Prediction of random effects

Linear mixed models

In general, a linear mixed model may be represented as

$$Y = X\beta + Zu + \varepsilon,$$

where

- Y is an $n \times 1$ vector of response;
- X is an $n \times p$ design matrix;
- β is a $p \times 1$ vector of “fixed” unknown parameter values;
- Z is an $n \times q$ model matrix of known constants;
- u is a $q \times 1$ random vector;
- ε is an $n \times 1$ random error.

Linear mixed models

We assume that

$$\begin{pmatrix} u \\ \varepsilon \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} G & 0 \\ 0 & R \end{pmatrix}\right)$$

Our goal is to predict the random effect u using the observed data.

Best linear unbiased prediction (BLUP)

- Assume that we would like to find a prediction $h(Y)$ for u, which minimizing the mean square error such that

$$\hat{h} = \arg\min_{h} E\{(u - h(Y))^2\}.$$

- It can be shown that the best prediction $h(Y)$ is $\hat{h} = E(u|Y)$.
- If we have normality assumption, $E(u|Y)$ is a linear function of Y. Hence $E(u|Y)$ is the BLUP of u.

BLUP

Based on our model assumption, it is easy to see that the joint distribution of u and Y is

$$\begin{pmatrix} u \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ X\beta \end{pmatrix}, \begin{pmatrix} G & GZ^T \\ ZG & R + ZGZ^T \end{pmatrix} \right).$$

Using a result in multivariate analysis, we know

$$E(u|Y) = E(u) + GZ^T\Sigma^{-1}\{Y - E(Y)\} = GZ^T\Sigma^{-1}(Y - X\beta),$$

where $\Sigma = R + ZGZ^T$.

BLUP

Since β is unknown, the BLUP of u is

$$\hat{u} = GZ^T\Sigma^{-1}(Y - X\hat{\beta}),$$

where $\hat{\beta} = (X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}Y$ is the BLUE of β.

Example

- Suppose that intelligence quotients (IQs) for a population of students are normally distributed with a mean μ and variance σ_u^2.
- Suppose that an IQ test was given to an IID sample of students in this population.
- Suppose that, given the IQ of a student, the test score of that student is normally distributed with a mean equal to the student's IQ and variance σ^2, and is independent of the test score of any other student.

Example continued

Assume that $\sigma_u^2/\sigma^2 = 9$. If the sample mean of the students' test scores was 100, what is the best prediction of the IQ of a student who scored 130 on the test?

Example continued

- Suppose that $u_1, \cdots, u_n$ are IID $N(0, \sigma_u^2)$ independent of $\varepsilon_1, \cdots, \varepsilon_n \overset{iid}{\sim} N(0, \sigma^2)$.
- Let $\mu + u_i$ be the IQ of the i-th student ($i = 1, \cdots, n$). Then the IQs of the students are $N(\mu, \sigma_u^2)$.
- Let $Y_i = \mu + u_i + \varepsilon_i$ be the test score of the i-th student ($i = 1, \cdots, n$). Then $Y_i | \mu + u_i \sim N(\mu + u_i, \sigma^2)$.

Example continued

In term of the matrix forms of the linear mixed effects models, we have

$$Y = X\beta + Zu + \varepsilon.$$

Here

$$X = \mathbf{1}_n, \beta = \mu, Z = I_n, G = \sigma_u^2 I_n, R = \sigma^2 I_n.$$

Then $\Sigma = (\sigma_u^2 + \sigma^2) I_n$. Thus,

$$\hat{\beta} = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} Y = \bar{Y} \text{ and } G Z^T \Sigma^{-1} = \frac{\sigma_u^2}{\sigma^2 + \sigma_u^2} I_n.$$

Example continued

The BLUP of u is

$$\hat{u} = GZ^T\Sigma^{-1}(Y - X\hat{\beta}) = \frac{\sigma_u^2}{\sigma_u^2 + \sigma^2}(Y - \mathbf{1}_n\bar{Y}).$$

The i-th element of $\hat{u}$ is $\hat{u}_i = \{\sigma_u^2/(\sigma_u^2 + \sigma^2)\}(Y_i - \bar{Y})$. Therefore the BULP of $\mu + u_i$ is

$$\begin{aligned}\hat{\mu} + \hat{u}_i &= \bar{Y} + (\sigma_u^2/\sigma_u^2 + \sigma^2)(Y_i - \bar{Y})\\ &= \frac{\sigma_u^2}{\sigma_u^2 + \sigma^2}Y_i + \frac{\sigma^2}{\sigma_u^2 + \sigma^2}\bar{Y}.\end{aligned}$$

Note that the BLUP is a linear combination of the individual score and the overall mean score.

Example continued

Because we assumed that $\sigma_u^2/\sigma^2 = 9$, the weights are

$$\frac{\sigma_u^2}{\sigma_u^2 + \sigma^2} = 0.9 \text{ and } \frac{\sigma^2}{\sigma_u^2 + \sigma^2} = 0.1.$$

As a result, we would predict the IQ of a student who scored 130 on the test to be

$$0.9 \times 130 + 0.1 \times 100 = 127.$$

Computation issue

- In the BLUE of β and the BLUP of u, both involve an inverse of an $n \times n$ matrix Σ.
- When the sample size is large, the inversion of Σ is challenging. This is often seen in animal breeding applications. For modern big data, the inverse of Σ is even impossible.
- We will introduce Henderson's estimating equation methods, which only involves the inverse of matrices G and R. Note that G is a $q \times q$ matrix. In practice, typically q is small and R is an $n \times n$ diagonal matrix, which is easy to invert.

Henderson's estimating equations method

- If we consider u as fixed effects, the estimating equations for β and u are

$$\begin{pmatrix} X^T R^{-1} X & X R^{-1} Z^T \\ Z^T R^{-1} X & Z R^{-1} Z^T \end{pmatrix} \begin{pmatrix} \beta \\ u \end{pmatrix} = \begin{pmatrix} X^T R^{-1} Y \\ Z^T R^{-1} Y \end{pmatrix}.$$

- The Henderson's mixed-effects equations are

$$\begin{pmatrix} X^T R^{-1} X & X R^{-1} Z^T \\ Z^T R^{-1} X & Z R^{-1} Z^T + G^{-1} \end{pmatrix} \begin{pmatrix} \beta \\ u \end{pmatrix} = \begin{pmatrix} X^T R^{-1} Y \\ Z^T R^{-1} Y \end{pmatrix}.$$

Henderson's estimating equations method

- By solving the Henderson's mixed-effects equations, we can obtain the BLUE of β and the BLUP of u simultaneously.
- The Henderson's method is computationally efficient because it only involves the inverse of small or diagonal matrices.
- Henderson (1963) showed that the mixed-effects equations do not actually depend on normality. The solutions $\hat{\beta}$ and $\hat{u}$ are BLUE and BLUP, respectively, under general conditions provided that the variances are known.