

PENALIZED ESTIMATION OF COVARIANCE MATRICES WITH FLEXIBLE AMOUNTS OF SHRINKAGE

Karin Meyer

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

SUMMARY

Penalized maximum likelihood estimation has been advocated for its capability to yield substantially improved estimates of covariance matrices, but so far only cases with equal numbers of records have been considered. We show that a generalization of the inverse Wishart distribution can be utilised to derive penalties which allow for differential penalization for different blocks of the matrices to be estimated. However, this requires multiple tuning factors to be determined and thus can increase computational requirements markedly. Simulation results are presented which indicate that the additional gains obtainable for estimates of genetic covariance components – over and above those from a simple, non-differential scheme – are moderate, even if numbers of records for different traits differ by orders of magnitude.

INTRODUCTION

Estimation of covariance components by restricted maximum likelihood (REML) subject to a penalty borrowing strength from the phenotypic covariance matrix, has been shown to yield estimates closer to the population values than their ‘standard’ unpenalized counterparts (Meyer 2011). So far, studies to evaluate the properties of penalized estimates only considered equal numbers of measurements for all traits. In practice, however, we may have subgroups of traits with greatly differing numbers of records. A particular type of penalty – motivated by Bayesian estimation – is given by minus the logarithmic value of the density of an inverse Wishart (IW) distribution added to the REML log likelihood. Using the phenotypic covariance as scale matrix, this shrinks individual, e.g. genetic, matrices towards the former (Meyer *et al.* 2011). A drawback of this ‘prior’ is the rigidity imposed by a single parameter for the degrees of freedom. Hence an extension to a generalized inverse Wishart (GIW) distribution (Brown 2006) has been proposed as a more flexible alternative. Munilla and Cantet (2012) give details together with an application to account for differential uncertainty in genetic parameters in a Bayesian analysis.

This paper describes a penalty based on the GIW distribution and presents a simulation study examining the effect of this penalty on sampling properties of penalized REML estimates of covariance matrices for unequal numbers of records between traits.

THE GIW PENALTY

Consider q traits with covariance matrix Σ , ordered so that 1 to q_1 are the subset of traits measured on a group of individuals without records for traits $q_1 + 1$ to q , while a second group has all q traits measured. Assume Σ has an IW distribution with scale matrix Ω . This gives ‘whole matrix’ penalty

$$\mathcal{P} = C \log |\Sigma| + \text{tr} (\Sigma^{-1} \Omega) \tag{1}$$

with $C \approx 1$ a constant depending on q and the degrees of freedom. Partition Σ and Ω into

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \quad \text{and} \quad \Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}$$

according to the subsets of traits. Σ_{11} is independent of $\Sigma_{22.1} = \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}$ and $\Sigma_{11}^{-1} \Sigma_{12}$ and has an IW distribution. The penalty for the first sub-matrix is then simply obtained ignoring

*AGBU is a joint venture of NSW Department of Department of Primary Industries and the University of New England

the remaining traits. Similarly, for $\Omega_{22.1} = \Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12}$ and C_{ij} similar to C the conditional distribution for the second block given the first is IW, which yields penalties

$$\mathcal{P}_{11} = C_{11} \log |\Sigma_{11}| + \text{tr} (\Sigma_{11}^{-1} \Omega_{11}) \quad \text{and} \quad \mathcal{P}_{22} = C_{22} \log |\Sigma_{22.1}| + \text{tr} (\Sigma_{22.1}^{-1} \Omega_{22.1}) \quad (2)$$

Expanding (1) in terms of the submatrices and subtracting \mathcal{P}_{11} and \mathcal{P}_{22} gives the penalty for the remaining covariance components (assuming $C_{11} = C_{22} = C$)

$$\mathcal{P}_{12} = \text{tr} (\Sigma_{22.1}^{-1} [(\Omega_{21}\Omega_{11}^{-1} - \Sigma_{21}\Sigma_{11}^{-1}) \Omega_{12} + (\Sigma_{21}\Sigma_{11}^{-1}\Omega_{11} - \Omega_{21}) \Sigma_{11}^{-1}\Sigma_{12}]) \quad (3)$$

It can be shown that \mathcal{P}_{12} is proportional to minus the log density for $\Sigma_{11}^{-1}\Sigma_{12}$ assumed to have a matrix-variate Normal distribution. These arguments are readily generalized to more subsets of traits; Brown (2006) summarizes the GIW as a series of sequential, conditional distributions.

MATERIAL AND METHODS

Data for 14 traits were simulated by sampling genetic and residual effects from appropriate multi-variate Normal distributions for a paternal half-sib design and different combinations of population heritabilities and correlations. For case A and B, all heritabilities were assumed equal, 0.4 and 0.2, respectively. For case C, values for traits 1 to 14 were 2×0.6 , 0.55, 2×0.5 , 0.45, 2×0.4 , 0.35, 2×0.3 , 0.25 and 2×0.2 . For scenario I, all correlations were assumed to be zero and all phenotypic variances were set to 1. For II, all genetic and residual correlations were equal, 0.5 and 0.2, respectively, and for III correlations between traits i and j were set to $0.5^{|i-j|}$ (genetic) and $0.2^{|i-j|}$ (residual), while phenotypic variances were set to $\text{mod}(i, 3) + 1$. This yielded nine sets of population parameters, referred to as A-I to C-III henceforth. Records for all traits were obtained for $s_1 = 400$ sires with 10 progeny each. In addition, records for the first $q_1 = 3, 5, 7, 9$ and 11 traits only were sampled for $s_2 = 400$ or $s_2 = 2000$ sires with 20 progeny. A total of 500 replicates per case were carried out.

Analyses. For each replicate, REML estimates of genetic (Σ_G) and residual (Σ_E) covariance matrices were obtained subject to five types of penalty, involving up to three different tuning factors (ψ_i)

$$\begin{aligned} \mathcal{P}_a &= \psi_1(\mathcal{P}_{22} + \mathcal{P}_{12} + \mathcal{P}_{11}) = \psi_1 \mathcal{P} & \mathcal{P}_c &= \psi_1 \mathcal{P}_{22} + \psi_2 \mathcal{P}_{12} \\ \mathcal{P}_b &= \psi_1(\mathcal{P}_{22} + \mathcal{P}_{12}) = \psi_1(\mathcal{P} - \mathcal{P}_{11}) & \mathcal{P}_d &= \mathcal{P}_e = \psi_1 \mathcal{P}_{22} + \psi_2 \mathcal{P}_{12} + \psi_3 \mathcal{P}_{11} \end{aligned}$$

and without penalization. Tuning factors were estimated by constructing matrices of means squares and cross-products corresponding to the data structure for the population parameters (which are unknown in practice), and maximizing the likelihood of estimates of Σ_G and Σ_E in these ‘validation data’. This was done using a derivative-free search as implemented in routine NEWUOA (Powell 2008), maximizing with respect to $\log \psi_i$ to ensure that estimates were positive. Any estimates exceeding 1,000 were set to this value. For \mathcal{P}_e maximization was performed in two steps by first estimating ψ_3 , considering records for traits 1 to q_1 only, and then (jointly) estimating ψ_1 and ψ_2 for ψ_3 fixed at its estimate from step 1.

Summary statistics. The deviation of estimated covariance matrices ($\hat{\Sigma}$) from their population values (Σ) was evaluated as the entropy loss (L_1) and, with $\bar{L}_1(\cdot)$ denoting the mean over replicates and $\hat{\Sigma}^\psi$ the estimate for a tuning factor of ψ , the percent reduction in average loss (PRIAL)

$$L_1(\Sigma, \hat{\Sigma}) = \text{tr}(\Sigma^{-1} \hat{\Sigma}) - \log |\Sigma^{-1} \hat{\Sigma}| - q \quad \text{PRIAL} = 100 [1 - \bar{L}_1(\Sigma, \hat{\Sigma}^\psi) / \bar{L}_1(\Sigma, \hat{\Sigma}^0)]$$

In addition, the deviation in likelihood from the (unpenalized) maximum ($\Delta \log \mathcal{L}$) was calculated.

RESULTS

Figure 1 shows the distribution of losses in estimates of Σ_G for different values of q_1 for one of the cases examined (C-III for $s_2 = 2000$). Patterns for other constellations were similar. As to be expected, losses in unpenalized estimates decreased substantially as the number of traits (q_1) with

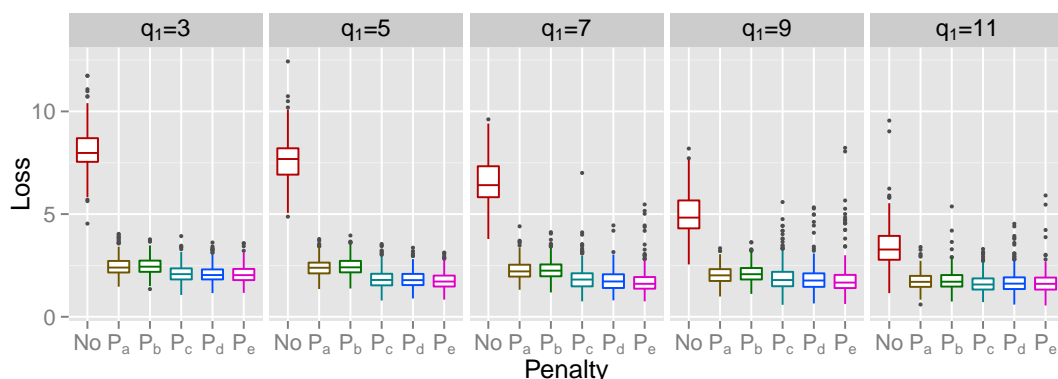


Figure 1. Distribution of loss in estimates of covariance matrices for case C-III ($s_2 = 2000$)

many records increased. Penalization reduced losses in $\hat{\Sigma}_G$ and their spread throughout with relatively small differences between types of penalty, especially for larger values of q_1 .

Means for tuning factors (across population values and q_1), PRIALs and $\Delta \log \mathcal{L}$ (across population values) for both sample sizes are summarized in Figures 2 and 3. With tuning factors obtained by exploiting knowledge of the population values, mean PRIALs were high, especially for small proportions of traits with many records. The number of sire families with records only for the first q_1 traits appeared unimportant until these represented at least half the traits. For $\hat{\Sigma}_G$, differences in mean PRIAL between penalties \mathcal{P}_a and \mathcal{P}_e increased with q_1 , amounting to 13 to 20%. Corresponding values for $\hat{\Sigma}_E$ ranged from 22 to 30% for $s_2 = 400$ and 13 to 25% for $s_2 = 2000$. Whilst only $\hat{\Sigma}_G$ was penalized directly, previous studies found marked associated improvements in $\hat{\Sigma}_E$, due to strong negative sampling correlations (Meyer 2011). For unequal numbers of records, the effect of penalties involving a single tuning factor (\mathcal{P}_a and \mathcal{P}_b) on $\hat{\Sigma}_E$ for low numbers of q_1 were substantially less than those with multiple factors. Again there was comparatively little difference between \mathcal{P}_c , \mathcal{P}_d and \mathcal{P}_e , suggesting that the main benefits were obtained by penalizing submatrices Σ_{22} and Σ_{12} differentially. Higher PRIALs for \mathcal{P}_c , \mathcal{P}_d and \mathcal{P}_e were accompanied by larger changes in likelihood. This was due to much more stringent penalization of block Σ_{12} . Similarly, estimating ψ_3 separately to ψ_1 and ψ_2 (\mathcal{P}_e) resulted in higher estimates of ψ_3 and more improvement in $\hat{\Sigma}_G$ than joint estimation (\mathcal{P}_d), suggesting that the three-dimensional search had some problems.

DISCUSSION

It has been shown that a generalization of the inverse Wishart distribution can be utilised to derive a penalty for penalized REML estimation of covariance components which allows differential shrinkage to be applied to different blocks of the covariance matrices to be estimated. A simulation study has been used to demonstrate that this can improve estimates more than non-differential penalties when there are substantially different numbers of records for different subsets of traits, especially those of residual covariances. However, this requires separate tuning factors to be determined. While not shown here, this can increase the complexity of analysis and computational burden markedly. The differential penalty employed utilizes sequential, conditional distributions of subsets of traits. Results suggest that estimation of tuning factors in an analogous fashion is advantageous.

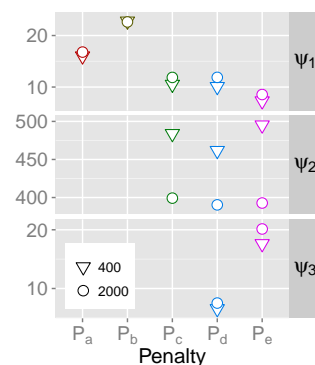


Figure 2. Tuning factors ($s_2 = 400$ and 2000)

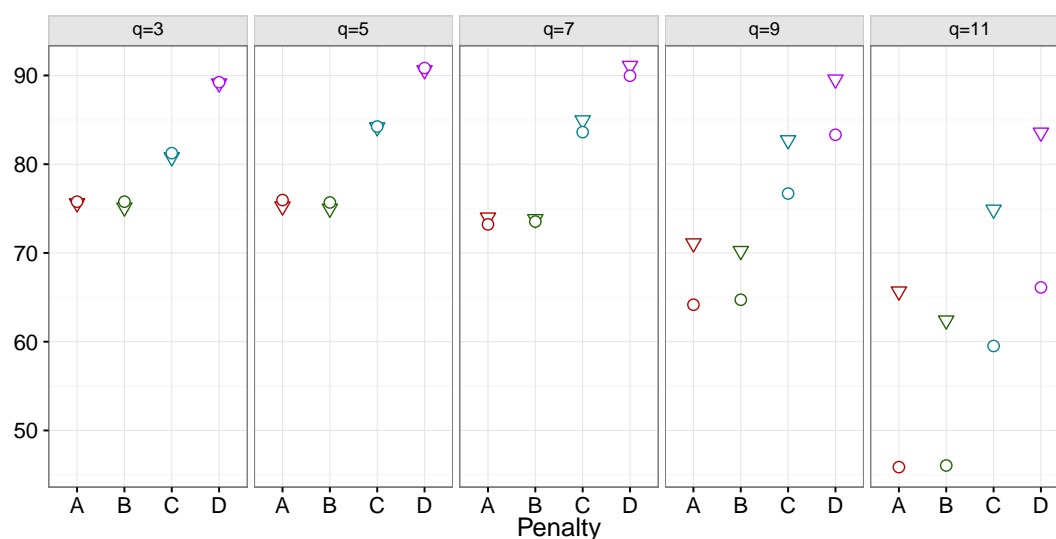


Figure 3. Mean percent reduction in average loss (PRIAL) for estimates of covariance matrices and corresponding change in log likelihood ($\Delta \log \mathcal{L}$) for $s_2 = 400$ (∇) and $s_2 = 2000$ (\circ)

Simulation results given represent a ‘best possible’ scenario as tuning factors were obtained utilizing the population values. Even so, additional improvements in estimates of genetic covariances, over and above those achieved by a simpler, non-differential penalty (\mathcal{P}_a), were moderate. Additional investigations (not shown) indicated that these decreased with the size of the subset of data with records for all traits. Somewhat surprisingly, benefits of penalties \mathcal{P}_c , \mathcal{P}_d and \mathcal{P}_e were most pronounced for the residual covariances. Whether in practice the extra gains possible warrant the additional effort required depends on how well multiple tuning factors can be estimated from data at hand. Future work should address this question. In the meantime, it is reassuring that the simple, non-differential penalty appears to be fairly robust against marked differences in information available for different traits, and can achieve a substantial proportion of the improvements feasible.

CONCLUSIONS

Differential shrinkage of different blocks of covariance matrices to be estimated is feasible, employing a penalty based on the generalised inverse Wishart distribution. However, this requires considerable effort to determine appropriate, multiple tuning factors whilst additional improvements in estimates of genetic covariances achievable appear quite moderate.

ACKNOWLEDGEMENTS

This work was supported by Meat and Livestock Australia under grant B.BFG.0050.

REFERENCES

- Brown P.J. (2006) In *Encyclopedia of Environmetrics*. John Wiley and Sons, Ltd, pp. 1079–1082.
 Meyer K. (2011) *Genet. Sel. Evol.* **43**:39.
 Meyer K., Kirkpatrick M. and Gianola D. (2011) *Proc. Ass. Advan. Anim. Breed. Genet.* **19**:87.
 Munilla S. and Cantet R.J.C. (2012) *J. Anim. Breed. Genet.* **129**:173.
 Powell M. (2008) *IMA J. Numer. Anal.* **28**:649.