Pooling estimates of covariance components using a penalized likelihood approach

Karin Meyer

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

The Task

- Estimate covariance matrices
  - MANY traits
  - Multiple sources of variation
- Combine results from part analyses

Problems

- 'Full' multivariate analysis not feasible
  → Use overlapping subsets of traits
  → Multiple estimates for individual σ_{ij}
- Pooled matrices
  → Must be positive definite
  → 'Preserve' phenotypic variances
  → Weight partial results differently
- Sampling variation

Penalized pseudo likelihood

$$-2\log\mathcal{L} \propto \sum_i d_i \left[\log\left|\mathbf{V}_i\right| + \mathrm{tr}(\mathbf{V}_i^{-1}\mathbf{M}_i)\right] + \psi\,\mathcal{P}$$

Penalty

Tuning factor

with

$$\mathbf{M}_i = \sum_x \mathbf{C}^x \otimes \mathbf{S}_i^x$$

$$\mathbf{V}_i = \sum_x \mathbf{C}^x \otimes \mathbf{\Sigma}_i^x$$

'Weight' for subset i

Coefficients for RE x in expect. Cov(Pseudo-relatives)

Estimates for RE x for subset i

Expectation of $\mathbf{S}_i^x$

Solution

Likelihood approach:

- Treat estimates from part analyses as 'data'
  → matrices of corrected MS/CP
- Pool matrices for all random effects simultaneously
  → Assume pseudo pedigree structure
- Impose penalty
  → borrow strength from phenotypic covariance
  → reduce sampling variation

Simulation

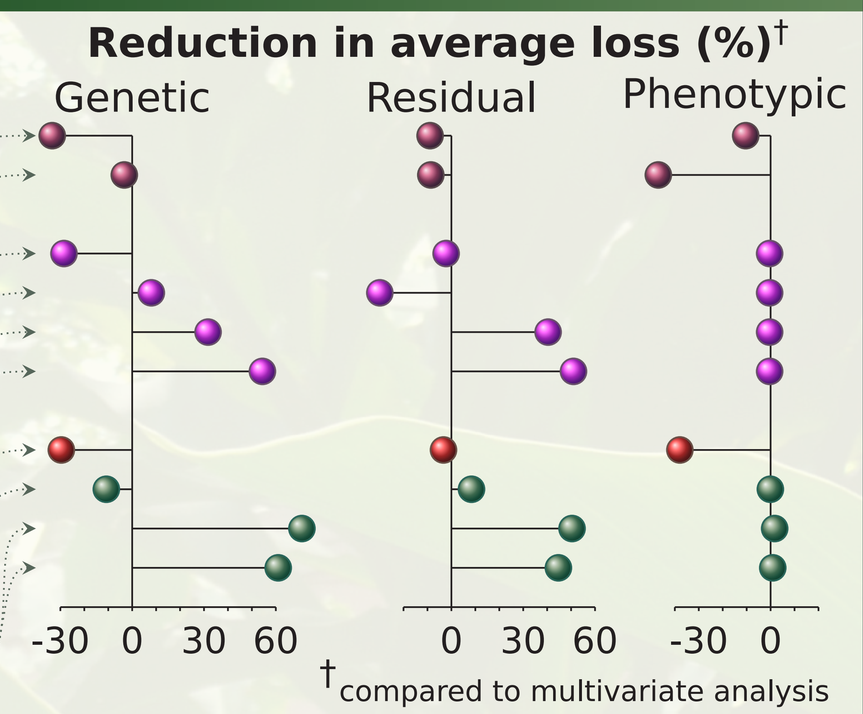
- Paternal half-sib design, 100 ♂
- q=10 traits, 1000 records/trait
  12 sets of population values
- REML estimates: $\hat{\mathbf{\Sigma}}_G$ & $\hat{\mathbf{\Sigma}}_E$
  - full q-variate analysis
  - bivariate → all pairs
- Pool pairwise estimates
  - Manipulate eigenvalues
  - ML approach
- **Loss** → Difference to pop. val.

$$L(\mathbf{\Sigma}, \hat{\mathbf{\Sigma}}) = \mathrm{tr}(\mathbf{\Sigma}^{-1}\hat{\mathbf{\Sigma}}) - \log\left|\mathbf{\Sigma}^{-1}\hat{\mathbf{\Sigma}}\right| - q$$

- Average loss over replicates
- Reduction: pooled *vs.* multiv.
  → positive: pooled is 'better'

Results

1. Matrix eigenvalues
   - truncate at ≈ 0
   - shrink to $\lambda_{min} \approx 0$
2. Canonical eigenvalues
   - truncate at ≈ 0
   - 'bend' to $\lambda_{min} \approx 0$
   - 'bend' further
3. ML approach
   - individual matrices
   - with pseudo pedigree
   - with pseudo ped. & penalty



Conclusions

- Analysis by parts ⇒ good estimates of large covariance matrices
  → some loss in efficiency, esp. for $\hat{\mathbf{\Sigma}}_G$
- Likelihood approach performs well
  → flexible alternative to current methods
  → pool matrices for all sources of variation simultaneously
  → pseudo pedigree structure approximates sampling correlation between $\hat{\mathbf{\Sigma}}_i$ for different sources ⇒ $\hat{\mathbf{\Sigma}}_P \approx$ constant
- Penalization can improve estimates substantially
  → recommend 'mild' penalty

Software

Add-on to WOMBAT

→ run option **– – pool**

→ general form of input
not restricted to WOMBAT estimates!

http://didgeridoo.une.edu.au/km/pool.php

Fourth International Congress on Quantitative Genetics, Edinburgh, Scotland, June 17–22, 2012

AGBU