MIXED MODEL APPROACHES FOR ESTIMATING GENETIC COVARIANCES BETWEEN TWO TRAITS WITH UNEQUAL DESIGN MATRICES

Zhu   Jun

(Agronomy Department, Zhejiang Agricultural University, Hangzhou 310029)

ABSTRACT

Mixed model approaches are proposed for estimating covariance components between two types of traits with unequal experimental design matrices. A system of equations presented can be used by either minimum norm quadratic unbiased estimation (MINQUE) method or restricted maximum likelihood (REML) method. An application is given to demonstrate how to estimate covariance components between seed traits and maternal plant traits. MINQUE(0/1) method, which   is a MINQUE method setting 0 for prior covariances between direct and maternal effects and 1 for other prior covariances, is suitable for estimating covariance components without using iteration.

Key Words: Mixed models, Unequal design matrices, Covariance estimation, Seeds and maternal plants

1. INTRODUCTION

Estimating genetic covariances is of importance for quantitative genetic research on multiple traits as well as for indirect selection in plant and animal breeding. There have been many ways for estimating genetic covariance components. By equating the sum of cross products to its expectation, estimation of covariance components can be obtained with the approaches of multivariate analysis of variance (MANOVA) (Grossman and Gall, 1968). Although these approaches have been widely used in quantitative genetics and index selection, they are only applicable for some simple genetic models with balanced data. The mixed model approaches have been developed to estimate covariance components for balanced or unbalanced data with complicated genetic models (Rao and Kleffe, 1980; Zhu, 1992). By now there is still no appropriate method for estimating covariance components between two traits with unequal design matrices.

The purpose of this paper is to develop new methods which can give unbiased estimation of covariance components for two traits with unequal design matrices. Methods for analyzing traits of seeds and their maternal plants are proposed as a specific application.

2 MIXED LINEAR MODELS FOR TRAITS WITH UNEQUAL DESIGN MATRICES

By the mixed model approaches, observation vector of any trait can be expressed as a linear model

including both fixed and random effects. Various traits will have different fixed effects and random effects. Most traits analyzed in a paper are usually measured on the same experimental units and have equal design matrices. Some traits measured on different experimental units, eg., seeds and plants, will have unequal design matrices. If the ith trait is affected by the fixed effects $\mathbf{b}_{(i)}$ and several random effects $\mathbf{e}_{(i)j}$ ($j = 1, 2, \ldots, k$), the vector $\mathbf{y}_i$ with n observations can be written as a general form of a mixed linear model,

$$\begin{aligned}\mathbf{y}_i &= \mathbf{X}\mathbf{b}_{(i)} + \mathbf{U}_{(i)1}\mathbf{e}_{(i)1} + \mathbf{U}_{(i)2}\mathbf{e}_{(i)2} + \cdots + \mathbf{U}_{(i)k}\mathbf{e}_{(i)k} \\ &= \mathbf{X}\mathbf{b}_{(i)} + \sum_{j=1}^{k}\mathbf{U}_{(i)j}\mathbf{e}_{(i)j}\end{aligned}$$

where $\mathbf{X}$ is an $n{\times}p$ matrix of known constants with rank min(n,p),

$\quad$ $\mathbf{b}_{(i)}$ is a $p{\times}1$ vector of unknown fixed parameters for trait i,

$\quad$ $\mathbf{e}_{(i)j}$ for $j = 1, 2, \ldots, k$ are $q_j{\times}1$ vectors of random variables for trait i with E($\mathbf{e}_{(i)j}$) = $\mathbf{0}$ and Var($\mathbf{e}_{(i)j}$) = $\sigma^2_{(i)j}\mathbf{I}$,

$\quad$ $\mathbf{U}_{(i)j}$ for $j = 1, 2, \ldots, k$ are $n{\times}q_j$ known incidence matrices to the jth random vector $\mathbf{e}_{(i)j}$ of trait i.

Since some random vectors in the model may be correlated, the variance-covariance matrix of observation vector $\mathbf{y}_i$ may include some covariance components as well as k variance components.

$\quad$ For trait i and trait i', covariances between observation vectors $\mathbf{y}_i$ and $\mathbf{y}_{i'}$ are as follows,

$$\begin{aligned}\mathrm{Cov}(\mathbf{y}_i, \mathbf{y}_i^T) &= \sigma_{1/1}\mathbf{F}_{(ii')1} + \sigma_{2/2}\mathbf{F}_{(ii')2} + \cdots + \sigma_{r/r}\mathbf{F}_{(ii')r} \\ &= \sum_{u=1}^{r}\sigma_{u/u}\mathbf{F}_{(ii')u}\end{aligned}$$

$$\begin{aligned}\mathrm{Cov}(\mathbf{y}_i, \mathbf{y}_i^T) &= \sigma_{1/1}\mathbf{F}_{(i'i)1} + \sigma_{2/2}\mathbf{F}_{(i'i)2} + \cdots + \sigma_{r/r}\mathbf{F}_{(i'i)r} \\ &= \sum_{u=1}^{r}\sigma_{u/u}\mathbf{F}_{(i'i)u}\end{aligned}$$

where $\quad \sigma_{u/u}$ for $u = 1, 2, \ldots, r$ ($r \times k$) are covariance components for two random vectors,

$\quad$ $\mathbf{F}_{(ii')u}$ for $u = 1, 2, \ldots, r$ are known incidence matrices to the uth covariance component between trait i and trait i',

$\quad$ $\mathbf{F}_{(i'i)u}$ for $u = 1, 2, \ldots, r$ are known incidence matrices to the uth covariance component between trait i' and trait i.

If trait i and trait i' have equal design matrices, $\mathbf{F}_{(ii')u} = \mathbf{F}_{(i'i)u}$. When two traits have unequal design matrices, $\mathbf{F}_{(ii')u} \neq \mathbf{F}_{(i'i)u}$.

3. METHODS FOR ESTIMATING COVARIANCE COMPONENTS

$\quad$ For traits with equal design matrices, unbiased estimation of covariance components between two traits can be obtained by equating quadratic functions of two observation vectors to their expectations (Zhu, 1992).

These methods can also be applied in estimation of covariance components between two traits with unequal design matrices. For trait i and trait i' with unequal design matrices, the following covariance matrix $\mathbf{V}$ is symmetric,

$$\mathbf{V} = \mathrm{Cov}(\mathbf{y}_i, \mathbf{y}_{i'}^T) + \mathrm{Cov}(\mathbf{y}_{i'}, \mathbf{y}_i^T)$$

$$= \sum_{u=1}^{r} \sigma_{u/u}(\mathbf{F}_{(ii')u} + \mathbf{F}_{(i'i)u})$$

$$= \sum_{u=1}^{r} \sigma_{u/u}\mathbf{F}_u$$

Define matrices $\mathbf{V}_{(\alpha)}$ and $\mathbf{Q}_{(\alpha)}$ as

$$\mathbf{V}_{(\alpha)} = \sum_{u=1}^{r} \alpha_{u/u}(\mathbf{F}_{(ii')u} + \mathbf{F}_{(i'i)u})$$

$$= \sum_{u=1}^{r} \alpha_{u/u}\mathbf{F}_u$$

$$\mathbf{Q}_{(\alpha)} = \mathbf{V}_{(\alpha)}^{-1} - \mathbf{V}_{(\alpha)}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{V}_{(\alpha)}^{-1}\mathbf{X})^{+}\mathbf{X}^T\mathbf{V}_{(\alpha)}^{-1}$$

where α_u $(u = 1, 2, \ldots, r)$ are some prior values which do not depend on observed data. Then the expectation of quadratic function $\mathrm{E}(\mathbf{y}_i^T\mathbf{Q}_{(\alpha)}\mathbf{F}_u\mathbf{Q}_{(\alpha)}\mathbf{y}_{i'}, \mathbf{y}_{i'}^T\mathbf{Q}_{(\alpha)}\mathbf{F}_u\mathbf{Q}_{(\alpha)}\mathbf{y}_i)$ can be represented as the product of a row vector and a column vector,

$$\mathrm{E}(\mathbf{y}_i^T\mathbf{Q}_{(\alpha)}\mathbf{F}_u\mathbf{Q}_{(\alpha)}\mathbf{y}_{i'}, \mathbf{y}_{i'}^T\mathbf{Q}_{(\alpha)}\mathbf{F}_u\mathbf{Q}_{(\alpha)}\mathbf{y}_i)$$

$$= \mathrm{E}(2\mathbf{y}_i^T\mathbf{Q}_{(\alpha)}\mathbf{F}_u\mathbf{Q}_{(\alpha)}\mathbf{y}_{i'})$$

$$= \mathrm{tr}(\mathbf{Q}_{(\alpha)}\mathbf{F}_u\mathbf{Q}_{(\alpha)}\mathbf{V})$$

$$= \mathrm{tr}[\mathbf{Q}_{(\alpha)}\mathbf{F}_u\mathbf{Q}_{(\alpha)}(\sum_{v=1}^{r}\alpha_{v/v}\mathbf{F}_v)]$$

$$= [\mathrm{tr}(\mathbf{Q}_{(\alpha)}\mathbf{F}_u\mathbf{Q}_{(\alpha)}\mathbf{F}_v)][\alpha_{v/v}], v = 1,2,\cdots,r.$$

When setting all the quadratic functions for $u = 1, 2, \ldots, r$ to their expectations, respectively, unbiased estimators of covariances [$\sigma_{u/u}$] can then be obtained by solving the system of equations ($u, v = 1, 2, \ldots, r$),

$$[\mathrm{tr}(\mathbf{Q}_{(\alpha)}\mathbf{F}_u\mathbf{Q}_{(\alpha)}\mathbf{F}_v)][\hat{\sigma}_{v/v}] = [2\mathbf{y}_i^T\mathbf{Q}_{(\alpha)}\mathbf{F}_u\mathbf{Q}_{(\alpha)}\mathbf{y}_{i'}]$$

The matrix $\mathbf{Q}_{(\alpha)}$ in Eq. (1) is a function of $\mathbf{V}_{(\alpha)}$ which consists of prior values $\alpha_{u/u}$ $(u = 1, 2, \ldots, r)$. The prior values may be chosen in some different ways. When setting $\alpha_{u/u} = 1$ for all u, estimation from solving Eq. (1) is called the MINQUE(1) estimation. MINQUE(1) is one of the methods of minimum norm quadratic unbiased estimation (MINQUE) (Rao, 1971). If $\alpha_{u/u}$ is replaced by the iterated estimates with the restriction that they are within parameter space, new estimates can be obtained by iteration until converging. This procedure is the restricted maximum likelihood (REML) method (Patterson and Thompsom, 1971; Corbeil and Searle, 1976).

4. AN APPLICATION FOR ANALYZING TRAITS OF SEEDS AND THEIR MATERNAL PLANTS

Evaluation of nutrition merit for plant seeds is now more and more important in crop breeding. Zhu (1992) proposed two genetic models for both diploid seeds and triploid endosperms. These models include direct genetic effects (A and D), cytoplasm effects (C), and maternal genetic effects (Am and Dm). Observations of seed traits $\mathbf{y}_S$ can be rewritten in a general form used by this paper,

$$\mathbf{y}_S = \mathbf{X}\mathbf{b}_{(S)} + \mathbf{U}_A\mathbf{e}_{(S)A} + \mathbf{U}_D\mathbf{e}_{(S)D} + \mathbf{U}_C\mathbf{e}_{(S)C} + \mathbf{U}_{Am}\mathbf{e}_{(S)Am} + \mathbf{U}_{Dm}\mathbf{e}_{(S)Dm} + \mathbf{U}_B\mathbf{e}_{(S)B} + \mathbf{e}_{(S)e}$$

$$\sim (\mathbf{X}\mathbf{b}_{(S)}, \Sigma_S = \sigma^2_{(S)A}\mathbf{V}_A + \sigma^2_{(S)D}\mathbf{V}_D + \sigma^2_{(S)C}\mathbf{V}_C + \sigma^2_{(S)Am}\mathbf{V}_C + \sigma^2_{(S)Dm}\mathbf{V}_{Dm} + \sigma^2_{(S)B}\mathbf{V}_B$$

$$+ \sigma_{(S)A.Am}\mathbf{V}_{A.Am} + \sigma_{(S)D.Dm}\mathbf{V}_{D.Dm} + \sigma^2_{(S)e}\mathbf{V}_e)$$

where $\mathbf{V}_A = \mathbf{U}_A\mathbf{U}_{Am}^T, \mathbf{V}_D = \mathbf{U}_D\mathbf{U}_{Dm}^T, \mathbf{V}_C = \mathbf{U}_C\mathbf{U}_C^T, \mathbf{V}_{Am}\mathbf{U}_{Am}\mathbf{U}_{Am}^T, \mathbf{V}_{Dm} = \mathbf{U}_{Dm}\mathbf{U}_{Dm}^T, \mathbf{V}_B = \mathbf{U}_B\mathbf{U}_B^T$

$\mathbf{V}_{A.Am} = \mathbf{U}_A\mathbf{U}_{Am}^T + \mathbf{U}_{Am}\mathbf{U}_A^T, \mathbf{V}_{D.Dm} = \mathbf{U}_D\mathbf{U}_{Dm}^T + \mathbf{U}_{Dm}\mathbf{U}_D^T, \mathbf{V}_e = \mathbf{I}$

In the seed model seven variance components are estimable for direct additive variance $\sigma^2_{(S)A}$, direct dominance variance $\sigma^2_{(S)D}$, cytoplasmic variance $\sigma^2_{(S)C}$, maternal additive variance $\sigma^2_{(S)Am}$, maternal dominance variance $\sigma^2_{(S)Dm}$, block variance $\sigma^2_{(S)B}$, and residual variance $\sigma^2_{(S)e}$. Since seed genes are contributed in part by its maternal plant, direct gene effects are therefore not independent to maternal gene effects. Estimated covariance components can also be obtained for additive covariance $\sigma_{(S)A.Am}$ and dominance covariance $\sigma_{(S)D.Dm}$.

The genetic model for maternal plants can be treated as a reduced model from the seed model in which direct additive effects and direct dominance effects are eliminated. Observations for maternal plants $\mathbf{y}_P$ can be expressed in a similar way,

$$\mathbf{y}_P = \mathbf{X}\mathbf{b}_{(P)} + \mathbf{U}_C\mathbf{e}_{(P)C} + \mathbf{U}_{Am}\mathbf{e}_{(P)Am} + \mathbf{U}_{Dm}\mathbf{e}_{(P)Dm} + \mathbf{U}_B\mathbf{e}_{(P)B} + \mathbf{e}_{(P)e}$$

$$\sim (\mathbf{X}\mathbf{b}_{(P)}, \Sigma_P = \sigma^2_{(P)C}\mathbf{V}_C + \sigma^2_{(P)Am}\mathbf{V}_{Am} + \sigma^2_{(P)Dm}\mathbf{V}_{Dm} + \sigma^2_{PS)B}\mathbf{V}_B + \sigma^2_{(S)e}\mathbf{V}_e)$$

Since traits for seeds and their maternal plants are based on different genetic models, the design matrices for these two types of traits are not completely equal. In spite of the inequality for matrices, there exist some correlations between traits of seeds and maternal plants. The covariance matrices for observation vectors of seeds and maternal plants are

$$\mathrm{Cov}(\mathbf{y}_S, \mathbf{y}_P) = \sigma_{A/Am}\mathbf{U}_A\mathbf{U}_{Am}^T + \sigma_{D/Dm}\mathbf{U}_D\mathbf{U}_{Dm}^T + \sigma_{C/C}\mathbf{U}_C\mathbf{U}_C^T$$

$$+ \sigma_{Am/Am}\mathbf{U}_{Am}\mathbf{U}_{Am}^T + \sigma_{Dm/Dm}\mathbf{U}_{Dm}\mathbf{U}_{Dm}^T + \sigma_{B/B}\mathbf{U}_B\mathbf{U}_B^T + \sigma_{e/e}\mathbf{I}$$

$$\mathrm{Cov}(\mathbf{y}_P, \mathbf{y}_S) = \sigma_{A/Am}\mathbf{U}_{Am}\mathbf{U}_A^T + \sigma_{D/Dm}\mathbf{U}_{Dm}\mathbf{U}_D^T + \sigma_{C/C}\mathbf{U}_C\mathbf{U}_C^T$$

$$+ \sigma_{Am/Am}\mathbf{U}_{Am}\mathbf{U}_{Am}^T + \sigma_{Dm/Dm}\mathbf{U}_{Dm}\mathbf{U}_{Dm}^T + \sigma_{B/B}\mathbf{U}_B\mathbf{U}_B^T + \sigma_{e/e}\mathbf{I}$$

where $\sigma_{A/Am}$ = additive covariance between seed and plant additive effects, $\sigma_{D/Dm}$ = dominance covariance between seed and plant dominance effects, $\sigma_{C/C}$ = cytoplasmic covariance, $\sigma_{Am/Am}$ = maternal additive

covariance, $\sigma_{Dm/Dm}$ = maternal dominance covariance, $\sigma_{B/B}$ = block covariance, and $\sigma_{e/e}$ = residual covariance.

The sum of these two covariance matrices can be represented as

with

$$\begin{aligned}
\mathbf{V} &= \mathrm{Cov}(\mathbf{y}_S, \mathbf{y}_P) + \mathrm{Cov}(\mathbf{y}_P, \mathbf{y}_S) \\
&= \sigma_{A/Am}\mathbf{F}_1 + \sigma_{D/Dm}\mathbf{F}_2 + \sigma_{C/C}\mathbf{F}_3 + \sigma_{Am/Am}\mathbf{F}_4 + \sigma_{Dm/Dm}\mathbf{F}_5 \\
&\quad + \sigma_{B/B}\mathbf{F}_6 + \sigma_{e/e}\mathbf{F}_7 \\
&= \sum_{u=1}^{7} \sigma_{u/u}\mathbf{F}_u
\end{aligned}$$

where $\mathbf{F}_1 = (\mathbf{U}_A\mathbf{U}_{Am}^T + \mathbf{U}_{Am}\mathbf{U}_A^T)$, $\mathbf{F}_2 = (\mathbf{U}_D\mathbf{U}_{Dm}^T + \mathbf{U}_{Dm}\mathbf{U}_D^T)$, $\mathbf{F}_3 = (2\mathbf{U}_C\mathbf{U}_C^T)$, $\mathbf{F}_4 = (2\mathbf{U}_{Am}\mathbf{U}_{Am}^T)$, $\mathbf{F}_5 = (2\mathbf{U}_{Dm}\mathbf{U}_{Dm}^T)$, $\mathbf{F}_6 = (2\mathbf{U}_B\mathbf{U}_B^T)$, $\mathbf{F}_7 = 2\mathbf{I}$.

MINQUE(0/1) method has been suggested for estimation variance components and covariance components for seed traits (Zhu, 1992). MINQUE(0/1) is a MINQUE method with all prior covariances between seed direct effects and maternal effects setting 0 and all other prior values setting 1. This approach is also suitable for estimating covariance components between seeds and plants. Define $\mathbf{V}_{(0/1)}$ and $\mathbf{Q}_{(0/1)}$ as

$$\mathbf{V}_{(0/1)} = 2[\mathbf{U}_C\mathbf{U}_C^T + \mathbf{U}_{Am}\mathbf{U}_{Am}^T + \mathbf{U}_{Dm}\mathbf{U}_{Dm}^T + \mathbf{U}_B\mathbf{U}_B^T + \mathbf{I}]$$

and

$$\mathbf{Q}_{(0/1)} = \mathbf{V}_{(0/1)}^{-1} - \mathbf{V}_{(0/1)}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{V}_{(0/1)}^{-1}\mathbf{X})^+\mathbf{X}^T\mathbf{V}_{(0/1)}^{-1}$$

Then solving the following Eq. (2)

$$[\mathrm{tr}(\mathbf{Q}_{(0/1)}\mathbf{F}_u\mathbf{Q}_{(0/1)}\mathbf{F}_v)][\hat{\sigma}_{u/u}] = [2\mathbf{y}_S^T\mathbf{Q}_{(0/1)}\mathbf{F}_u\mathbf{Q}_{(0/1)}\mathbf{y}_P]$$

will give unbiased estimation for covariance components between seed traits and maternal traits.

5. DISCUSSION

The methods proposed in this paper can be used for estimating covariance components between two traits with unequal design matrices. If two traits have equal design matrices, Eq. (1) will give the same solution as to those from the common approaches of MINQUE or REML. There are two requirements for using the new method: 1) two traits having correlated random vectors with equal number of effects; 2) observations for two traits being the same size. These two requirements are also needed for estimating covariances between traits with equal design matrices.

Sampling variances for the estimated covariance components may be obtained by the jackknife technique (Miller, 1974; Zhu, 1992), if the MINQUE approaches are used. Experimental blocks or genetic entries can be served as the resampling units. When the jackknife estimates and their standard errors are available, an approximate *t*-test is followed for hypothesis tests or for construction of confidence intervals. The REML

approaches are appropriate for estimating asymptotic sampling variances of covariance components (Rao and Kleffe, 1988). Then the z-test is applicable for hypothesis tests.

ACKNOWLEDGEMENTS

REFERENCES

[1]   Corbeil, R. R. and Searle, S. R. (1976). Restricted maximum likelihood (REML) estimation of covariance components in the mixed model. Technometrics 18:31-38.

[2]   Grossman, M and Gall, G. A. E. 1968. Covariance analysis with unequal subclass number: components estimation in quantitative genetics. Biometrics 24: 49-59.

[3]   Miller, R. G. (1974). The jackknife a review. Biometrika 61: 1-15.

[4]   Patterson, H. D. and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. Biometrika 58: 545-554.

[5]   Rao, C. R. (1971). Estimation of variance and covariance components MINQUE theory. J. Multivar. Anal. 1: 257-275.

[6]   Rao, C. R. and Kleffe, J. (1980). Estimation of variance components. In Handbook of Statistics. Vol. 1, 1-40. Krishnaiah, P. R. (ed.). North-Holland, New York.

[7]   Rao, C. R. and Kleffe, J. (1988). *Estimation of Variance Components and Applications*. North-Holland, New York.

[8]   Zhu, J. (1992). Mixed model approaches for estimating genetic variances and covariances. J. Biomath. 7: 1-11.