

## MILDLY PENALIZED MAXIMUM LIKELIHOOD ESTIMATION OF GENETIC COVARIANCES MATRICES WITHOUT TUNING

Karin Meyer

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

### SUMMARY

A scheme for penalized estimation of genetic covariance matrices free from tuning – using default settings for the strength or penalization – is described and its efficacy is demonstrated by simulation.

### INTRODUCTION

Estimates of genetic covariance matrices,  $\Sigma_G$ , are known to be afflicted by substantial sampling errors, increasing markedly with the number of traits considered. ‘Regularization’, i.e. modification of estimators to reduce sampling variation at the expense of a small, additional bias, has been advocated to obtain estimates closer to the population values. An early suggestion by Hayes and Hill (1981, ‘bending’) has been to shrink the canonical eigenvalues,  $\lambda_i$ , i.e. the eigenvalues of  $\Sigma_P^{-1}\Sigma_G$  (with  $\Sigma_P$  the phenotypic covariance matrix), towards their mean. As shown by Meyer and Kirkpatrick (2010), the analogue in a maximum likelihood framework is to maximize the likelihood subject to a penalty proportional to the variance among the estimates of  $\lambda_i$ . Neither authors provided guidelines on how to determine the amount of shrinkage to be applied. While cross-validation techniques allow estimation of so-called ‘tuning factors’, this proved laborious and only moderately successful (Meyer 2011).

A simple alternative is to apply a mild, default penalty which, while not providing maximum benefits, will yield stable estimates and worthwhile reductions in ‘loss’, i.e. the average deviations of estimates from population values. This is similar to the concept of weakly informative priors, which is gaining popularity in Bayesian estimation (e.g. Gelman 2006). This paper demonstrates the reductions in loss achievable using a default penalty on canonical eigenvalues.

### PRIORS AND PENALTIES

For a given prior distribution of some function of the parameters to be estimated, we can obtain a corresponding penalty as minus the logarithmic value of the pertaining probability density.

Shrinking canonical eigenvalues towards their mean,  $\bar{\lambda}$ , by applying a quadratic penalty,  $\mathcal{P} \propto \sum_i (\lambda_i - \bar{\lambda})^2$ , implies a Normal distribution,  $N(\bar{\lambda}, \sigma^2)$ , with  $\sigma^2$  the variance of  $\lambda_i$ . This gives penalty

$$\mathcal{P}_N = \frac{1}{2} \left[ q(\log(\sigma^2) + \log(2\pi)) + \frac{1}{\sigma^2} \sum_{i=1}^q (\lambda_i - \bar{\lambda})^2 \right] \quad (1)$$

with  $q$  denoting the number of traits. Similarly, assuming a log-Normal distribution, the penalty is obtained by substituting  $\log \lambda_i$  and  $(\sum_i \log \lambda_i)/q$  for  $\lambda_i$  and  $\bar{\lambda}$  in (Eq. 1). Earlier results showed that for such a prior it was advantageous to penalize both  $\log \lambda_i$  and  $\log(1 - \lambda_i)$  (Meyer 2011). We use  $\mathcal{P}_L$  to denote the penalty obtained by summing contributions for both, with the same variance  $\sigma^2$ .

A more flexible alternative is a Beta distribution,  $B(\alpha, \beta)$ , with scale parameters  $\alpha$  and  $\beta$  allowing for a wide range of shapes. For  $\alpha, \beta > 1$ , the distribution is unimodal. In Bayesian estimation,  $\nu = \alpha + \beta$  is interpreted as effective sample size, i.e. the number of ‘observations’ added by the prior. A Beta distribution with mode equal to  $\bar{\lambda}$  can be specified as  $\alpha = 1 + (\nu - 2)\bar{\lambda}$  and  $\beta = 1 + (\nu - 2)(1 - \bar{\lambda})$ , for  $\nu > 2$ . This allows us to quantify the degree of belief in the prior through the single parameter  $\nu$ . For  $m = \nu - 2$  and  $\Gamma(\cdot)$  denoting the Gamma function, the penalty for the Beta distribution is

$$\mathcal{P}_\beta = q \left[ \log \Gamma(\nu) - \log \Gamma(1 + m\bar{\lambda}) - \log \Gamma(1 + m(1 - \bar{\lambda})) \right] + m \left[ \bar{\lambda} \sum_{i=1}^q \log \lambda_i + (1 - \bar{\lambda}) \sum_{i=1}^q \log(1 - \lambda_i) \right] \quad (2)$$

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

As for  $\mathcal{P}_L$ ,  $\mathcal{P}_\beta$  involves functions of  $\log \lambda_i$  and  $\log(1 - \lambda_i)$ . The strength of penalization for all three penalties is regulated by the parameter  $\sigma^2$  or  $\nu$ . In contrast to previous formulations employing a tuning factor, this lends itself to attempts of direct estimation by maximizing the penalized likelihood with respect to this parameter (de los Campos 2013; pers. comm.).

### SIMULATION STUDY

Records for  $q = 9$  traits were sampled from multivariate normal distributions, assuming a balanced paternal half-sib design comprised of  $s = 100, 400$  or  $1000$  sire families of size 10. Population values for 72 scenarios, selected to represent an extensive range of possible – including unusual or ‘difficult’ – cases were obtained by combining 12 sets of heritabilities with 6 correlation structures. The variance among population values for canonical eigenvalues ranged from 0 to 0.099 (mean 0.046) on the original and 0 to 2.504 (mean 0.989) on the logarithmic scale. Restricted maximum likelihood estimates of genetic and residual ( $\Sigma_E$ ) covariance matrices were obtained fitting a simple animal model with means as the only fixed effects, for the three types of penalties described above. Penalties were applied using the same, default ‘strength parameter’ for all cases,  $\sigma^2 = 0.02$  to  $0.1$  for  $\mathcal{P}_N$ ,  $\sigma^2 = 0.5$  to  $2.0$  for  $\mathcal{P}_L$  and  $\nu = 2.5$  to  $10$  for  $\mathcal{P}_\beta$ . In addition,  $\sigma^2$  or  $\nu$  were estimated from the data by evaluating points on the profile likelihood and employing a quadratic approximation to determine its maximum. In doing so, parameter estimates were constrained to the interval  $[2.001, 50]$  for  $\nu$  and  $[0.001, 10]$  and  $[0.01, 25]$  for  $\sigma^2$  for  $\mathcal{P}_N$  and  $\mathcal{P}_L$ , respectively. A total of 500 replicates were carried out for each case. For each sample, the loss in estimates was determined as (for  $X = G, E$  and  $P$ )

$$L_1(\Sigma_X, \hat{\Sigma}_X) = \text{tr}(\Sigma_X^{-1} \hat{\Sigma}_X) - \log |\Sigma_X^{-1} \hat{\Sigma}_X| - q \quad (3)$$

with  $\Sigma_X$  the matrix of population values,  $\hat{\Sigma}_X$  the corresponding estimate, and  $\Sigma_P = \Sigma_G + \Sigma_E$ . The **Percentage Reduction In Average Loss** due to penalization was then evaluated as

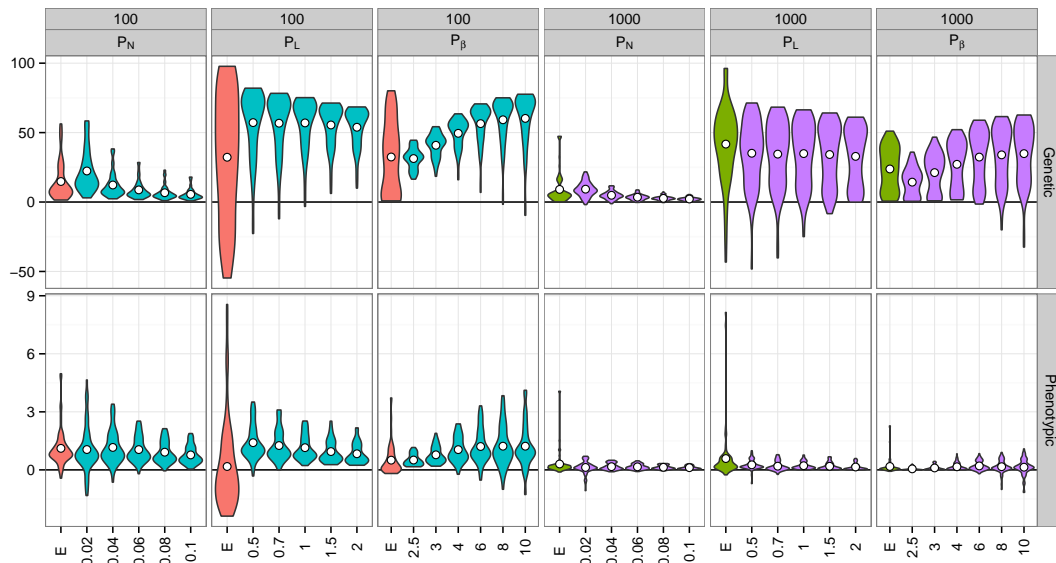
$$\text{PRIAL} = 100 [1 - \bar{L}_1(\Sigma_X, \hat{\Sigma}_X^\nu) / \bar{L}_1(\Sigma_X, \hat{\Sigma}_X^0)] \quad (4)$$

with  $\hat{\Sigma}_X^\nu$  and  $\hat{\Sigma}_X^0$  the penalized and unpenalized estimates of  $\Sigma_X$ , and  $\bar{L}_1(\cdot)$  the average loss over replicates. In addition, the mean reduction in unpenalized likelihood due to penalization (from its maximum for unpenalized estimates),  $\Delta\mathcal{L}$ , was calculated.

### RESULTS

Our main goal of penalized estimation is to reduce the loss in estimates of  $\Sigma_G$ . The rationale for shrinking canonical eigenvalues towards their mean or mode is that this reduces sampling variation by ‘borrowing strength’ from the estimate of  $\Sigma_P$ , which typically is estimated much more accurately than either of its components,  $\Sigma_G$  and  $\Sigma_E$  (Hayes and Hill 1981). Hence, loosely speaking, we attempt to redress the balance in partitioning skewed by sampling error. This implies that we expect the estimate of  $\Sigma_P$  to remain more or less unchanged. Too stringent penalization can result in reduced or even negative PRIAL for any of the covariance matrices estimated. In particular, a negative PRIAL for  $\Sigma_P$  represents a strong warning signal for over-penalization.

The distribution of PRIALs (higher values are better) across the 72 scenarios for  $\Sigma_G$  and  $\Sigma_P$  for two sample sizes is summarized in Figure 1. Central dots display mean values. Values on the x-axis are the fixed values for  $\sigma^2$  and  $\nu$  used, except for ‘E’ which denotes use of the estimated value. For clarity of scale, 5 values for  $\Sigma_G$  less than  $-60$ , occurring for  $\mathcal{P}_L$ , ‘E’ and  $s = 100$ , are omitted. PRIALs for  $\mathcal{P}_N$  were modest, but positive throughout for  $\Sigma_G$ . Except for the smallest, fixed value for  $\sigma^2$ , i.e. the most stringent penalty, there was little evidence for adverse effects on estimates for any of the cases. Estimating  $\sigma^2$  yielded marked improvements, especially for the larger sample size, for cases with low variance among the population values for  $\lambda_i$ , i.e. the cases which matched the prior. As there were few of the latter, however, mean PRIALs achieved remained quite low.



**Figure 1. Distribution of Percentage reduction in average loss for genetic and phenotypic covariance matrices (for  $s=100$  and  $1000$  sire families).**

As reported previously (Meyer 2011), penalizing  $\log \lambda_i$  yielded substantially larger PRIALs. Using a fixed value of  $\sigma^2 = 2$  proved to be a safe default for a mild penalty with mean PRIALs for  $\Sigma_G$  as high as 54, 44 and 33% for  $s = 100, 400$  and  $1000$ , respectively. Lower values for  $\sigma^2$  resulted in increasing numbers of unfavourable cases. Attempts to estimate  $\sigma^2$  for  $\mathcal{P}_L$  failed in a substantial proportion of replicates for a number of cases, with estimates close to the lower boundary. As this was set at 0.01, it resulted in far too stringent penalization. This held in particular for the smallest sample size, suggesting that this was, in part at least, attributable to insufficient information. Additional analyses (not shown) estimating separate values of  $\sigma^2$  for two the parts of  $\mathcal{P}_L$ , involving  $\log \lambda_i$  and  $\log(1 - \lambda_i)$ , respectively, reduced the incidence of problem cases, but, on the whole, was not satisfactory either.

A similar pattern emerged for a penalty based on the Beta distribution. However, for  $\mathcal{P}_\beta$  an estimate for  $\nu$  close to its lower boundary at 2.001 was equivalent to virtually no penalization. Hence, there were no negative PRIALs due to over-penalization. Yet, overall there was little advantage in estimating  $\nu$  compared to a default value for a mild penalty. Means and minimum values for PRIALs and the corresponding decrease in likelihood (from it's unpenalized maximum) for selected values of  $\nu$  are given in Table 1. Results for fixed  $\nu$  identified little adverse effects for any cases or sample sizes for values up to about 6. Average changes in likelihood were small, especially when considering that for  $q = 9$  traits there were 90 covariance components to be estimated. As shown in Table 1, repeating analyses with minimum values for  $\nu$  of 2.5 and 4 increased PRIALs by a few percent compared to corresponding results for a fixed  $\nu$ , but at the price of marked additional effort.

## DISCUSSION

Penalized estimation provides a powerful mechanism to improve estimates of genetic covariance matrices by reducing sampling variation. Large improvements can be obtained if population parameters approximately match the assumed prior distribution on which the penalty is based. In practice, however, true values are unknown and it is important that the procedure chosen is robust, i.e. unlikely to result in worse estimates. While estimation of the strength of penalization is possible in principle, it is computational demanding and may not be particularly advantageous or even successful.

**Table 1. Mean and minimum PRIAL for estimates of genetic ( $\Sigma_G$ ), residual ( $\Sigma_E$ ) and phenotypic ( $\Sigma_P$ ) covariance matrices and change in log likelihood ( $\Delta\mathcal{L}$ ) for penalty  $\mathcal{P}_\beta$** 

$\nu^a$	100 sires				400 sires				1000 sires			
	$\Sigma_G$	$\Sigma_E$	$\Sigma_P$	$\Delta\mathcal{L}$	$\Sigma_G$	$\Sigma_E$	$\Sigma_P$	$\Delta\mathcal{L}$	$\Sigma_G$	$\Sigma_E$	$\Sigma_P$	$\Delta\mathcal{L}$
	Mean											
F2.5	31.3	36.4	0.5	-0.27	21.9	15.6	0.1	-0.08	14.3	7.6	0.1	-0.03
F4.0	49.5	48.2	1.0	-1.25	37.5	23.3	0.4	-0.47	27.2	12.6	0.2	-0.22
F6.0	56.5	51.5	1.2	-2.52	43.6	26.3	0.5	-1.06	32.4	14.7	0.2	-0.55
F8.0	59.2	52.6	1.2	-3.72	45.7	25.5	0.5	-1.68	34.0	12.7	0.2	-0.92
E2.0	32.4	16.7	0.5	-1.25	31.9	11.8	0.4	-0.56	23.8	7.5	0.2	-0.27
E2.5	46.2	46.5	0.9	-1.35	38.8	24.2	0.5	-0.60	28.2	13.1	0.2	-0.29
E4.0	54.2	52.7	1.2	-2.01	46.7	29.3	0.6	-0.84	33.1	16.3	0.3	-0.41
	Minimum											
F2.5	16.2	8.7	0.2	-0.47	0.9	1.3	0.0	-0.23	0.2	0.4	0.0	-0.11
F4.0	16.0	22.5	-0.2	-1.98	4.0	5.0	0.0	-1.13	1.4	1.7	0.0	-0.69
F6.0	6.7	23.4	-0.5	-3.91	-0.4	9.4	-0.6	-2.41	-2.0	3.4	-0.2	-1.69
F8.0	-1.7	8.4	-1.0	-5.69	-17.1	13.2	-0.9	-3.72	-20.2	-1.3	-1.0	-2.67
E2.0	0.3	-3.1	-0.2	-5.21	0.5	-0.7	-0.1	-1.96	0.5	-0.2	-0.1	-1.18
E2.5	18.1	26.5	0.0	-4.99	14.5	7.1	-0.0	-2.01	3.4	3.0	-0.0	-1.17
E4.0	16.2	32.5	0.0	-5.38	15.8	9.8	0.1	-2.02	5.1	3.9	-0.1	-1.19

<sup>a</sup>Effective size, F: fixed value, E: estimated with this minimum value

Results show that a penalty encouraging shrinkage of canonical eigenvalue lends itself to a scheme using a default strength parameter to impose a mild penalty. Assuming a Beta distribution provides a more flexible prior than a Normal or log-Normal distribution. Moreover, the resulting penalty has an intuitive parameter – the so-called effective sample size – regulating its stringency. A value of  $\nu = 4$  to 6 yielded worthwhile reductions in loss without (non-negligible) negative PRIALs or substantial changes in likelihood and can be recommended for routine use. Additional computational requirements are small, but derivatives of the penalty may be needed. These are readily obtained by parameterising analyses to canonical eigenvalues and elements of the corresponding eigenvectors (Meyer and Kirkpatrick 2010), but this may have less favourable convergence rates than the standard parameterisation. Alternative penalties, e.g. to shrink genetic correlations towards their phenotypic counterparts may be preferable in this respect, and equally suited to default penalties (Meyer 2014).

## CONCLUSIONS

Maximum likelihood estimation subject to a penalty can markedly reduce sampling variation of estimates, and should be applied routinely in multivariate analyses involving more than a few traits.

## ACKNOWLEDGEMENTS

Work was supported by Meat and Livestock Australia grant B.BFG.0050

## REFERENCES

- Gelman A. (2006) *Bayesian analysis* **1**:515.  
 Hayes J.F. and Hill W.G. (1981) *Biometrics* **37**:483.  
 Meyer K. (2011) *Genet. Sel. Evol.* **43**:39.  
 Meyer K. (2014) In *Proc. Tenth World Congr. Genet. Appl. Livest. Prod.* Paper No. 217.  
 Meyer K. and Kirkpatrick M. (2010) *Genetics* **185**:1097.