REML estimation of variance components

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

Assume that (ε, u) are jointly normal and

$$E(\varepsilon) = 0,\ \text{Var}(\varepsilon) = R,\ E(u) = 0,\ \text{Var}(u) = G.$$

As a result,

$$\text{Var}(Y) = \Sigma = ZGZ^T + R.$$

Variance component estimation

Three basic methods:

- ANOVA methods (method of moments)
- Maximum likelihood (ML) method
- Restricted ML method (REML)

Maximum likelihood method

Assume that Σ is a function of γ. Here γ is a vector that contains all the variance components.

The likelihood function for β, γ is

$$L(\beta,\gamma) = (2\pi)^{-n/2}|\Sigma(\gamma)|^{-1/2}\exp\{-\frac{1}{2}(Y-X\beta)^T\Sigma(\gamma)^{-1}(Y-X\beta)\}.$$

For a fixed γ, the MLE of β is (if X is full rank)

$$\hat{\beta}(\gamma) = (X^T\Sigma^{-1}(\gamma)X)^{-1}X^T\Sigma^{-1}(\gamma)Y.$$

Maximum likelihood method

Plugging $\hat{\beta}(\gamma)$ back to the likelihood function, we have a profile likelihood for γ. That is

$$L^*(\gamma) = (2\pi)^{-n/2}|\Sigma(\gamma)|^{-1/2} \times \exp\left[-\frac{1}{2}\{Y - X\hat{\beta}(\gamma)\}^T\Sigma(\gamma)^{-1}\{Y - X\hat{\beta}(\gamma)\}\right].$$

Then the MLE for γ is

$$\hat{\gamma}^2 = \arg\max_{\gamma} L^*(\gamma).$$

Maximum likelihood method

- The consistency and asymptotic normality of MLEs are supported by the large sample theory.
- But in small sample case, MLE for variance components tend to underestimate variance components. This is due to the failure to account for a reduction in degrees of freedom associated with fixed effects.

Example of MLE bias

- For the case $\Sigma = \sigma^2 I_n$ and $\gamma = \sigma^2$, the MLE of σ^2 is

$$\hat{\sigma}^2 = \frac{1}{n}(Y - X\hat{\beta})^T(Y - X\hat{\beta}),$$

  which has expectation $\{n - rank(X)\}\sigma^2/n < \sigma^2$.
- The MLE is often criticized for "failing to account for the loss of degrees of freedom needed to estimate β".

Restricted ML method (REML)

- The idea of REML is to construct likelihood for a set of error contrasts whose distributions are unrelated to the fixed parameters β.
- REML is an approach that produces unbiased estimators for some special cases and produces less biased estimates than the ML estimators in general.

The REML method

- Find $n - rank(X) = n - p$ linearly independent vectors $b_1, \cdots, b_{n-p}$ such that $b_i^T X = 0$ for all $i = 1, \cdots, n-p$.
- Find the maximum likelihood estimate of γ using linear combinations of response $w_1 = b_1^T Y, \cdots, w_{n-p} = b_{n-p}^T Y$ as data. In matrix notations, the linear combinations are

$$w = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_{n-p} \end{pmatrix} = \begin{pmatrix} b_1' Y \\ \vdots \\ b_{n-p}' Y \end{pmatrix} = B^T Y,$$

where $B = (b_1, \cdots, b_{n-r})$.

The REML method

- If $b'X = 0$, $b'Y$ is known as an error contrast.
- $w_1, \cdots, w_{n-p}$ comprise a set of $n - p$ error contrasts.
- Recall that $P_X = X(X^T X)^- X^T$ and we know that

$$(I - P_X)X = X - P_X X = X - X = 0,$$

the elements of

$$(I - P_X)Y = Y - P_X Y = Y - \hat{Y}$$

are error contrasts.

The REML method

- Because $\text{rank}(I - P_X) = n - p$, there exists a set of $n - p$ linearly independent rows of $I - P_X$ that can be used in finding error contrasts.
- If we do use a subset of rows of $I - P_X$ to get $b_1, \cdots, b_{n-p}$, the error contrasts will be a subset of residual vector $Y - \hat{Y}$.
- This is the reason that the procedure is also called residual maximum likelihood estimator.

Error contrasts

Let $A = I_n - P_X$ be an $n \times n$ matrix. Define B, an $n \times (n-p)$ matrix, such that

$$BB^T = A \text{ and } B^T B = I_{n-p}.$$

Then the error contrasts is $w = B^T Y$. It can be shown that $B^T Y$ and $\hat{\beta} = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} Y$ are independent.

The REML method

The REML is defined as the maximizer of the following log-likelihood

$$\ell^*(\gamma) = -\frac{1}{2}\log(|\Sigma|) - \frac{1}{2}\log(|X^T\Sigma^{-1}X|) - \frac{1}{2}(Y - X\hat{\beta})^T\Sigma^{-1}(Y - X\hat{\beta}).$$

That is

$$\hat{\gamma} = \arg\max_{\gamma} \ell^*(\gamma).$$

Example

Assume $Y_1, \cdots, Y_n$ are independent and identically distributed $N(\mu, \sigma^2)$. The MLE of μ and σ^2 is

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} Y_i = \bar{Y} \text{ and } \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2.$$

But the REML estimator of σ^2 is

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2,$$

which is unbiased to σ^2.

Some remarks

- ANOVA methods: easy to compute in balanced case, widely known, unbiased, no requirement to completely specifying distributions but it may produce negative estimates.
- ML methods enjoy good large sample properties (efficiency), computation difficult and underestimate variance components.
- REML has the same estimate as the ANOVA method in simple balanced case when ANOVA estimates are inside parameter space. The REML estimates are typically less biased than the ML methods.