Once such an iterative process for an ordinary life table has been developed, the life table for a single cause [what Hoem (1969) calls the partial rate] can be obtained by inserting the observed age-cause-specific rate rather than the age-specific rate. For the multiple decrement table we again use the same computation, but insert the observed ratio of the given cause to all causes. Thus with very slight modification the same argument—indeed the same program—that produces the ordinary life table produces a table for an individual cause acting alone, and produces the multiple decrement table.

The present approach for a single or multiple decrement table that agrees with the data leads immediately to nuptiality, fertility, school attendance, labor force, and other tables. More difficult are methods for a combined table (needed, for example, in the study of fertility by birth order) and for treating the life table stochastically.

ACKNOWLEDGMENTS

I am grateful for advice and other help to Francisco Bayo, Elizabeth Eisenberg, Jair Ferreira-Santos, Griffith Feeney, Wilhelm Flieger, Jan Hoem, Samuel Preston, Robert Retherford, and Robert Schoen. The research has been supported by NSF grant GZ 995, by NIH research contract 69-2200, and by teaching grants to the Department of Demography, University of California at Berkeley, from the National Institute of General Medical Sciences (5T01 GM01240) and the Ford Foundation.

REFERENCES

Chiang, C. L. Introduction to Stochastic Processes in Biostatistics. New York: John Wiley & Sons, 1968.

Greville, T. N. E. "Short methods of constructing abridged life tables," *Record of the American Institute of Actuaries* XXXIII, Part I (1943), 29–42.

———. *Methodology of the National, Regional, and State Life Tables for the United States, 1959–61.* United States Department of Health, Education and Welfare: Public Health Service Publications No. 1252, Vol. I, No. 4 (1967).

Hoem, J. M. "Purged and partial Markov chains." (To appear in the *Skandinavisk Aktuarietidskrift*, 1969).

Jordan, C. W., Jr. Life Contingencies. Chicago: The Society of Actuaries, 1952.

Keyfitz, N. Introduction to the Mathematics of Population. Reading, Mass.: Addison-Wesley Publishing Co., 1968a.

———. "A life table that agrees with the data: II," *Journal of the American Statistical Association*, (1968), 1253–1268.

Reed, L. J. and M. Merrell, "A short method for constructing an abridged life table," *American Journal of Hygiene*, XXX (1939), 33–62.

Spiegelman, M. Introduction to Demography. Revised edition. Cambridge: Harvard University Press, 1968.

U. S. Department of Health, Education, and Welfare. *United States Life Tables: 1959–61.* Public Health Service Publication No. 1252, Vol. I, No. 1 (1964).

———. *Life Tables.* Vital Statistics of the United States, 1967, Vol. II, Section 5 (1969).

Foundations of Survey-Sampling*

V. P. GODAMBE
University of Waterloo

Dedicated to the late Professor J. B. S. Haldane who brought to my attention the following very significant story from the ancient Indian epic Mahabharat (Nala-Damayanti Akhyān): The king lost his way in a jungle and was required to spend the night in a tree. The next day he told some fellow traveller that the total number of leaves on the tree were "so many". On being challenged as to whether he counted *all* the leaves he replied; "No, but I counted leaves on a few branches of the tree and I know the science of die throwing". (I can vouch for accurateness of the reproduction only in the essential respects.)

1. During April 1968, a symposium on Foundations of Survey-sampling was organized, at my suggestion, by the University of North Carolina, at Chapel Hill. Eminent statisticians from all over the world participated in this symposium. The discussions were exciting, as well as illuminating. Unfortunately, it was impossible for the organizers, because of the limitation of space, to include these discussions in the publication of proceedings of the symposium which has just come out (Johnson and Smith, 1969). Hence, I am giving below some details concerning the 'central issue' which highlighted the discussions at the symposium. In doing so, it is natural if I emphasize more my viewpoint; yet I have presented the other side of the issue as clearly as possible. This 'central issue', I think, can be best understood with some historical perspective:

2. Several years ago (1955), I proved a somewhat puzzling result which was subsequently generalized by Joshi and myself (1965). A simple illustration of the puzzling result is as follows: Let a finite population consist of N individuals which are labelled by integers $i = 1, \cdots, N$. Each individual i has a variate value x_i, $(i = 1, \cdots, N)$, associated with it. The variate values x_i, $(i = 1, \cdots, N)$, are unknown. Hence to estimate the population mean,

$$\bar{x}_N = \sum_1^N x_i/N, \tag{1}$$

a sample (denoted by 's'), of n individuals is drawn by 'simple random sampling without replacement' and the variate values x_i for the individuals i included in the sample s are observed. Now due to a result of Halmos (1947) and Watson (1964), the sample mean,

$$\bar{x}_s = \sum_{i \epsilon s} x_i/n, \tag{2}$$

($\sum_{i \in s}$ meaning summation over all individuals i included in the samples) is the unbiased minimum variance, (UMV), estimator of the population mean (1): That this is true if and only if the individual *labels* i

* This article was prepared at the invitation of the Editor.

are *ignored*, follows from the result referred to at the beginning of this paragraph. I say 'ignored' because the process of drawing 'statistically' a random sample from a population *consisting* of a fixed number (finiteness is irrelevant) of individuals involves use of some random number tables which essentially implies that all the individuals of the population are already labelled in a manner *known* to the sampler. At a suggestion of the referee of this paper I elaborate on this point further: Without going into the details of the meaning of the word 'probability' I can say that with the *given* knowledge, we tend to believe that 'all events (say N in in number) that can possibly occur under a certain *phenomenon* occur with equal probabilities ($= 1/N$). The search for such a 'phenomenon' is clearly basic for the construction of random number tables. Fundamentally, then drawing an individual at random from a population of N individuals is equivalent to arranging a one-to-one correspondence of the N events of the above referred to 'phenomenon' and the N individuals of the population: the individual corresponding to the event that actually occurs is said to be selected at random. Thus statistical random sampling[1] presupposes a labelling of the individuals in the population. This knowledge of individual labels, as shown in the above papers, (1955, 1965), implies a general non-existence of UMV estimation. Indeed this general non-existence of UMV estimation is typically true for what is commonly called a 'survey-population' since it consists of a *fixed* number of individuals, and for practically *all* modes of randomization such as simple random sampling, stratified sampling, sampling with arbitrary probabilities and the like. Why then was UMV estimation emphasized,[2] especially in relation to the sample mean (2), in the early literature on survey-sampling, more specifically, since Neyman's (1934) paper? The answer to this question is to be found in the origin and the development of 'general statistical theory'.

3. It is generally believed that statistics or statistical theory originated with the investigations of biological and sociological phenomena, such as inheritance and the like, during the last century. Soon the statistical theory was directed towards studying the relationships between different factors that influenced those phenomena and toward studying the underlying chance mechanisms. For instance, the sizes of a dozen observed human skulls were supposed to have been produced by some kind of chance mechanism operating in the background. It was then but a small step to replace the chance mechanism by a hypothetical population generated by the independent repetitions of the chance mechanism. Thus, the dozen observed skulls were supposed to be a *random sample* from some hypothetical population. It was for the statistician to make inference about the properties of this hypothetical population on the basis of the given *random sample* with the help of the tools provided by the classical theory of probability. A basic assumption made here was that the chance mechanism uniquely determined the frequency function of some characteristic under study in the hypothetical population. For example, in the above referred to hypothetical population of human skulls, the sizes of the skull might have a frequency function which is normal with a specified mean and variance. Usually, however, this frequency function was known only partially to the statistician, and the inference from the given random sample about the hypothetical population and the underlying chance mechanism often meant inference about this partially known frequency function. One might, for instance, try to estimate some unknown parameter of the otherwise (assumed to be) known frequency function in the hypothetical population. If the frequency function was known to be normal with unknown mean and variance, on the basis of the given random sample, one would try to estimate the mean and the variance. This, one may say, has over the years resulted in the development of a mathematical theory of statistics or briefly statistical theory. Fisher often stressed in his writings (1956, 1959) the hypothetical nature of the populations dealt with in the *general statistical theory*. The earlier authors of statistical theory like Galton (1889) and Pearson (1920) were less clear on the subject. But even a cursory glance at the development of the subject suffices to show that Fisher's explicit postulation of the hypothetical population was the natural crystalization of what was underlying the earlier development of statistical theory.

4. Now in *contrast* to the hypothetical populations, which was the concern of the general statistical theory developed by Galton, Pearson, Fisher, etc., the survey-populations, as mentioned in paragraph 2, are *real* in the sense that they consist of a fixed number of real individuals. This basic *distinction* between a hypothetical and survey-population was not clearly understood for a long time. As a result, theorems true for hypothetical populations were also *assumed*, implicitly or explicitly, to be true for survey-populations. Here we see the answer to the question raised at the end of paragraph 2. Since for an hypothetical population, under very general conditions among the linear estimates, sample mean was known, from the days of Gauss, to be the unbiased minimum variance (UMV) estimate for the population mean, the same thing was *believed* to be true for a survey-population. For the convenience of presentation in paragraph 2, I have referred to Halmos' (1947) generalization, removing the restriction of linearity in Gauss' result. It is easy to see that for the illustration considered in paragraph

[1] Here I must say I find Barnard's [(1969, p. 708] interpretation of Martin–Löb, Kolmogoroff–type mathematical definitions of "randomness" in terms of the permutations of labels, statistically meaningless.

[2] There are numerous examples in the survey-sampling literature of this erroneous and confusing emphasis. Perhaps this confusion is best illustrated by an early paper of Hansen and Hurwitz (1943) where apparently, for the first time, the term 'best linear unbiased estimate' has been extensively used in relation to survey-sampling. A recent illustration of this confusion about the 'best estimator' is furnished by Hartley and Rao (1962, p. 351). Some of these authors may have meant, actually, something different than what they wrote.

2 if the individual labels $i = 1, \cdots, N$ are *ignored*, the hypothetical population generated by n draws without replacement has a joint n-variate frequency distribution such that each variate has a common unknown marginal distribution, determined by the unknown values x_i, $i = 1, \cdots, N$, the mean value of this marginal being the same as (1), the covariance between any two of the n variates being $-1/N$. Indeed for this hypothetical distribution, according to the Halmos (1947)–Watson (1964) Theorem, the sample mean (2) is the UMV estimate for (1). This result, however, becomes irrelevant or meaningless or false, as said before, if the individual labels i, $i = 1, \cdots, N$ are *not* ignored.

5. With the above background I can describe the 'central issue' (referred to in paragraph 1) in the discussions at the Symposium as follows: If the individual labels are entirely *uninformative* about the corresponding variate values, intuitively the sample mean (2) is the most *appropriate* point estimate for the population mean (1), provided simple random sampling without replacement is adopted. Corresponding to this *intuitive appropriateness* the only *formal* optimality property for the sample mean is its UMV-ness. Even this UMV-ness is not available if the individual labels are not ignored. This indeed is disturbing. If statistical theory could not explain such crucial statistical intuition as above, the theory would be seriously inadequate or unrealistic. One may try to get out of this disturbing situation by adopting one of the following two approaches;

(I) by extending the statistical theory with a new model and corresponding formal criteria of optimality or appropriateness,

(II) by interpreting survey-sampling in such a way that it would fit within the framework (model) of the *general statistical theory,* referred to in paragraph 3.

The 'central issue' in the discussions at the Symposium, I think, could be expressed as: Whether (I) or (II)? Since (1955), several authors, including myself, contributed to developing and formalizing (I) above. These contributions are far too many to list here. A reference to them may be found in my symposium paper (1969). These contributions, in my opinion, have clearly demonstrated that the conventional optimality criteria such as UMV or ML estimation and the like which arose in relation to hypothetical populations referred to in paragraph 3, could be replaced by some other equally reasonable criteria that are appropriate for survey-sampling: For instance we may refer to the latest criterion suggested by Kempthorne (1969, p. 678, 679) or the criteria previously suggested by Godambe (1955, 1969), Hájek (1959), Hanurav (1965). Even Bayesian analysis which can be implemented by both the approaches (I) and (II) above, becomes much more realistic by approach (I), [Ericson, (1969)], than by approach (II) for the reasons given in paragraph 7. Now for the convenience of further discussion I will use the following notation. A set U of individuals i, $U = \{i\}$, is called a population, $i = 1, 2, \cdots, N$. A subset s of U, $s \subset U$, is called a sample. If x_i' is the (real) variate value associated with the individual i, $i = 1, \cdots, N$,

$$\mathbf{x}' = (x_1', \cdots, x_N'), \tag{3}$$

is called a *population vector.* If S denotes the totality of the subsets s of U,

$$S = \{s\}, \qquad s \subset U,$$

then without any loss of theoretical generality a *sampling design* can be defined as a function

$$p \tag{4}$$

on S such that for all $s \epsilon S$, $1 \geq p(s) \geq 0$ and $\sum_S p(s) = 1$. Next,

$$(s, x_i : i \epsilon s) \tag{5}$$

will be called the *data.* It is already known [Godambe (1966)] that the approach (I) above is characterized by the function 'Prob. $(\cdot|\cdot)$' in (6), to follow, which defines the probability of obtaining the data (5) given the sampling design (4) and the population vector (3).

$$\text{Prob. } (s, x_i : i \epsilon s \mid \mathbf{x}', p) = \begin{cases} p(s) \text{ if } x_i = x_i' \text{ for all } i \epsilon s, \\ \qquad\qquad \text{for all } \mathbf{x}' \epsilon R_N,\ s \epsilon S, \\ 0 \text{ otherwise,} \end{cases} \tag{6}$$

where R_N is the Euclidean N-Space. With the appropriate prior knowledge, R_N in (6) could be replaced by $R_N{}^*$, some subset of R_N, [Godambe (1969), p. 40]. For a given sampling design p, in (6) above $\mathbf{x}'$ is the *unknown parameter,* R being the parametric space. Further given p and $\mathbf{x}'$ the function 'Prob.' in (6) is defined on the sample space,

$$A = \{(s, x_i : i \epsilon s) : -\infty < x_i < \infty, i \epsilon s : s \epsilon S\}. \tag{7}$$

{Thus given the data (5), (6) defines the likelihood function on the parameter space R_N or $R_N{}^*$. This simple likelihood function was not understood until recently. For instance, we find Barnard et al. (1962, p. 370) asserting, (in reply to Durbin's question), that in the present situation sample space, parameter space and kernel function are *not* defined! In fact, with apologies to Barnard and Durbin I should say, here the whole discussion on 'finite population' [Barnard et al. (1962), p. 353, 370] shows how much the entire subject was misunderstood.}

6. At the Symposium, as over the past several years, I, along with several others, upheld the approach (I), characterized by (6) above. The opposing viewpoint supporting approach (II), at the Symposium, was primarily based on two very recent works: One by Royall (1967–68) and the other by Hartley and Rao (1967–68, 1969). Of these two, I will first explain and comment on, Royall's work. It essentially consists of eliminating labels 's' in (6) above by appropriate summation and writing the probability of the *unlabelled* values $x_i : i \epsilon s$, denoted by

$$\text{Prob. } ([x_i : i \epsilon s] \mid \mathbf{x}', p). \tag{8}$$

Using (8) and assuming parameter space to be some symmetric subset R_N^* of R_N in (6), he proves the sample mean (2) to be the UMV estimate of the population mean (1), for *simple random sampling without replacement, with a fixed n umber of draws.* (Such sampling I will, from now on, denote by *SR-sampling.*) This is quite in line with my observation in paragraph 2 based on the Halmos (1947)–Watson (1964) theorem. Then following the idea behind maximum likelihood estimation, (MLE), Royall defines Best Supported Estimation, (BSE). Though Royall considers sampling designs p in (4) above where $p(s)$ is determined by the number of individuals i in s (and of course N), I think without loss of any conceptual generality I can, for the discussion here, present his BSE method, assuming SR-sampling. In (8), [refer to (3)], let $(x_1', \cdots, x_N')$ contain only M distinct values say,

$$z_1, \cdots, z_i, \cdots, z_M, \tag{9}$$

which, including M itself, may or may not be known. Let further for $j = 1, \cdots, M$

$$K_j = \sum_{i=1}^{N} \phi(z_j, x_i'), \quad \text{where } \phi(z_j, x_i') = \begin{cases} 1 & \text{if } z_j = x_i', \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

In (10) above, K_j, $j = 1, \cdots, M$ are assumed to be *unknown*. Further for the unlabelled values of the sample, i.e. $[x_i : i \epsilon s]$ in (8), for $j = 1, \cdots, M$, let

$$k_j = \sum_{i \epsilon s} \phi(z_j, x_i), \tag{11}$$

where ϕ is the same as in (10). From (10) and (11) we have in (8), assuming $p \equiv$ SR-sampling,

$$\text{Prob. } ([x_i : i \epsilon s] \,|\, \mathbf{x}') = \prod_{j=1}^{M} \binom{K_j}{k_j} \Big/ \binom{N}{n}, \tag{12}$$

where n is the sample-size. In (12) above K_j, $j = 1, \cdots, M$ are unknown but k_j, $j = 1, \cdots, M$ are known. Let $\hat{K}_j$ be the value of K_j, $j = 1, \cdots, M$ which maximizes (12) for the given values of k_j, $j = 1, \cdots, M$. Assuming for any j if $k_j = 0$, $\hat{K}_j = 0$, Royall defines the BSE for population mean (2) as

$$\hat{\bar{x}}_N = \sum_{1}^{M} \hat{K}_j z_j / N. \tag{13}$$

An interesting example given by Royall shows how the BSE (13) can differ considerably from the sample mean (2).

7. The above theory of Royall is based upon *deciding* right at the start that individual labels, after a SR-sample is drawn, are *irrelevant* concerning the problem of inference. Indeed it appears to me that Royall takes for granted that the individual labels in a population can at best be of use to draw a SR-sample or to stratify the population and then from each stratum to draw a SR-sample or perhaps to stratify the population after a SR-sample is drawn, and the like. He indeed gives the obvious extention of the BSE (13) appropriate for stratification. It is assumed that the stratification is sharp enough so that within each stratum, individual labels provide no further discrimination between the corresponding x-values in (3). Here I see a source of confusion. It is one thing to say that within a stratum individual labels are *irrelevant* if one wants to discriminate between the corresponding x-values and the other a very different thing, to say that the individual labels are *irrelevant* when one is concerned with *any* inference problem related to the population: The former statement looks intuitive to me but *not* the latter. Actually whether a certain aspect of the data is relevant for the purposes of inference or not will depend upon the tools of inference available at the time. I do not believe in the notion of 'absolute relevance' or 'absolute irrelevance'. There must always be an interplay between the concept of relevance (or irrelevance) and statistical theory. This interplay, I believe, is necessary for the development of both, on one side the 'concept of irrelevance' and on the other the 'statistical theory'. From a Bayesian viewpoint this interplay results in computations of priors on wider or richer parametric spaces, which are required to include the so called 'irrelevant data' in the model. An excellent illustration of this is provided by Empirical Bayes Methods, [H. Robbins (1956)]. Here, apparently, utterly *unrelated* data are pooled together to improve the inference related to each. The extensive development of sampling theory based on the approach (I) to which I referred in paragraph 5, provides another such illustration. The *general statistical theory* which was tied to hypothetical populations (refer to paragraph 3) was basically incapable of providing any inference utilizing the individual labels. Here I may emphasize that the idea of 'labelling' to draw a random sample is *not* new. Some historian may, extending Kempthorne's (1969, p. 673) guess of 40 years, trace the origin of the idea even to 4,000 years ago in history! What is essentially new is the understanding of the formal role of labels in the inference procedure. [Surely, I agree with Professor Kempthorne (1969, p. 685) that some aspects or *parts* of the 'sample' will have to be *ignored* in the process of inference. But this, as I said before, is due to our incapacity or the incapacity of the tools of inference at our disposal at the moment. At some later moment more sharpened tools might enable us to utilize *some* of the *ignored* part of the sample]. In fact it is clear from Royall's work that a sampling theory which is based on the reduced data $[x_i : i \epsilon s]$ [refer to (8)], ignoring individual labels, can at best explain stratified sampling and the like. More *sophisticated sampling designs,* (refer to paragraph 9) will be essentially outside the scope of such a theory. It is evident that one cannot at all formally study the more sophisticated sampling designs utilized by matured and experienced practitioners, except through the probability function (6), based on the approach (I). Actually Royall's work does not go any further than stratification. But I will have the occasion to comment on this in more detail in paragraph 9. Here I would like to emphasize one point, lest I may be misunderstood. I do *not* say that

Royall's probability function (8) refers to any hypothetical population, mentioned in paragraph 3. However, the former, (8), has a basic feature in common with the frequency functions related (refer paragraph 3) to hypothetical populations and that is the absence of individual lables.

8. Now I proceed to comment on the work of Hartley and Rao (1967–68) which also aims at implementing the approach (II), (refer paragraph 5). In this direction they attempt to prove that many estimators for the survey-population, in common use, are either UMV or ML estimators, ignoring the individual labels. They say:

> 'we confine ourselves to the estimators which do not functionally depend on the labels.' (1968, p. 547). — A

Hartley and Rao's, (from now on I will denote these authors by just 'HR') work is much less formal than Royall's. Royall, as I have shown in paragraph 7, formally reduces the data $(s, x_i : i \epsilon s)$ in (5) to $[x_i : i \epsilon s]$ and *starts* with the probability function (8) to obtain (12) and (13). On the other hand HR, assuming SR-sampling, *start* with the distribution (12) and obtain an estimator which is analogous to (13). Again, independently HR prove the sample mean (2) is the UMV-estimator for the population mean (1). This, as we have seen before, is in line with Royall's result. Concerning Stratified SR-sampling I must say HR-theory is very unclear. For instance they say:

> 'in most situations in which labels are attached to the units it is known that they cannot be informative beyond the design stage.' (1968, p. 549). — B

The above statements A and B, reproduced from HR (1968), do not enable me to obtain the *conventional estimator* for the population mean when stratified SR-sampling is adopted. It would be *inconsistent* with the two statements, A and B, quoted above to attach *different* weights for two *equal* x-values drawn from different strata; while the conventional estimator is obtained by multiplying each x-value by the corresponding stratum size; and yet HR say:

> 'In particular, the customary estimators are U.M.V. noting that each stratum is described by its separate set of parameters N_t.' (1968, p. 550). — C

Surely the assumption of different sets of parameters for different strata, (as stated in C above) at the *estimation stage,* contradicts HR-statements A and B above. [This remains true even after an elaborate treatment of stratification given by HR in (1969).] A simple way to avoid the contradiction is to say in addition to A above that 'the parameters (using HR-terminology) of different strata will be estimated separately or independently'. It clearly shows that even though HR claim to have developed a 'theory' of estimation, the simplest of the results they present can be validated only by some sort of *ad hoc* reasoning. Of course I do not mean to say that this *ad hoc* reasoning is implausible; but it illustrates that HR (1967–68) work considerably *lacks* in *theoretical* (*deductive*) *structure.* This lack of theoretical structure is all the more seriously felt in the authors' subsequent paper, [HR, (1969)].

9. To elaborate on the foregoing remarks, I comment in some detail on the HR (1969) paper. They say:

> 'Thus any identifying labels, i, that may be attached to the units may or may not be used for implementation of the sample design; however labels are not *directly used* in the computation of the estimators.' [HR (1969), p. 148] [italics mine]. — D

I think, the phrase 'however labels are not *directly used* . . .' in D above is of basic importance to HR-work; and yet nowhere do they formally define the underlying concept. Neither do they clearly explain what they mean by 'directly used'. I get a feeling, from their subsequent statement:

> 'Certain situations where labels of higher stage units are not informative also exist, for example identifiable subsets of certain lists. Both 'scale-load' and 'label-dependent' estimators are therefore required.' [HR (1969, p. 149] — E

that the authors possibly meant to say: at the estimation stage labels should be used only to the extent they are used at the design stage. This may have some intuitive *meaning,* though not necessarily a unique one. However, unless this *meaning* is made explicit, HR-work will not qualify itself to be called a 'theory'; whether a right or wrong theory or whether an adequate or inadequate theory is a different issue. Assuming a theory, ignoring partially, labels, is developed, still as I stated in paragraph 6 while commenting on Royall's work, this theory can at best explain stratified SR-sampling and the like. What such a theory, specifically cannot explain is the following: Using the notation in paragraph 5, I will say, when a general sophisticated sampling design p in (4) is employed with appropriate inclusion probabilities,

$$\pi_i = \sum_{s \ni i} p(s), \qquad i = 1, \cdots, N, \tag{14}$$

'$s \ni i$' meaning 'all subsets s which include the individual i', often the corresponding estimator for the population mean (1) is given by Horvitz–Thomson estimator,

$$\frac{1}{N} \sum_{i \epsilon s} x_i / \pi_i. \tag{15}$$

Indeed as *special* cases of (15) we get sample mean (2) or the conventional estimator in stratified SR-sampling, [refer HR, (1969), eq. (21)]. These special cases, I can imagine, could be explained by a possible theory, which partially ignores labels. However, such a theory, in principle, cannot explain the *general* estimator (15).

As a matter of fact in a 'single-stage' unequal probability sampling [HR, (1969), section 5], when m draws are made with replacement, the probability of selection for the individual i, at *each* draw being q_i, $i = 1, \cdots, N$, if m_i denotes the number of times the individual i, $i = 1, \cdots, N$, is drawn HR obtain the estimator, [HR (1969), eq. (37)],

$$(1/mN) \sum_{i \epsilon s} x_i m_i / q_i \tag{16}$$

(notation mine). It is well-known that the estimator (16) is *inadmissible* while the estimator (15) is always *admissible*, [Godambe and Joshi (1965)]. In the present case the estimator (15) is given by

$$(1/N) \sum_{i \epsilon s} x_i / [1 - (1 - q_i)^m], \tag{17}$$

[Godambe (1955)], which of course is different than the HR estimator (16). This, I believe, is a sufficient demonstration, of the serious intrinsic limitations of a possible theory which ignores partially (at least) the individual labels. Curiously enough I find HR [(1969) p. 162] after obtaining (16) above [or equation (37) in HR-1969], saying:

> 'Although only one single method of unequal probability sampling is examined in this section and although the method examined is known not to be particularly efficient . . .' } F

They seem to be completely unaware of the fact that it is not the method of sampling that is inefficient but what is inefficient is their estimator (16); for, with suitable values of the selection probabilities q_i, $i = 1, \cdots, N$ the method at least theoretically, will not be very objectionable. The inefficient estimator (16) clearly speaks by itself about the general HR-approach. In conclusion, to avoid misunderstanding, I would like to say that it is not the main purpose of this article to defend or advocate any particular estimator, or any property of optimality, etc.; rather the main purpose is to emphasize the unbiased (in the genuine sense) approach-(I), of paragraph 5, which provides a broad enough logical set-up within which most of the problems of survey-sampling could be formulated and discussed without any prejudice, particularly, without the prejudice, that 'such and such individual labels must be uninformative and therefore must be irrelevant for inference'.

I may conclude by expressing my appreciation for all efforts that Professors Nicholson, Johnson, Smith and others from the University of North Carolina have made for bringing about this unusually 'rewarding' [Barnard (1969), p. 711] Symposium.

REFERENCES

Barnard, G. A., Jenkins, G. M. and Winsten, C. B. (1962), Likelihood inference and time series, *J. R. Statist. Soc.* A, 125, 321–372.

Barnard, G. A. (1969), Summary remarks, New Developments in Survey Sampling, John Wiley and Sons, 696–741.

Ericson, W. A. (1969), Subjective Bayesian models in sampling finite populations: stratification, New Developments in Survey Sampling, John Wiley and Sons, 326–357.

Fisher, R. A. (1956), Statistical Methods and Scientific Inference, Oliver and Boyd, London, 31–36.

Fisher, R. A. (1959), The Cancer Controversy, Oliver and Boyd, London, 26–38.

Galton, F. (1889), Natural Inheritance, Macmillan and Co., London and New York.

Godambe, V. P. (1955), A unified theory of sampling from finite populations, *J. R. Statist. Soc.* B, 17, 268–278.

Godambe, V. P. and Joshi, V. M. (1965), Admissibility and Bayes estimation in sampling from finite populations—I, *Ann. Math. Statist.*, 36, 1707–1722.

Godambe, V. P. (1966), A new approach to sampling from finite populations—II, *J. R. Statist. Soc.* B, 28, 320–328.

Godambe, V. P. (1969), Some aspects of the theoretical developments in survey-sampling, New Developments in Survey Sampling, John Wiley and Sons, 27–58.

Hájek, J. (1959), Optimum strategy and other problems in probability sampling, *Casopis Pest. Mat.* 84, 387–423.

Hansen, Morris H. and Hurwitz, William N. (1943), On the theory of sampling from finite populations, *Annals of Math. Statist.*, 14, 333–362.

Hanurav, T. V. (1965), Hyper-admissibility and optimum estimators in sampling finite populations, Ph.D Thesis, Indian Statist. Inst., Calcutta.

Hartley, H. O. and Rao, J. N. K. (1962), Sampling with unequal probabilities and without replacement, *Annals of Math. Statist.*, 350–374.

Hartley, H. O. and Rao, J. N. K. (1967–68), A new estimation theory of sample surveys, (1967), reported first at the joint statistical meetings in Washington, D.C. and published subsequently (1968), *Biometrika*, 55, 547–557.

Hartley, H. O. and Rao, J. N. K. (1969), A new estimation theory for sample surveys II, New Developments in Survey Sampling, John Wiley and Sons, 147–169.

Kempthorne, Oscar (1969), Some remarks on statistical inference in finite sampling, New Developments in Survey Sampling, John Wiley and Sons, 671–695.

Pearson, Karl (1920), The fundamental problem of practical statistics, *Biometrika*, 13, 1–16.

Robbins, Herbert (1956), An empirical Bayes approach to statistics, Proc. Third Berkeley Symposium on Math. Statist. and Prob., Univ. of Calif. Press, Berkeley and Los Angeles, 157–163.

Royall, Richard (1967–68), An old approach to finite population sampling theory, (1967), reported first at the joint statistical meetings in Washington, D.C. and published subsequently (1968), *J. Amer. Statist. Assoc.* 63, 1269–1279.

Watson, G. S. (1964), Estimation in finite populations, (unpublished report).