“PDMATRIX” -- PROGAMS TO MAKE MATRICES POSITIVE DEFINITE

John M. Henshall[1] and Karin Meyer

Animal Genetics and Breeding Unit[*], University of New England,
Armidale, NSW 2351 Australia

INTRODUCTION

Quantitative genetic theory for continuous traits under the infinitesimal model assumes a multivariate normal distribution with positive (semi-) definite covariance matrix. In estimating such matrices, however, we frequently obtain matrices with negative eigenvalues due to sampling errors. “Bending”, i.e. regression of the eigenvalues to their mean so that the smallest value is zero, as suggested by Hayes and Hill (1981), has been widely used to constrain such estimated covariance matrices.

In practice, estimates of covariance components often originate from different analyses, and can be based on greatly differing amounts of information. “Bending” and related techniques do not allow for differential weighing of individual component estimates. PDMATRIX comprises two programs to generate positive definite covariance matrices from individual estimates, accounting for differences in accuracy. FLBEND modifies a non-positive definite matrix comprised of pair wise covariance estimates, where the estimates may or may not be based on experimental results. ITSUMCOV deals with the specific case where we want to combine results from multiple, multivariate analyses of various subsets of all traits.

COMPUTING METHODS

FLBEND finds a positive definite covariance matrix, which is ‘minimum distance’ from a non-positive definite matrix comprised of pair wise covariance estimates. The distance function is the sum of squared distances, weighted by any available standard error estimates, the weighted Euclidean norm. The minimization algorithm differs from other minimum distance implementations in that a reparameterisation is used to simplify the minimisation problem.

Finding a positive definite matrix $\mathbf{V}_{\mathrm{pd}}$ at a minimum distance from a non-positive definite matrix $\mathbf{V}_{\mathbf{npd}}$ is a constrained minimisation problem, and the boundary of the constraint is not a simple function. For a q x q matrix $\mathbf{B}$ of full rank, $\mathbf{B'B}$ is a positive definite matrix. A lower triangular matrix can be of full rank, and making $\mathbf{B}$ lower triangular has the advantage that only $q(q+1)/2$ parameters need to be estimated instead of q^2. The formulation of the minimisation problem then is, find lower triangular full rank matrix $\mathbf{B}$ such that the weighted Euclidean norm of ($\mathbf{B'B}$ - $\mathbf{V}_{\mathrm{npd}}$) is minimised.

---

[1] Current address :CSIRO Livestock Industries, Rockhampton, QLD 4702, Australia

[*] AGBU is a joint institute of NSW Agriculture and The University of New England

ITSUMCOV implements the ‘iterative summing of expanded part matrices’ approach suggested by Mäntysaari (1999), extended to allow for differential weighing of results from individual analyses. Consider q traits, and let $\mathbf{V}^t$ of size q x q denote the covariance matrix among all traits for the t-th iterate. Let $\mathbf{C}_s$ represent the matrix of covariance estimates from the s-th analysis, of size k x k. Define a transformation matrix $\mathbf{P}_s$, of size q x k, which has k elements of unity, $p_{ij}=1$, if the i-th trait overall is the j-th trait in the s-th analysis, and zero otherwise. The iterative summation scheme is then given as

$$\begin{aligned}\mathbf{V}^{t+1} = \sum_{s=1}^{S} w_s \{ & \mathbf{V}^t (\mathbf{P}_s \mathbf{P}_s' \mathbf{V}^t \mathbf{P}_s \mathbf{P}_s')^- \mathbf{P}_s \mathbf{C}_s \mathbf{P}_s' (\mathbf{P}_s \mathbf{P}_s' \mathbf{V}^t \mathbf{P}_s \mathbf{P}_s')^- \mathbf{V}^t \\ & + [(\mathbf{I} - \mathbf{P}_s \mathbf{P}_s')(\mathbf{V}^t)^{-1} (\mathbf{I} - \mathbf{P}_s \mathbf{P}_s')]^- \} / \sum_{s=1}^{S} w_s\end{aligned} \tag{1}$$

with $\mathbf{I}$ an identity matrix of size q, w_s the weight for the s-th analysis, and ‘$^-$’ denoting a generalised inverse. Starting values ($\mathbf{V}^0$) are obtained by averaging variance component estimates for each trait and correlations over individual analyses. If the combined correlation matrix is not positive definite, it is “bent” so that the smallest eigenvalue is equal to an operational zero. Multiplying the modified correlation matrix with the appropriate average standard deviations gives $\mathbf{V}^0$. Equation (1) is iterated until a given convergence criterion is met.

USE

FLBEND has been used to obtain covariance matrices for use in Australian sheep genetic evaluations. Variance components for groups of 42 traits were combined into genetic and residual matrices which were close to, but not positive definite. Using FLBEND positive definite matrices were obtained in which the largest deviations from the composite matrices were for those correlations with highest standard errors.

ITSUMCOV has been used to combine estimates from 231 bivariate analyses for records on Hereford cattle to estimate the complete covariance among the 22 traits considered in BREEDPLAN, the Australian genetic evaluation scheme for beef cattle (Meyer and Johnston, unpublished).

AVAILABILITY

PDMATRIX can be downloaded from :
`http://agbu.une.edu.au/~kmeyer/pdmatrix.html`

COMPUTING ENVIRONMENT}

PDMATRIX is written in standard FORTRAN 90 and is self-contained. Source code and compiled versions for Linux and Compaq Alpha stations are available.

REFERENCES

Hayes, J. F. and Hill, W. G. H. (1981) *Biometrics* **37**:483-493.
Mäntysaari, E. A. (1999) *50th Ann. Meet. Europ. Ass. Anim. Prod.*, Mimeo