BALANCED SAMPLES AND ROBUST BAYESIAN INFERENCE IN FINITE POPULATION SAMPLING

Richard M. Royall, Department of Biostatistics, Johns Hopkins University
Dany Pfeffermann, Department of Statistics, The Hebrew University

SUMMARY

Bayesian inference in finite populations uses probability models at two stages: (i) to describe relationships among population units and (ii) to express uncertainty concerning the values of parameters appearing at stage (i). Here we consider the Bayes posterior distribution of the population total when a multivariate normal regression model is used at stage (i), with a diffuse prior distribution on the regression coefficients. We study the situation where the stage (i) model is in error because an important regressor is omitted, and we show that in balanced samples such errors do not affect the posterior distribution. Cases where the covariance matrix contains an unknown scale parameter or is itself misspecified are also considered.

1. INTRODUCTION

We consider a population made up of N units labelled 1, 2, ..., N. Associated with unit i is an unknown number y_i and a known p-dimensional vector x_i. We observe the y values for a sample s of n of the units and seek to make inferences about the population total $T = \Sigma_1^N y_i$. If r is the set of N-n non-sample units, we can write $T = \Sigma_s y_i + \Sigma_r y_i$. After the sample is observed the first term is known, and inference about T is then equivalent to inference about the unknown sum, $\Sigma_r y_i$.

We treat the case where the vector $y = (y_1, y_2, \ldots, y_N)$ is a realization of a random vector Y which is related to the matrix $X = (x_1, x_2, \ldots, X_N)'$ through a linear regression model: $E(Y) = X\beta$ and $var(Y) = V$. Without loss of generality we list the n sample units first and partition Y, X, and V,

$$Y = \begin{pmatrix} Y_s \\ Y_r \end{pmatrix}, \quad X = \begin{pmatrix} X_s \\ X_r \end{pmatrix}, \quad V = \begin{pmatrix} V_{ss} & V_{sr} \\ V_{rs} & V_{rr} \end{pmatrix},$$

where Y_s is n×1, X_s is n×p, V_{sr} is n×(N-n), etc. We assume that X_s is of full rank p.

A Bayesian solution to our problem consists of finding the conditional distribution of T, given Y_s. If Y has a (multivariate) normal distribution $N(X\beta, V)$ with X and V known, and if the unknown parameter vector β has the diffuse prior distribution, $f(\beta) \propto$ constant, $-\infty < \beta_i < \infty$, then given Y_s the predictive distribution of Y_r is easily found to be normal with expectation vector

$$E(Y_r|Y_s) = X_r\hat{\beta} + V_{rs}V_{ss}^{-1}(Y_s - X_s\hat{\beta})$$

and covariance matrix

$$var(Y_r|Y_s) = (V_{rr} - V_{rs}V_{ss}^{-1}V_{sr}) + ADA'$$

where $A = (X_r - V_{rs}V_{ss}^{-1}X_s)$, $D = (X_s'V_{ss}^{-1}X_s)^{-1}$, and $\hat{\beta} = DX_s'V_{ss}^{-1}Y_s$. Now defining $1_s' = (1, 1, \ldots, 1)_n$, $1_r' = (1, 1, \ldots, 1)_{N-n}$ and $1_N' = (1_s', 1_r')$, it follows that the distribution of $T = 1_N'Y$, given Y_s, is normal with

$$E(T|Y_s) = 1_s'Y_s + 1_r'\{X_r\hat{\beta} + V_{rs}V_{ss}^{-1}(Y_s - X_s\hat{\beta})\} \quad (1)$$

and

$$var(T|Y_s) = 1_r'\{(V_{rr} - V_{rs}V_{ss}^{-1}V_{sr}) + ADA'\}1_r = \lambda, \quad (2)$$

where λ is defined by the last equality.

The Bayes posterior distribution of T depends on the sample s actually chosen and is otherwise independent of the sampling plan. Any sample s which minimizes (2) is optimal under the present model.

Example 1: If p = 1 and given β the Y's are independent with $E(Y_i|\beta) = \beta x_i$ and $var(Y_i|\beta) = \sigma^2 x_i$ then

$$E(T|Y_s) = (\Sigma_s y_i/\Sigma_s x_i)\sum_1^N x_i = \hat{T}_R,$$

the ratio estimator, and $var(T|Y_s) = (N/f)(1-f)\sigma^2 \bar{x}_r\bar{x}/\bar{x}_s = V_r$ where $f = n/N$, $\bar{x}_s = \Sigma_s x_i/n$, and $\bar{x}_r = \Sigma_r x_i/(N-n)$. Clearly $var(T|Y_s)$ is minimized when s consists of the n units whose x-values are largest. Thus this Bayesian model leads to the popular ratio estimator but prefers an extreme purposive sample to the randomly selected samples with which the estimator is most often used.

Real world relationships are invariably more complicated than those we can represent in mathematically tractable models. As Neyman and Pearson (1937) put it, "Mathematics deals with mathematical conceptions, not with real things and we can expect no more than a certain amount of correspondence between the two." This forces attention to the robustness of statistical procedures. If our model is imperfect, say in its specification that $E(Y|\beta) = X\beta$, then our calculated distribution of T, given Y_s might be misleading. Our Bayesian procedure is robust with respect to certain changes in the model if the posterior probability distribution of T is not greatly affected by the model changes. We refer to the model actually used in deriving the posterior distribution as the working model. To study the robustness of our inferences we imagine that there is a true model, different from the working model, and compare the true posterior distribution of T with that based on our working model.

In example 1 the working model specifies that the Y_i are independent $N(\beta x_i, \sigma^2 x_i)$ random variables, with β having a diffuse prior. There might, in fact, be another parameter γ which should be in the regression model, $EY_i = \gamma + \beta x_i$. Or perhaps the true model also includes another variable Z, with coefficient γ. Or the true variance might be proportional, not to x, but to x^2. In the following sections we study the effects of such errors in the working models and characterize samples for which these errors do not strongly affect the posterior distribution. The results obtained here help to clarify the role of probability sampling in Bayesian theory. We discuss this point in the last section.

It is tempting to advise that if one is seriously concerned about the possibility that $Z\gamma$ should appear in the regression function, then one should include this term in the working model, with an appropriate prior distribution for γ, cf Bernardo (1975). This counsel of perfection is impractical in many problems. Even after the working model has been enlarged, there often remain still more variables which should perhaps be included. If in example 1 the constant intercept term is added, $EY_i = \gamma_1 + \beta x_i$, we must admit that some degree of non-linearity, possibly approximated by a quadratic term $\gamma_2 x_i^2$, might also be present. Or the units might actually be of different types, and perhaps they should be partitioned according to type into strata, with different regression coefficients in different strata. The working model, even one containing many variables, is chosen on the basis of our judgement that it is an adequate approximation, not on knowledge that it is correct.

Another suggestion is that we should look at the sample and adjust our model, if necessary, to conform to the data. Whether or not such a procedure is in the spirit of Bayesian inference, it should be noted that: (i) An elaborate data analysis is not always possible, particularly in large scale sample surveys where the totals of many variables have to be predicted simultaneously within a short time and with limited manpower. (ii) Model failures which are not often apparent in samples can cause serious errors in our inferences. Huber (1975) credits Box and Draper (1959) with first recognizing the "shocking fact" that "subliminal deviations from the model" can distort inferences so severely that protecting against such distortions is oftem more important than minimizing variance. This is dramatically demonstrated in an empirical finite population study of Royall and Cumberland (1981). Although Huber's remarks were made in a non-Bayesian context, they apply with equal force to Bayesian inference.

2. OMMISSION OF EXPLANATORY VARIABLES

First we consider the effects of omitted regressors, when the covariance matrix, given the regression coefficients, is known. Then we consider models where this matrix contains an unknown scalar and models where it is incorrect.

2.1 Known covariance matrix

The working model is:

M1: Given β, $Y \sim N(X\beta, V)$, and β has a diffuse prior distribution, $f(\beta) \propto$ constant.

The true model is the same, except for the presence of an $(N\times q)$ matrix Z of additional regressors with a fixed coefficient vector γ:

M2: Given β, $Y \sim N(X\beta + Z\gamma, V)$, and β has a diffuse prior distribution.

Model M1, unaware of the very existence of γ, represents the purest state of uncompromised ignorance about γ. On the other hand, in effect, M1 asserts with perfect certainty that γ is precisely equal to the zero vector, and this error in specifying the value of γ is the only difference between M1 and M2. The following theorem states that two conditions together ensure that the posterior distribution for T derived under M1 remains correct under M2. One of the conditions restricts the covariance matrix:

Condition L. $V_{ss}1_s = X_s\delta_1$ and $V_{sr}1_r = X_s\delta_2$ for some vectors δ_1 and δ_2. The other condition, balance, restricts the sample:

Definition: A sample s is balanced on x if $1_s'X_s/n = 1_N'X/N$.

A sample is balanced on x if for each column of X the average value in the sample is the same as in the whole population. This condition is met if and only if the sample average, $1_s'X_s/n$, equals the non-sample average $1_r'X_r/(N-n)$. Both Condition L and balance have figured prominently in recent non-Bayesian sampling theory (Royall and Herson, 1973, Royall and Cumberland, 1978, Tallis, 1978).

Some immediate consequences of Condition L are given in the following lemma, in which P_s denotes the matrix $X_sDX_s'V_{ss}^{-1}$.

Lemma 1. Condition L implies that

(i) $1_s'(I_n - P_s) = 0$

(ii) $1_r'V_{rs}(I_n - P_s') = 0$

(iii) $1_r'V_{rs}V_{ss}^{-1}(I_n - P_s) = 0$

Theorem 1. If V satisfies Condition L and if the sample is balanced on x and z, then under both models M1 and M2 the posterior distribution of T, given Y_s, is normal with mean $N\bar{y}_s$, where $\bar{y}_s = \Sigma_s y_i/n$, and variance

$$\{(1-f)/f\}^2\, 1_s' V_{ss} 1_s - 2\{(1-f)/f\} 1_s' V_{sr} 1_r + 1_r' V_{rr} 1_r \,. \tag{3}$$

Proof. Under model M2, given β, $Y^* = Y - Z\gamma \sim N(X\beta, V)$, so the conditional distribution of T, given Y_s, is normal with variance given by (2). Using Lemma 1 and balance on x we see that $1_r' ADA' 1_r = \{(1-f)/f\}^2 1_s' V_{ss} 1_s$ $-2\{(1-f)/f\} 1_s' V_{sr} 1_r + 1_r' V_{rs} V_{ss}^{-1} V_{sr} 1_r$, so that (2) reduces to (3). Now

$$E(T|Y_s) = 1_s' Y_s + 1_r' X_r D X_s' V_{ss}^{-1} Y_s^* + 1_r' Z_r \gamma$$
$$+ 1_r' V_{rs} V_{ss}^{-1} (I_n - P_s) Y_s^* \tag{4}$$

where $Y_s^* = Y_s - Z_s\gamma$. Again using the lemma and balance on x we see that (4) equals $1_s' Y_s + \{(1-f)/f\} 1_s' Y_s^* + 1_r' Z_r \gamma$, and this equals $N\bar{y}_s$ because of balance on z. Thus under M2, given Y_s, T has a normal distribution whose mean and variance are independent of γ. Since M1 corresponds to the case of $\gamma = 0$, the proof is complete.

When V in Theorem 1 satisfies the stronger Condition L^*: $V1_N = X\delta$ for some δ, and $V_{rs} = 0$, the posterior variance (3) is simply $1_s' V_{ss} 1_s (1-f)/f^2$.

In example 1 it is easy to show that if the true model contains an intercept, that is, if $EY_i = \gamma + \beta x_i$, then the true posterior distribution of T has mean $E(T|Y_s) = \hat{T}_R + N\gamma(\bar{x}_s - \bar{x})/\bar{x}_s$ and variance $\text{var}(T|Y_s) = (N/f)(1-f)\sigma^2\, \bar{x}_r \bar{x}/\bar{x}_s$. The sample maximizing $\bar{x}_s$, optimal under the working model, is doubly disadvantageous in that (i) it leads to a large error in $E(T|Y_s)$ unless $\gamma = 0$, and (ii) it is obviously a bad sample for detecting when $\gamma \neq 0$. In this example Condition L^* is met since $V1_N = X\sigma^2$, and all samples are balanced on z because $Z = 1_N$. Thus if s is balanced on x then the theorem applies and we have $E(T|Y_s) = N\bar{y}_s$ and $\text{var}(T|Y_s) = (N/f)(1-f)\sigma^2\bar{x}$. Similarly, the posterior distribution remains unchanged by the further addition of a quadratic regressor, $EY_i = \gamma_1 + \beta x_i + \gamma_2 x_i^2$, if the sample is balanced on both x and x^2.

In this example the ratio of the minimum posterior variance to that for a balanced sample can entail a substantial loss of efficiency when the working model is correct. But this is not always the case. For some working models balanced samples are optimal, as the following easily proved lemma states.

Lemma 2. Under model M1 with $V = I_N\sigma^2$, if one column of X is the vector 1_N, then $\min_s \text{var}(T|Y_s) = \sigma^2 N(1-f)/f$ and this is achieved when s is balanced on x.

Example 2. Suppose that the working model is M1 but the units are actually grouped in H strata with a different regression coefficient $\beta_h = \beta + \gamma_h$ in each stratum $h = 1, 2, \ldots, H$. Then if the units are ordered according to strata we have $X' = (X_1', X_2', \ldots, X_H')$ where X_h is the matrix for stratum h,

$$Z = \begin{pmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & X_H \end{pmatrix}$$

and $\gamma' = (\gamma_1', \gamma_2', \ldots, \gamma_H')$. Balance on z and on x is achieved if (i) within each stratum the sample is balanced on x: $1_{s_h}' X_{s_h}/n_h = 1_h' X_h/N_h$ where s_h is the sample, n_h the sample size, and N_h the number of units in stratum h, and (ii) proportional allocation is used: $n_h = nN_h/N$. Thus, for a proportional stratified balanced sample the distribution of T given Y_s is the same for all γ if Condition L is met. For instance, if V is a diagonal matrix whose ith diagonal element has the form $v_{ii} = x_i'\delta$, then Condition L^* is satisfied and the posterior distribution of T derived under M1, $T \sim N(N\bar{y}_s, \Sigma_s v_{ii}(1-f)/f^2)$, is also correct under the true model M2.

Because the conclusion in Theorem 1 is true for every fixed γ, it is true when γ has any prior distribution for which the conditional distribution of γ, given Y_s, is a proper probability distribution. The model is now

$M2^*$: Given β and γ, $Y \sim N(X\beta + Z\gamma, V)$ and β and γ are independently distributed, with $f(\beta) \propto$ constant.

Since the prior distribution of γ is not necessarily normal, the distribution of T, given Y_s, need not be normal under $M2^*$. But under balance and Condition L the same normal distribution obtained under M1 applies under $M2^*$ as well:

Corollary 1. Under the conditions of Theorem 1, under both models M1 and $M2^*$ the posterior distribution of T, given Y_s, is normal with mean $N\bar{y}_s$ and variance given by (3).

2.2 Unknown σ^2

When the working model's covariance matrix contains an unknown scalar σ^2, and $\log\sigma$ is assigned a uniform prior distribution, balanced samples no longer ensure that the true posterior distribution is the same as that derived from the working model. However, they do ensure that the two distributions have the same mean value and that for large n the variance calculated using the working model is usually larger than the unknown true variance.

We refer to the difference between the erroneous posterior mean calculated under the working model and the correct value calculated under the true model as the Bayes bias. It is the distance by which our modelling errors cause the mean (the predictor under a quadratic loss function) to be displaced from its correct location. Using model M1 when M2 is correct introduces a Bayes bias, and Theorem 1 showed that balance protects against the Bayes bias in that situation.

Now the working model is:

M1a: Given β and σ, $Y \sim N(X\beta, V\sigma^2)$, and β and $\log \sigma$ have independent diffuse prior distributions, $f(\beta,\sigma) \propto 1/\sigma$.

Under this working model the total T, given Y_s, has a Student's t distribution with n-p degrees of freedom, mean given by (1) and variance $\lambda\hat{\sigma}_1^2(n-p)/(n-p-2)$, where λ is defined in (2) and

$$(n-p)\hat{\sigma}_1^2 = (Y_s - X_s\hat{\beta})' V_{ss}^{-1}(Y_s - X_s\hat{\beta}) = Y_s' B Y_s \quad (5)$$

where $B = (I_n - P_s)' V_{ss}^{-1}(I_n - P_s)$. That is, $\{T - E(T|Y_s)\}/(\lambda\hat{\sigma}_1^2)^{1/2}$ has a Student's t distribution with n-p degrees of freedom.

The true model is, for fixed γ,

M2a: Given β and σ, $Y \sim N(X\beta + Z\gamma, V\sigma^2)$, and β and $\log \sigma$ have independent diffuse prior distributions, $f(\beta,\sigma) \propto 1/\sigma$.

Under M2a, given Y_s, T again has a Student's t distribution with n-p degrees of freedom, but with mean given by (4) and variance $\lambda\hat{\sigma}_2^2(n-p)/(n-p-2)$ where

$$(n-p)\hat{\sigma}_2^2 = Y_s^{*\prime} B\, Y_s^{*} = (n-p)\hat{\sigma}_1^2 - 2\, Y_s' B Z_s \gamma + \gamma' Z_s' B Z_s \gamma. \quad (6)$$

Since balance and Condition L ensure that the mean, $E(T|Y_s)$, is the same under M1a and M2a, the only remaining difference between the two t-distributions is that between the scale factors $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$. When n is large the following lemma shows that the variance calculated under the working model will usually be the larger. We consider that the population grows so that N and $n \to \infty$.

Lemma 3. If there is a positive constant g such that as $n \to \infty$ $\lim \gamma' Z_s' B Z_s \gamma/n = g$, then given β and σ, $\lim \Pr(\hat{\sigma}_1^2 > \hat{\sigma}_2^2) = 1$.

Proof. It follows from (6) that $\hat{\sigma}_2^2 = \hat{\sigma}_1^2 - (2\gamma' Z_s' B\varepsilon_s + \gamma' Z_s' B Z_s \gamma)/(n-p)$, where $\varepsilon_s = Y_s - X_s\beta - Z_s\gamma$. Since $2\gamma' Z_s' B\varepsilon_s/(n-p)$ has mean zero and variance $4\sigma^2\gamma' Z_s' V_{ss}^{-1}(V_{ss} - X_s D X_s')V_{ss}^{-1} Z_s \gamma/(n-p)^2$, this term converges in probability to zero, so that $\hat{\sigma}_2^2 - \hat{\sigma}_1^2$ converges in probability to -g, which implies the desired result.

Condition L and balance do not ensure that the posterior distribution of T, given Y_s, is the same under M1a and M2a. They ensure only that the Bayes bias is zero. The same remains true when γ has a prior probability distribution, since in that case the posterior mean is $E(T|Y_s) = E\{E(T|Y_s, \gamma)|Y_s\} = E(N\bar{y}_s|Y_s) = N\bar{y}_s$.

2.3 Incorrect covariance matrix

When the working model is inaccurate in specifying that the covariance matrix is V as well as in setting $\gamma = 0$, balance continues to provide protection against the Bayes bias if the true covariance matrix satisfies Condition L. This is clear from Theorem 1, where the posterior mean, $N\bar{y}_s$, does not depend on V. We state the result as a corollary. The true model is now

M3. Given β, $Y \sim N(X\beta + Z\gamma, W)$, and β has a diffuse prior distribution.

Corollary 2. If both V and W satisfy Condition L and if the sample s is balanced on both x and z, then under the models M1 and M3, $E(T|Y_s) = N\bar{y}_s$.

As before, the conclusion continues to hold when γ has a prior probability distribution, since the posterior mean, for fixed γ, does not depend on γ.

When M1 is wrong only in its specification of V, that is, when $\gamma = 0$, and V and W satisfy Condition L, the Bayes bias is $1_r' X_r(\hat{\beta} - \hat{\beta}_W)$ where $\hat{\beta}_W = (X_s' W_{ss}^{-1} X_s)^{-1} X_s' W_{ss}^{-1} Y_s$. In this case Lemma 1(i) ensures that balance on x alone makes the Bayes bias zero.

In Corollary 2 the true covariance matrix W satisfies Condition L. What happens when W depends also on the omitted variables Z? Suppose W satisfies

Condition L_1.

$W_{ss}1_s = X_s\delta_{1x} + Z_s\delta_{1z}$, $W_{sr}1_r = X_s\delta_{2x} + Z_s\delta_{2z}$ for some vectors δ_{1x}, δ_{1z}, δ_{2x}, and δ_{2z}.

It follows immediately from Theorem 1 that for a diffuse prior on γ, if the sample is balanced on x and z, then $E(T|Y_s) = N\bar{y}_s$, and there is no Bayes bias. This result does not necessarily hold for fixed γ (or when γ has a "proper" prior distribution). Nevertheless, for large samples the following theorem shows that the Bayes bias will usually be small.

Theorem 2. Under model M3, if W satisfies Condition L_1 and the sample is balanced on x and on z, and if as $n \to \infty$

$$\lim t'W_{ss}^{-1} t = 0 \qquad (7)$$

where $t' = \{\delta_{2z}'/(N-n) - \delta_{1z}'/n\} Z_s'$, then then $E(T/N|Y_s) - \bar{Y}_s$ converges in probability to zero for each β and γ.

Proof. Condition L_1 implies that $1_s'W_{ss}B_W = \delta_{1z}' Z_s'B_W$ and that $1_r'W_{rs}B_W = \delta_{2z}'Z_s'B_W$, where B_W is the same matrix as B, but with V replaced by W throughout. From this we can write

$$E(T/N|Y_s) = \bar{Y}_s + t'B_W Y_s^*.$$

Now $B_W X_s = 0$, so $t'B_W Y_s^* = t'B_W \varepsilon_s$, and the result follows from Chebychev's inequality.

Condition (7) is mild. In the simple case where X_s and Z_s are column vectors with $W_{ss} = \mathrm{diag}(x_i\delta_{1x} + z_i\delta_{1z})$ and $W_{rs} = 0$, we have

$$t'W_{ss}^{-1}t = (\delta_{1z}/n)^2 \Sigma_s z_i^2/(x_i\delta_{1x} + z_i\delta_{1z}),$$

and when $\delta_{1x} = 0$ this is simply $\delta_{1z} \bar{z}_s/n$.

2.4 Models with proper prior distributions

The preceding results have been derived under models using improper diffuse prior distributions for β and log σ. They can also be obtained as approximations, for large n, when these parameters have proper "locally uniform" prior distributions. This is simply a facet of the well known principle of "precise measurement" or "stable estimation" Savage (1962) which states roughly that for large n all relatively smooth prior distributions, including the standard diffuse priors, lead to approximately the same posterior distributions, cf e.g. Zellner (1971, p. 46).

We will only sketch the argument in the simplest case. Let M1b and M2b be the models obtained by replacing the diffuse prior for β in M1 and M2 by any smooth prior probability distribution. Using techniques similar to those in Lindley (1965, p. 13) we can show that when n is large the posterior distributions of β, given Y_s, under M1b and M2b are approximately the same as under M1 and M2 respectively. Now this implies that the posterior distributions of T/N under M1b and M2b are approximately the same as under M1 and M2 respectively. Finally, Condition L and balance on x and z imply (Theorem 1) that the posterior distributions of T/N under M1 and M2 are the same.

3. DISCUSSION

These results help to answer two serious objections to the Bayesian approach to finite population inference. The first of these concerns the possibility that an imperfect working model might produce a posterior distribution which is seriously misleading. This objection has many facets, and an important one is the possibility that misspecification of the regression function $E(Y|\beta)$ might produce an important Bayes bias, or error in the posterior mean. Although a simple linear regression function might be used in the working model, if the sample is well balanced on various powers of the regressors, then it matters little whether some more elaborate polynomial regression model would be more realistic -- the posterior distribution of T would be the same as under the simple working model. By careful choice of his sample the Bayesian can ensure that his inference is robust in this sense.

The second reservation comes from the failure of random sampling to play an important general role in Bayesian theory (Basu, 1969). Justifications for random sampling have been described in terms of its psychological effects on respondents and on potential users of the results as well as in terms of protecting the sampler from his own subconscious biases. But in terms of the Bayesian sampler's formal statistical inferences, random sampling has been problematic. Ericson (1969) offered an argument based on approximate exchangeability, but his analysis does not apply to the many populations where every unit has its own unique value for an important auxiliary variable.

In the present results a role for random sampling in Bayesian inference appears: In practice there are always variables, such as those appearing in the matrix Z in model M2, which should be included in the working model, but which are omitted because their importance is not appreciated or because it is impractical to obtain the Z matrix. If Z is unknown then whether a given sample is well balanced on z cannot be determined. But simple random sampling provides (say via Chebyshev's Inequality) grounds for confidence that the selected sample is not badly unbalanced on z (Cornfield, 1971). Thus although the random sampling distribution does not play a central role in the Bayesian's inferences, it does have a secondary role in protecting against a Bayes bias, by providing samples which are approximately balanced.

REFERENCES

Basu, D. (1969). Role of the sufficiency and likelihood principles in sample survey theory. Sankhya A 31, 441-54.

Bernardo, J. M. (1975). In discussion of Dempster, A. P. (1975). A subjectivist look at robustness. Bull. Int. Statist. Inst. 46, Book 1, 386-7.

Box, G.E.P., and Draper, N. R. (1959). A basis for the selection of a response surface design. J. Am. Statist. Assoc. 54, 622-54.

Cornfield, J. (1971). The University Group Diabetes Program. J. Am. Med. Assoc. 217, 1676-87.

Ericson, W. A. (1969). Subjective Bayesian models in sampling finite populations. J. R. Statist. Soc. B 31, 195-224.

Huber, P. J. (1975). Robustness and designs. In A Survey of Statistical Designs and Linear Models, Ed. J. N. Srivastava, pp. 287-301. Amsterdam: North-Holland.

Lindley, D. V. (1965). Introduction to Probability and Statistics from a Bayesian Viewpoint, Part 2, Inference. Cambridge University Press.

Neyman, J., and Pearson, E. (1937). A note on some points in "Student's" paper on "Comparison between balanced and random arrangements in field plots." Biometrika 29, 380-8.

Royall, R. M., and Cumberland, W. G. (1978). Variance estimation in finite population sampling. J. Am. Statist. Assoc. 73, 351-8.

Royall, R. M. and Cumberland, W. G. (1981). An empirical study of the ratio estimator and estimators of its variance. J.Am. Statist. Assoc. 76, 66-77.

Royall, R. M., and Herson, J. H. (1973). Robust estimation in finite populations I. J. Am. Statist. Assoc. 68, 880-9.

Savage, L. J. (1962). Foundations of Statistical Inference. London: Methuen and Co.

Tallis, G. M. (1978). Note on robust estimation in finite populations. Sankhya C 40, 136-8.

Zellner, A. (1971). An Introduction to Bayesian Inference in Econometrics. New York: John Wiley and Sons.