BayesGWAS

May 12, 2015

1 Bayesian Regression Models for Whole-Genome Analyses

Meuwissen et al. (2001) introduced three regression models for whole-genome prediction of breeding value of the form

$$y_i = \mu + \sum_{j=1}^{k} X_{ij}\alpha_j + e_i,$$

where y_i is the phenotypic value, μ is the intercept, X_{ij} is j^{th} marker covariate of animal i, α_j is the partial regression coefficient of X_{ij}, and e_i are identically and independently distributed residuals with mean zero and variance σ_e^2. In most current analyses, X_{ij} are SNP genotype covariates that can be coded as 0, 1 and 2, depending on the number of B alleles at SNP locus j.

In all three of their models, a flat prior was used for the intercept and a scaled inverted chi-square distribution for σ_e^2. The three models introduced by Meuwissen et al. @Meuwissen.THE.ea.2001a differ only in the prior used for α_j.

1.1 BLUP

In their first model, which they called BLUP, a normal distribution with mean zero and known variance, σ_α^2, is used as the prior for α_j.

1.1.1 The meaning of σ_α^2

Assume the QTL are in the marker panel. Then, the genotypic value g_i for a randomly sampled animal i can be written as

$$g_i = \mu + \mathbf{x}_i'\boldsymbol{\alpha},$$

where $\mathbf{x}_i'$ is the vector of SNP genotype covariates and $\boldsymbol{\alpha}$ is the vector of regression coefficients. Note that randomly sampled animals differ only in $\mathbf{x}_i'$ and have $\boldsymbol{\alpha}$ in common. Thus, genotypic variability is entirely due to variability in the genotypes of animals. So, σ_α^2 is not the genetic variance at a locus (Fernando:2007, Gianola:2009:Genetics:19620397).

1.1.2 Relationship of σ_α^2 to genetic variance

Assume loci with effect on trait are in linkage equilibrium. Then, the additive genetic variance is

$$V_A = \sum_{j}^{k} 2p_j q_j \alpha_j^2,$$

where $p_j = 1 - q_j$ is gene frequency at SNP locus j. Letting $U_j = 2p_j q_j$ and $V_j = \alpha_j^2$,

$$V_A = \sum_{j}^{k} U_j V_j.$$

For a randomly sampled locus, covariance between U_j and V_j is

$$C_{UV} = \frac{\sum_j U_j V_j}{k} - (\frac{\sum_j U_j}{k})(\frac{\sum_j V_j}{k})$$

Rearranging this expression for C_{UV} gives

$$\sum_j U_j V_j = kC_{UV} + (\sum_j U_j)(\frac{\sum_j V_j}{k})$$

So,

$$V_A = kC_{UV} + (\sum_j 2p_j q_j)(\frac{\sum_j \alpha_j^2}{k}).$$

Letting $\sigma_\alpha^2 = \frac{\sum_j \alpha_j^2}{k}$ gives

$$V_A = kC_{UV} + (\sum_j 2p_j q_j)\sigma_\alpha^2$$

and

$$\sigma_\alpha^2 = \frac{V_A - kC_{UV}}{\sum_j 2p_j q_j},$$

which gives

$$\sigma_\alpha^2 = \frac{V_A}{\sum_j 2p_j q_j},$$

if gene frequency is independent of the effect of the gene.

1.1.3 Full-conditionals:

The joint posterior for all the parameters is proportional to

$$\begin{aligned}
f(\boldsymbol{\theta}|\mathbf{y}) &\propto f(\mathbf{y}|\boldsymbol{\theta})f(\boldsymbol{\theta}) \\
&\propto \left(\sigma_e^2\right)^{-n/2} \exp\left\{-\frac{(\mathbf{y} - \mathbf{1}\mu - \sum \mathbf{X}_j\alpha_j)'(\mathbf{y} - \mathbf{1}\mu - \sum \mathbf{X}_j\alpha_j)}{2\sigma_e^2}\right\} \\
&\times \prod_{j=1}^{k} \left(\sigma_\alpha^2\right)^{-1/2} \exp\left\{-\frac{\alpha_j^2}{2\sigma_\alpha^2}\right\} \\
&\times (\sigma_\alpha^2)^{-(\nu_\alpha+2)/2} \exp\left\{-\frac{\nu_\alpha S_\alpha^2}{2\sigma_\alpha^2}\right\} \\
&\times (\sigma_e^2)^{-(2+\nu_e)/2} \exp\left\{-\frac{\nu_e S_e^2}{2\sigma_e^2}\right\},
\end{aligned}$$

where $\boldsymbol{\theta}$ denotes all the unknowns.

1.1.4 Full-conditional for μ

The full-conditional for μ is a normal distribution with mean $\hat{\mu}$ and variance $\frac{\sigma_e^2}{n}$, where $\hat{\mu}$ is the least-squares estimate of μ in the model

$$\mathbf{y} - \sum_{\mathbf{j=1}}^{\mathbf{k}} \mathbf{X}_j\alpha_j = \mathbf{1}\mu + \mathbf{e},$$

and $\frac{\sigma_e^2}{n}$ is the variance of this estimator (n is the number of observations).

1.1.5 Full-conditional for α_j

$$
\begin{aligned}
f(\alpha_j|\text{ELSE}) &\propto \exp\left\{-\frac{(\mathbf{w}_j-\mathbf{X}_j\alpha_j)'(\mathbf{w}_j-\mathbf{X}_j\alpha_j)}{2\sigma_e^2}\right\} \\
&\times \exp\left\{-\frac{\alpha_j^2}{2\sigma_\alpha^2}\right\} \\
&\propto \exp\left\{-\frac{[\mathbf{w}_j'\mathbf{w}_j-2\mathbf{w}_j'\mathbf{X}_j\alpha_j+\alpha_j^2(\mathbf{x}_j'\mathbf{x}_j+\sigma_e^2/\sigma_\alpha^2)]}{2\sigma_e^2}\right\} \\
&\propto \exp\left\{-\frac{(\alpha_j-\hat{\alpha_j})^2}{\frac{2\sigma_e^2}{(\mathbf{x}_j'\mathbf{x}_j+\sigma_e^2/\sigma_\alpha^2)}}\right\},
\end{aligned}
$$

where

$$\mathbf{w}_j=\mathbf{y}-\mathbf{1}\mu-\sum_{l\neq j}\mathbf{X}_l\alpha_l.$$

So, the full-conditional for α_j is a normal distribution with mean

$$\hat{\alpha}_j=\frac{\mathbf{X}_j'\mathbf{w}_j}{(\mathbf{x}_j'\mathbf{x}_j+\sigma_e^2/\sigma_\alpha^2)}$$

and variance $\frac{\sigma_e^2}{(\mathbf{x}_j'\mathbf{x}_j+\sigma_e^2/\sigma_\alpha^2)}$.

1.1.6 Full-conditional for σ_α^2

$$
\begin{aligned}
f(\sigma_\alpha^2|\text{ELSE}) &\propto \prod_{j=1}^{k}\left(\sigma_\alpha^2\right)^{-1/2}\exp\left\{-\frac{\alpha_j^2}{2\sigma_\alpha^2}\right\} \\
&\times (\sigma_\alpha^2)^{-(\nu_\alpha+2)/2}\exp\left\{-\frac{\nu_\alpha S_\alpha^2}{2\sigma_\alpha^2}\right\} \\
&\propto (\sigma_\alpha^2)^{-(k+\nu_\alpha+2)/2}\exp\left\{-\frac{\sum_{j=1}^{k}\alpha_j^2+\nu_\alpha S_{\beta\alpha}^2}{2\sigma_\alpha^2}\right\},
\end{aligned}
$$

and this is proportional to a scaled inverted chi-square distribution with $\tilde{\nu}_\alpha=\nu_\alpha+k$ and scale parameter $\tilde{S}_\alpha^2=(\sum_k\alpha_j^2+\nu_\alpha S_\alpha^2)/\tilde{\nu}_\alpha$.

1.1.7 Full-conditional for σ_e^2

$$
\begin{aligned}
f(\sigma_e^2|\text{ELSE}) &\propto \left(\sigma_e^2\right)^{-n/2}\exp\left\{-\frac{(\mathbf{y}-\mathbf{1}\mu-\sum\mathbf{X}_j\alpha_j)'(\mathbf{y}-\mathbf{1}\mu-\sum\mathbf{X}_j\alpha_j)}{2\sigma_e^2}\right\} \\
&\times (\sigma_e^2)^{-(2+\nu_e)/2}\exp\left\{-\frac{\nu_e S_e^2}{2\sigma_e^2}\right\} \\
&\propto (\sigma_e^2)^{-(n+2+\nu_e)/2}\exp\left\{-\frac{(\mathbf{y}-\mathbf{1}\mu-\sum\mathbf{X}_j\alpha_j)'(\mathbf{y}-\mathbf{1}\mu-\sum\mathbf{X}_j\alpha_j)+\nu_e S_e^2}{2\sigma_e^2}\right\},
\end{aligned}
$$

which is proportional to a scaled inverted chi-square density with $\tilde{\nu}_e=n+\nu_e$ degrees of freedom and $\tilde{S}_e^2=\frac{(\mathbf{y}-\mathbf{1}\mu-\sum\mathbf{X}_j\alpha_j)'(\mathbf{y}-\mathbf{1}\mu-\sum\mathbf{X}_j\alpha_j)+\nu_e S_e^2}{\tilde{\nu}_e}$ scale parameter.

1.2 BayesB

1.2.1 Model

The usual model for BayesB is:

$$y_i = \mu + \sum_{j=1}^{k} X_{ij}\alpha_j + e_i,$$

where the prior μ is flat and the prior for α_j is a mixture distribution:

$$\alpha_j = \begin{cases} 0 & \text{probability}\,\pi \\ \sim N(0,\sigma_j^2) & \text{probability}\,(1-\pi) \end{cases},$$

where σ_j^2 has a scaled inverted chi-square prior with scale parameter S_α^2 and ν_α degrees of freedom. The residual is normally distributed with mean zero and variance σ_e^2, which has a scaled inverted chi-square prior with scale parameter S_e^2 and ν_e degrees of freedom. Meuwissen et al. @Meuwissen.THE.ea.2001a gave a Metropolis-Hastings sampler to jointly sample σ_j^2 and α_j. Here, we will show how the Gibbs sampler can be used in BayesB.

In order to use the Gibbs sampler, the model is written as

$$y_i = \mu + \sum_{j=1}^{k} X_{ij}\beta_j\delta_j + e_i,$$

where $\beta_j \sim N(0,\sigma_j^2)$ and δ_j is Bernoulli$(1-\pi)$:

$$\delta_j = \begin{cases} 0 & \text{probability}\,\pi \\ 1 & \text{probability}\,(1-\pi) \end{cases}.$$

Other priors are the same as in the usual model. Note that in this model, $\alpha_j = \beta_j\delta_j$ has a mixture distribution as in the usual BayesB model.

1.2.2 Full-conditionals:

The joint posterior for all the parameters is proportional to

$$\begin{aligned} f(\boldsymbol{\theta}|\mathbf{y}) &\propto f(\mathbf{y}|\boldsymbol{\theta})f(\boldsymbol{\theta}) \\ &\propto \left(\sigma_e^2\right)^{-n/2}\exp\left\{-\frac{(\mathbf{y}-\mathbf{1}\mu-\sum\mathbf{X}_j\beta_j\delta_j)'(\mathbf{y}-\mathbf{1}\mu-\sum\mathbf{X}_j\beta_j\delta_j)}{2\sigma_e^2}\right\} \\ &\times \prod_{j=1}^{k}\left(\sigma_j^2\right)^{-1/2}\exp\left\{-\frac{\beta_j^2}{2\sigma_j^2}\right\} \\ &\times \prod_{j=1}^{k}\pi^{(1-\delta_j)}(1-\pi)^{\delta_j} \\ &\times \prod_{j=1}^{k}(\sigma_j^2)^{-(\nu_\beta+2)/2}\exp\left\{-\frac{\nu_\beta S_\beta^2}{2\sigma_j^2}\right\} \\ &\times (\sigma_e^2)^{-(2+\nu_e)/2}\exp\left\{-\frac{\nu_e S_e^2}{2\sigma_e^2}\right\}, \end{aligned}$$

where $\boldsymbol{\theta}$ denotes all the unknowns.

1.2.3 Full-conditional for μ

The full-conditional for μ is a normal distribution with mean $\hat{\mu}$ and variance $\frac{\sigma_e^2}{n}$, where $\hat{\mu}$ is the least-squares estimate of μ in the model

$$\mathbf{y} - \sum_{\mathbf{j=1}}^{\mathbf{k}}\mathbf{X}_j\beta_j\delta_j = \mathbf{1}\mu + \mathbf{e},$$

and $\frac{\sigma_e^2}{n}$ is the variance of this estimator (n is the number of observations).

1.2.4 Full-conditional for β_j

$$
\begin{aligned}
f(\beta_j|\text{ELSE}) &\propto \exp\left\{-\frac{(\mathbf{w}_j - \mathbf{X}_j\beta_j\delta_j)'(\mathbf{w}_j - \mathbf{X}_j\beta_j\delta_j)}{2\sigma_e^2}\right\} \\
&\times \exp\left\{-\frac{\beta_j^2}{2\sigma_j^2}\right\} \\
&\propto \exp\left\{-\frac{[\mathbf{w}_j'\mathbf{w}_j - 2\mathbf{w}_j'\mathbf{X}_j\beta_j\delta_j + \beta_j^2(\mathbf{x}_j'\mathbf{x}_j\delta_j + \sigma_e^2/\sigma_j^2)]}{2\sigma_e^2}\right\} \\
&\propto \exp\left\{-\frac{(\beta_j - \hat{\beta}_j)^2}{\frac{2\sigma_e^2}{(\mathbf{x}_j'\mathbf{x}_j\delta_j + \sigma_e^2/\sigma_j^2)}}\right\},
\end{aligned}
$$

where

$$
\mathbf{w}_j = \mathbf{y} - \mathbf{1}\mu - \sum_{l\neq j}\mathbf{X}_l\beta_l\delta_l.
$$

So, the full-conditional for β_j is a normal distribution with mean

$$
\hat{\beta}_j = \frac{\mathbf{X}_j'\mathbf{w}_j\delta_j}{(\mathbf{x}_j'\mathbf{x}_j\delta_j + \sigma_e^2/\sigma_j^2)}
$$

and variance $\frac{\sigma_e^2}{(\mathbf{x}_j'\mathbf{x}_j\delta_j + \sigma_e^2/\sigma_j^2)}$.

1.2.5 Full-conditional for δ_j

$$
\Pr(\delta_j = 1|\text{ELSE}) \propto \frac{h(\delta_j = 1)}{h(\delta_j = 1) + h(\delta_j = 0)},
$$

where

$$
h(\delta_j) = \pi^{(1-\delta_j)}(1-\pi)^{\delta_j}\exp\left\{-\frac{(\mathbf{w}_j - \mathbf{X}_j\beta_j\delta_j)'(\mathbf{w}_j - \mathbf{X}_j\beta_j\delta_j)}{2\sigma_e^2}\right\}.
$$

1.2.6 Full-conditional for σ_j^2

$$
\begin{aligned}
f(\sigma_j^2|\text{ELSE}) &\propto \left(\sigma_j^2\right)^{-1/2}\exp\left\{-\frac{\beta_j^2}{2\sigma_j^2}\right\} \\
&\times (\sigma_j^2)^{-(\nu_\beta+2)/2}\exp\left\{-\frac{\nu_\beta S_\beta^2}{2\sigma_j^2}\right\} \\
&\propto (\sigma_j^2)^{-(1+\nu_\beta+2)/2}\exp\left\{-\frac{\beta_j^2 + \nu_\beta S_\beta^2}{2\sigma_j^2}\right\},
\end{aligned}
$$

and this is proportional to a scaled inverted chi-square distribution with $\tilde{\nu}_j = \nu_\beta + 1$ and scale parameter $\tilde{S}_j^2 = (\beta_j^2 + \nu_\beta S_\beta^2)/\tilde{\nu}_j$.

1.2.7 Full-conditional for σ_e^2

$$
\begin{aligned}
f(\sigma_e^2|\text{ELSE}) &\propto \left(\sigma_e^2\right)^{-n/2}\exp\left\{-\frac{(\mathbf{y} - \mathbf{1}\mu - \sum\mathbf{X}_j\beta_j\delta_j)'(\mathbf{y} - \mathbf{1}\mu - \sum\mathbf{X}_j\beta_j\delta_j)}{2\sigma_e^2}\right\} \\
&\times (\sigma_e^2)^{-(2+\nu_e)/2}\exp\left\{-\frac{\nu_e S_e^2}{2\sigma_e^2}\right\} \\
&\propto (\sigma_e^2)^{-(n+2+\nu_e)/2}\exp\left\{-\frac{(\mathbf{y} - \mathbf{1}\mu - \sum\mathbf{X}_j\beta_j\delta_j)'(\mathbf{y} - \mathbf{1}\mu - \sum\mathbf{X}_j\beta_j\delta_j) + \nu_e S_e^2}{2\sigma_e^2}\right\},
\end{aligned}
$$

which is proportional to a scaled inverted chi-square density with $\tilde{\nu}_e = n + \nu_e$ degrees of freedom and $\tilde{S}_e^2 = \frac{(\mathbf{y}-\mathbf{1}\mu-\sum \mathbf{X}_j\beta_j\delta_j)'(\mathbf{y}-\mathbf{1}\mu-\sum \mathbf{X}_j\beta_j\delta_j)+\nu_e S_e^2}{\tilde{\nu_e}}$ scale parameter.