Foundations of Survey Sampling (A Don Quixote Tragedy)*

H. O. HARTLEY
Texas A & M University
and
J. N. K. RAO
University of Manitoba

1. *The Rebuttal*

1.1 *The Aims of Our Papers.*

This paper is a reply to the paper "Foundations of Survey Sampling" by V. P. Godambe which appeared in the February, 1970, issue of the American Statistician. The latter paper discussed at length certain aspects of the April, 1968, Symposium on Foundations of Survey Sampling which was organized by the University of North Carolina at Chapel Hill. Specifically two of our papers, Hartley and Rao (1967–68) and Hartley and Rao (1969), were criticized by Dr. Godambe and since his discussion is based on considerable misunderstanding, it is necessary for us to reply to his assertions in detail.

Since this paper is a rebuttal, we refer the reader to the description by Dr. Godambe of what he considered the "central issue" of the Symposium. We quote from his section 5:

"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 5.

The 'central issue' in the discussions at the Symposium, I think, could be expressed as; Whether (I) or (II)?" (1970, p. 35).

In paragraph 6. he describes very briefly his and other statisticians work along the line of what he calls Approach (I), and then he says, "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".

Since we feel that Dr. Godambe has misunderstood the aims of our two papers we state these here: In both papers we were concerned with a new technique for sample surveys in which the k-vectors of characteristics y_i attached to the N units $(i = 1, \ldots, N)$ of a finite population are measured on discrete scales comprising a finite number of T scale vectors y_t $(t = 1, \ldots, T)$. In both papers we derived some distributional results (for certain classes of sample designs and estimators) resulting from this approach, our first paper (Hartley and Rao, 1967–68) being predominantly concerned with the so-called "optimality properties." Concerning these, we reiterate certain well-known principles of statistical theory concerning optimality properties of estimators. It is recognized that optimality properties depend (among others) on:

(1) the stochastic procedure supplying the observed data,
(2) the optimality criterion used,
(3) the class of estimators, i.e. the mathematical functions computed from the observed data admitted to the "competition for optimality",
(4) the parametric ranges for which the optimality property is claimed.

To illustrate the importance of (3) we should mention the well-known classical example of BLUE least squares estimators in which (3) is restricted to the class of linear unbiassed estimators. An interesting example of the importance of (4) is afforded by the fact that "admissibility" of the so-called Horvitz-Thompson estimator was proved by Godambe and Joshi (1965) for finite populations whose characteristics (parameters) may attain the value zero.

However, when the parameters are known to be strictly positive (which is the case for most populations occurring in practice), no such result has been proved. Indeed, Basu (1969) has shown that the Horvitz-Thompson estimator will be inadmissible if the parameters are known to be restricted to certain positive intervals.

With these preliminaries we are able to give more details about the aims of our first paper (Hartley and Rao, 1967–68). We were concerned with the derivation of certain optimality properties for a subset of estimators (called by us scale load estimators and perhaps alternatively described as estimators not depending on the indentifying labels) for certain specified sampling procedures. The optimality properties considered were UMV-ness and Maximum Likelihood. Briefly, therefore, we derived optimality properties for a *subset* of estimators which would presumably be admitted within the framework of Dr. Godambe's Approach (I). The procedure can be attacked on the grounds that the

* Supported by research grants of the U.S. Army Research Office and the National Research Council of Canada.

subclass of estimators considered by us is irrelevant. We discuss this point in sections 2–3.

In the second paper, Hartley and Rao (1969), we do *not* restrict our estimators to the above subclass but still use the above technique of discrete scaling (called by us reparametrization) to derive new results clearly outlined in the paper. The fact that we do not exclude the use of identifying labels is perhaps well illustrated by the fact that Dr. Godambe's likelihood (Godambe, 1970, equation 6) which he has used for some time and which he regards as fundamental to his Approach (I) arises as a special case of our likelihood formula (Hartley and Rao, 1969, equation 39) the special case being that when all $N_i = 1$, where N_i is the number of secondaries in the ith primary unit. Formula (39) with its parameter- and variate-ranges duly recognized then becomes a restatement of the sampling procedure $P(u_1 \ldots u_L)$ and is identical with Godambe's (6) as a restatement of the sampling procedure $p(s)$. However, our theory restricts the characteristics y_i to the scale points y_t.

Now it must be obvious to the reader of Dr. Godambe's paper that the aims he ascribes to us are quite different from the above stated aims. Indeed he quotes us as saying (statement A, p. 37)[1]

A { "we confine ourselves to the estimators which do not functionally depend on the labels".

Actually our statement is (1967–68, p. 547):—

A* { "We consider it therefore of interest to develop an estimation theory in which estimators are allowed to depend on labels only if these can be regarded as informative *concomitant* variables, and in the present paper we confine ourselves to estimators which do not functionally depend on the labels".

The first part of our statement is unfortunately omitted. Also the following clear statement made on page 149 of Hartley and Rao (1969) is ignored:

G* { "We consider that identifying labels of primary units (or all but the last stage units) will often be available as well as informative".

Clearly then the aims attributed to us by Dr. Godambe differ strikingly from those professed by us and the fact that the latter were well understood at the symposium is clear from Dr. G. A. Barnard's 'Summing Up' of the symposium (see particularly page 709) in which the technical content of our paper is summarized by the sentence, "The treatment of an observation in terms of scaling points was something which Fisher did very frequently with good effect (see, for example, his 1934 RSS paper); and it does help to view samples from this point of view."

We cannot help feeling that the misunderstanding of our aims by Dr. Godambe is reminiscent of the knight Don Quixote who tragically mistakes harmless windmills as hostile knights and launches violent and entirely victorious attacks against them! We now deal specifically with three of Dr. Godambe's more important charges against us.

1.2 *The Alleged Contradiction in the Treatment of Simple Stratified Sampling.*

For clarity, we recall that in the literature simple stratified sampling is described with the help of one of two concepts leading to an identical specification. They are:

S_1: The strata are regarded as separate populations, each described by its *separate* set of parameters, each sampled independently. Certain parameter functions to be estimated (such as the total of the original composite population) depend on all the parameters.
S_2: The strata are regarded as "primary units" of a single population which are all sampled.

The fact that we use notion S_1 is clear from the following quotation (Hartley and Rao, 1969, p. 155) and overlooked by Dr. Godambe. We say:

H* { "Notice that each stratum is described by its *separate* set of parameters i.e., we have an additional subscript i to *index the strata*".

Ignoring this statement, Dr. Godambe states that in contradiction to A we compute estimators that do in fact depend on the labels of the *units* while we clearly state that our subscript i is an index to a *stratum*, i.e., of a subpopulation. This fact is also stressed in our correctly (!) quoted statement C=C*. However, even if Dr. Godambe argues that we should have used the concept S_2, a contradiction can only be construed by his convenient truncation of our statement: Stratification is treated in our second (1969) paper, and in Dr. Godambe's truncation A of our statement A* (quoted from our 1967–68 paper) he has omitted the phrase "in this paper" and only goes on to say "we confine ourselves to the estimators which do not functionally depend on the labels." The fact that label dependence is considered often appropriate for the more general estimators of our second paper, is clearly stated on page 149 (quoted above as G*).

The presumed contradiction is ridiculed by describing it as a singular failure to achieve the (falsely alleged) aim of developing a theory *exclusively* confined to label-independent estimators.

Dr. Godambe then has second thoughts and admits that the "contradiction" could have been avoided by making the statement "the parameters of different strata will be estimated separately or independently." We leave it to the reader to judge whether this differs at all from the statements that we originally made.

[1] We refer to Dr. Godambe's truncations of our statements by his letters A, B, D, F while our associated complete statements are denoted by A*, C=C*, D*, F* with additional quotations referenced by G*, H*.

1.3 *The Alleged Lack of Clear Definitions.*

Dr. Godambe alleges that our concepts are not clearly defined by the device of ignoring the clear definition given by us and only quoting an explanatory rider following the definition. The complete definition (D*) and rider (D) we give in our paper are as follows (Hartley and Rao, 1969, p. 148):

D* { "In our previous paper we restricted (a) to simple random sampling and we confined the computation of estimators (b) to what we termed "scale-load" estimators. These were defined as mathematical functions of the scale vectors $\mathbf{y}_t$ and their sample-loads (frequencies) n_t = number of units in the sample having $\mathbf{y}_t$".

D D* { "Thus any identifying labels, i, that may be attached to the units may or may not be used for [the] implementation of the sample design; however, labels are not directly used in the computation of estimators".

In detaching the rider (D) from the definition (D*) the rider is made to appear vague. We adjoined it to the clear definition of "scale-load" estimators preceeding it to avoid a possible confusion in that labels *may* be used to implement the sample *design* but should *not* occur as functional arguments in the mathematical functions defining the *scale-load estimators.*

The lack of clear definition is used as an "elaboration" of Dr. Godambe's assertion that our work lacks theoretical structure. There are additional recriminations which we discuss in section 1.4.

1.4 *The Alleged Inefficiency in our Treatment of Unequal Probability Sampling.*

By truncating the summary of our short section on unequal probability sampling with replacement the aims of this section are distorted. The complete statement F* and Dr. Godambe's truncation F are shown below (1969, p. 162):—

F { F* { "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, the } discussion clearly indicates the possibility of deriving concrete likelihoods for other unequal probability sampling methods with the help of our technique of parametrization".

The second (omitted) part of the sentence clearly indicates our aim, namely to use our approach to derive concrete likelihoods and distributional properties in the area of unequal probability sampling. The result reported in Hartley and Rao (1969) is a proof of the maximum likelihood property of a well-known estimator, by considering a likelihood based on all the sample draws (leading to not necessarily distinct units). This estimator has been in use for decades and can be written in the form:[2]

$$\left(\frac{1}{mN}\right) \sum_{i \in s} y_i m_i / q_i \tag{1}$$

where q_i is the probability of drawing the ith unit at each individual draw and m_i is the number of times the ith unit is drawn ($\sum m_i = m$) $i = 1, 2, \ldots, N$. Godambe's objection to the estimator (1) is that it is inadmissible (as it depends on the m_i) unlike the Horvitz-Thompson estimator (for this method of sampling) namely:

$$\left(\frac{1}{N}\right) \sum_{i \in s} y_i / [1 - (1 - q_i)^m] \tag{2}$$

which is independent of the m_i. However, we had already demonstrated in our (1967–68) paper that, for simple random sampling with replacement, our approach, in fact, leads to a maximum likelihood estimator which is independent of the m_i. This result was obtained by considering the scale-load likelihood based on the n_t's, where n_t is the number of *distinct* units in the sample having the scale- point y_t. It is obvious that the maximum likelihood estimator for the present method of unequal probability sampling would also be independent of the m_i, provided the scale-load likelihood based on the n_t's is considered. In order to distinguish this latter estimator from (1) above, we took care to make the following statement (1969, p. 162):— "Finally it should be noted that (35) is the likelihood for the scores which do not necessarily represent counts of distinct units in the population.[3] However, it is possible to obtain the likelihood of the number of distinct units in the sample with scale ratio r_t which we denote by n_t. ...We intend to examine this distribution in more detail elsewhere". Unfortunately, Dr. Godambe has ignored the above statement and, instead, says "the inefficient estimator (6) [i.e., our (1) above] speaks by itself about the general H-R approach."

By contrast to this negative comment on estimator (1) (which is of course not "our" estimator) Dr. Godambe states about the Horvitz-Thompson estimator (2) that it is "always admissible" (page 38). This latter statement requires clarification. For, as pointed out earlier, the property of admissibility has not been proved for the Horvitz-Thompson estimator if the parameters attached to the units are known to be strictly positive.

Finally, we turn to Dr. Godambe's assertion of an erroneous statement in the truncated F. He says "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) [i.e., the above (1)]; for with suitable values of selection probabilities q_i, $i = 1, \ldots, N$, the method at least theoretically, will

[2] Godambe used x instead of y to denote a character of interest.

[3] To make our point even clearer we should perhaps have substituted "sample" for "population" here.

not be very objectionable".[4] We leave it to the reader to judge the merits of this statement vis a vis the following facts: Hanurav (1962, p. 429) states: "Since there does not exist a design[5] in which the variance[6] is uniformily minimum, the optimal designs are obtained by minimizing the expected variance under a realistic super-population set-up. These turn out to be designs in which the effective sample size is constant for all samples of the design". (see also Hanurav (1965, p. 199). Moreover in Godambe's own (1955) paper (section 7) a more general result is proved for a special case of the above super-population set-up, which provides a justification for preferring 'without replacement sampling' over 'sampling with replacement'. Finally ample evidence for the superiority of estimators in without-replacement sampling over those in whit-replacement sampling has already been provided by J. N. K. Rao (1966) and Ramakrishnan (1969), at least for the cases of equal probability sampling and Stevens' (1958) method of unequal probability sampling.

2. *The Relevance of Label-Independent (Scale-Load) Estimators*

2.1 *Definitions*

If we had been capable of writing papers that (to use Dr. Godambe's phrase) do not 'lack theoretical structure' we could have simply stated "let us consider the class of label independent estimators which are defined as follows...," and left it at that. However, as applied statisticians we felt compelled to at least discuss the 'need' for, or the 'relevance' of, this class of estimators. Now we have to confess immediately that the latter two notions depend on nonmathematical concepts such as the frequencies with which certain types of populations are 'encountered in practice' and the like. It is apparent that Dr. Godambe is somewhat allergic to such discussions (as he has every right to be). We believe that his misunderstanding of our aims is very strongly related to our raising these points in our papers and the consequential discussion at the symposium.

In order to avoid confusion we should commence with a definition of identifying labels and 'scale-load' estimators: Consider any set of observable attributes $l(j)$ attached to the N units of the population ($j = 1, \ldots, N$) where j is a non-observable conceptual index of a unit. If the observable $l(j)$ have the property $l(j) \neq l(j')$ if $j \neq j'$ then these attributes may be used as 'identifying labels'. Since the above property of labels remains invariant with regard to one to one mappings it is customary to use labels $i(j)$ which take the N integer values $i = 1, \ldots, N$. Labels may in certain cases be used to identify categorical concomitant variables that can then be attached to the units (e.g., the county in which an identified and labelled farm operator has his headquarters). Such categorical variables will however not in *general* be usable as labels since usually several units fall into the same category. However, in the special case in which only one unit falls into each category the categorical variables could be used as identifying labels of the units. We should note that in the general case such categorical variables are often used both, in the design stage (e.g., multistage sampling) as well as in the estimation stage, and we have in such cases referred to them as 'primary, secondary, or all but last stage unit labels'. The word label per se is therefore reserved for the identifying label of the last stage unit. Label-independent or scale-load estimators were defined by us as mathematical functions which only depend on the sample frequencies n_t with which the tth scale vector $\mathbf{y}_t$ is attained ($t = 1, \ldots, T =$ Total number of scale vectors). It must be made clear that since the $\mathbf{y}_i$ and $\mathbf{y}_t$ are allowed to be k-vectors of k attributes it could be argued (and we have discussed this point) that labels could be adjoined as the $(k + 1)$st element of the attribute vector $\mathbf{y}_i$. However, in that case we would have the priori information that $N_t =$ number of units in the population having $\mathbf{y}_t$ is either 0 or 1. In our definition of scale-load estimators such situations are excluded since it is assumed that no such prior information on our parameters N_t is available, or in other words, it is assumed that none of the elements of the $\mathbf{y}_i$ vectors is known to have the label property defined above.

Strata are regarded as separate populations with their separate designs and parameters. Historically speaking, practically all estimators used by practitioners are scale-load estimators, but we stress again that we have never recommended their *exclusive* use. However, there are many reasons why they are of considerable relevance and in the subsequent section we consider one such situation, namely the occurrence of finite populations with unlabelled units.

2.2 *Populations With Unlabelled Units.*

Dr. Godambe (1970, p. 34, left lines 2–30) says[7]: "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 (sic!) which essentially implies that all the individuals of the population are already labelled in a manner *known* to the sampler".

---

[4] Although the latter part of the statement is vague, we interpret it as follows: choose the q_i's such that the inclusion probability for the ith unit, π_i, is proportional to some known size z_i attached to the ith unit which is approximately proportional to $y_i = 1, \ldots, N$, (the so-called 'inclusion probabilities portional to size (IPPS) designs').

[5] Here Hanurav confines himself to IPPS designs.

[6] Variance of the Horvitz-Thompson estimator.

[7] For the context of this statement refer to Dr. Godambe's Section 2.

It would carry us too far afield to discuss here the concept of 'randomness' and we refer to the discussion by G. A. Barnard (1969) in the 'Summing Up' (pp. 707–708 and his references) where Dr. Godambe's maxim is rejected outright. However, we would like to raise the following points concerning the implementation of Dr. Godambe's maxim: It is known that all random number tables are imperfect simulations of random sequences. Tippett's (1927) table consists of central figures of British official statistics, the Fisher and Yates (1963 but numerous editions) tables consist of central decimals of a 24 decimal table of logarithms, the tables by Kendall and Smith (1939) have been generated by an electro-mechanical mechanism very similar to devices sometimes used for the drawing of a 'random sample' of unlabelled units. More recently, the 'product residue' method generating random digits in high speed computers and the resulting computer outputs have no mathematical guarantee of representing 'random sequences'. Of course, most of these tables have passed numerous searching 'tests for randomness'. If Dr. Godambe recommends the exclusive use of 'some random number tables' to draw samples of labeled units, the onus of the proof that this is a better simulation of a 'random sequence' than the physical processes and procedures of drawing random samples of unlabeled units is on Dr. Godambe. Next pre-labeling the units of an infinite population is an operation that is essentially impossible to implement and in any case requires the careful distinction of the concepts of 'enumerably infinite' and 'non-enumerably infinite', as well as a definition of his concept of a *'fixed non-finite* number of individuals'! (Is this a clear definition?)

However, it is apparent that Dr. Godambe wishes to exclude all populations of unlabeled units from his notion of 'survey populations' which are later described by him (see Section 4) as 'real in the sense that they consist of a fixed number of real individuals'. We now show below that this would prevent *his* theory (but not ours) to deal with a large family of problems in the sampling of finite populations. We have to confine ourselves to mentioning just a few examples.

2.2.1 *Acceptance Sampling*

The important area of acceptance sampling of finite lots of mass produced articles such as machine parts deals with finite populations of unlabeled and unidentified units. The number of units, always known to be finite, is often determined by count devices or bulk-weighing. It is well-known that the attachment of labels to such units is usually impractical. In many situations finite lots are stratified by categorical variables representing production characteristics. An interesting case of unequal probability sampling arises in the production of textile strands where certain procedures draw fibers with probabilities proportional to their length.

2.2.2 *Sampling from a Continuum.*

In many areas of both industrial and agricultural activities we are concerned with the sampling of what is essentially a continuum of 'size A' (e.g., a land area or a volume of a fluid or semi-fluid) in which the continuum is successfully treated as an atomistic finite population of N units by specifying units of size A/N. Once N and A/N have been chosen a procedure of splitting the population into its N 'real units' could indeed be implemented but the cost of doing this is usually astronomical. Sampling procedures have therefore been devised by which a sample of n 'reference points' in the continuum are selected by a specified random process and the associated n-sample of units actually constructed and their characteristics measured. The $N - n$ remaining units are never constructed let alone labeled. Particularly simple examples arise in the sampling of water reservoirs, grain silos, and the like. A somewhat more involved case is the well-known situation of 'area sampling' in agricultural surveys using either 'open' or 'closed' segments as sampling units. Because of the considerable amount of field work in delineating such segments by 'natural boundaries' it is imperative to confine this work to the sampled segments. We do not need to document the world-wide use of area sampling procedures. For an interesting instance of unequal probability sampling of an unlabeled population of farm operators using land acreages as sizes we refer to our (1969) paper (section 5) reporting on a method by R. J. Jessen.

2.2.3 *Sampling of Wild-Life Populations.*

It is well known that with these populations the number N (abundance) although known to be finite, is usually not known but is one of the target parameters. However, the units of the populations are clearly not found 'labeled'! Often area-units (which may be labeled) are used as higher stage units in multistage sampling procedures. Labels or 'tags' are sometimes attached to a 'first sample' of animals as with the well known capture-recapture procedures. Difficult problems of population coverage invariably arises.

3. *The Uninformativeness of Randomly Attached Labels*

In the previous section we discussed important situations in which label-independent estimators are not only relevant but vitally needed since labels are simply not available. We now turn to a class of situations in which the units of a population are labeled but labels are uninformative in the sense specified below. This situation arises when labels are randomly attached to units and it would certainly not appear to be surprising that such labels are uninformative. We confine ourselves here to the simplest case of 'random labeling' for which the mathematical definitions and concepts are as follows:

3.1 *The Two-Step Stochastic Process.*

Step 1. Random Labeling of Units: The N units of a population are conceptually identified by a non-observable index j ($j = 1, \ldots, N$). Before the sampling Step 2 commences, and unknown to the sampler, labels $i = 1, \ldots, N$ are attached to the units by choosing one of the $N!$ permutations $i(j)$ with equal probabilities $1/N!$.

Step 2. The Sample Selection: Given the set of N labeled units ($i = 1, \ldots, N$) a sample of fixed size n is drawn by what is called a 'size determined' design (see below).

3.2 *The Characteristics Attached to the Units*

We assume that N scalar[8] 'target attributes' y_j' ($j = 1, \ldots, N$) and N scalar[8] 'size attributes' x_j' are attached to the units as the elements of the two N-vectors $\mathbf{y}'$ and $\mathbf{x}'$.

3.3 *The 'Size Determined' Sample Design.*

One sample, s, out of the total number of possible S samples, comprising labels $i \in s$ is drawn by a process such that $\Pr(s) = p[x(s)]$ is a symmetric function of the numerical values x_i: $i \in s$.

3.4 *The Probability Distribution of the Stochastic Variables.*

The stochastic variables of the two-step process are: (a) the $N!$ permutations $i(j)$; (b) the set of sample labels s; (c) the attributes y_i, x_i found attached to the unit labeled i; (d) the 'variate transformations'

$$y_j^* = y_{i(j)},\ x_j^* = x_{i(j)},$$

that is the variate values y and x attached to the unit with index j through the permutation $i(j)$ and the sample draw $i \in s$.

Since the y_j^*, x_j^* are unique functions of the y_i, x_i and of $i(j)$ and since $i(j)$ represents a one-to-one mapping, the probability distribution can be formulated in terms of the variables $i(j)$, $i \in s$ and y_j^*, x_j^* and is given by:

$$\Pr\{i(j), s, y_j^*, x_j^* \mid \mathbf{y}', \mathbf{x}', p\} = \frac{1}{N!}\begin{cases} p[x(s)] & \text{if } y_j^* = y'_{i(j)},\ x_j^* = x'_{i(j)}, \\ & \text{for all } i \in s \text{ and for all } \mathbf{y}', \\ & \mathbf{x}' \text{ in } R_{2N}. \\ 0 & \text{Otherwise.} \end{cases} \quad (3)$$

where in (3) the factor $1/N!$ represents the Step 1 (marginal) probability of the variable $i(j)$ and the second factor represents the conditional probability of s, y_j^*, x_j^*, given $i(j)$ and uses notation analogous to Godambe's (1970, equation 6).

3.5 *The Likelihood Given the Data.*

The observable data are the set (s, $y_i = y'_i$, $x_i = x'_i$: $i \in s$) and hence using (3), the likelihood, given the data, is given by:

$$L(s, y_i, x_i: i \in s \mid y_j', x_j') = \frac{1}{N!}\sum_{j(i)}\begin{cases} p[x(s)] & \text{if } y^*_{j(i)} = y'_i,\ x^*_{j(i)} = x'_i \\ & \text{for all } i \in s \text{ and for all} \\ & \mathbf{y}', \mathbf{x}' \text{ in } R_{2N}. \\ 0 & \text{Otherwise.} \end{cases} \quad (4)$$

where $j(i)$ is the mapping inverse to $i(j)$ and $\sum_{j(i)}$ extends over all $N!$ permutations.

Clearly the likelihood (4) is completely determined by a specification of the observed numerical values $y_i = y'_i$, $x_i = x'_i$, irrespective of the set of observed labels s. This means that the information contained in s does not contribute anything to the information already contained in the observed numerical values of y_i and x_i. In this sense labels are uninformative. In the special case where the sizes, x_i, have the label property (i.e., no two units have the same size) the above statement is of course trivially correct since it states that the information contained in labels does not contribute anything to the information already contained in the y_i and the sizes x_i which already have the label property. If no sizes are available the above proof that observed labels are uninformative is maintained by the use of a constant set of x_i.

Concerning the concept of 'size determined' designs most survey designs fall into this category including all equal probability designs and all unequal probability 'draw by draw' designs. For brevity sake we do not enter here into a discussion of labeled populations in which it would be reasonable to make the assumption of a Step 1 random labeling, but it does seem to be often a reasonable assumption for the labels attached to the last stage units. Generalizations to restricted randomization of labeled units are clearly feasible (see C. R. Rao, 1970).

[8] Generalizations to vector attributes are obvious.

REFERENCES

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

Basu, D. (1969), An essay on the logical foundations of survey sampling, Part one, Technical Report No. 183, University of New Mexico, Albuquerque.

Fisher, R. A. and Yates, F. (1963), Statistical Tables for Biological, Agricultural, and Medical Research, 6th edition (1st edition 1938), Oliver and Boyd.

Godambe, V. P. (1955), A unified theory of sampling from finite population, *J. Roy. Statist. Soc., Series B, 17*, 269–277.

Godambe, V. P. and Joshi, V. M. (1965), Admissibility and Bayes estimation in sampling from finite populations-I, *Ann. Math. Statist., 26*, 1707–1722.

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

Godambe, V. P. (1970), Foundations of survey sampling, *The American Statistician, 24*, 33–38.

Hanurav, T. V. (1962), On Horvitz and Thompson estimator, *Sankhya, Series A, 24*, 429–436.

Hanurav, T. V. (1966), Some aspects of unified sampling theory, *Sankhya, Series A, 28*, 175–204.

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.
Kendall, M. G. and Smith, B. B. (1939), The problem of m rankings, *Ann. Math. Statist.*, *10*, 275–287.
Ramakrishnan, M. K., Some results on the comparison of sampling with and without replacement, *Sankhya*, *Series A*, *31*, 333–342.
Rao, C. R. (1970), Some aspects of statistical inference in problems of sampling from finite populations, Paper presented at the Symposium on Foundations of Statistical Inference, University of Waterloo, April 1970.
Rao, J. N. K. (1966), On the comparison of sampling with and without replacement, *Review of the International Statistical Institute*, *34*, 125–138.
Royall, R. M. (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.
Stevens, W. L. (1958), Sampling without replacement with probability proportional to size, *J. Roy. Statist. Soc.*, *Series B*, *20*, 393–397.

Mathematical Sciences and Social Sciences: Excerpts from the Report of a Panel of the Behavioral and Social Sciences Survey*

selected by WILLIAM H. KRUSKAL

Many problems in the behavioral and social sciences require mathematics, statistics, or computation for their solution. More and more frequently, social scientists are using methods and techniques from the mathematical sciences. Further, there is true interaction between the mathematical and the social sciences, for problems arising in the social sciences have motivated new theories and approaches in the mathematical sciences.

This interaction has a long record. Without trying to go back to its earliest history, a few examples are worth citing here. In the nineteenth century the psychologist G. T. Fechner was led to a variety of statistical problems through his early psychophysical investigations. Since at least the turn of the century, physical anthropology has both required results of, and made contributions to, the study of multivariate statistical methods as applied to body and skeleton measurements. Mathematics, more narrowly interpreted, has been an essential tool of economics at least since its use by Cournot and Léon Walras about one hundred years ago. Computation has become increasingly important to the social sciences with the advent of modern high-speed computing equipment, and computations are now routinely made that would have been impractical fantasies a few years ago. Two examples are input-output analysis in economics and the computations of quantitative linguistics. Geography has always had close connection with the making of maps, and cartography, in turn, has required mathematics, statistics, and computation in substantial ways.

The Mathematical Sciences Panel was established because of the close connections between the mathematical sciences and the social sciences. This panel report deals with germane problems of statistics, mathematics, and computation. We do not discuss relatively technical issues, important as they are, such as difficulties in carrying out true experiments for many social science problems.

The introductory material of the panel report next presents an outline of its contents, suggests other sources of information, and thanks the many mathematical and social scientists who helped in the report's preparation.

Chapter 1 then illustrates how the mathematical sciences interact with the social sciences in one interesting context. An extract from Chapter 1 follows.

The Mathematical Sciences at Work with the Social Sciences: Learning with Irregular Rewards

Conventional wisdom suggests that learning anything is best done if the learner is regularly and consistently rewarded for success, but not rewarded for failure. Indeed, much attention, both experimental and theoretical, has been given to learning situations with regular rewards. On the other hand, our lives have many aspects in which rewards are irregular, and in this chapter we discuss some ways in which the mathematical sciences help the social sciences to study learning with irregular rewards.

Two-Choice Experiments. Paychecks come at regular intervals for most of us, but other kinds of encouragement—being told that a job is well done, the joy of successfully finishing a long task, or seeing a child we have helped perform well—come at irregular intervals. Although some find these intervals too long, the ir-

* These selections are reprinted (with minor changes) by permission of the publisher from *Mathematical Sciences and Social Sciences*, edited by William H. Kruskal (copyright by Prentice-Hall, Inc., published in November 1970). This article appeared in the September 1970 issue of "Items" (Vol. 24, No. 3), published by the Social Science Research Council and reprinted here by permission. The Behavioral and Social Sciences Survey was jointly sponsored by the National Academy of Sciences–National Research Council and the Social Science Research Council and conducted by a central planning committee whose members were chairmen and cochairmen of panels in various social science fields. Mr. Kruskal was chairman of the Mathematical Sciences Panel; its other members were John P. Gilbert, Leo Katz, R. Duncan Luce, Alex Orden, and I. Richard Savage. Frederick Mosteller was an active participant in its work and a contributor to its report.