

Quantitative Genetics, House Sparrows and a Multivariate Gaussian Markov Random Field Model

Ingelin Steinsland & Henrik Jensen Norwegian University of Science and Technology, Trondheim, Norway

Of biological interest

- Quantitative genetics. How much does genetics influence traits?
- Trait = Genetic + environmental effects
 Speed and direction of evolution.

For the population:

- $y = X\beta + Zu + \epsilon = W\binom{u}{\beta} + \epsilon$
- X and Z: incidence matrices, $W = (Z, X)^T$. • u: genetic effect $u \sim N(0, \Sigma_u \otimes A)$

Results

- The Gibbs sampler ran for 10000 iterations (one hour), use GMRFLib.
- Results agree with results from the classical analysis (REML).

• Important for response to environmental changes, e.g. climate changes.



One of the study islands. **Data**

- Most house sparrows of five islands off the coast of Northern Norway are registered since 1993.
 Blood samples taken ⇒ DNA ⇒ Pedigree
- Traits are measured on adult birds, here tarsuslength (t), wing length (w), bill depth (bd) and

- A: genetic covariance matrix
- Structure of A^{-1} from moralising the pedigree $\Rightarrow A^{-1}$ sparse.
- ϵ : environmental effect $\epsilon \sim N(0, \Sigma_{\epsilon} \otimes I)$



Of interest: • u, β, Σ_u and Σ_ϵ • Heritability; $h_j = \frac{\sigma_{uj}^2}{\sigma_{uj}^2 + \sigma_{\epsilon j}^2}, j \in \{t, w, bd, bl\}$ <u>Posterior distributions:</u>

• Mixing seems to be good



bill length (bl) are used:



• There are 2563 birds in the pedigree, of these 746 have (some) adult traits measurements.

Model

For bird $i \ (i = 1, \dots, 746)$:

• $\beta, u | y, \Sigma_u, \Sigma_\epsilon \sim \text{MGMRF}$, a multivariate Gaussian Markov Random Field.

• $\Sigma_u, \Sigma_\epsilon | y, \beta, u \sim \text{Inverted Wishart}$

Gibbs sampler

Use a Gibbs sampler with three blocks for making inference.

Algorithm

• For each iteration

1. Sample Missing data 2. Sample (β, u) of dimension 10280! 3. Sample $(\Sigma_u, \Sigma_{\epsilon})$

Step 2 is the challenge:

Complexity $< \mathcal{O}(n^{1.5})!$

Fast sampling of MGMRF

• Want to sample $x \sim N(0, Q^{-1})$.



Summary

• Breeding model of multiple traits = MGMRF model.

- Use a Gibbs sampler with three blocks
- Because of the sparse structure the pedigree impose, sampling from a MGMRF of dimension 10000 is fast.

Further work

• Use the breeding values u to explore whether

observed traits =fixed +genetic +environmental $y_i = \beta_i + u_i + \epsilon_i$

y_i: observations (traits), y_i = (y_l, y_w, y_{bd}, y_{bl})_i
β_i: "fixed effects" (sex and island), β ~ N(0, σ²_βI)
u_i: genetic effects, u_i ~ N(0, Σ_u)
ε_i: Environmental effects: ε_i ~ N(0, Σ_ε)
Conjugate prior for Σ_u and Σ_ε (inverted Wishart).

• Find Choleskey factor; $Q = LL^T$, and sample $z \sim N(0, I)$. Solve $L^T x = z$.

• Complexity for general multivariate problem $\mathcal{O}(n^3)$.

• Sparse $Q \Rightarrow$ cheaper calculations.

Reorder elements in Q to get sparse L. Original Q Reordered Q Choleskey factor $\begin{array}{c}
 \end{array}$

selection is genetic or environmental.

• Response to selection.

