Outline for Prediction of Random Effect in a Simple Mixed Model

Ed Stanek

Introduction

We derive predictors of random effects for a mixed model. First, parameters are defined in a population. We assume response has been reported for subjects selected via a two stage cluster sample. At the first stage, a simple random sample of clusters (which we will also refer to as primary sampling units (PSU)) is selected. At the second stage, a simple random sample of subjects (which we will also refer to as second stage sampling units (SSU)) in the selected clusters will be selected. We index the selection of the clusters in the sample by \( i = 1, \ldots, n \), and the selection of the subjects in a selected cluster as \( j = 1, \ldots, m \). We assume that there are \( N \) clusters in the population, and \( M \) SSUs in each cluster. Much of this development is drawn from Robinson, G.K. (1991).

Study Population and Parameters

We assume the study population consists of a large number of clusters that are numbered \( s = 1, \ldots, N \), where each cluster contains a large number of subjects \( t = 1, \ldots, M \). As an example, the clusters may be hospitals, with the subjects representing patients in a hospital. Alternatively, the clusters may be subjects, with the repeated measures corresponding to time points where the subjects are measured. We represent the potentially observable response of a subject in a cluster as a fixed constant \( \mu_{st} \). Using this representation of response, we define parameters for the mean and variance of cluster \( s \) as

\[
\mu_s = \frac{1}{M} \sum_{t=1}^{M} \mu_{st} \quad \text{and} \quad (\frac{M-1}{M}) \sigma^2_s = \left( \frac{M-1}{M} \right) \left( \frac{M \sum_{t=1}^{M} (\mu_{st} - \mu) \right)^2}{M-1}
\]

for \( s = 1, \ldots, N \).

We describe parameters for the population of clusters next. As for the individual clusters, we define the mean and variance as

\[
\mu = \frac{1}{N} \sum_{s=1}^{N} \mu_s \quad \text{and} \quad (\frac{N-1}{N}) \sigma^2 = \left( \frac{N-1}{N} \right) \left( \frac{N \sum_{s=1}^{N} (\mu_s - \mu) \right)^2}{N-1}
\]

We use these definitions of parameters to define a model for a response in the population. The potentially observable response for a subject is given by

\[
\mu_{st} = \mu + (\mu_s - \mu) + (\mu_{st} - \mu_s)
\]

\[
= \mu + \beta_s + \epsilon_{st}
\]

where \( \beta_s = (\mu_s - \mu) \) represent the deviation of a cluster mean from the overall mean, and \( \epsilon_{st} = (\mu_{st} - \mu_s) \) represents the deviation of a subject's response from the cluster mean.
The Study

The study is conducted by randomly selecting \( i = 1, \ldots, n \) clusters (without replacement), and within each selected cluster, randomly selecting \( j = 1, \ldots, m \) subjects (without replacement) and then observing response. The index \( i \) represents the position of a cluster in the sample, while the index \( j \) represents the position of a subject in the selected cluster. Since the subject assigned to position \( j \) in the cluster assigned to position \( i \) is random, we represent response as the random variable

\[ Y_{ij} = \text{response for the subject in position } j \text{ in the cluster in position } i. \]

The response can also be represented via the model:

\[ Y_{ij} = \mu + B_i + E_{ij} \]

where \( B_i \) is a random variable that represents the difference between the expected response for the cluster in position \( i \) from the population mean, and \( E_{ij} \) is a random variable that represents the difference between response for the subject in position \( j \) from the expected response of the cluster in position \( i \).

Simplifying Assumption

We simplify the problem using some additional assumptions. The assumptions are:

1. \( N \) is very large, so that \( \sigma^2 \) represents the variance of the cluster means. Also, terms in the variance that are proportional to \( \frac{1}{N} \) can be assumed to be zero.
2. \( M \) is very large, so that \( \sigma_s^2 \) represents the variance of the SSUs in a cluster. Also, terms in the variance that are proportional to \( \frac{1}{M} \) can be assumed to be zero.

The Model for a Sample of \( m \) subjects selected from the cluster Selected in Position \( i \)

We assume that a simple random sample of \( n \) clusters is selected, and that within each selected cluster, we select a simple random sample of \( m \) subjects. For the cluster in position \( i \), let \( \mathbf{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{im})' \) represent the response vector. We represent the model as

\[ \mathbf{Y}_i = \mathbf{X}_i \mathbf{\alpha} + \mathbf{Z}_i \mathbf{B} + \mathbf{E}_i \]
where \( \mathbf{B} = \left( \left( B_i \right) \right) = \left( B_1, B_2, \cdots, B_n \right) \). For the simple model \( Y_{ij} = \mu + B_i + E_{ij} \), \( \mathbf{X}_i = \mathbf{1}_n \), where \( \mathbf{1}_n \) represents an \( n \times 1 \) column vector, and \( \mathbf{Z}_i \mathbf{B} = \mathbf{1}_n B_i \). We refer to the cluster selected in position \( i \) as PSU \( i \), and the subject selected in position \( j \) as SSU \( j \).

Let us represent expectation with respect to selection of PSUs with the subscript \( \xi_1 \), and expectation with respect to selection of SSUs with the subscript \( \xi_2 \). Since \( E_{\xi_1} \left( B_i \right) = 0 \), and \( E_{\xi_2|\xi_1} \left( E_{ij} \right) = 0 \), then \( E_{\xi_2|\xi_1} \left( Y_{ij} \right) = \mathbf{X}_i \alpha \). We make use of expressions the conditional expansion of the variance, \( \text{var}_{\xi_2|\xi_1} \left( Y_i \right) = E_{\xi_1} \left[ \text{var}_{\xi_2|\xi_1} \left( Y_i \right) \right] + \text{var}_{\xi_1} \left[ E_{\xi_2|\xi_1} \left( Y_i \right) \right] \) to evaluate the variance. Then

\[
\text{var}_{\xi_2|\xi_1} \left( Y_i \right) = E_{\xi_1} \left[ \sigma_e^2 \mathbf{I}_m \right] + \text{var}_{\xi_1} \left[ \mathbf{X}_i \alpha + \mathbf{Z}_i \mathbf{B} \right]
= \sigma_e^2 \mathbf{I}_m + \mathbf{1}_n \text{var}_{\xi_1} \left[ B_i \right] \mathbf{1}_m'
= \sigma_e^2 \mathbf{I}_m + \sigma_B^2 \mathbf{1}_m \mathbf{1}_m'
= \mathbf{V}_i
\]

**The Model for the Sample**

Combining response for all sampled clusters, the sample response vector is given by

\[
\mathbf{Y} = \begin{pmatrix}
\mathbf{Y}_1 \\
\mathbf{Y}_2 \\
\vdots \\
\mathbf{Y}_n
\end{pmatrix}, \text{ where } \mathbf{Y} = \mathbf{X} \alpha + \mathbf{Z} \mathbf{B} + \mathbf{E}, \text{ where } E_{\xi_2|\xi_1} \left( \mathbf{Y}_{nm} \right) = \mathbf{X} \alpha, \text{ and}
\]

\[
\text{var}_{\xi_2|\xi_1} \left( \mathbf{Y}_{nm} \right) = \mathbf{V} = \begin{pmatrix}
\mathbf{V}_1 & 0 & \cdots & 0 \\
0 & \mathbf{V}_2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \mathbf{V}_n
\end{pmatrix} = \mathbf{R} + \mathbf{Z} \mathbf{G} \mathbf{Z}' \text{ where var}_{\xi_1} \left[ \mathbf{B} \right] = \mathbf{G} = \sigma_B^2 \mathbf{I}_n \text{ and } \mathbf{R} = \sigma_e^2 \mathbf{I}_{nm}.
\]

Example: Suppose we have an overall mean model, with random effects corresponding to the deviation in the PSU mean from the overall mean. We can explicitly write expressions for the matrices corresponding to \( \mathbf{X}_i, \alpha, \mathbf{Z}_i, \mathbf{B}, \mathbf{X}, \) and \( \mathbf{Z} \). Thus, \( \mathbf{X}_i = \mathbf{1}_n \), \( \mathbf{Z}_i = \mathbf{e}_i' \otimes \mathbf{1}_m \) where \( \mathbf{e}_i = \left( \left( e_i \right) \right) = \left( 0 \ 1 \ \cdots \ 0 \right)' \) with one in position \( i \) and zero otherwise, \( \alpha = \mu \),

\[
\mathbf{B} = \left( \left( B_i \right) \right) = \left( B_1 \ B_2 \ \cdots \ B_n \right)', \ \mathbf{X} = \mathbf{1}_{nm} \text{ and } \mathbf{Z} = \begin{pmatrix}
\mathbf{1}_m & 0 & \cdots & 0_m \\
0_m & \mathbf{1}_m & \cdots & 0_m \\
\vdots & \vdots & \ddots & \vdots \\
0_m & 0_m & \cdots & \mathbf{1}_m
\end{pmatrix} = \mathbf{1}_n \otimes \mathbf{1}_m.
\]
We proceed to develop predictors of the random effects, $B$. The development is an expansion of the derivations of the best linear unbiased predictor (BLUP) given by Robinson (1991) in section 4.

**Henderson’s Justification of BLUP**

Our interest is in prediction of the random variables $B$. For the model

$$Y_{nm} = X \alpha + (ZB + E),$$

where $E(Y|B = b) = X \alpha + Zb$, and $\text{var}_{\xi_{i_1}}(Y_{nm}) = V$, we make the following assumption concerning the joint distribution of $Y|B = b$ and $B$:

1. the random variables $\left( \begin{array}{c} Y | B = b \\ B \end{array} \right)$ are normally distributed with expected value $E \left( \begin{array}{c} Y | B = b \\ B \end{array} \right) = \left( \begin{array}{c} X\alpha + Zb \\ 0 \end{array} \right)$, and $\text{var} \left( \begin{array}{c} Y | B = b \\ B \end{array} \right) = \left( \begin{array}{cc} R & 0 \\ 0 & G \end{array} \right)$.

This formulation is closely related to Bayes theorem, given by

$$P(B | Y = y) = \frac{P(Y | B = b)P(B = b)}{\int_{b}P(Y | B = b)P(B = b)db}.$$

Let us represent the denominator by the normalizing constant $A = \int_{b}P(Y | B = b)P(B = b)db$. Then $P(B | Y = y) = \frac{1}{A}P(Y | B = b)P(B = b)$. This probability function describes the posterior distribution of $B$. In a Bayesian framework, some characteristic of this distribution, such as the mean, mode, or median, can be used to predict $B$.

Notice that in the formulation above, $\text{cov} \left( \begin{array}{c} Y | B = b \\ B \end{array} \right) = 0$.

With these assumptions, the likelihood is given by

$$L = \exp \left[ -\frac{1}{2} \left( \begin{array}{c} Y - (X\alpha + Zb) \\ B \end{array} \right)' \left( \begin{array}{cc} R & 0 \\ 0 & G \end{array} \right)^{-1} \left( \begin{array}{c} Y - (X\alpha + Zb) \\ B \end{array} \right) \right].$$

so that $\ell = \log(L)$ where

$$\ell = -\frac{1}{2} \left( \begin{array}{c} Y - (X\alpha + Zb) \\ B \end{array} \right)' \left( \begin{array}{cc} R & 0 \\ 0 & G \end{array} \right)^{-1} \left( \begin{array}{c} Y - (X\alpha + Zb) \\ B \end{array} \right)$$

$$-\frac{nm + n}{2} \log(2\pi) - \frac{1}{2} \log \left( \begin{array}{cc} R & 0 \\ 0 & G \end{array} \right).$$

or
\[
\ell = -\frac{1}{2} \left( \left[ Y - (X\alpha + Zb) \right] R^{-1} \left[ Y - (X\alpha + Zb) \right] + B'G^{-1}B \right) \\
- \frac{nm + n}{2} \log(2\pi) - \frac{1}{2} \log \left| \begin{pmatrix} R & 0 \\ 0 & G \end{pmatrix} \right| 
\]

We differentiate this expression with respect to \( \alpha \) and \( B \). First notice that
\[
\left[ Y - (X\alpha + Zb) \right] R^{-1} \left[ Y - (X\alpha + Zb) \right] = \\
= -2\alpha'X'Y + \alpha'X'X\alpha + 2\alpha'X'R^{-1}Zb - 2b'Z'R^{-1}Y + b'Z'R^{-1}Zb + Y'R^{-1}Y 
\]

Then
\[
\frac{\partial \ell}{\partial \alpha} = -\frac{1}{2} \left[ -2X'R^{-1}Y + 2X'R^{-1}X\alpha + 2X'R^{-1}Zb \right], \text{ while} \\
\frac{\partial \ell}{\partial b} = -\frac{1}{2} \left[ 2Z'R^{-1}X\alpha - 2Z'R^{-1}Y + 2Z'R^{-1}Zb + 2G^{-1}B \right].
\]

Setting these equations simultaneously to zero,
\[
\begin{pmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{pmatrix} \begin{pmatrix} \hat{\alpha} \\ \hat{b} \end{pmatrix} = \begin{pmatrix} X'R^{-1}Y \\ Z'R^{-1}Y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.
\]

These are Henderson’s mixed model equations. Notice that in this derivation, we have treated \( B \) and \( b \) as equivalent in the differentiation.

Solving for \( \hat{b} \)

From basic matrix algebra, notice that Let \( X = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \). Then
\[
X^{-1} = \begin{pmatrix} A^{-1} + A^{-1}BQ^{-1}CA^{-1} & -A^{-1}BQ^{-1} \\ -Q^{-1}CA^{-1} & Q^{-1} \end{pmatrix} \text{ where } Q = D - CA^{-1}B.
\]

The maximum likelihood is determined by taking the derivative of this expression with respect to \( \alpha \), setting the derivative to zero (which results in the ‘estimating equations’) and then solving for an expression for \( \alpha \) (which is the estimator). Note that in this process, the expression for the variance is assumed to be a fixed constant. The resulting estimator is the generalized least squares estimator given by
\[
\hat{\alpha} = (X'V^{-1}X)^{-1}X'V^{-1}Y.
\]
 estimator must minimize the sum of squared errors given by
\[
(Y - X\hat{\alpha})'V^{-1}(Y - X\hat{\alpha}).
\]
The minimum is determined by taking the derivative of this expression with respect to \( \alpha \), setting the derivative to zero (which results in the ‘estimating equations’) and then solving for an expression for \( \hat{\alpha} \) (which is the estimator).
We illustrate this process. First we expand the expression to be minimized:

\[ S = (Y - X\alpha)'V^{-1}(Y - X\alpha) = Y'V^{-1}Y + 2Y'V^{-1}X\alpha + \alpha'(X'V^{-1}X)\alpha. \]

This expression is a function of \( \alpha \). Next, we evaluate the derivative with respect to \( \alpha \), resulting in

\[ \frac{\partial S}{\partial \alpha} = 2(Y'V^{-1}X)' + 2(X'V^{-1}X)\alpha. \]

We set this expression equal to zero, resulting in

\[ \hat{\alpha} = (X'V^{-1}X)^{-1}X'V^{-1}Y. \]

This is the resulting estimator is the generalized least squares estimator (or weighted least squares estimator; see Searle et al. (1992), Judge et al (1985)). The variance of the estimator can be directly evaluated in terms of

\[ \text{var}_s(Y_{(n_m+1)}) = V. \]

Confidence intervals for the estimators can be estimated using non-parametric procedures (such as using bootstrap methods).

**Maximum Likelihood Estimate of \( \alpha \)**

Maximum likelihood estimators are evaluated by postulating a distribution for the random variables. With the postulated distribution, given the parameter values, we can write an equation for the probability of a particular realization of sample values. Looking at the problem another way, considering the parameters as unknown and conditioning on the realized sample vector, the probability function is defined to be the likelihood. We use as ‘estimates’ the values of the parameters that will maximize the likelihood function. These are the maximum likelihood estimates.

**Assumptions**

We make two assumptions to develop maximum likelihood estimators. The distributional assumptions are as follows:

1. Given a particular selected cluster, \( Y_i | i = s \overset{\text{indep.}}{\sim} N(X_i\alpha + Z_i\beta, \sigma^2 I_m) \)
2. \( B \) is normally distributed, \( B \overset{\text{indep.}}{\sim} N(0, G) \)

These assumptions deserve some further explanation. For simplicity, let us assume that the design matrix \( X \) corresponds to a single column of ones, such that \( \alpha \) represent the overall population mean, and that \( Z \) corresponds to a matrix of indicators for the different cluster means. The first assumption states that if we fix a particular cluster, then the response corresponding to SSUs in the cluster are normally distributed. If these responses are independent, then \( R_i = \sigma^2 I_m \). The vector \( \beta' = (\beta_{s_1}, \beta_{s_2}, \ldots, \beta_{s_i}) \) is a vector of deviations of realized cluster means from the population mean, where \( \beta_{s_i} \) represents the difference between the cluster that was realized in position \( i \) in the sample and the population mean. Note that the clusters in the sample are considered to have arisen from a simple random sample of the clusters in the population. Also notice that variance of response in the cluster depends on the cluster selection.
in the position. For example, we use $R_s = \sigma_s^2 I_m$ when cluster $s$ was selected in position $i$. These variances are conditional on the selected cluster.

The second assumption is that the distribution of cluster means in the population are normally distributed with variance $G$. When clusters are selected using simple random sampling from a very large population of clusters, $G = \sigma^2 I_n$.

**Estimation**

With these assumptions, the likelihood is given by

$$L = \exp \left[ -\frac{1}{2} (Y - X\alpha)' V^{-1} (Y - X\alpha) \right]$$

so that $\log (L) = -\frac{n}{2} \log (2\pi) - \frac{1}{2} \log |V| - \frac{1}{2} (Y - X\alpha)' V^{-1} (Y - X\alpha)$. The maximum likelihood is determined by taking the derivative of this expression with respect to $\alpha$, setting the derivative to zero (which results in the ‘estimating equations’) and then solving for an expression for $\alpha$ (which is the estimator). Note that in this process, the expression for the variance is assumed to be a fixed constant. The resulting estimator is the generalized least squares estimator given by $\hat{\alpha} = (X'V^{-1}X)^{-1} X'V^{-1}Y$. The variance can be evaluated, and confidence intervals can be constructed using the normality assumption.

**Survey Sampling Prediction Based Estimate of $\alpha$**

The development of an estimate of $\alpha$ where $\alpha$ corresponds to the population mean in a simple random sample is given in c01ed11.doc. The framework for the problem is somewhat simpler, but is related to the development of Goldberger (1962). Recall that the basic model we have been discussing can be stated as $Y = X\alpha + (ZB + E)$. We represent this model as

$$Y_i = X_i\alpha + E_i.$$ We limit ourselves here to a setting where the design matrix

$$X_i = 1, \quad X_i = 1_{nm}$$ corresponds to a single column of ones, such that $\alpha$ represent the overall population mean. The vector $Y_i$ corresponds to the random variables that are potentially realized in a sample. Let us represent a random permutation of the random variables in the population as

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}, \quad \text{where} \quad Y_2 \quad \text{corresponds to the portion of the population not included in the sample}.$$ The parameter in the model is the overall mean. This can be expressed as a linear function of $Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$ where $\mu = \begin{pmatrix} g_1' \\ g_2' \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$ and $\begin{pmatrix} g_1' \\ g_2' \end{pmatrix} = \frac{1}{NM} \begin{pmatrix} I_{nm} \\ 1_{NM-nm} \end{pmatrix}$. Since the random variables in $Y_2$ represent the remaining values in a random permutation of the
population, we can express a model simultaneously for the random variables such that
\[
\begin{pmatrix}
    Y_1 \\
    Y_2
\end{pmatrix} = \begin{pmatrix}
    X_1 \\
    X_2
\end{pmatrix} \alpha + \begin{pmatrix}
    E_1 \\
    E_2
\end{pmatrix},
\]
where
\[
\text{var}_{ni\hat{\xi}_2} \left( \begin{pmatrix}
    E_1 \\
    E_2
\end{pmatrix} \right) = \begin{pmatrix}
    V_{11} & V_{12} \\
    V_{21} & V_{22}
\end{pmatrix},
\]
where
\[
V = V_{11} = \sigma^2 \left( I_n - \frac{J_m}{N} \right) \otimes \mathbf{J} + \sigma_e^2 \mathbf{I}_n \otimes \left( I_m - \frac{J_m}{M} \right)
\]
and
\[
V_{12}' = \begin{pmatrix}
    \sigma^2 \left( I_n - \frac{J_n}{N} \right) \otimes \left( \mathbf{J} - \frac{J_m}{(M-m)M} \mathbf{M} \right) - \sigma_e^2 I_n \otimes \left( \frac{J_m}{M} \right) \\
    -\sigma^2 \frac{(N-n)J_m}{N} \otimes \mathbf{J}
\end{pmatrix}.
\]
Note that if both \( N \) and \( M \) are large enough that terms divided by them are negligible, then we can approximate
\[
V_{11} = I_n \otimes \left( \sigma_e^2 I_m + \sigma^2 \mathbf{J}_m \right).
\]
After selecting the sample, we will know the value of \( Y_1 \). An estimate of \( \alpha \) then requires predicting the function \( L_2' Y_2 \). For this reason, this problem is called the prediction approach to survey sampling. We require the predictor to be:

1. Linear in the sample data, \( \hat{P} = \left( g_1' + \theta \right) Y_1 \)

2. Unbiased, such that expected bias is zero, \( E_{ni\hat{\xi}_2} \left( \hat{P} - \mu \right) = 0 \)

3. Have minimum variance, \( \text{var}_{ni\hat{\xi}_2} \left( \hat{P} - \mu \right) \) is a minimum.

Note that \( g_1' = \frac{I_{nm}}{NM} \) and \( g_2' = \frac{I_{NM-nm}}{NM} \). The expression for the expected bias simplifies to \( a'X_1 - g_2'X_2 = 0 \). The expression for the variance can be expanded such that \( \text{var}_{ni\hat{\xi}_2} \left( \hat{P} - \mu \right) = a'V_1 a - 2a'V_{12}g_2 + g_2'V_{22}g_2 \). As a result, the Lagrange function to be minimized is given by \( \Phi = a'V_1 a - 2a'V_{12}g_2 + g_2'V_{22}g_2 + 2 \left( a'X_1 - g_2'X_2 \right) \lambda \). Differentiating with respect to \( a \) and \( \lambda \), and setting the derivative to zero results in the estimating equations given by
\[
\begin{pmatrix}
    V_{11} & X_1 \\
    X_1' & 0
\end{pmatrix}\begin{pmatrix}
    \hat{a} \\
    \lambda
\end{pmatrix} = \begin{pmatrix}
    V_{12}g_2 \\
    X_1'g_2
\end{pmatrix}.
\]
Since
\[
\begin{pmatrix}
    V_{11} & X_1 \\
    X_1' & 0
\end{pmatrix}^{-1} = \begin{pmatrix}
    V_{11}^{-1} - V_{11}^{-1}X_1 \left( X_1'V_{11}^{-1}X_1 \right)^{-1}X_1'V_{11}^{-1} & V_{11}^{-1}X_1 \left( X_1'V_{11}^{-1}X_1 \right)^{-1} \\
    \left( X_1'V_{11}^{-1}X_1 \right)^{-1}X_1'V_{11}^{-1} & -\left( X_1'V_{11}^{-1}X_1 \right)^{-1}
\end{pmatrix},
\]
\[
\hat{a} = \left( V_{1i}^{-1} - V_{1i}^{-1} X_1 \left( X_1^T V_{1i}^{-1} X_1 \right)^{-1} X_1^T V_{1i}^{-1} \right) V_{12} g_2 + V_{1i}^{-1} X_1 \left( X_1^T V_{1i}^{-1} X_1 \right)^{-1} X_2^T g_2. 
\]

Using expression for the partitioned inverse, and re-arranging terms \( \hat{P} = g_1' Y_1 + g_2' \left[ X_1^T \hat{\alpha} + V_{12}^{-1} V_{11}^{-1} (Y_1 - X_1 \hat{\alpha}) \right] \)

\( \hat{\alpha} = \left( X_1^T V_{11}^{-1} X_1 \right)^{-1} X_1^T V_{11}^{-1} Y_1. \)

**Simplification**

We simplify the solution when we can represent \( V_{11} = I_n \otimes \left( \sigma_e^2 I_m + \sigma_e^2 J_m \right) \). Note from matrix results, \( (R + s u')^{-1} = R^{-1} - \left( 1 + u'R^{-1} s \right)^{-1} R^{-1} s u' R^{-1} \). Let \( R = I_n, s = \frac{\sigma_e^2}{\sigma_e}, \) and \( u = I_m \). Then

\[
\left( \sigma_e^2 I_m + \sigma_e^2 J_m \right)^{-1} = \left[ \sigma_e^2 \left( I_m + \frac{\sigma_e^2}{\sigma_e} I_m \right)^{-1} \right]^{-1} = I_m - \frac{r}{1 + mr} J_m \text{ where } r = \frac{\sigma_e^2}{\sigma_e}. 
\]

Then, since \( X_1 = I_{nm}, \hat{\alpha} = \left( X_1^T V_{11}^{-1} X_1 \right)^{-1} X_1^T V_{11}^{-1} Y_1 = \left( \frac{1 + mr}{nm} \right) \left( \frac{1}{1 + mr} \right) I_{nm} Y_1 = \overline{Y}. \)

Using these results,

\[
\hat{P} = g_1' Y_1 + g_2' \left[ X_1^T \hat{\alpha} + V_{12}^{-1} V_{11}^{-1} (Y_1 - X_1 \hat{\alpha}) \right] 
= \frac{nm}{NM} \overline{Y} + \frac{NM - nm}{NM} \overline{Y} + g_2' V_{12}^{-1} V_{11}^{-1} (Y_1 - X_1 \overline{Y}) 
= \overline{Y} + g_2' V_{12}^{-1} V_{11}^{-1} (Y_1 - X_1 \overline{Y}) 
\]

Now \( V_{12}^{-1} V_{11}^{-1} = \frac{1}{1 + mr} \left[ \sigma^2 \left( I_n - \frac{J_n}{N} \right) - \frac{\sigma^2}{M} \right] \otimes I_{M-m} \) and hence

\[
g_2' V_{12}^{-1} V_{11}^{-1} = \frac{-1}{1 + mr} \left[ \left( \frac{N-n}{N} \right) m \sigma^2 + \left( \frac{M-m}{M} \right) \frac{\sigma^2}{M} \right] \otimes I_{M-m}. 
\]

However, \( (Y_1 - X_1 \overline{Y}) = \left( I_{nm} - \frac{J_{nm}}{nm} \right) Y_1 \).

Since \( I_{nm} \left( I_{nm} - \frac{J_{nm}}{nm} \right) = 0, g_2' V_{12}^{-1} V_{11}^{-1} (Y_1 - X_1 \overline{Y}) = 0 \) and \( \hat{P} = \overline{Y}. \)

**A Bayesian Framework for Predicting \( \alpha \)**
Consider the model given by where $Y_{nm} = X\alpha + (ZB + E)$, where $E_{\text{iid}}\left(Y_{nm}\right) = X\alpha$, and \[ \text{var}_{\text{iid}}\left(Y_{nm}\right) = V = \begin{pmatrix} V_1 & 0 & \cdots & 0 \\ 0 & V_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & V_n \end{pmatrix} = R + ZGZ' \] where $R = \sigma^2 e_{nm}$. Let the prior density of $\alpha$ be given by $p(\alpha)$ where $Y_{nm}$ represents some data. We are interested in evaluating $p(\alpha \mid Y_{nm})$. To do so, we make use of Bayes Theorem which states that

$$ p(\alpha \mid Y_{nm}) = \frac{p(\alpha) p(Y_{nm} \mid \alpha)}{\int p(\alpha) p(Y_{nm} \mid \alpha) d\alpha}. $$

This is the posterior distribution of $\alpha$, given the data. It is this distribution that is the target for Bayesian inference.

For example, suppose that $\alpha \sim N(0, \Gamma)$, and $Y_{nm} \mid \alpha \sim N(X\alpha, V)$. With these assumptions, $p(\alpha) p(Y_{nm} \mid \alpha) = \left(\frac{1}{2\pi} \frac{1}{\Gamma} \right)^{n/2} \left| V \right|^{n/2} \exp\left[ -\frac{1}{2} \alpha' \Gamma^{-1} \alpha + \frac{1}{2} \left( Y - X\alpha \right)' V^{-1} \left( Y - X\alpha \right) \right]$, or taking the log,

$$ \log[p(\alpha) p(Y_{nm} \mid \alpha)] = -\log\left(\frac{1}{2\pi} \frac{1}{\Gamma} \right)^{n/2} \left| V \right|^{n/2} - \frac{1}{2} \left[ \alpha' \Gamma^{-1} \alpha + \left( Y - X\alpha \right)' V^{-1} \left( Y - X\alpha \right) \right] $$

$$ = -\log\left(\frac{1}{2\pi} \frac{1}{\Gamma} \right)^{n/2} \left| V \right|^{n/2} - \frac{1}{2} \left[ \begin{pmatrix} \alpha' \\ (Y - X\alpha)' \end{pmatrix} \begin{pmatrix} \Gamma & 0 \\ 0 & V \end{pmatrix}^{-1} \begin{pmatrix} \alpha \\ (Y - X\alpha) \end{pmatrix} \right]. $$

We determine an expression for $\alpha$ that will maximize this density. To maximize this expression, we need to differentiate with respect to $\alpha$ and set the resulting derivative to zero. This results in the estimating equations given by $\left( \Gamma^{-1} + X'V^{-1}X \right) \hat{\alpha} = X'V^{-1}Y$, such that the Bayesian estimator is given by $\hat{\alpha} = \left( \Gamma^{-1} + X'V^{-1}X \right)^{-1} X'V^{-1}Y$. 


References:


