SOME THOUGHTS ON THE FOUNDATIONS OF SURVEY SAMPLING

J. N. K. Rao
University of Manitoba, Canada

Introduction

The development of sampling theory of surveys has progressed more or less inductively. Estimators which appeared to be reasonable have been proposed and their relative properties carefully studied by analytical and/or empirical methods, mainly through comparisons of biases and mean square errors. Similar methods have been employed in the selection of survey designs. Most of the text books on sampling reflect the above useful approach.

The need for basic deductive theories and for bringing inference in sample surveys more into the mainstream of statistical inference requires no emphasis. It is indeed gratifying that several leading statisticians, realizing this need, have made important contributions in recent years. It is my hope that before too long we will achieve this goal and thus make the study of sampling theory more exciting and useful.

In Section 2, we examine the contributions of Godambe and others towards a unified theory and demonstrate that their results on the unique choice of an estimator have doubtful value. We briefly consider some alternative approaches inSection 3, which have been put forth recently and which seem to lead to more satisfactory results. It must be emphasized, however, that no one theory is likely to provide completely satisfactory answers for all situations.

Throughout the paper we confine ourselves to estimation of population mean or total and neglect non-sampling errors. This does not imply, however, that non-sampling errors or estimation of other parameters are not important. The need for suitable methods to measure and control non-sampling errors and to handle analytic statistics from complex samples (for instance, regression and correlation coefficients, analysis of variance, contingency table analysis) is being increasingly felt and, in fact, estimation of totals or means is no longer of primary interest. Several ingeneous techniques, such as the 'jack-knife' and 'balanced half-sample replication,' have been proposed for analytic statistics (see Kish and Frankel [1970]) but much remains to be done in this area.

---

* Supported by a research grant from the National Research Council of Canada.

2. Godambe's Set-up

2.1. Formulation

A survey population U consists of a known number N of distinct units which are identified through a set of labels, say the integers 1, 2, . . ., N. The unit receiving label 'j' is denoted by $U_j (j=1, \ldots, N)$. Thus all finite populations of unlabeled units (with known or unknown N) are excluded. One reason usually given for the exclusion of such populations is that it is not possible to draw probability samples from unlabeled population. This reasoning, however, is not justifiable and we refer the reader to Hartley and Rao [1971] for details on the occurrence of populations with unlabeled units and drawing probability samples from such populations (see also Barnard [1969], p. 707-8).

With each unit U_j an unknown quantity y_j (possibly vector-valued) is associated which can be made *exactly* known by observing U_j. The unknown vector $\underline{\theta}=(y_1, \ldots, y_N)$ is called a parameter and $\underline{\theta}$ belongs to a well-defined set Ω (called the parameter space). A sample $s \in S$ is an ordered finite sequence $\{u_1, \ldots, u_{n(s)}\}$ of units from U, where u_i=some U_j, $i=1, \ldots, n(s)$ and the u_i's may not be distinct (*i.e.*, same unit may occur two or more times in the sample), $n(s)$ is the sample size and S denotes the collection of all possible samples s. A probability measure p (.) on S, according to which a sample s is chosen and observed[1] is defined and the pair (S, p), or simply $p(s)$, is called the sampling design. Any real-valued function $e(s, \underline{\theta})$ which depends on $\underline{\theta}$ only through those y_j for which U_j is in s is called an estimator. The purpose is to estimate the population total $Y=\Sigma\, y_j$ or the mean $\overline{Y}=Y/N$ with the help of an estimator.

As Basu [1969] has pointed out, the above formulation seems to imply the following :

(*a*) A sample space S with just one probability measure p (.) on S ;

(*b*) a statistic is a special kind of a function of s and the parameter $\underline{\theta}$.

This formulation is obviously confusing because a statistic should be a function defined on the sample space and there can be no statistical inference with a single probability measure. Basu cleared up this confusion by showing that a reformulation[2] leads to the conventional set-up $(X, A, P_{\underline{\theta}})$ where X is the sample space, A is the σ-field generated by subsets of X and $P_{\underline{\theta}}$ is a family of probability measures

1. We confine ourselves here to non-sequential sampling.
2. Pathak (see Basu [1969]' P. 453) and Hanurav ([1966], P. 196) have also given similar reformulations.

indexed by the parameter $\underline{\theta}$. The sample is $\underline{x}_s=(s, \underline{y}_s)$ where $\underline{y}_s=(y_1', \ldots, y_{n'(s)})$ and y_i' is the y-value of u_i, *i.e.* s *together* with the associated observation $\underline{y}_s$. For instance, if

$$s=(U_2, U_4, U_2) \text{ and } y_2=3, y_4=6 \text{ (say)},$$

then

$$\underline{y}_s=(3, 6, 3).$$

The sample space X is the set of all possible samples $\underline{x}_s$ and A is the σ-field generated by one-point sets of X. Let Ω_x denote the set of parameter points $\underline{\theta}$ which are consistent with a given sample $\underline{x}_s$ *i.e.*, for all $\theta \varepsilon \Omega_x$, the co-ordinates of $\underline{\theta}$ corresponding to the units in s are equal to the observed y-values for these units. Then the probability of observing $\underline{x}_s$ is giveen by

$$P_{\underline{\theta}}(\underline{x}_s)=\begin{cases}P(s) \text{ for } \underline{\theta}\ \varepsilon\ \Omega_x \\ 0 \quad \text{otherwise}\end{cases} \qquad \ldots(1)$$

Equation (1) also defines the likelihood of $\underline{\theta}$ given $\underline{x}_s$, denoted by $L(\underline{\theta} \mid \underline{x}_s)$.

There seems to be agreement among statisticians that the likelihood function plays a basic role in statistical inference. Let us, therefore, examine Godambe's likelihood (1). It is clear that (1) merely says that all sets of population values θ containing observed sample values are equally tenable, *i.e.*, the likelihood is completely uniformative on the unobserved y-values and, hence, on Y or $\overline{Y}$. The following quotation throws more light on the difficulties associated with Godambe's likelihood : "In situations like the one we are considering where the full likelihood does not satisfy our purpose, we may have to depend on a statistic which for every observed value supplies information (however poor it may be) on parameters of interest. In choosing a statistic for this purpose we may be guided by $I(\underline{\theta}, T)$, the information due to randomization. Unfortunately, no unique choice T which maximizes $I(\underline{\theta}, T)$ may be possible unless some further restrictions are placed on the class of statistics to be considered", C.R. Rao [1970] (see also Kempthorne [1969, p. 685] and Dempster [1968, p. 24] for similar remarks). It is clear, therefore, that no meaningful likelihood inference on $\underline{\theta}$ or Y is possible unless we are prepared to ignore some aspects of the sample $\underline{x}_s$ and thus make the sample non-unique. The decision on which aspects of the sample to be ignored depends on the situation at hand and, moreover, there is no unique way of going from the sample to the population (see Section 3).

Basu has rigorously proved that the minimal sufficient statistic is the set of distinct units s together with the associated y-values :

$\underline{x}_s^*=[s^*, \underline{y}^*_s]$ where s^* is the set of *distinct* units in s, say $(U_{j_1}, \ldots, U_{j_{v(s)}})$ where $j_1 < j_2 < \ldots < j_{v(s)}$, $v(s)$ is the number of distinct units in s (called the effective sample size) and $\underline{y}_s^* = (y_{j_1}, \ldots, y_{j_\nu}(s))$. However, $\underline{x}_s^*$ is not 'complete' so that infinitely many unbiased estimators of Y (or $\bar{Y}$) which are all functions of $\underline{x}_s^*$ exist. Moreover, for many of the sampling designs used in practice, $\underline{x}_s^*$ and $\underline{x}_s$ (without regard to order of selection) are equal and, even for the other designs, the minimal sufficient statistic is usually too 'wide'.

The possibility that for some sampling designs $\underline{x}_s^* \neq \underline{x}_s$ has lead to quite a few papers in which the main idea is as follows : Given an unbiased estimator $e(s, \underline{\theta})$ of Y, the estimator $e_1(s^*, \underline{\theta}) = E[e(s, \theta) \mid \underline{x}_s^*]$ is also unbiased and $V(e_1) \leqslant V(e)$ with strict inequality at least once by the Rao-Blackwell theorem. This idea has lead to useful estimators in some situations, for instance, unordering Des Raj's estimator gives Murthy's estimator which is as simple to compute as the former when the sample size $n = 2$. However, oftentimes the authors of some of these papers seem to be somewhat over-enthusiastic in this pursuit as little attention has been paid to the following facts : (1) Often e_1 may be computationally much more cumbersome than e ; for instance, in pps sampling with replacement ; (2) Corresponding sampling designs (with $\underline{x}_s = \underline{x}_s^*$), which are equally feasible and which lead to estimators more efficient than e_1 for the same expected cost, might exist ; for instance, the sample mean $\bar{y}$ in simple random sampling (*srs*) without replacement has smaller variance than Basu's estimator $e_1 = \bar{y}_{\nu(s)} = (y_{j_1} + \ldots + y_{j\nu(s)})/\nu(s)$ in srs with replacement for the same expected cost, assuming a linear cost function ; (3) Considerations other than efficiency might dictate the selection of a sampling design with $\underline{x}_s \neq \underline{x}_s^*$ and an estimator $e(s, \underline{\theta})$; for instance, in multistage sampling the primaries are often selected with pps with replacement and the customary estimator of Y based on *all* the primary draws is used mainly because an *unbiased* variance estimator is obtained simply from the estimators of the selected primary totals, provided subsampling is done independently each time a primary is selected.

2.2. Choice of Estimator

Following Horvitz and Thompson [1952], Godambe [1955] defined the 'most general' linear[3] estimator of Y as

$$e_b(s, \underline{\theta}) = \sum_{U_j \varepsilon s} b_{sj} y_j \qquad .. (2)$$

where the coefficients b_{sj} are defined in advance for all possible s and all $U_j \varepsilon s$. It is customary, however, to consider the class defined by

$$e_b(s^*, \underline{\theta}) = \sum_{U_j \varepsilon s^*} b_{s^*j}\, y_j \qquad .. (3)$$

instead of (2), because for each unbiased estimator of the form (2) there exists, by the Rao-Blackwell theorem, an unbiased estimator of type (3) which is at least as efficient as the former. Godambe proved the non-existence of a best (minimum variance) estimator in the class of linear unbiased estimators (2) (or 3) for any

3. Basu [1970] objected the usage of the term 'linear' because the sample space X is not linear and any estimator is a function on X, He proposed the term 'generalized linear' instead.

sampling design (excepting those in which no two s with $p(s) > 0$ have at least one common and one uncommon unit). In this connection, it may be appropriate to comment on Godambe's constant criticism of Neyman [1934], Cochran (1953), Sukhatme [1954] and others : "the fallacy of Neyman's [1934] argument" etcetra. Neyman considered simple random sampling [$n(s)=n$ for all s] and linear unbiased estimators of the form

$$\bar{e}_b = b_1 y_1' + \ldots + b_n y_n' \qquad \ldots (4)$$

where b_i is a preselected constant to be attached to y'_i. He proved (by appealing to the Gauss-Markov theorem) that the sample mean $\bar{y}$ is the best estimator in the above class (which is a sub-class of (2).). This positive result for a sub-class (however trivial it may be) in no way contradicts Godambe's negative result for the wider class (2). It is obvious that Neyman was not considering the class (2) (or 3) for which Godambe (rightly) claims priority ! Similarly, Cochran and Sukhatme have clearly stated their assumptions and defined the class of estimators they were considering while dealing with conditions under which ratio or regression estimator is optimum. It may be noted that Godambe's class of estimators cannot be implemented, unlike Neyman's, for sampling from populations with unlabeled units.

We now examine the criteria of optimality that have been proposed for the selection of estimator.

Admissibility. In a class of unbiased estimators of Y, an estimator t_1 belonging to the class is said to be admissible if for every other estimator t in the class $V(t_1)<V(t)$ for at least one $\theta \varepsilon \Omega$. In view of the minimal sufficiency of $\underline{x}_s^*$ it, therefore, follows that any admissible estimator is necessarily a function of $\underline{x}_s^*$ and that any estimator which is not a function of $\underline{x}_s^*$ is inadmissible (by the Rao-Blackwell theorem). However, as pointed out earlier in Section 2.1, one should not discard inadmissible estimators without objective examination.

Godambe proved that the Horvitz-Thompson ($H-T$) estimator of Y, given by $\hat{Y}_{HT} = \sum_{U_j \varepsilon s} y_j/\pi_j$, is admissible in the class of linear unbiased estimators (2) for any sampling design $p(s)$ with inclusion probabilities $\pi_j>0$ for all $U_j \varepsilon U$, provided $\Omega = R^N$ (the N-dimensional Eucledian space) or any interval around the origin of R^N (i.e., $-\alpha_j \leqslant y_j \leqslant \beta_j$ where $\alpha_j \leqslant 0$, $\beta_j > 0$). This result has been considerably generalized by V.M. Joshi in a series of papers, by relaxing linearity and/or unbiasedness. Joshi has also shown the admissibility of a number of other estimators (including Lahiri's estimator, regression estimator, sample mean) and, I understand, he has recently proved the admissibility of Murthy's and related estimators. One might be tempted, therefore, to conjecture that any linear unbiased estimator of Y which is a non-zero function of x_s^* is admissible, but this is not true (Dharmadhikari [1969], Joshi [1969]) ; however, the counter examples proposed are artificial. In any case, the criterion of admissibility does not appear to be sufficiently selective for distinguishing between the merits of estimators.

Since the criterion of admissibility has not been conclusive, several new criteria, which give rise to a *unique* choice of estimator, have been put forth in recent years. We now examine some of these criteria in turn.

Necessary bestness and hyper-admissibility. The criterion of necessary bestness boils down to the following : an estimator t_1 belonging to a class of unbiased estimators of Y is said to be the necessary best estimator if for every other unbiased estimator in that class $V(t_1) \leqslant V(t)$ in *each* of the N principal hyper-surfaces (phs) of dimension one, viz., $(y_1, 0,...,0),..., (0, 0,...,0, y_N)$ with strict inequality for at least one phs of dimension one. It is easily seen that the H-T estimator, $\hat{Y}_{HT}$, is the necessary best estimator in the linear class (2) for any sampling design with $\pi_i > 0$ for all $U_i \in U$. The criterion of hyper-admissibility, on the other hand, is defined as follows : an estimator t_1 in a class of unbiased estimators of Y is hyper-admissible in that class if it is admissible inside *each* of the $2^N - 1$ possible phs's in R^N. It is evident that the H-T estimator is the unique hyper-admissible estimator in class (2) because $\hat{Y}_{HT}$ is hyper-admissible and, from the necessary bestness of $\hat{Y}_{HT}$, every other unbiased estimator in (2) is inadmissible in at least one of the phs's of dimension one.

As a justification for hyper-admissibility, Hanurav (1968) stated that, in practice, often one is interested in estimating not only Y but also the totals of sub-populations or domains and that these domains should all be admissibly estimable by a single estimator. I don't think anyone questions the importance of domain estimation, but it is obviously unrealistic to consider *all* the $2^N - 1$ possible phs's (domains) because the number of domains of practical interest will be very much smaller than $2^N - 1$ and the sizes of such domains will be large, although the actual number of units in a domain is unknown. The domains of size one (which are utterly uninteresting in practice) play the key role in arriving at the unique choice. If we exclude these domains, the uniqueness result may no longer hold and, in fact, even the 'necessary best estimator of second order' (using only phs's of dimension two) does not exit[4].

Further evidence on the weakness of hyper-admissibility is provided by the result that the Horvitz-Thompson estimator of variance of $\hat{Y}_{HT}$ say $v(\hat{Y}_{HT})$, which is known to have some undesirable properties (including taking negative values often), is *uniquely* hyper-admissible Rao and Singh [1969] in a general class of quadratic unbiased estimators of $V(\hat{Y}_{HT})$. It may be noted that $V(\hat{Y}_{HT})$ is well-behaved always non.negative in phs of dimension one .

4. Prabhu-Ajgaonkar's [1969] result that $\hat{Y}_{HT}$ is also the necessary estimator of second order is incorrect.

If we take these criteria seriously, we should be using $\hat{Y}_{HT}$ for any sampling design irrespective of whether there is any positive correlation between the y_j and π_j or not. It is obvious, however, that $\hat{Y}_{HT}$ could lead to nonsensical results when y_j and π_j are unrelated or poorly correlated—see Rao [1966] for some practical situations and Basu [1970] for a delightful 'circus' example. In such situations one should, of course, employ alternative estimators such as the sample mean even if they are biased (Rao [1966]). Horvitz and Thompson [1952] proposed $\hat{Y}_{HT}$ only for those situations where y_j and π_j are strongly related (positively).

Linear sufficiency. Barnard [1963] introduced the concept of linear sufficiency in least squares which reduced to the following definition in the case of estimating a scalar parameter μ ; a statistic t_1 is linear sufficient for μ if for every other statistic t_2 with $cov(t_1, t_2)=0$ we have $E(t_2)=0$. Clearly, this definition is applicable to our situation without any change, even when t_1 and/or t_2 are non-linear. Godambe [1966], however, modified it, without any reason, by changing "$cov(t_1, t_2)=0$" to the following (see his definition 4.2) : two estimators $e_b(s, \theta) = \Sigma b_{sj} y_j$ and $e_b'(s, \underline{\theta}) = \Sigma b'_{sj} y_j$ belonging to class (2) are said to be independent of each other if $\Sigma b_{sj} b'_{sj}=0$ for every $s \varepsilon S$ with $p(s)>0$. The two definitions are obviously different. Using his definition, Godambe proved that, for any fixed sample size design ($n(s)=n$ for all s), the *only* unbiased estimator of Y in class (2) which is linear sufficient and which satisfies the 'principle of censoring' (viz., inference should not depend on the probabilities of the undrawn s) is given by

$$e_b=\left[\binom{N-1}{n-1} p(s)\right]^{-1} \sum_{U_j \varepsilon s} y_i \qquad \ldots(5)$$

On the other hand, using Barnard's original definition, M. K. Ramakrishnan (a former student of the author) has shown that $\hat{Y}_{HT}$ is an unbiased linear sufficient estimator of Y in class (2), but *not unique*. I do not know how appropriate the concept of linear sufficiency is in survey sampling, but I see no reason for changing the original definition.

Godambe [1966] proposed the concept of distribution-free sufficiency to handle non-linear estimators $e(s, \theta)$, by defining the independence between two estimators $e_1(s, \underline{\theta})$ and $e_2(s, \underline{\theta})$ as $E_\xi(e_1 e_2 | s) = E_\xi(e_1 | s) E_\xi(e_2 | s)$ for every $s \varepsilon S$ where $E(\,.\,| s)$ denotes the expectation when s is fixed and ξ is a priori distribution for $\underline{\theta}$ which assumes that the unknown variate values $y_1, \ldots, y_N$ are *a priori* independent. He proved that $e_{\bar{b}}$ is an unbiased distribution-free sufficien estimator. However, the class C of priors ξ is unreasonable because it implies that one's prior information on y_1 (say) would be precisely the same whether $y_2, \ldots, y_N$ were known or unknown. Even if one is willing to accept Godambe's definition, his result has little value as Kale [1967] has shown that *any* linear estimator belonging to class (3) is distribution-free sufficient.

Bayesian sufficiency. Let $h(\underline{\theta})$ denote some prior distribution on $\underline{\theta}$, then, using the likelihood (1), the posterior distribution of $\underline{\theta}$ given $\underline{x_s}$ is

$$h(\underline{\theta} \mid x_s) \begin{cases} \underline{h(\theta)} & \text{for } \theta \in \Omega_x \\ 0 & \text{otherwise,} \end{cases} \tag{6}$$

i.e., the posterior distribution is nothing but the restriction of prior to the set Ω_x. A statistics $t(\underline{xs})$ is Bayes sufficient for y if the posterior distribution depends on x_s only through $t(\underline{xs})$. This definition, of course, is simply a Bayesian version of the classical definition of sufficiency. Assuming the class C of priors ξ, Godambe showed that the statistic.

$$(s^*, \underset{U_j \varepsilon s}{\Sigma\Omega y_j})$$

is Bayes sufficient for Y and then imposing origin and scale invariance, he arrived at the following result : $N\bar{y}_{v(s)}$ in *uniquely* the Bayes sufficient, origin and scale invariant estimator of Y for any sampling design. He has also considered the case where supplementary information $z_1, \ldots\ldots, z_N$ (attached to the N units in U) is available and, assuming the class C of priors ξ conditional on $z_1, \ldots\ldots\ldots\ldots, z_N$, he proved that for any sampling design the ratio estimator $(\bar{y}_{v(s)}/\bar{z}_{v(s)})\,Z$ is *uniquely* the Bayes sufficient, scale invariant estimator and the difference estimator $N(\bar{y}_{v(s)} - \bar{z}_{v(s)}) + Z$ is *uniquely* the Bayes sufficient origin invariant estimator, where Z is the population total of z_j's. These impressive results, however, heavily depend on the choice of prior and, as mentioned before, the class C of priors ξ is unreasonable. To illustrate this, we consider two reasonable classes of priors which lead to different results. First, when labels j are not expected to carry information regarding the associated y_j's, it may be reasonable to assume an exchangeable prior on $\underline{\theta}$ (*i,e.*, a prior distribution symmetric in the co-ordinates y_j) ; then the posterior distribution of $\underline{\theta}$ given $\underline{x_s}$ does not depend on the label-set s but only on $\underline{y_s}^* = \{y_j : U_j \varepsilon s\}$ and, in this case, $\{y_j : U_j \varepsilon s\}$ is Bayes sufficient for Y (Solmon and Zacks [1970]). A broad class of exchangeable priors C_1 can be generated by assuming that the y_j are probabilistically independent and indentically distributed conditional on some parameter $\underline{\mu}$ and then averaging over the marginal (prior) distribution of $\underline{\mu}$ (Ericson [1969]). Similarly, if supplementary information $(z_1, \ldots, z_N)$ is available, one could get a broad class of priors C_2 assuming that the y_j are probabilistically independent conditional on $z_1, \ldots, z_N$ *and* some parameter $\underline{\mu}$ and then averaging over the marginal (prior) distribution of $\underline{\mu}$. For instance, the often-used super-population model

$$E(y_j \mid z_j)=\beta z_j,\ V(y_j \mid z_j)=a z_j^g,\ g \geqslant 0$$
$$cov\ (y_j, y_k \mid z_j, z_k)=0, j \neq k=1, \ldots, N \qquad \ldots(7)$$

where $\underline{\mu}'=(\beta, a, g)$ is unknown, would lead to a prior for $\underline{\theta}$ which belongs to the above class, by assuming a prior on $\underline{\mu}$ and normality for the y_j (conditional on the z_j).

Since the sampling design $p(s)$ does not enter into the definition of Ω_x, it is clear from (6) that Bayesian analysis does not depend on the way the data has been collected. The 'likelihood principal' (which says that the information supplied by x_s is the likelihood function) when applied to (1) also leads to the same conclusion. This revolutionary conclusion is in direct conflict with current practice. The alternative approaches in Section 3 throw additional light on this dilemma, but I don't think that the proponents of the likelihood principle intended that it should ever be applied to likelihoods not involving the parameter(s) of interest.

Fiducial estimation. Since the posterior of $\underline{\theta}$ is independent of the sampling design $p(s)$, Godambe [1969] introduced fiducial estimation into survey sampling and demonstrated that, under certain conditions, the fiducial distribution depends on $p(s)$, for simple random sampling without replacement. Barnard [1969, p. 709], however, says, "...I must admit I find it difficult to see how Dr. Godambe's conditions could arise in practice". Zacks [1970] argued that fiducial distributions should be derived within a general group invariance framework and proved that no fiducial estimator of Y exists, under the group of real translations and squared-error loss.

2.3. Choice of design

Suppose e is an estimator of Y associated with a sampling design d, then $H=(e, d)$ is called a strategy. Let $L(H)$ denote a class of equi-cost strategies involving designs d_1 and d_2 and unbiased estimators e belonging to class (2) (or 3). The design d_1 is said to be *better* than d_2 if for every strategy $H_2=(e_2, d_2) \in L(H)$, there exists a strategy $H_1=(e_1, d_1) \in L(H)$ such that $V(e_1 \mid d_1) \leqslant V(e_2 \mid d_2)$ for all $\underline{\theta} \in \Omega$ with strict inequality at least once. Using this definition, it is unlikely that one could even show that *srs* without replacement is *better* than *srs* with replacement. It is, therefore, sensible to use a practical approach by comparing the relative efficiencies of 'good' estimators in d_1 and d_2 for the same expected cost. This approach could lead to clear-cut answers for some simple designs (*e.g.*, Seth and Rao [1964]), but, more often than not, it would be necessary to carry out extensive empirical or semi-empirical studies on natural populations with a view to spotting out reasonably good strategies under different situations commonly met with in practice (Rao and Bayless [1969]). Other meaningful criteria, such as efficiency of variance estimator, could throw further light on the choice of strategy.

Using the super-population model (7) with $g=2$ and the criterion of average variance, Godambe [1955] has shown that the strategy satisfying $v_d(s)=v_0$ for all s, $\pi_i(d)=v_0 z_i/Z$ and $e=\hat{Y}_{HT}(d)$ is 'optimum' in $L(H)$, where v_0 is the expected sample size for a given budget and the average is over all finite populations that can be drawn from the infinite super-population. This result, however, has limited scope because no 'optimum' strategy exists for $g\neq 2$. Moreover, assuming (7) with $g=2$, Rao and Bayless [1969] have shown that the loss in average efficiency of Murthy's strategy over the optimum strategy is negligible for a wide variety of natural z-populations.

Recently, Basu [1970] and Royall [1970] advocated strategies involving purposive selection of units, the former using Bayesian argument and the latter employing the super-population model (7) with known g and the criterion of average mean square error. We plan to examine these results in a subsequent paper.

3. Alternative Approaches

We now briefly consider some alternative approaches; the reader is referred to the relevant papers for details.

srs without replacement. In the Hartley-Rao [1968, 1969] approach, it is assumed that the character y is measured on a *known* scale with a finite set of scale points $\{y_1^*, \ldots, y_T^*\}$ where T can be arbitrarily large. This, of course, simply corresponds to the realities of practice. The population mean $\bar{Y}$ may be written as $\bar{Y} = \sum_{t=1}^{T} (N_t/N)\, y_t^*$ where N_t=number of units U_j in the population with $y_j=y_t^*$, $(\Sigma N_t=N)$. Let n_t denote the number of units having the value y_t^* in the sample of fixed size n, so that $n_t\geqslant 0$ and $\Sigma n_t=n$. If we ignore the label-set s^* from the sample x_s^* and consider the likelihood based on $y_s^*=(n_1, \ldots, n_T)$, we get

$$P\left(n_1, \ldots\, n_T \,\middle|\, N_1 \ldots\, N_T\right)=\frac{\prod \binom{N_t}{n_t}}{\binom{N}{n}} \qquad \ldots (8)$$

which depends on *all* the parameters of interest, N_t. The loss of information due to ignoring s^* may be regarded as negligible if there is no evidence of any relationship between the labels j and corresponding y_j; there is no loss of information if the N units are labeled at random (see Hartley and Rao [1971]). C. R. Rao [1970] has given another justification for ignoring s^*. In the present situation of *srs*, it seems appropriate to ignore s^*, but it should be noted that there is no unique way of going from sample to population.

Since (8) is the multi-dimensional hypergeometric distribution, the statistic $(n_1, \ldots, n_T)$ is sufficient and 'complete'. Consequently, the sample mean $\bar{y} = \Sigma(n_t/n)y_t^*$ is the U. M. V. unbiased estimator of $\bar{Y}$ in the class of estimators which depend only on $(n_1, \ldots, n_T)$. Royall [1968] has obtained a similar result.

Maximization of the likelihood (8) subject to

$$N_t \geqslant 0, \; \Sigma N_t = N \qquad \ldots(9)$$

leads to the maximum likelihood estimators (m.l.e.) $\hat{N}_t = N(n_t/n)$ and $\hat{\bar{Y}} = \Sigma(\hat{N}_t/N)y_t^* = \bar{y}$, provided N/n is an inteteger ; when N/n is not integral, the m.l.e. of N_t's will be found to be rounded up and down version of $\hat{N}_t$'s given above and $\bar{y}$ will no longer be the m.l.e. of $\bar{y}$. One might quite rightly say that maximization of (8) subject to (9) is not very realistic as additional information on the N_t may be available, especially when T is small and the N_t are parameters of interest. Conceptually there is no problem, however, because one could maximize (8) subject to (9) and any additional restrictions on the N_t ; however, the actual solution might be quite formidable.

Assuming a compound-multinomial prior for the N_t, Hartley and Rao obtained the posterior distribution of the N_t using the likelihood (8) ; we refer the reader to Hartley and Rao [1968, 1969] for details on the Bayesian estimation of $\bar{Y}$. Ericson [1969] assumed an exchangeable prior for $(y_1, \ldots, y_N)$ in whichcase the prior for $(N_1, \ldots, N_T)$ is compound-multinomial and the conditional distribution $P(n_1, \ldots, n_T / N_1, \ldots, N_T)$ is given by (8) for any sampling design $p(s)$. Consequently, the posterior distribution of the N_t obtained by Hartley and Rao is the same as that of Ericson; however, note that the two approaches are fundamentally different (Solomon and Zacks [1970], p. 659).

Kempthorne [1969] has given another justification for the sample mean $\bar{y}$ in *srs* without replacement. He has shown that $\bar{y}$ has minimum average variance, for permutations of values attached to the units, in Godambe's class of unbiased estimators (3). Minimization of average variance over the $N!$ permutations of $(y_1, \ldots, y_N)$ makes sense because in the present situation one could suppose that the unknown set of N y-values are associated with the labels of the units in an unknown way.

***srs* with replacement.** A simple random sample s of fixed size n is drawn with replacement and let $v(s)$ and $v_t(s)$ respectively denote the number of distinct units in s and the number of distinct units having the value y_t^* in s, $(\Sigma v_t(s) = v(s))$

If we ignore the label-set s^* and consider the likelihood based on $y_s^* = [v_1(s), \ldots, v_T(s)]$, the m.l.e. of $\overline{Y}$ is identical to Basu's estimator $\bar{y}_{v(s)} = \Sigma[v_t(s)/v(s)]y_t^*$ provided $N/v(s)$ is integral ; the m.l.e of N_t's will be rounded up and down version of the $N_t = N[v_t(s)]$ when $N/v(s)$ is not integral. C. R. Rao [1970] has given a slightly different argument to justify Basu's estimator.

It is important to note that the identification labels are used at the identical state to get $\bar{y}_s^*$,but the label-set s^* is ignored at the estimation stage. Similarly, if a concommitant variate z with a finite set of scale-points $\{z_1^*, \ldots, z1^*\}$ is attached to the units, the labels are used to arrive at the sample values n_{it}=number of units in the sample which have z_i^* and y_t^* attached to them ($i=1, \ldots, I$; $t=1, \ldots, T$, $\Sigma\Sigma n_{it}=n$), but the label-set s^* is ignored at the estimation stage. Assuming that *only* the population mean $\overline{Z}$ of the z_j's is known, Hartley and Rao [1968] have shown that the likelihood based on the n_{it} leads to an m.l.e. of $\overline{Y}$ which is closely related to the customary regression estimator for large n and $N>>n$.

Unequal probability sampling (uni-stage designs). C.R. Rao [1970] used Kempthorne's approach to justify the H-T estimator in those situations where the y_j are approximately proportional to the corresponding π_j (for instance, when the z_j are approximately proportional to the y_j and π_j is chosen proportional to z_j). He considered the set of parameter values $(\pi_1 w_{i1}, \ldots, \pi_N w_{iN})$ by obtainedpermutations of $w_1, \ldots, w_N$ holding $\pi_1, \ldots, \pi_N$ fixed, where $w_j = y_j/\pi_j$. This obviously makes sense because when $y_j \propto \pi_j$ (approximately), w_j's will be unrelated to the π_j—a situation similar to *srs*. If we suspect that y_j may be related to $\pi_j^{\frac{1}{2}}$ (say) or unrelated to π_j (which could happen in p.p.s. sampling with multiple characters, Rao [1966]), the average is taken over a different set of parameter values appropriate to the situation at hand ; for instance, when y_j is expected to be unrelated or poorly correlated to π_j, it would make sense to take $w_j = y_j$ and consider the set $(\pi_1 y_{i1}, \ldots, \pi_N y_{iN})$.

Hartley and Rao [1969] have made a start with maximum likelihood estimation for unequal probability sampling, using their approach and a parametrization, similar to C. R. Rao's, appropriate to the situation at hand. They considered only p.p.s. sampling with replacement in their 1969 paper, but further work is in progress.

Stratified sampling. Strata are regarded as separate populations, each described by its separate set of parameters, *i.e.*, and additional subscript, say h, is used to index the strata and these strata labels h are regarded as informative because of strata differences. Extension of the above results to stratified sampling is, therefore, straightforward.

Hartley and Rao [1969] briefly considered two-stage sampling in which the primary labels may be regarded as informative. They indicated some difficulties and the need for further work.

Parametric models. The above approaches are essentially distribution-free. Barnard [1969] suggested that we should go a step further by assuming super-population models of the type (7) with specific distributional assumptions on the y_i and then applying standard likelihood procedures for comparing alternative estimators. Kalbfleisch and Sprott [1969] have used this approach in some simple situations. It would be difficult to implement Barnard's suggestion in large-scale surveys, but we should exploit it at least in smaller surveys of specialized scope.

4. Summary

The paper presents a critical discussion on the contributions of Godambe and others towards a unified theory of sampling from finite populations vis-a-vis some alternative approaches which have been put forth recently by several other researchers.

REFERENCES

Barnard G. A. [1963]. "The logic of least squares." *J. Roy. Statist. Soc. B, 25*, 124-7.

Barnard, G. A. [1969]. "Summary remarks", *New Developments in Survey Sampling :* Wiley-Inter Science, New York.

Basu, D. [1970]. "Role of the sufficiency and likelihood principles in sample survey theory". *Sankhya A, 51*, 441-54.

Basu, D. [1969]. "An essay on the logical foundations of survey sampling, part one". *Symposium on Foundations of Statistical Inference*, Waterloo, Canada.

Cochran, W. G. [1953]. *Sampling Techniques*: Wiley, New York. Dempster, 'A. P. [1968]. "A generalization of Bayesian inference (with discussion)". *J. Roy Statist. Soc. B, 30*, 205-47.

Dharmadhikari, S. W. [1969]. "A note on some inadmissible estimators". *Sankhya A, 31*, 361-4.

Ericson, W. A. [1969]. "Subjective Bayesian models in sampling finite populations". *New Developments in Survey Sampling :* Wiley-Interscience, New York.

Godambe, V. P. [1955]. "A unified theory of sampling from finite populaticns." *J. Roy. Statist. Soc., B, 17*, 268-78.

Godambe, V. P. [1966]. "A new approach to sampling finite populations, I & II" *J. Roy. Statist. Soc., B, 28*, 310-28.

Godambe, V. P. [I968]. "Bayesian sufficiency in survey sampling". *Ann. Inst. Statist. Math., 20*, 363-73.

Godambe, V. P. [1969]. "Some aspects of the theoretical developments in survey sampling". *New Developments in Survey Sampling :* Willey Interscience, New York.

Hanurav, T. V. [1966]. "Some aspects of unified sampling theory". *Sankhya A, 28*, 175-204.

Hánurav, T. V. [1968]. "Hyper-admissibility and optimum estimators for sampling finite populations". *Ann. Math. Statist., 39*, 621-42.

Hartley, H. O. and Rao, J. N. K. [1968]. "A new estimation theory for sample surveys, *Biometrika, 55*, 547-57.

Hartley, H. O. and Rao, J. N. K. [1969]. "A new estimaticn theory for sample surveys, II". *New Developments in Survey Sampling :* Wiley-Interscience, New York.

Hartley, H. O. and Rao, J. N. K. [1971]. "Foundations of survey sampling (a Don Quixote tradegy)". *American Statstician* (in press).

Horvitz, D. G. and Thompson, D. J. [1952]. "A generalization of sampling without replacement from a finite universe". J. *Amer. Statist. Assoc., 47*, 663-85.

Joshi, V. M. [1969]. "Admissibility of estimates of the mean of a finite population". *New Developments in Survey Sampling :* Wiley-InterScience, New York.

Kalbfleisch, J. D. and Sprott, D. A. [1969]. "Applications of likelihood and fiducial probability to sampling finite populations" *New Developments in Survey Sampling :* Wilev-InterScience, New York.

Kale, B. K. [1968]. "Distribution-free sufficiency in sampling from finite populations". *Ann. Math. Statist.* (abstract), *38*, 1935.

Kempthorne, O [1969]. "Some remarks on statistical inference in finite sampling". *New Developments in Survey Sampling :* Wiley-InterScience, New York.

Kish, L. and Frankel, M. R. [1970]. "Balanced repeated replications for standard error". *J. Amer. Statist. Assoc., 65*, 1071-94.

Neyman, J. [1934]. "On the two different aspects of the representative method : the method of stratified sampling and the method of purposive selection". *J. Roy. Statist. Soc., 97*, 558-606.

Prabhu-Ajgaonkar, S. G. [1969]. ,'On the non-existence of a best estimator for the entire class of linear estimators". *Sankhya A, 31*, 455-62.

Rao, C. R. [1970]. Some aspects of statistical inference in problems of sampling from finite populations. *Foundations of Statistical Inference :* A Symposium Edited by V. P. Godambe and D. A. Sprott, 177-202, *Holt, Reinhardt and W:nston,* Toranto, CANADA

Rao, J. N. K. [1966]. "Alternative estimators in pps. sampling for multiple characteristics." *Sankhya A, 28*, 47-60.

Rao, J. N. K. and Bayless, D. L. [1969]. "An empirical study of the stabilities of estimators and variance estimators in p.p.s. sampling". *J. Amer. Statist. Assoc., 64*, 540-59.

Rao, J. N. K. and Singh, M. P. [1969]. "On the criterion of 'hyper-admissibility' and 'necessary bestness' in survey sampling". Tech. Report No. 40, Sept. 12, 1969, R. T. S , Indian Statistical Institute.

Royall, R. M. [1968]. "An old approach to finite population sampling theory". *J. Amer. Statist, Assoc., 63*, 1269-79.

Royall, R. M. [1970]. "On finite population sampling theory under certain linear regression models". *Biometrika, 57*, 377-87.

Seth, G. R. and Rao, J. N. K. [1964]. "On the comparison of simple random sampling with-and without replacement". *Sankhya A, 26*, 85-6.

Solomon, H and Zacks, S. [1970]. "Optimal design of sampling from finite populations : a critical review and indication of new research areas". *J. Amer. Statist. Assoc., 65*, 653-677.

Sukhatme, P V. [1954]. *Sampling Theory of Surveys with Applications :* Indian Society of Agricultural Statistics, New Delhi.

Zacks, S. [1970]. "On fiducial estimation of parameters of finite populations". Technical Report No. 191, University of New Mexico.