because they have less bearing on this work.

are all equal. However, practical limitations (in the way of resources both human and monetary) may set a limit as to the way existing aggregates of ultimate units (e.g., farms in a minor civil division or houses found in city block) may be arranged to have as far as possible equally sized clusters in terms of the ultimate units considered.

It is well known that if clusters, treated as sampling units, are selected with probabilities proportionate to their aggregate characteristics, and with replacement, the sampling variance of the mean of these aggregates, each weighted by the reciprocal of its respective selection probability, is zero. Further, in most practical situations, the aggregate characteristic of a cluster is highly correlated with the number of ultimate units (or size of the cluster) which is often known in advance.

All these considerations prompted Hansen and Hurwitz to select units with probabilities proportionate to some measure of size in order to reduce the variance of their estimator over that which would have been obtained on the basis of equal probabilities of selection. However, because only one first-stage unit was selected per stratum, it was not possible to calculate the sampling variance of the estimator.

The special theories of Midzuno (1950) and Horvitz and Thompson (1951, 1952) were attempts to generalize this approach in the selection of units. About this time, and later, various statisticians from India took up the problems of unequal probability sampling and proposed various sorts of estimators. Among them, Narain (1951) arrived at the same type of linear estimator of the population total as proposed by Horvitz and Thompson; Lahiri (1951), by a different approach, obtained essentially the same estimator as Midzuno's, which, in

functional form, retains the character of a ratio estimate; Das (1951) proposed an entirely different linear estimator. In a subsequent section all this work will be reviewed.

In any situation once the clusters (first-stage units) have been sampled with unequal probabilities (in the sense explained) the ultimate units (second-stage units) in each selected cluster can be sampled with equal or unequal probabilities. If there is a hierarchy of units, one nested within the other, then in a multi-stage procedure, sampling at one or more stages may be with equal or unequal probabilities, and with or without replacement. In this thesis we shall not be concerned with sampling beyond the first stage. Once results for the first stage are obtained the extension of the theory to cover two or more stages is almost immediate.

1.2 Statement of the Problem

In this thesis we shall be concerned with sampling a finite population without replacement of the units sampled after each draw and with unequal probabilities of selection.

To be specific the solution of the following questions, or their discussion, will be attempted as a contribution to the special theory (in the sense which will be explained in the next section) already initiated by Horvitz and Thompson (1952).

- (i) During the past seven years various estimators alluded to above have been proposed. Can they all be placed in a logical scheme starting with a few simple characteristics (let us say axioms) on the way a sample is formed? Also can general linear estimators and a most general linear

estimator (from which it may be possible to derive all known estimators) be formulated on the basis of these axioms?

- (ii) Can minimum variance estimators in the sense of Gauss and Markoff (Neyman, 1934, 1952) be determined for each general estimator?
- (iii) From certain known results in the theory for infinite populations (which will be discussed) there are indications that these minimum variance estimators cannot be used because they may have weights which may not be independent of the properties of the population. Attempts will be made to approximate or simulate them in all such cases.
- (iv) Are there any more unbiased estimators still undiscovered, and if so what are their properties?
- (v) The difficulties in comparing the efficiencies of known unbiased estimators will be explored.
- (vi) The conditions under which certain known unbiased estimators yield negative estimates of variance will be determined. Also whether or not probability systems exist for certain classes of estimators which will always yield positive estimates of variance will be explored.
- (vii) An attempt to solve the problem of optimum probabilities for a given estimator will be made by way of an example. The problem in a general way will be formulated taking into consideration factors of cost in the sampling of more than one characteristic, i.e., multivariate sampling.

1.3 Review of Literature

First and foremost it is instructive to review the work on equal probability sampling. The theory of sampling finite populations with equal probabilities and without replacement of the elements selected after each draw took shape with Isserlis' (1916) paper, "On the Conditions Under Which the 'Probable Errors' of Frequency Distributions Have a Real Significance." Among other results he obtained the first four moments of the mean. It appears that this paper did not come into prominence among statisticians concerned with the problems of sampling because of its very title. Isserlis (1918) reproduced these results in another paper more suggestive of its contents, and Edgeworth (1918), noting his work, derived his results in a novel way.

A Russian statistician, Tschuprow (1923), reported the same results. He also gave the formula for the optimum allocation of units in stratified sampling. He was aware of Isserlis' work and in a footnote to his 1923 paper he also referred to the work of an Italian statistician, G. Mortara (1917), in which, to use his own words, "the special formula for the standard error of the average in the case of 'unreplaced tickets' " was given. He further explained that his results were obtained about the same time as Isserlis' and were subsequently published in a Scandanavian journal (1918). In this paper he considered the population to be changing at each draw and he obtained the first four moments for the mean as a special case of his theory when the elements are not replaced after each draw.

Neyman (1925)^{1/}, unaware of previous work, obtained the same formulas for the first four moments of the mean and the first two

^{1/}He then wrote under the name of J. Splawa-Neyman.

moments of the sampling variance. His method of derivation of these moments was much more direct than either Isserlis' or Tschuprow's and without the complexity of notation and algebraic involvement which was characteristic of the latter's work. Indeed it seems that the work of Tschuprow was overlooked among writers in the English language for this reason.

Church (1926) obtained the third and fourth moments of the sampling variance of the mean. It is of interest to note that he started out to derive the third moment by first determining the uncorrected third moment as

$$2^M_3 = \frac{1}{M_{C_N}} \sum \left\{ \left[\frac{S(x_1^2)}{N} - \left(\frac{S(x_1)}{N} \right)^2 \right]^3 \right\}$$

where S represents summation over the N values of the sample and \sum represents the summation over the M_{C_N} samples. (There were M elements for his universe, and N in the sample.) It will be noted that $1/M_{C_N}$ is the probability of obtaining a given sample of N . Hence this uncorrected moment is merely the expectation of the expression in curled brackets. In the algebra associated with the derivation of expectations in unequal probability sampling we perform similar operations.

Carver (1930) also obtained the first four moments of sample aggregate values by precisely the same approach as Church. If one reads a little deeper into Carver's work it appears that he may have recognized that the sample aggregate weighted by $N C_n / N-1 C_{n-1}$ provides an unbiased estimator of the population total (and indeed the best linear unbiased estimator). This statement may appear a little naive because in the context of equal probability sampling the weight in

question reduces to N/n but in the context of unequal probability sampling it assumes significance as we shall see later.

The work of Hansen and Hurwitz (1943) has already been referred to in some detail.

Midzuno (1950) extended the Hansen-Hurwitz approach to sampling a combination of n elements with a probability proportionate to some measure of size of the combination. From a formula given in his subsequent paper (1952) and also as noted by Horvitz and Thompson (1952) and Sen (1952), this probability is equal to the total probability of selecting the first unit with a probability proportionate to a measure of size of the first unit and the remainder with equal probabilities and without replacement.

Lahiri (1951) was concerned with the bias in the ratio estimate and he set out to find an estimator which while retaining the character of a ratio estimate was also to be free from bias. The usual ratio estimate of the total for single-stage sampling is

$$\left[\frac{\sum_{i=1}^n x_i}{\sum_{i=1}^n y_i} \right] \cdot T_y,$$

where the y 's are the auxiliary characteristics which in many situations are known, T_y is their total, and the x 's are the characteristics under study of the n elements sampled. He saw that if a sampling scheme could be devised in which the probability of obtaining a given group of n elements is proportional to the total measure of their corresponding auxiliary characteristics, then the estimate would be unbiased. He devised such a scheme. It must be said his approach was much more direct than that of Midzuno, of whose work he was unaware, and he arrived at his sampling scheme, as he confesses, by geometrical intuition.

Sen (1952) worked out the consequences of the ideas contained in Midzuno's 1950 paper, and among other results he obtained independently the same type of ratio estimate as Lahiri's in the case of sampling without replacement and in two stages. He was concerned mostly with samples of size $n = 2$ (Sen et al., 1954).

The foregoing account will show that there was a unity of theme in the works of Midzuno, Lahiri, and Sen. In their investigations they assumed from the beginning that the probability with which the sample should be drawn should be some function of known measurable characteristics related to the elements under study. Subsequently it will be shown that their estimator is a particular case of a more general estimator formulated by Horvitz and Thompson (1952).

A more general approach to the problem of sampling finite populations with unequal probabilities and without replacement of elements was formulated by Horvitz and Thompson (1951) and subsequently this paper was published in full in 1952. In their theory the probabilities of selection were completely arbitrary at each draw. As we shall see later the generality of this approach is of considerable theoretical advantage because it provides the basis for determining the optimum probabilities in any defined sense. They formulated three classes of linear estimators of the population total with coefficients for each class depending on the order of draw, the presence or absence of an element in a sample and the particular sample in question. Subsequently the general estimators corresponding to each of these features of sample formation will typify what will be designated as class one, class two and class three respectively. However, the logical consequences of these ideas were not explored except in

respect of the estimator with coefficients depending on the appearance or nonappearance of the elements in the sample. These coefficients were determined (a) by imposing the condition of unbiasedness, i.e., requiring the expected value of the estimator to be equal to the quantity under estimation (the total in this case); and (b) by requiring that they shall be independent of the properties of the population, i.e., insisting further that the expected value of the estimator shall be identically equal to the quantity under estimation. In sequel the distinction between conditions (a) and (b) will become more evident. The sampling variance of this estimator was given as well as its estimator. The extension of the theory for the case of two-stage sampling was also given. They also attempted to determine those values of the initial selection probabilities that would result in a low variance on the basis of information on an auxiliary correlated variable y , known for each element of the universe.

Yates and Grundy (1953) noted that the Horvitz-Thompson estimator of variance can assume negative values. They proposed another unbiased estimator believed to be less often negative.

Narain (1951) among other questions considered the estimation of the mean value of a population of clustered elements by sampling in two stages, sampling first the cluster (treated as the primary unit) with unequal probabilities and without replacement, and then a constant number of elements in each cluster at the second stage with equal probabilities. He chose as his estimator the unweighted mean of the characteristic of the elements under study, and for sampling primary units of size two he determined from the condition of unbiasedness the expression for the initial probabilities. These turned out to be

complicated functions of the cluster sizes. To circumvent this difficulty he proposed the selection of groups of primary units, which he termed "group sampling," from among all possible combinations with certain predetermined probabilities for each group or sample. In order that the estimator may be unbiased, he chose probabilities of including a given unit proportional to the product of the sample size and cluster size. He wrote down the expressions for these probabilities for the case of samples of two units from a universe of four. Narain's estimator is of the same type as Horvitz and Thompson's estimator described earlier. The sampling variance of his estimator is also of the same type as the one obtained by Horvitz and Thompson for their two-stage case.

Das (1951) also proposed an unbiased estimator (for the case of single-stage sampling) which was entirely different from any described in the foregoing account. The weights attached to each observed characteristic were dependent partly on the order of the draw and partly on the element appearing at the draw. He also obtained the expression for the sampling variance and its estimator, which can assume negative values.

Des Raj (1956) proposed two new unbiased estimators with coefficients which depend partly on the element itself and partly on the order of the draw. In the same paper he proposed a third but he was apparently unaware that it was essentially the same as Das'. The estimates of variance for these estimators can take negative values except for a special form of one of his estimators. He further gave the extensions of the theory for multi-stage sampling.

Earlier (1954) he had explored the special form of the Midzuno-Lahiri-Sen estimator and he stated that it was the only unbiased estimator in the class. Subsequently it will appear that this statement is not true. He also gave the appropriate theory for the multi-stage and two-phase extensions of this estimator, but in regard to the former problem in a more straightforward manner than Sen. The estimators of the variance in each case discussed in this 1954 paper can also be negative. The conditions under which the estimator of variance shall be positive even for single-stage sampling have not been explored.

Godambe (1955) put forward a unified theory of sampling finite populations. He defined symbolically a system of probabilities for the selection of each element which are functions specified by the following arguments: the individual selected, the particular draw, and the sequence of elements preceding the individual element selected. For example, probabilities associated with drawing any element in a stratified sampling scheme can be expressed with a unity of notation. On the basis of these probabilities the total probability of any sequence of elements appearing can be uniquely determined. What may perhaps be termed a weakness in the generality of approach manifests itself here; the question as to what is the total number of possible samples, given that there are N distinct elements in the universe, and which are themselves the units of sampling, is not clear. Clearly the answer to this question will depend on how the elements are split up into strata and whether when sampling they are, or are not, replaced. Ignoring the problem of stratification, let us enumerate the number of possible samples appropriate to the following situations when n drawings are made:

- (i) The elements are not replaced after each draw.
- (ii) For the first r_1-1 draws, the elements are replaced, but at the r_1^{th} draw the element drawn is not replaced; next, at the $(r_1+1)^{\text{th}}$ and up to $(r_1+r_2-1)^{\text{th}}$ draw the elements are replaced but at the $(r_1+r_2)^{\text{th}}$ draw the element drawn is not replaced and so on up to the last set of r_k draws. In all there are $\sum_{q=1}^k r_q = n$ draws and $1 \leq r_q < n$.

- (iii) The elements are replaced after each draw.

A "chaotic" situation can also arise if, prior to any draw, after deciding not to replace, we reverse our decision and put back some or all of the elements previously removed. Conceptually, a theory of sampling taking into account such type of "erratic" behavior on the part of the sampler is possible. In the climate of a unified theory such a suggestion is admissible, even though the appropriate situation would occur infrequently. However, in view of its nature, we shall not discuss this question further.

In situation (i) there are only $N(N-1)\cdots(N-n+1)$ possible samples; in (ii) there are $N^{r_1}(N-1)^{r_2}\cdots(N-k+1)^{r_k}$ possible samples; and in (iii) there are N^n possible samples. In regard to (iii) there may be cases when we desire, say, d distinct units in the sample and the drawing may proceed up to the point when d distinct units are obtained. In this case it is possible for $n > N$. Sukhatme and Narain (1952) have discussed the point in relation to certain multi-stage designs. The same remarks apply to situation (ii). I have enlarged on this topic since it helps to bring into relief the special theories possible within the total field of a unified theory. In this thesis I shall limit myself to situation (i), i.e., where the elements are not replaced after each draw.

To continue with Godambe's ideas, it must be said that the general estimator which he proposed is referable to any of the above situations relating to the question of replacement and nonreplacement. He proves that in such a class of estimators there is no unique minimum variance estimator; i.e., one having coefficients which are entirely independent of the properties of the population.

In my view his proof is not convincing for the following reasons. Prior to minimizing the variance function subject to N restraints (stemming from the condition of unbiasedness) he has eliminated from one part of his variance function the unknown coefficients or weights (which are the subject of study) by the use of the equations of restraint. Therefore the position is that he may not have been studying the variance function of his most general type of linear estimator, but some other modified function. Also if the classical method of Lagrange is strictly adhered to, i.e., simply augmenting the function under study without in any way tampering with its form, then a different set of equations may be obtained.

1.4 Theory Treated

In the special theory which is the concern of this thesis, I will not follow the above procedure of minimization for the reasons already given.

In the title of the thesis the words "general theory" appear. Here there are semantic difficulties in how this theory should be termed. I have referred to it in the body of the thesis merely as "the special theory" to be logically consistent with the words "a unified theory" already appropriated, and I believe rightly so, by Godambe. It is to be interpreted

as general partly in the sense that the numerical values of the probabilities to be used in any practical situation are discretionary and partly in the sense that the system of ideas on the basis of which the estimators are constructed are relevant to any scheme of sample formation where the elements, as units of sampling, are drawn one at a time. I stress this because it is possible to construct a theory of sampling--somewhat "odd" but nevertheless rational--where groups or clusters not necessarily having the same number of elements are simultaneously sampled. Effectively, therefore, "sample size" differs in the drawing of each group. It is possible to construct an unbiased estimator of the total and an unbiased estimate of the sampling variance. It will be noted that such a theory is not encompassed by the present unified theory.

1.5 Notation

I have followed the standard notation of denoting an element by u_m with the subscript denoting the particular element in question. Further, a sub-subscript n attached to a subscript m , such as in u_{m_n} , indicates that the m^{th} element appears at the n^{th} draw.

Next, x_i denotes the value of a characteristic of u_i ; and sometimes a sub-subscript t attached to i , such as in x_{i_t} , indicates that u_i which bears the characteristic x_i appears at the t^{th} draw.

In regard to arbitrary probabilities, I have followed the notation of Horvitz and Thompson, and of Des Raj (1956), by writing "p" in a lower case letter with the subscript denoting the element and the sub-subscript denoting the order of the draw. The letters attached to p as superscripts indicate the elements which have already appeared.

For example, $p_{i_3}^{st}$ denotes the probability of drawing u_i at the third draw after u_s and u_t have appeared at the first and second draws.

The unconditional probability for drawing a given element at a particular draw is denoted by a capital "P" with the subscript denoting the element and the sub-subscript the order of the draw under consideration. Thus P_{i_t} denotes the probability of drawing u_i at the t^{th} draw. A similar notation P_{i_t, j_w} denotes the unconditional probability of drawing u_i and u_j at the t^{th} and w^{th} draws respectively, the t^{th} draw preceding the w^{th} , also a comma separates i and j which appear on the same line.

Horvitz and Thompson's notation for the probability of including u_i in a sample of n is $P(u_i)$; I have changed this simply to P_i to be notationally consistent with the result

$$\sum_{t=1}^n P_{i_t} = P_i,$$

the meaning of which will be apparent. Incidentally Thompson (1952) used P_i in his thesis. Similarly for their $P(u_i, u_j)$, which denotes the probability that u_i and u_j appear in a sample of n , I have written simply P_{ij} , the subscripts i and j appearing in the same line. This again is notationally consistent with the result

$$P_{ij} = \sum_{t,w} P_{i_t, j_w}.$$

Cansado (1955) developed a system of notation in which the u 's always appear in the representation of arbitrary probabilities and unconditional probabilities. However, his formulas require more space,

and typographically speaking more labor in composition. The omission of the u 's saves much space.

Also P_{ij} is essentially Singh's (1954) notation for Horvitz and Thompson's $P(u_i, u_j)$ and P_{i_t} is Des Raj's (1956) notation for unconditional probabilities.

The lower case letter \underline{s} identifies a given sample of \underline{n} elements in the context used. Thus $P_{\underline{s}}$ represents the total probability of a given sample, \underline{s} , of \underline{n} elements appearing and $P_{\underline{s}(o)}$ represents the joint probability of a given sample of \underline{n} elements appearing in a specified order. Also there is notational consistency in the result

$$P_{\underline{s}} = \sum P_{\underline{s}(o)}$$

where the summation is taken over all the $n!$ possible ways in which the \underline{n} elements can appear.

The capital letter \underline{S} indicates the total number of samples in all possible orders of appearance and \underline{S} the total number of distinct samples.

Regarding summation symbols, $\sum_{i \in \underline{s}}$ denotes summation over all \underline{n} elements which are included in sample \underline{s} , a typical element being u_i . Also $\sum_{\underline{s} \ni i}$ denotes summation over all samples which include u_i . Symbols such as

$$\sum_{i < j \in \underline{s}}, \sum_{i \neq j \in \underline{s}}, \sum_{s > i, j}, \text{ and } \sum_{s > i_t}$$

bear similar meanings.

Seven Greek letters, $\alpha, \beta, \gamma, \delta, \theta, \phi, \psi$, which denote weights to be attached to the seven different classes of estimators, have subscripts and sub-subscripts whose meaning will be indicated when they are used.

The population total of a characteristic in question is denoted by \underline{T} . However, this \underline{T} with a numerical subscript, such as T_2 , represents a class of estimators indicated by the numeral each of which is an estimator of this population total.

The symbols not mentioned here except those in standard usage are explained in each particular context.

II. DEFINITION OF THE PROBABILITY SYSTEM

Consider a population of N elements, u_1, u_2, \dots, u_N . Each element has a certain number of measurable characteristics. By taking a sample in a way to be specified, and observing their characteristics, it is proposed to estimate the aggregate of each of these measurable characteristics pertaining to the population. It is given that the probabilities of selection at each draw are arbitrary in the sense that the choice of the numerical values which they may assume are discretionary. Later we shall see how these probabilities may be chosen. Also the reason for leaving them arbitrary will reveal itself in the course of the discussion.

The notion of arbitrary probabilities of selection for each draw in sampling finite populations is due to Horvitz and Thompson (1952), and the definition of the probability system which follows is implicit in their 1952 paper. The later elaborations in the display of this idea and notation (the need for which was noted by the two authors) were made independently by Singh (1954), Cansado (1955) and Des Raj (1956).

Prior to the first draw let the probability of selection of the i^{th} element be p_{i_1} ($i=1,2,\dots,N$) where

$$p_{i_1} \geq 0, \text{ and } \sum_{i=1}^N p_{i_1} = 1.$$

Again prior to the second draw, when one element has already been removed, let the probability of selecting the i^{th} element at the second

draw be either

$$p_{i_2}^1 \text{ or } p_{i_2}^2 \dots, p_{i_2}^t, \dots \text{ or } p_{i_2}^N ; \sum_{i(\neq t)=1}^N p_{i_2}^t = 1, t=1,2,\dots,N$$

depending on which of the \underline{N} elements have been removed at the first draw. The superscript \underline{t} in the symbol $p_{i_2}^t$ indicates the element which has been removed from the universe. Thus there are \underline{N} possible sets of selection probabilities, speaking in an operational sense, or \underline{N} possible sets of probability distributions, speaking in an abstract sense. Finally, prior to the n^{th} draw, let there be ${}^N C_{n-1}$ sets of selection probabilities for the remaining $\overline{N-n-1}$ elements, depending on which of the $n-1$ elements have been selected at the preceding $n-1$ draws; symbolically these ${}^N C_{n-1}$ sets of selection probabilities may be written as

$$p_{i_n}^{t_1 t_2 t_3 \dots t_{n-1}}$$

where

$$\sum_{i(\neq t_1 \neq t_{n-1})}^N p_{i_n}^{t_1 t_2 t_3 \dots t_{n-1}} = 1$$

The superscripts $t_1 t_2 \dots t_{n-1}$ indicate that the elements t_1, t_2, \dots, t_{n-1} have already been drawn prior to i_n .

The above probability system defines the way in which the sample is to be drawn. It is abstract and we may say that the frame which describes the elements and the technique of sampling by which these elements, as units of sampling, are selected with probabilities prescribed by the system, are its real counterparts.

Thus associated with drawing a sample of size n there are in all

$$D = 1 + {}^N C_1 + {}^N C_2 + \dots + {}^N C_{n-1}$$

possible sets of selection probabilities to be defined prior to each draw. In practice, of course, only n sets out of the D possible sets are used. We shall hereafter speak of these D sets of selection probabilities (or probability distributions) as the probability system, or sometimes simply as the system when there is no danger of ambiguity. It will be noted that in all there are

$$G = N + (N-1) {}^N C_1 + (N-2) {}^N C_2 + \dots + (N-n+1) {}^N C_{n-1} = \sum_{r=1}^n r {}^N C_r$$

arbitrary selection probabilities in the system when n elements are selected without replacement. Obviously there is an infinite number of probability systems including the one where the selection probabilities are equal at each draw. Subsequently it will appear that some are better than others in the sense that for the estimation of a given characteristic, a given estimator (irrespective of whatever properties it may already possess), using a given probability system, will have smaller variance than when an alternative system is used. One topic of research in the theory of unequal probability sampling is the search for such systems.

Given the above probability system, the a priori probabilities for the selection of a given element at a given draw, for the selection of a group of elements in a specified order, etc., are needed for the development of the theory and can all be evaluated.

The various formulas on probabilities given below are contained in the articles of Singh (1954), Cansado (1955), Des Raj (1956) and

those of a less involved nature in Horvitz and Thompson (1952), Narain (1951), and in Sukhatme's book (1953). They are reproduced here in order that the discussion may be coherent.

Let a sample of size n be selected (without replacement) according to the above probability system and let

$$x_{i_1}, x_{j_2}, \dots, x_{l_{n-1}}, x_{m_n}$$

be the values of a given characteristic observed on the i^{th} , j^{th} , ..., l^{th} , m^{th} selected elements (or sample units) respectively, according to the order indicated in the sub-subscripts; i.e., element u_{i_1} bearing the characteristic x_{i_1} is drawn first, u_{j_2} bearing the characteristic x_{j_2} is drawn second and so on.

Denoting by P_{i_t} the probability of u_{i_t} being selected at the t^{th} draw, the expressions for these unconditional probabilities are as follows:

$$P_{i_1} = p_{i_1}$$

$$P_{i_2} = \sum_{j(\neq i_1)=1}^N p_{j_1} p_{i_2}^j$$

(Here there are ${}^{N-1}C_1$ terms under summation.)

$$P_{i_3} = \sum_{s \neq t (\neq 1)=1}^N p_{s_1} p_{t_2}^s p_{i_3}^{st}$$

(Here there are $2!{}^{N-1}C_2$ terms under summation.)

⋮

Finally

(Here there are $(n-1)!{}^{N-1}C_{n-1}$ terms under the summation sign.)

$$P_{i_n} = \sum_{j \neq k \neq \dots (\neq i)=1}^N p_{j_1} p_{k_2}^j \dots p_{l_{n-1}}^{jk \dots s} p_{i_n}^{jk \dots t}$$

In the expression for P_i above, the symbols j, k, \dots, t appearing as superscripts in $p_{i_n}^{jk\dots t}$ denote the $n-1$ elements u_j, u_k, \dots, u_t which are drawn prior to u_i , the element drawn at the n^{th} draw. If we sum these unconditional probabilities we have the probability of u_i appearing in a sample of n . Denoting this probability by P_i we have

$$P_i = \sum_{t=1}^n P_{i_t}$$

Further

$$\sum_{i=1}^N P_{i_t} = 1, \text{ for } t=1, 2, \dots, n.$$

If P_{i_t, j_w} denotes the probability that the i^{th} element is selected at the t^{th} draw and the j^{th} element at the w^{th} draw ($w > t$), then

$$P_{i_t, j_w} = \sum_{\substack{a \neq b \neq \dots \\ (\neq i) \neq \dots \\ (\neq j)}}^N p_a p_b^a \dots p_{i_t}^{ab\dots e} p_{k_{t+1}}^{ab\dots ei} p_{j_w}^{ab\dots eif\dots lm}$$

In the above expression there are $(w-2)! \binom{N-2}{w-2}$ terms under the summation sign. The symbols appearing as superscripts in $p_{j_w}^{ab\dots eif\dots lm}$ denote the $t-1$ elements u_a, u_b, \dots, u_e which were drawn prior to u_i , the one element u_i which appeared at the t^{th} draw, and the $w-(t+1)$ elements $u_f \dots u_m$ which appeared after the t^{th} but preceding the w^{th} draw. Also we will find

$$P_{i_t} = \sum_{j(\neq i)=1}^N P_{i_t, j_w}$$

and since

$$\sum_{i=1}^N P_{i_t} = 1, \text{ for all } t$$

we have

$$\sum_{i=1}^N \sum_{j(\neq i)=1}^N P_{i_t, j_w} = 1.$$

Denoting by P_{ij} the probability of including the i^{th} and the j^{th} elements in a sample of size \underline{n} , it will be found that

$$P_{ij} = \sum_{t=1}^n \sum_{w(\neq t)=1}^n P_{i_t, j_w}$$

and further we will find

$$(n-1)P_i = \sum_{j(\neq i)=1}^N P_{ij}$$

Further, if $P_{s(o)}$ denotes the probability that a given sample \underline{s}

of elements $u_{a_1}, u_{b_2}, \dots, u_{k_{n-1}}, u_m$ appear in the order indicated by the sub-subscripts, then

$$P_{s(o)} = P_{a_1}^a P_{b_2}^a \dots P_{m_n}^{ab\dots k}$$

and the total probability of obtaining \underline{s} irrespective of order is given by

$$P_s = \sum P_{s(o)}$$

where the summation is taken over all possible orders of appearance of $u_a u_b \dots u_k u_m$, there being $n!$ terms in all. Taking into account order it will be evident that there are $n! {}^N C_n = S$ samples of \underline{n} elements

each set of \underline{n} having probability $P_{s(o)}$ of appearing, which are not necessarily the same even for samples composed of the same elements. Also it will be evident that there are only ${}^N C_n = S'$ distinct samples, each distinct set of \underline{n} elements having a total probability P_s of appearing. These results have a bearing on what is to follow.

Finally, it will be apparent that

$$P_i = \sum_{s \supset i} P_s ,$$

and

$$P_{i_t} = \sum_{s \supset i_t} P_{s(o)}$$

where $s \supset i_t$ means all samples which include u_i , when it appears at the t^{th} draw.

III. ENUMERATION AND CONSTRUCTION OF ESTIMATORS

3.1 The Axioms of Sample Formation

In drawing a sample according to the probability system defined above (i.e., one element at a time and without replacement from the finite population of N), three features inherent in the nature of the process of selection are evident. They are as follows:

- (i) the order of appearance of the elements,
- (ii) the presence or absence of any given element (in the sample) which is a member of the population (or universe),
and
- (iii) the set of elements composing the sample considered as one of the total number possible (in repeated sampling according to the given probability system).

In regard to (ii) it may be pointed out that if an element is assigned an arbitrary probability of zero at each of the n draws, then it can never appear in a sample. The statements at (i), (ii), (iii), of course, are perfectly general and apply to sampling from any finite or infinite population, one element at a time, and with or without replacement. Further, as their veracity is self-evident, we may designate the three of them as axioms.

These features, inherent in the process of selection (and as a result sample formation), supply the bases for the construction of estimators. Thus we may say that the approach here in constructing estimators is a deductive one.

3.2 The Seven General Estimators

Let

$$T = \sum_{i=1}^N x_i$$

where x_i is the measure of a certain characteristic of u_i ($i=1,2,\dots,N$),
 be the population total to be estimated on the basis of the sample
 $x_{i_1}, x_{j_2}, \dots, x_{m_n}$.

Considering the order of appearance of the elements we have the
 estimator

$$T_1 = \sum_{t=1}^n \alpha_t x_t$$

where α_t ($t=1,2,\dots,n$) is the weight to be attached to the element
 selected at the t^{th} draw. (For convenience the subscripts identifying
 elements have been dropped from the x 's.)

Considering the presence or absence of an element in the sample,
 we have

$$T_2 = \sum_{i \in s} \beta_i x_i$$

where β_i ($i=1,2,\dots,N$) is the weight to be attached to the i^{th} element
 whenever it appears in the sample.

Considering the sample obtained as one of the set of all possible
distinct samples we have

$$T_3 = \gamma_s \sum_{i \in s} x_i ,$$

where $\sum_{i \in s}$ means summation over all elements included in the s^{th} sample,

and where γ_s ($s=1,2,\dots,S'$) is the weight to be attached to the s^{th} sample whenever it is selected. It will be recalled that there are only $S' = {}^N C_n$ distinct samples of \underline{n} elements.

Further four other estimators are possible, taking into consideration two or all of the features at a time. Thus we have

$$T_4 = \sum_{t=1}^n \delta_{i_t} x_t,$$

where δ_{i_t} ($i=1,2,\dots,N$; $t=1,2,\dots,n$) is the weight to be attached to the i^{th} element whenever it appears at the t^{th} draw.

$$T_5 = \sum_{i \in s} \theta_{s_i} x_i,$$

where θ_{s_i} ($i=1,2,\dots,N$; $s=1,2,\dots,S'$) is the weight to be attached to the i^{th} element whenever it appears in the s^{th} sample, and where $\sum_{i \in s}$ means summation over all \underline{i} elements (there are \underline{n} of them) included in \underline{s} , considered without regard to the order of appearance of the elements.

$$T_6 = \phi_s \sum_{i \in s} x_i$$

where ϕ_s ($s=1,2,\dots,S$)^{1/} is the weight to be attached to the s^{th} sample whose elements appear in a specified order and where $\sum_{i \in s}$ means summation over those elements in \underline{s} .

^{1/} It may be recalled that there are $S=n! {}^N C_n$ samples, taking into consideration order of appearance, each of which has a probability $P_{s(o)}$ of occurring.

$$T_7 = \sum_{i \in S} \psi_{s_{i_t}} x_i$$

where $\psi_{s_{i_t}}$ ($t=1,2,\dots,n$; $s=1,2,\dots,S$) is the weight to be attached to the i^{th} element appearing at the t^{th} draw in the s^{th} sample (whose elements, of course, appear in a specified order).

Thus there are seven different estimators of the population total. Each estimator, which is a linear function of the characteristics of the observed elements of the population, constitutes a class in the sense that for any given probability system there exists a set or sets of weights which can be determined from conditions arising out of the application of the criterion that the estimator shall be unbiased. In each class there will be as many sets of weights as there are linear functions which satisfy the condition of unbiasedness, and we shall attempt to choose (in each class) the one which has minimum variance in the sense of Gauss and Markoff; i.e., the best linear unbiased estimator for the class in question. For reasons which will appear in the subsequent discussion, only these seven classes of estimators are possible and no others. In the next paragraph we shall enumerate the total number of weights in each class.

3.3 The Enumeration of Weights

In class one and class two it will be evident that there are n and N weights respectively. As there are ${}^N C_n$ distinct samples there will be ${}^N C_n$ weights in class three, one applicable to each sample. Regarding class four there will be Nn weights as each element has a different weight at each draw. In class five there are ${}^{N-1} C_{n-1}$ weights, since each element appears in only ${}^{N-1} C_{n-1}$ of the ${}^N C_n$

distinct samples. In enumerating the weights in class six we first note that the order of appearance of the elements is taken into consideration so that for any given set of n elements there are $n!$ ways in which samples can be formed; thus there will be $n!^N C_n$ weights for this class. For class seven we note first that there are $n!^N C_n$ samples considering order; since each element has a weight depending on its order, there will be in all $n \cdot n!^N C_n$ weights.

3.4 General Comments

The estimators T_1 , T_2 , and T_3 , in their most general form, which specify classes one, two, and three, were proposed by Horvitz and Thompson (1951, 1952)^{1/}. Earlier Midzuno (1950) proposed an estimator which will be shown to belong to class three. Independently, Lahiri (1951) gave a procedure for sampling a collection of units with probability proportional to the sum of their sizes and his estimator, like Midzuno's, retained the character of a ratio estimate. Narain (1951) also independently proposed an estimator which belongs to class two, and his approach in arriving at it was essentially the same as Horvitz and Thompson's. Das (1951) also arrived at an estimator which on examination will be found to belong to class four. Sen (1952) elaborated on a particular form of Midzuno's estimator. All his estimators will be shown to belong to class three. Godambe (1955) formulated a unified theory, and his most general estimator, in the context of this special theory for a finite population when the units are replaced after each draw, is logically equivalent to T_7 . Des Raj

^{1/} They used the term "subclass." I have used the term class for reasons which will become apparent in Chapter IV.

(1956) proposed two estimators which will be shown to belong to class four.

Regarding Godambe's way of arriving at his most general type of linear estimator, it must be said that he did so after recognizing, as I also do, the deductive approach implicit in Horvitz and Thompson's work. However, he did not posit the three features of sample formation as axioms.

Finally we note that the following statement which appears on page 668 of Horvitz and Thompson's paper is pregnant with suggestions:

We have indicated above only three of the possible subclasses of linear estimators of T when sampling a finite universe without replacement.

The application of these axioms has resulted in four more general ones, including Godambe's.

IV. UNBIASED ESTIMATORS AND THE PROBLEM OF MINIMUM VARIANCE ESTIMATORS IN EACH CLASS

4.1 Introductory Statement

Having defined the estimators, we shall now determine the weights which give unbiased estimators in each class. As stated above, in each class there will be as many sets of weights as there are linear functions which satisfy the condition of unbiasedness. In each class we shall attempt to determine the linear function which has minimum variance, i.e., the best linear unbiased estimator.

Before we proceed it is pertinent to remark that the criteria of unbiasedness and minimum variance, when applied to populations non-stationary in the stochastic sense, do not always yield unbiased estimators which can be computed from sample data. Briefly the following example illustrates this proposition. Consider an infinite population; it is desired to estimate the mean value of a certain character of this population, which is continuously changing during the period when observations are taken. One element is drawn randomly at a certain epoch of time, another at a succeeding epoch, and so on. Altogether n stochastically independent elements are drawn. At the t^{th} epoch when x_t is the character observed, the population has a certain mean, μ_t , and a certain variance, σ_t^2 ($t=1,2,\dots,n$), which are both unknown. Let μ be the mean value of the character (during the period the observations are taken) which it is desired to estimate. Consider a linear function of the x 's,

$$F = a_1x_1 + \dots + a_nx_n$$

to estimate μ . We shall determine the weights a_t , such that

$$E(F) = \mu, \text{ and}$$

$$V(F) = E(F - \mu)^2 = \min.$$

It can easily be shown that^{1/}

$$F = \mu \sum_{t=1}^n \mu_t (x_t / \sigma_t^2) \sum_{t=1}^n (\mu_t^2 / \sigma_t^2)$$

and

$$V(F) = \mu^2 \sum_{t=1}^n (\mu_t^2 / \sigma_t^2).$$

Thus clearly the sample values alone are not sufficient to obtain an unbiased estimate of μ . We can, however, obtain some estimate of μ ; e.g., $\bar{x} = \sum x/n$, but we would not know how precise it would be. One may therefore ask, does a best linear unbiased estimator independent of the properties of the population (μ , μ_t , σ_t^2 , etc.) exist? The answer is certainly no to this question. Formally it does exist in some sense, although from a practical point of view it is useless. When the population is stationary, i.e., when all the respective moments relating to each epoch are equal, which implies that $\mu_1 = \dots = \mu_n = \mu$ and $\sigma_1^2 = \dots = \sigma_n^2 = \sigma^2$, then we have the well-known classic results:

$$F = (1/n) \sum x \text{ and}$$

$$V(F) = \sigma^2/n.$$

^{1/}When we consider the estimation of a mean value with data from samples of size one each coming from n infinite stationary populations with different means and variances the same results are also obtained. But it will be recognized that the problems are not logically equivalent.

Logically, therefore, these results could not have been possible without the formal existence of F . There is nothing new in the above formulas, but in the theory of estimation it appears that the underlying implications have little or no significance. However, in the theory of sampling finite populations (in the present context) there are somewhat analogous results which are of interest in a number of ways. In what follows we shall comment on these types of results when they appear.

4.2 Class One Estimators

Let us consider the general estimator in class one; i.e.,

$$T = \sum_{t=1}^n \lambda_t x_t .$$

First it is pertinent to inquire whether weights can be obtained for this estimator which are independent of the properties of the population. We have

$$\begin{aligned} E(T_1) &= \sum_{t=1}^n \lambda_t E(x_t) \\ &= \sum_{t=1}^n \lambda_t \sum_{i=1}^N P_{i_t} x_i = \sum_{i=1}^N x_i \sum_{t=1}^n \lambda_t P_{i_t} \end{aligned}$$

To satisfy the condition of unbiasedness we must have

$$\sum_{i=1}^N x_i \sum_{t=1}^n \lambda_t P_{i_t} = T = \sum_{i=1}^N x_i ,$$

and for the λ 's to be independent of the properties of the population we must insist on the condition of identity, i.e.,

$$\sum_{i=1}^N x_i \sum_{t=1}^n \lambda_t P_{i_t} = \sum_{i=1}^N x_i$$

which leads to the conditions

$$\sum_{t=1}^n \lambda_t P_{i_t} = 1, \text{ for } i=1,2,\dots,N.$$

These conditions are clearly impossible since n ($< N$) unknowns cannot simultaneously satisfy N equations. Thus this estimator cannot have weights which are independent of the properties of the population except in a special case which we shall derive as follows. Initially we shall attempt to find weights without insisting on the condition of identity, i.e., $E(T_1) = T$, and also such that $V(T_1) = \min$. Next from results arising out of this consideration we shall show that the best linear unbiased estimator exists with weights independent of the population values when the selection probabilities are made equal at each draw.

We have^{1/}

$$\begin{aligned} V(T_1) &= E \left[T_1 - E(T_1) \right]^2 \\ &= E \left[\sum_{t=1}^n \lambda_t (x_{i_t} - \sum_{i=1}^N P_{i_t} x_i) \right]^2 \\ &= \sum_{t=1}^n \lambda_t^2 E(x_{i_t} - \sum_{i=1}^N P_{i_t} x_i)^2 \\ &\quad + 2 \sum_{t < w} \lambda_t \lambda_w E(x_{i_t} - \sum_{i=1}^N P_{i_t} x_i)(x_{j_w} - \sum_{j=1}^N P_{j_w} x_j). \end{aligned}$$

^{1/}

I have reintroduced the subscripts i identifying the elements in the x 's as this is necessary for the sake of clarity in defining the moments.

Defining the following moments^{1/}

$$\mu_1(t) = E(x_{i_t}) = \sum_{i=1}^N P_{i_t} x_i,$$

for $t=1,2,\dots,n$

$$\mu_2(t) = E [x_{i_t} - \mu_1(t)]^2 = \sum_{i=1}^N P_{i_t} [x_i - \mu_1(t)]^2$$

$$\begin{aligned} \mu_{11}(t,w) &= E [x_{i_t} - \mu_1(t)] [x_{j_w} - \mu_1(w)] \\ &= \sum_{i=1}^N \sum_{j(\neq i)}^N P_{i_t, j_w} [x_i - \mu_1(t)] [x_j - \mu_1(w)], \end{aligned}$$

for $t \neq w = 1,2,\dots,n$,

which are analogues of those in equal probability sampling, we find

$$V(T_1) = \sum_{t=1}^n \lambda_t^2 \mu_2(t) + \sum_{t \neq w} \lambda_t \lambda_w \mu_{11}(t,w),$$

and the condition of unbiasedness becomes,

$$T_1 = \sum_{t=1}^n \lambda_t \mu_1(t).$$

To determine the λ 's which make T_1 a minimum variance estimator we set up the function,

$$H = V(T_1) + 2\lambda \left[\sum_{t=1}^n \lambda_t \mu_1(t) - T \right],$$

^{1/}The definitions of higher moments and product moments follow on the same lines, thus

$$\mu_a(t) = \sum_{i=1}^N P_{i_t} [x_i - \mu_1(t)]^a.$$

where λ is an undetermined Lagrange multiplier, and we solve the following equations:

$$\delta H / \delta \alpha_t = 0 = 2\alpha_t \mu_2(t) + 2 \sum_{w(\neq t)=1} \alpha_w \mu_{11}(t,w) + 2\mu_1(t), \quad (t=1,2,\dots,n)$$

together with the condition of unbiasedness. There are $n+1$ independent equations and $n+1$ unknowns (together with λ), and therefore we can solve for the μ 's. They are as follows

$$\begin{array}{cccc} \alpha_1 \mu_2(1) + \alpha_2 \mu_{11}(1,2) + \dots + \alpha_n \mu_{11}(1,n) + \lambda \mu_1(1) & = & 0 & \\ \vdots & & \vdots & \\ \alpha_1 \mu_{11}(n,1) + \alpha_2 \mu_{11}(n,2) + \dots + \alpha_n \mu_2(n) + \lambda \mu_1(n) & = & 0 & \\ \alpha_1 \mu_1(1) + \alpha_2 \mu_1(2) + \dots + \alpha_n \mu_1(n) & = & T & \end{array}$$

If μ represents the $n \times n$ product-moment and moment matrix,

$$\begin{array}{cccc} \mu_2(1), & \mu_{11}(1,2) & \dots & \mu_{11}(1,n) \\ \vdots & & & \\ \mu_{11}(n,1), & \mu_{11}(n,2) & \dots & \mu_2(n) \end{array}$$

α the column vector $(\alpha_1, \dots, \alpha_n)$, μ_0 the row vector $\mu_1(1), \dots, \mu_1(n)$

and O the null column vector, we have concisely the equations in matrix notation as

$$\left[\begin{array}{c|c} \mu & \mu_0' \\ \hline \mu_0 & O \end{array} \right] \begin{bmatrix} \alpha \\ \lambda \end{bmatrix} = \begin{bmatrix} O \\ T \end{bmatrix}, \text{ and thus}$$

$$\begin{bmatrix} \alpha \\ \lambda \end{bmatrix} = \left[\begin{array}{c|c} \mu & \mu_0' \\ \hline \mu_0 & O \end{array} \right]^{-1} \begin{bmatrix} O \\ T \end{bmatrix}.$$

Clearly the α 's will be functions of the moments and product-moments, say f_t , each multiplied by \underline{T} . Thus symbolically

$$\alpha_t = T \cdot f_t$$

Hence

$$V(T_1) = T^2 \sum_{t=1}^n f_t^2 \mu_2(t) + 2 \sum_{t \neq w} f_t f_w \mu_{11}(t, w)$$

or in matrix notation

$$V(T_1) = T^2 \cdot f \mu f'$$

where $f = [f_1, f_2, \dots, f_n]$ is a row vector.

When the selection probabilities at each draw are made equal, i.e.,

$$P_{i1} = 1/N, \quad \text{for } i=1, 2, \dots, N$$

$$P_{j2}^i = 1/(N-1), \quad \text{for } j=1, 2, \dots, N \text{ but } j \neq i$$

etc.,

we find

$$P_{i_t} = 1/N \text{ and } P_{i_t, j_w} = 1/N \cdot 1/(N-1)$$

so that

$$\left. \begin{aligned} \mu_1(t) = \mu_1 &= (1/N) \sum_{i=1}^N x_i, \\ \mu_2(t) = \mu_2 &= (1/N) \sum_{i=1}^N (x_i - \mu_1)^2 \end{aligned} \right\} \text{for } t=1, 2, \dots, n$$

$$\mu_{11}(t, w) = \mu_{11} = \frac{1}{N(N-1)} \sum_{i=1}^N \sum_{j(\neq i)=1}^N (x_i - \mu_1)(x_j - \mu_1) = \frac{-\mu_2}{N-1} \text{ for } t, w=1, 2, \dots, n.$$

Under these circumstances (of equal selection probabilities) the above results simplify to the following:

$$\alpha_t = N/n, \text{ for } t=1,2,\dots,n$$

so that

$$T_1^i = (N/n) \sum_{i=1}^n x_i$$

and after some algebra we will find

$$V(T_1^i) = N^2(\mu_2/n) (N-n)/(N-1) ,$$

which are familiar results, the formula for the variance being due to Isserlis (1916).

Before we pass on we make the following observations:

(a) When the selection probabilities are unequal the populations of elements existing prior to each draw, as evidenced by the values of the moment coefficients, are non-stationary in the stochastic sense and they become stationary only when the selection probabilities are equal.

(b) Evidently this is the analogue of the infinite population case noted earlier, and we note again that the property of stationarity confers advantages in that under these particular circumstances of sampling (the main feature being one element drawn at a time) unbiased estimation becomes possible.

It may be of interest to write down the general estimator for the case when $n=2$. The equations for solving α_1 , α_2 and λ are as follows:

$$\begin{bmatrix} \mu_2(1) & \mu_{11}(1,2) & \mu_1(1) \\ \mu_{11}(1,2) & \mu_2(2) & \mu_1(2) \\ \mu_1(1) & \mu_1(2) & 0 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \lambda \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ T \end{bmatrix}$$

and we find

$$T_{1(n=2)}'' = T \cdot \frac{[\mu_{11}(1,2)\mu_1(2) - \mu_2(2)\mu_1(1)]x_1 + [\mu_{11}(1,2)\mu_1(1) - \mu_1(2)\mu_2(1)]x_2}{\mu_1(1)[\mu_{11}(1,2)\mu_1(2) - \mu_2(2)\mu_1(1)] + \mu_1(2)[\mu_{11}(1,2)\mu_1(1) - \mu_1(2)\mu_2(1)]}.$$

It may be noted that $E[T_{1(n=2)}''] = T$.

One may ask under what circumstances the estimator $T_{1(n=2)}''$ has prospects of being useful. Suppose it is known (or assumed) that the characteristic \underline{x} is related to \underline{y} by the approximate relationship $x = c y$, and that the y -values are known for all N elements in the population. Then we will have

$$\mu_1(t) = c\mu_1(t)_y$$

and

$$\mu_2(t) = c^2\mu_2(t)_y \quad \text{for } t = 1, 2,$$

and

$$\mu_{11}(t,w) = c^2\mu_{11}(t,w)_y \quad \text{for } t \neq w = 1, 2$$

where the moments and product-moments, on the right-hand side of the above three equations refer to those of \underline{y} computed by the formulas given on page 35. (The suffix \underline{y} is added merely to distinguish them from those for \underline{x} .)

Also we note

$$T = c \sum_{i=1}^N y_i = cT_y.$$

Substituting for T and the moments of \underline{x} with these values in the formula for $T_{1(n=2)}''$, we get exactly the same type of estimator, since the c 's in the numerator and denominator cancel out. We may write this estimator as

$$T_{1(n=2)}''' = \alpha_1(y)x_1 + \alpha_2(y)x_2$$

where $\alpha_1(y)$ and $\alpha_2(y)$ are weights known in terms of the y 's, and their functional form can be identified in $T_{1(n=2)}''$. We define this to be a simulated minimum variance estimator. We have

$$V[T_{1(n=2)}'''] = \alpha_1^2(y)\mu_2(1) + \alpha_2^2(y)\mu_2(2) + 2\alpha_1(y)\alpha_2(y)\mu_{11}(1,2).$$

An unbiased estimate of $V[T_{1(n=2)}''']$ can be found if we can find unbiased estimates for $\mu_2(1)$, $\mu_2(2)$ and $\mu_{11}(1,2)$.

We will find unbiased estimators in the general case, i.e., for $\mu_2(t)$ and $\mu_{11}(t,w)$. Now

$$\begin{aligned} \mu_2(t) &= E(x_{i_t}^2) - [E(x_{i_t})]^2 \\ &= E(x_{i_t}^2) - \left[\sum_{i=1}^N P_{i_t} x_i \right]^2 \\ &= E(x_{i_t}^2) - \left[\sum_{i=1}^N P_{i_t}^2 x_i^2 + \sum_{i \neq j} P_{i_t} P_{j_t} x_i x_j \right]. \end{aligned}$$

We note before proceeding further that x_{i_t} is the characteristic of the i^{th} element appearing at the t^{th} draw and x_{j_w} is the characteristic of the j^{th} element appearing at the w^{th} draw. $\mu_2(t)$ can be rewritten as

$$\mu_2(t) = E(x_{i_t}^2) - \left\{ E(P_{i_t} x_{i_t}^2) + E \left[\frac{P_{i_t} P_{j_t} x_{i_t} x_{j_w}}{P_{i_t, j_w}} \right] \right\}.$$

Thus,

$$\hat{\mu}_2(t) = x_{i_t}^2 - P_{i_t} x_{i_t}^2 - \frac{P_{i_t} P_{j_t} x_{i_t} x_{j_w}}{P_{i_t, j_w}}$$

is an unbiased estimator of $\mu_2(t)$. We note that only the i^{th} and the j^{th} elements enter into the expression for the estimator. Also P_{j_t} , a coefficient of $x_{i_t} x_{j_w}$, is the unconditional probability of the j^{th} element appearing at the t^{th} draw, and is not the same as P_{j_w} . Considered as a quadratic form $\hat{\mu}_2(t)$ is not positive definite and therefore we may sometimes have negative estimates of $\mu_2(t)$. Similarly we find

$$\mu_{11}(t, w) = E(x_{i_t} x_{j_w}) - \left\{ E(P_{i_t} x_{i_t}^2) + E \left[P_{i_t} P_{j_w} \frac{x_{i_t} x_{j_w}}{P_{i_t, j_w}} \right] \right\}$$

Hence one unbiased estimate of $\mu_{11}(t, w)$ would be

$$\hat{\mu}_{11}(t, w) = x_{i_t} x_{j_w} - \left(P_{i_t} x_{i_t}^2 + P_{i_t} P_{j_w} \frac{x_{i_t} x_{j_w}}{P_{i_t, j_w}} \right).$$

We note that

$$E(P_{j_t} x_{j_w}^2) = \sum_{j=1}^N P_{j_w} (P_{j_t} x_{j_w}^2) = \sum_{i=1}^N P_{i_w} (P_{i_t} x_{i_t}^2) = E(P_{i_w} x_{i_t}^2)$$

and therefore

$$c_1 P_{j_t} x_j^2 + c_2 P_{i_w} x_i^2$$

is an unbiased estimate of

$$\sum_{l=1}^N P_{l_t} P_{l_w} x_l^2$$

whenever $c_1 + c_2 = 1$. Perhaps it is simpler to choose $c_1 = c_2 = 1/2$

so that another unbiased estimate of $\mu_{11}(t,w)$ is

$$\mu_{11}(t,w)' = x_i x_j \left(1 - \frac{P_{i_t} P_{j_w}}{P_{i_t, j_w}} \right) - (1/2)(P_{j_t} x_j^2 + P_{i_w} x_i^2).$$

Hence, for the case $n = 2$, where we recall that the i^{th} and the j^{th} elements have appeared at the first and second draws respectively (for the sake of brevity they have been written x_1 and x_2 previously),

we find

$$\begin{aligned} \hat{V}(T_{1(n=2)}^{11}) &= \alpha_1^2(y) \left(x_i^2 (1 - P_{i_1}) - \frac{P_{i_1} P_{j_1}}{P_{i_1, j_2}} x_i x_j \right) \\ &+ \alpha_2^2(y) \left(x_j^2 (1 - P_{j_2}) - \frac{P_{i_2} P_{j_2}}{P_{i_1, j_2}} x_i x_j \right) \\ &+ 2\alpha_1(y)\alpha_2(y) \left[1 - \frac{P_{i_1} P_{j_2}}{P_{i_1, j_2}} x_i x_j \right. \\ &\left. - (1/2)(P_{j_1} x_j^2 + P_{i_2} x_i^2) \right]. \end{aligned}$$

I have discussed the case of $n = 2$ because, by and large, it is the most usual case considered in unequal probability sampling. The

arguments for the case of n elements follow along the same lines. Of course, some technical difficulties will be experienced in constructing the estimator T_1'' , given symbolically as

$$T_1'' = \sum_{t=1}^n \alpha_t(y) x_t,$$

since in the calculation of the $\alpha_t(y)$'s the first step will be to calculate the $n(n+1)/2$ moments and product moments of the y 's plus the n first moments. The next step will be to solve for the $\alpha(y)$'s according to the set of $n+1$ simultaneous equations given on page 36. Equally, before all this is done, the computation of the unconditional probabilities will be laborious.

Before we leave this topic we note that the $\alpha(y)$'s are approximations of the true α_t 's. They will represent the true values only when x is directly proportional to y . Thus, T_1'' will not be free from bias. However, this is the price we pay for attempting to construct an estimator having the appearance of a minimum variance estimator in the hope that the property of minimum variance will be conserved. If the assumed relationship holds for a large number of elements we have reason to believe that this property will be conserved.

4.3 Class Two Estimators

The general estimator in class two is given by

$$T_2 = \sum_{i \in S} \beta_i x_i .$$

We recall that β_i is the weight to be attached to the i^{th} element whenever it appears in the sample. The weights which are independent of the properties of the population are well known (Horvitz and Thompson, 1952), but those which are not independent have not been considered. We know from the case of T_1'' that such an estimator has possibilities of becoming useful whenever each of the characteristics of the population currently under investigation is related to some other known for a long time among writers on probability, and it appears to have been first used by Isserlis (1916).

Let z_i be a characteristic random variable such that

$$\begin{aligned} z_i &= 1, \text{ whenever element } u_i \text{ appears in the sample;} \\ &= 0, \text{ whenever element } u_i \text{ does not appear in the sample.} \end{aligned}$$

Thus we find

$$E(z_i^A) = 1^A \cdot P_i + 0^A(1-P_i) = P_i \quad (A=1,2,3\dots)$$

and

$$\begin{aligned} E(z_i^A z_j^B) &= 1^A \cdot 1^B \left[\text{Probability that both } u_i \text{ and } u_j \text{ appear in the} \right. \\ &\quad \left. \text{sample} \right] + 0^A \cdot 0^B \left[\text{Probability that } u_i \text{ and } u_j \text{ do} \right. \\ &\quad \left. \text{not appear in the sample} \right] \\ &\quad + 0^A \cdot 1^B \left[\text{Probability that } u_i \text{ appears but } u_j \right. \\ &\quad \left. \text{does not appear} \right] \\ &\quad + 1^A \cdot 0^B \left[\text{Probability that } u_j \text{ appears but } u_i \right. \\ &\quad \left. \text{does not appear} \right] \\ &= P_{ij}, \text{ for all } A, B = 1, 2, 3, \dots \text{ and } i \neq j. \end{aligned}$$

We only need the above expectations for cases when $A = 1, 2$ and $B = 1$, but for the determination of higher moments the more general cases, as presented here, are needed.

Hence

$$V(z_i) = P_i - P_i^2 = P_i(1 - P_i),$$

$$\text{Cov}(z_i, z_j) = P_{ij} - P_i P_j \quad \text{for } i \neq j.$$

Now the estimator can be rewritten as

$$T_2 = \sum_{i=1}^N z_i \beta_i x_i = \sum_{i \in S} \beta_i x_i$$

since for the remaining $N-n$ elements which do not appear in the sample, the z 's are all zero. We have

$$E(T_2) = \sum_{i=1}^N E(z_i) \beta_i x_i = \sum_{i=1}^N P_i \beta_i x_i.$$

For the estimator to be unbiased as well as for β 's to be independent of the properties of the population we insist that

$$\sum_{i=1}^N P_i \beta_i x_i = \sum_{i=1}^N x_i$$

so that we must have

$$P_i \beta_i = 1 \quad \text{for all } i,$$

and thus

$$\beta_i = 1/P_i.$$

We note that there is one and only one set of such weights. Therefore, as noted by Horvitz and Thompson,

$$T_2' = \sum_{i \in S} (x_i / P_i)$$

is the only type of unbiased linear estimator possible in this class, and in this very special sense it is the best since it is the only estimator which we can choose.

Next we have

$$\begin{aligned} V(T_2) &= E \left[\sum_{i=1}^N z_i \beta_i x_i - E \left(\sum_{i=1}^N z_i \beta_i x_i \right) \right]^2 \\ &= E \left[\sum_{i=1}^N \beta_i x_i (z_i - P_i) \right]^2 \\ &= \sum_{i=1}^N \beta_i^2 x_i^2 V(z_i) + 2 \sum_{i < j} \beta_i \beta_j x_i x_j \text{Cov}(z_i, z_j) \\ &= \sum_{i=1}^N \beta_i^2 x_i^2 P_i (1 - P_i) + 2 \sum_{i < j} \beta_i \beta_j x_i x_j (P_{ij} - P_i P_j) \end{aligned}$$

We note that when we put $\beta_i = 1/P_i$, Horvitz and Thompson's formula is immediately obtained, i.e.,

$$V(T_2') = \sum_{i=1}^N x_i^2 \cdot \frac{1 - P_i}{P_i} + 2 \sum_{i < j} x_i x_j \cdot \frac{P_{ij} - P_i P_j}{P_i P_j}$$

Obviously one cannot obtain an estimator such that $V(T_2)$ is a minimum subject to the formal condition $E(T_2) = T$, since, as pointed out above, this condition leads to one and only one estimator to choose from.

We proceed to find the minimum variance estimator which, as will be seen subsequently, will not be independent of the properties of the population. For this we minimize $V(T_2)$ subject to the restriction $E(T_2) = T$. Setting

$$H = V(T_2) + 2\lambda \left[\sum_{i=1}^N P_i \beta_i x_i - T \right]$$

we solve the equations

$$\delta H / \delta \beta_i = 0 = 2\beta_i x_i^2 P_i (1 - P_i) + 2 \sum_{j(\neq i)=1}^N \beta_j x_i x_j (P_{ij} - P_i P_j) + 2\lambda x_i P_i,$$

for $i=1, 2, \dots, N$,

together with

$$\sum_{i=1}^N P_i \beta_i x_i = T.$$

The above N equations reduce to

$$x_i \beta_i + (1/P_i) \sum_{j(\neq i)=1}^N x_i P_{ij} \beta_j + \lambda = T.$$

These, with the restricting equation, constitute a system of $N+1$ independent equations with $N+1$ unknowns. Writing them in full we have in matrix notation,

$$\begin{bmatrix} 1 & P_{12}/P_1 & \dots & P_{1N}/P_1 & 1 \\ P_{21}/P_2 & 1 & \dots & P_{2N}/P_2 & 1 \\ \vdots & & & & \\ P_{N1}/P_N & P_{N2}/P_N & \dots & 1 & 1 \\ P_1 & P_2 & \dots & P_N & 0 \end{bmatrix} \begin{bmatrix} \beta_1 x_1 \\ \beta_2 x_2 \\ \vdots \\ \beta_N x_N \\ \lambda \end{bmatrix} = \begin{bmatrix} T \\ T \\ \vdots \\ T \\ T \end{bmatrix}$$

Hence the solution of these equations will be of the form

$$\beta_i = T/x_i \cdot f_i$$

where f_i is purely a function of probabilities and to be specific it is the ratio of determinant of the matrix on the left-hand side of the equations with 1's substituted in the i^{th} column divided by the same determinant. To illustrate when $N = 3$ and we have samples of $n = 2$, then

$$\beta_1 = \frac{T}{x_1} \left[\frac{P_{23}}{P_1} \left(\frac{P_{13}}{P_3} + \frac{P_{12}}{P_2} \right) - \frac{P_{23}^2}{P_2 P_3} \right] / D$$

$$\beta_2 = \frac{T}{x_2} \left[\frac{P_{13}}{P_2} \left(\frac{P_{23}}{P_3} + \frac{P_{12}}{P_1} \right) - \frac{P_{13}^2}{P_1 P_3} \right] / D$$

$$\beta_3 = \frac{T}{x_3} \left[\frac{P_{12}}{P_3} \left(\frac{P_{13}}{P_1} + \frac{P_{23}}{P_2} \right) - \frac{P_{12}^2}{P_1 P_2} \right] / D$$

where

$$D = \frac{P_{12}}{P_1 P_2} (2P_{13} P_2 - P_3 P_{12}) + \frac{P_{23}}{P_2 P_3} (2P_{12} P_3 - P_1 P_{23}) + \frac{P_{13}}{P_1 P_3} (2P_{23} P_1 - P_2 P_{13}).$$

Again there are possibilities of using the estimator when we know that \underline{x} is related to some characteristic \underline{y} , say, approximately by the relationship

$$\underline{x} = c \underline{y} .$$

In this situation the β 's become known, and we find

$$\beta_i = (T/y_i) f_i .$$

The labor of computing from these probability functions will of course be considerable. It may be noted that the estimator assumes the form of a weighted ratio estimate

$$T_2'' = T_y \sum_{i=1}^n (x_i f_i) / y_i .$$

When the probabilities of selection are equal, then $f_i = 1/n$ and this estimator reduces to $T_y \sum \frac{x}{y} / n$, a well known ratio estimate. The variance of T_2'' is given by

$$V(T_2'') = T_y^2 \left[\sum_{i=1}^N \frac{f_i^2}{y_i} x_i^2 P_i (1-P_i) + 2 \sum_{i < j} \frac{f_i f_j}{y_i y_j} x_i x_j (P_{ij} - P_i P_j) \right] .$$

We have merely substituted the expression for the β 's in $V(T_2'')$. The unbiased estimate of $V(T_2'')$ is given by

$$V(T_2'') = T_y^2 \left[\sum_{i=1}^n \frac{f_i^2}{y_i} x_i^2 (1-P_i) + 2 \sum_{i < j} \frac{f_i f_j}{y_i y_j} x_i x_j \frac{P_{ij} - P_i P_j}{P_{ij}} \right] .$$

These results appear to be mainly of theoretical interest and, of course, conditional upon the appropriateness of the functional relationship assumed to hold. In practice the formulas may be unwieldy for use except for sample sizes of $n = 2$ from populations of $N = 3$ or 4 .

Lastly we note that this estimator, which has the appearance of a minimum variance estimator, without strictly having its properties, will certainly not be free from bias since the β 's used are not the true β 's. But it is likely that the property of minimum variance will be conserved to some extent as in the case of T_1''' . Again we pay a price for simulation.

4.4 Class Three Estimators

The estimators in class three will now be considered. We recall that for

$$T_3 = \gamma_s \sum_{i \in s} x_i$$

there are $S' = {}^N C_n$ weights, one for each distinct sample. For the T_3 to be unbiased and for the γ 's to be independent of the properties of the population we must have

$$E(T_3) = \sum_{s=1}^{S'} P_s \gamma_s \sum_{i \in s} x_i \equiv \sum_{i=1}^N x_i,$$

i.e.,

$$\sum_{i=1}^N x_i \sum_{s \ni i} P_s \gamma_s \equiv \sum_{i=1}^N x_i,$$

where $\sum_{s \ni i}$ indicates summation over all those products $P_s \gamma_s$ which are the coefficients of aggregates of sample values which include the element u_i . There are ${}^{N-1} C_{n-1}$ such product terms. Thus we must have

$$\sum_{s \ni i} P_s \gamma_s = 1, \quad \text{for } i=1,2,\dots,N.$$

There will be many sets of solutions to these N equations.

Now if for each i , $\gamma_s = \gamma_i$ for all $s \ni i$, then the condition of unbiasedness becomes

$$\gamma_i \sum_{s \ni i} P_s = 1,$$

i.e.,

$$\gamma_i P_i = 1 \quad \text{or} \quad \gamma_i = 1/P_i,$$

and this leads to the unbiased estimator in class two proposed by Horvitz and Thompson (1951, 1952). We note that one set of solutions is given by

$$\gamma_s = 1/(P_s^{N-1} C_{n-1}) \quad \text{for all } s, i \text{ and all } i.$$

Hence one subclass of unbiased estimators is given by

$$T'_3 = [1/(P_s^{N-1} C_{n-1})] \sum_{ies} x_i,$$

which includes Midzuno's (1950), Lahiri's (1951) and Sen's (1952) estimators for the case of one-stage sampling. Des Raj (1954) stated that this is the only unbiased estimator of the aggregate of the population, of the particular form which has been defined in this thesis as class three. Clearly this is one among many.

Next we have

$$V(T_3) = \sum_{s=1}^{S'} P_s (\gamma_s \sum_{ies} x_i)^2 - \left[\sum_{s=1}^{S'} P_s \gamma_s \sum_{ies} x_i \right]^2.$$

We shall determine the γ 's which minimize $V(T_3)$ subject to the N restrictions on the unbiasedness of the estimate given above. Then the best linear unbiased estimate will be obtained. For this we set

$$\begin{aligned} H &= V(T_3) + 2 \sum_{i=1}^N \lambda_i \left(\sum_{s,i} P_s \gamma_s - 1 \right) \\ &= V(T_3) + 2 \left[\sum_{s=1}^{S'} P_s \gamma_s \sum_{ies} x_i - \sum_{i=1}^N \lambda_i \right], \end{aligned}$$

where the λ 's are the Lagrange multipliers, and solve the S' equations

$$\frac{\partial H}{\partial \gamma_s} = 0 = 2P_s \gamma_s \left(\sum_{ies} x_i \right)^2 - 2 \left[\sum_{s=1}^{S'} P_s \gamma_s \sum_{ies} x_i \right] P_s \sum_{ies} x_i + 2P_s \sum_{ies} \lambda_i,$$

for all s ,

together with the N equations representing the restrictions.

Substituting unity for $\sum_{s=1} P_s \gamma_s$, which is the coefficient of each x_i , when the expression in square braces is rewritten, the above equation simplifies to

$$P_s \gamma_s \left(\sum_{ies} x_i \right)^2 + P_s \sum_{ies} \lambda_i = T P_s \sum_{ies} x_i \quad \text{for all } s,$$

or

$$\gamma_s \left(\sum_{ies} x_i \right)^2 + \sum_{ies} \lambda_i = T \sum_{ies} x_i, \quad s=1,2,\dots,S'.$$

The restricting equations which we have already used to simplify the above equations are

$$\sum_{s>1} P_s \gamma_s = 1, \quad \text{for } i=1,2,\dots,N.$$

There are $\binom{N}{S'} + N$ independent equations and the same number of unknowns (including the N λ 's) and therefore a unique solution for these unknowns exists. This solution will involve all the x 's. Technically speaking the algebraic solution of these linear simultaneous equations will be cumbersome. Of course, the solution can be written in matrix notation, but in this connection the symbolism will be barren. Therefore let us simply say ${}_o\gamma_s$ for $s=1,2,\dots,S'$ and ${}_o\lambda_i$, $i=1,2,\dots,N$ represents the values (weights) obtained as a result of solving the

above (${}^N C_n + N$) equations. Then the minimum variance estimator T_3'' will be symbolically represented by

$$T_3'' = \sum_{s=1}^{S'} \gamma_s \sum_{i \in s} x_i$$

and its variance by

$$V(T_3'') = \sum_{s=1}^{S'} P_s \gamma_s^2 \left(\sum_{i \in s} x_i \right)^2 - T^2.$$

Symbolically the estimate of this variance will be

$$\hat{V}(T_3'') = \sum_{s=1}^{S'} \gamma_s^2 \left(\sum_{i \in s} x_i \right)^2 - (1/P_s) \left[\frac{\sum_{i \in s} x_i^2}{{}^{N-1} C_{n-1}} + 2 \frac{\sum_{i < j \in s} x_i x_j}{{}^{N-2} C_{n-2}} \right]$$

where \underline{s} is the particular sample in question and γ_s is its corresponding weight. The expectation of the expression in braces multiplied by $1/P_s$ is equal to T^2 . This is easy to show if we note that when summing $\sum_{i \in s} x_i^2$ over all \underline{s} , x_i^2 occurs ${}^{N-1} C_{n-1}$ times, and also when summing $\sum_{i < j \in s} x_i x_j$ over all \underline{s} , x_i and x_j , when $i < j$, occur together ${}^{N-2} C_{n-2}$ times. Clearly $\hat{V}(T_3'')$ can assume negative values.

Also $V(T_3'')$ can be given in terms of the λ_i 's. If we add the set of S' simultaneous equations where P_s is present we will find

$$\sum_{s=1}^{S'} P_s \gamma_s \left(\sum_{i \in s} x_i \right)^2 + \sum_{s=1}^{S'} P_s \sum_{i \in s} \lambda_i = T \sum_{i=1}^{S'} P_s \sum_{i \in s} x_i$$

i.e.,

$$V(T_3'') + T^2 + \sum_{i=1}^N \lambda_i \sum_{s:i} P_s = T \sum_{i=1}^N x_i \sum_{s:i} P_s,$$

and therefore

$$V(T_3'') = T \sum_{i=1}^N x_i P_i - \sum_{i=1}^N \lambda_i P_i - T^2,$$

where λ_i , for all i , is the solution given above.

If there is some approximate relationship between x_i and a characteristic y_i , known for all i , such as $x = cy$, then we can obtain a simulated minimum variance estimator by substituting for the x 's in the above system of simultaneous equations and solving for the γ 's and λ 's. Thus for any sample s we will have a simulated value γ_s' , of the true γ_s . The closeness of each of these values to the true value will of course depend on the degree of closeness of the relationship $x = cy$. This estimator will then be

$$T_3''' = \gamma_s' \sum_{i \in s} x_i,$$

and the estimate of its sampling variance will be obtained simply by substituting γ_s' for γ_s in the expression for $\hat{V}(T_3'')$. For samples of size $n = 2$, and $N = 4, 5, 6, 7$, and 8 , the number of unknowns involved in the solution of the linear simultaneous equations will be 10, 15, 21, 28 and 36 respectively. The calculations involved will also be very simple. The only operations needed are sums of combinations of numbers taken $n(=2)$ at a time and their corresponding squares

and the total for y . In solving the equations^{1/} the constant c^2 can be absorbed into the λ 's.

Among the three simulated minimum variance estimators presented so far, T_1'' , T_2'' , and T_3'' , the last one will be the least cumbersome to handle in regard to practical applications.

4.5 Class Four Estimators

The general estimator in class four is given by

$$T_4 = \sum_{t=1}^n \hat{c}_{i_t} x_t$$

where we recall \hat{c}_{i_t} ($i=1,2,\dots,N$; $t=1,2,\dots,n$) is the weight to be attached to the i^{th} element whenever it appears at the t^{th} draw.

There are two approaches in deriving the expected value of T_4 . In one approach we use the unconditional probabilities, i.e., the probabilities of the elements appearing at specified draws; in the other we use the probability distributions at each draw and in this sense the derivation rests much more on first principles, although much more cumbersome. The former approach is also geared to the use of the characteristic random variable, but the latter is not.

We shall first find $E(T_4)$ using unconditional probabilities.

Let z_{i_t} be a characteristic random variable such that

$$\begin{aligned} z_{i_t} &= 1, \text{ whenever the } i^{\text{th}} \text{ element appears at the } t^{\text{th}} \text{ draw.} \\ &= 0, \text{ whenever the } i^{\text{th}} \text{ element does not appear at the } \\ &\quad t^{\text{th}} \text{ draw.} \end{aligned}$$

^{1/} It is easy to solve such equations by electronic machines.

Then we may write T_4 as

$$T_4 = \sum_{t=1}^n \sum_{i=1}^N \delta_{i_t} x_i z_{i_t}.$$

We note again that the x 's are mere numbers. To specify a characteristic of element u_i fully we should write x_{i_t} . But here the subscript t has been suppressed in view of the presence of \underline{z} . Now

$$E(T_4) = \sum_{t=1}^n \sum_{i=1}^N \delta_{i_t} x_i E(z_{i_t}).$$

We have

$$\begin{aligned} E(z_{i_t}^r) &= 1^r [\text{Probability that } u_i \text{ appears at } t^{\text{th}} \text{ draw}] \\ &\quad + 0^r [\text{Probability that } u_i \text{ does not appear at } \\ &\quad \quad \quad t^{\text{th}} \text{ draw}] \\ &= 1^r \cdot P_{i_t} + 0^r (1 - P_{i_t}) = P_{i_t}, \text{ for } r=1, 2, \dots, \end{aligned}$$

so that

$$E(T_4) = \sum_{t=1}^n \sum_{i=1}^N \delta_{i_t} x_i P_{i_t} = \sum_{i=1}^N x_i \sum_{t=1}^n P_{i_t} \delta_{i_t}.$$

For T_4 to be unbiased we must have

$$E(T_4) = \sum_{i=1}^N x_i \sum_{t=1}^n P_{i_t} \delta_{i_t} = \sum_{i=1}^N x_i,$$

and for the δ 's to be independent of the unknown x 's we further insist on the condition of identity; i.e.,

$$\sum_{i=1}^N x_i \sum_{t=1}^n P_{i_t} \delta_{i_t} = \sum_{i=1}^N x_i.$$

Hence,

$$\sum_{t=1}^n P_{i_t} \delta_{i_t} = 1 \text{ for } i=1, 2, \dots, N.$$

We can find many sets of solutions to these equations. One set of solutions is given by

$$f_{i_t} = c_t / P_{i_t} \quad \text{for } i=1,2,\dots,N$$

provided

$$\sum_{t=1}^n c_t = 1.$$

Thus we have

$$T''_4 = \sum_{i=1}^n c_t (x_t / P_{i_t}).$$

This is the estimator proposed by Des Raj (1956). It appears to be simpler to take $c_1 = \dots = c_n = 1/n$ so that one estimator in this subclass will be

$$(1/n) \sum_{t=1}^n (x_t / P_{i_t}).$$

Let us attempt to find the minimum variance estimator. We have

$$V(T_4) = \sum_{i=1}^N \sum_{t=1}^n \frac{x_i^2}{i} V(z_{i_t}) + 2 \sum_{\substack{i=j, t \leq w \\ i < j, t=w \\ i < j, t < w}} i_t j_w x_i x_j \text{Cov}(z_{i_t}, z_{j_w})$$

Now

$$\begin{aligned} V(z_{i_t}) &= E(z_{i_t}^2) - [E(z_{i_t})]^2 \\ &= P_{i_t} - P_{i_t}^2 = P_{i_t} (1 - P_{i_t}). \end{aligned}$$

When $i = j$, $t < w$,

$$\text{Cov}(z_{i_t}, z_{j_w}) = E(z_{i_t} z_{j_w}) - E(z_{i_t})E(z_{j_w})$$

Now $E(z_{i_t} z_{j_w}) = 0$, since i^{th} unit cannot appear at the w^{th} draw when it has already appeared at the t^{th} . Therefore

$$\text{Cov}(z_{i_t}, z_{j_w}) = 0 - P_{i_t} \cdot P_{i_w} = -P_{i_t} P_{i_w}.$$

Similarly when $i < j$, $t = w$, we find

$$\begin{aligned} \text{Cov}(z_{i_t}, z_{j_t}) &= E(z_{i_t} z_{j_t}) - E(z_{i_t})E(z_{j_t}) \\ &= 0 - P_{i_t} P_{j_t} = -P_{i_t} P_{j_t}. \end{aligned}$$

Finally when $i < j$, $t < w$,

$$\begin{aligned} \text{Cov}(z_{i_t}, z_{j_w}) &= E(z_{i_t} z_{j_w}) - E(z_{i_t})E(z_{j_w}) \\ &= P_{i_t, j_w} - P_{i_t} P_{j_w}. \end{aligned}$$

Hence, with these results $V(T_4)$ is known. We shall now determine the δ 's subject to the restrictions on unbiasedness of T_4 . For this we set up a function

$$H = V(T_4) + 2 \sum_{i=1}^N \lambda_i \left(\sum_{t=1}^n P_{i_t} \delta_{i_t} - 1 \right)$$

where the λ 's are the undetermined Lagrange multipliers and solve the following equations:

$$\begin{aligned} \partial H / \partial \delta_{i_t} = 0 = & 2 \delta_{i_t} x_i^2 V(z_{i_t}) + 2x_i \left[\sum_{w(\neq t)=1}^n \delta_{i_w} x_i \text{Cov}(z_{i_t}, z_{i_w}) \right. \\ & + \sum_{j(\neq i)=1}^N \delta_{j_t} x_j \text{Cov}(z_{i_t}, z_{j_t}) \\ & \left. + \sum_{j(\neq i)=1}^N \sum_{w(\neq t)=1}^n \delta_{j_w} x_j \text{Cov}(z_{i_t}, z_{j_w}) \right] + 2\lambda_i P_{i_t}, \end{aligned}$$

for $\begin{cases} i=1, 2, \dots, N \\ t=1, 2, \dots, n, \end{cases}$

together with the N restricting equations. Substituting for the variance and covariance expressions and rearranging terms we will find

$$x_i^2 P_{i_t} \delta_{i_t} - x_i P_{i_t} \sum_{i=1}^N x_i \sum_{t=1}^n P_{i_t} \delta_{i_t} + x_i \sum_{j(\neq i)=1}^N \sum_{w(\neq t)=1}^n x_j P_{i_t} \delta_{j_w} + \lambda_i P_{i_t} = 0,$$

for all i and t .

The restricting equations are

$$\sum_{t=1}^n P_{i_t} \delta_{i_t} = 1, \quad \text{for } i=1, 2, \dots, N.$$

There are $Nn+N$ independent equations and the same number of unknowns (including the N λ 's). Clearly the solution for the δ 's will involve the the unknown x 's. Thus at least formally a minimum variance estimator T'_L exists. If (as before) we attempt to approximate for these unknown values of the x 's by a function of known values of a related characteristic y , then it will be possible to obtain simulated values of the δ 's and thus to have an estimator (as before) having the appearance of a minimum variance estimator. However, in view of the

complexity of the solutions for the ξ 's we will not write the actual expressions for this estimator and its variance.

Next we shall determine the expected value of T_4 , using the probabilities relevant to each draw. We recall that the probability system defines all these probabilities which are somewhat analogous to conditional probabilities. We argue from first principles. Suppose element u_i appears at the first draw and u_j at the second. Let x_{j_2} represent the character observed on u_j . Then

$$E(x_{j_2}) = E [E(x_{j_2} | u_i \text{ appeared at the first draw.})].$$

We can find $E(x_{j_2} | u_i)$ since we have defined $[p_{j_2}^i]$. Next we can find the expectation of the resulting expression, since we have defined $[p_{i_1}]$. Thus

$$E(x_{j_2}) = E \sum_{j(\neq i)=1}^N p_{j_2}^i x_j = \sum_{i=1}^N p_{i_1} \left(\sum_{j(\neq i)=1}^N p_{j_2}^i x_j \right).$$

Now for three or more draws the above notation for expectations applicable to each draw becomes cumbersome. Thus if x_{k_3} is the character observed on u_k at the third draw then

$$E(x_{k_3}) = E \left\{ E \left[E(x_{k_3} | u_i \text{ appeared at the 1}^{\text{st}} \text{ draw, and } u_j \text{ at the 2}^{\text{nd}} \text{ draw}) \middle| u_i \text{ appeared at the 1}^{\text{st}} \text{ draw} \right] \right\}$$

For this reason we denote the above operations of taking expectations simply by

$$E(x_{j_2}) = E_1 E_2(x_{j_2}) \quad \text{and} \quad E(x_{k_3}) = E_1 E_2 E_3(x_{k_3})$$

where the symbols E_1 , E_2 and E_3 refer to the operation of taking expectations applicable to the first, second and third draws respectively.

Generally for the expectation of the variable appearing at the n^{th} draw we can write

$$E(x_{m_n}) = E_1 E_2 \dots E_n(x_{m_n})$$

meaning that the operation is carried out in n steps using probability distributions relevant to each of the n draws. Symbolically

$E = E_1 E_2 \dots E_n$. (These remarks apply equally to functions of the variables involved.)

Thus we have

$$\begin{aligned} E(T_4) &= E\left(\sum_{t=1}^n \delta_{i_t} x_{i_t}\right) \\ &= E_1(\delta_{i_1} x_{i_1}) + E_1 E_2(\delta_{j_2} x_{j_2}) + \dots + E_1 E_2 \dots \\ &\quad E_n(\delta_{m_n} x_{m_n}). \end{aligned}$$

Now

$$\begin{aligned} E_1(\delta_{i_1} x_{i_1}) &= \sum_{i=1}^N p_{i_1} x_i \delta_{i_1}^i, \\ E_1 E_2(x_{j_2} \delta_{j_2}^i) &= E_1 \sum_{j(\neq i)=1}^N p_{j_2}^i x_j \delta_{j_2}^i = \sum_{i=1}^N p_{i_1} \sum_{j(\neq i)=1}^N p_{j_2}^i x_j \delta_{j_2}^i, \end{aligned}$$

and so on up to

$$E_1 E_2 \dots E_n(x_{m_n} \delta_{m_n}^i) = \sum_{i=1}^N p_{i_1} \sum_{j(\neq i)=1}^N p_{j_2}^i \dots \sum_{m(\neq 1 \neq \dots \neq j \neq i)=1}^N p_{m_n}^{ij \dots l} x_m \delta_{m_n}^i.$$

When we arrange terms we will find

$$E_1 E_2 \dots E_t(\delta_{k_t} x_{k_t}) = \sum_{i=1}^N p_{i_1} \dots \sum_{k(\neq \dots \neq j \neq i)=1}^N p_{k_t}^{ij \dots f} x_k \delta_{k_t}^i = \sum_{i=1}^N x_i \delta_{i_t}^i p_{i_t}.$$

Hence, as before,

$$E(T_4) = \sum_{t=1}^n \sum_{i=1}^N x_i \delta_{i_t} P_{i_t} = \sum_{i=1}^N x_i \sum_{t=1}^n \delta_{i_t} P_{i_t},$$

and the same condition of unbiasedness, i.e.,

$$\sum_{t=1}^n \delta_{i_t} P_{i_t} = 1 \text{ for all } \underline{i},$$

follows.

Further if for any \underline{i} , $\delta_{i_t} = \delta_i$ for $t=1, 2, \dots, n$, then the condition of unbiasedness becomes

$$\delta_i \sum_{t=1}^n P_{i_t} = 1, \text{ for all } \underline{i}$$

i.e.,

$$\delta_i P_i = 1$$

or

$$\delta_i = 1/P_i, \text{ for all } \underline{i}$$

and this leads to the unbiased estimator in class two.

We shall now obtain another unbiased estimator in class four.

Consider the expressions for the following expectations:

$$E_1(\delta_{i_1} x_{i_1}), E_1 E_2(\delta_{j_2} x_{j_2}), \dots, E_1 E_2 \dots E_n(\delta_{m_n} x_{m_n}),$$

given on page 61, the sum of which is $E(t_4)$. Let

$$\delta_{i_1} = c_1/p_{i_1}, \text{ for } i=1, 2, \dots, N;$$

$$\delta_{j_2} = c_2/(N-1)p_{i_1} p_{j_2}^i, \text{ for } j=1, 2, \dots, N, \text{ but } j \neq i;$$

$$\delta_{k_3} = c_3/(N-1)(N-2)p_{i_1} p_{j_2}^i p_{k_3}^{ij}, \text{ for } k=1, 2, \dots, N, \text{ but } k \neq j \neq i;$$

$$\delta_{m_n} = c_n/(N-1)\dots(N-n+1)p_{i_1} p_{j_2}^i \dots p_{m_n}^{ij\dots l}, \text{ for } m=1, 2, \dots, N, \text{ but } m \neq \dots \neq j \neq i,$$

where c_1, c_2, \dots, c_n are constants which will be chosen to satisfy the condition of unbiasedness. Denote the resulting estimator by $T_4^{(c)}$. Substituting these δ 's in the relevant expressions for the above expectations and adding them we obtain, after some algebra

$$\begin{aligned} E(T_4^{(c)}) &= c_1 \sum_{i=1}^N x_i + [c_2/(N-1)] \sum_{i=1}^N \sum_{j(\neq i)=1}^N x_j + \dots \\ &\quad + [c_n/(N-1)(N-2)\dots(N-n+1)] \sum_{i=1}^N \dots \sum_{m(\neq 1\dots\neq i)=1}^N x_m \\ &= [c_1 + c_2 + \dots + c_n] \sum_{i=1}^N x_i. \end{aligned}$$

Hence,

$$\begin{aligned} T_4^{(c)} &= [c_1/p_{i_1}] x_{i_1} + [c_2/(N-1)p_{i_1} p_{j_2}^i] x_{j_2} + \dots \\ &\quad [c_n/(N-1)(N-2) \dots (N-n+1)p_{i_1} p_{j_2}^i \dots p_{m_n}^{ij\dots l}] x_{m_n} \end{aligned}$$

is an unbiased estimator of

$$T = \sum_{i=1}^N x_i$$

provided

$$\sum_{t=1}^n c_t = 1.$$

The estimator $T_4^{(c)}$ was first obtained by Das (1951), but the probabilities he used at the second and subsequent draws were based on the initial probabilities p_{i_1} ($i=1, 2, \dots, N$). He gave the variance of this estimator and the unbiased estimator of this variance when $c_t = 1/n$ for

all t , and he noted also that the values of the c 's which minimized $V(T_4''')$ were functions of all the parameters of population. Later Des Raj (1956) reported the same estimator but using arbitrary probabilities at each draw.

In the same paper Des Raj also reported another unbiased estimator which now will be shown to belong to class four. He first wrote

$$t_1 = x_{i_1} / p_{i_1},$$

$$t_2 = x_{i_1} + x_{j_2} / p_{j_2}^i,$$

$$t_3 = x_{i_1} + x_{j_2} + x_{k_3} + \dots + x_{m_n} / p_{m_n}^{ij\dots l}$$

and stated that the unbiased estimator was

$$T_4'''' = c_1 t_1 + c_2 t_2 + \dots + c_n t_n$$

where

$$\sum_{t=1}^n c_t = 1.$$

Now this estimator can be rewritten as

$$T_4'''' = [(c_1/p_{i_1}) + c_2 + \dots + c_n]x_{i_1} + [(c_2/p_{j_2}^i) + c_3 + \dots + c_n]x_{j_2} \\ + [(c_3/p_{k_3}^{ij}) + c_4 + \dots + c_n]x_{k_3} + \dots + [c_n/p_{m_n}^{ij\dots l}]x_{m_n}.$$

To prove that it belongs to class four we need only show that when

$$\delta_{i_1} = (c_1/p_{i_1}) + c_2 + \dots + c_n, \text{ for } i=1,2,\dots,N$$

$$\delta_{j_2} = (c_2/p_{j_2}^i) + c_3 + \dots + c_n, \text{ for } j=1,2,\dots,N, \text{ but } j \neq i$$

$$\vdots$$

$$\delta_{m_n} = (c_n/p_{m_n}^{ij\dots l}), \text{ for } m=1,2,\dots,N, \text{ but } m \neq l \dots \neq i,$$

$$E(T_4''') = \sum_{i=1}^N x_i = T,$$

provided

$$\sum_{i=1}^n c_t = 1.$$

We have, after substituting for the δ 's,

$$\begin{aligned} E(T_4''') &= E_1 [(c_1/p_{i_1}) + c_2 + \dots + c_n] x_{i_1} \\ &+ E_1 E_2 [(c_2/p_{j_2}^i) + c_3 + \dots + c_n] x_{j_2} \\ &+ \dots + E_1 E_2 \dots E_n (c_n/p_{m_n}^{ij\dots l}) x_{m_n}. \end{aligned}$$

Now

$$\begin{aligned} E_1 [(c_1/p_{i_1}) + c_2 + \dots + c_n] x_{i_1} &= E_1 [(c_1/p_{i_1}) x_{i_1}] + [c_2 + \dots + c_n] E_1(x_{i_1}) \\ &= c_1 \sum_{i=1}^N x_i + [c_2 + \dots + c_n] E_1(x_{i_1}), \end{aligned}$$

$$\begin{aligned}
& E_1 E_2 [(c_2/p_{j_2}^i) + c_3 + \dots + c_n] x_{j_2} \\
&= E_1 E_2 ((c_2/p_{j_2}^i) x_{j_2}) + (c_3 + \dots + c_n) E_1 E_2 (x_{j_2}) \\
&= E_1 \sum_{j(\neq i)=1}^N (c_2/p_{j_2}^i) p_{j_2}^i x_j + (c_3 + \dots + c_n) E_1 E_2 (x_{j_2}) \\
&= c_2 E_1 \sum_{j(\neq i)=1}^N x_j + (c_3 + \dots + c_n) E_1 E_2 (x_{j_2}) \\
&= c_2 E_1 \left(\sum_{i=1}^N x_i - x_i \right) + (c_3 + \dots + c_n) E_1 E_2 (x_{j_2}) \\
&= c_2 \sum_{i=1}^N x_i - c_2 E_1 (x_{i_1}) + (c_3 + \dots + c_n) E_1 E_2 (x_{j_2}),
\end{aligned}$$

$$\begin{aligned}
& E_1 E_2 E_3 [(c_3/p_{k_3}^{ij}) x_{k_3}] + (c_4 + \dots + c_n) E_1 E_2 E_3 (x_{k_3}) \\
&= c_3 \sum_{i=1}^N x_i - c_3 E_1 (x_{i_1}) - c_3 E_1 E_2 (x_{j_2}) + (c_4 + \dots + c_n) E_1 E_2 E_3 (x_{k_3}),
\end{aligned}$$

and finally

$$\begin{aligned}
& E_1 E_2 \dots E_n [(c_n/p_{m_n}^{ij\dots l}) x_{m_n}] \\
&= c_n \sum_{i=1}^N x_i - c_n E_1 (x_{i_1}) - c_n E_1 E_2 (x_{j_2}) - c_n E_1 E_2 E_3 (x_{k_3}) - \dots \\
&\quad - c_n E_1 E_2 \dots E_{n-1} (x_{l_{n-1}}).
\end{aligned}$$

Collecting all these expectations together we find

$$E(T_4''''') = c_1 \sum_{i=1}^N x_i + c_2 \sum_{i=1}^N x_i + \dots + c_n \sum_{i=1}^N x_i = \left(\sum_{t=1}^n c_t \right) T.$$

Hence, provided

$$\sum_{t=1}^n c_t = 1, \quad E(T_4^{(n)}) = T.$$

One estimator in this subclass, for which $c_1 = c_2 = \dots = c_n = 1/n$, has an estimated variance which is always positive (Des Raj, 1956).

4.6 Class Five Estimators

Next we consider the general estimator in class five given by

$$T_5 = \sum_{i \in s} \theta_{s_i} x_i$$

where θ_{s_i} is the weight to be attached to the i^{th} element whenever it appears in the s^{th} sample. We recall that there are altogether $S' = N \binom{N-1}{n-1} = n \binom{N}{n}$ weights. We shall determine these weights such that

$$E(T_5) = T,$$

and further we shall attempt to find the best linear unbiased estimator, i.e., an estimator with weights such that

$$V(T_5) = \min.,$$

subject to the restrictions on unbiasedness.

We have

$$E(T_5) = \sum_{s=1}^{S'} P_s \sum_{i \in s} \theta_{s_i} x_i = \sum_{i=1}^N x_i \sum_{s=1}^{S'} P_s \theta_{s_i}.$$

Hence, for the estimator to be unbiased we must have

$$\sum_{s>i} P_s \theta_{s_i} = 1, \text{ for all } \underline{i}.$$

One set of solutions is given by $\theta_{s_i} = c_{s_i} / P_s$, provided

$$\sum_{s>i} c_{s_i} = 1$$

and another set by choosing $c_{s_i} = 1 / \binom{N-1}{n-1} C_{n-1}$ for all \underline{i} , so that

$$\theta_{s_i} = 1 / \binom{N-1}{n-1} C_{n-1} P_s, \text{ for all } \underline{i}$$

which leads to an estimator belonging also to class three. Further if for any \underline{i} , $\theta_{s_i} = \theta_i$ for all $s>i$ we will have

$$\theta_i \sum_{s>i} P_s = 1$$

i.e.,

$$\theta_i P_i = 1$$

or

$$\theta_i = 1 / P_i$$

and this leads to the estimator given by Horvitz and Thompson.

We have

$$V(T_5) = \sum_{s=1}^{S'} P_s \left(\sum_{i \leq s} \theta_{s_i} x_i \right)^2 - \left(\sum_{s=1}^{S'} P_s \sum_{i \leq s} \theta_{s_i} x_i \right)^2.$$

Next, to find the minimum variance estimator, we construct the function

$$\begin{aligned} H &= V(T_5) + 2 \sum_{i=1}^N \lambda_i \left[\sum_{s:i} P_s \theta_{s_i} - 1 \right] \\ &= V(T_5) + 2 \sum_{s=1}^{S'} P_s \left(\sum_{i \in s} \lambda_i \theta_{s_i} \right) - 2 \sum_{i=1}^N \lambda_i \end{aligned}$$

where the λ 's are undetermined Lagrange multipliers, and attempt to solve the $n^N C_n$ equations

$$\partial H / \partial \theta_{s_i} = 0 = P_s^2 \left(\sum_{i \in s} x_i \theta_{s_i} \right) - 2 \left[\sum_{s=1}^{S'} P_s \sum_{i \in s} x_i \theta_{s_i} \right] P_s x_i + 2 P_s \lambda_i$$

for $s=1, 2, \dots, S'$ and for all i in each s ,

together with the N restricting equations.

Substituting unity for $\sum_{s:i} P_s \theta_{s_i}$, which is the coefficient of each x_i , when the expression in braces is rewritten, the above equation reduce to

$$P_s x_i \sum_{i \in s} x_i \theta_{s_i} + P_s \lambda_i = T P_s x_i, \text{ for } s=1, 2, \dots, S' \text{ and all } i \text{ in each } s$$

or

$$x_i \sum_{i \in s} x_i \theta_{s_i} + \lambda_i = T x_i.$$

The restricting equations which we have used to simplify the above equations are

$$\sum_{s:i} P_s \theta_{s_i} = 1, \text{ for } i=1, 2, \dots, N.$$

There are $n \binom{N}{C_n} + N$ independent equations and the same number of unknowns (including the $N \lambda$'s) and therefore a unique solution exists. Again the algebraic solution of this system of linear simultaneous equations is cumbersome. Therefore let us simply say

$$o^{\theta}_{s_i}, \text{ for } s=1,2,\dots,S' \text{ and all } \underline{i} \text{ in each } \underline{s}$$

and

$$o^{\lambda}_i, \text{ for } i=1,2,\dots,N$$

represent the values obtained as a result of solving the above $n \binom{N}{C_n} + N$ equations. The minimum variance estimator T'_5 will then be symbolically represented by

$$T'_5 = \sum_{i \in s} o^{\theta}_{s_i} x_i.$$

Let us determine the expression for its variance. Multiplying the fourth equation on the previous page by θ_{s_i} , then summing over all \underline{i} in each \underline{s} and then over all \underline{s} , we find

$$\sum_{s=1}^{S'} P_s \left(\sum_{i \in s} x_i \theta_{s_i} \right)^2 + \sum_{s=1}^{S'} P_s \sum_{i \in s} \lambda_i \theta_{s_i} = T \sum_{s=1}^{S'} P_s \sum_{i \in s} x_i \theta_{s_i}.$$

This reduces to

$$V(T'_5) + T^2 + \sum_{i=1}^N \lambda_i \sum_{s \ni i} P_s \theta_{s_i} = T \sum_{i=1}^N x_i \sum_{s \ni i} P_s \theta_{s_i};$$

substituting the \underline{N} conditions of unbiasedness, we find

$$V(T'_5) + \sum_{i=1}^N \lambda_i = 0.$$

Now taking the set of equations

$$x_i \sum_{i \in s} x_i \theta_{s_i} + \lambda_i = T x_i$$

and summing over all \underline{i} in each \underline{s} and then over all \underline{s} , we will find

$$\sum_{s=1}^{S'} \left(\sum_{i \in s} x_i \right) \left(\sum_{i \in s} x_i \theta_{s_i} \right) + \sum_{s=1}^{S'} \sum_{i \in s} \lambda_i = T \sum_{s=1}^{S'} \sum_{i \in s} x_i$$

i.e.,

$$\sum_{s=1}^{S'} \left(\sum_{i \in s} x_i \right) \left(\sum_{i \in s} x_i \theta_{s_i} \right) + {}^{N-1}C_{n-1} \sum_{i=1}^N \lambda_i = {}^{N-1}C_{n-1} T^2,$$

or

$$\left(1/{}^{N-1}C_{n-1} \right) \sum_{s=1}^{S'} \left(\sum_{i \in s} x_i \right) \left(\sum_{i \in s} x_i \theta_{s_i} \right) - T^2 + \sum_{i=1}^N \lambda_i = 0$$

Eliminating $\sum_{i=1}^N \lambda_i$ between the two equations concerned, and writing θ_{s_i} for θ_{s_i} , we will find

$$\begin{aligned} V(T_5^2) &= \left(1/{}^{N-1}C_{n-1} \right) \sum_{s=1}^{S'} \left(\sum_{i \in s} x_i \right) \left(\sum_{i \in s} x_i \theta_{s_i} \right) - T^2 \\ &= \left(1/{}^{N-1}C_{n-1} \right) \sum_{s=1}^{S'} \left[\sum_{i \in s} x_i^2 \theta_{s_i} + \sum_{i \neq j \in s} x_i \theta_{s_i} x_j \right] - T^2. \end{aligned}$$

The unbiased estimate of $V(T_5^{\hat{}})$ will be given by

$$V(T_5^{\hat{}}) = (1/N^{n-1} C_{n-1}) \left[\frac{\sum_{i \in s} x_i^2 o_{s_i}}{N^{n-1} C_{n-1}} + \frac{\sum_{i < j \in s} x_i o_{s_i} x_j}{N^{n-2} C_{n-2}} \right]$$

$$= (1/P_s) \left[\frac{\sum_{i=1}^N x_i^2}{N^{n-1} C_{n-1}} + 2 \frac{\sum_{i < j \in s} x_i x_j}{N^{n-2} C_{n-2}} \right]$$

Clearly $\hat{V}(T_5^{\hat{}})$ can assume negative values.

That $V(T_5^{\hat{}})$ is the minimum variance of

$$T_5^{\hat{}} = \sum_{i \in s} o_{s_i} x_i$$

can be easily verified. When we set $o_{s_i} = N/n$ for all i in each s and for all s , the resulting expression for $V(T_5^{\hat{}})$ reduces to

$$N^2 (1/n) \cdot \frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N} [1 - (n-1)/(N-1)]$$

where

$$\bar{x} = (1/N) \sum_{i=1}^N x_i$$

the variance of the best linear unbiased estimate of the total in the case of sampling with equal probability and without replacement.

Again if there is some approximate relationship between x_i and a characteristic y_i , known for all i , such as $x = cy$, we can find a simulated minimum variance estimator by exactly the same procedures

as for the estimator in class three. Thus if θ'_{s_i} is the simulated value of θ_{s_i} , then this estimator will be

$$T_5^w = \sum_{i \in s} \theta'_{s_i} x_i,$$

and the estimate of its variance can be obtained by inserting θ'_{s_i} for θ_{s_i} in the expression for $\hat{V}(T_5^w)$. However, the number of equations for obtaining θ'_{s_i} for each s increases. For example, for samples of size $n = 2$ and $N = 4$ and 5 , the number of equations ^{1/} involved will be 16 and 25, respectively.

4.7 Class Six Estimators

We consider now the estimators in class six, the general form being given by

$$T_6 = \phi_s \sum_{i \in s} x_i.$$

We recall that there are in all $n! \binom{N}{n} = S$ possible samples taking into account the order of appearance of the elements and therefore S weights.

For T_6 to be an unbiased estimate whatever the values of the x 's, we must have

$$E(T_6) = \sum_{s=1}^S P_{s(o)} \phi_s \sum_{i \in s} x_i = \sum_{i=1}^N x_i \sum_{s>i} \phi_s P_{s(o)} \stackrel{1/}{=} \sum_{i=1}^N x_i,$$

^{1/} Still with electronic machines the solution of such equations is no problem.

and therefore we must have

$$\sum_{s \ni i} P_{s(o)} \phi_s = 1 \text{ for all } \underline{i}.$$

We recall that $P_{s(o)} = p_{i_1}^i p_{j_2}^i \dots p_{k_{n-1}}^{ij\dots l} p_{m_n}^{ij\dots k}$, and when summed for all possible $n!$ permutations

$$\sum_{i, j, \dots, m} P_{s(o)} = P_s.$$

Suppose $\phi_s = \phi_{s_i}$ for each possible set of n elements u_i, u_j, \dots, u_m which always include u_i ; this implies that the order of appearance of the elements is immaterial. Then the condition of unbiasedness becomes

$$\sum_{s \ni i} \phi_{s_i} \sum_{i, j, \dots, m} P_{s(o)} = 1$$

i.e.,

$$\sum_{s \ni i} \phi_{s_i} P_s = 1 \text{ for all } \underline{i}.$$

This leads to the unbiased estimator in class five. From this step we can obtain the condition leading to the estimator in class three. Next if each $\phi_s = \phi_{i_t}$ for every \underline{s} which contains \underline{i} at the t^{th} draw, then we will have

$$\sum_{t=1}^n \phi_{i_t} \sum_{s \ni i_t} P_{s(o)} = 1,$$

i.e.,

$$\sum_{t=1}^n \phi_{i_t} P_{i_t} = 1$$

for all i and this leads to the estimator in class four. Further if for any i , $\phi_s = \phi_i$ for all $s > i$ we will have $\phi_i P_s = 1$ and this, as considered above, leads to the unbiased estimator in class two.

One set of solutions satisfying the above set of N equations is also

$$\begin{aligned} \phi_s &= 1/n! \binom{N-1}{n-1} P_s(o) \\ &= \frac{N/n}{[Np_{i_1}] [(N-1)p_{j_2}^i] \dots [(N-n+1)p_{m_n}^{ij\dots k}]} \end{aligned}$$

Hence, one unbiased estimator is

$$T_6^j = \frac{(N/n) \sum_{i \in s} x_i}{[Np_{i_1}] [(N-1)p_{j_2}^i] \dots [(N-n+1)p_{m_n}^{ij\dots k}]}$$

One feature of this new estimator is that only those selection probabilities of the probability distributions used at each draw enter into its expression. This contrasts with other known estimators, excepting T_4^i and T_4^m , where other probability distributions in the system having no actual bearing on the formation of the sample enter into its expression. From a computational point of view this estimator appears to enjoy considerable advantage over all others proposed so far.

Lastly we note that when the selection probabilities are such that the expression in the denominator for T'_6 is close to 1, then the estimator is effectively like the one used in the equal probability sampling case. Indeed when the probabilities at each draw are equal the numerical value is 1.

The procedure for finding the minimum variance estimator in class six is along the same lines as for the one in class three. First we have

$$V(T'_6) = \sum_{s=1}^S P_{s(0)} (\phi_s \sum_{i \in s} x_i)^2 - \sum_{s=1}^S P_{s(0)} (\phi_s \sum_{i \in s} x_i)^2 \quad 2$$

The equations for solving ϕ for all s and λ_i ($i = 1, 2, \dots, N$) the Lagrange multipliers, are

$$\phi_s \sum_{i \in s} x_i + \sum_{i \in s} \lambda_i = T \sum_{i \in s} x_i, \quad s = 1, 2, \dots, S$$

and

$$\sum_{s \ni i} P_{s(0)} \phi_s = 1, \quad i = 1, 2, \dots, N$$

There are $n! \binom{N}{n} + N$ independent equations and therefore we can solve for the ϕ 's and the λ 's. If ϕ_s , for all s , and λ_i , for all i , are the solutions, then the minimum variance estimator will be given by

$$T''_6 = \phi_s \sum_{i \in s} x_i$$

and its variance by

$$V(T''_6) = \sum_{s=1}^S P_{s(0)} \phi_s^2 \left(\sum_{i \in s} x_i \right)^2 - T^2 = T \sum_{i=1}^N x_i P_i - T^2 - \sum_{i=1}^N \lambda_i P_i$$

The unbiased estimate of $V(T_6^u)$ will be

$$\hat{V}(T_6^u) = \sigma_s^2 \left(\sum_{i \in s} x_i \right)^2 = (1/n) P_{s(o)} \left[\frac{\sum_{i \in s} x_i^2}{N-1} + 2 \frac{\sum_{i < j \in s} x_i x_j}{N-2} \right],$$

it can assume negative values.

Here also the simulated minimum variance estimator can be worked out along the same lines as for the one in class three. For $n = 2$ and $N = 4, 5$, or more, the same number of equations will be involved in the solution of the β 's and λ 's as in the case of T_5^u . But for $n > 2$, the number of equations will necessarily be more. The estimate of the variance of the simulated minimum variance estimator, for any given s , can be obtained by substituting the simulated value of ϕ in the expression for $\hat{V}(T_6^u)$. It also can assume negative values.

4.8 Class Seven Estimators

Finally we come to the estimators in class seven, generally given by

$$T_7 = \sum_{i \in s} \psi_{s_{i_t}} x_i$$

where it will be recalled that $\psi_{s_{i_t}}$ is the weight to be attached to the i^{th} element appearing at the t^{th} draw in the s^{th} sample (whose elements appear in a specified order). The summation is only over the elements i appearing in s . Again s is used to identify the individual samples of n and the sub-subscript t in the weight is used to indicate the weight appropriate at the t^{th} draw. There are in all $n! N C_n$ possible samples and $nn! N C_n$ weights to be determined.

We have

$$E(T_7) = \sum_{s=1}^S P_{s(0)} \sum_{i \in S} \psi_{s_{i_t}} x_i = \sum_{i=1}^N x_i \sum_{s=1}^S P_{s(0)} \psi_{s_{i_t}}.$$

For T_7 to be unbiased and at the same time for the weights to be independent of the x 's we must have

$$\sum_{i=1}^N x_i \sum_{s=1}^S P_{s(0)} \psi_{s_{i_t}} = \sum_{i=1}^N x_i$$

and therefore we must have

$$\sum_{s=1}^S P_{s(0)} \psi_{s_{i_t}} = 1, \text{ for all } \underline{i}.$$

Clearly one set of solutions will be given by

$$\psi_{s_{i_t}} = 1/n!^{N-1} C_{n-1}^P s(0)$$

for all $i \in S$ and for all corresponding \underline{t} , and this leads to the estimator in class six. Similarly from the above condition of unbiasedness we can derive conditions leading to unbiased estimators in class five, four, three and two, and, when all selection probabilities are equal, also one.

The determination of the minimum variance estimator for class seven proceeds on the same lines as for class five. We have

$$V(T_7) = \sum_{s=1}^S P_{s(0)} \left(\sum_{i \in S} \psi_{s_{i_t}} x_i \right)^2 = \left[\sum_{s=1}^S P_{s(0)} \sum_{i \in S} \psi_{s_{i_t}} x_i \right]^2$$

We set up the function

$$H = V(T_7) + \sum_{i=1}^N \lambda_i \left(\sum_{s>i} \psi_{s_i_t} P_{s(o)} = 1 \right)$$

where the λ_i 's are the Lagrange multipliers specific to each of the N restrictions stemming from the condition of unbiasedness, and solve the equations

$$\partial H / \partial \psi_{s_i_t} = 0 = 2P_{s(o)} \left(\sum_{i \in s} \psi_{s_i_t} x_i \right) x_i - 2 \left[\sum_{s=1}^S P_{s(o)} \sum_{i \in s} \psi_{s_i_t} x_i \right] P_{s(o)} x_i$$

$$+ P_{s(o)} \lambda_i, \text{ for } s = 1, 2, \dots, S \text{ and for all } \underline{i} \text{ in each } \underline{s},$$

together with the \underline{N} independent restricting equations

$$\sum_{s>i} P_{s(o)} \psi_{s_i_t} = 1, \text{ for } i = 1, 2, \dots, N.$$

These $n \sum_{i=1}^N C_n + N$ equations are all independent and therefore a unique solution exists. Hence, a minimum value of $V(T_7)$ exists. Let the solution be symbolically represented as

$$\circ \psi_{s_i_t}, \text{ for all } \underline{s} \text{ and for all } \underline{i} \text{ in each } \underline{s}$$

$$\circ \lambda_i, \text{ for all } \underline{i}.$$

Thus, the minimum variance estimator $T_7^{\#}$ is given by

$$T_7^{\#} = \sum_{i \in s} \circ \psi_{s_i_t} x_i$$

and its variance by

$$V(T_7^2) = (1/n) \binom{N-1}{n-1} \sum_{s=1}^S \left(\sum_{i \in s} x_i \right) \left(\sum_{i \in s} x_i \psi_{s_{i,t}} \right) - T^2,$$

the unbiased estimate of which is

$$\begin{aligned} \hat{V}(T_7^2) &= (1/n) \binom{N-1}{n-1} \left[\left(\sum_{i \in s} x_i^2 \psi_{s_{i,t}} \right) / \binom{N-1}{n-1} \right. \\ &\quad \left. + \left(\sum_{i \neq j \in s} x_i \psi_{s_{i,t}} x_j \right) / \binom{N-2}{n-2} \right] \\ &= \frac{1}{n \binom{N-1}{n-1}} \left[\frac{\sum_{i \in s} x_i^2}{\binom{N-1}{n-1}} + 2 \frac{\sum_{i < j \in s} x_i x_j}{\binom{N-2}{n-2}} \right] \end{aligned}$$

for any given sample s . $\hat{V}(T_7^2)$ can assume negative values.

Needless to say, as in the case for the previous six minimum variance estimators, we cannot use T_7^2 unless auxiliary information is available. For example, if there is an approximate relationship between x and y such as $x_i = cy_i$ for all y in the universe, we can construct a simulated minimum variance estimator on exactly the same lines as for the estimator, say, in class five. In this case there will be more simulated values to determine than in all the previous cases.

4.9 Summary

In summary we find that in each of the seven classes, unbiased estimators exist; in the case of class one such an estimator exists

only when the selection probabilities are equal. Also all unbiased estimators reduce to the one in class one when all selection probabilities are equal at each draw. All known unbiased estimators have been classified. These estimators have weights which are functions of the probabilities. An estimator, believed to be new, has been derived (from the condition of unbiasedness) $\frac{1}{n}$ in class six. So far it has not been possible to derive analogous types of unbiased estimators in class five and seven any different from those already derived. From the condition of unbiasedness in class seven every estimator can be derived. From six, every estimator except seven can be derived. From five, estimators in class three, two and one can be derived. From four, estimators in class two and one can be derived. From three, two and one can be derived and finally from two, one can be derived (when the selection probabilities are all equal at each draw).

Further minimum variance estimators exist in each class, though with weights not independent of the properties of the population. From a severely practical point of view they are useless, but they provide the means for the construction of other estimators. Thus estimators with simulated weights, obtained by the use of auxiliary information, called simulated minimum variance estimators have been obtained; those in class one, two and four appear cumbersome, but those in the remaining classes are much easier to construct. Both formal and simulated variances and their corresponding estimates have been obtained.

$\frac{1}{n}$ It can be derived also intuitively.

How closely the simulated minimum variance approaches the true minimum variance will depend on the exactitude of the relationship between the characteristic under study and the characteristic chosen to determine the simulated weights.

V. FURTHER NOTES ON UNBIASED ESTIMATORS

5.1 General Notes

In Chapter IV unbiased estimators in each class with weights which are explicit functions of the probabilities were derived. In class one the best linear unbiased estimator was derived as a special case when all probabilities are equal.

Considering these special types of estimators, we have a unique one in class two (given on page 46), a sub-class in class three (given on page 51), three sub-classes in class four (given on pages 57, 63 and 64, respectively), and a sub-class in class six (given on page 75). Thus 12 interclass comparisons are possible among the estimators in these four classes; and within class four, three intraclass (or inter-sub-class) comparisons are possible. Alternatively one might ask which of these six estimators has the least variance, given the probability system and sample size.

5.2 Comparison of Estimators

From what follows it will appear that we are only able to state definitely that the variance of the unbiased estimator in class three is always less than that in class six. Regarding the 14 remaining comparisons, we are, in the present state of knowledge, only able to say that the relative magnitudes of the variances in question are dependent on the probability system.

Let us compare the special estimators in class two and class three. We recall that these estimators are given by

$$T_2^* = \sum_i (x_i/P_i)$$

and

$$T_3^* = (1/N-1 C_{n-1} P_s) \sum_{i \in S} x_i \quad .$$

Their variances are respectively

$$V(T_2^*) = \sum_{i=1}^N (x_i^2/P_i) + 2 \sum_{i < j} x_i x_j (P_{ij}/P_i P_j) - T^2,$$

and

$$\begin{aligned} V(T_3^*) &= (1/N-1 C_{n-1})^2 \sum_{s=1}^{S:} \frac{\left(\sum_{i \in S} x_i \right)^2}{P_s} - T^2 \\ &= (1/N-1 C_{n-1})^2 \sum_{s=1}^{S:} \frac{\sum_{i \in S} x_i^2 + 2 \sum_{i < j \in S} x_i x_j}{P_s} - T^2 \\ &= (1/N-1 C_{n-1})^2 \left[\sum_{i=1}^N x_i^2 \sum_{s>i} (1/P_s) + 2 \sum_{i < j} x_i x_j \sum_{s>i,j} (1/P_s) - T^2 \right] \end{aligned}$$

where $\sum_{s>i,j}$ denotes summation over all samples which contain u_i and u_j .

Now T_2^* is more efficient than T_3^* if $V(T_3^*) - V(T_2^*) > 0$. We find

$$\begin{aligned} V(T_3^*) - V(T_2^*) &= \left\{ \sum_{i=1}^N x_i^2 \left[\frac{1}{N-1 C_{n-1}} \sum_{s>i} (1/P_s) - \frac{1}{P_i} \right] \right. \\ &\quad \left. + 2 \sum_{i < j} x_i x_j \left[\frac{1}{N-1 C_{n-1}} \sum_{s>i,j} (1/P_s) - \frac{1}{P_i P_j} \right] \right\} / N-1 C_{n-1} \end{aligned}$$

Before proceeding to simplify the terms in braces in the expression above, with a view to obtaining more meaningful quantities, we note that the total number of samples containing u_i is ${}^{N-1}C_{n-1}$ and therefore the total number of P_s with \underline{s} which include \underline{i} is ${}^{N-1}C_{n-1}$. Similarly the total number of P_s with \underline{s} which include \underline{i} and \underline{j} is ${}^{N-2}C_{n-2}$. Clearly therefore $\sum_{s \supset i} (1/P_s) / {}^{N-1}C_{n-1}$ is the reciprocal of the harmonic mean of all P_s which include \underline{i} and also $\sum_{s \supset i, j} (1/P_s) / {}^{N-2}C_{n-2}$ is the reciprocal of the harmonic mean of all P_s which include \underline{i} and \underline{j} . Let us denote these two harmonic means by H_i^P and H_{ij}^P , respectively. Further we can write

$$P_i = \sum_{s \supset i} P_s = {}^{N-1}C_{n-1} A_i^P$$

and

$$P_{ij} = \sum_{s \supset i, j} P_s = {}^{N-2}C_{n-2} A_{ij}^P$$

where A_i^P and A_{ij}^P denote the arithmetic means of all P_s which include \underline{i} and all P_s which include \underline{i} and \underline{j} respectively. Substituting these expressions into the formula for $V(T_3^i) - V(T_2^i)$ we have

$$\begin{aligned} V(T_3^i) - V(T_2^i) &= \sum_{i=1}^N x_i^2 \left[(1/H_i^P) - (1/A_i^P) \right] \\ &+ 2 \left(\frac{n-1}{N-1} \right) \sum_{i < j} x_i x_j \left(\frac{1}{H_{ij}^P} - \frac{A_{ij}^P}{A_i^P A_j^P} \right) \Bigg/ {}^{N-1}C_{n-1} . \end{aligned}$$

Now the harmonic mean is always less than the arithmetic mean (Hardy et al., 1934) and therefore for any given i ,

$$(1/H^P_i) - (1/A^P_i) > 0.$$

However, given a probability system, for some pairs of i and j

$$(1/H^P_i) - (A^P_{ij})/(A^P_i A^P_j)$$

may be positive and for others negative or zero.

Treating the expression for $V(T^1_3) - V(T^1_2)$ as a quadratic form, we can write out the N conditions under which it will be positive definite but the algebra is quite cumbersome.

We can definitely state, however, that for some probability systems it will be positive definite and for others indefinite. This is because there are $G - D > N$ values of arbitrary probabilities which we are free to choose in order to make the quadratic form either positive definite or indefinite. In this particular comparison it can never be negative definite because the special harmonic-arithmetic inequality, given above, is always positive. Thus for some probability systems T^1_2 will be more efficient than T^1_3 and for others vice versa. What these systems are is another matter.

The examination as to whether $V(T^1_3) - V(T^1_2)$ is positive definite, for any particular probability system, will involve the examination of all inequalities (resulting from the expansion of all principal minors of the determinant of the matrix) made up from various kinds of harmonic and arithmetic means described above. Certainly the mathematical problems here are very complex. Further, an examination may not lead to any meaningful interpretation.

Similar conclusions will be arrived at and similar difficulties will be encountered for other comparisons in respect to the entire set of probability systems, or any particular one thereof, except for the case discussed below.

Next let us compare the unbiased estimator in class six with that in class three. We have

$$T_0^3 = \frac{1}{n! \binom{N-1}{n-1} P_{s(o)}} \sum_{i \in s} x_i$$

and

$$V(T_0^3) = \frac{1}{(n! \binom{N-1}{n-1})^2} \sum_{s=1}^S \frac{\left(\sum_{i \in s} x_i \right)^2}{P_{s(o)}} = T^2$$

We find

$$\begin{aligned} V(T_0^3) - V(T_3^3) &= \left[\sum_{s=1}^S \frac{\left(\sum_{i \in s} x_i \right)^2}{(n!)^2 P_{s(o)}} - \sum_{s=1}^S \frac{\left(\sum_{i \in s} x_i \right)^2}{P_s} \right] / \binom{N-1}{n-1}^2 \\ &= \sum_{s=1}^S \left(\sum_{i \in s} x_i \right)^2 \left(\frac{1}{(n!)^2} \sum_{i \neq j \neq \dots \neq m} \frac{1}{P_{s(o)}} - \frac{1}{P_s} \right) / \binom{N-1}{n-1}^2 \end{aligned}$$

where $\sum_{i \neq j \neq \dots \neq m}$ denotes summation of the reciprocal of $P_{s(o)}$ (which is given by $p_i^i p_{j_2}^i \dots p_{m_n}^{ij \dots l}$) for any given sample \underline{s} (made up of elements $u_i, u_j, \dots, u_l, u_m$) for all the $n!$ possible orders of appearance of the elements. Now

$$(1/n!) \sum_{i \neq j \neq \dots \neq m} \frac{1}{p_i^i p_{j_2}^i \dots p_{m_n}^{ij \dots l}}$$

is the reciprocal of the harmonic mean of all $P_{s(o)}$'s for a given \underline{s} . Let us denote this quantity by $H^P_{s(o)}$. Similarly we have

$$A^P_{s(o)} = (1/n!) \sum_{i \neq j \neq \dots \neq m} P_{s(o)} = (1/n!) P_s$$

where $A^P_{s(o)}$ is the arithmetic mean of all $P_{s(o)}$'s for a given \underline{s} .

Thus we find

$$V(T_0^{\dagger}) - V(T_3^{\dagger}) = \sum_{s=1}^{S^{\dagger}} \left(\sum_{i \in S} x_i \right)^2 \left(\frac{1}{H^P_{s(o)}} - \frac{1}{A^P_{s(o)}} \right) / (n!) \binom{N-1}{n-1}^2.$$

Since $H^P_{s(o)} < A^P_{s(o)}$ for all \underline{s} , $V(T_0^{\dagger}) - V(T_3^{\dagger}) > 0$ whatever the value of the x 's, and, of course, whatever the probability system. Hence T_3^{\dagger} is always more efficient than T_0^{\dagger} , a very pleasing result.

In the discussion so far I have omitted the expressions for the variance of the three unbiased estimators in class four. They are all given by Des Raj (1956). Also one of them is found in Das' (1951) paper. The expressions are very complicated and the problem of comparing their variances, for any given probability system, appears to be intractable as in the case of T_2^{\dagger} and T_3^{\dagger} .

5.3 Multiple Comparison of Efficiency

Finally if a multiple comparison of efficiency is attempted we note that we can always find a probability system when, for example, the following situation obtains:

$$V(T_2^{\dagger}) > V(T_0^{\dagger}) > V(T_4^{\dagger}) > V(T_4^{\dagger}) > V(T_4^{\dagger}) > V(T_3^{\dagger}),$$

provided

$$G - D = \sum_{r=2}^{n-1} (r-1) N C_r + n N C_n - 1 > 5N.$$

since there are five times the number of conditions to be satisfied here; the second inequality is true for $2 \leq n < N$, when $N \geq 7$. ^{1/}

We note that $V(T_3)$ must always be less than $V(T_6)$ in this system of inequalities. The order of magnitude of the variances in the inequality can be permuted except in respect to this restriction, and for each permissible permutation there corresponds a probability system.

Thus, except for the cases noted, we conclude that for some probability systems the magnitude of the variances will be in some unknown order, and for others in a different order but still unknown. In other words, the order of efficiency of the six estimators is dependent on the probability system. What the probability system is in relation to a given order of efficiency is something we are unable to answer at present.

^{1/} When $N = 3$ and $n = 2$, we find $G - D = 5$, so that the multiple inequality of six members (variances) will not hold. But for a simple inequality a probability system which meets the requirement can certainly be found. Other exceptional cases for $3 < N \leq 6$ and for the relevant values of n can easily be worked out.

VI. THE PROBLEM OF NEGATIVE ESTIMATES OF VARIANCE

6.1 Introductory Remarks

In sampling with unequal probabilities, negative estimates of variance have been found for almost all estimators. It appears that such estimates have been rejected because of the lack of meaningful interpretations to be given in each specific situation. On purely logical grounds, once we have accepted the premises behind a theory, we cannot reject its conclusions (provided of course they are validly deduced). One difficulty here is that there is a tendency in our thinking, perhaps unconsciously, to equate a quantity such as s^2 (the square of the standard error) to the estimator of a variance, and when the numerical result is negative we find it unsatisfactory, partly because in contrast to equal probability sampling we are always used to finding positive results, and partly because its square root is not meaningful in the context of current statistical theories used for the purpose of determining confidence intervals and tests of significance.

In what follows attempts will be made to interpret negative estimates of variance in certain classes of estimators. However, as we shall soon see, not all classes of estimators even allow the easy discovery, in meaningful terms, of the conditions under which their estimates of variance are positive. The main source of this difficulty lies in the complexity of the probability functions which are enmeshed in the expressions for these estimators.

6.2 The Case for Class Two Estimators

Regarding class two, Horvitz and Thompson's estimator of the variance is given by

$$\sum_{i=1}^n \left[\frac{(1-p_i)}{p_i^2} \right] x_i^2 + 2 \sum_{i < j} \left[\frac{(p_{ij} - p_i p_j)}{p_i p_j p_{ij}} \right] x_i x_j,$$

and Yates and Grundy's (1953) version of it by

$$\begin{aligned} & \sum_{i < j} \left[\frac{(p_i p_j - p_{ij})}{p_{ij}} \right] \left[\left(\frac{x_i}{p_i} \right) - \left(\frac{x_j}{p_j} \right) \right]^2 \\ & = \sum_{i=1}^n \left(\frac{x_i^2}{p_i} \right) \sum_{j(\neq i)=1}^n \left[\frac{(p_i p_j - p_{ij})}{p_{ij}} \right] + 2 \sum_{i < j} \left[\frac{(p_{ij} - p_i p_j)}{p_i p_j p_{ij}} \right] x_i x_j. \end{aligned}$$

Both these estimators can lead to negative estimates for certain samples, except in the case of the latter estimator for two special situations noted by Sen (1953). For each estimator we can, regarding its expression as a quadratic form, write down the conditions under which, for a given sample, it will be positive definite. However, it is not easy to interpret these conditions in a meaningful way. In this connection, Sen in the same paper advocated taking zero as the estimate whenever this estimator yields a negative value. As a practical measure it is certainly appropriate but we do not know what meaning it carries. For a given sample, the interpretation to be given to these sorts of conditions appears to be a basis for interpreting negative in relation to positive estimates of variances, as we shall see from the case of the unbiased estimator in class three.

6.3 The Case for $\hat{V}(T_3)$

This unbiased estimator in class three is given by

$$\begin{aligned} \hat{V}(T_3) &= \left(\frac{\sum_{i \in S} x_i}{N-1 \binom{N-1}{n-1} P_s} \right)^2 - (1/P_s) \left[\frac{\sum_{i \in S} x_i^2}{N-1 \binom{N-1}{n-1}} + 2 \frac{\sum_{i < j \in S} x_i x_j}{N-2 \binom{N-1}{n-2}} \right] \\ &= \frac{1}{N-1 \binom{N-1}{n-1} P_s} \left[\left(\frac{1}{N-1 \binom{N-1}{n-1} P_s} - 1 \right) \sum x_i^2 + 2 \left(\frac{1}{N-1 \binom{N-1}{n-1} P_s} - \frac{N-1}{n-1} \right) \sum_{i < j} x_i x_j \right] \end{aligned}$$

It can be shown that $\hat{V}(T_3) > 0$ when $(1/N \binom{N}{n}) \geq P_s$. To prove this result we establish the following lemma.

Lemma. The quadratic form

$$Q = a \sum_{i=1}^n y_i^2 + 2(a-b) \sum_{i < j} y_i y_j,$$

where $b > 0$, is positive definite, if, and only if, $a \geq [(n-1)/n]b$.

To prove this we note that the $n \times n$ determinant of the matrix of this quadratic form is

$$\begin{vmatrix} a & a-b & a-b & \dots & a-b \\ a-b & a & a-b & \dots & a-b \\ \cdot & & & & \\ \cdot & & & & \\ \cdot & & & & \\ a-b & & & & a \end{vmatrix}$$

The necessary and sufficient condition for Q to be positive definite is that

$$a > 0, \quad \begin{vmatrix} a & a-b \\ a-b & a \end{vmatrix} > 0, \quad \begin{vmatrix} a & a-b & a-b \\ a-b & a & a-b \\ a-b & a-b & a \end{vmatrix} > 0, \quad \text{and so on.}$$

Since $b > 0$, the second inequality reduces to

$$a > (1/2)b,$$

the third to

$$a > (2/3)b,$$

and finally the n^{th} inequality to

$$a > [(n-1)/n]b.$$

Since $b > 0$, the necessary and sufficient condition reduces to

$$a > [(n-1)/n] b.$$

When

$$a = [(n-1)/n] b,$$

then

$$Q = b \sum (y - \bar{y})^2 > 0,$$

where

$$\bar{y} = \sum y/n.$$

To apply this lemma we note that

$$a = (1/N^{n-1} C_{n-1} P_s) - 1,$$

and

$$b = (N-n)/(n-1) > 0$$

in the formula for $\hat{V}(T_3)$, disregarding the constant term outside the square braces. Hence $\hat{V}(T_3)$ is positive definite if and only if

$$(1/N^{n-1} C_{n-1} P_s) - 1 \geq (n-1)/n \cdot (N-n)/(n-1), \text{ i.e.}$$

$$(1/N^{n-1} C_{n-1} P_s) \geq (N/n), \text{ i.e.,}$$

$$(1/N C_n) \geq P_s$$

If for any \underline{s} , it happens that $P_s = 1/N C_n$, then by applying the lemma or even by substituting $1/N C_n$ for P_s , it can be easily shown that

$$\hat{V}(T_3) = N^2 \left\{ \sum (x - \bar{x})^2 / n(n-1) \right\} [1 - (n/N)] > 0.$$

We note that $1/N C_n$ is the probability of drawing any sample of size n when all selection probabilities are equal. Also this quantity is the arithmetic average of all P_s . Thus the practical upshot of this result is that all samples with high probabilities of appearing (in the sense of being greater than $1/N C_n$) will have negative estimates of variance. This also means that even if the selection probabilities, which make up each P_s , are optimum (in the sense that $V(T_3)$ is a minimum for these values whatever they may be) the above condition still holds.

6.4 An Example for $\hat{V}(T_3)$

Sometimes samples with high probabilities of appearing may also yield estimates which are near the true value as the following example given below illustrates.

We have the following data for a population of four units

\underline{i}	\underline{x}	\underline{y}
1	4	2
2	5	4
3	9	6
4	6	9

It is desired to estimate the total for the x -population by sampling two units. Suppose we know \underline{y} but not \underline{x} and \underline{y} is in some way related

to \underline{x} . Now in general it is extremely difficult to obtain the optimum probabilities (in the sense explained above) with which the x -units are to be selected. But for the estimator which we shall use, $T_3^!$, it is not difficult to determine the optimum P_s . This optimum P_s is given by

$$P_s = \sum_{i \in S} x_i / {}^{N-1}C_{n-1} T$$

where

$$T = \sum_{i=1}^N x_i.$$

The proof is as follows:

Here it is of interest to use the Cauchy inequality. Now $V(T_3^!)$ is a minimum, for variations in P_s , if

$$\sum_{s=1}^{S!} \frac{\left(\sum_{i \in S} x_i \right)^2}{P_s} = \sum_{s=1}^{S!} \left(\frac{\sum_{i \in S} x_i}{\sqrt{P_s}} \right)^2 \sum_{s=1}^{S!} (\sqrt{P_s})^2$$

is a minimum. This can be easily seen from the formula for $V(T_3^!)$. Now

$$\sum_{s=1}^{S!} \left(\frac{\sum_{i \in S} x_i}{\sqrt{P_s}} \right)^2 \sum_{s=1}^{S!} (\sqrt{P_s})^2 \geq \left[\sum_{s=1}^{S!} \frac{\sum_{i \in S} x_i}{\sqrt{P_s}} \cdot (\sqrt{P_s}) \right]^2$$

Equality is only attained when

$$\frac{\sum_{i \in S} x_i}{\sqrt{P_s}} / \sqrt{P_s} = \text{a constant for all } s,$$

i.e.,

$$\sum_{i \in S} x_i = \lambda P_s,$$

where λ is some constant which can be determined by summing over all s . We do this and find

$$\lambda \sum_{s=1}^{S'} P_s = \sum_{s=1}^{S'} \sum_{i \in s} x_i = {}^{N-1}C_{n-1} T,$$

so that

$$\lambda = {}^{N-1}C_{n-1} T,$$

and thus

$$P_s = \frac{\sum_{i \in s} x_i}{{}^{N-1}C_{n-1} T}.$$

Des Raj (1954) also stated this result, but he did not appear to be aware that it was an optimum for the P_s . One set of selection probabilities appropriate to this situation is to select the first unit with probability proportional to the size of the x -character and the remaining $n-1$ units with equal probabilities and without replacement. For these values of P_s , $V(T_3^*) = 0$. This result is not in itself of value. But it provides the basis for determining the near-optimum P_s . And coming back to our example, we must select the first unit with probability proportional to the x -value and the remaining unit with equal probabilities. In our case we do not know the x 's, but we know the corresponding y values which are in some way related to the x 's. Hence all we can do is to take

$$P_s = \frac{\sum_{i \in s} y_i}{{}^{N-1}C_{n-1} T_y}.$$

Again we have to "unscramble" this probability to obtain the "scrambled" selection probabilities in terms of the known y 's in order to select

the sample. This means that we select the first unit with probability proportional to its y -measure and the remaining unit with equal probability.

Then the formula for the estimate is

$$T_{3(n=2)}^{\dagger} = \sum_{i \in s} x_i / N-1 C_{n-1} P_s = \left(\frac{\sum_{i \in s} x_i}{\sum_{i \in s} y_i} \right) \cdot T_y$$

which is Lahiri's (1951) ratio estimate. ^{1/}

In our example, the possible samples, the P_s values in terms of y , the estimates, and the estimated variances are as follows.

Samples	P_s	$T_{3(n=2)}^{\dagger}$	$\hat{V}[T_{3(n=2)}^{\dagger}] = T_3^{\dagger 2} - (T_y / \sum y)(x_1^2 + x_j^2 + 6x_1 x_j)$
1,2	6/60	30	363
1,3	8/60	32 1/2	273 3/4
1,4	10/60	20	8
2,3	10/20	28	32
2,4	12/60	18 1/3	-65 5/9
3,4	14/60	21 3/7	-170 40/49

$V [T_{3(n=2)}^{\dagger}] = 26.53$

The total for x is 24. The sample (3,4) which yields an estimate of 21 3/7 which is the closest to the true value has a very large negative variance. The samples (1,4) and (2,3) which have the same probabilities of selection yield estimates which deviate equally from the true value but the estimates of their variances are different. Samples (1,2) and (1,3) for which each corresponding $P_s < 1/4 C_2$, have very large positive estimates of variance. The true variance $V [T_{3(n=2)}^{\dagger}] = 26.53$. Like in equal probability sampling, a low estimate of variance does not

^{1/} In his sampling procedure there is a direct method for selecting the sample s proportional to $\sum_{i \in s} y_i$, and P_s need not be "unscrambled."

necessarily imply that the estimate is always closer to its true value than one having a high estimate of variance. But in this example it will be noted that accuracy (meaning absolute deviation from true values) is associated with low estimates of variance and high probabilities of selecting the sample in question.

6.5 A Further Example for $\hat{V}(T_3)$ Illustrative of an Unwise Choice of Selection Probabilities

It is possible to manufacture an example, so that samples with high or low probabilities of selection may yield estimates deviating greatly from their true values by suitably manipulating the probabilities. Consider the same data again for \underline{x} , but this time the auxiliary information is not so "good" (in the sense that there is no high positive relationship between \underline{x} and \underline{y} in turn leading to a much higher variance than in 6.4).

\underline{i}	\underline{x}	\underline{y}	Samples	P_s	$T_{3(n=2)}^{\dagger}$	$\hat{V} [T_{3(n=2)}^{\dagger}]$
1	4	4	1,2	7/30	12 6/7	- 64 34/49
2	5	3	1,3	5/30	26	50
3	9	1	1,4	6/30	16 2/3	- 48 8/9
4	6	2	2,3	4/30	35	285
			2,4	5/30	22	2
			3,4	3/30	50	1030

$\hat{V} [T_{3(n=2)}^{\dagger}] = 124.79$

In the selection of samples of size $n = 2$, as in the first example, the first unit is selected with probability proportional to the new y -measure and with no replacement, and the second with equal probability. The results are given above. The samples with the highest and

lowest probabilities of selection have yielded the worst estimates. The cause of this state of affairs is to be found in the selection probabilities; we have associated large values of \underline{x} with small values of \underline{y} and based our selection probabilities on these \underline{y} 's. Also we have used what are believed to be optimum probabilities for the selection of each sample, but their use has not been of much help as evidenced by $V \left[T_{3(n=2)}^{\dagger} \right] = 124.79$; with the previous set of \underline{y} 's it was 26.53.

6.6 Some Interpretations of Negative Estimates of Variance for T_{3}^{\dagger}

In view of the above examples, we know that when T_{3}^{\dagger} is to be used as an estimator, negative estimates of variance are likely to occur whatever the probability system, except for those systems where the selection probabilities are such that each $P_s = 1/N C_n$. The question we have arrived at is this: How are negative estimates of variance to be treated? The above examples seem to give a clue to the answer. When the selection probabilities are derived from the near-optimum probabilities (optimum in the sense of P_s being optimum), then there appears to be meaning in using them as indicators of the precision attained in sampling. We have not introduced any quantity like s^2 to be equated to the estimate of the variance partly because in a sampling theory of this kind it is unnecessary (at least in the writer's view) $\frac{1}{s}$ and partly because in the present state of knowledge we are unable to give a meaningful interpretation to imaginary numbers in relation to the theory of statistical sampling. If there is basis for

$\frac{1}{s}$ When we are concerned with very small samples, as here, there is no basis for considering s^2 , since even if we do, we cannot use it for the purpose of determining confidence intervals in view of the fact that we do not have the advantages of large sample theory at our disposal.

belief that the selection probabilities are not appropriate (in the sense that the P_s which are derived from them are not near the optimum P_s) then there appears to be no way out of the difficulty. Indeed, a negative estimate may sometimes mean that the selection probabilities may not be appropriate, as revealed by our second example. This remark may be at least as meaningful as the assertion, necessitated by practical circumstances, that its estimate is zero.

6.7 Estimators in Class Four

There are three sub-classes of unbiased estimators in class four, T_4''' proposed by Das (1951) and T_4'' and T_4'''' by Des Raj (1956). Little can be said about the estimators of $V(T_4''')$ and $V(T_4''''')$ as quadratic forms in view of their complexity. Generally they all yield negative estimates of variance. However, Des Raj has shown that for a special form of T_4''''' given in his notation as

$$t_{\text{mean}} = (1/n) \sum_{i=1}^n t_i$$

(where the t 's are as defined on page 64) the estimate of the variance is

$$\hat{V}(t) = \left[(1/n) \sum_i t_i \right]^2 - \hat{T}^2 = (1/n) \cdot \frac{\sum_{i=1}^n (t_i - t_{\text{mean}})^2}{n-1}$$

and is therefore always positive. This property is partly a consequence of the result that $\text{Cov}(t_a, t_b) = 0$ for $a \neq b$, so that $E(t_a t_b) = T^2$, which leads to the result that an unbiased estimate of T^2 is

$$\hat{T}^2 = \frac{2}{n(n-1)} \sum_{a < b} t_a t_b .$$

6.8 Estimators in Class Five and Class Seven

There are no especial types of unbiased estimators for class five and class seven. Let us consider an unbiased estimator for class five. It will be given by, say

$$T_5^{u'} = \sum_{i=1}^n \theta_{s_i} x_i$$

where θ 's satisfy the condition of unbiasedness; i.e.,

$$\sum_{s \supset i} P_s \hat{\theta}_{s_i} = 1 \quad \text{for all } \underline{i}.$$

The estimate of the variance of $T_5^{u'}$ will be

$$\hat{V}(T_5^{u'}) = \left(\sum_{i=1}^n \hat{\theta}_{s_i} x_i \right)^2 - (1/P_s) \left[\frac{\sum x_i^2}{N-1 C_{n-1}} + 2 \frac{\sum_{i < j} x_i x_j}{N-2 C_{n-2}} \right].$$

We see that $\hat{V}(T_5^{u'})$ can be negative. Here also the conditions under which it will be positive are hard to determine. Similar formulas and remarks apply to any unbiased estimator in class seven.

6.9 The Case of $\hat{V}(T_6)$

The special type of unbiased estimator in class six is given by

$$T_6 = \frac{1}{n \binom{N-1}{n-1} P_s(o)} \sum_{i \in s} x_i,$$

with estimated variance, which is unbiased, by

$$\begin{aligned} \hat{V}(T_0) &= (T_0)^2 - \frac{1}{n!^{N-1} C_{n-1} P_S(o)} \left\{ \sum x_i^2 + 2 \left[\frac{(N-1)}{(n-1)} \right] \sum_{i < j} x_i x_j \right\} \\ &= \frac{1}{n!^{N-1} C_{n-1} P_S(o)} \left[\left(\frac{1}{n!^{N-1} C_{n-1} P_S(o)} - 1 \right) \sum x_i^2 \right. \\ &\quad \left. + 2 \left(\frac{1}{n!^{N-1} C_{n-1} P_S(o)} - \left[\frac{(N-1)}{(n-1)} \right] \right) \sum_{i < j} x_i x_j \right]. \end{aligned}$$

Applying the lemma established earlier, we will find that $\hat{V}(T_0)$ is positive definite if

$$\frac{1}{n!^{N-1} C_{n-1} P_S(o)} - 1 > \frac{n-1}{n} \cdot \frac{N-n}{n-1}$$

i.e.,

$$1/n!^N C_n \geq P_S(o)$$

or if

$$1/N \cdot 1/(N-1) \dots 1/(N-n+1) \geq p_{i_1}^i p_{j_2}^i \dots p_{i_1}^{ij\dots f} \dots p_{m_n}^{ij\dots l}.$$

It will be noted that $1/N \cdot 1/(N-1) \dots 1/(N-n+1)$ is the probability of obtaining a given sample in a given order when all probabilities are equal as also the p-product on the right-hand side of the inequality, but, when all selection probabilities at each draw are unequal.

When

$$P_{s(o)} = 1/n! {}^N C_n \text{ for any given } s,$$

we will find

$$\hat{V}(T_2^1) = N^2 \frac{\sum (x_i - \bar{x})^2}{n(n-1)} [1 - (n/N)] > 0$$

where $\bar{x} = \sum x/n$.

There is one distinctive feature about this estimator. The same set of elements drawn in a different order will give rise to a different estimate of the total and a different estimate of the variance despite the fact that the sample aggregate $\sum_{ies} x_i$ remains the same. Thus the same set of elements may sometimes have a positive estimate of variance and sometimes a negative one.

By the use of the Cauchy inequality we can, as in the case of T_3^1 , show that the optimum value of $P_{s(o)}$ is

$$\sum_{ies} x_i/n! {}^{N-1} C_{n-1} T.$$

In view of the similarity of conditions with respect to T_3^1 , negative estimates of variance here may also be interpreted as described in sections 6.4 and 6.5.

6.10 The Existence of Probability Systems for Which

$\hat{V}(T_2^1)$, $\hat{V}(T_3^1)$ and $\hat{V}(T_5^1)$ are Always Positive

One might ask whether for each class of unbiased estimator there are probability systems for which the estimate of the variance for every possible sample is positive.

We already know the answer for class one. The system where the elements are sampled with equal probability and without replacement yields estimates of variance which are always positive.

In class two there is only one unbiased estimator, and Sen has noted the two special cases in which the estimate of variance is positive; viz., (i) when the first unit is selected with an arbitrary probability and the remaining $n-1$ with equal probability at each draw and without replacement, and (ii), applicable only for the case of two elements when the first unit is selected with an arbitrary probability and the second with probabilities proportional to the probability masses of the remaining elements. (The latter statement is often phrased as "probability proportional to size.") We will show that there are probability systems for which the Horvitz-Thompson estimator of variance, or Yates and Grundy's version of it, is positive for every possible sample. These estimators can be written in a general way as

$$V = \sum_{i \in S} q_i x_i^2 + 2 \sum_{i < j \in S} q_{ij} x_i x_j$$

where q_i and q_{ij} are functions of the selection probabilities specific to the sample in question and specific to the type of estimator. Now there are only ${}^N C_n$ distinct samples possible, and therefore there are only ${}^N C_n$ distinct quadratic forms. With each estimator there are \underline{n} conditions to be satisfied for its corresponding quadratic form to be positive definite. Thus in all there are at most $n \cdot {}^N C_n$ conditions to be satisfied, perhaps not all distinct. Each of these \underline{n} conditions are that the principal minors of \underline{V} should all be positive, and the

are therefore in the form of inequalities. The members of these inequalities are made up of the q 's and therefore all the selection probabilities in the system. There are in all

$$G = N + (N-1) {}^N C_1 + (N-2) {}^N C_2 + \dots + (N-n+1) {}^N C_{n-1}$$

$$= \sum_{r=1}^n r {}^N C_r$$

arbitrary probabilities in the entire system. These arbitrary probabilities are subject to

$$D = 1 + {}^N C_1 + {}^N C_2 + \dots + {}^N C_{n-1} = 1 + \sum_{r=1}^{n-1} {}^N C_r$$

restrictions of the type

$$\sum_{j(\neq t)=1}^N p_j^t = 1, \text{ etc.}$$

Viewing the arbitrary probabilities as quantities which we are free to choose (subject of course to the restrictions mentioned above) we have altogether

$$G - D = \sum_{r=2}^{n-1} (r-1) {}^N C_r + n {}^N C_n - 1 \quad (\text{for } n \geq 3)$$

of them. Now we have noted above that for the ${}^N C_n$ quadratic forms (and therefore the estimators) to be positive definite, $n {}^N C_n$ conditions, at most, must be satisfied. Clearly we can choose the various p 's such that these conditions are all satisfied since

$$G - D > n \cdot {}^N C_n, \quad \text{for } n \geq 3$$

For the case $n = 2$, $G - D = N^2 - N - 1$, and the exact number of conditions (regarding the principal minors) to be satisfied is

$N + {}^N C_2 = [N(N+1)] / 2$. Here also $G - D > [N(N+1)] / 2$. This proves the above assertion. Thus the two probability systems on which Sen's two cases rest are part of a larger group of systems all of which yield positive estimates of variance for every possible sample. The identification of these systems is another matter; certainly the task is not easy.

Regarding class three we have already shown that when P_s happens to be equal to $1/{}^N C_n$, then $\hat{V}(T_3^s) > 0$. Now consider the case when

$$P_s = \sum_{i_1, j_1, \dots, i_n, j_n} p_{i_1} p_{j_1} \dots p_{i_n} p_{j_n} = 1/{}^N C_n$$

There are ${}^N C_n$ such equations and $G - D$ arbitrary probabilities which we are free to choose. Since $G - D > {}^N C_n$, there is a group of systems (of which the equal probability case is just one) for which every possible $\hat{V}(T_3^s) > 0$, but only arising from the condition $P_s = 1/{}^N C_n$ for all s .

It is difficult to establish the proposition for estimators T_4^u , T_4^b and T_4^c in class four; as also for class six and class seven. For the unbiased estimator T_5^b , in class five, where there are ${}^N C_n$ distinct quadratic forms corresponding to each estimator, the proof is exactly as given above for the class two estimator.

Dev Raj's special estimator T_4^u (when $c_i = 1/n$) is an example of an estimator accommodating itself to any system. Whatever the probability system, the estimate of its variance is always positive.

VII. A NOTE ON OPTIMUM PROBABILITIES

7.1 Introductory Note

In the foregoing account we have been speaking about arbitrary selection probabilities in a most general way. We mentioned in section 1.3 that optimum values of these probabilities may be chosen in any defined sense.

The notion of optimum probabilities is due to Hansen and Hurwitz (1949). They spoke of it to mean the set of selection probabilities that minimized "the variance for a fixed cost of obtaining sample results," or alternatively "the cost for a fixed sampling error."

7.2 A Simple Example

Leaving aside the question of costs, which is admittedly unrealistic, we can determine at least theoretically the set of probabilities that will minimize the variance of any estimator, say T_a , having a variance, $V(T_a)$. For this we set

$$H = V(T_a) + \sum \lambda (\sum p - 1)$$

and solve the equation

$$\partial H / \partial p = 0$$

for all p together with the entire set of restrictions, one of which is symbolized by $\sum p = 1$. The actual solutions are not easy in practice. Further we can only proceed if we have some knowledge of the x 's through some related characteristic y .

The following example relating to Horvitz and Thompson's estimator for the case $N = 3$ and $n = 2$, with probabilities for the selection of

the second unit dependent on the initial probabilities, illustrates the problem. We have

$$\begin{aligned}
 V(T) = & \frac{x_1^2}{p_1(1 + \frac{p_2}{1-p_2} + \frac{p_3}{1-p_3})} + \frac{x_2^2}{p_2(1 + \frac{p_1}{1-p_1} + \frac{p_3}{1-p_3})} \\
 & + \frac{x_3^2}{p_3(1 + \frac{p_1}{1-p_1} + \frac{p_2}{1-p_2})} + \frac{2x_1x_2}{p_1p_2(\frac{1}{1-p_1} + \frac{1}{1-p_2})} \\
 & + \frac{2x_2x_3}{p_2p_3(\frac{1}{1-p_2} + \frac{1}{1-p_3})} + \frac{2x_3x_1}{p_3p_1(\frac{1}{1-p_3} + \frac{1}{1-p_1})} - T^2.
 \end{aligned}$$

Let us assume we know about the x 's by an approximate relationship $x = cy$. To find the optimum p 's we set

$$H = V(T) + \lambda(p_1 + p_2 + p_3 - 1),$$

where λ is the undetermined Lagrange multiplier, and put

$$\begin{aligned}
 \partial H / \partial p_1 = 0 = & \lambda + \frac{(-1)x_1^2}{p_1^2(1 + \frac{p_2}{1-p_2} + \frac{p_3}{1-p_3})} + \frac{(-1)x_2^2}{p_2^2(1 + \frac{p_1}{1-p_1} + \frac{p_3}{1-p_3})^2} \cdot \frac{1}{(1-p_1)^2} \\
 & + \frac{(-1)x_3^2}{p_3^2(1 + \frac{p_1}{1-p_1} + \frac{p_2}{1-p_2})^2} \cdot \frac{1}{(1-p_1)^2} \\
 & + \frac{(-2)x_1x_2}{p_2(\frac{p_1}{1-p_1} + \frac{p_1}{1-p_2})^2} \left[\frac{1}{(1-p_1)^2} + \frac{1}{1-p_2} \right] \\
 & + \frac{(-2)x_3x_1}{p_3(\frac{p_1}{1-p_3} + \frac{p_1}{1-p_1})^2} \left[\frac{1}{(1-p_1)^2} + \frac{1}{1-p_3} \right]^2
 \end{aligned}$$

We have two similar equations corresponding to setting

$$\partial H / \partial p_2 = 0 = \partial H / \partial p_3$$

There are four equations and four unknowns (including λ) and therefore, theoretically at least, they can be solved. Technically, of course, the solution is not easy. We may solve them by Newton's method.

Let the function obtained by eliminating λ between $\partial H / \partial p_1 = 0$ and $\partial H / \partial p_2 = 0$, and then writing $p_3 = 1 - (p_1 + p_2)$ be denoted by

$$f_{12}(p_1, p_2) = 0.$$

Also let the function obtained by eliminating λ between $\partial H / \partial p_2 = 0$ and $\partial H / \partial p_3 = 0$ and then writing $p_3 = 1 - (p_1 + p_2)$ be denoted by

$$f_{23}(p_1, p_2) = 0.$$

These two equations involve p_1 and p_2 only. If (p_1^i, p_2^i) is an approximate solution sufficiently close to the true solution (p_1, p_2) then we will have by Taylor's theorem,

$$f_{12}(p_1, p_2) = 0 = f_{12}(p_1^i, p_2^i) + (p_1 - p_1^i)(\partial f_{12} / \partial p_1^i) + (p_2 - p_2^i)(\partial f_{12} / \partial p_2^i),$$

and

$$f_{23}(p_1, p_2) = 0 = f_{23}(p_1^i, p_2^i) + (p_1 - p_1^i)(\partial f_{23} / \partial p_1^i) + (p_2 - p_2^i)(\partial f_{23} / \partial p_2^i),$$

neglecting second and higher order derivatives which will be insignificant compared with those of the first order. Thus we can solve for p_1 and p_2 from the above simultaneous equations which are linear in p_1, p_2 . Using these values again, we can obtain an improved solution. A prolonged iterative procedure, however, will not be necessary if the right set of values can be guessed. We may start by guessing that

$p_1^i = y_i / (y_1 + y_2 + y_3)$. Obviously this is about the right order of magnitude. When we have obtained sets of (p_1, p_2) which do not differ appreciably then we may stop and take p_3 as given by

$$p_3 = 1 - (p_1 + p_2)$$

These will be the initial probabilities of selection which will make $V(T_a)$ a minimum.

7.3 A Formal Statement of the Problem of Optimum Probabilities in Multivariate Sampling with General Cost Functions

When we take cost into consideration the problem can be formulated most generally as follows. Let C_i be the cost for sampling the i^{th} element ($i=1, 2, \dots, N$), whatever the form of the cost function. For example, C_i may be a function of the population, size of physical area, if the unit of sampling is a given delineated area, etc. The expected cost will be given by

$$E = \sum_{i=1}^N P_i C_i.$$

Suppose we wish to use some unbiased estimator T_a belonging to one of the seven classes for the estimation of k characters. Let the allowable amount of variance to be fixed for the g^{th} character be F_g . Then we set up the function

$$H = E + \sum_{g=1}^k \lambda_g [V(T_a) - F_g] + \sum \lambda (\sum p - 1),$$

where the λ 's are the undetermined Lagrange multipliers, and solve the resulting simultaneous equations,

$$\delta H / \delta p = 0,$$

for all selection probabilities, p , together with the equations

$$V(\underset{g}{T}_a) = \underset{g}{F},$$

and the entire set of equations symbolized by

$$\sum p = 1.$$

When $V(\underset{g}{T}_a)$ is differentiated with respect to the p 's, functions of $\underset{g}{x}$'s will appear and therefore we must have auxiliary information about each $\underset{g}{x}$ through some other variable, say $\underset{g}{y}$, for all $\underset{g}{g}$, to solve these equations. The optimum values of the p 's obtained will thus be optimum for the multivariate sampling of \underline{k} characters.

VIII. SUMMARY AND CONCLUSIONS

8.1 Summary

An examination of a unified theory of sampling for a finite universe propounded by Godambe, relevant only to the case when the elements, as units of sampling, or in clusters of non-overlapping elements, are drawn one at a time, shows that it encompasses several special fields each depending on whether or not the units are replaced after each draw. When the sampling is one stage the following situations are conceivable: (i) when the elements are not replaced after each draw, (ii) when they are occasionally replaced, (iii) when they are always replaced, and (iv) a "chaotic" situation when some or all of the elements, none of which have been previously replaced, are replaced prior to a given draw. Theories of sampling specific to each situation can be developed. In this thesis the special theory of sampling appropriate to situation (i) has been constructed. There appears to be a need for such a theory for this particular field because during the past seven years various estimators have been proposed and the logical connections between them in the wider perspective of such a theory have not been explored. A deductive approach is necessary for such a task.

As a prerequisite the most general probability system appropriate to this situation has been defined in Chapter II and expressions for all a priori probabilities needed for the development of the theory have been given.

In Chapter III, the three axioms of sample formation have been stated. Seven general linear estimators were constructed on the basis of these axioms.

T_1 has weights which are based on the order of appearance, or, of drawing the elements, T_2 on the presence or absence of a given element in the sample, T_3 on the set of elements composing the sample, T_4 on the appearance of a given element at a given draw, T_5 on the given element and the particular sample in which it appears, T_6 on the set of elements appearing in a specified order, and finally T_7 on the element, the order of its draw and the particular sample in which it appears. Each typifies a class of estimator.

From the condition of unbiasedness special estimators were determined in each class. One new special unbiased estimator was obtained in class six. Also minimum variance estimators were obtained in each class. However, the weights of each of these minimum variance estimators were not independent of the properties of the population; i.e., they were functions of the x-values.

By the use of auxiliary information, in the form of an approximate relationship, $x = cy$, presumed to hold and known for all y's in the universe, all weights in each class were simulated. In practice these weights will be obtained by solving certain systems of linear simultaneous equations specific to each class. Thus seven simulated minimum variance estimators were obtained.

The efficiencies of six known special types of unbiased estimators (T_2^I , T_3^I , T_4^I , T_4^{II} , T_4^{III} , T_6^I) were compared. It was found that $V(T_6^I) > V(T_3^I)$ whatever the probability system. In regard to the fourteen remaining comparisons, the relative efficiencies depend on the probability system.

On the question of multiple comparisons of efficiency, in the present state of knowledge we can only say that the order of efficiency

depends on the kind of probability system used. This means that for some systems it will be in one order and for others in another order.

The problem of negative estimates of variance was discussed. Some meaningful interpretations were given for the estimators T_3^1 and T_6^1 . For others, such kinds of interpretations could not be found in view of the refractory nature of the inequalities emerging from their respective quadratic forms.

The problem of optimum probabilities was discussed for a simple case and a solution given by way of an example. Also the problem of optimum probabilities in multivariate sampling, taking into account costs, was formulated and the necessary equations given.

8.2 Conclusions

Some of the statements in section 8.1 are in the form of conclusions. Before we proceed to other conclusions it should be noted that in any study dealing with the problems of unequal probability sampling it is of the utmost importance to define the probability system in sufficient detail for the following reasons:

- (a) It is the key to the problem of relationships between the various classes of estimators.
- (b) It determines whether or not one kind of estimator has a larger variance than another except for the one comparison of $V(T_6^1)$ and $V(T_3^1)$.

Regarding the estimators, the most general estimator T_7 includes all others as special cases. Neither T_7 nor T_5 yield any distinct special forms. Also the order of generality decreases as we proceed downwards from T_7 to T_1 in the following sense. From the condition of

unbiasedness of T_7 we can derive all other estimators. Next, from the condition of unbiasedness of T_6 we can derive all others except T_7 , and so on down to T_1 , one special form of which is the best linear unbiased estimator in the case of equal probabilities of selection.

A new special unbiased estimator, T_6^i , belonging to class six, was found. Of those known, it is the simplest to compute. Unfortunately it is less efficient than T_3^i . However, when the probabilities of selection at each draw are not too unequal it will be almost as efficient as T_3^i .

The simulated minimum variance estimators like the other unbiased estimators with few exceptions can have negative estimates of variance.

Regarding the estimates of variance for T_3^i and T_6^i , even before they are computed we can determine whether the outcome will be positive or negative simply by computing the probability, P_s or $P_{s(o)}$, as the case may be, and determining whether $P_s > 1/N C_n$ or $P_{s(o)} > 1/[N(N-1)\dots(N-n+1)]$. We note that $P_{s(o)}$ is far easier to compute than P_s , which involves the summation of $n!$ products. When there is an excess the estimate in both cases will be negative; otherwise it will be positive. The most probable samples have negative estimates of variance. In the situation where the probabilities of the appearance of each of the possible samples are optimum or near-optimum, then negative estimates of variance appear to be associated with estimates which are closer to the true population value under study. The first example discussed shows this. When there is an "unwise" or misconceived choice of selection probabilities anything can happen; this is also evident on purely theoretical grounds.

No meaningful interpretation of negative estimates of variance of T_2^I , T_4^{II} , T_4^{III} , T_4^{IV} , in all their special forms, appears to be possible in the present state of knowledge. The cause for this has already been stated above.

8.3 Suggestions for Future Work

The exploration of all remaining special fields, already indicated at length in Chapter I, encompassed by the present unified theory will certainly be of interest.

What Sen (1952) and Durbin (1953) were doing when they attempted comparisons of efficiency when sampling with, as against sampling without replacement can be termed as inter-field comparisons. Many more such comparisons between estimators in the different fields (and of course within each field) are possible. The possibility of interesting and fruitful conclusions arising from such studies cannot be excluded.

We note again that the axioms of sample formation, posited here, can produce all the possible estimators, in each field, in all their ramifications.

The identification of more probability systems which yield positive estimates of variance for all possible samples, when T_2^I or T_3^I is used, should be of interest since it appears that only these two special estimators have this possibility.

On a more practical level, with all needed information available, it is of importance to compare the efficiencies of the seven simulated minimum variance estimators of this special theory. With the use of electronic computing machines, the solution of the set of linear

simultaneous equations for the simulated weights specific to each estimator, and other calculations involved (including the evaluation of probabilities) should not be difficult.

We have only considered the simplest approximate relationship $x = cy$, to obtain the simulated minimum variance estimators. Equally, other functional forms, either algebraic or even transcendental, can be used, provided they are appropriate.

For each estimator if we know the simulated weights we would be in a position to compute each possible estimate and its estimated variance. Again their computation is no problem with electronic machines. Here it would be instructive to examine the relationship between the estimator and its estimated variance, e.g., between $T_{\frac{1}{3}}$ and $\hat{V}(T_{\frac{1}{3}})$, with a view to securing meaningful interpretations for negative variances in relation to the probability system used.

Also for any probability system considered above, the empirical distribution of the estimated variance would be available and it would be of interest to see what kinds of functional relationships result in distributions with wide spread or with narrow spread. Such studies would throw light on the appropriateness of the functional relationship used.

LITERATURE CITED

- Cansado, E. 1955. Sampling without replacement from finite populations. Proceedings of the International Statistical Institute, Rio de Janeiro (1955) (to appear in Trabajos de Estadística).
- Carver, H. C. 1930. Fundamentals of the theory of sampling. Ann. Math. Stat. 1(1): 101-121.
- Church, A. E. R. 1926. On the means and squared standard deviations of small samples from any population. Biometrika 18: 321-394.
- Das, A. C. 1951. On two-phase sampling and sampling with varying probabilities. Bull. Inter. Stat. Inst. 33: 105-112.
- Des Raj. 1954. Ratio estimation in sampling with equal and unequal probabilities. Jour. Ind. Soc. Agric. Stat. 6(2): 127-138.
- Des Raj. 1956. Some estimators in sampling with varying probabilities without replacement. Jour. Amer. Stat. Assoc. 51: 269-284.
- Durbin, J. 1953. Some results in sampling theory when the units are selected with unequal probabilities. Jour. Roy. Stat. Soc., B 15: 262-269.
- Edgeworth, F. Y. 1918. On the value of a mean as calculated from a sample. Jour. Roy. Stat. Soc. 81: 624-632.
- Godambe, V. P. 1955. A unified theory of sampling from finite populations. Jour. Roy. Stat. Soc., B 17(2): 269-278.
- Hansen, M. H., and Hurwitz, W. N. 1943. On the theory of sampling from finite populations. Ann. Math. Stat. 14(4): 333-362.
- Hansen, M. H., and Hurwitz, W. N. 1949. On the determination of optimum probabilities in sampling. Ann. Math. Stat. 20(3): 426-432.
- Hardy, G. H., Littlewood, J. E., and Polya, G. 1934. Inequalities. Cambridge University Press, Cambridge. pp. 143-145.
- Horvitz, D. G., and Thompson, D. J. 1951. A generalization of sampling without replacement from a finite universe. Abstract. Ann. Math. Stat. 22(2): 315.
- Horvitz, D. G., and Thompson, D. J. 1952. A generalization of sampling without replacement from a finite universe. Jour. Amer. Stat. Assoc. 47: 663-685.

- Isserlis, L. 1916. On the conditions under which the "probable errors" of frequency distributions have a real significance. Proc. Roy. Soc. Lond. A, 92: 23-41.
- Isserlis, L. 1918. On the value of a mean as calculated from a sample. Jour Roy. Stat. Soc. 81: 75-81.
- Lahiri, D. B. 1951. A method of sample selection providing unbiased ratio estimates. Bull. Inter. Stat. Inst. 33: 133-140.
- Midzuno, H. 1950. An outline of the theory of sampling systems. Ann. Inst. Stat. Math. (Japan) 1(2): 149-156.
- Midzuno, H. 1952. On the sampling system with probability proportionate to sum of sizes. Ann. Inst. Stat. Math. (Japan) 3(2): 99-107.
- Mortara, G. 1917*. Elementi di statistica. Appunti sulle lezioni di statistica metodologica dettate nel R. Istituto Superiore di studi commerciali di Roma. Rome. p. 356.
- Narain, R. D. 1951. On sampling without replacement with varying probabilities. Jour. Ind. Soc. Agric. Stat. 3(2): 169-174.
- Neyman, J. 1934. On two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. Jour. Roy. Stat. Soc. 97: 558-606.
- Neyman, J. 1952. Lectures and conferences on mathematical statistics and probability. U. S. Dept. Agric., Washington, D. C. pp. 227-228.
- Sen, A. R. 1952. Further developments of the theory and application of primary sampling units with special reference to the North Carolina agricultural population. Ph.D. thesis, North Carolina State College, Raleigh.
- Sen, A. R. 1953. On the estimate of the variance in sampling with varying probabilities. Jour. Ind. Soc. Agric. Stat. 5(2): 119-127.
- Sen, A. R., Anderson, R. L., and Finkner, A. L. 1954. A comparison of stratified two-stage sampling systems. Jour. Amer. Stat. Assoc. 49: 539-558.
- Singh, D. 1954. On efficiency of the sampling with varying probabilities without replacement. Jour. Ind. Soc. Agric. Stat. 6(1): 48-57.
- Splawa-Neyman, J. 1925. Contributions to the theory of small samples drawn from a finite population. Biometrika 17: 472-479.
- Sukhatme, P. V. 1953. Sampling theory of surveys with applications. The Indian Society of Agricultural Statistics, New Delhi, India, and The Iowa State College Press, Ames, Iowa. pp. 60-72.

- Sukhatme, P. V., and Narain, R. D. 1952. Sampling with replacement. Jour. Ind. Soc. Agric. Stat. 4(1): 42-49.
- Thompson, D. J. 1952. A theory of sampling finite universes with arbitrary probabilities. Unpublished Ph.D. thesis, Iowa State College, Ames, Iowa.
- Tschuprow, A. 1918*. Zur theorie der stabilität statistischer reihen. Skan. Akt. 1: 219.
- Tschuprow, A. 1923. On the mathematical expectation of the moments of frequency distributions in the case of correlated observations. Metron 2(3): 461-293; 2(4): 646-683.
- Yates, F., and Grundy, P. M. 1953. Selection without replacement from within strata with probability proportional to size. Jour. Roy. Stat. Soc., B, 15(2): 253-261.

Note: The two references in asterisk are, at present, not accessible to me. As mentioned in Chapter I, they were given by Tschuprow himself in his 1923 paper.

INSTITUTE OF STATISTICS
NORTH CAROLINA STATE COLLEGE

(Mimeo Series available for distribution at cost)

265. Eicker, Friedhelm. Consistency of parameter-estimates in a linear time-series model. October, 1960.
266. Eicker, Friedhelm. A necessary and sufficient condition for consistency of the LS estimates in linear regression. October, 1960.
267. Smith, W. L. On some general renewal theorems for nonidentically distributed variables. October, 1960.
268. Duncan, D. B. Bayes rules for a common multiple comparisons problem and related Student-t problems. November, 1960.
269. Bose, R. C. Theorems in the additive theory of numbers. November, 1960.
270. Cooper, Dale and D. D. Mason. Available soil moisture as a stochastic process. December, 1960.
271. Eicker, Friedhelm. Central limit theorem and consistency in linear regression. December, 1960.
272. Rigney, Jackson A. The cooperative organization in wildlife statistics. Presented at the 14th Annual Meeting, South-eastern Association of Game and Fish Commissioners, Biloxi, Mississippi, October 23-26, 1960. Published in Mimeo Series, January, 1961.
273. Schutzenberger, M. P. On the definition of a certain class of automata. January, 1961.
274. Roy, S. N. and J. N. Shrivastava. Inference on treatment effects and design of experiments in relation to such inferences. January, 1961.
275. Ray-Chaudhuri, D. K. An algorithm for a minimum cover of an abstract complex. February, 1961.
276. Lehman, E. H., Jr. and R. L. Anderson. Estimation of the scale parameter in the Weibull distribution using samples censored by time and by number of failures. March, 1961.
277. Hotelling, Harold. The behavior of some standard statistical tests under non-standard conditions. February, 1961.
278. Foata, Dominique. On the construction of Bose-Chaudhuri matrices with help of Abelian group characters. February, 1961.
279. Eicker, Friedhelm. Central limit theorem for sums over sets of random variables. February, 1961.
280. Bland, R. P. A minimum average risk solution for the problem of choosing the largest mean. March, 1961.
281. Williams, J. S., S. N. Roy and C. C. Cockerham. An evaluation of the worth of some selected indices. May, 1961.
282. Roy, S. N. and R. Gnanadesikan. Equality of two dispersion matrices against alternatives of intermediate specificity. April, 1961.
283. Schutzenberger, M. P. On the recurrence of patterns. April, 1961.
284. Bose, R. C. and I. M. Chakravarti. A coding problem arising in the transmission of numerical data. April, 1961.
285. Patel, M. S. Investigations on factorial designs. May, 1961.
286. Bishir, J. W. Two problems in the theory of stochastic branching processes. May, 1961.
287. Konsler, T. R. A quantitative analysis of the growth and regrowth of a forage crop. May, 1961.
288. Zaki, R. M. and R. L. Anderson. Applications of linear programming techniques to some problems of production planning over time. May, 1961.
289. Schutzenberger, M. P. A remark on finite transducers. June, 1961.
290. Schutzenberger, M. P. On the equation $a^{2+n} = b^{2+m}c^{2+p}$ in a free group. June, 1961.
291. Schutzenberger, M. P. On a special class of recurrent events. June, 1961.
292. Bhattacharya, P. K. Some properties of the least square estimator in regression analysis when the 'independent' variables are stochastic. June, 1961.
293. Murthy, V. K. On the general renewal process. June, 1961.
294. Ray-Chaudhuri, D. K. Application of geometry of quadrics of constructing PBIB designs. June, 1961.
295. Bose, R. C. Ternary error correcting codes and fractionally replicated designs. May, 1961.
296. Koop, J. C. Contributions to the general theory of sampling finite populations without replacement and with unequal probabilities. September, 1961.
297. Foradori, G. T. Some non-response sampling theory for two stage designs. Ph.D. Thesis. November, 1961.
298. Mallios, W. S. Some aspects of linear regression systems. Ph.D. Thesis. November, 1961.
299. Tauber, R. C. On sampling with replacement: an axiomatic approach. Ph.D. Thesis. November, 1961.
300. Gross, A. J. On the construction of burst error correcting codes. August, 1961.
301. Srivastava, J. N. Contribution to the construction and analysis of designs. August, 1961.
302. Hoeffding, Wassily. The strong laws of large numbers for u-statistics. August, 1961.
303. Roy, S. N. Some recent results in normal multivariate confidence bounds. August, 1961.
304. Roy, S. N. Some remarks on normal multivariate analysis of variance. August, 1961.
305. Smith, W. L. A necessary and sufficient condition for the convergence of the renewal density. August, 1961.
306. Smith, W. L. A note on characteristic functions which vanish identically in an interval. September, 1961.
307. Fukushima, Kozo. A comparison of sequential tests for the Poisson parameter. September, 1961.
308. Hall, W. J. Some sequential analogs of Stein's two-stage test. September, 1961.
309. Bhattacharya, P. K. Use of concomitant measurements in the design and analysis of experiments. November, 1961.