

Subjective Bayesian Models in Sampling Finite Populations

By W. A. ERICSON

University of Michigan

[Read before the ROYAL STATISTICAL SOCIETY at a meeting organized by the RESEARCH METHODS SECTION on Wednesday, March 12th, 1969, Professor R. L. PLACKETT in the Chair]

SUMMARY

A general and basic model for inference about characteristics of a finite population of distinguishable elements is presented from a subjectivistic-Bayesian point of view. A subjectivist analogue to simple random sampling, based on the notion of exchangeable random variables, is discussed and the inputs and assumptions underlying the model are shown to involve nothing more than is required for inference under Bayesian models for infinite populations. The model is illustrated by a number of particular examples including one based on the multinomial distribution which incorporates a prior distribution representing an extreme position of initial ignorance. Inferences under this particular model are shown to agree closely in several respects with usual "classical" results. Finally, an extension of the results is presented involving the use of concomitant measurements, and under this Bayesian model several common ratio and regression estimators are shown to arise as means of posterior distributions.

1. PRELIMINARIES

FOLLOWING several recent writers, Godambe (1955), Hájek (1959), Godambe (1965), and others, we define a finite population of N distinguishable elements labelled by the integers $1, 2, \dots, N$. We let $\mathcal{N} = \{1, 2, \dots, N\}$, the label set, and $\mathbf{X} \equiv (X_1, X_2, \dots, X_N)$, where X_i is the unknown value of some characteristic possessed by the i th population element. The unknown X_i can be taken as vector valued, a case of considerable practical importance, though only the scalar case will be treated here. Inference concerns the N -dimensional real vector parameter \mathbf{X} or, more realistically, some simple function $g(\mathbf{X})$ of \mathbf{X} say. Here we will be mainly concerned with the simple functions $T = \sum_1^N X_i$, $\mu = T/N$, and $\sigma^2 = 1/N \sum_1^N (X_i - \mu)^2$, the population total, mean, and variance.

A *sample* of size m , s^* is defined quite generally to be an ordered sequence of m of the population elements $i_1^*, i_2^*, \dots, i_m^*$ ($i_j^* \in \mathcal{N}$, $j = 1, \dots, m$, repetitions allowed) together with the sequence of their associated observed characteristic values $\mathbf{x}^* = (x_{i^*(1)}, x_{i^*(2)}, \dots, x_{i^*(m)})$, i.e. we observe for each i_j^* that $X_{i^*(j)} = x_{i^*(j)}$, $j = 1, 2, \dots, m$. While it is therefore assumed that if element j is included in the sample then the value of X_j becomes known with certainty the model is being extended to weaken this restriction and thereby incorporate response error, bias, and non-response.

A *sample design* is then defined by some countable set S^* of ordered sequences, s^* , together with a probability measure assigned by choosing a function $p(s^*) \geq 0$, $\sum_{s^* \in S^*} p(s^*) = 1$, where $p(s^*)$ is the probability of choosing the sample s^* . That any such sample design may be implemented by an element by element sampling procedure has been demonstrated by Hanurav (1962).

For any such sample (s^*, \mathbf{x}^*) we define the statistic (s, \mathbf{x}) to be the set of indices of *distinct* population elements, $s = \{i_1, \dots, i_n\} \subseteq \mathcal{N}$, included in the observed sequence s^* together with the observed values x_j of $X_j, j \in s$. For notational convenience, given any sample s^* containing the n distinct units $s = \{i_1, \dots, i_n\}$, we define the (matrix) operator \mathbf{S} such that $\mathbf{S}(\mathbf{X}) = (X_{i(1)}, \dots, X_{i(n)})$ (for definiteness we assume $i_1 < i_2 < \dots < i_n$); the complementary operator $\bar{\mathbf{S}}$ such that

$$\bar{\mathbf{S}}(\mathbf{X}) = (X_{j(1)}, X_{j(2)}, \dots, X_{j(N-n)})$$

for all $j_i \in \mathcal{N} - s, (j_1 < j_2 < \dots < j_{N-n})$; and the vector $\mathbf{x} = (x_{i(1)}, \dots, x_{i(n)})$ of observed values of $\mathbf{S}(\mathbf{X})$.

It is obvious under this model that for any joint prior probability distribution of the vector parameter \mathbf{X} given by the general density $p'(X_1, \dots, X_N)$ the posterior probability distribution of \mathbf{X} given the sample (s^*, \mathbf{x}^*) is precisely the same as that given only the statistic (s, \mathbf{x}) . It then follows immediately from the Bayesian definition of sufficiency that (s, \mathbf{x}) is a sufficient statistic, a fact previously demonstrated or noted using the more usual (and equivalent) definitions by Basu (1958), Hájek (1959), and others.

It also follows that given (s^*, \mathbf{x}^*) the likelihood function of \mathbf{X} (the likelihood function being unique only up to an arbitrary positive multiplicative constant) is given by

$$l\{\mathbf{X}; (s^*, \mathbf{x}^*)\} = l\{\mathbf{X}; (s, \mathbf{x})\} = \begin{cases} kp(s^*), & \text{for } \mathbf{X} | \mathbf{S}(\mathbf{X}) = \mathbf{x}, \\ 0, & \text{otherwise,} \end{cases} \tag{1}$$

where $k > 0$ is an arbitrary constant.

Hence given a joint N -dimensional prior on \mathbf{X} , with density $p'(\mathbf{X})$, posterior distribution of \mathbf{X} conditional on a sample of the sort described above has a density given by

$$p\{\mathbf{X} | (s, \mathbf{x})\} \propto \begin{cases} p(s^*) p'(\mathbf{X}) & \text{for } \mathbf{X} | \mathbf{S}(\mathbf{X}) = \mathbf{x}, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

(Here and in the sequel, for notational convenience and where it will not lead to confusion, we will use the same symbol, for instance, \mathbf{X} , to indicate both a random variable and its generic value.)

Finally, if $p(s^*)$ is independent of \mathbf{X} (the case to be assumed here) then this posterior density is given by

$$p\{\mathbf{X} | (s, \mathbf{x})\} = \begin{cases} p'(\mathbf{X})/p'_{\mathbf{S}(\mathbf{X})}(\mathbf{x}), & \text{for } \mathbf{X} | \mathbf{S}(\mathbf{X}) = \mathbf{x}, \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

where $p'_{\mathbf{S}(\mathbf{X})}(\mathbf{x}) \neq 0$ is just the marginal prior density of $\mathbf{S}(\mathbf{X})$.

Several writers, notably Godambe (1966), have viewed the likelihood function, (1), as being almost, if not completely, uninformative and have adopted principles of inference eschewing the likelihood principle. Our view here, quite to the contrary, is that when reasonable prior distributions are introduced, their revision by sample data can lead to meaningful and useful inferences on those functions of \mathbf{X} which are typically of interest. No new principles of inference are necessary.

In this paper we examine several instances of what we consider to be the simplest useful class of prior distributions—those reflecting exchangeable or symmetrically dependent opinions regarding the X_j 's. Such priors seem, in some sense, to yield a subjectivist analogue to inference under simple random sampling. Some basic notions and discussion of subjective Bayesian views in sampling are given in the next section while specific examples and results are given in Sections 3 and 4. Section 5 presents some results on the use of auxiliary information in setting priors.

Before proceeding we note the following earlier relevant work. One of the earliest published examples of a posterior distribution of a mean or total of a finite population appears in a paper of Karl Pearson (1928). In this paper Pearson gives the posterior distribution of the number or proportion of elements possessing some attribute using a "diffuse" prior, by essentially normalizing the hypergeometric likelihood function. Pearson's result is a special case of the results of Section 4. More recently, Aggarwal (1959, 1966) has used some normal distribution priors, mainly as a technique for generating minimax estimators in finite populations. Some of his Bayes estimators are special cases of those given in Section 3.

Hill (1968) has considered the posterior distributions of means and percentiles of finite and infinite populations using quite a different model from that of the present paper. Also, independently, Roberts (1967), using a model similar to that given below, has obtained a few of the results given in Sections 3.1 and 3.2.

Other writers, notably Cochran (1939, 1946), Godambe (1955 and later), Hájek (1959), have used prior distributions (or their equivalents) in sampling theory. These uses have been almost exclusively in discussion of optimum design strategy and not with a view to obtaining useful posterior distributions.

2. EXCHANGEABLE PRIORS

The view taken here is a purely subjectivistic Bayesian one which essentially views statistical inference as a process of revision, by relevant evidence, of one's degrees of belief or ignorance as measured by subjective probability. That is, in the model outlined in the preceding section, $p'(X)$ represents one's initial betting odds on the X_i 's in the subjective probability sense of de Finetti, Savage, and others. Perhaps the simplest class of prior distributions is that given by taking $p'(X) = \prod_1^N p'_i(X_i)$, that is, by viewing the X_i 's as independent *a priori*. The meaning and consequences of such a prior are clear—given a sample we learn for certain that $X_i = x_i$ for $i \in s$ but such information does not alter our opinions in any way regarding the non-sampled X_i 's. It seems doubtful that there are many, if any, real problems in which one's real opinions would be so expressible.

The simplest useful class of prior distributions in this situation would seem to be that in which the X_i 's are viewed as exchangeable random variables. This notion was introduced by de Finetti and discussions are available in de Finetti (1937) and briefly in Feller (1966). De Finetti's monograph is available in English translation and with new notes in Kyburg and Smokler (1964.) The notion is one of symmetry: random variables X_1, \dots, X_N are said to be exchangeable if each of the $N!$ permutations, $X_{i(1)}, \dots, X_{i(N)}$, has the same joint probability distribution. Exchangeability thus expresses the prior knowledge that while the units of the finite population are identifiable by their labels (here the integers $1, 2, \dots, N$) there is no information carried by these labels regarding the associated X_i 's; that is, under exchangeability, given $1 \leq r \leq N$, one's initial betting behaviour regarding events defined by the unknown quantities $X_{i(1)}, \dots, X_{i(r)}$ is invariant over the ordered sets of indices i_1, \dots, i_r .

Alternatively, exchangeability is akin to viewing the finite population as being effectively randomized. I believe that the notion of exchangeability and exchangeable prior distributions very closely approximates the real opinions of thoughtful “classical” practitioners in many situations where they deem simple random sampling to be appropriate.

Assuming that the X_i 's are viewed as exchangeable, one still faces the problem of assigning the N -dimensional joint prior distribution, $p'(\mathbf{X})$. One may be aided in this task by noting that a wide class of prior distributions $p'(\mathbf{X})$ representing exchangeability may be generated by viewing the X_i 's as independent, identically distributed *conditional* on some real or hypothetical parameter $\theta = (\theta_1, \dots, \theta_k)$ with general density $p(X_i|\theta)$ and where θ is assigned the probability distribution function $F(\theta)$. The joint prior density $p'(\mathbf{X})$ is then taken as the marginal distribution of \mathbf{X} given by the mixture

$$p'(\mathbf{X}) = \int_{\theta} \prod_{i=1}^N p(X_i|\theta) dF(\theta). \quad (4)$$

The generation of a joint prior distribution by this approach is, barring differences in probabilistic interpretation, equivalent to viewing the finite population as a sample from an infinite superpopulation having unknown parameter θ . This notion has been used previously in sampling theory. (See Cochran, 1939, 1946, for example.) Here, unlike earlier uses of the superpopulation concept, the parameter θ is itself assigned a subjective probability distribution. This “superpopulation” notion seems more generally reasonable from the present viewpoint than under objectivist interpretations of probability.

2.1. Posterior Distribution

Combining a prior distribution of the sort generated by (4) with the likelihood function (1) resulting from a sample (s^*, \mathbf{x}^*) which yields the sufficient statistic (s, \mathbf{x}) we find that the posterior distribution of \mathbf{X} is given by the density

$$p\{\mathbf{X}|(s, \mathbf{x})\} \propto \begin{cases} \int_{\theta} \prod_{i \notin s} p(X_i|\theta) \prod_{i \in s} p(x_i|\theta) dF(\theta), & \text{for } \mathbf{X}|\mathbf{S}(\mathbf{X}) = \mathbf{x}, \\ 0, & \text{elsewhere.} \end{cases} \quad (5)$$

Noting that the posterior distribution of the “superpopulation” parameter θ , is proportional to $\prod_{i \in s} p(x_i|\theta) dF(\theta)$, the posterior on \mathbf{X} may be expressed as

$$p\{\mathbf{X}|(s, \mathbf{x})\} = \begin{cases} \int_{\theta} \prod_{i \notin s} p(X_i|\theta) dF(\theta|\mathbf{x}), & \mathbf{X}|\mathbf{S}(\mathbf{X}) = \mathbf{x}, \\ 0, & \text{otherwise,} \end{cases} \quad (6)$$

where $F(\theta|\mathbf{x})$ is the posterior distribution function of θ . Thus it is seen that one's posterior opinions regarding the unobserved X_i 's are also exchangeable but now with a density yielded by a mixing distribution equal to the usual posterior on the parameters, θ . Note that under this formulation both the prior and posterior distributions of \mathbf{X} can be viewed as predictive distributions of the unobserved components of \mathbf{X} under the parametric model specified by $p(X_i|\theta)$ and the prior $F(\theta)$. The notion of predictive distributions has been used and discussed by Roberts (1965).

2.2. *Discussion*

Several points should be noted explicitly. First, the basic finite population model which yields the likelihood function given in (1) above is general and virtually free from any subjective element. From a Bayesian point of view the subjectivity enters solely through the choice of the prior, $p'(\mathbf{X})$. When one's prior knowledge regarding the X_i 's is such that they may be viewed as exchangeable then, as indicated in expression (4), the assessment of such a prior may be thought of as involving (a) the choice of a parametric family of distributions, $p(X_i|\boldsymbol{\theta})$, and (b) the choice of a prior distribution of $\boldsymbol{\theta}$. The *same* two specifications, where *both* typically involve some degree of subjectivity, are precisely the ones required for a Bayesian approach in most parametric inference problems.

The second point is that the basic notions given above can be extended in several directions to give formal expression for a much wider class of realistic prior knowledge. Two fairly obvious extensions are to cases where the elements of the population can be partitioned into k subsets such that within each such subset one views the X_i 's as exchangeable and to cases where one's opinions regarding the X_i 's are exchangeable conditional on some known auxiliary measures y_i , $i = 1, \dots, N$. These extensions have obvious connections with "classical" notions of stratification and regression and ratio estimation. Some specific results along the latter lines will be given in Section 5. Stratification will be dealt with in a subsequent paper.

A final point, to be noted explicitly here and used in the sequel, concerns intimate similarities between subjective prior knowledge expressible by exchangeable prior distributions of \mathbf{X} and objective distributions induced by simple random sampling. Recall that under the general model put forth above the units of the population are distinguished by an associated tag (name, address, etc., of the element)—here taken to be coded by the integers $1, \dots, N$. Thus X_i is the (unknown) variate value possessed by the population element tagged by the integer i (or with name, address, etc., coded by i).

From an objectivist point of view if a simple random sample of n of the N population elements is drawn then each of the $\binom{N}{n}$ subsets, $s = \{i_1, \dots, i_n\} \subseteq \mathcal{N}$ has probability $p(s) = 1/\binom{N}{n}$. Thus conditional on $\mathbf{X} = \mathbf{x}$ the probability that the n sample units assume the values $x_{i(1)}, \dots, x_{i(n)}$ is given by

$$p(x_{i(1)}, \dots, x_{i(n)} | \mathbf{X} = \mathbf{x}) = p(s) = 1/\binom{N}{n}.$$

This distribution, of course, yields the sampling distributions and moments of functions of the sample observations upon which standard sampling theory inferences are based.

From a Bayesian view suppose that a joint prior distribution, $p'(\mathbf{X})$, has been assessed which reflects exchangeable prior knowledge of the X_i 's. Let s' be any prespecified subset of n of the population elements—for definiteness we will suppose $s' = \{1, 2, \dots, n\}$. Then the marginal prior distribution of X_i , $i \in s'$ is the same as that for any other subset of n elements, that is,

$$p'(X_1, X_2, \dots, X_n) = p'(X_{i(1)}, \dots, X_{i(n)}) \quad \text{for every } \{i_1, \dots, i_n\} \subseteq \mathcal{N}.$$

This in turn implies, by simply looking at mixtures, that one's prior distribution of X_1, \dots, X_n is precisely the same as that on the variate values attached to n units selected by simple random sampling (or indeed by any fixed size probability sampling design). Further we note the property of such priors that given $\mathbf{X} = \mathbf{x}$ then

$$p'(X_1, \dots, X_n | \mathbf{X} = \mathbf{x}) = \begin{cases} 1, & \text{if } X_i = x_i, \quad i = 1, \dots, n, \\ 0, & \text{otherwise.} \end{cases}$$

While this stands in contrast to the objectivist's sampling distribution there is another aspect of an exchangeable prior which is tantamount to the sampling distribution. Let $\mathbf{X}^* = (X_1^*, \dots, X_N^*)$ denote the collection of values of \mathbf{X} without their distinguishing labels (or under an arbitrary labelling). Note that $\mathbf{X}^* = \mathbf{x}^*$ whenever \mathbf{X} equals any permutation of the co-ordinates of $\mathbf{x}^* = (x_1^*, \dots, x_N^*)$. It then follows immediately from the exchangeability of $p'(\mathbf{X})$ that conditional on $\mathbf{X}^* = \mathbf{x}^*$ the X_i 's are still exchangeable random variables and hence

$$p'(\mathbf{X} | \mathbf{X}^* = \mathbf{x}^*) = 1/N!$$

for \mathbf{X} equal to any permutation of the co-ordinates of \mathbf{x}^* ; and for any n co-ordinates of \mathbf{x}^* , $x_{i^*(1)}, \dots, x_{i^*(n)}$

$$p'(X_1 = x_{i^*(1)}, \dots, X_n = x_{i^*(n)} | \mathbf{X}^* = \mathbf{x}^*) = \sum p'(\mathbf{X} | \mathbf{X}^* = \mathbf{x}^*) = (N-n)!/N!$$

Finally, if we ignore the ordering by letting $\mathbf{X}_n = \{x_{i^*(1)}, \dots, x_{i^*(n)}\}$ denote the event that $x_{i^*(1)}, \dots, x_{i^*(n)}$ are the values assumed by X_1, \dots, X_n in any order then

$$p'(\mathbf{X}_n = \{x_{i^*(1)}, \dots, x_{i^*(n)}\} | \mathbf{X}^* = \mathbf{x}^*) = 1 / \binom{N}{n}.$$

Thus it follows strictly from an exchangeable prior on \mathbf{X} that given the collection, \mathbf{X}^* , of the N population variate values, but not knowing the units to which they are attached, the probability that any prespecified subset of n of the population *units* will assume any subset of these N values is precisely the same as if the subset were drawn by simple random sampling from the N elements of \mathbf{X}^* .

As immediate consequences of this result we have that, if the prior on \mathbf{X} is exchangeable and if s denotes any specified set of n of the population elements, and letting \bar{X} and s_n^2 be the mean and variance of these X_i 's ($i \in s$), then

$$E(\bar{X} | \mu) = E(\bar{X} | \mu, \sigma^2) = \mu, \quad (7)$$

$$V(\bar{X} | \mu, \sigma^2) = \frac{N-n}{N-1} \frac{\sigma^2}{n}, \quad (8)$$

and

$$E(s_n^2 | \mu, \sigma^2) = \sigma^2 \frac{n}{N-1}, \quad (9)$$

where μ and σ^2 are the finite population mean and variance, respectively, as defined earlier. These may also be seen directly by observing that conditional on μ (or on μ and σ^2) the X_i 's remain exchangeable random variables and hence have common means, variances, and covariances, and using the conditions:

$$E(\mu | \mu) = \mu,$$

$$V(\mu | \mu, \sigma^2) = 0,$$

and

$$E(s_N^2 | \mu, \sigma^2) = \sigma^2 \frac{N}{N-1}.$$

2.3. A Result on the Posterior Mean of μ

In this section we show that under reasonably general conditions the posterior expectation of μ , $E(\mu | (s, \mathbf{x}))$ is a weighted average of \bar{x} and the prior expectation of μ , $E(\mu)$, with weights respectively inversely proportional to the prior variance of μ , $V(\mu)$, and the prior expectation of the conditional "sampling" variance of \bar{X} . This appealing form follows as a corollary of a similar result for "infinite" populations given by Ericson (1969).

By way of preliminaries observe that under the model above the posterior expectation of $\mu = \sum_1^N X_i / N$ is given by

$$E\{\mu | (s, \mathbf{x})\} = \frac{1}{N} [n\bar{x} + (N-n) E\{X_i | (s, \mathbf{x})\}],$$

and since conditional on θ the X_i 's are independent identically distributed it follows that

$$E\{X_i | (s, \mathbf{x})\} = E\{\mu(\theta) | (s, \mathbf{x})\},$$

where $\mu(\theta) \equiv E(X_i | \theta)$, the common superpopulation mean. Thus we have

$$E\{\mu | (s, \mathbf{x})\} = \frac{1}{N} [n\bar{x} + (N-n) E\{\mu(\theta) | (s, \mathbf{x})\}]. \quad (10)$$

Assuming the existence of the expectations used below and that $V\{\mu(\theta)\} > 0$ we have the following theorem.

Theorem. If $E\{\mu(\theta) | (s, \mathbf{x})\} = \alpha\bar{x} + \beta$ where α and β are independent of the x_i 's then

$$E\{\mu | (s, \mathbf{x})\} = \frac{\bar{x}V(\mu) + m'E_\mu V(\bar{X} | \mu)}{V(\mu) + E_\mu V(\bar{X} | \mu)}, \quad (11)$$

where $V(\mu)$ is the prior variance of μ , $m' \equiv E(\mu)$, the prior mean of μ and $E_\mu V(\bar{X} | \mu)$ is the prior expectation of the conditional variance of \bar{X} given μ .

Proof. It was shown by the author (1969) that under the conditions of the present theorem

$$E\{\mu(\theta) | (s, \mathbf{x})\} = \frac{\bar{x}V\{\mu(\theta)\} + E\{\mu(\theta)\} E_\theta V(\bar{X} | \theta)}{V\{\mu(\theta)\} + E_\theta V(\bar{X} | \theta)}. \quad (12)$$

Now note that

$$m' \equiv E(\mu) = E_\theta E(\mu | \theta) = E\{\mu(\theta)\},$$

and by the conditional independence of the X_i 's

$$V(\mu) = E_\theta V(\mu | \theta) + V_\theta E(\mu | \theta) = \frac{1}{N} E_\theta V(X_i | \theta) + V\{\mu(\theta)\}$$

and

$$E_\theta V(\bar{X} | \theta) = \frac{1}{n} E_\theta V(X_i | \theta).$$

Substituting these results in (12) and simplifying we find

$$E\{\mu|(s, \mathbf{x})\} = \frac{\bar{x}V(\mu) + m'\{(N-n)/Nn\} E_{\theta} V(X_i|\theta)}{V(\mu) + \{(N-n)/Nn\} E_{\theta} V(X_i|\theta)}.$$

The asserted result then follows using (8) and noting that

$$V(\bar{X}|\mu) = E_{\sigma^2, \mu} V(\bar{X}|\mu, \sigma^2)$$

and

$$\begin{aligned} E_{\mu} V(\bar{X}|\mu) &= E_{\mu, \sigma^2} V(\bar{X}|\mu, \sigma^2) = \frac{N-n}{(N-1)n} E(\sigma^2) \\ &= \frac{N-n}{(N-1)n} E_{\theta} E(\sigma^2|\theta) = \frac{N-n}{nN} E_{\theta} V(X_i|\theta). \end{aligned} \quad (13)$$

It will be noted that in all of the examples discussed in Sections 3 and 4 below, as well as many others, the condition of the theorem holds, and thus the form (11) holds for a variety of distribution assumptions for finite populations. The usefulness of this theorem is that the condition $E\{\mu(\theta)|(s, \mathbf{x})\} = \alpha\bar{x} + \beta$ is very easily verified, while to show (11) directly has been found often to require considerable manipulation. The last form in (13) also provides an easy way to evaluate $E_{\mu} V(\bar{X}|\mu)$ rather than using the conditional distribution of \bar{X} given μ .

2.4. *The Role of Randomization*

A few fragmentary comments are in order on the subtle subject of the role of randomization in inference and particularly as it pertains to the specific model outlined above. These comments are in the spirit of general comments on randomization in Bayesian inference given by Savage (1962).

The first point to be re-emphasized here is that under the present model and any prior on \mathbf{X} the form of the posterior conditional on the sample sufficient statistic (s, \mathbf{x}) is given by (3) no matter how the sample was selected. (Provided, of course, that the measure $p(s^*)$ is independent of \mathbf{X}). The problem of an initial choice of sample design still remains and one might, for example, adopt a criterion such as choosing the design to minimize the prior expectation of the posterior variance of μ . The choice of design under this model will quite realistically depend on one's prior knowledge of the population as reflected here in $p'(\mathbf{X})$ as well as on the economics of implementing various alternative designs.

It is also immediately evident by symmetry that if one's initial knowledge regarding the X_i 's is truly reflected by an exchangeable prior on \mathbf{X} then, economics aside, there are no *a priori* grounds for preferring a sample consisting of any particular subset of n of the N units over any other subset of the same size. Under this state of prior knowledge and for the purposes of one's own personal inference, *ceteris paribus*, one may feel rather indifferent regarding randomization and depending on its cost may have preferences against it. When observational costs are introduced there would still remain the question of economic sample size.

Randomization seems compelling because the real world differs from the idealization of the preceding paragraph in at least two important respects. First, sample information is seldom obtained solely for one's own personal inference purposes. Clearly even if one has truly exchangeable opinions regarding the X_i 's the use of randomization may more than offset its cost through the increased utility of the resulting sample to others.

Second, it would appear that one's true prior opinions are rarely *exactly* represented by an exchangeable prior distribution of $\mathbf{X} = (X_1, \dots, X_N)$. Note, however, that for *any* prior on \mathbf{X} , $p'_0(\mathbf{X})$ say, if the original identifiability of the population units is to be destroyed by a replacement of their labels (integers $1, \dots, N$) by one of the $N!$ permutations of these integers chosen with equal probability then the prior distribution of the resulting $\mathbf{X}' = (X'_1, \dots, X'_N)$ (under the random labelling) is given by the *exchangeable* prior $p'(\mathbf{X}')$. ($p'(\mathbf{X}')$ being merely the equally weighted mixture of $p'_0(\mathbf{X})$ over all $N!$ permutations of labels.) In view of this, if one's initial opinions regarding \mathbf{X} were roughly exchangeable there may be considerable economy of thought and effort through symmetry and little loss in directly assessing the exchangeable prior $p'(\mathbf{X}')$ rather than agonizing in detail over the contingencies needed to assess $p'_0(\mathbf{X})$. In this case the model discussed above is exactly applicable for inference regarding \mathbf{X}' and functions of \mathbf{X}' , and as pointed out above one would be indifferent among samples consisting of any subset of n of the randomly relabelled units. Finally, for inference regarding permutation-invariant functions of \mathbf{X} (mean, total, variance, percentiles, etc.) this procedure is tantamount to taking $p'(\mathbf{X}) = p'(\mathbf{X}')$, that is, the exchangeable prior on the original identifiable units, and then selecting the sample of n by simple random sampling.

3. RESULTS FOR SPECIFIC DISTRIBUTIONS

In this section we give illustrative results for two special cases obtained by taking $p(X_i|\theta)$ to be normal with either known or unknown variance and taking $dF(\theta)$ to be either natural conjugate or diffuse. The case where $p(X_i|\theta)$ is binomial is a special case of results given in Section 4. Similar results are readily obtained under various other distributional assumptions. In carrying out this program much of the notation follows that of Raiffa and Schlaifer (1961). The reader will also find that a good deal of the required distribution theory used below is given in their monograph.

3.1. *Normal-Superpopulation Variance Known*

As a first example which may not be useful in characterizing real prior opinion but which sets the pattern for many of the results persisting under different assumptions, suppose $p(X_i|\theta)$ is taken as a normal density with unknown mean θ and *known* variance v and suppose further that $dF(\theta)$ is taken as the conjugate prior—normal with mean m' and variance v' . It follows readily that the prior on \mathbf{X} is an N -dimensional symmetric normal distribution with common means, m' , common variances, $v + v'$, and common covariances, v' . The prior distribution on $\mu = N^{-1} \sum_1^N X_i$ is thus normal with mean $E(\mu) = m'$ and variance

$$V(\mu) = v' + v/N. \quad (14)$$

As one expects under such a model, the prior distribution of $\sigma^2 = \sum_1^N (X_i - \mu)^2/N$, the variance of the finite population, is such that $N\sigma^2/v$ has a χ^2 distribution on $N-1$ degrees of freedom—reflecting little prior uncertainty regarding σ^2 .

Given a sample consisting of n distinct units and observed variate values \mathbf{x} , the posterior on $\mu(\theta)$, here equal to θ , is, by well-known normal distribution theory, normal with mean

$$E\{\theta|(s, \mathbf{x})\} \equiv m'' = \frac{\bar{x}v' + m'v/n}{v' + v/n} \quad (15)$$

and variance

$$v'' = \frac{1}{n/v + 1/v'} = \left(\frac{v/n}{v' + v/n} \right) v', \quad (16)$$

that is, the posterior mean of the superpopulation mean θ is a weighted average of the sample mean, \bar{x} , and the prior mean, m' , with weights inversely proportional to the variances $V(\bar{X}|\theta)$ and $V(\theta)$, while the posterior variance of θ is the reciprocal of the sum of these weights. These forms carry over to the finite population in a very natural way.

From the form (6) it is clear that the posterior distribution of \mathbf{X} is an N -dimensional singular normal distribution with all its probability concentrated in the subspace where $\mathbf{S}(\mathbf{X}) = \mathbf{x}$, or equivalently the posterior distribution on the unobserved co-ordinates $\bar{\mathbf{S}}(\mathbf{X})$ is $(N-n)$ -dimensional normal with common means, m'' , common variances and covariances given by $v+v''$ and v'' respectively, where v'' is as in (16).

It then follows immediately that the finite population mean μ , which we write in the form

$$\mu = N^{-1} \left(n\bar{x} + \sum_{i \neq s} X_i \right),$$

being a linear combination of normal random variables has a normal posterior distribution with mean

$$E\{\mu|(s, \mathbf{x})\} = N^{-1}\{n\bar{x} + (N-n)m''\}$$

and variance

$$V\{\mu|(s, \mathbf{x})\} = N^{-2}\{(N-n)(v+v'') + (N-n)(N-n-1)v''\}.$$

Substituting from (14), (15), and (16) these quantities may be rewritten as

$$E\{\mu|(s, \mathbf{x})\} = \frac{n(Nv' + v)\bar{x} + (N-n)vm'}{N(nv' + v)} \quad (17)$$

and

$$V\{\mu|(s, \mathbf{x})\} = \frac{N-n}{N^2} \frac{v(Nv' + v)}{(nv' + v)} = \frac{N-n}{N} \frac{v/n}{v' + v/n} V(\mu). \quad (18)$$

This last expression for the posterior variance is instructive, for the first factor is a finite population correction factor and the second factor is just that by which the data reduce the prior variance of the superpopulation mean (compare with (16)). It is also clear from the theorem of Section 2.3 and the linearity of $E\{\theta|(s, \mathbf{x})\}$ in \bar{x} (formula (15)) that the weights in (17) have the interpretation given in (11). This is also directly verified using (14) and from (13) noting that $E_{\mu} V(\bar{X}|\mu) = \{(N-1)/nN\}v$.

Alternatively, under the present model, it is easily verified that \bar{X}_n , the mean of any n of the X_i 's, and μ have a bivariate normal distribution. From this and the sufficiency of \bar{X}_n it follows that the distribution of \bar{X}_n given μ is $N[\mu, \{(N-n)/N\}(v/n)]$, while the marginal on μ is $N\{m', (Nv' + v)/N\}$. Thus using standard normal prior-to-posterior results it follows that the posterior distribution of μ given (s, \mathbf{x}) is normal with mean and variance given by (15) and (16), with v' replaced by $(Nv' + v)/N$ and v by $(N-n)v/N$. It also follows immediately that in the present case

$$V(\bar{X}|\mu) = E_{\mu} V(\bar{X}|\mu).$$

Finally, note that if the prior on θ is taken as degenerate uniform on the whole real line than the posterior on the finite population mean μ is normal with mean \bar{x} and variance $\{(N-n)/N\}(v/n)$.

3.2. *Normal-Superpopulation Mean and Variance Unknown*

Suppose now, somewhat more realistically, that the finite population is viewed as a sample from a normal superpopulation with unknown mean θ and unknown variance $1/h$, thus we here take $\theta = (\theta, h)$. Hence

$$p(\mathbf{X}|h, \theta) \propto h^{1/2N} \exp(-\frac{1}{2}hN\sigma^2) \exp(-\frac{1}{2}hN(\mu - \theta)^2) = h^{1/2N} \exp\left\{-\frac{1}{2}h \sum_1^N (X_i - \theta)^2\right\}, \quad (19)$$

where again μ and σ^2 are the mean and variance of the finite population. Note that here σ^2 is not equal to $1/h$.

Suppose further that $F(\theta)$ is taken to be non-degenerate normal-gamma with density

$$f(\theta, h) \propto \exp\left\{-\frac{1}{2}hn'(\theta - m')^2\right\} h^{1/2(n')} \exp(-\frac{1}{2}hv'v') h^{1/2} h^{-1}, \quad 0 < h < \infty, \quad -\infty < \theta < \infty, \\ v', n', v' > 0. \quad (20)$$

(See Raiffa and Schlaifer, 1961, for further discussion.)

By defining $\delta(n') = 0$ if $n' = 0$ and $\delta(n') = 1$ for $n' > 0$ various degenerate priors can be had as special cases of this density: for example, taking $n' = v' = 0$ (20) reduces to

$$f(\theta, h) \propto h^{-1}, \quad 0 < h < \infty, \quad -\infty < \theta < \infty. \quad (21)$$

This is merely the often used prior discussed by Jeffreys (1948) and Savage (1961), which seems a successful representation of vagueness in that it is uniform in θ and in the logarithm of the variance, h^{-1} .

It is shown in the Appendix that under this model the prior distribution of the finite population mean, μ , is "Student" that is, μ is distributed like

$$m' + t_{v'} \left\{ \frac{(N+n')v'}{n'N} \right\}^{1/2}, \quad (22)$$

where $t_{v'}$ is a standard t random variable on v' degrees of freedom. Thus the prior mean and variance of μ are m' and $v'v'(N+n')/n'N(v'-2)$ respectively.

The main result concerning the posterior distribution of μ is given in the following:

Theorem. (a) With the joint prior on \mathbf{X} implied by (19) and (20) and given the sample (s, \mathbf{x}) with mean \bar{x} and variance s^2 the posterior distribution of μ is "Student", that is, μ is distributed like the quantity

$$\frac{V(\mu)\bar{x} + V(\bar{X}|\mu)E(\mu)}{V(\mu) + V(\bar{X}|\mu)} + t_{v''} \left[\frac{N-n}{N} \frac{V\{\theta|(s, \mathbf{x})\}}{V(\theta)} V(\mu) \frac{v''-2}{v''} \right]^{1/2}, \quad (23)$$

where $v'' = v' + n$, $V(\theta|(s, \mathbf{x}))$ and $V(\theta)$ are the posterior and prior variances of θ , $V(\bar{X}|\mu)$ is the prior variance of the mean of a sample of size n given the finite population mean, $V(\mu)$ and $E(\mu)$ are the prior variance and mean of μ , respectively, and $t_{v''}$ is a standard t random variable on v'' degrees of freedom.

(b) Under the prior on \mathbf{X} generated by (19) and (21) the posterior distribution of μ given (s, \mathbf{x}) is such that μ is distributed like the quantity

$$\bar{x} + t_{n-1} \left(\frac{N-n}{N} \frac{s^2}{n} \right)^{1/2}. \quad (24)$$

A proof of this result, along with a number of alternative expressions, explicit formulae, and some results on the prior and posterior distributions of the finite population variance, σ^2 , are given in the Appendix. We note here merely that the posterior mean of μ is again of the weighted average form given in equation (11), and where, in this particular instance, $V(\bar{X}|\mu)$ is independent of μ . Also from (23) it follows that the posterior variance of μ , $\{(N-n)/N\} [V(\theta|(s, \mathbf{x}))/V(\theta)] V(\mu)$, is simply the prior variance reduced by the same two factors as in the preceding example (compare (18)). Finally, for a diffuse prior on the superpopulation parameters the posterior mean and variance of μ are seen to be \bar{x} and $\{(N-n)/N\} \{(n-1)/(n-3)\} s^2/n$ respectively, in close agreement with that which one might expect from traditional sampling theory.

4. EXTREME PRIOR VAGUENESS AND THE MULTINOMIAL

Analyses of the sort given in the preceding section can, of course, be carried through under various alternative assumptions regarding $p(X_i|\theta)$, for example, by assuming the superpopulations to be Poisson, gamma, binomial (a special case of the results given below), etc. Under these models one's prior uncertainty is represented in terms of his imperfect knowledge regarding θ and, to a lesser extent especially for large finite populations, by viewing the population as a sample from a superpopulation. While such models may often adequately approximate one's prior uncertainty regarding the unknown \mathbf{X} , nevertheless they do assume strong prior knowledge regarding the shape of the finite population distribution by taking the superpopulation form as known. Weaker prior knowledge of the shape of the finite population can be modelled by taking the $p(X_i|\theta)$ in (4) to be a member of a more flexible family of distributions, for example, the power distributions used by Box and Tiao (1962). Other approaches might take $p'(\mathbf{X})$ as a more complicated mixture. An alternative approach, based on the multinomial distribution and incorporating extreme vagueness regarding the shape of the finite population, is developed in this section. Special cases of the results below are the early result of Pearson (1928) suitable for dichotomous populations (using (4) with a binomial superpopulation), and a generalization using more general beta priors on θ than the uniform one used by Pearson.

4.1. *The Basic Model*

Suppose that each X_i can only assume one of the finite set of numerical values $\mathcal{Y} = \{y_1, y_2, \dots, y_k\}$, $y_1 < y_2 < \dots < y_k$, where k may be an extremely large integer having no relation whatever to N , the population size. This assumption clearly recognizes the inherent discreteness of almost all observation due to limitations of measuring instruments, etc. Suppose that the probability that X_i equals y_j is p_j , that is,

$$P\{X_i = y_j | \mathbf{p}\} = p_j, \quad j = 1, \dots, k, \quad \sum_{j=1}^k p_j = 1, \tag{25}$$

and where $\mathbf{p} = (p_1, \dots, p_{k-1})$. Here \mathbf{p} is assumed unknown and plays the role of θ in (4). The X_i 's are assumed independent and identically distributed with the distribution (25). Thus for any $\mathbf{y} = (y_{i(1)}, \dots, y_{i(N)})$, $y_{i(j)} \in \mathcal{Y}$, a joint prior on \mathbf{X} can be given by

$$p'(\mathbf{X} = \mathbf{y}) = \int_{\mathbf{p}} \prod_{j=1}^N p_{i(j)} f'(\mathbf{p}) d\mathbf{p}, \tag{26}$$

where $f'(\mathbf{p})$ is a prior density on the superpopulation parameter \mathbf{p} . The class of prior distributions of \mathbf{X} of the form (26) again reflects exchangeability regarding the X_i 's.

We shall find it more convenient in dealing with certain aspects of this model to consider it in slightly different terms. Let N_j be the unknown number of the N population elements for which $X_i = y_j, j = 1, \dots, k$. In this notation $\mu = N^{-1} \sum_{j=1}^k y_j N_j$ and $\sigma^2 = N^{-1} (\sum_{j=1}^k y_j^2 N_j - N\mu^2)$. In this manner inferences regarding these and other symmetric functions of the unknown finite population X_j 's are expressible in terms of the unknown N_j 's. From (26) it follows that the joint prior distribution on the N_j 's is given in non-singular form by

$$p'(\mathbf{N}) = \int_{\mathbf{p}} \frac{\Gamma(N+1)}{\prod_{i=1}^{k-1} \Gamma(N_i+1) \Gamma\left(N - \sum_1^{k-1} N_i + 1\right)} \prod_{i=1}^{k-1} p_i^{N_i} \left(1 - \sum_1^{k-1} p_i\right)^{N - \sum_1^{k-1} N_i} f'(\mathbf{p}) d\mathbf{p}, \quad (27)$$

where $\mathbf{N} = (N_1, N_2, \dots, N_{k-1})$, $\mathbf{p} = (p_1, p_2, \dots, p_{k-1})$, and the integral is over the simplex $\{p_i | 0 \leq p_i \leq 1, \sum_1^{k-1} p_i \leq 1\}$.

To summarize briefly and in slightly different form: Under the present model both the real finite population and the hypothetical "superpopulation" are defined by unknown distributions over the k distinct and ordered values y_1, \dots, y_k . The real finite population is defined by the unknown distribution function (d.f.)

$$F_X(x) = P(X \leq x) = \begin{cases} 0, & x < y_1, \\ \sum_{\mathcal{J}(x)} N_i/N, & y_1 \leq x \leq y_k, \\ 1, & x > y_k, \end{cases}$$

where $\mathcal{J}_x = \{i | y_i \leq x\}$. In addition this finite population is assumed to have been generated by N independent observations on a random variable having unknown distribution function defined by

$$F_Z(z) = P(Z \leq z) = \begin{cases} 0, & z < y_1, \\ \sum_{\mathcal{J}(z)} p_i, & y_1 \leq z \leq y_k, \\ 1, & z > y_k, \end{cases}$$

where $\mathcal{J}(z) = \{i | y_i \leq z\}$.

Choice of $f'(\mathbf{p})$

To this point the model outlined above seems straightforward and realistic. The difficulty, however, is in choosing appropriate and useful prior distributions for the parameter \mathbf{p} . Pending further study and only as a tentative and convenient approximation in certain special cases, we will take the prior on \mathbf{p} to be a $(k-1)$ -dimensional Dirichlet distribution, (see Wilks, 1962) with density

$$f'(\mathbf{p}) = \frac{\Gamma(\epsilon)}{\prod_1^{k-1} \Gamma(\epsilon_i) \Gamma\left(\epsilon - \sum_1^{k-1} \epsilon_i\right)} \prod_{i=1}^{k-1} p_i^{\epsilon_i-1} \left(1 - \sum_1^{k-1} p_i\right)^{\epsilon - \sum_1^{k-1} \epsilon_i-1}, \quad (28)$$

for parameter values $\epsilon_i > 0, i = 1, \dots, k, (\epsilon = \sum_1^k \epsilon_i)$. It is well known that this distribution has means, variances, and covariances given by

$$E(p_i) = \epsilon_i/\epsilon, \quad i = 1, \dots, k,$$

$$V(p_i) = \frac{\epsilon_i(\epsilon - \epsilon_i)}{\epsilon^2(\epsilon + 1)}, \quad i = 1, \dots, k,$$

$$\text{cov}(p_i, p_j) = \frac{-\epsilon_i \epsilon_j}{\epsilon^2(\epsilon + 1)}, \quad i \neq j = 1, \dots, k.$$

It then follows immediately that the mean and variance of the distribution function $F_Z(z)$ of the “superpopulation” are given by

$$E\{F_Z(z)\} = \begin{cases} 0, & z < y_k, \\ \frac{1}{\epsilon} \sum_{\mathcal{J}(z)} \epsilon_i, & y_1 \leq z \leq y_k, \\ 1, & z > y_k, \end{cases}$$

where \mathcal{J}_z was defined above, and

$$V\{F_Z(z)\} = \begin{cases} 0, & z < y_1 \text{ or } z > y_k, \\ \frac{1}{\epsilon + 1} E\{F_Z(z)\} [1 - E\{F_Z(z)\}], & y_1 \leq z \leq y_k. \end{cases}$$

It then follows that by the choice of the ϵ_i 's (up to an arbitrary positive multiplicative constant) the prior expectation of the d.f., $F_Z(z)$, can assume almost any desired shape, while by choosing that constant so that $\epsilon = \sum_1^k \epsilon_i$ is small, the variance of $F_Z(z)$ can be made large—representing vagueness regarding this unknown “superpopulation” d.f. It is in fact only for such small values of ϵ_i and ϵ that we believe such a Dirichlet prior is at all tenable in this setting and even then it is only tentatively put forth to represent an initial state of extreme vagueness.

Before discussing this further we note that the resulting prior on \mathbf{N} is given, using (27) and (28), as

$$p'(\mathbf{N}) = f_{DM}^{(k-1)}(\mathbf{N} | N, \epsilon, \epsilon) \equiv \frac{\Gamma(N+1) \prod_{i=1}^{k-1} \Gamma(N_i + \epsilon_i) \Gamma\left(N - \sum_1^{k-1} N_i + \epsilon - \sum_1^{k-1} \epsilon_i\right) \Gamma(\epsilon)}{\prod_{i=1}^{k-1} \Gamma(N_i + 1) \Gamma\left(N - \sum_1^{k-1} N_i + 1\right) \Gamma(N + \epsilon) \prod_{i=1}^{k-1} \Gamma(\epsilon_i) \Gamma\left(\epsilon - \sum_1^{k-1} \epsilon_i\right)},$$

$$N_i = 0, 1, \dots, N, \quad \sum_{i=1}^{k-1} N_i \leq N, \quad (29)$$

and where $\epsilon = (\epsilon_1, \dots, \epsilon_{k-1})$.

This joint distribution is merely a straightforward generalization of what Raiffa and Schlaifer (1961) have termed the beta-binomial distribution. It might analogously be termed the Dirichlet-multinomial distribution; Mosimann (1962) has termed it the

compound multinomial distribution. We will adopt the notation $f_{DM}^{(k-1)}(N|N, \epsilon, \epsilon)$ to denote the $(k-1)$ -dimensional Dirichlet-multinomial distribution whose density is given in (29). It is also easily seen that this joint distribution arises by: assuming that the N_i 's are independent Poisson random variables with parameters λ_i , $i = 1, \dots, k$, assuming that the λ_i 's are independent gamma random variables with parameters ϵ_i and common scale parameter α , and then finding the joint marginal distribution of the N_i 's conditioned on $\sum N_i = N$. Thus this prior has the property that the only relationship among the N_i 's is due to the constraint that they sum to N . When the finite population is viewed microscopically as envisaged under the present model (i.e. when k is huge and the collection of y_j 's represents all possible observations on any X_i to the practical limitations of one's measuring ability), such a feature of one's prior opinions does seem approximately realistic. However, this property persists even when large numbers of adjacent cell frequencies, N_j , are grouped. This class of priors thus seems incapable of giving formal expression to prior knowledge which is characterized by vagueness concerning the shape of the finite population and a belief that the grouped frequencies, at least, are "smooth".

We proceed to examine some consequences arising from the adoption of such a Dirichlet-multinomial prior with emphasis on the case where the ϵ_i 's and their sum are chosen to be small. In such a case this prior seems to represent an extreme position of prior vagueness, not even incorporating prior belief in the "smoothness" of grouped frequencies. As will be seen below, even though in any sample from the finite population almost all cells have an observed frequency of zero, seemingly realistic and useful posterior inferences can be made about those properties of the population typically of interest in sampling. Such inferences depend little on the actual choice of the ϵ_i 's, providing they and their sum are small. However, since there are aspects of this class of priors which are not at all realistic, it is to be expected that certain features of the resulting posteriors will be disturbing.

Properties of the prior on N

Before proceeding with the analysis of aspects of the posterior distributions it is useful for future reference to catalogue some properties of the prior distribution of several characteristics of the finite population. Since, given \mathbf{p} , the N_i 's are multinomially distributed while the p_i 's have means and variances given in the preceding section, it follows readily that

$$E(\mu) = \sum_1^k y_j \epsilon_j / \epsilon \quad (30)$$

and

$$V(\mu) = \frac{N + \epsilon}{N(\epsilon + 1)} \left\{ \sum_1^k y_i^2 \epsilon_i / \epsilon - \left(\sum_1^k y_i \epsilon_i / \epsilon \right)^2 \right\}, \quad (31)$$

where again μ is the finite population mean.

Similarly, the prior expectation of the population variance, σ^2 , is given by

$$E(\sigma^2) = \frac{\epsilon(N-1)}{N(\epsilon+1)} \left\{ \sum_1^k y_i^2 \epsilon_i / \epsilon - \left(\sum_1^k y_i \epsilon_i / \epsilon \right)^2 \right\}. \quad (32)$$

Also if ξ_π is defined as the π th percentile of the finite population, that is, $\xi_\pi = y_j$ if j is the smallest integer for which $\sum_{i=1}^j N_i/N \geq \pi$, then since $\xi_\pi \leq y_j$ whenever $\sum_{i=1}^j N_i \geq N\pi$ the prior distribution function of ξ_π is given by

$$p'(\xi_\pi \leq y_j) = \begin{cases} \sum_{c_j = \{N\pi\}}^N \binom{N}{c_j} \frac{\Gamma(\epsilon)\Gamma\left(c_j + \sum_1^j \epsilon_i\right)\Gamma\left(N + \epsilon - c_j - \sum_1^j \epsilon_i\right)}{\Gamma\left(\sum_{i=1}^j \epsilon_i\right)\Gamma\left(\epsilon - \sum_{i=1}^j \epsilon_i\right)\Gamma(N + \epsilon)}, & j = 1, \dots, k-1, \\ 1, & j = k, \end{cases} \tag{33}$$

where $\{N\pi\}$ denotes the smallest integer not less than $N\pi$. This follows since, given \mathbf{p} , $\sum_{i=1}^j N_i \equiv c_j$ has a binomial distribution with parameters N and $\sum_{i=1}^j p_i$ while, by properties of the Dirichlet distribution on \mathbf{p} , $\sum_{i=1}^j p_i$ has a beta distribution. The successive terms in the sum (33) are merely the terms of the beta-binomial distribution. For further discussion of this distribution see Raiffa and Schlaifer (1961).

There is a further interesting and useful property implicit in the use of the prior (29) which we have assumed. If s denotes any subset of n of the N distinct population elements and if n_j denotes the number of those X_i 's ($i \in s$) equal to y_j then the joint prior distribution on $\mathbf{n} = (n_1, \dots, n_{k-1})$ given \mathbf{N} is simply the generalized hypergeometric

$$p'(\mathbf{n} | \mathbf{N}) = \frac{\prod_{i=1}^{k-1} \binom{N_i}{n_i} \binom{N - \sum_1^{k-1} N_i}{n - \sum_1^{k-1} n_i}}{\binom{N}{n}}, \quad n_i = 0, \dots, N_i, \quad \sum_1^{k-1} n_i \leq n. \tag{34}$$

This result follows readily from (26) since the joint distribution of \mathbf{N} and \mathbf{n} is given, in singular form, by

$$\begin{aligned} p'(\mathbf{n}, \mathbf{N}) &= \int_{\mathbf{p}} \frac{n!}{k} \frac{1}{\prod_1^k n_i!} \prod_1^k p_i^{n_i} \frac{(N-n)!}{\prod_{i=1}^k (N_i - n_i)!} \prod_1^k p_i^{N_i - n_i} f'(\mathbf{p}) d\mathbf{p} \\ &= \frac{n!}{\prod_1^k n_i!} \frac{(N-n)!}{\prod_1^k (N_i - n_i)!} \frac{\Gamma(\epsilon) \prod_1^k \Gamma(N_i + \epsilon_i)}{\prod_1^k \Gamma(\epsilon_i) \Gamma(N + \epsilon)}. \end{aligned}$$

The result (34) then follows by dividing this expression by $p'(\mathbf{N})$ (formula (29)).

This conditional ‘‘sampling’’ distribution is a property of the exchangeable prior distribution and does not, in the usual sense, depend on random sampling. We record, for future reference, some well-known properties of this conditional distribution:

$$E(n_i | \mathbf{N}) = \frac{n}{N} N_i, \tag{35}$$

$$V(n_i | \mathbf{N}) = \frac{(N-n)nN_i(N-N_i)}{N^2(N-1)}, \tag{36}$$

and

$$\text{cov}(n_i, n_j | \mathbf{N}) = \frac{-(N-n)nN_iN_j}{N^2(N-1)} \tag{37}$$

4.2. *Posterior Inference*

We now proceed to develop some important aspects of the posterior distribution of several characteristics of the finite population typically of interest in survey work. Given any sample, no matter how selected, consisting of n distinct population units with their associated observed variate values, that is, given (s, \mathbf{x}) , it follows immediately from (5) and (26) that the posterior distribution of \mathbf{X} is given by

$$p\{\mathbf{X} = \mathbf{y} | (s, \mathbf{x})\} \propto \begin{cases} \int \prod_{p, j \neq s} p_{i(j)} \prod_{j=1}^k p_j^{n_j} f'(\mathbf{p}) d\mathbf{p}, & \mathbf{y} | \mathbf{S}(\mathbf{y}) = \mathbf{x}, \\ 0, & \text{otherwise,} \end{cases} \tag{38}$$

where $\mathbf{y} = (y_{i(1)}, \dots, y_{i(N)})$ and n_j ($j = 1, \dots, k$) is the number of the n observed x_i 's equal to y_j . Combining the second product in the integral above with the Dirichlet density $f'(\mathbf{p})$ it is clear that the posterior distribution on the unsampled units is of the same form as the prior, (29). It is immediately evident that the posterior distribution on the parameter \mathbf{p} is of the same form as the prior—a $(k-1)$ -dimensional Dirichlet with the ϵ_i 's replaced by $(\epsilon_i + n_i)$'s.

Thus letting $M_j = N_j - n_j$ be the number of unsampled X_i 's which are equal to y_j , $j = 1, \dots, k$, $\mathbf{n} = (n_1, \dots, n_{k-1})$, and $M = N - n$ it follows that the posterior distribution of $\mathbf{M} = (M_1, \dots, M_{k-1})$ is also $(k-1)$ -dimensional Dirichlet-multinomial with parameters M , $\epsilon + \mathbf{n}$, and $\mathbf{\epsilon} + \mathbf{n}$, that is, has a density given by

$$p\{\mathbf{M} | (s, \mathbf{x})\} = f_{DM}^{(k-1)}(\mathbf{M} | M, \epsilon + \mathbf{n}, \mathbf{\epsilon} + \mathbf{n}) \tag{39}$$

in the notation of formula (29). The posterior distribution of N is then immediately obtainable by substitution.

This posterior distribution, of course, is exactly that resulting by taking (34) as a likelihood function using the observed \mathbf{n} , with the prior (29): that is, (39) is, as a function of \mathbf{N} , proportional to the product of (34) and (29). Also it should be emphasized that *a posteriori* the M_j 's are distributed multinomially given \mathbf{p} while \mathbf{p} has the $(k-1)$ -dimensional Dirichlet distribution defined by the parameters $\epsilon_i + n_i$.

The posterior mean of μ

Given any sample (s, \mathbf{x}) with observed sample mean, \bar{x} , the posterior distribution of the finite population mean, μ , has a mean $E\{\mu | (s, \mathbf{x})\}$ which is again of the weighted average form displayed in (11): also, for all intents and purposes, $E\{\mu | (s, \mathbf{x})\} \doteq \bar{x}$.

We note first that

$$E\{M_i | (s, \mathbf{x})\} = \frac{(N-n)(\epsilon_i + n_i)}{\epsilon + n}, \tag{40}$$

$$V\{M_i | (s, \mathbf{x})\} = \frac{(N-n)(N+\epsilon)(\epsilon_i + n_i)(\epsilon + n - \epsilon_i - n_i)}{(\epsilon + n)^2(\epsilon + n + 1)} \tag{41}$$

and

$$\text{cov}\{M_i, M_j | (s, \mathbf{x})\} = \frac{-(N-n)(N+\epsilon)(\epsilon_i + n_i)(\epsilon_j + n_j)}{(\epsilon + n)^2(\epsilon + n + 1)}. \tag{42}$$

These formulae follow most easily by conditioning on \mathbf{p} initially, using moments of the multinomial and Dirichlet distributions. Since $\mu = N^{-1} \sum_1^k y_i (M_i + n_i)$ and since $n^{-1} \sum_1^k n_i y_i = n^{-1} \sum_{i \in S} x_i = \bar{x}$, it follows readily using (40) and (30) that

$$E\{\mu | (s, \mathbf{x})\} = \frac{n(N + \epsilon) \bar{x} + (N - n) \epsilon E(\mu)}{N(\epsilon + n)} \tag{43}$$

It is then clear that if ϵ is small, as assumed here, then

$$E\{\mu | (s, \mathbf{x})\} \doteq \bar{x}; \tag{44}$$

even if ϵ is near unity the relative weight assigned to $E(\mu)$ is the typically small fraction $(N - n)/N(n + 1)$.

It is here again easily verified that the condition of the theorem of Section 2.3 holds with $\theta = \mathbf{p}$ under the present model, and hence the posterior mean, (43), can be recast in the form (11). We then have

$$\mu(\mathbf{p}) = E(X_i | \mathbf{p}) = E_{N|\mathbf{p}} E(X_i | \mathbf{N}, \mathbf{p}) = E_{N|\mathbf{p}} \sum_1^k y_i N_i / N = \sum_1^k y_i p_i, \tag{45}$$

and using the fact that \mathbf{p} has, *a posteriori*, a Dirichlet distribution it follows that

$$E\{\mu(\mathbf{p}) | (s, \mathbf{x})\} = \sum_1^k y_i \frac{n_i + \epsilon_i}{n + \epsilon} = \frac{n\bar{x} + \epsilon E(\mu)}{n + \epsilon}, \tag{46}$$

which is linear in \bar{x} , as required.

We can thus deduce by equating (43) with (11) and using (31) and (32) that

$$E\{V(\bar{X} | \mu)\} = \frac{N - n}{N - 1} \frac{E(\sigma^2)}{n}, \tag{47}$$

where $E(\sigma^2)$ is as given in (32).

The posterior variance of μ

Several results can now be demonstrated concerning the variance of the posterior distribution μ , $V(\mu | (s, \mathbf{x}))$. Clearly

$$V\{\mu | (s, \mathbf{x})\} = \frac{1}{N^2} V\left(\sum_{j=1}^k M_j y_j | (s, \mathbf{x})\right),$$

and using the results (41) and (42) we find the expression

$$V\{\mu | (s, \mathbf{x})\} = \frac{(N - n)(N + \epsilon)}{N^2} \left\{ \sum_1^k y_i^2 \frac{(\epsilon_i + n_i)(\epsilon + n - \epsilon_i - n_i)}{(\epsilon + n)^2(\epsilon + n + 1)} - \sum_{i \neq j} y_i y_j \frac{(\epsilon_i + n_i)(\epsilon_j + n_j)}{(\epsilon + n)^2(\epsilon + n + 1)} \right\}. \tag{48}$$

The first result, in analogy with results given in Section 3, is that if $\mu(\mathbf{p}) \equiv \sum_1^k y_i p_i$, the superpopulation mean, then using (31), and the facts that the posterior variance of $\mu(\mathbf{p})$ is merely the quantity in curly brackets in (48) and that the prior variance of $\mu(\mathbf{p})$ is the same quantity with $n_i = n = 0$ for all i , it follows from (48) that

$$V\{\mu | (s, \mathbf{x})\} = \frac{N - n}{N} \left[\frac{V\{\mu(\mathbf{p}) | (s, \mathbf{x})\}}{V\{\mu(\mathbf{p})\}} \right] V(\mu). \tag{49}$$

The other, more interesting, results come from re-expressing (48) in a more enlightening form. By letting $s^2 = (n-1)^{-1}(\sum_{i=1}^k y_i^2 n_i - n\bar{x}^2)$, the observed sample variance, then after some algebraic manipulation it may be verified that (48) is equivalent to

$$V\{\mu | (s, \mathbf{x})\} = \frac{(N-n)(N+\epsilon)(n-1)s^2}{N^2(n+\epsilon)(n+\epsilon+1)} + \frac{\epsilon(\epsilon+1)(N-n)V(\mu)}{N(n+\epsilon)(n+\epsilon+1)} + \frac{n\epsilon(N-n)(N+\epsilon)\{\bar{x}-E(\mu)\}^2}{N^2(n+\epsilon)^2(n+\epsilon+1)}. \quad (50)$$

From this form it follows that if the prior parameter ϵ is chosen small (approaching zero) then

$$V\{\mu | (s, \mathbf{x})\} \doteq \frac{N-n}{N} \frac{n-1}{n+1} \frac{s^2}{n} \doteq \frac{N-n}{N} \frac{s^2}{n}. \quad (51)$$

Even if ϵ is near unity the first term in (50) may be an adequate approximation if n is large, for the latter two terms of (50) are of $O(1/n^2)$. Thus under the extreme diffuseness of prior knowledge captured in this model the posterior distribution of the finite population mean, μ , has approximate mean \bar{x} and variance given by (51). This provides a Bayesian interpretation for the usual unbiased estimates in traditional sample survey theory. Conclusions like this, especially regarding the mean, under an interesting alternative Bayesian model have been obtained earlier by Hill (1969).

The final result concerns the exact form (50). At first glance it appears similar to that obtained for the posterior variance of μ assuming a normal superpopulation with unknown mean and variance (Section 3.2). On closer examination it turns out that $V\{\mu | (s, \mathbf{x})\}$, (50), is of precisely the same form as under that normal distribution model! This may be seen by comparing (50) with the result (A21) of the Appendix and using (A7) and taking $\epsilon = n'$ and $\epsilon+3 = \nu'$. This coincidence is being studied further.

The posterior mean of σ^2

It is also interesting to note how the data change the expectation of the variance, σ^2 , of the finite population. Since $\sigma^2 = N^{-1}(\sum_1^k y_i^2 N_i - N\mu^2)$ it follows that the posterior expectation of σ^2 is given by

$$E\{\sigma^2 | (s, \mathbf{x})\} = \frac{1}{N} \left[\sum_1^k y_i^2 [n_i + E\{M_i | (s, \mathbf{x})\}] - NV\{\mu | (s, \mathbf{x})\} - NE^2\{\mu | (s, \mathbf{x})\} \right]. \quad (52)$$

Substituting from (40), (43), and (50) and after some further algebra using (32) and (33) it may be verified that

$$E\{\sigma^2 | (s, \mathbf{x})\} = \frac{1}{N} \left[\frac{(N+\epsilon)\{N\epsilon+n(N+1)\}(n-1)s^2}{N(n+\epsilon)(n+\epsilon+1)} + \frac{(N-n)(\epsilon+1)\{Nn+\epsilon(N-1)\}E(\sigma^2)}{(N-1)(n+\epsilon)(n+\epsilon+1)} + \frac{(N-n)(N+\epsilon)(n\epsilon)\{\bar{x}-E(\mu)\}^2}{N(n+\epsilon)(n+\epsilon+1)} \right]. \quad (53)$$

Using (32) it follows that if ϵ is very small then

$$E\{\sigma^2 | (s, \mathbf{x})\} \doteq \frac{N+1}{N} \frac{n-1}{n+1} s^2, \quad (54)$$

where s^2 is, as before, the observed sample variance. For ϵ near unity the two other terms in (53) play a more important role than in $V\{\mu|(s, \mathbf{x})\}$, for here they are each of $O(1/n)$.

Here too it may be verified that the expression (53) is formally identical to that (formula (A26) of the Appendix) under the normal model, by making the correspondence $\epsilon = n'$ and $\epsilon + 3 = v'$.

Posterior distribution of percentiles

In this section we obtain quite easily the exact posterior distribution of ξ_π , the π th percentile of the finite population. It is then briefly demonstrated that standard confidence intervals for percentiles agree approximately with posterior probability intervals under the model being treated here. An alternative subjectivistic approach to the distribution of percentiles has been given by Hill (1968).

To obtain the posterior distribution of ξ_π , as defined below equation (32), we note that $\xi_\pi \leq y_j$ whenever $\sum_1^j (M_i + n_i) \geq N\pi$ and thus

$$p\{\xi_\pi \leq y_j | (s, \mathbf{x})\} = p\left\{\sum_1^j M_i \geq N\pi - \sum_1^j n_i | (s, \mathbf{x})\right\}.$$

As with the prior distribution on ξ_π the posterior distribution of $\sum_{i=1}^j M_i$ is given by the beta-binomial distribution:

$$\begin{aligned} p\left\{\sum_1^j M_i = u | (s, \mathbf{x})\right\} &= \int_0^1 \binom{N-n}{u} w^u (1-w)^{N-n-u} \frac{\Gamma(n+\epsilon) w^{\sum_1^j (n_i+\epsilon_i)-1} (1-w)^{n+\epsilon-\sum_1^j (n_i+\epsilon_i)-1}}{\Gamma\{\sum_1^j (n_i+\epsilon_i)\} \Gamma\{n+\epsilon-\sum_1^j (n_i+\epsilon_i)\}} dw \\ &= \binom{N-n}{u} \frac{\Gamma(n+\epsilon) \Gamma\{u+\sum_1^j (n_i+\epsilon_i)\} \Gamma\{N+\epsilon-u-\sum_1^j (n_i+\epsilon_i)\}}{\Gamma\{\sum_1^j (n_i+\epsilon_i)\} \{\Gamma n+\epsilon-\sum_1^j (n_i+\epsilon_i)\} \Gamma(N+\epsilon)}, \end{aligned}$$

(55)

It follows that

$$p\{\xi_\pi \leq y_j | (s, \mathbf{x})\} = \begin{cases} 0, & \text{if } \{N\pi\} - \sum_1^j n_i > N-n, \quad j = 1, \dots, k, \\ \sum_{u=\{N\pi\}-\sum_1^j n_i}^{N-n} p\left\{\sum_1^j M_i = u | (s, \mathbf{x})\right\}, & 0 < \{N\pi\} - \sum_1^j n_i \leq N-n, \\ & j = 1, \dots, k-1, \\ 1, & \text{if } j = k \text{ and/or } \{N\pi\} - \sum_1^j n_i \leq 0, \end{cases}$$

(56)

where again $\{N\pi\}$ denotes the smallest integer not less than $N\pi$, and where the terms in the sum are given by (55). This expression of course yields the posterior distribution on ξ_π for any configuration of sample observations. It is clear from (55) and (56) that with the extremely diffuse prior (including no smoothness beliefs) under this model, that is, ϵ_i and ϵ small, most of the posterior probability is unrealistically concentrated on those values, y_i , observed in the sample. Nonetheless, posterior probability is attached to intervals in approximate agreement with standard confidence intervals.

To show briefly the relationship between inferences based on the above posterior distribution and standard confidence intervals for percentiles we assume, for simplicity, that the n sample observations are distinct, in other words that n_j is either zero or one for all j . We denote by $x_{(i)}$ the i th sample order statistic, thus $x_{(1)} < x_{(2)} < \dots < x_{(n)}$; and suppose also that $x_{(i)} = y_i < y_m = x_{(j)}$ for any $i < j$. We finally approximate the distribution in (56) by taking $\epsilon_i = 0, i = 1, \dots, k$, in (55).

Under these assumptions it is first clear that if $n_j = 0$ then $p\{\xi_\pi = y_j | (s, \mathbf{x})\} = 0$. Additionally if $n_j = 1$ then using (55) and (56) and for notational convenience letting

$$v_j \equiv \sum_{i=1}^{j-1} n_i, \quad \left(v_j + 1 = \sum_{i=1}^j n_i \right),$$

one has

$$p\{\xi_\pi = y_j | (s, \mathbf{x})\} = \frac{1}{\binom{N-1}{n-1}} \left\{ \sum_{u=\{N\pi\}-v_j-1}^{N-n} \binom{u+v_j}{v_j} \binom{N-u-v_j-2}{n-v_j-2} - \sum_{n=\{N\pi\}-v_j}^{N-n} \binom{u+v_j-1}{v_j-1} \binom{N-u-v_j-1}{n-v_j-1} \right\}.$$

Using the well-known identity $\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r}$, this expression reduces to the hypergeometric term

$$p\{\xi_\pi = y_j | (s, \mathbf{x})\} = \frac{\binom{\{N\pi\}-1}{\sum_{i=1}^{j-1} n_i} \binom{N-1-(\{N\pi\}-1)}{n-1-\sum_{i=1}^{j-1} n_i}}{\binom{N-1}{n-1}}, \tag{57}$$

whenever this expression is defined and zero otherwise. Using these results it follows that

$$p\{x_{(i)} < \xi_\pi \leq x_{(j)} | (s, \mathbf{x})\} = \sum_{j=i+1}^m p\{\xi_\pi = y_j | (s, \mathbf{x})\} = \sum_{v=i}^{j-1} \frac{\binom{\{N\pi\}-1}{v} \binom{N-1-(\{N\pi\}-1)}{N-1-v}}{\binom{N-1}{n-1}}. \tag{58}$$

Finally, if N is large relative to n and since $(\{N\pi\}-1)/(N-1) \doteq \pi$, using the binomial approximation to the hypergeometric we have

$$p\{x_{(i)} < \xi_\pi \leq x_{(j)} | (s, \mathbf{x})\} \doteq \sum_{v=i}^{j-1} \binom{n-1}{v} \pi^v (1-\pi)^{n-1-v} = I_\pi(i, n-i) - I_\pi(j, n-j), \tag{59}$$

where $I_\pi(u, v)$ is the usual incomplete beta function. This expression is then recognized as approximately the confidence coefficient attached to the statement that ξ_π is trapped within the random interval $\{x_{(i)}, x_{(j)}\}$ (see Wilks, 1948).

5. SOME RESULTS ON THE USE OF AUXILIARY MEASUREMENTS

In this section we give a simple extension of some of the earlier results to indicate how the basic model may be extended to incorporate *a priori* knowledge of some concomitant measurements y_1, y_2, \dots, y_N and one's prior knowledge regarding the relation of the unknown X_i 's to these values, by using them to help assess the requisite N -dimensional prior distribution of \mathbf{X} . Under various assumed relationships between \mathbf{y} and \mathbf{X} some commonly used ratio and regression estimators turn out to be the means of the posterior distribution of μ under diffuse priors.

5.1. *Regression Model*

In the following it is assumed that the y_i 's are known positive values associated with the distinguishable population elements. Let $z_i = g(y_i)$, $i = 1, \dots, N$, where g is some pre-specified positive valued function. We consider the model obtained by assuming that given y_i , α , and h , the X_i 's are independent normally distributed with means αy_i and variances z_i/h , $i = 1, \dots, N$. The parameter (α, h) is assumed unknown and assigned a normal-gamma prior.

We will be interested in three special cases obtained by taking z_i to be y_i^2 , 1, and y_i respectively. Each of these cases is equivalent to a regression through the origin with three different assumptions regarding the error variances.

Letting $v = 1/h$ note that these three cases are respectively equivalent to:

(a)
$$X_i/y_i = \alpha + \epsilon_i, \text{ where } \epsilon_i \sim N(0, v), \tag{60}$$

that is, it is assumed that the unknown ratios X_i/y_i , given y_i , are equal to an unknown constant, α , plus a normally distributed error having unknown variance, v .

(b)
$$X_i = \alpha y_i + \epsilon_i, \text{ where } \epsilon_i \sim N(0, v), \tag{61}$$

a regression through the origin with constant unknown error variance.

(c)
$$X_i/y_i = \alpha + \epsilon_i, \text{ where } \epsilon_i \sim N(0, v/y_i), \tag{62}$$

that is, the ratios X_i/y_i equal an unknown constant plus a normal error having an unknown variance proportional to y_i^{-1} .

Before proceeding one further comment seems appropriate. Although it is assumed here that the investigator can, on the basis of his prior information and knowledge, adequately approximate his prior distribution by appropriately choosing z_i , this is not completely required. He can let the data (sample) point the way by his choosing only a family of functions, say $z_i = y_i^r$, where r is unknown and then assigning a joint prior distribution on (α, h, r) . The basic form of the analysis to derive the posterior distribution of μ remains unchanged. This remark is in the spirit of the important paper of Box and Cox (1964.)

5.2. *Analysis of Model: Prior to Posterior*

Under this model for setting a joint prior on \mathbf{X}

$$p(\mathbf{X} | \alpha, h) \propto \prod_{i=1}^N \left(\frac{h}{z_i}\right)^{\frac{1}{2}} \exp\left\{-\frac{1}{2} \frac{h}{z_i} (X_i - \alpha y_i)^2\right\}, \tag{63}$$

the conditioning on the y_i 's being implicit. It is assumed that a normal-gamma prior distribution is assigned (α, h) having density

$$f'(\alpha, h) \propto \exp\left\{-\frac{1}{2} hn'(\alpha - \bar{\alpha}')^2\right\} h^{\frac{1}{2} \delta(n')} \exp\left(-\frac{1}{2} hv'v'\right) h^{\frac{1}{2} \nu' - 1}, \tag{64}$$

where $\delta(n') = 0$ if $n' = 0$ and one otherwise. The “diffuse” prior,

$$f'(\alpha, h) \propto h^{-1}, \tag{65}$$

will be considered as a special case obtained by putting $n' = \nu' = 0$ in (64). The joint prior distribution of \mathbf{X} can then be obtained by multiplying (63) by (64) and integrating out (α, h) .

Given α and h , $\mu = \sum_1^N X_i/N$ is clearly normally distributed with mean $\alpha \sum_1^N y_i/N$ and variance $\sum_1^N z_i/(N^2 h)$, and thus has density

$$p(\mu | h, \alpha) \propto (N^2 h/z)^{\frac{1}{2}} \exp\{-\frac{1}{2}(N^2 h/z)(\mu - \alpha \bar{y})^2\},$$

where $z = \sum_{i=1}^N z_i$ and $\bar{y} = \sum_{i=1}^N y_i/N$. Multiplying by $f'(\alpha, h)$ and integrating out (α, h) we find, after some simplification and assuming $n' > 0$, $\nu' > 0$, that the prior density of μ is given by

$$p'(\mu) \propto \left\{ \nu' + \frac{n' N^2}{v'(n'z + y^2)} (\mu - \bar{\alpha}' \bar{y})^2 \right\}^{-((\nu'+1)/2)},$$

where $y = N\bar{y} = \sum_1^N y_i$. From well-known properties of the “Student” distribution it follows that

$$E(\mu) = \bar{\alpha}' \bar{y}$$

and

$$V(\mu) = \frac{\nu' v'}{\nu' - 2} \frac{n' z + y^2}{n' N^2}, \quad \nu' > 2.$$

To obtain the posterior distribution of μ given the sample (s, \mathbf{x}) we recall the fact that $\mu = (n\bar{x}_s + (N-n)\mu_{\bar{s}})/N$ where $\mu_{\bar{s}} = \sum_{i \notin s} X_i/(N-n)$ and \bar{x}_s is the sample mean. Proceeding conditionally we have that, given (s, \mathbf{x}) , α and h , $\mu_{\bar{s}}$ has a normal density

$$p(\mu_{\bar{s}} | (s, \mathbf{x}), \alpha, h) \propto \left\{ \frac{(N-n)^2 h}{z_{\bar{s}}} \right\}^{\frac{1}{2}} \exp\left\{ -\frac{1}{2} \frac{h(N-n)^2}{z_{\bar{s}}} (\mu_{\bar{s}} - \alpha \bar{y}_{\bar{s}})^2 \right\}, \tag{66}$$

where $z_{\bar{s}} = \sum_{i \notin s} z_i$ and $\bar{y}_{\bar{s}} = \sum_{i \notin s} y_i/(N-n)$.

Next the posterior distribution of (α, h) given (s, \mathbf{x}) has density

$$f\{\alpha, h | (s, \mathbf{x})\} \propto h^{n/2} \exp\left\{ -\frac{1}{2} h \sum_{i \in s} \left(\frac{x_i - \alpha y_i}{\sqrt{z_i}} \right)^2 \right\} \\ \exp\left\{ -\frac{1}{2} h n' (\alpha - \bar{\alpha}')^2 \right\} h^{\frac{1}{2} \delta(n')} \exp(-\frac{1}{2} h \nu' v') h^{\frac{1}{2} \nu' - 1}.$$

This, after some simplification, can be written as

$$f\{\alpha, h | (s, \mathbf{x})\} \propto \exp\left\{ -\frac{1}{2} h n'' (\alpha - \bar{\alpha}'')^2 \right\} h^{\frac{1}{2} \delta(n'')} \exp\left\{ -\frac{1}{2} h \nu'' v'' \right\} h^{\frac{1}{2} \nu'' - 1}, \tag{67}$$

where

$$n'' = \sum_{i \in s} y_i^2/z_i + n', \tag{68}$$

$$\bar{\alpha}'' = \frac{1}{n''} \left(\sum_{i \in s} x_i y_i/z_i + n' \bar{\alpha}' \right), \tag{69}$$

$$\nu'' v'' = \left\{ \nu' v' + \frac{n'}{n''} \sum_{i \in s} \left(\frac{x_i - \bar{\alpha}' y_i}{\sqrt{z_i}} \right)^2 + \frac{1}{n''} \sum_{i \in s} y_i^2/z_i \sum_{i \in s} x_i^2/z_i - \frac{1}{n''} \left(\sum_{i \in s} x_i y_i/z_i \right)^2 \right\}, \tag{70}$$

$$\nu'' = n + \nu' + \delta(n') - \delta(n''), \tag{71}$$

and $\delta(n'') = 0$ if $n'' = 0$ and unity otherwise.

Writing the distribution in this fashion it is immediately obvious that it is still of the normal-gamma family. Thus the derivation of the posterior distribution of $\mu_{\bar{s}}$ is analogous to that of the prior on μ . The result is again a “Student” density,

$$f\{\mu_{\bar{s}} | (s, \mathbf{x})\} \propto \left\{ \nu'' + \frac{n''(N-n)^2}{\nu''(n''z_{\bar{s}} + y_{\bar{s}}^2)} (\mu_{\bar{s}} - \bar{\alpha}''\bar{y}_{\bar{s}})^2 \right\}^{-\frac{1}{2}(\nu''-1)},$$

where $y_{\bar{s}} = (N-n)\bar{y}_{\bar{s}} = \sum_{i \notin s} y_i$; and hence the posterior mean and variance of $\mu_{\bar{s}}$ are given by

$$E\{\mu_{\bar{s}} | (s, \mathbf{x})\} = \bar{\alpha}''\bar{y}_{\bar{s}}$$

and

$$V\{\mu_{\bar{s}} | (s, \mathbf{x})\} = \frac{\nu''\nu''}{\nu''-2} \frac{n''z_{\bar{s}} + y_{\bar{s}}^2}{n''(N-n)^2}, \quad \nu'' > 2.$$

It now follows immediately that the posterior distribution of μ given (s, \mathbf{x}) is like that of the quantity

$$\frac{n}{N}\bar{x}_s + \bar{\alpha}'' \frac{N-n}{N}\bar{y}_{\bar{s}} + t_{\nu''} \left\{ \frac{\nu''(n''z_{\bar{s}} + y_{\bar{s}}^2)}{n''N^2} \right\}^{\frac{1}{2}}, \tag{72}$$

where $t_{\nu''}$ is a standard t random variable on ν'' degrees of freedom. Thus

$$E\{\mu | (s, \mathbf{x})\} = \frac{n}{N}\bar{x}_s + \frac{N-n}{N}\bar{\alpha}''\bar{y}_{\bar{s}} \tag{73}$$

and

$$V\{\mu | (s, \mathbf{x})\} = \frac{\nu''\nu''}{\nu''-2} \frac{(n''z_{\bar{s}} + y_{\bar{s}}^2)}{n''N^2}. \tag{74}$$

The posterior distribution of the finite population mean, μ , with only diffuse prior information on (α, h) is immediately obtainable from the above results simply by letting $\nu' = n' = \delta(n') = 0$. It then follows that given the n distinct sample elements, (s, \mathbf{x}) , μ is distributed like a linear function of a standard t random variable on $n-1$ degrees of freedom. This situation will be discussed explicitly for the three special cases mentioned earlier.

5.3. *Special Cases of Diffuseness*

Several interesting results obtain by reconsidering the three cases introduced in Section 5.1 with only vague prior information on α and h , as represented by the improper prior density of (65). Under each of these models the posterior “Student” distribution of μ turns out to be centred on (has mean equal to) a familiar and natural “classical” estimator. This then gives some formal subjective Bayesian interpretation or justification for these estimators and some insight into the sort of prior distributions which might result in their use. Exact Bayesian credible intervals on μ are immediately obtainable using (72). In each case below we merely state the posterior mean and variance of μ obtained from (73) and (74) using the definitions (67)–(71) and taking $\nu' = n' = \delta(n') = 0$.

In the first case (a) ($z_i = y_i^2$) one finds

$$E\{\mu | (s, \mathbf{x})\} = \frac{1}{N}(n\bar{x}_s + \bar{r}y_{\bar{s}}) = \frac{1}{N}\{n(\bar{x}_s - \bar{r}\bar{y}_{\bar{s}}) + \bar{r}y\}, \tag{75}$$

where $\bar{r} = \sum_{i \in s} r_i/n$, $y = \sum_1^Y y_i$ and $r_i = x_i/y_i$. Also

$$V\{\mu | (s, \mathbf{x})\} = \left(\frac{n-1}{n-3}\right) \frac{s_r^2}{n} \frac{n \sum_{i \in s} y_i^2 + y_s^2}{N^2}, \quad (76)$$

where $s_r^2 = \sum_{i \in s} (r_i - \bar{r})^2 / (n-1)$. This posterior mean is a natural average of ratios estimator.

In case (b) ($z_i = 1$) one finds

$$E\{\mu | (s, \mathbf{x})\} = \frac{1}{N} \{n\bar{x}_s + \hat{\alpha}(y - n\bar{y}_s)\}, \quad (77)$$

where $\hat{\alpha} = \sum_s x_i y_i / \sum_s y_i^2$ is the usual least-squares estimator of α , and $\bar{y}_s = \sum_s y_i/n$. Also

$$V\{\mu | (s, \mathbf{x})\} = \frac{n-1}{n-3} s_{\hat{\alpha}}^2 \frac{(N-n) \sum_s y_i^2 - y_s^2}{N^2}, \quad (78)$$

where

$$s_{\hat{\alpha}}^2 = \frac{\sum_s y_i^2 \sum_s x_i^2 - (\sum_s x_i y_i)^2}{(n-1)(\sum y_i^2)^2}$$

is also the usual "classical" estimator of the variance of $\hat{\alpha}$.

Finally, in the last model (c) ($z_i = y_i$) one finds

$$E\{\mu | (s, \mathbf{x})\} = \frac{\bar{x}_s}{\bar{y}_s} \bar{y}, \quad (79)$$

the usual ratio estimator, and

$$V\{\mu | (s, \mathbf{x})\} = \frac{1}{N^2} \frac{s^2}{y_s} (y_s \bar{y}), \quad (80)$$

where

$$s^2 = \frac{\sum_s y_i \sum_s x_i^2 / y_i - (\sum_s x_i)^2}{(n-3) \sum_s y_i}$$

y and y_s are the totals over the population and sample respectively of the y_i 's, and $y_s = y - y_s$.

ACKNOWLEDGEMENTS

I am deeply grateful for the many useful comments on earlier drafts of this paper by G. R. Antelman, V. P. Godambe, J. W. Pratt, H. V. Roberts, J. Sedransk, and a referee.

This research was partially sponsored by the U.S. Bureau of the Census and by the National Science Foundation under Grants NSF-GP-6008 and NSF-GP-7350.

REFERENCES

- AGGARWAL, O. P. (1959). Bayes and minimax procedures in sampling from finite and infinite populations, I. *Ann. Math. Statist.*, **30**, 206–218.
 — (1966). Bayes and minimax procedures for estimating the arithmetic mean of a population with two-stage sampling. *Ann. Math. Statist.*, **37**, 1186–1195.
 BASU, D. (1958). On sampling with and without replacement. *Sankhyā*, **20**, 287–294.
 BOX, G. E. P. and COX, D. R. (1964). An analysis of transformations (with discussion). *J. R. Statist. Soc. B*, **26**, 211–252.
 BOX, G. E. P. and TIAO, G. C. (1962). A further look at robustness via Bayes's theorem. *Biometrika*, **49**, 419–432.

COCHRAN, W. G. (1939). The use of the analysis of variance in enumeration by sampling. *J. Amer. Statist. Ass.*, **34**, 492–510.

— (1946). Relative accuracies of systematic and stratified random samples for a certain class of population. *Ann. Math. Statist.*, **17**, 164–179.

DE FINETTI, B. (1937). La prévision: ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré*, **7**, 1–68.

ERICSON, W. A. (1969). A note on the posterior mean of a population mean. *J. R. Statist. Soc. B*, **31**, 332–334.

FELLER, W. (1966). *An Introduction to Probability Theory and its Applications*, Vol. II. New York: Wiley.

GODAMBE, V. P. (1955). A unified theory of sampling from finite populations. *J. R. Statist. Soc. B*, **17**, 267–278.

— (1965). A review of the contributions towards a unified theory of sampling from finite populations. *Inter. Statist. Inst. Rev.*, **33**, 242–258.

— (1966). A new approach to sampling from finite populations, I and II. *J. R. Statist. Soc. B*, **28**, 310–328.

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

HANURAV, T. V. (T. V. HANUMANTHA RAO) (1962). An existence theorem in sampling theory. *Sankhyā*, **24**, 327–330.

HILL, B. M. (1968). Posterior distribution of percentiles: Bayes's theorem for sampling from a population. *J. Amer. Statist. Ass.*, **63**, 677–691.

— (1969). Foundations for the theory of least squares. *J. R. Statist. Soc. B*, **31**, 89–97.

JEFFREYS, H. (1948). *Theory of Probability*, 2nd ed. Oxford: Clarendon Press.

KYBURG, H. E. and SMOKLER, H. E. (Ed.) (1964). *Studies in Subjective Probability*. New York: Wiley.

MOSIMANN, J. E. (1962). On the compound multinomial distribution, the multivariate β -distribution, and correlations among proportions. *Biometrika*, **49**, 65–82.

PEARSON, K. (1928). On a method of ascertaining limits to the actual number of marked members in a population of given size from a sample. *Biometrika*, **20**, 149–174.

RAIFFA, H. and SCHLAIFER, R. O. (1961). *Applied Statistical Decision Theory*. Boston: Graduate School of Business Administration, Harvard University.

ROBERTS, H. V. (1965). Probabilistic prediction. *J. Amer. Statist. Ass.*, **60**, 50–62.

— (1967). Statistical inference and decision, Chapter 14: Sampling from finite populations. (Private communication.)

SAVAGE, L. J. (1961). The subjective basis of statistical practice. (Private communication.)

— (1962). *The Foundations of Statistical Inference*. London: Methuen.

WILKS, S. S. (1948). Order statistics. *Bull. Amer. Math. Soc.*, **54**, 6–50.

— (1962). *Mathematical Statistics*. New York: Wiley.

APPENDIX

We give here a proof of the theorem of Section 2.3, as well as a number of related results under the normal distribution model of that section.

Prior on (μ, σ^2)

We begin by obtaining the prior distribution on \mathbf{X} . Multiplying (19) by (20) and integrating out θ and h one finds that the prior on \mathbf{X} is “Student” with density

$$p'(\mathbf{X}) \propto \{v' + (\mathbf{X} - \mathbf{m}')\mathbf{H}(\mathbf{X} - \mathbf{m}')\}^{-\frac{1}{2}(N+v')}, \quad -\infty < \mathbf{X} < \infty, \tag{A1}$$

where the superscript t denotes transpose, $\mathbf{m}' = (m', \dots, m')$ and

$$\mathbf{H} = \frac{1}{v'(N+n')} \begin{bmatrix} N+n'-1 & -1 & \dots & -1 \\ -1 & N+n'-1 & \dots & -1 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & \dots & N+n'-1 \end{bmatrix}. \tag{A2}$$

The prior distribution of μ can be readily deduced from this distribution, but since we wish the prior on σ^2 as well we take the following seemingly easier approach. Note that given h and θ the joint distribution of μ and σ^2 has density

$$f(\mu, \sigma^2 | h, \theta) \propto h^{\frac{1}{2}} \exp\left\{-\frac{1}{2}hN(\mu - \theta)^2\right\} (\sigma^2)^{\frac{1}{2}\nu-1} h^{\frac{1}{2}\nu} \exp\left(-\frac{1}{2}N\sigma^2\right), \quad (\text{A3})$$

where $\nu \equiv N-1$. This follows since given h and θ , $N\sigma^2 h \sim \chi^2_\nu$ independently of $(hN)^{\frac{1}{2}}(\mu - \theta)$ which is distributed as $N(0, 1)$. Multiplying (A3) by (20) and integrating out h and θ one finds that the joint prior on μ and σ^2 has density

$$f(\mu, \sigma^2) \propto (N\sigma^2)^{\frac{1}{2}\nu-1} \left\{N\sigma^2 + \nu'v' + \frac{n'N}{N+n'}(\mu - m')^2\right\}^{-\frac{1}{2}(\nu+\nu'+\delta(n'))}. \quad (\text{A4})$$

This may be integrated with respect to μ and σ^2 as shown in Raiffa and Schlaifer (1961) and yields the result that the prior distribution of μ has a "Student" density given by

$$f(\mu) \propto \left\{v' + \frac{n'N}{(n'+N)v'}(\mu - m')^2\right\}^{-\frac{1}{2}(\nu'+\delta(n'))}, \quad -\infty < \mu < \infty. \quad (\text{A5})$$

Expression (22) then follows immediately. Hence

$$E(\mu) = m' \quad (\text{A6})$$

and

$$V(\mu) = \frac{\nu'v'}{\nu'-2} \left(\frac{1}{n'} + \frac{1}{N}\right) = \left(\frac{1}{n'} + \frac{1}{N}\right) \frac{N}{N-1} E(\sigma^2), \quad (\text{A7})$$

the last equality follows from (A9) below.

Similarly, by integrating μ out of the joint density (A4) one finds that the prior density of σ^2 is given by

$$f(\sigma^2) \propto \frac{(N\sigma^2)^{\frac{1}{2}\nu-1}}{(N\sigma^2 + \nu'v')^{\frac{1}{2}(\nu+\nu'+\delta(n')-1)}}, \quad 0 < \sigma^2 < \infty, \quad (\text{A8})$$

i.e. $N\sigma^2/(N\sigma^2 + \nu'v')$ has a prior beta distribution, or $N\sigma^2$ has what Raiffa and Schlaifer term an "inverted beta-2" distribution. The prior mean and variance of σ^2 are given by

$$E(\sigma^2) = \frac{\nu'v'}{\nu'-2} \frac{N-1}{N}, \quad \nu' > 2, \quad (\text{A9})$$

and

$$V(\sigma^2) = \frac{2(\nu+\nu'-2)(\nu'v')^2}{\nu(\nu'-2)(\nu'-4)} \left(\frac{N-1}{N}\right)^2, \quad \nu' > 4. \quad (\text{A10})$$

Posterior on μ

Turning to the posterior distribution of μ we proceed as follows: given the sample (s, \mathbf{x}) with observed mean \bar{x} and variance $s^2 = (n-1)^{-1} \sum_{i \in s} (x_i - \bar{x})^2$, the posterior distribution of the superpopulation parameters (with the prior given in (20)) is also normal-gamma with density given by

$$f\{h, \theta | (s, \mathbf{x})\} \propto \exp\left\{-\frac{1}{2}hn''(\theta - m'')^2\right\} h^{\frac{1}{2}} \exp\left(-\frac{1}{2}\nu''v''h\right) h^{\frac{1}{2}\nu''-1}, \quad (\text{A11})$$

where

$$m'' = \frac{n'm' + n\bar{x}}{n' + n}, \quad (\text{A12})$$

$$n'' = n' + n, \quad (\text{A13})$$

$$v''v'' = v'v' + (n-1)s^2 + \frac{nm'}{n' + n}(\bar{x} - m')^2, \quad (\text{A14})$$

and

$$v'' = v' + \delta(n') + n - 1. \quad (\text{A15})$$

From well-known properties of the normal-gamma distribution it follows that the posterior expectation of $\theta = E(X_i | \theta, h)$ is m'' which, being linear in \bar{x} , implies by the theorem of Section 2.3 that the posterior mean of μ must be of the weighted average form displayed in (11). By the same argument used in obtaining (A1), the posterior distribution of \mathbf{X} is of the same form as the prior on \mathbf{X} and is a degenerate N -dimensional "Student" distribution concentrated in the subspace of dimension $N-n$ where $\mathbf{S}(\mathbf{X}) = \mathbf{x}$. Equivalently, the posterior distribution of the $N-n$ unobserved X_j 's is a non-singular $N-n$ dimensional "Student" distribution having a density of precisely the same form as the prior on \mathbf{X} , (A1), obtained by replacing \mathbf{X} by $\bar{\mathbf{S}}(\mathbf{X})$, N by $(N-n)$, and the primed parameters by the double-primed parameters given in (A12)–(A15). The derivation of the posterior distribution of μ is analogous to that of the prior. Let

$$\mu_{\bar{s}} = \frac{1}{N-n} \sum_{i \notin s} X_i$$

and

$$\sigma_{\bar{s}}^2 = \frac{1}{N-n-1} \sum_{i \notin s} (X_i - \mu_{\bar{s}})^2$$

be the mean and variance of the unobserved X_i 's, i.e. of the elements of $\bar{\mathbf{S}}(\mathbf{X})$. It is then clear that

$$f\{\mu_{\bar{s}}, \sigma_{\bar{s}}^2 | h, \theta, (s, \mathbf{x})\} \propto h^{\frac{1}{2}} \exp\{-\frac{1}{2}(N-n)(\mu_{\bar{s}} - \theta)^2\} \\ (\sigma_{\bar{s}}^2)^{\frac{1}{2}(N-n-1)-1} h^{\frac{1}{2}(N-n-1)} \exp\{-\frac{1}{2}(N-n-1)\sigma_{\bar{s}}^2 h\}. \quad (\text{A16})$$

Multiplying this expression by the posterior density of (h, θ) given (s, \mathbf{x}) , (A11), and integrating out (h, θ) we find, in direct analogy with (A4), that

$$f\{\mu_{\bar{s}}, \sigma_{\bar{s}}^2 | (s, \mathbf{x})\} \propto \{(N-n-1)\sigma_{\bar{s}}^2\}^{\frac{1}{2}(N-n-1)-1} \\ \left\{ (N-n-1)\sigma_{\bar{s}}^2 + v''v'' + \frac{n''(N-n)}{n'' + (N-n)}(\mu_{\bar{s}} - m'')^2 \right\}^{-\frac{1}{2}(N-n+v'')}. \quad (\text{A17})$$

Integrating out $\sigma_{\bar{s}}^2$, one finds

$$f\{\mu_{\bar{s}} | (s, \mathbf{x})\} \propto \left\{ v'' + \frac{n''(N-n)}{(n'' + N-n)v''}(\mu_{\bar{s}} - m'')^2 \right\}^{-\frac{1}{2}(v''+1)}, \quad (\text{A18})$$

or the posterior distribution of $\mu_{\bar{s}}$ is like that of $m'' + t_{v''}\{(N-n+n'')v''/(N-n)n''\}^{\frac{1}{2}}$, where $t_{v''}$ is a standard t random variable on v'' degrees of freedom.

Since $\mu = \{n\bar{x} + (N-n)\mu_{\bar{s}}\}/N$, it follows readily that the posterior distribution of μ given (s, \mathbf{x}) is like that of

$$\frac{n}{N}\bar{x} + \frac{N-n}{N}m'' + t_{v'} \left\{ \frac{(N-n+n'')(N-n)v''}{n''N^2} \right\}^{\frac{1}{2}}, \quad (\text{A19})$$

and thus using (A12)-(A15)

$$E\{\mu | (s, \mathbf{x})\} = \frac{n}{N}\bar{x} + \frac{N-n}{N}m'' = \frac{n(N+n')\bar{x} + (N-n)n'm'}{N(n'+n)}, \quad (\text{A20})$$

and

$$\begin{aligned} V\{\mu | (s, \mathbf{x})\} &= \frac{N-n+n''}{n''N^2} \frac{(N-n)v''v''}{v''-2} \\ &= \frac{N-n}{N^2} \frac{N+n'}{n+n'} \frac{\{v'v' + (n-1)s^2 + nn'(\bar{x}-m')^2/(n+n')\}}{v' + \delta(n') + n - 3}. \end{aligned} \quad (\text{A21})$$

Since the condition of the theorem of Section 2.3 holds here, we may identify (A20) and (11). Making use of the fact that $V(\mu)$ has the form in (A7) we deduce that

$$EV(\bar{X} | \mu) = \frac{N-n}{nN} \frac{v'v'}{v'-2}. \quad (\text{A22})$$

Further in this case it follows from the fact that the joint prior on X is "Student" (formula (A1)) that the conditional distribution of \bar{X} given μ has density

$$f(\bar{X} | \mu) \propto \left\{ v' + \frac{nN}{v'(N-n)} (\bar{X} - \mu)^2 \right\}^{-\frac{1}{2}(v' + \delta(n'))} \quad (\text{A23})$$

and hence it follows (Raiffa and Schlaifer, 1961) that

$$V(\bar{X} | \mu) = \frac{N-n}{nN} \frac{v'v'}{v'-2} = EV(\bar{X} | \mu),$$

and thus the posterior mean of μ , (A20), may be recast as

$$E\{\mu | (s, \mathbf{x})\} = \frac{V(\mu)\bar{x} + V(\bar{X} | \mu)m'}{V(\mu) + V(\bar{X} | \mu)}. \quad (\text{A24})$$

Finally, it may be observed, using (A7), (A21) and known properties of the normal-gamma distribution, that

$$V\{\mu | (s, \mathbf{x})\} = \frac{N-n}{N} \left\{ \frac{v''v''}{n''(v''-2)} \frac{n'(v'-2)}{v'v'} \right\} V(\mu) = \frac{N-n}{N} \frac{V\{\theta | (s, \mathbf{x})\}}{V(\theta)} V(\mu). \quad (\text{A25})$$

Part (a) of the theorem is thus established using these results and (A19).

In the special case of the diffuse prior on (θ, h) given by (21), all of the above results hold merely by taking $n' = v' = \delta(n') = 0$. This yields part (b) of the theorem.

Posterior on σ^2

While the posterior distribution of σ^2 is relatively messy, low moments may be found. For example, the posterior expectation of σ^2 may be found as follows: First we note that

$$N\sigma^2 - (n-1)s^2 = \sum_{i \neq \bar{s}} (X_i - \mu_{\bar{s}})^2 + \frac{n(N-n)}{N} (\bar{x} - \mu_{\bar{s}})^2,$$

where $\mu_{\bar{s}}$ is the mean of the $N-n$ unobserved X_i 's. Taking the expectation of this expression, first conditional on θ, h , and (s, \mathbf{x}) we find

$$E\{N\sigma^2 - (n-1)s^2 \mid \theta, h, (s, \mathbf{x})\} = \frac{(N-n)(N-1)}{N} h^{-1} + \frac{n(N-n)}{N} (\bar{x} - \mu)^2.$$

Then taking the expectation of this quantity with respect to the posterior distribution of (θ, h) given (s, \mathbf{x}) , using known properties of that normal-gamma distribution, and after some manipulation we have

$$\begin{aligned} E\{\sigma^2 \mid (s, \mathbf{x})\} &= \frac{(n-1)s^2}{N} \left[1 + \frac{(N-n)\{N(n+n')-n'\}}{N(n+n')\{\nu' + \delta(n') + n - 3\}} \right] \\ &+ \frac{(\nu' - 2)(N-n)\{N(n+n')-n'\} E(\sigma^2)}{N(N-1)(n+n')\{\nu' + \delta(n') + n - 3\}} \\ &+ \frac{nn'(N-n)}{N^2(n+n')^2} \frac{[n'\{\nu' + \delta(n') + n - 4\} + N(n+n')]}{\{\nu' + \delta(n') + n - 3\}} \{\bar{x} - E(\mu)\}^2. \quad (\text{A26}) \end{aligned}$$

For the diffuse prior obtained by taking $\nu' = n' = \delta(n') = 0$

$$E\{\sigma^2 \mid (s, \mathbf{x})\} = \left(\frac{n-1}{n-3} \frac{N-3}{N} \right) s^2.$$

DISCUSSION ON PROFESSOR ERICSON'S PAPER

Professor M. R. SAMPFORD (University of Liverpool): It gives me great pleasure to see Professor Ericson here, and to have the opportunity of proposing the vote of thanks on his extremely interesting paper. I find myself in a slight difficulty here. By Society tradition the proposer of the vote of thanks is expected to be kind to the speaker: I thus find myself, a non-Bayesian, having to be kind to a Bayesian! Perhaps we should have a film of this rare and interesting event!

As I have said, I am not myself a convinced "Bayesian"—or, at least, not a convinced practitioner of Bayesian methods. I agree, of course, that prior beliefs cannot safely be ignored in the interpretation of experimental or survey results—no "thoughtful classical practitioner", to borrow a happy phrase from our speaker, can reasonably think otherwise. Further, I accept that the use of Bayes's theorem provides an apparently less arbitrary method of taking account of such beliefs than the classical "method" of viewing with extreme suspicion, and in sufficiently desperate situations repeating, any experiment whose results conflict too violently with one's prior notions. However, I remain unconvinced of the practical value of the Bayesian argument—at least, in the field of biological and agricultural statistics that I know best.

My reasons for remaining unconvinced about Bayesian methods in survey analysis are rather different from my reasons relating to experimentation. It is arguable that experimentation is a private activity (I am well aware that this is an oversimplification!), so that if the experimenter wishes to interpret his results subjectively, he has every right to do so, and no one should be able to quarrel with his interpretation. The difficulty here, in my experience, is that most biological experiments involve so much extraneous, uncontrollable variation that an honest prior must be so diffuse as to be of little value (and even then one must be prepared for results in a particular experiment to be so anomalous as to suggest that the experimental conditions achieved were quite different from those envisaged, so