ACTA ET COMMENTATIONES UNIVERSITATIS TARTUENSIS DE MATHEMATICA
Volume 8, 2004

On reparameterization of random effects in linear mixed models

Märt Möls

Abstract. The Empirical Best Linear Unbiased Predictor of random effects in linear mixed model may be non-unique. For fixed effects two approaches are used to derive unique solutions – one is based on using estimable linear combinations of parameters and the other one uses reparameterization constraints. It is shown in this article that both approaches can be applied in a similar manner to derive unique prediction results for random effects.

1. Notation

Consider the model

$$Y = X\beta + Z\gamma + \varepsilon,$$

where Y is a vector of n observable random variables, β is a vector of p unknown parameters having fixed values (fixed effects), γ is a random vector of length k (random effects) and ε is a random n-vector of errors. Matrix X is an $n \times p$ and matrix Z is an $n \times k$ matrix. Both X and Z are assumed to be known. We suppose that $E(\gamma) = 0$, $E(\varepsilon) = 0$ and

$$\operatorname{Var}\begin{bmatrix} \gamma \\ \varepsilon \end{bmatrix} = \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix}.$$

The covariance matrix of Y can be expressed as

$$V = R + ZGZ^T. \tag{1}$$

Throughout the paper it is assumed that V and R are non-singular matrices. Generalized inverse of a square matrix A is denoted by A^-. The Best Linear Unbiased Estimator of fixed effects (BLUE) $\hat{\beta}(V)$ is defined as

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

Received December 10, 2003.
2000 *Mathematics Subject Classification.* 62J05, 62J10, 62H12, 62J07.
Key words and phrases. Mixed models, BLUP, BLUE, random effects.
Supported partially by Estonian Science Foundation Grant 5203.

The Best Linear Unbiased Predictor (BLUP) of random effects is defined as

$$\hat{\gamma}(R, G) = GZ^T V^{-1}(Y - X\hat{\beta}(V)), \tag{2}$$

where V is a function of R and G as given in (1).

2. Reparameterization of random effects

The proofs in this article use results from matrix algebra, two of which are emphasized as Lemma 1 and Lemma 2.

Lemma 1. *If matrices Z and H satisfy the condition*

$$\operatorname{rank}\begin{pmatrix} Z \\ H \end{pmatrix} = \operatorname{rank}(Z) + \operatorname{rank}(H), \tag{3}$$

then there exists a symmetric positive definite matrix A such that

$$HAZ^T = 0.$$

Proof. See Harville (1997), Theorems 17.2.4 and 17.7.1. □

Lemma 2. *Let A be a positive definite matrix. Then, for any matrix Z for which the multiplication ZA is defined, the following equality holds:*

$$ZAZ^T(ZAZ^T)^- ZA = ZA.$$

Proof. See Harville (1997), Theorem 14.12.25. □

Now consider two separate mixed models which differ from each other only by having different covariance matrices for γ. For model 1 (reference model) $\operatorname{Var}(\gamma) = G$ and for model 2 (alternative model) $\operatorname{Var}(\gamma) = G_*$. The first theorem quarantees the existence of such alternative model for which BLUP predictors satisfy "reparameterisation constraints" in the form $H\hat{\gamma} = 0$ and for which the BLUP predictors for linear combinations of $Z\gamma$ ("predictable linear combinations") would be the same as for the reference model.

Theorem 1. *Let H be a matrix, for which the equality (3) holds. Then there exists a matrix G^* such that*

$$H\hat{\gamma}(R, G^*) = 0,$$
$$Z\hat{\gamma}(R, G^*) = Z\hat{\gamma}(R, G).$$

Proof. From Lemma 1 the existence of a positive definite matrix A, such that

$$HAZ^T = 0,$$

follows. Using matrix A, define an idempotent matrix P projecting into column space of Z^T:

$$P^T = AZ^T(ZAZ^T)^- Z.$$

Obviously

$$HP^T = HAZ^T(ZAZ^T)^- Z = 0, \tag{4}$$

and from Lemma 2 it follows that

$$ZP^T = ZAZ^T(ZAZ^T)^- Z = ZAZ^T(ZAZ^T)^- ZAA^{-1} = ZAA^{-1} = Z.$$

As the final step, define matrix G^* as $\mathrm{Var}(P^T\gamma)$:

$$G^* = P^T GP. \tag{5}$$

Then

$$ZG^*Z^T = ZP^TG(ZP^T)^T = ZGZ^T$$

and, hence,

$$V(G^*) = ZG^*Z^T + R = ZGZ^T + R = V.$$

If $V(G^*) = V$ then also $\hat\beta(V(G^*)) = \hat\beta(V)$, and

$$\begin{aligned} Z\hat\gamma(R, G^*) &= ZG^*Z^TV^{-1}(Y - X\hat\beta(V(G^*))) \\ &= ZGZ^TV^{-1}(Y - X\hat\beta(V)) \\ &= Z\hat\gamma(R, G). \end{aligned}$$

Vector $\hat\gamma(R, G^*)$ of predicted random effects satisfies the condition $H\hat\gamma(R, G^*) = 0$ because of $HP^T = 0$:

$$\begin{aligned} H\hat\gamma(R, G^*) &\overset{(2)}{=} HG^*V(G^*)^{-1}(Y - X\hat\beta(V(G^*)) \\ &\overset{(5)}{=} HP^TGPV(G^*)^-(Y - X\hat\beta(V(G^*)) \\ &\overset{(4)}{=} 0. \end{aligned}$$

□

3. Identifiability of random effects

The likelihood function of Y for normally distributed Y is

$$\begin{aligned} L(y, G, \ldots) =& |2\pi(R + ZGZ^T)|^{-1/2} \\ & \times \exp\left(-\frac{1}{2}(y - X\beta)^T(R + ZGZ^T)^{-1}(y - X\beta)\right). \end{aligned}$$

Replacing G with some other covariance matrix may, but does not have to, change the likelihood function. It is possible to consider a class of covariance matrices $\mathcal{G}(G)$, which all lead to the same likelihood function (for any possible value of y):

$$\mathcal{G}(G) = \{G_i : L(y, G, \ldots) = L(y, G_i, \ldots) \mid \text{for all } y\}.$$

The class $\mathcal{G}(G)$ consists of covariance matrices which are in the equally good agreement with observed data and, hence, can not be preferred one over another on the basis of observed data alone. Note, that the matrix G^* defined in (5) also belongs to this set, $G^* \in \mathcal{G}(G)$.

Two covariance matrices $G_1, G_2 \in \mathcal{G}(G)$ may lead to different predictors for γ. However, if $G_1, G_2 \in \mathcal{G}(G)$, then $Z\hat\gamma(R, G_1) = Z\hat\gamma(R, G_2)$, as is proved in the following lemma.

Lemma 3. *If $G_1, G_2 \in \mathcal{G}(G)$, then $Z\hat{\gamma}(R, G_1) = Z\hat{\gamma}(R, G_2)$.*

Proof. Instead of the likelihood function, it is easier to work with log-likelihood function

$$\begin{aligned} l(y, G, \ldots) &= log(L(y, G, \ldots)) \\ &= log(|2\pi V|^{-1/2}) - \frac{1}{2}(y - X\beta)^T V^{-1}(y - X\beta), \qquad (6) \end{aligned}$$

where $V = V(G) = R + ZGZ^T$. If $G_1, G_2 \in \mathcal{G}(G)$, then $l(y, G_1, \ldots) = l(y, G_2, \ldots)$ for any value of y. Choosing $y = X\beta$ one gets the equality

$$log(|2\pi V(G_1)|^{-1/2}) = log(|2\pi V(G_2)|^{-1/2}), \qquad (7)$$

where $V(G_1) = R + ZG_1Z^T$ and $V(G_2) = R + ZG_2Z^T$. Choosing $y = X\beta + v$, where v is an arbitrary vector, and using (7), it follows from the equality of likelihood functions:

$$v^T V(G_1)^{-1} v = v^T V(G_2)^{-1} v.$$

The equation

$$v^T(V(G_1)^{-1} - V(G_2)^{-1})v = 0$$

holds for any v. In addition, the matrix $V(G_1)^{-1} - V(G_2)^{-1}$ is symmetric. Consequently,

$$V(G_1)^{-1} = V(G_2)^{-1},$$

and, hence, also $\hat{\beta}(V(G_1)) = \hat{\beta}(V(G_2))$. The predictor of $Z\gamma$ obtained using G_1 is

$$Z\hat{\gamma}(R, G_1) = ZG_1Z^T V(G_1)^{-1}(y - X\hat{\beta}(V(G_1))).$$

From (1) it follows, that $ZG_1Z^T = V(G_1) - R$ and therefore

$$\begin{aligned} Z\hat{\gamma}(R, G_1) &= (y - X\hat{\beta}(V(G_1))) - RV(G_1)^{-1}(y - X\hat{\beta}(V(G_1))) \\ &= (y - X\hat{\beta}(V(G_2))) - RV(G_2)^{-1}(y - X\hat{\beta}(V(G_2))) \\ &= Z\hat{\gamma}(R, G_2), \end{aligned}$$

which proves the lemma. □

Lemma 4. *Let an $n \times k$ matrix Z and a $v \times k$ matrix H satisfy the condition*

$$\text{rank}\begin{pmatrix} Z \\ H \end{pmatrix} = \text{rank}(Z) + \text{rank}(H) = k.$$

Then the condition $H\hat{\gamma} = 0$ uniquely determines the predictor of γ within the set of covariance matrices $\mathcal{G}(G)$.

Proof. From Theorem 1 it follows that there exists a matrix $G^* \in \mathcal{G}(G)$ such that the condition $H\hat{\gamma}(R, G^*) = 0$ holds. Let two matrices $G_1, G_2 \in \mathcal{G}(G)$ satisfy the condition $H\hat{\gamma}(R, G_1) = H\hat{\gamma}(R, G_2) = 0$. Then one can write

$$\begin{bmatrix} Z \\ H \end{bmatrix} \hat{\gamma}(R, G_1) = \begin{bmatrix} Z \\ H \end{bmatrix} \hat{\gamma}(R, G_2). \qquad (8)$$

Because the rank of $(n+v) \times k$ matrix $\begin{bmatrix} Z \\ H \end{bmatrix}$ is k, it is of full column rank and therefore the inverse of the matrix $\begin{bmatrix} Z \\ H \end{bmatrix}^T \begin{bmatrix} Z \\ H \end{bmatrix}$ exists. Therefore it is sufficient to multiply the equation (8) from the left with the matrix

$$\left[\begin{bmatrix} Z \\ H \end{bmatrix}^T \begin{bmatrix} Z \\ H \end{bmatrix}\right]^{-1} \begin{bmatrix} Z \\ H \end{bmatrix}^T$$

to derive the conclusion

$$\hat{\gamma}(R, G_1) = \hat{\gamma}(R, G_2),$$

which proves the lemma. □

4. Discussion

For fixed effects it is well known that if the matrix X is not of full column rank, then one needs to apply additional constraints to define the unknown parameters in β uniquely. These constraints can be represented in the form $H\beta = 0$, where

$$\text{rank}\begin{pmatrix} X \\ H \end{pmatrix} = \text{rank}(X) + \text{rank}(H).$$

The validity of these constraints on observed data cannot be assessed using the observed data. Different constraints will lead to different but mathematically equal parameterizations of the linear model. Nevertheless, the linear combinations of parameters in the form $X\beta$ will remain the same for all parameterizations. Interested reader is referred to Scheffé (1959) and Searle (1971) for more detailed coverage of the topic.

It was shown in this article, that based on the observed data alone cannot exactly identify the covariance matrix $G = \text{Var}(\gamma)$. Unidentifiability of G causes also the unidentifiability of γ, if the matrix Z does not have a full column rank. To determine γ uniquely one has to imply some assumptions that can not be tested. These assumptions may be presented as assumptions on the structure of G but they may also be represented in the form $H\gamma = 0$, where ranks satisfy (3).

The restrictions $H\gamma = 0$ are preferable since they are less restrictive – namely there exist several covariance matrices G yielding exactly the same predictors $\hat{\gamma}$. Choosing an exact form for the covariance matrix among those yielding identical results may be unnecessary in practical applications.

As in the case of fixed effects, the correctness of implied restrictions cannot be assessed using only the observed data.

There are some aspects of Theorem 1 which should be pointed out. First, one can consider matrices G and R as some functions of unknown parameters $\sigma_1^2, \ldots, \sigma_k^2$. Then the matrix functions

$$V = ZG(\sigma_1^2, \ldots, \sigma_k^2)Z^T + R(\sigma_1^2, \ldots, \sigma_k^2)$$

and

$$V(G^*) = ZG^*(\sigma_1^2, \dots, \sigma_k^2)Z^T + R(\sigma_1^2, \dots, \sigma_k^2),$$

where

$$G^*(\sigma_1^2, \dots, \sigma_k^2) = P^T G(\sigma_1^2, \dots, \sigma_k^2)P, \qquad (9)$$

are equal. Therefore, replacing G with G^* does not change any function which depends on G only via the covariance matrix V. For example, if Y is normally distributed, then the likelihood function of Y depends on the covariance parameters via the matrix function V only. Hence, the likelihood function of Y does not change if one replaces G with G^*. In situations, where the covariance parameters are unknown and one uses the likelihood function of Y to derive the maximum likelihood estimates of covariance parameters, the estimates remain unaffected if one uses G^* instead of G, because the likelihood function does not change.

The other frequently used method for estimating unknown covariance parameters in mixed models is the restricted maximum likelihood (REML) method. Under REML, one maximizes the likelihood of a linearly transformed vector kY instead of the likelihood of Y. But the likelihood of the transformed vector kY depends on G only via the matrix V. Hence, replacing G with G^* does not affect the estimates of covariance parameters (if Y is normally distributed and one uses REML or ML method). Therefore, for normally distributed data and matrices G and G^* related by (9), the EBLUE (estimated BLUE) estimates will be equal

$$X\hat{\beta}(\hat{V}) = X\hat{\beta}(\hat{V}(\hat{G}^*)),$$

and EBLUP (estimated BLUP) predictors will also be equal

$$Z\hat{\gamma}(\hat{R}, \hat{G}) = Z\hat{\gamma}(\hat{R}, \hat{G}^*).$$

It is worth to note here that the derivation of G^* is not necessarily unique, because there may exist more than one matrix A satisfying the key condition $HAZ^T = 0$.

References

Harville, D. A. (1997). *Matrix Algebra from a Statistician's Perspective*. Springer-Verlag, Berlin.

Scheffé, M. (1959). *The Analysis of Variances*. Wiley, New York.

Searle, S. R. (1971). *Linear Models*. Wiley, New York.

Institute of Mathematical Statistics, J. Liivi 2, 50409 Tartu, Estonia

E-mail address: martm@kodu.ee