Post-Estimation Penalization: more ‘pep’ for estimates of genetic covariance matrices

Karin Meyer

Animal Genetics and Breeding Unit
University of New England
Armidale, Australia

20th Conference of the Association for Animal Breeding and Genetics, Napier, NZ, October 20-23, 2013

Objectives
- Propose new procedure to modify estimates of covariance matrices
- Demonstrate efficacy
  → results closer to population values

Introduction
- Estimates of genetic covariance matrices are subject to substantial sampling variation
- Penalty on likelihood in multivariate REML
  - reduces sampling errors
  - yields ‘better’ estimates
- Drawback: need multiple multivariate analyses
- Alternative: 2-step procedure
  - impose penalty afterwards in second step

The PEP Method
a) Treat standard, multivariate estimates \( S_x \) as matrices of mean squares/cross products
b) Feed into likelihood assuming simple, pseudo pedigree structure
c) Maximise likelihood subject to penalty

\[
\log L \propto \left[ \log |V| + \text{tr}(V^{-1}M) \right] - \psi \Psi \Psi
\]

\[
M = \sum \Sigma \otimes \Sigma \Psi\Psi\Psi
\]

\[
V = \sum \Sigma \otimes \Sigma \Psi\Psi\Psi
\]

\[
\text{E}(|S_x|) \Psi\Psi\Psi
\]

\[
\text{E}([\text{Cov}(\text{Pseudo-relatives})]) \Psi\Psi\Psi
\]

Results
- Optimum \( \psi \) (for population values known)
  - Penalty reduces average loss by \( \sim 60 - 70\% \)
  - 2-step method (PEP) almost as effective as penalised estimation (MVP)
- Pick \( \psi \) so that \( \Delta L \approx -\frac{1}{2} \chi^2_{5\%},1 \)
  - change in \( \log L \) not significant
  - under-penalisation, but achieve substantial proportion of reduction in loss possible
  - results for PEP more variable than for MVP
- Little difference between types of penalty

Simulation
- 10 traits, 2000 records each, 250 replicates
- Pseudo pedigree: Paternal half-sibs, \( s = 2, n = 2 \)
- Compare estimates
  - \( MV0 \): Standard multivariate REML analysis
  - \( MVP \): Multivariate analysis subject to penalty
  - \( PEP \): \( MV0 \) + separate ‘pseudo’-REML penalty
- \( P \): Penalties to encourage shrinkage of
  i) canonical eigenvalues \( \rightarrow \) mean
  ii) genetic \( \rightarrow \) phenotypic correlation matrix
- Pick tuning factor \( \psi \)
  - \( \text{Opt.} \): Minimise loss in \( \hat{E}_G \) \& \( \hat{E}_E \)
  - \( \Delta L \): Limit change in log \( L \) to \( -\frac{1}{2} \chi^2_{5\%},1 \)
- \( \text{Loss} \): Measure of divergence between matrices of estimates \( \hat{E} \) \& population values \( \Sigma \)

\[
L(\hat{E}, \Sigma) = \text{tr}(\Sigma^{-1}\hat{E}) - \log |\Sigma^{-1}\hat{E}| - q
\]

Loss in estimates of genetic covariance matrix

Distribution over replicates

Conclusions
- Substantial reduction of ‘loss’ in estimates of genetic parameters is feasible
- Post-Estimation Penalisation offers a pragmatic alternative to penalized REML
- Mild penalisation for covariance matrices recommended on routine basis