A Simple Examples of an Exchangeable Superpopulation Model

Related to Ericson's Superpopulation

Ed Stanek

Introduction

We describe a simple example of an exchangeable superpopulation model. The description is meant to help clarify the ideas of Ericson (Ericson (1969)). We relate these ideas to the projection of a superpopulation model corresponding to single stage simple random sampling given by Stanek (c00ed48.doc). The problem that we consider is very simple. We assume the finite population consists of $N=3$ units, where $n=2$ units are to be selected via simple random sampling without replacement.

The Population and Superpopulation

The finite population consists of $N=3$ primary sampling units (PSUs) with non-stochastic population values given by $\mathbf{y}'=\begin{pmatrix} y_1 & y_2 & y_3 \end{pmatrix}$. The population values are the population parameters, $\boldsymbol{\mu}'=\begin{pmatrix} m_1 & m_2 & m_3 \end{pmatrix}$, and thus $\mathbf{y}=\boldsymbol{\mu}$. We denote the unit labels as $s=1,\ldots,3$.

We define an expanded population to be a set of N^2 random variables that arise as a result of permutations of the finite population units. Each realization of the expanded population has associated with it the probability $1/N!$. In the example, the expanded population is given by $\mathbf{Y}'=\begin{pmatrix} \mathbf{Y}_1' & \mathbf{Y}_2' & \mathbf{Y}_3' \end{pmatrix}$ where $\mathbf{Y}_s'=\begin{pmatrix} Y_{1s} & Y_{2s} & Y_{Ns} \end{pmatrix}$ such that

$\mathbf{Y}' = \left(\left(Y_{11} \quad Y_{21} \quad Y_{31} \right) \quad \left(Y_{12} \quad Y_{22} \quad Y_{32} \right) \quad \left(Y_{13} \quad Y_{23} \quad Y_{33} \right) \right)$, where $Y_{is} = U_{is} y_s$. The random variables U_{is} are indicator sampling random variables that have a value of one when the i^{th} PSU in the permutation corresponds to PSU s, and zero otherwise.

We project the expanded population onto a sub-space, representing the value in the projected dimension by the value of the unit selected in the permutation. The projection matrix is given by $\mathbf{P}' = \mathbf{1}_N{}' \otimes \mathbf{I}_N$ resulting in $\mathbf{X} = \left(\left(X_i \right) \right) = \mathbf{P}'\mathbf{Y} = \begin{pmatrix} \sum_{s=1}^{N} U_{1s} y_s \\ \sum_{s=1}^{N} U_{2s} y_s \\ \sum_{s=1}^{N} U_{3s} y_s \end{pmatrix}$.

Using the assumption that each permutation of the units is equally likely, the expected value and variance of the expanded population vector $\mathbf{Y}$ are developed in c00ed27.doc, and are given by $E_{x_1}\left(\mathbf{Y}\right) = \mathbf{y} \otimes \frac{\mathbf{1}_N}{N}$ and $\text{var}_{x_1}\left(\mathbf{Y}\right) = \frac{1}{(N-1)} \mathbf{D}_{\mathbf{y}} \left(\mathbf{I}_N - \frac{\mathbf{J}_N}{N} \right) \mathbf{D}_{\mathbf{y}} \otimes \left(\mathbf{I}_N - \frac{\mathbf{J}_N}{N} \right)$. The expected value and variance of the projected expanded population are given by $E_{x_1}\left(\mathbf{X}\right) = \mathbf{1}_N \bar{m}$, while $\text{var}_{x_1}\left(\mathbf{X}\right) = s^2 \left(\mathbf{I}_N - \frac{\mathbf{J}_N}{N} \right)$ where $\bar{m} = \frac{\sum_{s=1}^{N} y_s}{N}$ and $s^2 = \frac{\sum_{s=1}^{N} \left(y_s - \bar{m} \right)^2}{N-1}$.

We call the vector $\mathbf{X} = \left(X_1 \quad X_2 \quad X_3 \right)'$ the superpopulation, and represent an element in the superpopulation as X_i.

Exchangeable Random Variables

The vector $\mathbf{X}$ is a vector of random variables. The random variables are said to be exchangeable if each of the $N! = 3! = 6$ permutations of the random variables has the same joint probability distribution. The vector $\mathbf{X}$ satisfies this property, and hence is exchangeable.

Bayesian Inference

Bayesian inference requires a prior distribution on the vector $\mathbf{X}$. We trace the basic model described by Ericson (Ericson (1969)) and summarized by (Ghosh and Meeden (1997)) (section 1.4) to develop this prior distribution. Also, see the discussion by (Cassel, Sarndal et al. (1977)) (page 135, case B). In the example, since the finite population consists of three discrete values, each random variable can assume one of a finite set of values. However, when specifying the prior, Ericson (p206) states that the number of possible values,

> ‘ k may be an extremely large integer having no relation whatever to N, the population size.’

In keeping with this idea for this example, let each random variable X_i result in a single value from the ordered set $Y = \{y_1, y_2, y_3, y_4, y_5\}$ of $k = 5$ discrete values. These values are constants. A subset of these values will be known as a result of realizing the random variables in the sample.

Let the elements of $\mathbf{p} = (p_1, p_2, p_3, p_4)$ represent $P\{X_i = y_j \mid \mathbf{p}\} = p_j$ for $j = 1,\ldots,k-1 (= 4)$ and $\sum_{j=1}^{k} p_j = 1$. These values are generally unknown. In our example, if the values of y_s correspond to the values of y_j for $j = \{1,3,4\}$, then $P\{X_i = y_j \mid \mathbf{p}\} = 1/3$ for $j = \{1,3,4\}$ for all $i = 1,\ldots,3$ and $P\{X_i = y_j \mid \mathbf{p}\} = 0$ for $j = 2,5$ and for all $i = 1,\ldots,3$. Ericson considers $\mathbf{p} = (p_1, p_2, p_3, p_4)$ to be a parameter of the superpopulation. Note that the dimension of this parameter may be larger than the number of units in the population.

Now, to form the prior distribution of $\mathbf{X}$, Ericson first assumes a distribution of $\mathbf{p} = (p_1, p_2, p_3, p_4)$, say $F(\mathbf{p})$. Then, Ericson (Ericson (1969)) (p206) assumes that given $\mathbf{p} = (p_1, p_2, p_3, p_4)$, the random variables $X_i \mid \mathbf{p}$ are identically distributed (which is true in our example) and independent (which is not the case in our example).

[Aside: It seems to me that the independence assumption is not really necessary, since the purpose of the assumption is to express the joint distribution of $\mathbf{X} \mid \mathbf{p}$ as a product, $\prod_{i=1}^{N} X_i \mid \mathbf{p}$ when calculating the marginal distribution. Connected to this thought is the discussion of Heath and Sudderth (Heath and Sudderth (1976)) where de Finetti's theorem states that

> "every sequence of exchangeable (not necessarily 0-1) variables is a mixture of sequences of independent, identically distributed variables" (p189).]

Since the distribution $F(\mathbf{p})$ is assumed known, Ericson proposes integrating over this distribution to form the prior distribution of $\mathbf{X}$. Thus, the prior distribution of $\mathbf{X}$ is given by

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

Alternative Motivation of the Prior

The goal as stated by Ericson (Ericson (1969)) (p207) is to arrive at a plausible prior distribution of $\mathbf{X}$. Assuming that the values in the finite population can be represented by a discrete set of k values, let N_j represent the number of actual finite population units that have the value y_j. Let $\mathbf{N} = (N_1, N_2, \ldots, N_k)$ denote the vector of unknown values of N_j in the population. Suppose $\mathbf{p}$ represents a $k \times 1$ dimensional parameter representing the proportion (or probability) of elements with the values of N_j in a possible infinite superpopulation, and let $F(\mathbf{p})$ represent a (assumed) known probability density. Then, the joint prior on $\mathbf{N}$ is given by

$$p(\mathbf{N}) = \int_{\mathbf{p}} \prod_{j=1}^{k} \left(\frac{N!}{N_j!} p_j^{N_j} \right) F(\mathbf{p})\, d\mathbf{p}$$

where we have the usual constraints that $N = \sum_{j=1}^{k} N_j$ and $1 = \sum_{j=1}^{k} p_j$. With this formulation, both the population and superpopulation have the same number of distinct values (i.e. k). The difference between the population and the superpopulation is the probability of these values.

Ericson suggests using a Dirichlet distribution for $F(\mathbf{p})$. Note that $\Gamma(N+1) = N!$. The Dirichlet distribution is given by $F(\mathbf{p}) = \prod_{j=1}^{k} \left(\frac{\Gamma(e)}{\Gamma(e_j)} p_j^{e_j - 1} \right)$ where $e_k = e - \sum_{j=1}^{k-1} e_j$ and $p_k = 1 - \sum_{j=1}^{k-1} p_j$. For the Dirichlet distribution, note that $E(p_j) = e_j / e$, $V(p_j) = \frac{e_j (e - e_j)}{e^2 (e+1)}$, and $\text{cov}(p_j, p_{j^*}) = \frac{-e_j e_{j^*}}{e^2 (e+1)}$. This distribution has the desirable property that by choosing e to be small, the variance can be made arbitrarily large, implying vague prior information.

References

Cassel, C. M., C. E. Sarndal, et al. (1977). Foundations of inference in survey sampling. New York, John Wiley and Sons.

Ericson, W. A. (1969). "Subjective Bayesian models in sampling finite population I." Journal of the Royal Statistical Society B **31**: 195-234.

Ghosh, M. and G. Meeden (1997). Bayesian methods for finite population sampling. New York, Chapman Hall.

Heath, D. and W. Sudderth (1976). "De Finetti's theorem on exchangeable variables." The American Statitistician **30**: 188-189.