Ann. Inst. Statist. Math.
31 (1979), Part A, 253-269

ON STATISTICAL INFERENCE IN SAMPLE SURVEYS AND THE UNDERLYING ROLE OF RANDOMIZATION

J. C. KOOP

(Received Sept. 22, 1978; revised May 28, 1979)

Summary

Assuming only the existence of a universe and a frame that identifies its members (ultimate sampling units), and a minimum number of necessary but nonrestrictive assumptions, the writer derives a basic proposition which shows that for sample surveys any form of inference about any universe characteristic must depend on the sampling distribution of estimates generated by randomization, and, by direct implication, the sampling design. Unbiasedness is validated by this proposition, but likelihood appears to be in conflict with it, and by implication with randomization. The relation between high or low variance and correspondingly low or high probability for an estimator is also investigated in the paper. Finally it is argued that restrictions on randomization may be injurious to normality.

1. Introduction

Randomization, replication and control are the cornerstones of Fisher's [1] work, *The Design of Experiments*. Indeed, both the theory and practice of sample surveys and experimental designs have these three fundamental principles in common. In sample surveys control is usually exercised in the way of stratification, clustering and multistage sampling, and both Bowley [2] and Mahalanobis [3] regarded these procedures as restrictions on randomization. We note that Neyman in his 1934 paper [4] advocated support for all three principles; see also Fisher's comments in the discussion following this paper.

The early work of Gini and Galvani [5], which is commendably frank, points to the imperative need for proper randomization in order that selective bias may be avoided. Yates [6] discussed the rationale for randomization and replication in his 1946 paper.

This writer is of the view that for statistical inference in sample surveys the underlying role and the consequences of randomization,

replication, and restrictions on randomization still need to be investigated. Furthermore, Kempthorne's [7] recent remark that "there is a partial logical basis for accepting randomization," and that he does not understand this problem in the context of a finite population, adds justification for this investigation.

Except for a few necessary assumptions like the existence of the finite universe and the frame, no prior restrictive assumptions and appeals to superpopulation models are made. Recent work on estimation problems based on superpopulation models is adequately summarized in the book by Cassel, Särndal and Wretman [8].

We shall not be concerned with problems relating to the properties or the choice of estimators already dealt with by Midzuno [9], [10], Horvitz and Thompson [11], Koop [12], Godambe [13] and others who have avoided the use of model assumptions, regarding which, inference is impugned to the extent that such models are inappropriate. Specifically, the objectives of this paper are given in its summary.

2. Definitions, notation and explanations

To make this paper consistent with its objectives, and self-contained, it will be necessary to restate some of the terminology, definitions and notation given elsewhere (e.g., Godambe [13], [14], Koop [12], [15]). Some brief notes on aspects of randomization that have so far remained unnoticed will also be necessary. We shall also consider whether the probability of realizing a distinct sample, viz. $p(s)$ as defined later, can qualify as a likelihood function. Effectively, the entire section constitutes a form of basic theory needed for the development of the remaining sections.

2.1. *Definitions and notation*

There is a finite *universe* U (Bowley [16]) consisting of N different identifiable units where u_i is the ith unit, i.e.,

$$U=\{u_i:\ i=1, 2, \cdots, N\}\ , \tag{2.1}$$

with a set of N corresponding vectors of l real-valued components $(x, y, z, \cdots)$ for each unit, i.e.

$$\{(x_i, y_i, z_i, \cdots):\ i=1, 2, \cdots, N\}\ . \tag{2.2}$$

We shall assume that all these l-component vectors can be observed without error.

The *frame* F (United Nations [17]), which substantively exists in the way of lists and maps, identifies the units of U. We assume that F is perfect, i.e. through it every $u_i \in U$ can be identified as an *ulti-*

mate sampling unit.

Denote by s' a *sequence* of units from U formed through F according to some *rule* H. Remark 2 in Subsection 2.2 explains this rule.

The total number of units in s' will be denoted by $n(s')$.

Now we make the following

ASSUMPTION 2a. The rule H shall be such that every unit is contained in at least one s'.

A typical sequence s' that is realized by applying H may be given by

$$s' = \{u_{i_1}, u_{j_2}, \cdots, u_{m_{n(s')}}\} \tag{2.3}$$

in which the numbers $1, 2, \cdots, n(s')$ attached to the unit identification labels $i, j, \cdots, m$, respectively, indicate the order in the sequence. We shall call these numbers *marks of order.*

Conceptually, and apart from the question of randomization, *barring the complete census of N units*, that is the set (2.1), there are at most

$$\mathcal{N} = N + N^2 + \cdots + N^{N-1} = \{N(N^{N-1} - 1)\}/(N-1) \tag{2.4}$$

sequences of units with $n(s')$ varying from 1 to $N-1$ that could qualify as samples, in the usual sense of the term, for making inferences about U. However, because of restrictions on the way in which the units can be identified through F, like for example, the division of F itself into disjoint parts (resulting in strata for U), only a smaller collection of the number in (2.4) will qualify as possible samples when H is applied. Let us denote this collection of *logically possible sequences* by $\mathcal{S}'$, i.e.,

$$\mathcal{S}' = \{s' : (F, H)\} . \tag{2.5}$$

Although theoretically admissible, it is utterly impracticable, unless N is very small, to enumerate all members of $\mathcal{S}'$, then assign nonzero probability measures to each of them in order to select one or more of them, by a randomization procedure.

All that is needed for selecting any s' is to prescribe, in the context of the frame F and the rule H, sets of (nonzero) selection (or sometimes rejection) probabilities for the relevant units specific to each draw, and the appropriate *physical randomization procedure* R (e.g. devices such as fair dice and tested random numbers) to implement these probabilities. Such a finite collection of sets of probabilities, say P, may be called a *probability system* following an earlier terminology (Yamamoto [18]). It is easy to see that the number of such probability systems is unlimited. When P is given in detail then the probability of selecting a sequence s', i.e.,

(2.6) $$q(s')>0\ , \qquad \text{for all } s' \in \mathcal{S}'\ ,$$

can be computed for all relevant sequences.

Underlying the physical randomization procedure R there is the *principle of randomization*, which we shall denote by $\mathcal{R}$. Thus in some sense there is a connection between $\mathcal{R}$ and the probability system P through R. We have gone into these details because there is a need to distinguish between acts of physical randomization in selecting s', which may or may not be carried out with the probabilities prescribed by P, from the principle itself. We shall recur to this point in Section 3.

The choice of P is in some way restricted or influenced by F and H so that we may express it as an abstract function

(2.7) $$P=P(F, H)\ .$$

We can now say that the combination

(2.8) $$(P, R)$$

defines the *sampling procedure* for selecting samples from U with the object of estimating some function of the values in (2.2), e.g., $x_1+x_2+\cdots+x_N$. Without getting unduly involved, both in terminology and its underlying technicalities, let us say that there is a class of sampling procedures denoted simply by $\{(P, R)\}$.

In the context of (2.6), (2.7) and (2.8), s' as a realization specified by (2.3) is called a *particular sample.*

The collection $\mathcal{S}'$, now viewed in the context of (2.8), may be partitioned into a finite number of different subcollections, such that each subcollection say s^*, consists of one or more particular samples each having the same set of distinct units designated as s.

Any two subcollections are deemed to be *different* if and only if their respective constituent particular samples have at least one unit that is not common.

ASSUMPTION 2b. To avoid trivialities, we shall assume that there are at least two different subcollections in $\mathcal{S}'$.

Following Hájek [19], the relation between s and s' may be expressed by the abstract function

(2.9) $$s=s(s')\ , \qquad \text{for all } s' \in s^*\ .$$

The set s may be called a *distinct sample*, i.e., a sample in which repeated units and/or marks of order associated with the formation of s', through the sampling procedure (P, R), are disregarded.

The number of distinct units in s, denoted by $n(s)$, is called the

effective sample size.

The entire collection of s is designated $\mathcal{S}$, i.e.,

$$\mathcal{S}=\{s:\ s\subset U\}\ . \tag{2.10}$$

In view of (2.9), the total probability of realizing s, denoted by $p(s)$, is given by $p(s)=\sum_{s'\in s^*} q(s')$, for all $s\in\mathcal{S}$. We may also regard p as a function defined on $\mathcal{S}$ such that $\sum_{s\in\mathcal{S}} p(s)=1$. Godambe [20] calls p a *sampling design*; we shall follow his terminology.

2.2. *Explanatory and critical remarks*

Remark 1. In Godambe's work the notion of frame is almost totally ignored and appears to be completely ignored in all other papers which follow his approach. On the other hand, Basu [21] and Rao [22], each in his own way, recognizes this concept. It is the units identified by the frame that are randomized so that *without a frame no sample survey based on probability can take place.* Furthermore, in my view, any discussion of randomization without incorporating the frame as part of the logical set-up is, to say the least, unrealistic.

Remark 2. The rule H for the formation of sequences may incorporate one or more draw mechanisms such as the selection of units one at a time, either without replacement, with replacement, or both so that repeated units are allowed in the formation of s'. The rule may also include rejection procedures, for example, leading to what Hájek [23] describes as rejective sampling.

Remark 3. Without Assumption 2a, which is in the nature of a condition, the possibility of some units of U remaining unselected could arise, and this could vitiate the primary objective of sampling because, for example, we would not be able to estimate meaningfully the mean value of the units for some specified characteristic.

Remark 4. If F and H are completely specified then the number of sequences comprising $\mathcal{S}'$ can be computed. For example, if F shows that U is divided up into L strata each having N_h units ($h=1,2,\cdots,L$), for which $N=\sum_1^L N_h$, and H is the rule which specifies that n_h ($<N_h$ but >1) units shall be drawn one at a time without replacement from stratum h, for which $n=\sum_1^L n_h$, then the number of possible sequences is $\mathcal{N}'=\prod_1^L N_h(N_h-1)\cdots(N_h-\overline{n_h-1})$. Thus in the context of the given F and H, sequences other than this set of $\mathcal{N}'$ sequences are logically impossible and are naturally not members of the $\mathcal{S}'$ in question. It is easily seen that $\mathcal{N}\gg\mathcal{N}'$.

Remark 5. *It is important to note that each distinct sample s can only be realized through a selected particular sample s' that is an outcome of randomization which is the physical component of the sampling procedure* (P, R).

Remark 6. Godambe ([14], p. 249) notes that sampling designs or methods of randomization can be considered just different functions p defined on $\mathcal{S}$. However in the context of the foregoing discussion it is really q, given at (2.6), rather than p which is directly connected with the physical randomization procedure R. We may say that these observations relate to one aspect of randomization. An account of other related aspects follow in the next subsection.

2.3. *The extent of randomization*

Because randomization underlies much of sampling theory and practice, it is fitting in a paper such as this to at least define the extent of randomization and also consider the effect of restrictions on randomization, both of which are of deep significance to inference as will appear in Sections 3 and 4. Furthermore, the interplay of these three basic principles of randomization, replication and control will be evident in the notes which follow. There is no claim to originality. What appears to be interesting is *their interplay in a different form* in sample surveys culminating in the use of the fundamental concept of cardinality to describe the extent of randomization.

One aspect of the cardinality of a given $\mathcal{S}$ is that it is a comparative indicator of the extent of randomization. With a given type of restricted randomization, e.g., stratified random sampling, it is $\prod_1^L \binom{N_h}{n_h}$, whereas with more extended randomization, e.g., simple random sampling of an equal number of units from the undivided universe it is $\binom{N}{n}$.

Merely as a convention we shall say that in the entire class of sampling procedures $\{(P, R)\}$, unrestricted or *U-randomization* is achieved by a subclass of sampling procedures appropriate to any form of random sampling with replacement with $n(s)$ varying from 1 to $N-1$.

Almost in polarity with U-randomization, minimal or m-randomization is achieved by another subclass of sampling procedures usually described as "controlled selection" (Goodman and Kish [24]); systematic sampling in all its ramifications may be included in this subclass as also the methods investigated by Sukhatme and Avadhani [25] and Jessen [26]. The twin characteristics of m-randomization are that (i) the cardinalities of the respective $\mathcal{S}$ corresponding to the constituent sampling procedures are relatively very low and (ii) by and large, assumption free unbiased variance estimation is not possible with a single sample for

linear as well as nonlinear estimators.

Needless to stress there are many more subclasses of sampling procedures characterized by various levels of randomization.

Note that *purposive sampling* might have been considered as constituting a degenerate subclass with cardinality one if it had not been excluded by Assumption 2b.

2.4. *Does $p(s)$ qualify as a likelihood function?*

At least for the purpose of this paper it is necessary to examine this question even at the risk of arguments appearing to be facile. Godambe [13] in his 1966 paper initiated this line of thinking in the estimation of the population total $T=x_1+x_2+\cdots+x_N$. He says that the likelihood takes the value $p(s)$, for all populations corresponding to points in Euclidean space R_N, of which $x=(x_1, x_2, \cdots, x_N)$ is an element, which could have given rise to the sample data $\{s, x_i: i \in s\}$, and zero otherwise. Later on page 312 of this paper he remarks that "this by itself is of no avail from the point of view of estimating the population total T." Obviously this is because $p(s)$ is not a function of T.

Much earlier Fairfield Smith ([27], p. 29) said: "If single elements are 'sampled with equal probability' the probability distribution is

$$p(x)=\frac{1}{N}, \qquad x=X_1, \cdots, X_N."$$

Then on page 32 he states: "Since neither the specific parameters nor that single function of them ('variance') to be estimated, enters into the probability function, $p(x)$, *the method of maximum likelihood cannot be used to indicate a preferred parameter . . . .*" This contrasting viewpoint appears to have escaped notice; the italics are mine.

Clearly also $p(s)$ which is quite independent of parameters, or other universe values of interest, cannot be viewed as a likelihood function in the sense explained by Barnard, Jenkins and Winsten [28] on pp. 321–322 of their 1962 paper.

In contradistinction to this attempt to endow $p(s)$ with likelihood status, it is a recognizable fact in sample surveys that every member of the entire set of probabilities, $\{p(s): s \in \mathcal{S}\}$, is calculable as a numerical value. *As such, at least in principle, the $p(s)$-values qualify as data that are to be used for the purpose of inference until it can be proved that such values are irrelevant for the same purpose.*

These considerations rule out *likelihood as a principle for technical use* regarding inference about any characteristic of the universe U.

3. A basic proposition on inference

Sample surveyors are concerned most of the time with the estimation of means, totals, and other simple real functions of the multivariate values of the u_i's, viz. those values given by (2.2). As the concern of this paper is with principles and their consequences, it is appropriate to consider *inference about some general real valued function* $A(U)$, that is a function of all the values in (2.2) and *which subsumes all the estimands considered in the literature of sample surveys.*

We have as survey data the variate values $(x_i, y_i, z_i, \cdots)$ for each of the $n(s')$ members of the particular sample s' selected through the sampling procedure (P, R) with probability $q(s')$. In general, the possibility of repeated units in s' cannot be excluded. To estimate $A(U)$, suppose we construct an estimator $b(s')$, defined for all $s' \in \mathcal{S}'$, as a real function of all the variate values of s', recognizing order and multiplicity as information which cannot be ignored until it can be proved that both are irrelevant; this function will have a certain number of undetermined constants. Then using the same method of proof followed by Koop [15] it can be shown that, say

$$\bar{b} = \mathrm{E}\{b(s') \mid s\} = \sum_{s' \in s^*} q(s')b(s') \Big/ \sum_{s' \in s^*} q(s') ,$$

has a mean square error which is smaller or at most equal to that of $b(s')$, and in this sense an improvement on $b(s')$.

Therefore, in estimating $A(U)$ we choose a real function $a(s)$, defined for all $s \in \mathcal{S}$, that is a function of all the variate values of the distinct sample s and with a certain number of undetermined constants, and also such that $a(s)_{\min} < A(U) < a(s)_{\max}$. Examples of such estimating functions for which constants have already been determined by some principle or theory are means and regression coefficients. This function cannot be improved since its *conditional expectation, given the distinct sample*[1] s, is still $a(s)$.

By definition, the *sampling distribution* of $a(s)$, i.e., the *distribution generated by the randomization procedure* R, is given by the entire set of such values with their corresponding $p(s)$-values, viz.

$$\{(a(s), p(s)) : s \in \mathcal{S} \text{ and } a(s)_{\min} < A(U) < a(s)_{\max}\} . \tag{3.1}$$

Unlike its classical analogues, it is not characterized by parameters, unless by convention we choose to regard $A(U)$ as a parameter. Note

[1] The writer in his 1963 paper ([12], p. 201) viewed the concept of the *distinct sample* as the primitive analogue of the concept of sufficiency. However, it should be reported that Sir Ronald Fisher told him in 1961 that he did not think sufficiency applied to finite populations, but only to populations with distributional forms specified by parameters.

that the $p(s)$-values are automatically recognized as data in (3.1). Despite its structural difference from the conventional density functions, (3.1) in principle, has all the necessary elements for the expression of probabilities respecting $A(U)$; following Fisher [29], [30] it may also be called the *reference set.*

It should be noted that if the randomization component R in (2.8) is compromised, then the probability system P will also be compromised, so that for a given s we will not have $p(s)$ but some other value $p(s)+\Delta p(s)$ where $\Delta p(s)$ is unknown; this would impugn the claim that $a(s)$ is distributed according to (3.1). The extent to which the claim is impugned would depend on the extent to which R is compromised. All that can be done in the context of real world applications is to ensure that the devices comprising R, and equally the persons involved, perform faithfully, and this amounts to the adoption of the assumption that the principle of randomization is not violated.

To initiate inference about $A(U)$ we need in principle a *direct probabilistic criterion.* We have already ruled out likelihood on technical grounds in Section 2. We shall keep an open mind about the principles of unbiasedness and minimum variance and comment on them later, but we cannot use them now because they are not criteria which can be directly applied to the general estimating function $a(s)$ in order to yield immediate inferential results respecting $A(U)$.

The principle of the criterion advocated by Pitman ([31], p. 212) is reasonable and naturally qualifies as a direct probabilistic criterion. Given an estimator $a(s)$, we want to be able to estimate its constants and/or obtain guidance for the choice of P such that as many as possible of the $a(s)$ in (3.1) are as close as possible to $A(U)$. Needless to elaborate, this is the dominant practical requirement in sample surveys. Therefore consistent with this requirement, for some real positive number c we require

$$\text{(3.2)} \qquad \mathrm{P}\{|a(s)-A(U)|\leqq c\}$$

to be always nonzero and as close as possible to 1. A little later we shall evaluate (3.2).

By *statistical inference* or simply inference, we shall mean any statement consistent with the principle embodied by (3.2); this includes estimation of $A(U)$ and probability statements about $A(U)$.

Note that the statements leading to (3.2) carry two significant implications, viz.,

(i) The requirement at (3.2) is just a simple probability statement about the closeness of $a(s)$ to $A(U)$.

(ii) Additional to (i), the estimating function $a(s)$ may be chosen so that its corresponding probability can be made larger relative to

the choice of another function.

These implications are crucial to the developments which now follow.

Let us obtain an evaluation of (3.2) in the sense of implication (i) and then follow its logical consequences. According to the sampling distribution (3.1), the $(2r)$th moment of $a(s)$ about $A(U)$ is given by

$$\text{(3.3)}\quad \mathrm{E}\{a(s)-A(U)\}^{2r}=\sum_{s\in\mathcal{S}} p(s)\{a(s)-A(U)\}^{2r}, \qquad (r=1, 2, 3, \cdots).$$

The motivation for choosing the $(2r)$th moment is because it has $a(s)-A(U)$ as an argument and it is positive like (3.2).

Consider the division of $\mathcal{S}$ into disjoint parts $\mathcal{S}_e$, $\mathcal{S}_f$ and $\mathcal{S}_g$ such that

$$\text{(3.4)}\quad \begin{aligned} \mathcal{S}_e&=\{s:\ a(s)-A(U)<-c\}\ ,\\ \mathcal{S}_f&=\{s:\ |a(s)-A(U)|\leqq c\}\ ,\quad \text{and}\\ \mathcal{S}_g&=\{s:\ a(s)-A(U)>c\}\ . \end{aligned}$$

For (3.2) itself to be nonzero and in view of Assumption 2b, we must assume that for any given $a(s)$ the number c is chosen such that $\mathcal{S}_f$ has at least one member and so also $\mathcal{S}_e$ and/or $\mathcal{S}_g$. We have

$$\text{(3.5)}\quad \begin{aligned} &\sum_{s\in\mathcal{S}} p(s)[1-\{a(s)-A(U)\}^{2r}/c^{2r}]\\ &\quad=\sum_{s\in\mathcal{S}_e} p(s)[1-\{a(s)-A(U)\}^{2r}/c^{2r}]\\ &\qquad+\sum_{s\in\mathcal{S}_g} p(s)[1-\{a(s)-A(U)\}^{2r}/c^{2r}]\\ &\qquad+\sum_{s\in\mathcal{S}_f} p(s)[1-\{a(s)-A(U)\}^{2r}/c^{2r}]\ . \end{aligned}$$

On the right-hand side of (3.5) each of the first two expressions relating to summations over $\mathcal{S}_e$ and $\mathcal{S}_g$ are clearly negative; the term $-\sum_{s\in\mathcal{S}_f} p(s)\{a(s)-A(U)\}^{2r}/c^{2r}$ is either zero or negative. Let us factor out -1 in each of these three expressions and denote their sum, which is a positive valued polynomial of degree $2r$ in the entire set of $a(s)$-values, with c and appropriate $p(s)$-values playing the role of constants, by λ_{2r}.

Then we have

$$\text{(3.6)}\quad 1-\mathrm{E}\{a(s)-A(U)\}^{2r}/c^{2r}=\sum_{s\in\mathcal{S}_f} p(s)-\lambda_{2r}\ .$$

Furthermore noting that λ_{2r} is positive, we find from (3.6) that $0<\lambda_{2r}<\mathrm{E}\{a(s)-A(U)\}^{2r}/c^{2r}$. Thus in the sense of implication (i), for $r=1, 2, 3, \cdots$,

$$(3.7) \qquad \sum_{s \in S_f} p(s) = \mathrm{P}\{|a(s)-A(U)| \leqq c\} = 1+\lambda_{2r}-\mathrm{E}\{a(s)-A(U)\}^{2r}/c^{2r} .$$

We note that every specialization of the r-values in (3.7) always gives a statement of probability about the closeness of $a(s)$ to $A(U)$, that still functionally depends on all the variable elements of (3.1). Thus we arrive at the following

PROPOSITION. For sample surveys, any inference about $A(U)$, either by way of an estimate or by a statement of probability, must depend on the *sampling distribution of estimates*, and, by direct implication, the *sampling design*.

The statistician's traditional reliance on these two basic elements of sample survey methodology is therefore justified by this proposition, which so far, to my knowledge, has not been derived by anyone.

Incidentally, it may be noted that if we assume $\mathrm{E}\{a(s)\} = A(U)$ and set $r=1$ in (3.7) we will obtain almost immediately the Bienaymé-Tchebycheff inequality. Other analogues of this inequality immediately follow by setting $r=2, 3, \cdots$, including Karl Pearson's [32] generalization which was derived in a very different context. However, in order not to distract attention from the considerations which follow, these exercises are left to the reader.

The bearing of the proposition on the concepts of unbiasedness, variance and likelihood will now be considered.

Unbiasedness. Viewed in historical perspective unbiasedness is an independent principle of considerable utility; the entire theory of k-statistics invented by Fisher and its extensions to a finite universe by Irwin and Kendall [33], Tukey [34] and Wishart [35] rests on this principle. Still it is sometimes said that the principle, (i) is arbitrary and (ii) that it is not invariant under non-linear transformations. Now in sample surveys, to my knowledge, non-linear transformations of variables in (2.2) have not been used, and if at all used, such use must be extremely rare so that the criticism under (ii) is not really of practical significance. In regard to (i) the classical theory of sample surveys shows that the application of this principle in the estimation of polynomials always yields estimators whose functional forms are dependent on the sampling design. Therefore, on the basis of the proposition just stated inference through unbiasedness is justified so that *the principle itself can no longer be considered arbitrary for a finite universe for which a meaningful $A(U)$ always exists.* It is in this sense that the proposition validates the principle of unbiasedness giving it added significance for sample survey theory.

Variance. Next let us consider implication (ii) in respect to the choice of $a(s)$ hinging on the magnitude of the variance. Assuming

that $a(s)$ is unbiased, for a given sampling design p and for any two sets of estimators $\{a_1(s): s \in \mathcal{S}\}$ and $\{a_2(s): s \in \mathcal{S}\}$, or for simplicity a_1 and a_2 we find in an obvious notation, after putting $r=1$ in (3.7) with respect to each of these estimating functions, that

$$(3.8) \quad \mathrm{P}\{|a_1-A| \leqq c\} - \mathrm{P}\{|a_2-A| \leqq c\} = \{\mathrm{V}(a_2)-\mathrm{V}(a_1)\}/c^2+\lambda_2(a_1)-\lambda_2(a_2) .$$

Hence if $\mathrm{P}\{|a_1-A| \leqq c\} \geqq \mathrm{P}\{|a_2-A| \leqq c\}$, then $\mathrm{V}(a_1) \leqq \mathrm{V}(a_2)$ if and only if $\lambda_2(a_2) \geqq \lambda_2(a_1)$. The converse is also true with the λ-condition reversed. Note that the λ's will remain indeterminate. (Note also that this proposition and its converse are still true if instead of p we had p_1 and p_2 corresponding to a_1 and a_2.) *The point to be made here is that, it is only true that lower variance implies a relatively higher probability for the closeness of the estimate to the estimand and vice versa, only with the corresponding conditions on the λ's.* Similar conclusions hold for the mean square error (and indeed for all higher moments), when unbiasedness for the a's is no longer assumed. Some of these results are perhaps known in other contexts, but for sample surveys, this is an important clarification and provides further evidence for the value of the criterion (3.2). In *practice* all that can be done is to choose $a(s)$ so that its variance is as low as practicable or possible, hoping at the same time that the appropriate λ-condition is favorable for the achievement of a high probability. For linear estimators this objective is perhaps achieved by methods of sampling with probabilities proportional to known measures of size, but for nonlinear estimators appropriate methods have yet to be investigated.

Minimum variance. The reader may now ask: What is the status of the Gauss-Laplace principle of minimum variance vis-à-vis the proposition just proved? Koop [36] showed that when this principle is applied in linear estimation the resulting estimators depend both on the sampling design and the unknown universe values. Godambe and Joshi [37] showed that except for the uni-cluster sampling design (which is achieved by m-randomization), minimum variance unbiased estimators do not exist. The sense in which these best estimators do not exist was explained by Koop [15] in his 1974 paper for the more realistic problem of estimation when observations are subject to measurement and/or response errors. It is that the estimators depend both on the sampling design and unknown universe values. What we may usefully say is that this principle, in some sense, supports the foregoing proposition, because the results of attempted applications of the principle in the papers cited above all show that the sampling design is in no way redundant. We shall see later that this is contrary to what the likelihood principle asserts.

Likelihood principle. For the sake of argument let us accept likeli-

hood and the resulting likelihood principle. We reproduce the definition of this principle given by Barnard, Jenkins and Winsten [28] on page 323 of their paper. "The result x from (S, Ω, f) and the result y from (T, Ω, g) will give equivalent likelihood functions $f(x, \theta)$ and $g(y, \theta)$ if a number c exists independent of θ, such that

$$f(x, \theta)=cg(y, \theta)\ , \qquad (1.6)$$

and the likelihood principle says that, in this case, the inference from x about θ would be the same as the inference from y about θ." Here S and T represent the sample spaces of x and y respectively and Ω the parameter space of θ; f and g are the relevant likelihood functions. After announcing this principle the three authors hasten to sound a note of caution regarding its uncritical use. In the last three lines of this page we find Barnard's famous statement which forms the nucleus of Godambe's 1966 argument: "$\cdots$ *before* the experiment is carried out$\cdots$probabilities are relevant, *after* the experiment likelihoods are relevant." It may be noted that E. S. Pearson has questioned this reasoning on page 365 of the same paper. This contrary viewpoint also appears to have escaped notice. In regard to our sample survey problem, for any two sampling designs p_1 and p_2, yielding an identical distinct sample s, the ratio of the likelihoods, $p_1(s)/p_2(s)$, is a value depending only on F, H and P and obviously independent of $A(U)$. Nonetheless, according to the likelihood principle, as interpreted by Godambe [13] in his 1966 paper[2], inference about $A(U)$ should not depend on the sampling design thus contradicting the foregoing proposition, which is founded partly on randomization. This would seem to imply that the principles of likelihood and randomization are in conflict as far as sample surveys are concerned. *However, as we have not admitted likelihood status for $p(s)$ in Section 2, this disconcerting result does not really arise.* It is interesting to note that Fisher *did not mention likelihood* in his exposition of estimation problems in *The Design of Experiments, one of the cornerstones of which is randomization.*

Finally, as a conclusion we may say that any theory of statistical inference for a finite universe, which has places within its logical structure for the two elements generated by randomization, viz., (a) the sampling distribution of estimates, (b) the sampling design, and (c) the established principle of unbiasedness and (d) the notion of the distinct sample, is acceptable. Of course to any such theory other *desiderata* or principles, like for example consistency, admissibility, conditionality and ancillarity, which are not in logical contradiction to (a), (b), (c) and (d) and among themselves, can be admitted; note that (a), (b), (c) and (d)

[2] It should be noted that he still has serious reservations on this question as evidenced by his vigorous reply [38] to Basu [39].

are necessary though not sufficient for a viable theory of statistical inference for sample surveys that would include rules for the making of probability statements about any estimand. Such a theory remains to be formulated, and the considerations which follow in the next section will indicate that it is really needed.

4. Effect of restrictions on randomization

Despite Cochran's note of caution ([40], pp. 38–43) the traditional view persists that because many hundreds of ultimate sampling units are observed in most sample surveys, the distribution of an estimate, particularly when linear in the variates involved, will be somewhat like the normal distribution thereby justifying its use, for example, in setting confidence limits.

Regarding this traditional view, for complex large-scale sample surveys (considered in the logical framework of Section 2) the problem of $n(s)$ and N tending both to infinity, to provide the desired conditions for the generalized estimator $a(s)$ to achieve asymptotic normality, is complicated by the initial restrictions implicit in F and H and the resulting restrictions on randomization the nature of which was described in Subsection 2.3. In the writer's view it appears to be very difficult to derive central limit theorems within this framework.

The work of Hastings [41] for one-stage stratified sampling, and Koop [12] for two-stage sampling, both with equal probabilities without replacement, shows that normality can be seriously disturbed, as judged by the values of the skewness and/or kurtosis coefficients of the estimates involved for varying first-stage and second-stage sample sizes. Thus even from this evidence the effect of restrictions on randomization cannot be ignored.

An investigation of the kurtosis of the distribution of the generalized estimator $a(s)$, given by (3.1), throws considerable light into the nature of the problem for large-scale surveys where the effective sample size $n(s)$ is usually large, but certainly nowhere near infinity.

By the identity of Lagrange,

$$
\begin{aligned}
(4.1)\qquad & \sum_{s\in S} p(s) \sum_{s\in S} p(s)(a(s)-\mathrm{E}(a(s)))^4 \\
&= \Big\{ \sum_{s\in S} p(s)(a(s)-\mathrm{E}(a(s)))^2 \Big\}^2 \\
&\quad + \sum_{i>j} \{(a(s_i)-\mathrm{E}(a(s)))^2 (p(s_i)p(s_j))^{1/2} \\
&\qquad - (a(s_j)-\mathrm{E}(a(s)))^2 (p(s_j)p(s_i))^{1/2}\}^2 .
\end{aligned}
$$

In (4.1) we have attached subscripts i and j to s; these subscripts

should run from 1 to the value of the cardinality of $\mathcal{S}$ which will depend ultimately on F and H. If this cardinality is C then the summation indicated by $\sum_{i>j}$ is over $C(C-1)/2$ terms.

On the right-hand side of (4.1) the second term is zero if and only if

$$(4.2) \qquad (a(s)-\mathrm{E}(a(s)))^2=\text{a constant for all } s\in\mathcal{S}\ .$$

Dividing both sides of (4.1) by the square of the variance of $a(s)$ we find on simplification that the kurtosis of $a(s)$ is given by

$$(4.3) \qquad \beta_2\{a(s)\}=1+\Big\{\sum_{i>j}p(s_i)p(s_j)(a(s_i)-a(s_j))^2 \cdot (a(s_i)+a(s_j)-2\,\mathrm{E}(a(s)))^2\Big\}\Big/\mathrm{V}^2\{a(s)\}\ .$$

Note that the term following unity in (4.3) is always positive except when (4.2) holds in which case $\beta_2\{a(s)\}=1$; this result is certainly not trivial. For example when (i) the cardinality of $\mathcal{S}$ is $4M$ for which M is a very large integer, (ii) every $p(s)=1/4M$, (iii) M of the $a(s)$-values are each equal to h, (iv) M of them are equal to $k>h$, (v) M of them are equal to $h+(k-h)/g$, $g>1$ and (vi) the rest are each equal to $k+(k-h)/g$, then (4.2) holds when $g\to\infty$ and the required computations will show that $\beta_2=1$.

Generally (4.3) shows that the value of β_2 will depend on the sampling design p and the estimator $a(s)$, both of which are at choice. One is hard put to it to specify the conditions under which the second term of (4.3), say β_2', is equal to 2 in order that one of the prime requirements for approximate normality is achieved. At any rate it must be admitted that in the class

$$(4.4) \qquad \{U,(P,R)\colon (x_i,y_i,z_i,\cdots) \text{ is fixed for } i=1,2,\cdots,N\}$$

there must be many members for which $\beta_2'=2$. On the other hand there must also be many members for which $1\leqq\beta_2'<2$ or $\beta_2'>2$, and for this subclass we may say that it is the restrictions on randomization, resulting from the initial restrictions on F and H, which prevent the attainment of one of the conditions for approximate normality. Thus, even if the effective sample size relevant to (4.3) is very large, restrictions on randomization may very well prevent the attainment of normality for the estimator $a(s)$.

Acknowledgement

I thank Professor D. J. Finney, Dr. John W. Pratt, U Khint Maung and Professor William H. Kruskal for an exchange of correspondence

over the last few years all of which was most helpful and encouraging. The term *U*-randomization was suggested by Professor Finney in order to avoid derogatory implications respecting sampling procedures with severe constraints. Finally, I am very grateful to the editor and the referee for comments which improved the original paper and stimulated the discussion presented in Section 4.

RESEARCH TRIANGLE INSTITUTE, NORTH CAROLINA

REFERENCES

[1] Fisher, R. A. (1935). *The Design of Experiments*, Oliver and Boyd, Edinburgh.
[2] Bowley, A. L. (1926). Measurement of the precision attained in sampling, *Bull. Int. Statist. Inst.*, **22**, 6-62.
[3] Mahalanobis, P. C. (1944). On large-scale sample surveys, *Phil. Trans. R. Soc.*, B, **231**, 329-451.
[4] Neyman, J. (1934). On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection, *J. R. Statist. Soc.*, **97**, 588-625.
[5] Gini, C. and Galvani L. (1929). Di una applicazione del metoda rappresentativo all' ultimo censimento italiano della populazione, *Annali di Statistica*, **4**, 1-107.
[6] Yates, F. (1946). A review of recent statistical developments in sampling and sampling surveys, *J. R. Statist. Soc.*, **109**, 12-43.
[7] Kempthorne, O. (1973). Inference from experiments and randomization, *A Survey of Statistical Design and Linear Models* (ed. J. N. Srivastava), North Holland, Amersterdam, 303-331.
[8] Cassel, C.-M., Särndal, C.-E. and Wretman, J. H. (1977). *Foundations of Inference in Survey Sampling*, Wiley, New York.
[9] Midzuno, H. (1950). An outline of the theory of sampling systems, *Ann. Inst. Statist. Math.*, **1**, 149-156.
[10] Midzuno, H. (1952). On the sampling system with probability proportionate to the sum of sizes, *Ann. Inst. Statist. Math.*, **3**, 99-107.
[11] Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe, *J. Amer. Statist. Ass.*, **47**, 663-685.
[12] Koop, J. C. (1963). On the axioms of sample formation and their bearing on the construction of linear estimators in sampling theory for finite universes I, II, III, *Metrika*, **7**, 81-114 and 165-204.
[13] Godambe, V. P. (1966). A new approach to sampling from finite populations I, *J. R. Statist. Soc.*, B, **28**, 310-319.
[14] Godambe, V. P. (1969). A fiducial argument with applications in survey sampling, *J. R. Statist. Soc.*, B, **31**, 240-260.
[15] Koop, J. C. (1974). Notes for a unified theory of estimation for sample surveys taking into account response errors, *Metrika*, **21**, 19-39.
[16] Bowley, A. L. (1920). *Elements of Statistics*, King, London.
[17] United Nations (1948). *Recommendations for the Preparation of Sample Survey Reports*, Series C, No. 1, United Nations, New York.
[18] Yamamoto, S. (1955). On the theory of sampling with probabilities proportionate to given values, *Ann. Inst. Statist. Math.*, **7**, 25-38.
[19] Hájek, J. (1959). Optimum strategy and other problems in probability sampling, *Čás. pěst. Math.*, **84**, 387-423.
[20] Godambe, V. P. (1960). An admissible estimate for any sampling design, *Sankhyā*,

22, 285-288.

[21] Basu, D. (1971). An essay on the logical foundations of survey sampling, part one, *Foundations of Statistical Inference* (ed. V. P. Godambe and D. A. Sprott), Holt, Rinehart and Winston, Toronto, 203-242.

[22] Rao, C. R. (1975). Some problems in sample surveys, *Suppl. Adv. Appl. Prob.*, **7**, 50-61.

[23] Hájek, J. (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite population, *Ann. Math. Statist.*, **35**, 1491-1523.

[24] Goodman, R. and Kish, L. (1950). Controlled selection—a technique in probability sampling, *J. Amer. Statist. Ass.*, **45**, 350-372.

[25] Sukhatme, B. V. and Avadhani, M. S. (1965). Controlled selection a technique in random sampling, *Ann. Inst. Statist. Math.*, **17**, 15-28.

[26] Jessen, R. J. (1970). Probability sampling with marginal constraints, *J. Amer. Statist. Ass.*, **65**, 776-796.

[27] Smith, H. F. (1955). Variance components, finite populations, and experimental inference, *N. C. Inst. Statistics Mimeo Series*, 135.

[28] Barnard, G. A., Jenkins, G. M. and Winsten, C. B. (1962). Likelihood inference and time series (with discussion), *J. R. Statist. Soc.*, A, **125**, 321-372.

[29] Fisher, R. A. (1956). *Statistical Methods and Scientific Inference*, Oliver and Boyd, Edinburgh.

[30] Fisher, R. A. (1959). Mathematical probability in the natural sciences, *Metrika*, **2**, 1-10.

[31] Pitman, E. J. G. (1937). The "closest" estimates of statistical parameters, *Proc. Camb. Phil. Soc.*, **33**, 212-222.

[32] Pearson, K. (1919). On the generalized Tchebycheff theorem in the mathematical theory of statistics, *Biometrika*, **12**, 284-296.

[33] Irwin, J. O. and Kendall, M. G. (1944). Sampling moments for a finite population, *Ann. Eugenics*, **12**, 138-142.

[34] Tukey, J. W. (1950). Some sampling simplified, *J. Amer. Statist. Ass.*, **45**, 501-519.

[35] Wishart, J. (1952). Moment coefficients of the k-statistics in samples from a finite population, *Biometrika*, **39**, 1-13.

[36] Koop, J. C. (1957). Contributions to the general theory of sampling finite populations without replacement and with unequal probabilities, Ph.D Thesis, N. C. State Univ. (Also in *N. C. Inst. Statistics Mimeo Series*, 296, 1961.)

[37] Godambe, V. P. and Joshi, V. M. (1965). Admissibility and Bayes estimation in sampling from finite populations: I, *Ann. Math. Statist.*, **36**, 1707-1722.

[38] Godambe, V. P. (1975). A reply to my critics, *Sankhyā*, C, **37**, 53-76.

[39] Basu, D. (1969). Role of sufficiency and likelihood principles in sample survey theory, *Sankhyā*, A, **31**, 441-454.

[40] Cochran, W. G. (1963). *Sampling Techniques*, Wiley, New York.

[41] Hastings, W. K. (1974). Variance reduction and nonnormality, *Biometrika*, **61**, 143-149.