

PENALIZED MAXIMUM LIKELIHOOD ESTIMATES OF GENETIC COVARIANCE MATRICES WITH SHRINKAGE TOWARDS PHENOTYPIC DISPERSION

Karin Meyer¹, Mark Kirkpatrick² and Daniel Gianola³

¹Animal Genetics and Breeding Unit*, University of New England, Armidale, NSW 2351

²Section of Integrative Biology, University of Texas, Austin, TX 78712

³University of Wisconsin-Madison, Madison, WI 53706

SUMMARY

A simulation study examining the effects of ‘regularization’ on estimates of genetic covariance matrices for small samples is presented. This is achieved by penalizing the likelihood, and three types of penalties are examined. It is shown that regularized estimation can substantially enhance the accuracy of estimates of genetic parameters. Penalties shrinking estimates of genetic covariances or correlations towards their phenotypic counterparts acted somewhat differently to those aimed reducing the spread of sample eigenvalues. While improvements of estimates were found to be comparable overall, shrinkage of genetic towards phenotypic correlations resulted in least bias.

INTRODUCTION

Estimates of genetic covariance matrices are inherently subject to substantial sampling variation, especially if more than just a few traits are considered and if sample sizes are small. Recently, there has been increasing interest in ‘regularized’ estimation to reduce sampling variation and thus mean square error, albeit usually at the expense of some additional bias. In quantitative genetic analyses covariances between traits are partitioned into their genetic (Σ_G) and environmental (Σ_E) components. Typically, this results in strong sampling correlations between corresponding estimates, so that their sum – the phenotypic covariance matrix Σ_P – is estimated much more accurately than Σ_G . This has led to suggestions to borrow strength from $\hat{\Sigma}_P$ in estimating Σ_G . A specific proposal, referred to as ‘bending’, has been to regress the eigenvalues of $\hat{\Sigma}_P^{-1}\hat{\Sigma}_G$ (λ_i) towards their mean (Hayes and Hill 1981). In a maximum likelihood (ML) framework, this can be achieved by penalizing the likelihood by a term proportional to the variance of the estimates of λ_i (Meyer and Kirkpatrick 2010).

A penalty to the likelihood can be derived from a Bayesian prior probability with the penalty proportional to minus the logarithmic value of the prior’s density. A quadratic penalty on the λ_i thus implies a prior that assumes the λ_i are normally distributed. A standard prior used in Bayesian estimation of covariance matrices is an Inverse Wishart (IW) distribution. This paper examines the scope for improved estimation of Σ_G via ML using penalties derived invoking such assumption.

MATERIAL AND METHODS

Penalties. Let $\log \mathcal{L}(\theta)$ denote the (unpenalized) log likelihood for a given model of analysis with vector of parameters θ . For a penalty $\mathcal{P}(\theta)$, the penalized likelihood is $\log \mathcal{L}_P(\theta) = \log \mathcal{L}(\theta) - \frac{1}{2} \psi \mathcal{P}(\theta)$, where ψ is a tuning factor which determines the amount of regularization to be applied. We consider:

i. A quadratic penalty on the deviation of the canonical eigenvalues (log scale) from their mean

$$\mathcal{P}_\lambda(\theta) \propto \text{tr}(\mathbf{\Lambda} - \bar{\lambda}\mathbf{I})^2 \quad \text{with } \mathbf{\Lambda} = \text{Diag}\{\log(\hat{\lambda}_i)\} \quad \text{and } \bar{\lambda} = \text{tr}(\mathbf{\Lambda})/q \quad (1)$$

ii. A penalty on the genetic covariance matrix (with $\tilde{\Sigma}_P^0$ the estimate of Σ_P for $\psi=0$)

$$\mathcal{P}_\Sigma(\theta) \propto C \log |\hat{\Sigma}_G| + \text{tr}(\hat{\Sigma}_G^{-1}\tilde{\Sigma}_P^0) \quad (2)$$

iii. A penalty on the genetic correlation matrix \mathbf{R}_G (with $\tilde{\mathbf{R}}_P^0$ the estimate of \mathbf{R}_P for $\psi=0$)

$$\mathcal{P}_R(\theta) \propto C \log |\hat{\mathbf{R}}_G| + \text{tr}(\hat{\mathbf{R}}_G^{-1}\tilde{\mathbf{R}}_P^0) \quad (3)$$

*AGBU is a joint venture of NSW Department of Industry and Investment and the University of New England

where q denotes the number of traits. Using unpenalized estimates of Σ_P and the phenotypic correlation matrix, \mathbf{R}_P , for the scale parameter in the IW prior, penalties $\mathcal{P}_\Sigma(\theta)$ and $\mathcal{P}_R(\theta)$ imply an empirical Bayes procedure which shrinks estimates of Σ_G and \mathbf{R}_G towards their phenotypic counterparts. The IW prior gives $C=(\psi + q + 1)/\psi$. Approximating C with unity yields penalties proportional to the Kullback-Leibler divergence between the genetic and phenotypic matrices.

Data. A simulation study was carried out for a paternal half-sib design, considering $q=5$ traits recorded on each of $n=10$ progeny of $s=100$ unrelated sires. Population parameters were obtained by combining 12 sets of heritabilities (A to L; see Table 1) with 5 scenarios for genetic (r_G) and residual (r_E) correlations (S1 to S5). This resulted in 60 different cases, labelled as 1A to 5L in the following. For S1, $r_{Gij}=r_{Eij}=0$

Table 1. Population heritability values ($\times 100$)

A	B	C	D	E	F	G	H	I	J	K	L
40	50	60	70	90	70	80	90	20	30	50	60
40	45	50	55	50	70	30	30	20	25	20	10
40	40	40	40	30	40	30	10	20	20	15	10
40	35	30	25	20	10	30	10	20	15	10	10
40	30	20	10	10	10	30	10	20	10	5	10

for all $i \neq j$, so that canonical eigenvalues were equal to the heritabilities. In addition all phenotypic variances were assumed to be equal, $\sigma_{P_i}^2=1$ for $i=1, q$. For S2, $r_{Gij}=0.8$ and $r_{Eij}=0$, with $\sigma_{P_1}^2=1$, $\sigma_{P_2}^2=1.5$, $\sigma_{P_3}^2=2.25$, $\sigma_{P_4}^2=3.375$ and $\sigma_{P_5}^2=5.065$. For S3 and S4, correlations were assumed to follow an approximately auto-regressive structure, i.e. $r_{Gij}=0.6^{|i-j|}$ for S3 and $r_{Gij}=0.02i + (-0.8)^{|i-j|}$ for S4, with $r_{Eij}=0.5+(-0.4)^{|i-j|}$ for both ($i \neq j$). Finally, for S5 correlations were $r_{Gij}=0.5+(-1)^i 0.05j$ and $r_{Eij}=0.2+(-1)^j 0.1i$. Phenotypic variances for S3 to S5 were $\sigma_{P_1}^2=\sigma_{P_5}^2=3$, $\sigma_{P_2}^2=\sigma_{P_4}^2=2$ and $\sigma_{P_3}^2=1$. Data were generated by sampling matrices of crossproducts between and within sires from appropriate Wishart distributions. A total of 1000 replicates per case were carried out.

Analyses. Restricted ML (REML) estimates of Σ_G and Σ_E were obtained using a combination of Method of Scoring and simple derivative-free algorithms to locate the maximum of $\log \mathcal{L}_P(\theta)$. To determine the ‘optimal’ tuning factor ($\hat{\psi}$) for each analysis, estimates $\hat{\Sigma}_G^\psi$ and $\hat{\Sigma}_E^\psi$ were obtained for a range of values of ψ : 0 to 2 in steps of 0.1, 2.2 to 5 in steps of 0.2, 5.5 to 10 in steps of 0.5, 11 to 100 in steps of 1, 102 to 250 in steps of 2, 255 to 500 in steps of 5 and 510 to 1000 in steps of 10, 311 in total. For each ψ the unpenalized log likelihood was then calculated as $\log \mathcal{L}(\theta)^\psi = -[(s-1)(\log |\Sigma_B| + \text{tr}(\Sigma_B^{-1} \mathbf{M}_B)) + s(n-1)(\log |\Sigma_W| + \text{tr}(\Sigma_W^{-1} \mathbf{M}_W))]/2$ with $\Sigma_W = \hat{\Sigma}_E^\psi + \frac{3}{4} \hat{\Sigma}_G^\psi$ and $\Sigma_B = \Sigma_W + \frac{1}{4} n \hat{\Sigma}_G^\psi$. The validation ‘data’ used for this, i.e. the corresponding matrices of mean squares \mathbf{M}_W and \mathbf{M}_B , were not sampled but simply constructed using the population parameters. This can be thought of as equivalent to sampling an infinite number of additional data sets for the same data structure. The value of ψ which maximised $\log \mathcal{L}(\theta)^\psi$ was then chosen as $\hat{\psi}$.

Summary statistics. The percentage reduction in average loss (PRIAL) of a covariance matrix is

$$\text{PRIAL} = 100 [\bar{L}_1(\Sigma_X, \hat{\Sigma}_X^0) - \bar{L}_1(\Sigma_X, \hat{\Sigma}_X^{\hat{\psi}})] / \bar{L}_1(\Sigma_X, \hat{\Sigma}_X^0) \quad (4)$$

with $\hat{\Sigma}_X^0$ and $\hat{\Sigma}_X^{\hat{\psi}}$ the unpenalized and penalized estimates of Σ_X , respectively, and $\bar{L}_1(\cdot)$ the entropy loss, $L_1(\Sigma, \hat{\Sigma}) = \text{tr}(\Sigma^{-1} \hat{\Sigma}) - \log |\Sigma^{-1} \hat{\Sigma}| - q$, averaged over replicates. In addition, the relative bias (in %) for parameter θ_i is calculated as $100(\hat{\theta}_i - \theta_i)/\theta_i$.

RESULTS

Mean PRIAL values across the 60 cases examined are summarized in Table 2. On average, the reduction in loss for $\hat{\Sigma}_G$ was about 70%, with little difference between the types of penalties employed. However, as shown in Figure 1 there were substantial differences in individual cases. As noted by Meyer and Kirkpatrick (2010), penalty $\mathcal{P}_\lambda(\theta)$ performed best when the

Table 2. Mean PRIAL

Penalty	$\hat{\Sigma}_G$	$\hat{\Sigma}_E$	$\hat{\Sigma}_P$
$\mathcal{P}_\lambda(\theta)$	71.3	43.4	1.2
$\mathcal{P}_\Sigma(\theta)$	70.6	13.3	1.2
$\mathcal{P}_R(\theta)$	72.0	37.3	2.2

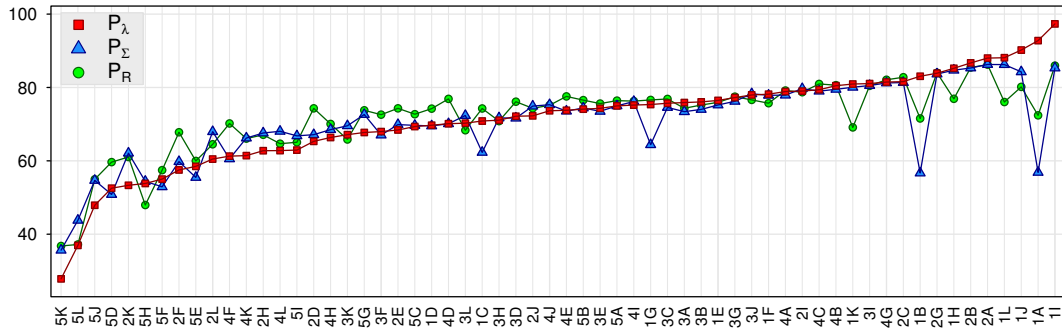


Figure 1. PRIAL for estimates of Σ_G (ordered by values for penalty $\mathcal{P}_\lambda(\theta)$).

population canonical eigenvalues where close together, but tended to over-shrink sample eigenvalues when they were spread apart. $\mathcal{P}_\Sigma(\theta)$ yielded substantially less improvements than the other penalties for cases with similar λ_i , in particular 1A, 1B, 1C, 1G, 1I and 1J. With some exceptions, $\mathcal{P}_\Sigma(\theta)$ and $\mathcal{P}_R(\theta)$ tended to out-perform $\mathcal{P}_\lambda(\theta)$ for cases with a substantial spread of the population λ_i . As the canonical eigenvalues are a function of both $\hat{\Sigma}_G$ and $\hat{\Sigma}_E$, penalty $\mathcal{P}_\lambda(\theta)$ resulted in a substantial improvement in $\hat{\Sigma}_E$ while $\mathcal{P}_\Sigma(\theta)$ had only a modest effect on $\hat{\Sigma}_E$. Somewhat surprisingly, $\mathcal{P}_R(\theta)$ decreased loss in $\hat{\Sigma}_E$ by almost as much as $\mathcal{P}_\lambda(\theta)$. As to be expected from the nature of penalties imposed, estimates of Σ_P were little affected by penalized estimation.

Bias. The mean relative bias in estimates of individual canonical eigenvalues, genetic variances (σ_{Gi}^2) and heritabilities (h_i^2) is given in Table 3. As expected from theory, unpenalized estimation resulted in systematic overestimates of the largest and underestimates of the smallest λ_i . While all three penalties alleviated this bias, they acted in a different fashion. This is illustrated in Figure 2 for case 1K. With most of the 60 cases examined representing scenarios with a substantial spread of population λ_i , $\mathcal{P}_\lambda(\theta)$ resulted on average in over-shrinkage. On the relative scale this was most pronounced for λ_5 , for which half the population values were less than 0.05. Penalty $\mathcal{P}_\Sigma(\theta)$ predominantly affected the estimates of the smallest λ_i . Whilst $\mathcal{P}_R(\theta)$ also over-shrank the smallest λ_i , this was less pronounced than for the other penalties and estimates of the largest, most important values were least biased.

It has to be emphasized that standard, unpenalized REML estimates are biased, as estimates are constrained to the parameter space. This is most evident in the upward bias in estimates of the lowest heritability, \hat{h}_5^2 , and a small downwards bias in the largest value, \hat{h}_1^2 . Shrinking canonical eigenvalues towards their mean exacerbated these biases. Penalty $\mathcal{P}_\Sigma(\theta)$ affected the lower heritabilities in a similar way to $\mathcal{P}_\lambda(\theta)$ but tended to exaggerate estimates of the higher values. Again, $\mathcal{P}_R(\theta)$ resulted in the least bias in the penalized estimates. As penalized estimation had negligible effects on estimates of the phenotypic components, the pattern of relative bias in estimates of genetic variances closely followed that for the corresponding heritabilities.

Similarly, standard estimates of genetic correlations (r_G) can be biased. Figure 3 shows the mean estimate of r_G between traits 4 and 5 for scenario S2. The population value is 0.8, shown by the top line. With a corresponding population value for the residual correlation of zero, the phenotypic correlation (r_P , shown by the bottom line) ranges from 0.3 to

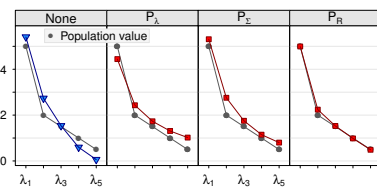


Figure 2. Mean estimates (▼ ■) of canonical eigenvalues for case 1K

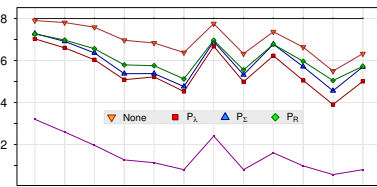


Figure 3. Mean estimates of r_G 45

Table 3. Mean bias (in %; $\hat{\lambda}_i$ canonical eigenvalue, $\hat{\sigma}_{G_i}^2$ genetic variance, \hat{h}_i^2 heritability)

Penalty	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_3$	$\hat{\lambda}_4$	$\hat{\lambda}_5$	$\hat{\sigma}_{G1}^2$	$\hat{\sigma}_{G2}^2$	$\hat{\sigma}_{G3}^2$	$\hat{\sigma}_{G4}^2$	$\hat{\sigma}_{G5}^2$	\hat{h}_1^2	\hat{h}_2^2	\hat{h}_3^2	\hat{h}_4^2	\hat{h}_5^2
None	9.4	26.5	16.7	-19.5	-78.8	-0.9	4.1	4.7	7.3	12.5	-1.1	3.8	4.5	7.2	12.3
$\mathcal{P}_\lambda(\theta)$	-3.7	16.3	28.8	57.7	101.4	-7.0	4.6	11.4	23.5	45.3	-6.5	4.6	11.5	23.4	44.9
$\mathcal{P}_\Sigma(\theta)$	8.1	24.9	24.7	39.1	75.3	0.8	10.4	15.7	26.1	45.1	0.7	10.0	15.4	25.6	44.3
$\mathcal{P}_R(\theta)$	1.3	16.2	20.8	37.3	57.2	-2.3	2.1	4.8	8.6	17.2	-2.1	2.2	4.9	8.8	17.2

0.06. Unpenalized estimates of r_G were the more subject to sampling variation and thus the more biased, the lower the corresponding heritabilities. All three penalties shrunk \hat{r}_G towards \hat{r}_P , with $\mathcal{P}_\lambda(\theta)$ resulting in most additional bias. For this scenario, estimates using $\mathcal{P}_R(\theta)$ were consistently closer to the population values than those from $\mathcal{P}_\Sigma(\theta)$, but for other constellations of correlations differences were less clear cut. Across all 10 correlations amongst the 5 traits and all 60 cases, mean deviations of estimates \hat{r}_G from their population values were -0.019 , -0.046 , -0.039 and -0.039 for standard estimates and estimates employing penalties $\mathcal{P}_\lambda(\theta)$, $\mathcal{P}_\Sigma(\theta)$ and $\mathcal{P}_R(\theta)$, respectively.

DISCUSSION

Results show that regularized estimation of genetic covariances matrices can result in estimates which, on average, have greatly reduced loss, i.e. are closer to the population values and have lower mean square errors than standard, unpenalized estimates. This can be achieved by penalizing the likelihood function with penalties aimed at reducing the spread of sample eigenvalues or at shrinking genetic covariance and correlation matrices towards their phenotypic counterparts. While a penalty targeting eigenvalues worked best when population eigenvalues were similar, this is a scenario not likely to be encountered very often in practical applications. Overall, penalty $\mathcal{P}_R(\theta)$ performed best with a slightly higher PRIAL for $\hat{\Sigma}_G$ than the other penalties and somewhat lower biases arising from penalization. This penalty ‘works’ by making estimates of r_G similar to those for r_P and thus reducing sampling variation. Interestingly, this can be interpreted as a modern and flexible adaptation of the suggestion, due to Cheverud (1988), to substitute estimates of r_P for r_G when the latter can not be determined reliably.

Simulation results presented used knowledge of the population values to select the tuning parameter ψ and should thus be viewed as ‘best possible’. Appropriate choice of ψ presents the main challenge for practical use of penalized ML estimation. Suitable techniques are cross-validation and strategies limiting the change in likelihood values. While we need to expect a reduction in efficacy when the tuning parameter is estimated with error, initial simulation results (Meyer 2011) indicate that mild penalization can improve estimates of genetic parameters for most multivariate analyses where sample sizes are limited.

CONCLUSIONS

Regularized estimation of genetic parameters can result in ‘better’ estimates by reducing sampling variation. In a maximum likelihood framework (using either full ML or REML), this is readily implemented by penalizing the likelihood function. A penalty encouraging shrinkage of genetic towards phenotypic correlations appears especially suited to ‘borrowing strength’. It is an appealing strategy to make the most of limited and often precious data which is currently under-utilized.

REFERENCES

- Cheverud J.M. (1988) *Evolution* **42**:958.
 Hayes J.F. and Hill W.G. (1981) *Biometrics* **37**:483.
 Meyer K. (2011) *Proc. Ass. Advan. Anim. Breed. Genet.* **19**:83.
 Meyer K. and Kirkpatrick M. (2010) *Genetics* **185**:1097.