BAYESIAN SUFFICIENCY IN SURVEY-SAMPLING*

V. P. GODAMBE

(Received Dec. 7, 1967)

1. Introduction

If $p(\boldsymbol{x}|\boldsymbol{\theta})$ is the frequency function of the random variate $\boldsymbol{x}$ which is *known* excepting for the *unknown* parameter $\boldsymbol{\theta}$, assuming for convenience that $\boldsymbol{x}=(x_1, \cdots, x_n)$ takes values in the n-dimensional Euclidian space X and $\boldsymbol{\theta}=(\theta_1, \cdots, \theta_k)$ takes values in the k-dimensional Euclidian space Ω, we may ask the question: What part of the random variate $\boldsymbol{x}$ is relevant for inference concerning an arbitrarily specified (parametric) function ϕ on Ω, disregarding individual values $\theta_1, \cdots, \theta_k$ excepting to the extent they determine the value of the function ϕ? In other words, is there such a thing as a sufficient statistic $t(\boldsymbol{x})$, for $\phi(\boldsymbol{\theta})$, different than a sufficient statistic for $\boldsymbol{\theta}$ itself? Clearly this is a generalization of the situation of nuisance-parameter. The convertional sufficiency is defined for $\boldsymbol{\theta}$ itself as such an alternative definition of sufficiency, for the nuisance-parameter situation, based on a (or the?) group-structural relationship between the random-variate $\boldsymbol{x}$ and the parameter $\boldsymbol{\theta}$, has been put forward by Barnard [1] and Sprott [10]. In the following we give yet another definition of sufficiency for an arbitrary parametric function ϕ, which seems appropriate for the situations when our prior knowledge concerning parameter $\boldsymbol{\theta}$, can be characterized by a class C of prior distributions ξ, on the parametric space Ω. This new sufficiency would be called *Bayesian sufficiency* for the reasons which would be evident later on. This paper discusses applications of Bayesian sufficiency primarily for the problem of estimation as it arises in survey-sampling.

2. Preliminaries

With the same notation as above, for any prior distribution ξ on Ω

* Research supported in part by National Science Foundation Grant No. GP-6225 awarded to the Department of Statistics, The Johns Hopkins University. A part of this work was carried out during the author's employment (summer 1966) at the Department of Mathematics, University of Michigan.

we can write the posterior density (assuming it exists), of $\boldsymbol{\theta}$ as

$$(2.1) \qquad p_\xi(\boldsymbol{\theta}\,|\,\boldsymbol{x})=p(\boldsymbol{x}\,|\,\boldsymbol{\theta})\Big/\int_\Omega p(\boldsymbol{x}\,|\,\boldsymbol{\theta})d\xi\,, \qquad \text{for all } \boldsymbol{x}\in X,\ \boldsymbol{\theta}\in\Omega.$$

The posterior distribution of the parametric function ϕ on Ω can be obtained by integrating (2.1) on the set of relevant values of $\boldsymbol{\theta}$. We assume this posterior distribution of ϕ admits the density function,

$$(2.2) \qquad p_\xi(\phi\,|\,\boldsymbol{x})\,, \qquad \text{for all } \boldsymbol{x}\in X,\ \boldsymbol{\theta}\in\Omega\,.$$

DEFINITION 2.1. A (statistic) function $t(\boldsymbol{x})$ on the (sample) space X, is said to be *Bayes sufficient* for the (parametric) function ϕ on Ω, with respect to a class C of prior distributions ξ on the (parameter) space Ω, if the posterior density $p_\xi(\phi\,|\,\boldsymbol{x})$ in (2.2) depends on $\boldsymbol{x}$ only through $t(\boldsymbol{x})$ for all $\boldsymbol{x}\in X$ and $\xi\in C$.

The motivation for a *class* of prior distributions in the above Definition 2.1 is as follows: In practice almost never the prior knowledge completely specifies a prior distribution. On most of the occasions the prior knowledge just characterizes a few *properties* (such as bell-shapedness etc.) of a possible prior distribution. These *properties* define a *class*, C say, of prior distributing. This idea was first utilized by the author in [3] and subsequently in [7], but not via Bayes theorem. The Bayesian Sufficiency above aims at defining the part of Bayesian inference that is *common* for *all* the prior distributions contained in the class C. A realistic illustration from survey-sampling is discussed in sections 4 and 5.

After the reduction of the data by application of Bayesian sufficiency in section 6 we have suggested application of the principle of invariance. In spirit it is analogous to the usual practice of reduction of data through (conventional) sufficiency and invariance.

It is easy to check that if we put

$$p(\boldsymbol{x}\,|\,\boldsymbol{\theta})=(\sqrt{2\pi\sigma^2})^{-n}\exp\left\{-\frac{1}{2}\sum_1^n(x_i-\mu)^2/\sigma^2\right\}$$

where

$$\boldsymbol{x}=(x_1,\cdots,x_n) \quad\text{and}\quad \boldsymbol{\theta}=(\mu,\sigma^2)\,,$$

and suppose $\phi(\boldsymbol{\theta})=\sigma^2$, then if $\bar{x}=\sum_1^n x_i/n$ according to the above Definition $t(\boldsymbol{x})=\sum_1^n(x_i-\bar{x})^2$ is a *Bayes sufficient* statistic for $\phi(\boldsymbol{\theta})=\sigma^2$, for any class C of prior distributions ξ, such that μ and σ^2 when distributed as ξ, are independent, μ having uniform (degenerate) distribution over $(-\infty,\infty)$ and σ^2 having any arbitrary distribution over $(0,\infty)$.

Similarly if we put $p(\boldsymbol{x}|\boldsymbol{\theta})=(\sqrt{2\pi\sigma^2})^{-(n_1+n_2)}\exp\left\{-\frac{1}{2}\left[\sum_1^{n_1}(x_i-\mu_1)^2+\sum_{n_1+1}^{n_1+n_2}(x_i-\mu_2)^2\right]\Big/\sigma^2\right\}$ where $\boldsymbol{x}=(x_1, \cdots, x_{n_1+n_2})$ and $\boldsymbol{\theta}=(\mu_1, \mu_2, \sigma^2)$ and suppose $\phi(\boldsymbol{\theta})=\sigma^2$, then if $\bar{x}_1=\sum_1^{n_1}x_i/n_1$ and $\bar{x}_2=\sum_{n_1+1}^{n_1+n_2}x_i/n_2$, according to the above definition $t(\boldsymbol{x})=\sum_1^{n_1}(x_i-\bar{x}_1)^2+\sum_{n_1+1}^{n_1+n_2}(x_i-\bar{x}_2)^2$ is a *Bayes sufficient* statistic for $\phi(\boldsymbol{\theta})=\sigma^2$, for any class C of prior distributions ξ, such that μ_1, μ_2 and σ^2 when distributed as ξ are independent, μ_1, μ_2 having uniform (degenerate) distribution over $(-\infty, \infty)$ and σ^2 having some arbitrary distribution over $(0, \infty)$.

In the preceeding example, the exponent in $p(\boldsymbol{x}|\boldsymbol{\theta})$ can be written as

$$-\frac{1}{2\sigma^2}\left\{\sum_1^{n_1}(x_i-\bar{x}_1)^2+\sum_{n_1+1}^{n_1+n_2}(x_i-\bar{x}_2)^2+[A(\mu_1+\mu_2)+B]^2\right.$$
$$\left.+C[(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)]^2\right\}$$

where $A^2=n_1+n_2$, $B=-\{n_1[\bar{x}_1-(\mu_1-\mu_2)]+n_2[\bar{x}_2(\mu_1-\mu_2)]\}/\sqrt{n_1+n_2}$ and $C=n_1n_2/n_1+n_2$. Hence for the same class C of the priors as before, by transferring (μ_1, μ_2) into $(\mu_1+\mu_2, \mu_1-\mu_2)$ in the posterior and then integrating out $\mu_1+\mu_2$ we have the result: A Bayes sufficient statistic for the parametric function $\phi(\boldsymbol{\theta})=[(\mu_1-\mu_2), \sigma^2]$ is given by $t(\boldsymbol{x})=\left[(\bar{x}_1-\bar{x}_2), \sum_1^{n_1}(x_i-\bar{x}_1)^2+\sum_{n_1+1}^{n_1+n_2}(x_i-\bar{x}_2)^2\right]$.

A some what striking reduction of the data is achieved by the application of Bayes sufficiency to an estimation problem arising in survey-sampling. This is shown in the following sections.

3. Survey-sampling

Let the survey population U, consist of N units, denoted by the integers $i=1, \cdots, N$. If the variate value associated with the unit i is z_i, $i=1, \cdots, N$, then,

$$z=(z_1, \cdots, z_i, \cdots, z_N) \tag{3.1}$$

is a point in the N-dimensional Euclidian space R_N, and the *population total* $T(z)$ is a function on R_N given by

$$T(z)=\sum_1^N z_i\,. \tag{3.2}$$

Now in survey-sampling the word *sample* is used with some what different meaning than in the rest of statistical literature. We formally define a *sample* and a *sampling design* as follows:

DEFINITION 3.1. Any sub-set s of U (i.e. integers $1, \cdots, N$) $s \subset U$ is a *sample*.

DEFINITION 3.2. If S denotes the set of all possible samples s, then any real function p on S such that $p(s) \geqq 0$ and $\sum p(s)=1$ for $s \in S$ is called a *sampling design*.

It has been repeatedly shown by the author [3], [4] that all survey-sampling designs such as simple random sampling, stratification, arbitrary probability sampling and the like are special cases of the sampling design p in the Definition 3.2 above.

Here a general problem of estimation is to estimate the *population total* given by (3.2) on the basis of a sample s, (Definition 3.1) and the values z_i, $i \in s$ where s is drawn with a specified sampling-design (Definition 3.2).

DEFINITION 3.3. Any function $f(s, z)$ on $S \times R_N$ which depends on z in (3.1) only through those z_i for which $i \in s$ is called a *statistic*.

Now to obtain a Bayes sufficient *statistic* for the population total in (3.2) we note that in the present case the observation consists of s, $(z_i, i \in s)$. And the probability of s, $(z_i, i \in s)$ is completely determined by the sampling design p (Definition 3.2) and z in (3.1) as follows:

$$P[s, (z_i, i \in s) \mid z'] = \begin{cases} p(s) & \text{if } z' \in R_N(z_i, \ i \in s) \\ 0 & \text{if } z' \notin R_N(z_i, \ i \in s) \end{cases} \tag{3.3}$$

for all $s \in S$ and $z' \in R_N$, where $R_N(z_i, i \in s)$ is a sub-set of R_N, $R_N(z_i, i \in s) \subset R_N$, such that $z' = (z'_1, \cdots, z'_i, \cdots, z'_N) \in R_N(z_i, i \in s)$ if and only if $z'_i = z_i$ for all $i \in s$.

Remark. Thus here our sample-space X consists of all points s, $(z_i, i \in s)$ where $s \in S$ and z_i, $i \in s$ are any real numbers. The parametric space $\Omega = R_N$. For a specified sampling design p, (3.3) determines the probability distribution.

For a given prior ξ on R_N, the posterior distributions for z in (3.1) and $T(z)$ in (3.2) are derived in the next section. These correspond to the posterior distributions in (2.1) and (2.2) respectively. Subsequently, a Bayes sufficient statistic for $T(z)$ is obtained.

4. Posterior distribution of the population total

Without any loss of (statistical) generality we assume a *discrete* prior distribution ξ on R_N. Then denoting by $P_\xi(\cdot)$ and $P_\xi(\cdot|\cdot)$, the corresponding probability and conditional probability respectively when ξ obtains and the sampling design by p in (3.3), we have

$$P_\xi[z'|s,(z_i,\ i\in s)]=\frac{P_\xi[z',s,(z_i,\ i\in s)]}{P_\xi[s,(z_i,\ i\in s)]}=\frac{\Psi_\xi[z',z_i,\ i\in s]p(s)}{P_\xi[z_i,\ i\in s]p(s)} \tag{4.1}$$

where $p(s)$ is the same as in (3.3), $P_\xi[z_i\ i\in s]$ is the corresponding marginal probability obtained from $P_\xi(z')$ and

$$\Psi_\xi[z',(z_i,\ i\in s)]=\begin{cases}P_\xi(z') & \text{for } z'\in R_N(z_i,\ i\in s)\\ 0 & \text{for } z'\notin R_N(z_i,\ i\in s)\,.\end{cases} \tag{4.2}$$

$R_N(z_i,\ i\in s)$ being the same set as in (3.3). Now for samples s with $p(s)\neq 0$ we have from (4.1) and (4.2),

$$P_\xi[z'|s,(z_i,\ i\in s)]=\begin{cases}P_\xi(z')/P_\xi(z_i,\ i\in s) & \text{for } z'\in R_N(z_i,\ i\in s)\\ 0 & \text{for } z'\notin R_N(z_i,\ i\in s)\,.\end{cases} \tag{4.3}$$

Now (4.3) clearly defines the posterior distribution of z' given the prior ξ, the observation s, $(z_i, i\in s)$ and the sampling design* (3.3). Next, from the posterior distribution of z in (4.3) we can obtain the posterior distribution $T(z)$ in (3.2) in the usual manner by summing $P_\xi[z'|s,(z_i,\ i\in s)]$ over the relevant subset of R_N. An interesting simplification results by assuming ξ to be such that $z_1,\cdots,z_i,\cdots,z_n$ in (3.1) are mutually *independent* i.e.

$$P_\xi(z)=\prod_1^N P_\xi(z_i)\,. \tag{4.4}$$

Substituting (4.3) in (4.4) we have for samples s with $p(s)\neq 0$,

$$P_\xi[z'|s,(z_i,\ i\in s)]=\begin{cases}\prod_{i\notin s} P_\xi(z_i') & \text{for } z'\in R_N(z_i, i\in s)\\ 0 & \text{for } z'\notin R_N(z_i,\ i\in s)\,.\end{cases} \tag{4.5}$$

Now the posterior probability that $T(z)$ in (3.2) belongs to a set of values A is the same as $\sum_{i\notin s} z_i$ belongs to the set of values $A(\sum_{i\in s} z_i)$, where the set $A(\sum_{i\in s} z_i)$ is obtained from A by substracting from each element of A, $\sum_{i\in s} z_i$. Thus we have from (4.5)

* Note the posterior distribution of z in (4.3) is independent of the sampling design p. But we would not discuss the problem of randomization in this paper. For a relevant commet see the Appendix.

THEOREM 4.1. *For every prior* ξ *on* R_N, *such that* $z_1, \cdots, z_N$ *when distributed as* ξ *are mutually independent, the posterior distribution of the population total* T *in* (3.2) *conditional on* $[s, (z_i, i \in s)]$, *depends on* $(z_i, i \in s)$ *only through* $\sum_{i \in s} z_i$.

(4.6) Further suppose C is a class of priors ξ on R_N such that when $z_1, \cdots, z_N$ are distributed as ξ they are mutually independent, for all $\xi \in C$.

Then from Theorem 4.1, Definitions 2.1 and 3.3 we have

THEOREM 4.2. *With respect to the class* C *in* (4.6) *of the prior distribution on* R_N, *the statistic* $(s, \sum_{i \in s} z_i)$ *is Bayes sufficient* (*Definition* 2.1) *for the population total* T *in* (3.2).

The above theorem raises the following question: If as suggested by the theorem $(s, \sum_{i \in s} z_i)$ is the only *relevant* information in the data concerning the estimation of the population total T what is the logical meaning of the conventional standard errors of estimates of T? This question is not answered in this paper. However, the next section contains some related discussion.

5. A general discussion

The prior knowledge, characterized by a class C of the prior distributions in (4.6) implies that the variates $z_1, \cdots, z_N$ are causally independent. Now, the concept of causally independent variates is at the root of all physical sciences. The autonomy of a physical experiment is based on the assumption that the variates not controlled by the experiment are causally independent with the controlled variates. In fact one could probabilistically *define* the knowledge of *causal independence* of the variates $z_1, \cdots, z_N$ by *any* class C of prior distributions ξ on R_N such that $P_\xi(z_{i_1}, \cdots, z_{i_R} | z_{j_1}, \cdots, z_{j_l}) = P_\xi(z_{i_1}, \cdots, z_{i_R})$ where $z_{i_1}, \cdots, z_{i_R}$ and $z_{j_1}, \cdots, z_{i_R}$ are two non-intersecting subsets of the variates $z_1, \cdots, z_N$. One may now raise the question as to whether under the hypothesis of causal independence any inference can be made about the unseen values $z_{i_1}, \cdots, z_{i_R}$ on the basis of the observations $z_{j_1}, \cdots, z_{j_R}$. That such an inference is possible can be quickly illustrated in case the class C consists of *all* the prior distributions ξ such that $z_1, \cdots, z_N$ are independently and identically distributed. On the basis of the observations $z_{j_1}, \cdots, z_{j_R}$ one can have some estimate of ξ which can be used for inferring the unseen values $z_{i_1}, \cdots, z_{i_R}$. Again the class C is a parametric family of distributions, and, we can estimate some of the parameters from the observed values. In general, we can say that in spite of the assumption of the causal independence of the variates $z_1, \cdots, z_N$, observing some of these variates may sharpen our prior knowledge from class C of prior distributions to the class C^* where $C^* \subset C$. And this fact implies the possibility of inference concerning the unobserved values. The consideration of invariance discussed in the next section may suggest an alternative approach. Now in many cases of survey-sampling, the variates $x_1, \cdots, x_N$ are causally independent, in the sense described above, some possible exceptions being the situations where the populations are stratified.

6. Origin and scale invariance

Intuitively it is more appealing to do the reduction of the data *first* according to certain natural invariance properties of the model and *then* apply the sufficiency criteria for the further reduction. But in practice, generally, the reverse procedure is more convenient. An excellent reference in this direction is Hall, et al. [8].

In this section we consider the problem of finding a suitable point estimator for the population total T in (3.2). With slight change in the Definition 3.3, we have

DEFINITION 6.1. Any *real* function $e(s, z)$ on $S \times R_N$, which depends on z only through those z_i for which $i \in s$, is called an *estimator*.

DEFINITION 6.2. Any estimator e is called a Bayes sufficient estimator for the population total T in (3.2) if e depends on $[s, (z_i,\ i \in s)]$ only through the Bayes sufficient statistic $[s, \sum_{i \in s} z_i]$ in Theorem 4.2.

We now introduce the concept of the origin and scale invariant estimation for the population total as follows:

DEFINITION 6.3. Any estimator $e(s, z)$ is said to be *origin invariant* for the population total $T(z)$ if and only if for any real constant k, $(-\infty < k < \infty)$, $e(s, z + k\bar{z}) = e(s, z) + Nk$.

DEFINITION 6.4. Any estimator $e(s, z)$ is said to be *scale invariant* for the population total $T(z)$ if and only if for any real constant k, $(-\infty < k < \infty)$, $e(s, kz) = k \cdot e(s, z)$.

It seems that certain symmetries implied in some possible prior knowledge about the population may make the above properties of invariance intruitively appealing. Any way, it is interesting to note that Bayes sufficiency together with the concept of invariance introduced by Definitions 6.3. and 6.4 implys a *unique* point estimator $\bar{e}(s, z) = N\bar{z}$, $\bar{z}$ being the usual *sample mean*. This is proved below.

By Definition 6.2 any *Bayes sufficient estimator* for the population total T is given by $e(s, \sum_{i\in s} z_i)$. The *origin invariance* by Definition 6.3 inplies $e(s, \sum_{i\in s}(z_i+k))=e(s, \sum_{i\in s} z_i)+Nk$ for all values of z_i, $i \in s$ and k. Putting $\sum_{i\in s} z_i=0$, we have $e(s, n(s)k)=e(s, o)+Nk$ for all k, $n(s)$ being the number of individuals i, $(i=1, \cdots, N)$, such that $i \in s$. Hence, *origin invariant Bayes sufficient* estimator for the population total must be of the form:

$$e(s, z)=\text{const.}+N\bar{z}(s) \tag{6.1}$$

where $\bar{z}(s)=\sum_{i\in s} z_i/n(s)$. Next it is easy to see that among the class of estimators obtained from (6.1) by giving "const." different values the only one which is *scale invariant* by Definition 6.4, is given by

$$e(s, z)=N\bar{z}(s)\,. \tag{6.2}$$

Hence we have the following,

THEOREM 6.1. *The estimator* $N\bar{z}(s)$ *is uniquely the Bayes sufficient, origin and scale invariant* (*Definitions* 6.2, 6.3, *and* 6.4) *estimator for the population total* T *in* (3.2).

On the other hand, if our prior knowledge about the population contains knowledge of some ancillary variate y taking values y_i, $i=1, \cdots, N$ for different individuals i of the population, we may wish to find an estimator which is *Bayes sufficient, origin invariant* (Definitions 6.2, 6.3) and which assumes *exactly* the value $T(\boldsymbol{y})$ at $z=\boldsymbol{y}=(y_1, \cdots, y_N)$ i.e. the estimator should be such that at the point $z=\boldsymbol{y}$, $e(s, z)=T(z)$. A unique estimator satisfying this condition is obtained from (6.1) as follows,

$$e(s, \boldsymbol{y})=\text{const.}+N\bar{y}(s) \tag{6.3}$$

where $\bar{y}(s)$ is the mean of y_i, $i \in s$. And since we have $e(s, \boldsymbol{y})=T(\boldsymbol{y})$ we have from (6.3)

$$\text{const.}=T(\boldsymbol{y})-N\bar{y}(s)\,. \tag{6.4}$$

Putting $T(\boldsymbol{y})=N\bar{Y}$, $\bar{Y}$ denoting the population mean of y's, we have from (6.1), (6.3) and (6.4)

$$e(s, z)=N(\bar{z}(s)-\bar{y}(s))+N\bar{Y}\,. \tag{6.5}$$

This is what is conventionally known as the *difference estimator* for the population total. Thus we have the following,

THEOREM 6.2. *The* (*difference*) *estimator, given by* (6.5), *is the unique Bayes sufficient, origin invariant* (*Definitions* 6.2, 6.3) *estimator which is exactly equal to the population total at* $z=\boldsymbol{y}$.

Now conventionally the estimator

$$e(s, z)=\frac{\bar{z}(s)}{\bar{y}(s)}T(y) \tag{6.6}$$

is called the ratio-estimator of the population total. Following arguments similar to those of Theorem 6.2, we have

THEOREM 6.3. *The ratio-estimator given by* (6.6) *is the unique Bayes sufficient scale invariant* (*Definitions* 6.2, 6.4) *estimator which is exactly equal to the population total at* $z=y$.

Evidently the Theorems 6.2 and 6.3 above would be relevant for the situations where we have a reason to believe that the values z_i, $i=1, \cdots, N$ are obtained from some previously observed values y_i, $i=1, \cdots, N$ by just changing the origin or scale of measurement and by adding to these changed values some *small* random fluctuations *independently* for each i, $i=1, \cdots, N$.

We conclude this paper by just raising a question: Are the concepts of *origin* and *scale* invariance discussed above logically equivalent to choosing some special class of prior distributions in (4.6)?

(The relationship of the above Theorems 6.1, 6.2 and 6.3 with the author's previous results ([7], Theorem 4.1) is obvious).

7. Stratification

The case of what is conventionally called a stratified sampling is specially interesting. Clearly the usual estimator

$$\sum N_k \bar{z}_k \, ,$$

($\bar{z}_k$ being the sample mean and N_k the number of units, in the kth stratum $k=1, 2, \cdots$) for the population total T in (3.2) is not a function of the statistic $(s, \sum_{i \in s} z_i)$ referred to the Theorem 4.2.

To understand the above situation one should note that stratification invariably is based on some kind of *prior knowledge* concerning the variate values z_i associated with different units $i=1, \cdots, N$. On the basis of this prior knowledge we *expect* the variate values z_i, associated with some of the units to be larger or smaller than these associated with some other units in the population. (Thus broadly speaking, a stratum consists of units i which we *expect* to have more or less equal variate values z_i). Clearly such a *prior knowledge* implies some prior distribution ξ on R_N so that different coordinates of $z=(z_i, \cdots, z_N)$ in (3.1), are probabilistically *dependent*. That is, $P_\xi(z)$ would *not* satisfy the equation (4.4). This situation therefore is *not* covered by the Theorems 4.1 and 4.2.

On the other hand, the situations described in the section 5 were different. These are generally characterized *either* by more or less complete absence of any prior knowledge concerning the variate values associated with different units *or* when the prior knowledge provides the values y_i for units $i=1, \cdots, N$, of some auxiliary variate. The prior knowledge of y_i, $i=1, \cdots, N$ suggests a prior distribution ξ on R_N such that

$$(7.1) \qquad P_\xi(z_{i_1}, \cdots, z_{i_R} | z_{j_1}, \cdots, z_{j_l}; y_1, \cdots, y_N) = P_\xi(z_{i_1}, \cdots, z_{i_R} | y_1, \cdots, y_N)$$

where $z_{i_1}, \cdots, z_{i_R}$ and $z_{j_1}, \cdots, z_{j_l}$ denote *any* two non-interesting subsets of the variates $z_1, \cdots, z_N$ in (3.1). And (7.1) implies

$$(7.2) \qquad P_\xi(z | y_1, \cdots, y_N) = \prod_{i=1}^{N} P_\xi(z_i | y_1, \cdots, y_N) .$$

According to (7.2), the prior distribution ξ on R_N is such that conditioned on the values $(y_1, \cdots, y_N)$, $z_1, \cdots, z_N$ are probabilistically independent. Hence, we get the validity of Theorems 4.1 and 4.2.

Now the absence of any auxiliary variate values y_i, $i=1, \cdots, N$, as is generally the case in connection with stratified sampling, may some times mean that our prior knowledge implies some distribution on the values $y_1, \cdots, y_N$. With respect to this distribution we may integrate out $y_1, \cdots, y_N$ in (7.2). Usually the resulting distribution $P_\xi(z)$ would be such that $z_1, \cdots, z_N$ are probabilistically *dependent*. This case as stated before, is not covered by Theorems 4.1 and 4.2.

Of course the Theorems 4.1 and 4.2 could clearly be applied for each stratum separately on the assumption that we have no prior knowledge to distinguish between the values of the components of z belonging to any given stratum. This condition is usualy fulfilled when stratification is carried out to the fullest extent.

Acknowledgement

A some what restrictive version of the definition of Bayesian Sufficiency in the section 2 of this paper was previously given by Raiffa and schlaifer [9]. Some discussions with Professor A. Birnbaum and L. J. Savage were helpful. Needless to say, however, that none of the above persons are responsible for the views expressed in this paper.

Appendix

As we have already said (foot-note, p. 367) the problem of randomization is not discussed in this paper. But to make this paper self-contained the following interpretation (suggested by Joshi (ref. Godambe [6])) and Beran [2], of random sampling as *means* of validating the assumption of independence in (4.6) may suffice. Consider a possibility (and this can happen in many practical situation) of $z_1, \cdots, z_N$, the realized values of the random variates satisfying (4.6), being *arraged* in some *order* by an intelligent agency before a 'sample' is drawn. Clearly this arrangement may offset the assumption of probabilistic *independence* in (4.6) on which all the subsequent analysis is based. A simple way to restore this independence is through random permutations of the units $(1, \cdots, N)$, before the sample is drawn. But this is logically equivalent to a *simple random sampling.*

THE JOHNS HOPKINS UNIVERSITY, BALTIMORE, MARYLAND

REFERENCES

[1] G. A. Barnard, "Some logical aspects of the fiducial argument," *J. R. Statist. Soc., Ser. B,* 25 (1963), 111-115.

[2] R. J. W. Beran, *A Personal Communication,* 1966.

[3] V. P. Godambe, "A unified theory of sampling from finite populations," *J. R. Statist. Soc., Ser. B,* 17 (1955), 268-278.

[4] V. P. Godambe, "A review of the contributions towards a unified theory of sampling from finite populations," *Rev. Int. Statist. Inst.,* 33:2 (1965), 238-257.

[5] V. P. Godambe, "Bayesian sufficiency in survey-sampling," *Ann. Math. Statist.,* 37, No. 5 (1966), 1414-1415, Abstract.

[6] V. P. Godambe, "Bayes and empirical Bayes estimation in sampling finite populations," *Technical Report,* Department of Statistics, The Johns Hopkins University, No. 41 (1966a).

[7] V. P. Godambe, "A new approach to sampling from finite population—I, II," *J. R. Statist. Soc., Ser. B,* 28, No. 2 (1966b), 310-328.

[8] W. J. Hall, R. A. Wijsman and J. K. Ghosh, "The relationship between sufficiency and invariance with applications in sequential analysis," *Ann. Math. Statist.,* 36 (1965), 575-614.

[9] H. Raiffa and R. O. Schlaifer, *Applied Statistical Decision Theory,* Boston: Division of Research, Graduate School of Business Administration, Harvard University, 1961.

[10] D. A. Sprott, "Transformations and sufficiency," *J. R. Statist. Soc,, Ser. B,* 27 (1965), 479-485.