Bayesian Analysis in the Context of Simple Random Without Replacement Sampling

Ed Stanek

Introduction

We have discussed a problem where data are available on two subjects indexed by \( j = 1, \ldots, n = 2 \). There was no mention of other subjects in the context of the analysis that we discussed. There is mention, however, of other possible latent values. In fact, we assumed that the underlying latent value parameters may correspond to those for the two study subjects (i.e. \( \pi = y \)) or may correspond to those for the subjects in the population (i.e. \( \pi = y_N \)). The way this was presented, the second setting involves all the subjects in the population. Since the latent values in the \( N \times 1 \) vector only need to include those in \( y \), if there are \( N \) latent values in \( \pi \), the additional latent values don’t actually have to include latent values for other subjects in the population who are not in the data, but could correspond to any \( N - n \) additional latent values.

We refer to the description given in Chapter 7 of Gelman et al. (2004) in our discussion of sampling and Bayesian analysis here. It is valuable to review what these authors state regarding Bayesian analysis and sampling.

A naïve student of Bayesian inference might claim that because all inference is conditional on the observed data, it makes no difference how those data were collected. This misplaced appeal to the likelihood principle would assert that given (1) a fixed model (including the prior distribution) for the underlying data and (2) fixed observed values of the data, Bayesian inference is determined regardless of the design for the collection of the data. Under this view there would be no formal role for randomization in either sample surveys or experiments. The essential flaw in the argument is that a complete definition of ‘the observed data’ should include information on how the observed values arose, and in many situations such information has a direct bearing on how these values should be interpreted. Formally then, the data analyst needs to incorporate the information describing the data collection process in the probability model used for the analysis.

The notion that the method of data collection is irrelevant to Bayesian analysis can be dispelled by the simplest of examples. Suppose for instance that we, the authors, give you, the reader, a collection of the outcomes of ten rolls of a die and all are 6’s. Certainly your attitude toward the nature of the die after analyzing these data would be different if we told you (i) these were the only rolls we performed, versus (ii) we rolled the die 60 times but decided to report only the 6’s, versus (iii) we decided in advance that we were going to report honestly that ten 6’s appeared but would conceal how many rolls it took, and we had to wait 500 rolls to attain that result. In simple situations such as these, it is easy to see that the observed data follow a different distribution form that for the underlying ‘complete data’. p198.

In the ‘simplest of examples’, the authors describe three different situations, and say that it is ‘easy to see that the observed data follow a different distribution’. The three situations actually correspond to different data. In the first setting, there are three observed values. In the second setting, there are 60 observed values, and the data is a set of values on 60 rolls. In the
third situation, there are 600 values corresponding to a sequence of 600 rolls. It is easy to see that if data differ, what we think the data mean (i.e., inference) may also differ.

By the complete data, these authors refer to “Values from all \( N \) units in the population” p199. We define the complete data as an \( N \times 1 \) vector of latent values, \( \mathbf{y}_N = (y_1, y_2, \cdots, y_N)' \), where \( y_s, \ s = 1, \ldots, N \) represents response for the subject with label \( s \). This seems like a simple starting point, but it already introduces a subtle feature of the notation. We have described the elements of \( \mathbf{y}_N \) by \( y_s \), the response for the subject with label \( s \). In order for this description to hold, the position in the vector \( \mathbf{y}_N \) defines a subject in the population. If this is true, then permuting the values (and the corresponding subjects) will result in another vector which also could be considered to be the complete data for the same population. Since the subject labels are assumed not to contain information (i.e., a person may be labeled by their name, their social security number, or their medical record number, each of which serves to identify the person), the values of subjects in the population could be described by listing subject’s values in different orders.

Suppose now that we define the complete data as an \( N \times 1 \) vector of latent values, \( \mathbf{y}_N = (12 \ 4 \ \cdots \ 33)' \). Notice that if \( \mathbf{y}_N' = (4 \ 12 \ \cdots \ 33)' \), then in general, \( \mathbf{y}_N \neq \mathbf{y}_N' \). The idea is that if the subjects are \( s = (Lily \ Rose \ \cdots \ Daisy)' \), then the complete data are different if \( \mathbf{y}_N = (y_{Lily} = 12 \ | \ y_{Rose} = 4 \ | \cdots \ | \ y_{Daisy} = 33)' \), or if \( \mathbf{y}_N' = (y_{Lily} = 4 \ | \ y_{Rose} = 12 \ | \cdots \ | \ y_{Daisy} = 33)' \). The missing information in representing the complete data by \( \mathbf{y}_N = (y_1, y_2, \cdots, y_N)' \) is the connection between the subject labels and the positions.

We can introduce information that is missing in two ways. One way to introduce it is to assume that labels are ordered (from smallest to largest), and that the order of the label corresponds to the position in \( \mathbf{y}_N \). We have used \( j \) to index the labels with this idea. A second way to introduce a connection between labels and positions is via a permutation matrix. Suppose \( \mathbf{v}_m^{(N)} \) is an \( N \times N \) permutation matrix. If \( N = 3 \) and \( s = (Lily \ Rose \ Daisy)' \), we can describe

\[
\mathbf{y}_N = \begin{pmatrix} y_{Lily} & y_{Rose} & y_{Daisy} \end{pmatrix}' \quad \text{as} \quad \mathbf{v}_m^{(N)} \mathbf{y}_N = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_{Rose} & y_{Lily} & y_{Daisy} \end{pmatrix}'.
\]

Since \( \mathbf{y}_N \) and \( \mathbf{v}_m^{(N)} \mathbf{y}_N \) describe the same response for the same subjects, the same population is described.

These issues are valuable to consider when we think of the \( N \) subjects being selected from \( N > N \) subjects in a superpopulation. The \( N \) subjects can be represented as a set of subjects, or as a sequence of subjects. These two different representations relate to the issues described above.

In order to distinguish the data from the population, a vector of indicator random variables is introduced given by \( \mathbf{I}^B = (I_1^B \ I_2^B \ \cdots \ I_N^B)' \) where the value of \( I_s^B \) is one if subject
is included in a sample, and zero otherwise. We include the superscript $B$ to indicate that these indicator random variables are used in the context of Bayesian analyses. These indicator random variables corresponding to $I_s^B = \sum_{i=1}^{n} U_{is}$ where $U_{is}$ is an indicator random variable that has a value of one if subject $s$ is in position $i$ in a sample sequence, and zero otherwise.

**Relationship with Other Notation for Samples**

In the context of finite population sampling, a sample may be represented as a set of subjects, or as a sequence of subjects. The sequence notation is more general, since for a given set of subjects, many sequences can be formed. However, information describing which subjects are included in a sample is contained in a sample set, so that commonly representing a sample by a set of subjects is sufficient.

One way to describe samples is to list all possible samples. Let $\tilde{h} = 1, \ldots, \tilde{H}$ index all possible sample sequences of $n$ subjects from a finite population. A particular sample sequence is vector of pairs of values, where for each position in the sequence a pair consists of the label for the subject in the position, and the response. We define the labels for subjects explicitly in the population by $\mathbf{c} = (1 \ 2 \ \cdots \ N)'$ with elements $c_s = s$. For example, when $n = 2$ and $N = 3$, the data for sequence $\tilde{h}$ consisting of subject $s = 3$ with response $y_3$, followed by subject $s = 1$ with response $y_1$ is $((s = 3, y_3) \ (s = 1, y_1)).$
Sample sequences can be represented as permutations of elements in a sample set. Let us index possible sample sets by \( h \) where \( h = 1, \ldots, H \), and let the set \( S \) represent all possible sample sets. We define the subjects in sample set \( h \) by 
\[
\delta_h' = \begin{pmatrix}
\delta_{h11} & \delta_{h12} & \cdots & \delta_{h1N} \\
\delta_{h21} & \delta_{h22} & \cdots & \delta_{h2N} \\
\vdots & \vdots & \ddots & \vdots \\
\delta_{hn1} & \delta_{hn2} & \cdots & \delta_{hnN}
\end{pmatrix},
\]
\( \delta_{hj} \) has a value of one if the \( j^{th} \) smallest subject’s label in set \( h \) is for subject \( s \), and zero otherwise. The subscript \( j \) uniquely identifies a subject in a given sample set, where \( j = 1, \ldots, n \).

A sample sequence corresponds to a particular permutation of subjects in a sample set. Let \( v_m \) be an \( n \times n \) permutation matrix where elements in each row and column all sum to one. The subscript \( m = 1, \ldots, M \) indexes possible permutations of sample subjects, and we define 
\[
\delta_h = \delta_h' v_m.
\]

### Complete Data in a Bayesian Analysis

We represent the vector of latent values in sample set \( h \) by \( y_{h,obs} = \delta_h' y_N \). A Bayesian analysis defines the complete data as the potentially observed data plus the missing data. This is not included in Godambe’s (1955) framework. The idea is included in Ericson’s (1969) development, where there is a complementary matrix, \( \delta_h', \text{miss} \), of dimension \((N - n) \times N\)
\[
\delta_h', \text{miss} = \begin{pmatrix}
\delta_{h(n+1)1} & \delta_{h(n+1)2} & \cdots & \delta_{h(n+1)N} \\
\delta_{h(n+2)1} & \delta_{h(n+2)2} & \cdots & \delta_{h(n+2)N} \\
\vdots & \vdots & \ddots & \vdots \\
\delta_{hN1} & \delta_{hN2} & \cdots & \delta_{hNN}
\end{pmatrix}
\]
such that \( (\delta_h, \delta_h', \text{miss}) \) is a permutation matrix with all rows and all columns summing to one. The columns of \( \delta_h', \text{miss} \) are ordered such that the subjects identified have labels that are ordered from smallest to largest. Using this notation, we define 
\[
y_{h, \text{miss}} = \delta_h', \text{miss} y_N.
\]

We represent response for a sample sequence as random variables by defining indicator random variables, \( I_h \) for \( h = 1, \ldots, H \), where \( I_h \) has a value of one if sequence \( h \) is realized, and zero otherwise. In this context, \( U'_f = \sum_{h=1}^{H} I_h \tilde{u}_h' \) where \( I_h \) is an indicator random variable for sequence \( h \), and elements of \( U_f \) are \( U_{is} \). Response for a sample set is represented by introducing a vector of random variables \( I_h \), \( h = 1, \ldots, H \), that have a value of one if set \( h \) is selected, and a value of zero otherwise where 
\[
S = \begin{pmatrix}
I_1 & I_2 & \cdots & I_H
\end{pmatrix}.'
We represent the sample latent values by \( Y_{\text{obs}} = \sum_{h=1}^{H} I_h y_{h,\text{obs}} \). We have represented a realized value of this random variable as \( y \) in c09ed39.doc, and now substitute the notation \( y_{h,\text{obs}} \) for a realized value of \( Y_{\text{obs}} \). Notice that we can express

\[
Y_{\text{obs}} = \sum_{h=1}^{H} I_h \delta_h' y_N
\]

\[
= \left( I'_H \otimes I_n \right) \left( \bigoplus_{h=1}^{H} \delta_h' y_N \right) S
\]

where \( \left( \bigoplus_{h=1}^{H} \delta_h' y_N \right) S = \left( I_1 y'_1, I_2 y'_2, \ldots, I_H y'_H \right)' \) is an \( nH \times 1 \) vector of possible sets of sample latent values.

Sample Data with Labels and Latent Values

Ericson (1969) describes the data to include not only the subject values but also the subject labels. For a sample sequence, Ericson represents the data as \( (s_{h,\text{obs}}, y_{h,\text{obs}}) \). This definition recognizes the labels and values as part of the observed data. In the problems considered by Ericson, there was no response error.

Prior Distributions

Ericson (1969) assumes that \( y_N \) is a realization of the vector of random variables, \( Y_N \), and represents the joint prior distribution of \( Y_N \) by \( p'(Y_N) \). (Notice that we have not explicitly defined \( Y_N \).) He then states that the probability of \( Y_N \) given \( (s_{h,\text{obs}}, y_{h,\text{obs}}) \) (i.e. the posterior probability) is the same regardless of whether or not we condition on \( (s_{h,\text{obs}}, y_{h,\text{obs}}) \), or on \( (s_{h,\text{obs}}, y_{h,\text{obs}}) \). This is true since \( (s_{h,\text{obs}}, y_{h,\text{obs}}) \) and \( (s_{h,\text{obs}}, y_{h,\text{obs}}) \) are linearly related via a non-stochastic matrix \( v_m \) where \( s_{h,\text{obs}} = v_m s_{h,\text{obs}} \) and \( y_{h,\text{obs}} = v_m y_{h,\text{obs}} \). Since the posterior distributions are the same, \( (s_{h,\text{obs}}, y_{h,\text{obs}}) \) is a sufficient statistic.

As a result, given \( (s_{h,\text{obs}}, y_{h,\text{obs}}) \), Ericson considered the likelihood as (see equation (1), p196 in Ericson (1969)) as

\[
l(Y_N; (s_{h,\text{obs}}, y_{h,\text{obs}})) = l(Y_N; (s_{h,\text{obs}}, y_{h,\text{obs}})) = \begin{cases} kp(\tilde{I}_h) & \text{for } Y_N | \delta_h' Y_N = y_{h,\text{obs}}, \\ 0 & \text{otherwise} \end{cases}
\]

where \( k > 0 \) is an arbitrary constant.

We discuss this expression since the notation used by Ericson is slightly different. Ericson represents \( \delta_h' Y_N = y_{h,\text{obs}} \) by defining an operator \( S \) such that \( S(Y_N) = y_{h,\text{obs}} \), which we interpret as the realized values for a particular sample set. In an earlier section, he refers to the
set $S^*$ as a set of all possible ordered sample sets. Ericson represents $\delta_h' Y_N = y_{h,obs}$ by $S(Y_N) = y_{obs}$ (where $S$ is not bold, i.e. $S$), but does not define $S$. Suppose that $S$ represents a set of all possible operators, $S$, that define sample sets. We have represented the operators in $S$ by $\delta_h$ where $h = 1, \ldots, H$. Notice that Ericson does not include an index for the operator. If $Y_N$ is a vector of random variables, it is possible that different sets would result in the same values of $y_{obs}$. If this is the case, the notation $S$ may indicate this difference. However, if such a subtle meaning was intended, Ericson did not discuss it. For this reason, the interpretation of $S(Y_N) = y_{obs}$ is not clear.

In a likelihood, the realized values are fixed, and the parameters may be considered to be random variables. The realized values are given by $(s_{h,obs}, y_{h,obs})$, and the parameters $Y_N$ are random variables. Since we use a likelihood, we do not use the conditional notation since in the likelihood, $(s_{h,obs}, y_{h,obs})$ is fixed. Since nothing is lost by representing a sample as a set as opposed to a sequence, (i.e. $(s_{h,obs}, y_{h,obs})$ is a sufficient statistic) the posterior distribution given $(s_{h,obs}, y_{h,obs})$ or given $(s_{h,obs}, y_{h,obs})$ are the same. As a result, for a given value of $Y_N$, we find that $l(Y_N; (s_{h,obs}, y_{h,obs})) = l(Y_N; (s_{h,obs}, y_{h,obs}))$. These expressions will be equal to some constant if the value of $\delta_h' Y_N = y_{h,obs}$, and otherwise it will be equal to zero. The constant will be equal to $p'(Y_N | \delta_h' Y_N = y_{h,obs}) = k$, where $k$ could be equal to zero.

Prior Distributions

Ericson discusses a possible prior distribution assuming that the values of $Y_s$ are exchangeable (p198). Suppose there is a parameter vector $\theta = (\theta_1, \theta_2, \ldots, \theta_F)'$, with elements $\theta_f$, $f = 1, \ldots, F$, which we assume has a probability distribution function given by $F(\theta)$ (multivariate). One example occurs when $F = 2$, where $\theta_1 = \mu$ and $\theta_2 = \sigma^2$, and $\theta = (\mu, \sigma^2)'$ are the mean and variance of a finite population. There are two parameters in this prior distribution, corresponding to the mean and variance.

Since this example is simple, we discuss two different settings. Suppose $N = 3$. Using the idea of a superpopulation, let $N = 4$. Assume that the $\binom{N}{N} = 4$ possible subsets, each consisting of $N$ subjects from $\mathbb{N}$, are equally likely, and that the elements of $\theta$ correspond to the mean and variance of each subset. There are four possible realizations of $\theta$, with each equally likely. Next, suppose $N = 3$, there is a superpopulation of $\mathbb{N} = 4$, and we assume that in addition to the four possible realizations of $\theta$ (which we assume are equally likely with probability of 1/8 each) we also consider one additional possible values of $\theta$ (specified subjectively), with probability 1/2. Ericson makes the point that this possibility may not be able to be represented in terms of a superpopulation (see discussion in Ericson 1969, p197-200).
Another example occurs when $\theta = (\beta_0, \beta_1, \sigma^2)'$ correspond to the intercept, slope, and variance of a linear regression. For different realized populations, the values of $\theta$ will differ. If associated with these values is the probability of the population occurring, then these probabilities form the prior distribution of $\theta$. A third example occurs when $\theta = (\gamma_1, \gamma_2, \cdots, \gamma_N)'$, the $N$ values of subjects in a finite population. Suppose the population is a realized sequence from a superpopulation with $N = \mathbb{N}$, with each sequence equally likely. Different sequences result in different values of $\theta$, and specify the prior distribution of $\theta$.

Now, let us also assume that the $Y_s$, $s = 1,...,N$ are independent and identically distributed conditional on $\theta$. This means that $Y_s \mid \theta_{\beta}$ for $s = 1,...,N$ is independent and identically distributed. Let us represent the probability of a particular realization of $Y_s$ by $p\left(Y_s = y_s^{*} \mid \theta_{\beta}\right)$. Since $Y_s$, $s = 1,...,N$ are independent, $p\left(Y_N \mid \theta_{\beta}\right) = \prod_{s=1}^{N} p\left(Y_s = y_s^{*} \mid \theta_{\beta}\right)$.

Ericson then defines the prior distribution as $p'(Y_N) = \int \prod_{s=1}^{N} p\left(Y_s = y_s^{*} \mid \theta\right) dF(\theta)$.

**A Prior Distribution where the Population is All Possible Subjects**

Prior Distribution Includes Response Error

One possible way of considering this prior distribution is as follows. Suppose that $F = N$ so that $\theta = \theta_0 = (\theta_1 = y_1, \theta_2 = y_2, \cdots, \theta_N = y_N)'$, where $y_s$ is the latent value for subject $s$, and $p(\theta = \theta_0) = 1$. Now assume that we do not observe the latent value directly, but that there is response error. Let us represent $Y_s = y_s + E_s$, where $E_s$ corresponds to the response error. We represent a realization of $Y_s$ by $y_s^{*}$, and note that this corresponds to a realization of $E_s$ which we represent by $e_s^{*}$. Suppose now that $E_s$, $s = 1,...,N$ are independent and identically distributed. Let us represent $E_r(E_s) = 0$, and $\text{var}_r(E_s) = \sigma^2$. Let $y_n^{*}$ denote a realization of $Y_N$, and $e_n^{*}$ denote the corresponding realization of $E_N = (E_1, E_2, \cdots, E_N)'$. One setting where $Y_s$, $s = 1,...,N$ will be independent and identically distributed will occur if we assume that $y_s = \mu$ for all $s = 1,...,N$. This problem is described by Ericson (1988) in section 8.1, and Ericson uses it to develop a linear Bayes estimator.

We consider another possible setting here. Suppose that $Y_i^{*}$ represents the response for the subject in position $i$ in a permutation of subjects, where $E_r(Y_i^{*}) = \mu_i$, and $\text{var}_r(Y_i^{*}) = \sigma^2$. Let us further assume that $\mu_i = x_i' \beta$. Notice that $x_i$ is a vector of dimension $1 \times p$ that depends only on position $i$ in a permutation, where $X = (x_1, x_2, \cdots, x_N)'$. For a given value of $\beta$, 

C10ed01.doc 4/28/2011 6:05 PM 7
and conditional on \( \mathbf{X} \),

\[
\mathbf{E}_R \left( Y_1^* \ Y_2^* \ \cdots \ Y_N^* \right)' = \mathbf{X}\boldsymbol{\beta}.
\]

Since the expected response is different for each position, the random variables \( Y_i^* \), \( i = 1, \ldots, N \) are not exchangeable.

We consider a second possibility for understanding such a prior distribution where there is no response error. Suppose that \( F = 1 \) so that \( \theta = \mu_0 \), where \( y_s = \mu + \beta_s \) is the latent value for subject \( s \), \( s = 1, \ldots, N \) and \( p(\theta = \mu_0) = 1 \). We define \( \mu = \frac{1}{N} \sum_{s=1}^{N} y_s = \mu_0 \). Suppose the random variables \( Y_s \), \( s = 1, \ldots, N \) are exchangeable. This implies that possible realizations of \( Y_s \) correspond to \( y_s = \mu + \beta_s \), where \( s = 1, \ldots, N \). Notice that for any realization,

\[
\mu = \frac{1}{N} \sum_{s=1}^{N} y_s = \mu_0.
\]

Under this assumption, \( p\left( Y_s = y_s^* \mid \theta = \mu_0 \right) = \frac{1}{N} \). Using the subscript \( \xi \) to index expectation with respect to this prior distribution,

\[
E_\xi (Y_s) = \mu \text{ for all } s = 1, \ldots, N.
\]

Let us represent \( \mathbf{Y}_s = (Y_1 \ Y_2 \ \cdots \ Y_N)' \), and assume a realization of \( \mathbf{Y}_N \) is \( \mathbf{y}_N = (y_1 \ y_2 \ \cdots \ y_N)' \).

Representing each possible permutation as equally likely (corresponding to the assumption that \( Y_s \), \( s = 1, \ldots, N \) are exchangeable), \( p\left( \mathbf{Y}_N = \mathbf{y}_N \mid \theta \right) = \frac{1}{N!} \). We have assumed \( p\left( \theta = \mu_0 \right) = 1 \) so that

\[
p'(\mathbf{Y}_N) = \int \prod_{s=1}^{N} p\left( Y_s = y_s^* \mid \theta \right) dF(\theta) = \frac{1}{N!}.
\]

**Prior Distributions and Superpopulations**

We can consider this slightly differently by assuming that the \( N \) subjects were selected from a larger superpopulation of \( \mathbb{N} > N \) subjects. Let us assume that the selection is made via simple random without replacement sampling, so that there are \( \binom{\mathbb{N}}{N} \) equally likely sets of \( N \) subjects which we index by \( f = 1, \ldots, F \). We add a subscript \( f \) to indicate a realization of the subjects in the population, so that we represent a realization of \( \mathbf{Y}_N \) by

\[
\mathbf{y}_f = (y_{f1} \ y_{f2} \ \cdots \ y_{fN})'.
\]

Elements in this vector are ordered with consecutively increasing subscripts which we represent by \( j = 1, \ldots, N \) such that the value for a subject is given by \( y_{fj} \).

We can alter this description of a superpopulation slightly if we assume that response is continuous so that a particular response is unique for a subject. In this case, a set of \( N \) responses will identify a set of \( N \) subjects. We assume the possible realizations of response are exchangeable, so that each \( N \times 1 \) vector of responses (representing a sequence) is equally likely with probability \( \frac{\mathbb{N}!}{(\mathbb{N} - N)!} \). Since a response identifies a subject, the response vector identifies \( N \) subjects, which, when re-ordered, correspond to a set which we index by \( f = 1, \ldots, F \). We do
not explicitly defined the random variables, $Y_i$, for $i = 1, ..., N$, in $Y_N$, but discuss their definition later.

Let us index the parameters corresponding the mean latent value for set $f$ by $\theta_f$ for $f = 1, ..., F$ where $\theta_f = \frac{1}{N} \sum_{j=1}^{N} y_{ij}$, and define $\sigma_f^2 = \frac{1}{N-1} \sum_{j=1}^{N} (y_{ij} - \theta_f)^2$. We consider the parameters $\theta_f$ for $f = 1, ..., F$ to be the parameters for the superpopulation.

We have assumed that the subjects in the population are selected from a larger superpopulation. We can represent the population of subjects via a permutation model for the superpopulation. The population is a random vector, where a subject label and value is associated with each position in the vector. Suppose all possible permutations of the first $N$ elements of the superpopulation, which we index by $\tilde{f} = 1, ..., \tilde{F}$, are equally likely. We identify a particular permutation via an indicator random variable, $\tilde{I}_f$, which has a value of one if permutation $\tilde{f}$ is selected, and zero otherwise. Note that $P(\tilde{I}_f = 1) = \frac{1}{(N!)^{(N-N)!}}$. Let us designate the first $i = 1, ..., N$ elements in a permutation of the superpopulation as the population, and refer to an element by $Y_i$, $i = 1, ..., N$. We represent $E_\eta (Y_i) = \theta_0$, and note that $E_\eta (Y_N) = \theta_0 I_N$. Also, note that defining $\bar{Y}_N = \frac{1}{N} \sum_{i=1}^{N} Y_i$, $E_\eta (\bar{Y}_N) = \theta_0$. Since the elements of $Y_N$ correspond to the first $N$ elements in a permutation of the superpopulation, $\text{var}_\eta (Y_i) = \frac{1}{N} \sum_{i=1}^{N} (y_{is} - \theta_0)^2$ and hence $\text{var}_\eta (Y_N) = v \left( I_N - \frac{1}{N} J_N \right)$.

Consider the set of permutations, $\tilde{f}$, where the elements in permutation $\tilde{f}$ correspond to the same set of subjects, $f$. Given $f$, we express the variance of an element $Y_i$, $i = 1, ..., N$, in such a set of permutations as $\sigma_f^2$ and the mean as $\theta_f$. These parameters are conditional on the set of subjects, $f$. Unconditionally, we represent $\sigma_f^2 = \sum_{f=1}^{\tilde{F}} \tilde{I}_f \sigma_f^2$, and the mean of the first $N$ elements in a permutation by $\bar{Y}_N = \sum_{f=1}^{\tilde{F}} \tilde{I}_f \theta_f$.

**Expected Value and Variance of** $\bar{Y}_N = \frac{1}{N} \sum_{i=1}^{N} Y_i$ and $\bar{Y} = \frac{1}{n} (1^n \ 0'_{N-n}) Y_N$

We can use these results to develop expressions for linear combinations of random variables associated with a population, and the mean of the first $n$ elements in the population, where $\bar{Y} = \frac{1}{n} (1^n \ 0'_{N-n}) Y_N$. Notice that $\bar{Y}_N = \frac{1}{N} 1'_N Y_N$ so that
\[ E_{\xi} (\bar{Y}^2_N) = \frac{1}{N} 1'_{N} E_{\xi} (Y_N) = \theta_0, \]
\[ E_{\xi} (\bar{Y}) = \frac{1}{n} 1'_{n} E_{\xi} (Y_N) = \theta_0. \]

We evaluate the variance in a similar manner such that
\[ \text{var}_{\xi} (\bar{Y}^2_N) = \frac{1}{N^2} 1'_{N} \text{var}_{\xi} (Y_N) 1_N \]
\[ = \frac{v}{N^2} 1'_{N} \left( I_N - \frac{1}{N} J_N \right) 1_N \]
\[ = v \frac{1}{N} \left( 1 - \frac{N}{N} \right) \]

and
\[ \text{var}_{\xi} (\bar{Y}) = \frac{1}{n^2} (1'_{n} 0'_{n-N}) \text{var}_{\xi} (Y_N) \left( \begin{array}{c} 1_n' \\ 0_{n-N} \end{array} \right) \]
\[ = \frac{v}{n^2} (1'_{n} 0'_{n-N}) \left( I_N - \frac{1}{N} J_N \right) \left( \begin{array}{c} 1_n' \\ 0_{n-N} \end{array} \right) \]
\[ = v \frac{1}{n} \left( 1 - \frac{n}{N} \right) \]

**Expected Value of** \( \sigma^2 = \sum_{j=1}^{F} \bar{I}_j \sigma_j^2 \)

We evaluate the expected value using the definition of \( \sigma_j^2 \) such that
\[ E_{\xi} (\sigma^2) = \frac{1}{N!} \sum_{j=1}^{F} \sigma_j^2 \]
\[ = \frac{N!}{(N-N)!} \sum_{j=1}^{F} \sigma_j^2 \]
\[ = \frac{N!(N-N)!}{N!} \frac{1}{N-1} \sum_{j=1}^{F} \sum_{f=1}^{N} (y_{jf} - \theta_j)^2 \]

We simplify this expression. Now \( \sum_{j=1}^{F} \sum_{f=1}^{N} (y_{jf} - \theta_j)^2 = \sum_{f=1}^{F} \sum_{j=1}^{N} y_{jf}^2 - N \sum_{j=1}^{N} \theta_j^2 \) and
\[ \nu = \frac{1}{N-1} \sum_{x=1}^{N} y_x^2 - \frac{N}{N-1} \theta_0^2. \] Also, \( \sum_{f=1}^{F} \sum_{j=1}^{N} y_{jf}^2 = \frac{(N-1)!}{(N-N)!(N-1)!} \sum_{x=1}^{N} y_x^2 \) and since \( \bar{Y}_N = \sum_{j=1}^{F} \bar{I}_j \theta_j \),
\[
\text{var}_s(\bar{X}_N) = \left( \frac{N!}{(N-N)!} \right) \sum_{j=1}^{F} \theta_j^2 - \left[ \text{E}_s(\bar{X}_N) \right]^2
\]
\[
= \frac{(N-N)!N!}{N!} \sum_{j=1}^{F} \theta_j^2 - \theta_0^2
\]

Since we can also express \( \text{var}_s(\bar{X}_N) = \frac{1}{N} \left( 1 - \frac{N}{N} \right) \), equating these expressions,
\[
\sum_{j=1}^{F} \theta_j^2 = \frac{N!}{(N-N)!N!} \left( \frac{N-N}{N} \right)^v + \frac{N!}{(N-N)!N!} \theta_0^2.
\]

We substitute these expressions into
\[
\text{E}_s(\sigma^2) = \frac{N!(N-N)!}{N!} \left( \frac{1}{N-1} \right) \sum_{j=1}^{F} \sum_{j=1}^{N} (y_j - \theta_j)^2
\]
\[
= \frac{N!(N-N)!}{N!} \left( \frac{1}{N-1} \right) \sum_{j=1}^{F} \sum_{j=1}^{N} y_j^2 - N \left[ \frac{N!(N-N)!}{N!} \left( \frac{1}{N-1} \right) \sum_{j=1}^{F} \theta_j^2 \right].
\]

As a result,
\[
\text{E}_s(\sigma^2) = \frac{N!(N-N)!}{N!} \left( \frac{1}{N-1} \right) \left( \frac{(N-1)!}{(N-N)!} \right) \sum_{j=1}^{N} y_j^2
\]
\[
- N \left[ \frac{N!(N-N)!}{N!} \left( \frac{1}{N-1} \right) \left( \frac{N-N}{N} \right)^v + \frac{N!}{(N-N)!N!} \theta_0^2 \right]
\]
\[
= \frac{1}{(N-1)} \left[ \frac{(N-1)!}{N!} \right] \sum_{j=1}^{N} y_j^2 - \frac{1}{N-1} \left( \frac{N-N}{N} \right)^v - N \frac{\theta_0^2}{N-1}
\]
\[
= \frac{N}{N-1} \left[ \frac{1}{N} \sum_{j=1}^{N} y_j^2 - \theta_0^2 \right] - \frac{1}{N-1} \left( \frac{N-N}{N} \right)^v
\]

Now \( \frac{1}{N} \sum_{j=1}^{N} y_j^2 - \theta_0^2 = \frac{N-1}{N} \). As a result,
\[
\text{E}_s(\sigma^2) = \frac{1}{N} \sum_{j=1}^{N} y_j^2 - \theta_0^2 = \frac{N-1}{N} \frac{N}{N-1}
\]
\[
= \frac{N}{N-1} \frac{N-1}{N} \left( \frac{N-N}{N} \right)^v
\]
\[
= \frac{N(N-N)-N+N}{(N-1)N} \frac{N}{N-1} \left( \frac{N-N}{N} \right)^v
\]
\[
= \frac{N}{N-1} \frac{N}{N} \left( \frac{N-N}{N} \right)^v
\]
\[
= v
\]

**Explicit Definition of** \( Y_i \)
Notice that we have represented \( \bar{Y}_N = \sum_{f=1}^{F} \tilde{I}_f \theta_f \) and defined \( \theta_f = \frac{1}{N} \sum_{j=1}^{N} y_{fj} \), but have not explicitly defined \( Y_i \). Since in general, the realization of \( Y_i \) can be the value for any subject in the superpopulation, referring to these values explicitly requires identifying the subjects in the superpopulation. This is awkward since the subjects in the superpopulation are conceptual, not an explicit list of subject. We have introduced some explicit notation for such subjects when we expressed \( v = \frac{1}{N-1} \sum_{s=1}^{N} (y_{s} - \theta_0)^2 \), but note that Ericson did not define subscripts \( s^* = 1, \ldots, N \), or even indicate the size of the superpopulation (which is usually taken to be essentially infinite). By conditioning on \( \theta_f \) and \( \sigma_f^2 \), it is not necessary to refer to all the elements in the superpopulation.

We explicitly define \( Y_i \) next. To do so, we first define possible realizations of \( Y_i \) as \( y_{fj} \) for \( f = 1, \ldots, F \). We have already introduced representation of realizations of all possible sets of \( N \) subjects from the \( N \) subjects in the superpopulation which we index by \( f = 1, \ldots, F \). The value of subject \( j \) in set \( f \) is represented by \( y_{fj} \) for \( j = 1, \ldots, N \). A vector of these responses for set \( f \) is given by \( y_f = \left( y_{f1}, y_{f2}, \ldots, y_{fN} \right)' \). Let \( m^* = 1, \ldots, M^* \) index all possible permutations of the subjects in set \( f \), and let us represent an \( N \times N \) permutation matrix \( v_{m}^{(N)} \) such that \( y_{fm} = v_{m}^{(N)} y_f \) represents a permutation of subjects in set \( f \). Elements of \( v_{m}^{(N)} \) are given by \( v_{m_{ij}} \) where

\[
\begin{pmatrix}
V_{m_{11}} & V_{m_{12}} & \cdots & V_{m_{1N}} \\
V_{m_{21}} & V_{m_{22}} & \cdots & V_{m_{2N}} \\
\vdots & \vdots & \ddots & \vdots \\
V_{m_{N1}} & V_{m_{N2}} & \cdots & V_{m_{NN}}
\end{pmatrix}
\]

I would like to introduce some notation for the elements of \( y_{fm} = \left( y_{fm1}, y_{fm2}, \ldots, y_{fMN} \right)' \), where \( y_{fm_i}, i = 1, \ldots, N \) is response for the subject in set \( f \) in position \( i \) in permutation \( m^* \).

We have represented the first \( N \) elements in a permutation of the superpopulation by \( y_i, i = 1, \ldots, N \), and indexed realizations of these permutations by \( \tilde{f} \). There is a mapping of permutation \( \tilde{f} \) to permutation \( m^* \) of set \( f \), so that \( y_{\tilde{f}_i} = y_{fm_i} \). Representing a realization of \( Y_N \) by \( y_f = \left( y_{f1}, y_{f2}, \ldots, y_{fN} \right)' \), \( y_f = y_{fm^*} \). Using this notation,

\[
Y_N = \sum_{f=1}^{F} \tilde{I}_f y_f = \sum_{m^*=1}^{M^*} \sum_{f=1}^{F} I_{m^*}^{(f)} I_f y_{fm^*}.
\]

As a result,

\[
Y_i = \sum_{m^*=1}^{M^*} \sum_{f=1}^{F} I_{m^*}^{(f)} I_f y_{fm^*}.
\]
Conditional Expected Value of the Sample Mean
\[ \overline{Y} = \frac{1}{n} \left( \mathbf{1}_n' \mathbf{0}'_{N-n} \right) \mathbf{Y}_N \]

Next, we evaluate expressions for the expected value of the sample mean, given a particular set of subjects identified by \( \theta_f \) and \( \sigma_j^2 \). The sample mean is given by
\[ \overline{Y} = \frac{1}{n} \left( \mathbf{1}_n' \mathbf{0}'_{N-n} \right) \mathbf{Y}_N. \] We evaluate \( E \left( \overline{Y} \mid \theta_f, \sigma_j^2 \right) \) and \( \text{var} \left( \overline{Y} \mid \theta_f, \sigma_j^2 \right) \). Notice that conditioning on the mean and variance is similar to, but not the same as conditioning on the set of subjects in the population. We assume that if response is continuous, there is a one-to-one correspondence between conditioning on the mean and variance, and conditioning on \( I_f \). This implies that
\[ \left( \mathbf{Y}_N \mid \theta_f, \sigma_j^2 \right) = \left( \mathbf{Y}_N \mid I_f = 1 \right). \]

Now \( \left( \mathbf{Y}_N \mid I_f = 1 \right) = \sum_{m=1}^M f^{(f)}_m \mathbf{y}_{fm}. \) As a result, the expectation is with respect to permutations of subjects in the finite population, which we index by \( \xi_i \). The expectation is given by
\[ E \left( \overline{Y} \mid \theta_f, \sigma_j^2 \right) = \frac{1}{n} \left( \mathbf{1}_n' \mathbf{0}'_{N-n} \right) E_{\xi_i} \left( \mathbf{Y}_N \mid \theta_f, \sigma_j^2 \right). \]

Now \( \mathbf{y}_{fm} = \mathbf{v}_{m; f}^{(N)} \mathbf{y}_f \) so that
\[ E \left( \overline{Y} \mid \theta_f, \sigma_j^2 \right) = \frac{1}{n} \left( \mathbf{1}_n' \mathbf{0}'_{N-n} \right) E_{\xi_i} \left( \sum_{m=1}^M f^{(f)}_m \mathbf{y}_{fm}^{(N)} \right) \mathbf{y}_f. \]

\[ = \frac{1}{n} \left( \mathbf{1}_n' \mathbf{0}'_{N-n} \right) \left( \frac{1}{N} \mathbf{J}_N \mathbf{y}_f \right) \quad = \theta_f. \]

Conditional Variance of the Sample Mean
\[ \overline{Y} = \frac{1}{n} \left( \mathbf{1}_n' \mathbf{0}'_{N-n} \right) \mathbf{Y}_N \]

We evaluate the conditional variance by expressing,
\[ \text{var} \left( \overline{Y} \mid \theta_f, \sigma_j^2 \right) = \frac{1}{n^2} \left( \mathbf{1}_n' \mathbf{0}'_{N-n} \right) \text{var}_{\xi_i} \left( \mathbf{Y}_N \mid \theta_f, \sigma_j^2 \right) \begin{pmatrix} \mathbf{1}_n \\ \mathbf{0}'_{N-n} \end{pmatrix}. \]

\[ = \frac{1}{n^2} \left( \mathbf{1}_n' \mathbf{0}'_{N-n} \right) \text{var}_{\xi_i} \left( \mathbf{Y}_N \mid I_f = 1 \right) \begin{pmatrix} \mathbf{1}_n \\ \mathbf{0}'_{N-n} \end{pmatrix}. \]

\[ = \frac{1}{n^2} \left( \mathbf{1}_n' \mathbf{0}'_{N-n} \right) \text{var}_{\xi_i} \left( \sum_{m=1}^M f^{(f)}_m \mathbf{y}_{fm}^{(N)} \right) \begin{pmatrix} \mathbf{1}_n \\ \mathbf{0}'_{N-n} \end{pmatrix}. \]
Using $y_{mf} = v_{m}^{(N)}y_{f}$,

$$
\sum_{m=1}^{M} f^{(f)}_m y_{mf} = \left( \sum_{m=1}^{M} f^{(f)}_m v_{m}^{(N)} \right) y_{f}
$$

$$
= V y_{f}
$$

where $V = \sum_{m=1}^{M} f^{(f)}_m v_{m}^{(N)}$. Now $V y_{f} = (I_N \otimes y_{f'}) vec(V')$. Since $V$ is a permutation matrix, from Argentina2007-lec1.doc, $\text{var}_{\xi_{ni}} \left[ vec(V') \right] = \frac{1}{N-1} P_N \otimes P_N$. Hence,

$$
\text{var}_{\xi_{ni}} (V y_{f}) = (I_N \otimes y_{f'}) \text{var}_{\xi_{ni}} \left[ vec(V') \right] (I_N \otimes y_{f})
$$

$$
= \frac{1}{N-1} (I_N \otimes y_{f'}) P_N \otimes P_N (I_N \otimes y_{f})
$$

$$
= \sigma_f^2 P_N
$$

since $\sigma_f^2 = \frac{1}{N-1} y_{f'}' P_N y_{f'}$. Using these results,

$$
\text{var}(\bar{Y} | \theta_f, \sigma_f^2) = \frac{1}{n^2} (1_n' \quad 0_{N-n}') \text{var}_{\xi_{ni}} (V y_{f}) \left( \begin{array}{c} 1_n \\ 0_{N-n} \end{array} \right)
$$

$$
= \sigma_f^2 \frac{1}{n^2} (1_n' \quad 0_{N-n}') P_N \left( \begin{array}{c} 1_n \\ 0_{N-n} \end{array} \right)
$$

$$
= \frac{\sigma_f^2}{n} \left( \frac{N-n}{N} \right)
$$

Using the notation given above, Ericson’s RESULT 4.1 (1988, p218) can be stated as:

RESULT 4.1. If the random variables $Y_1, Y_2, \ldots, Y_N$, are exchangeable with means, variances and covariances as defined above, where $\bar{Y}_N = \frac{1}{N} I_N' Y_N$ and $\bar{Y} = \frac{1}{n} (1_n' \quad 0_{N-n}') Y_N$, then

$$
E_{\xi} (\bar{Y}_N) = E_{\xi} (\bar{Y}) = \theta_0,
$$

$$
\text{var}_{\xi} (\bar{Y}_N) = \nu \frac{1}{N} \left( 1 - \frac{N}{N} \right),
$$

$$
E_{\xi} (\sigma^2) = \nu,
$$

$$
\text{var}_{\xi} (\bar{Y}) = \nu \frac{1}{n} \left( 1 - \frac{n}{N} \right)
$$

$$
E \left( \bar{Y} | \theta_f, \sigma_f^2 \right) = \theta_f,
$$

and

$$
\text{var} \left( \bar{Y} | \theta_f, \sigma_f^2 \right) = \frac{\sigma_f^2}{n} \left( \frac{N-n}{N} \right).
$$

Relationship with Finite Population Sampling
Ericson (1988, p219) states that

“Note that these properties are those of the prior distribution only, no sampling design has been specified. The conditional mean and variance of \( \bar{Y} \), (given above), are in the exact form as for a simple random sample design. The analogy is even closer, for it may also be shown that under an exchangeable prior on \( Y_N \), given the collection of the \( N \) population variate values, \( Y_i \)’s, but not the units to which they are attached, the probability that any prespecified subset of \( n \) of the population elements will assume the values given by any collection of \( n \) of the \( N \) population values is precisely the same as the objective probability that the subset was selected by simple random sampling, namely \( 1 / \binom{N}{n} \).”

In Ericson’s paper, there is often reference to “an exchangeable prior on \( Y \)”. Typically, the dimension of \( Y \) is not indicated. We can, for example, consider \( Y \) to be an \( N \times 1 \) vector, an \( N \times 1 \) vector, or an \( n \times 1 \) vector and still refer to “an exchangeable prior on \( Y \)”. The difference between these definitions occurs when evaluating the expected value or variance.

In fact, it is tempting to view the prior distribution as simply a permutation distribution of the \( N \) subjects. This is the distribution used for simple random sampling without replacement. With this understanding, although no sampling design has been specified, the exact sample probability distribution is used, so that the difference is simply whether the distribution is referred to as a sampling distribution, or an exchangeable distribution.

Ericson emphasizes that only the value, and ‘not the units to which they are attached’ are pre-specified. If response is continuous and the population is finite, a value will uniquely identify a subject. When this is the case, a collection of values is given by \( y_f \). This collection of values corresponds to identifiable subjects. However, associated with this collection of values are \( N! \) sequences, with values in the sequence given by \( y_f = y_{fm} = y^{(N)}_{m}y_f \) for \( m = 1, \ldots, N! \).

Relative to these sequences, Ericson’s comment above makes perfect sense. Given a collection of \( N \) population values (namely \( y_f \)), if we prespecify the sample subjects, there is one sequence that will have the sample subject’s values match those for the pre-specified sample.

Ericson (1988) states that

“The superpopulation concept seems to be nothing but a partial specification of a prior distribution. The ‘superpopulation’ is a subjective artifact of no practical importance per se. It is true that one might, at times, specify an exchangeable prior distribution on \( Y \) via

\[
p(Y \mid X) = \int \prod_{i=1}^{N} f(Y_i \mid X, \theta) dF(\theta \mid X)
\]

- a mixture of independent, identically distributed random variables conditional on \( \theta \). The view here is that the specification of a superpopulation via \( \prod_{i=1}^{N} f(Y_i \mid X, \theta) \) is rather subjective and should be part of a prior distribution and not be combined with \( p_X(s) \) in order to yield a more informative ‘likelihood’ than that of (2.1).”
where (2.1) is given by

$$I(Y; (s, y_s)) = \begin{cases} kp_x(s) \text{ for } Y = y_s \\ 0 \end{cases}. $$

Bayes Linear Estimator

Ericson (1988, p219) states a theorem and corollary and uses them to evaluate a linear Bayes estimator in the context of an exchangeable prior distribution. We develop these results using the previous notation. Suppose we have an exchangeable prior on $Y_N$ in the context of the previous discussion of a superpopulation. Let us represent a sample of $n$ subjects by $(s_{h,obs}, y_{h,obs})$ where $s_{h,obs} = v'_m s_{h,obs}$ and $y_{h,obs} = v'_m y_{h,obs}$. We are now ready to apply Ericson’s Theorem 3.1 (page 216). We first state the theorem using Ericson’s notation, and then discuss it with respect to previous work on finite population sampling.
Ericson's Theorem 3.1 Suppose that $X_1$ ($n_1 \times 1$) and $X_2$ ($n_2 \times 1$) are any jointly distributed random vectors having means $\mu_1$ and $\mu_2$ and covariance matrix

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$ 

If

$$E(X_1 | X_2) = PX_2 + \gamma$$

for some $P$ ($n_1 \times n_2$) and $\gamma$ ($n_1 \times 1$) not depending on $X_2$, or if $P$ and $\gamma$ are chosen to minimize

$$E_{X_2} \left\| E(X_1 | X_2) - (PX_2 + \gamma) \right\|^2,$$

then

$$PX_2 + \gamma = \Sigma_{12} \Sigma_{22}^{-1} X_2 + (\mu_1 - \Sigma_{12} \Sigma_{22}^{-1} \mu_2).$$

Furthermore,

$$E_{X_2} \left[ \text{var}(X_1 | X_2) \right] - \left( \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \right) \leq 0,$$

that is, the matrix on the left is negative, semi-definite, with equality holding if $E(X_1 | X_2) = PX_2 + \gamma$.

Ericson’s Theorem 3.1 is very similar to the derivation of the BLUP in a mixed model (following Royall (1976)). In order to see the parallel, we switch the labels for the first and second set of variables, stating Ericson’s theorem as follows.

Ericson’s Theorem 3.1v1 Suppose that $Y_1$ ($n_1 \times 1$) and $Y_2$ ($n_2 \times 1$) are any jointly distributed random vectors having means $\mu_1$ and $\mu_2$ and covariance matrix

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$ 

If

$$E(Y_2 | Y_1) = PY_1 + \gamma$$

for some $P$ ($n_2 \times n_1$) and $\gamma$ ($n_2 \times 1$) not depending on $Y_1$, or if $P$ and $\gamma$ are chosen to minimize

$$E_{Y_1} \left\| E(Y_2 | Y_1) - (PY_1 + \gamma) \right\|^2,$$

then

$$PY_1 + \gamma = \Sigma_{21} \Sigma_{11}^{-1} Y_1 + (\mu_2 - \Sigma_{21} \Sigma_{11}^{-1} \mu_1).$$

Furthermore,

$$E_{Y_1} \left[ \text{var}(Y_2 | Y_1) \right] - \left( \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \right) \leq 0,$$

that is, the matrix on the left is negative, semi-definite, with equality holding if $E(Y_2 | Y_1) = PY_1 + \gamma$. 

C10ed01.doc 4/28/2011 6:05 PM 17
In terms of this version of Ericson’s theorem (Note: Ericson expresses vectors as row vectors), the Corollary is given by

Ericson’s Corollary 3.1v1. Suppose that $Y (= Y_1)$ $(n \times 1)$ and $\xi (= Y_2)$ $(1 \times 1)$ are any $n + 1$ jointly distributed random variables such that $E(\xi) = m$, var$(\xi) = v(\xi) < \infty$, var$(Y) = \Sigma_{11}$ is positive definite, and $E(Y_i | \xi) = \xi$, for $i = 1, ..., n$. If either

$$E(\xi | Y) = Y'a + b$$

where $a = (a_1, a_2, \ldots, a_n)'$, or $a$ and $b$ are chosen to minimize

$$E_\nu \left[ E(\xi | Y) - (Y'a + b) \right]^2,$$

then

$$Y'a + b = \frac{E(\xi) E[V(\hat{\xi} | \xi)] + \hat{\xi}v(\xi)}{V(\hat{\xi}) + E[V(\hat{\xi} | \xi)]},$$

where $\hat{\xi} = \hat{\xi}(Y) = (1', \Sigma_{11}^{-1} 1_n)^{-1} 1_n, \Sigma_{11}^{-1} Y$ is the usual BLUE or weighted least squares estimator of $\xi$ with respect to either the variance-covariance matrix $\Sigma_{11}$ or $E_\nu[ V(Y | \xi) ] = E_\nu[ \Sigma_{11} | \xi ]$. Also,

$$E_\nu \left[ V(\xi | Y) \right] \leq v(\xi) \left[ 1 - 1_n' \Sigma_{11}^{-1} 1_n \right] = v(\xi) \left( \frac{E[V(\hat{\xi} | \xi)]}{v(\xi) + E[V(\hat{\xi} | \xi)]} \right)$$

with equality if $E(\xi | Y) = Y'a + b$.

Interpreting $E(\xi | Y)$

Two conditions are given by Ericson in the corollary. These conditions implicitly define the target quantity of interest. This target is $E(\xi | Y)$. In order to interpret this target, we need to be able to interpret $Y$ and $\xi$. Ericson provides such an interpretation. To place the problem in the context of simple random sampling of a finite population, we use the notation given in Result 4.1 (p218 of Ericson 1988, as given previously on page 14).

RESULT 4.1. If the random variables $Y_1, Y_2, ..., Y_N$, are exchangeable with means, variances and covariances as defined above, where $\bar{Y}_N = \frac{1}{N} 1_N Y_N$ and $\bar{Y} = \frac{1}{n} (1_n' 0'_{N-n}) Y_N$, then

$$E_\xi(\bar{Y}_N) = E_\xi(\bar{Y}) = \theta_0,$$

$$\text{var}_\xi(\bar{Y}_N) = v \left( \frac{1}{N} \left( 1 - \frac{N}{N} \right) \right),$$

$$E_\xi(\sigma^2) = v,$$

$$\text{var}_\xi(\bar{Y}) = v \left( \frac{1}{n} \left( 1 - \frac{n}{N} \right) \right),$$

$$E(\bar{Y} | \theta_f, \sigma^2) = \theta_f,$$

and

$$\text{var}(\bar{Y} | \theta_f, \sigma^2) = \frac{\sigma^2}{n} \left( \frac{N - n}{N} \right).$$
Let us assume that the prior distribution of $N_Y$ is exchangeable, where $N$ corresponds to the number of subjects in a population. We select a subset of $n$ of the $N$ subjects, and refer to the subset by $h$. The labels of the subjects in the subset $h$ are given by $s_{h,obs} = \delta_{s}$. The values of the subjects in the subset are given by $y_{h,obs} = \delta_{y}$. One way of considering the $N$ subjects in the population is to consider them to be a realization of $N$ random variables defined relative to a superpopulation. Let $f$ index realizations of such populations, and $m^*$ index permutations of the subjects in population $f$. With this idea, we represent the elements of $Y_N$ by

$$Y = \sum_{m=1}^{M} \sum_{f=1}^{F} I_{m}^{(f)} I_{f} y_{f m^*}^{'}. \text{ The random variables, } Y_i, i = 1, \ldots, N \text{ are exchangeable. Note that } h \text{ indexes different sets of subjects in a sample, while } f \text{ indexes different finite populations.}
$$

In this context, $\xi = \bar{Y}_N$, the finite population mean. The values in subset $h$ of population $f$ that are observed in the sample are $y_{h,obs}$. Ericson describes these values as a realization of $Y$. This is not correct, since $Y$ corresponds to a sequence of random variables, and not a set. A realization of $Y$ is a realization of a sequence of random variables, given by $y_{h,obs} = \bar{u}_h y_N$, not $y_{h,obs}$.

In order to make this distinction clear, it is helpful to represent the problem using explicit notation for random variables whose realization is the finite population. Recall that in the context of finite population mixed models, we have defined $U = \sum_{h=1}^{H} \bar{I}_h \bar{u}_h$, where $\bar{I}_h = I_m^{(h)} I_{h}$ is an indicator random variable for sequence $h$, elements of $U$ are $U_{is}$, and $\bar{u}_h = v_m \delta_{h}$. Let $S = (\bar{I}_1, \bar{I}_2, \ldots, \bar{I}_H)$ where $\bar{H} = \frac{N!}{(N-n)!}$. We have represented a realization of $Y_N$ by $y_f = (y_{f1}, y_{f2}, \ldots, y_{fN})^{'},$ but have not explicitly defined random variables that represent the connection between these realizations and the superpopulation. For example, in the previous paragraph, we considered $Y_N$ to be a sequence of random variables. When the elements represent random variables in a sequence, $Y_i = \sum_{m=1}^{M} \sum_{f=1}^{F} I_{m}^{(f)} I_{f} y_{f m^*}^{'}$, they are exchangeable.

Possibly, $Y_h$ could refer to a set. If it did, then the elements would not be exchangeable.

For ease of discussion, we assume that there are $N$ subjects in a superpopulation and represent the response of these subjects (in increasing label order) by $z = (z_1, z_2, \ldots, z_N)^{'}$, where $z_t$ is response for subject $t$ in the superpopulation, and $t = 1, \ldots, N$. We imagine all possible permutations of the $N$ subjects. Let us index distinct sequences of the first $N$ elements of these permutations by $\tilde{f} = 1, \ldots, \tilde{F}$, and identify a particular permutation via an indicator random variable, $\tilde{I}_{\tilde{f}}$, which has a value of one if permutation $\tilde{f}$ is selected, and zero otherwise.
Note that \( P(\tilde{I}_j = 1) = \frac{1}{\binom{N}{N!}} \). Let us define an \( \mathbb{N} \times N \) matrix \( \tilde{q}_j \) with elements equal to zero or one such that the sum of each column is equal to one, and the sum of each row is zero or one. Using this notation, \( Y_N = \sum_{f=1}^{F} \tilde{I}_j \tilde{q}_j' \mathbf{z} \). We can express \( \tilde{q}_j' = r'_p t'_f \) where \( t'_f \) is an \( \mathbb{N} \times N \) matrix of indicator variables that identify the subjects in population \( f \), \( f = 1, ..., F \), where \( F = \frac{N!}{(N-N)!N!} \) and \( r_p \), \( p = 1, ..., P = N! \) is an \( N \times N \) permutation matrix that identifies the sequence (permutation) of the subjects in population \( f \). [Notice that this represents a change in notation from the earlier part of this document where we previously let \( m' = 1, ..., M' \) index all possible permutations of the subjects in set \( f \).] Also, let \( \tilde{I}_j = I^{(f)}_p I_f \) represent indicator random variables for population \( f \) and sequence \( p \) of population \( f \). Using this notation, \( Y_N = \sum_{p=1}^{P} \sum_{f=1}^{F} I^{(f)}_p I_f r'_p t'_f \mathbf{z} \). We represent a realization of \( Y_N \) by \( y_{pN} = r'_p t'_f \mathbf{z} \), where we include the subscripts \( p \) and \( f \) to represent the particular sequence \( p \) of the subjects in population \( f \). All elements, \( Y_i, i = 1, ..., N \), of \( Y_N \) have the same expectation and variance. Notice that a different vector, \( Y_{fN} = \sum_{f=1}^{F} I_f t'_f \mathbf{z} \), contains the same subjects as \( Y_N \), but the elements of this vector given by \( Y_{f_i}, i = 1, ..., N \), do not have the same expectation and variance.

I would like to relate this to \( Y \), the \( n \times 1 \) vector that represents the sample sequences of subjects. To do so, we use the representation of \( Y = \sum_{h=1}^{H} I_h \tilde{u}_h' Y_N \) or \( Y = \sum_{m=1}^{M} \sum_{h=1}^{H} I^{(h)}_m I_h \tilde{v}_m \delta_h' y_{pN} \). Noting that \( y_{N} \) is just one possible realization of a finite population selected from a superpopulation, we identify the particular sequence \( p \) of the subjects in population \( f \) by \( y_{pN} = r'_p t'_f \mathbf{z} \). As a result, \( Y = \sum_{m=1}^{M} \sum_{h=1}^{H} I^{(h)}_m I_h \tilde{v}_m \delta_h' y_{pN} \). This representation of the \( n \) random variables in a sample is conditional on the sequence of subjects in the finite population, \( y_{pN} \).

There are several possible ways to define random variables in a sample. When the realized value of \( I_f \) is one, the sample random variables are limited to values taken on by subjects in population \( f \). When the realized value of \( I_h \) is one, the random variables are limited to values taken on by subjects in set \( h \). When the realized value of \( I^{(f)}_p \) is one, a particular permutation of elements of population \( f \) is specified. When the realized value of \( I^{(h)}_m \) is one, a particular permutation of elements of set \( h \) is specified. In order for the random variables in the sample to be exchangeable, we need to include at a minimum one of the random variables, \( I^{(f)}_p \) or \( I^{(h)}_m \).
Some General Thoughts

We provide some general discussion here. We want the random variables giving rise to a realization in a sample to be exchangeable. This requirement does not define a population or superpopulation. For example, we can define exchangeable random variables for a sample if \( n = N = \mathbb{N} \), as well as if \( n < N \leq \mathbb{N} \). With Godambe’s ideas, the first setting may be the one we need to focus on. Notice that the setting \( n = N = \mathbb{N} \) may be interesting even in the context of a superpopulation, since we can imagine realization of a sequence of subjects in the superpopulation, and a sample of subjects that are identifiable. More generally, the different settings may have different definitions of \( \xi \).

A subtle variation of the issue of exchangeability is whether the position in a sample corresponds to an identifiable subject, where the random variable for the subject is exchangeable is some other sense. For example, suppose with \( n = 2 \), we can identify the subjects in position \( i = 1 \) and \( i = 2 \). Suppose however, that these subjects are considered to be realizations of subjects from a superpopulation. In the sample, the subjects are identifiable, while in the superpopulation, they are exchangeable.

Permutations Tied to Positions, not Subjects

We have used nested notation to define the random variables \( I_p^{(f)} \) and \( I_m^{(h)} \). Is this nested notation necessary? Could these random variables be specified without nesting? It seems that nesting is not necessary. We consider this question next. To do so, let us focus on a single random variable, \( I_m^{(h)} \), that is an indicator for permutation \( m \) of subjects in set \( h \). It may be easiest to consider a simple example.

Suppose that \( n = 2 \) and \( N = 3 \), where the subjects in the finite population have values given by \( y_1, y_2, \) and \( y_3 \). Consider the permutation matrices \( v_m' \) for \( m = 1, \ldots, M = 2 \) where

\[
v_1' = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad v_2' = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.
\]

We have associated a random variable with these matrices given by \( I_m^{(h)} \), where \( h \) indexes possible sets of \( n = 2 \) subjects. The \( h \) is not needed, and it is sufficient to index the random variable solely by \( m \), i.e. \( I_m \). A similar consideration applied to permutations of subjects in a realized population from a superpopulation, where we replace \( I_p^{(f)} \) by \( I_p \).

Possible Representations of Exchangeable Random Variables

We list possible ways of defining exchangeable random variables in a sample, where the elements of \( Y \) are exchangeable. The first possibilities is
\[ Y = \sum_{m=1}^{M} \sum_{h=1}^{H} I_m I_h Y_{m} V'_{m} \delta_{h} Y_{N} \]
\[ = \sum_{m=1}^{M} \sum_{h=1}^{H} I_m I_h V'_{m} \delta_{h} \sum_{p=1}^{F} I_p I_{f} r'_{j} z, \]
\[ = \sum_{m=1}^{M} \sum_{h=1}^{H} \sum_{p=1}^{F} I_p I_{f} I_m I_h V'_{m} \delta_{h} r'_{j} z \]

where each sample random variable could have a realized value equal to any value in the superpopulation. The second possibility is
\[ Y_f = \sum_{m=1}^{M} \sum_{h=1}^{H} I_m I_h V'_{m} \delta_{h} r'_{j} z, \]
where each sample random variable could have a realized value equal to any value among the values in the set \( f \) of \( N \) subjects in a finite population. The third possibility is
\[ Y_{pf} = \sum_{m=1}^{M} \sum_{h=1}^{H} I_m I_h V'_{m} \delta_{h} r'_{j} z, \]
where each sample random variable could have a realized value equal to any value among the values in the set \( f \) of \( N \) subjects in a finite population, but where the subjects in the population are placed in a different order (note that it appears that there is a one to one correspondence between sample points in \( Y_{pf} \) and \( Y_{r} \), with the exception of different assignment of subjects to axis labels). The fourth possibility is
\[ Y_{hpf} = \sum_{m=1}^{M} I_m V'_{m} \delta_{h} r'_{j} z, \]
where each sample random variable could have a realized value equal to any value among the values in set \( h \) of \( n \) subjects selected from the \( N \) subjects in finite population \( f \). The fifth possibility is
\[ Y_{m} = \sum_{h=1}^{H} \sum_{p=1}^{F} I_p I_{f} I_h V'_{m} \delta_{h} r'_{j} z, \]
where the permutation of subjects in a sample is fixed, but the order of subject in the population is random, and the population of subjects are random. We illustrate this last possibility with an example.

**An Illustration of** \( Y_{m} = \sum_{h=1}^{H} \sum_{p=1}^{F} I_p I_{f} I_h V'_{m} \delta_{h} r'_{j} z \)

We illustrate the random variables in a sample, \( Y_{m} = \left( Y_{1}^{(m)} \quad Y_{2}^{(m)} \right)^{t} \) of size \( n = 2 \) from a population of size \( N = 3 \) where \( N = N \). Let \( z = (z_{1} \quad z_{2} \quad z_{3})^{t} \), where \( z_{t}, t = 1, \ldots, N \) are labels for subjects in the superpopulation. In this example, since \( N = N \), the only subset of the superpopulation is a proper subset, where \( t_{f} = I_{f} \) and \( f = 1, \ldots, F = 1 \), with \( P(I_{f} = 1) = 1 \). For this reason, when \( N = N \)
\[ Y_m = \sum_{h=1}^{H} \sum_{p=1}^{P} I_p I_h v'_m \delta'_h r'_p z. \]

There are \( p = 1, \ldots, P = N! = 6 \) possible permutations of the elements of \( z \), and each are equally likely. These permutations result in the following superpopulation response vectors:

\[
\begin{align*}
\mathbf{r}_1' &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, & \mathbf{r}'_2 &= \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}, & \mathbf{r}'_3 &= \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \\
\mathbf{r}'_4 &= \begin{pmatrix} z_2 \\ z_1 \\ z_3 \end{pmatrix}, & \mathbf{r}'_5 &= \begin{pmatrix} z_2 \\ z_3 \\ z_1 \end{pmatrix}, & \mathbf{r}'_6 &= \begin{pmatrix} z_3 \\ z_2 \\ z_1 \end{pmatrix}.
\end{align*}
\]

Each of these vectors is a population which we represent as \( r'_p z = z^{(p)} \), where

\[
z^{(p)} = \begin{pmatrix} z_1^{(p)} \\ z_2^{(p)} \\ z_3^{(p)} \end{pmatrix}
\]

with elements \( z_s^{(p)} \) for \( s = 1, \ldots, N \). Notice that if we start with a finite population, we essentially condition on a value of \( p \), and hence drop \( p \) from the notation, representing the response for element \( s \) in the population by \( y_s \). While certain parameters, such as the mean of the population, are identical for all populations, and the set of subjects in each population is identical, the populations are different. The difference is due to the fact that the elements \( z_s^{(p)} \) of a population vector differ between populations. The subjects corresponding to positions in a vector are different for different populations.

The next random variable identifies subsets of a population. Since \( n = 2 \) and \( N = 3 \), there are three possible subsets of a population corresponding to \( \delta'_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \), \( \delta'_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \), \( \delta'_3 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \), and \( \delta'_4 = \begin{pmatrix} z_1^{(p)} \\ z_2^{(p)} \\ z_3^{(p)} \end{pmatrix} \). We index these subsets by \( h = 1, \ldots, H = 3 \), and represent the vector corresponding to subset \( h \) of population \( p \) by \( \delta'_h r'_p z = z^{(hp)} \), where \( z^{(hp)} = \begin{pmatrix} z_1^{(hp)} \\ z_2^{(hp)} \end{pmatrix} \), with elements \( z_j^{(hp)} \) for \( j = 1, \ldots, n \). Since the data corresponding to the sample is observed, we typically condition on \( h \) and \( p \), and represent the elements by \( y_j, j = 1, \ldots, n \).

The final term is a permutation of the sample. Only one permutation is considered, which we represent by \( v'_m \). For example, assume \( v'_m = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \). We represent the vector corresponding to permutation \( m \) of subset \( h \) of population \( p \) by \( v'_m \delta'_h r'_p z = z^{(mph)} \), where \( z^{(mph)} = \begin{pmatrix} z_1^{(mph)} \\ z_2^{(mph)} \end{pmatrix} \), with elements \( z_i^{(mph)} \) for \( i = 1, \ldots, n \). Using this notation, we represent
\[ Y_m = \sum_{h=1}^{H} \sum_{p=1}^{P} I_p I_h z_{(mhp)} = \begin{pmatrix} \sum_{h=1}^{H} \sum_{p=1}^{P} I_p I_h z_{1}^{(mhp)} \\ \vdots \\ \sum_{h=1}^{H} \sum_{p=1}^{P} I_p I_h z_{n}^{(mhp)} \end{pmatrix} = \begin{pmatrix} Y_1^{(m)} \\ \vdots \\ Y_n^{(m)} \end{pmatrix}, \]

with elements of the \( Y_m \) represented by \( Y_i^{(m)} \).

In summary, there are 18 possible realizations of \( Y_m \) when \( n = 2 \) from a population of size \( N = 3 \) where \( N = N \). We express each realization in terms of the original notation for the subjects in the superpopulation. The first three realizations, corresponding to population \( p = 1 \) are given by

\[ z^{(m11)} = v'_m \delta_i r'_i z = \begin{pmatrix} z_1^{(m11)} \\ z_2^{(m11)} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} z_1^{(11)} \\ z_2^{(11)} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & z_1 \\ z_2 \\ z_3 \end{pmatrix}, \]

\[ z^{(m21)} = v'_m \delta_i r'_i z = \begin{pmatrix} z_1^{(m21)} \\ z_2^{(m21)} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} z_1^{(21)} \\ z_2^{(21)} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & z_1 \\ z_2 \\ z_3 \end{pmatrix}, \]

\[ z^{(m31)} = v'_m \delta_i r'_i z = \begin{pmatrix} z_1^{(m31)} \\ z_2^{(m31)} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} z_1^{(31)} \\ z_2^{(31)} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & z_1 \\ z_2 \\ z_3 \end{pmatrix}. \]

Notice that although each of these realizations are equally likely, conditional on the population, the elements in the sample are not exchangeable.

The next three possible realizations for population \( p = 2 \) are given by
\[ \mathbf{z}^{(m_{12})} = \mathbf{v}_r^{m_{12}} \mathbf{r}_m \mathbf{z} = \begin{pmatrix} z_{1}^{(m_{12})} \\ z_{2}^{(m_{12})} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} z_{1}^{(12)} \\ z_{2}^{(12)} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} z_{1}^{(2)} \\ z_{2}^{(2)} \\ z_{3}^{(2)} \end{pmatrix}, \]

\[
= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} z_{1} \\ z_{2} \\ z_{3} \end{pmatrix} = \begin{pmatrix} z_{3} \\ z_{1} \end{pmatrix}
\]

\[ \mathbf{z}^{(m_{22})} = \mathbf{v}_r^{m_{22}} \mathbf{r}_m \mathbf{z} = \begin{pmatrix} z_{1}^{(m_{22})} \\ z_{2}^{(m_{22})} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} z_{1}^{(22)} \\ z_{2}^{(22)} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} z_{1}^{(2)} \\ z_{2}^{(2)} \\ z_{3}^{(2)} \end{pmatrix}, \]

\[
= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} z_{1} \\ z_{2} \\ z_{3} \end{pmatrix} = \begin{pmatrix} z_{2} \\ z_{1} \end{pmatrix}
\]

\[ \mathbf{z}^{(m_{32})} = \mathbf{v}_r^{m_{32}} \mathbf{r}_m \mathbf{z} = \begin{pmatrix} z_{1}^{(m_{32})} \\ z_{2}^{(m_{32})} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} z_{1}^{(32)} \\ z_{2}^{(32)} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} z_{1}^{(2)} \\ z_{2}^{(2)} \\ z_{3}^{(2)} \end{pmatrix}, \]

\[
= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} z_{1} \\ z_{2} \\ z_{3} \end{pmatrix} = \begin{pmatrix} z_{3} \\ z_{2} \end{pmatrix}
\]

The three possible realizations for population \( p = 3 \) are given by

\[ \mathbf{z}^{(m_{13})} = \mathbf{v}_r^{m_{13}} \mathbf{r}_m \mathbf{z} = \begin{pmatrix} z_{1}^{(m_{13})} \\ z_{2}^{(m_{13})} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} z_{1}^{(13)} \\ z_{2}^{(13)} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} z_{1}^{(3)} \\ z_{2}^{(3)} \\ z_{3}^{(3)} \end{pmatrix}, \]

\[
= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} z_{1} \\ z_{2} \\ z_{3} \end{pmatrix} = \begin{pmatrix} z_{1} \\ z_{2} \end{pmatrix}
\]
\[ z^{(m23)} = v'_m \delta_z^* r'_4 z = \begin{pmatrix} \begin{pmatrix} z_1^{(m23)} \\ z_2^{(m23)} \end{pmatrix} \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} \begin{pmatrix} z_1^{(23)} \\ z_2^{(23)} \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \begin{pmatrix} z_1^{(3)} \\ z_2^{(3)} \\ z_3^{(3)} \end{pmatrix}, \]

\[ = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} = \begin{pmatrix} z_3 \\ z_2 \end{pmatrix} \]

\[ z^{(m33)} = v'_m \delta_z^* r'_4 z = \begin{pmatrix} \begin{pmatrix} z_1^{(m33)} \\ z_2^{(m33)} \end{pmatrix} \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} \begin{pmatrix} z_1^{(33)} \\ z_2^{(33)} \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} z_1^{(3)} \\ z_2^{(3)} \\ z_3^{(3)} \end{pmatrix}, \]

\[ = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} = \begin{pmatrix} z_3 \\ z_1 \end{pmatrix} \]

The three possible realizations for population \( p = 4 \) are given by

\[ z^{(m14)} = v'_m \delta_z^* r'_4 z = \begin{pmatrix} \begin{pmatrix} z_1^{(m14)} \\ z_2^{(m14)} \end{pmatrix} \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} \begin{pmatrix} z_1^{(14)} \\ z_2^{(14)} \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} \begin{pmatrix} z_1^{(4)} \\ z_2^{(4)} \\ z_3^{(4)} \end{pmatrix}, \]

\[ = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} = \begin{pmatrix} z_3 \\ z_2 \end{pmatrix} \]

\[ z^{(m24)} = v'_m \delta_z^* r'_4 z = \begin{pmatrix} \begin{pmatrix} z_1^{(m24)} \\ z_2^{(m24)} \end{pmatrix} \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} \begin{pmatrix} z_1^{(24)} \\ z_2^{(24)} \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} \begin{pmatrix} z_1^{(4)} \\ z_2^{(4)} \\ z_3^{(4)} \end{pmatrix}, \]

\[ = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \]
\[ z^{(m34)} = v_m' \delta_4 r_4' z = \begin{pmatrix} z_1^{(m34)} \\ z_2^{(m34)} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} z_1^{(34)} \\ z_2^{(34)} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{pmatrix} = \begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{pmatrix} . \]

The three possible realizations for population \( p = 5 \) are given by

\[ z^{(m15)} = v_m' \delta_5 r_5' z = \begin{pmatrix} z_1^{(m15)} \\ z_2^{(m15)} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} z_1^{(15)} \\ z_2^{(15)} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{pmatrix} = \begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{pmatrix} . \]

\[ z^{(m25)} = v_m' \delta_5 r_5' z = \begin{pmatrix} z_1^{(m25)} \\ z_2^{(m25)} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} z_1^{(25)} \\ z_2^{(25)} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{pmatrix} = \begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{pmatrix} . \]

\[ z^{(m35)} = v_m' \delta_5 r_5' z = \begin{pmatrix} z_1^{(m35)} \\ z_2^{(m35)} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} z_1^{(35)} \\ z_2^{(35)} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{pmatrix} = \begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{pmatrix} . \]

The three possible realizations for population \( p = 6 \) are given by
We summarize the realizations of permutation $m$ of subset $h$ of population $p$ in this setting as follows:

$$
\mathbf{z}^{(m16)} = \mathbf{v}_m^\top \mathbf{\delta}_1 \mathbf{r}_0 \mathbf{z} = \begin{pmatrix} z_1^{(m16)} \\ z_2^{(m16)} \\ z_3^{(m16)} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} z_1^{(16)} \\ z_1^{(16)} \\ z_1^{(16)} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} z_1^{(6)} \\ z_2^{(6)} \\ z_3^{(6)} \end{pmatrix},
$$

$$
= \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} = \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix},
$$

$$
\mathbf{z}^{(m26)} = \mathbf{v}_m^\top \mathbf{\delta}_2 \mathbf{r}_0 \mathbf{z} = \begin{pmatrix} z_1^{(m26)} \\ z_2^{(m26)} \\ z_3^{(m26)} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} z_1^{(26)} \\ z_1^{(26)} \\ z_1^{(26)} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} z_1^{(6)} \\ z_2^{(6)} \\ z_3^{(6)} \end{pmatrix},
$$

$$
= \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} = \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix},
$$

$$
\mathbf{z}^{(m36)} = \mathbf{v}_m^\top \mathbf{\delta}_3 \mathbf{r}_0 \mathbf{z} = \begin{pmatrix} z_1^{(m36)} \\ z_2^{(m36)} \\ z_3^{(m36)} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} z_1^{(36)} \\ z_1^{(36)} \\ z_1^{(36)} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} z_1^{(6)} \\ z_2^{(6)} \\ z_3^{(6)} \end{pmatrix},
$$

$$
= \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} = \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix}.
$$

Each row of possible outcomes corresponds to a realized population, $p$. Given a population, $p$, the expected value of the average response is equal to the population mean. Since all
populations include the same subjects, these averages are the same for all populations, and are
equal to the average response of subjects in the superpopulation.

Notice that for a given population (represented as a sequence corresponding to
permutation $p$ of the set $f$ of $N$ values from a superpopulation), $p$, random variables
representing elements in the sample are not exchangeable. Recall that we can express the
elements of $Y_m$ by $Y_i^{(m)}$, where $Y_i^{(m)} = \sum_{h=1}^{H} \sum_{p=1}^{p} I_h z_i^{(m,p)}$. Conditional on a population, the
expected value of these elements are given as $E(Y_m | p)$

$$E(Y_m | p) = \frac{1}{3}(z_1 + 2z_2, 2z_1 + z_3, z_1 + 2z_3, 2z_1 + z_2, z_1 + 2z_2, z_1 + 2z_3).$$

Next Steps

An interesting feature of the previous discussion is the importance of permutations of the
superpopulation in defining exchangeable random variables, assuming that a set is taken for the
finite population. This specification of random variables, which is different from the usual
sampling setting, should be possible in the context of Ericson’s inference.

Several aspects of the setting are valuable to notice based on the simple example
considered earlier. First, note that although there are $N = 3$ subjects, there are six distinct
populations that result from the superpopulation. All the populations have the same mean and
variance.

Development of a BLUP

Suppose that $E\left(\begin{array}{c} Y_1 \\ Y_2 \end{array}\right) = \left(\begin{array}{c} X_1 \\ X_2 \end{array}\right) a$ where $Y_i (n \times 1)$ represents a vector of response for a
sample, and $Y_2 ((N - n) \times 1)$ represents potential response for the remainder of a population,

$$\text{var}\left(\begin{array}{c} Y_1 \\ Y_2 \end{array}\right) = \left(\begin{array}{cc} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{array}\right),$$

and assume that the target of interest is $P = g\left(\begin{array}{c} Y_1 \\ Y_2 \end{array}\right)$ where we partition
g as $g = \left(\begin{array}{c} g_1' \\ g_2' \end{array}\right)$ so that $P = g_1' Y_1 + g_2' Y_2$. We wish to predict the target conditional on
realizing $Y_i$ (which corresponds to selecting the sample and observing the results). The basic
inference problem is prediction of $g_2' Y_2$. We require the predictor to be:

Linear in the sample: $\hat{P} = (g_1' + a') Y_1$

Unbiased: $E(\hat{P} - P) = 0$

Minimum MSE: $\text{var}(\hat{P} - P)$ is minimized.

The unbiased constraint implies that $a'X_1 - g_2'X_2 = 0$. We minimizing the variance,

$$\text{var}(\hat{P} - P) = a'\Sigma_{11}a - 2g_1'\Sigma_{12}a + g_2'\Sigma_{22}g_2$$

subject to this unbiased constraint using Lagrangian multipliers. The quantity to be minimized is
\[ f(a, \lambda) = a'\Sigma_{a}a - 2g_2'\Sigma_{22}a + 2g_2'\Sigma_{22}g_2 + 2(a'X_i - g_2'X_2)\lambda. \]

The solution is given by the best linear unbiased predictor,
\[ \hat{P} = g_1'Y_i + g_2' \left[ X_i\hat{\alpha} + \Sigma_{21i}^{-1} (Y_i - X_i\hat{\alpha}) \right], \]
where \( \hat{\alpha} = \left( X_i'\Sigma_{11i}^{-1}X_i \right)^{-1} X_i'\Sigma_{11i}^{-1}Y_i. \) The coefficients of \( a \) that minimize this quantity are given by
\[ \hat{a} = \left[ \Sigma_{11i}^{-1} - \Sigma_{11i}^{-1}X_i \left( X_i'\Sigma_{11i}^{-1}X_i \right)^{-1} X_i'\Sigma_{11i}^{-1} \right] \Sigma_{12}g_2 + \Sigma_{11i}^{-1}X_i \left( X_i'\Sigma_{11i}^{-1}X_i \right)^{-1} X_2'g_2. \]
References


