Part II

Bayesian Inference: Application to Whole Genome Analyses

Model

Model:

$$y_i = \mu + \sum_j X_{ij}\alpha_j + e_i$$

Priors:

- $\mu \propto$ constant (not proper, but posterior is proper)
- $e_i \sim$ (iid)N$(0, \sigma_e^2)$; $\sigma_e^2 \sim \nu_e S_e^2 \chi_{\nu_e}^{-2}$
- Consider several different priors for α_j

Normal

- Prior: $(\alpha_j|\sigma^2_\alpha) \sim$ (iid)N$(0, \sigma^2_\alpha)$; σ^2_α is known
- What is σ^2_α?
- Assume the QTL genotypes are a subset of those available for the analysis
  - Then, the genotypic value of i can be written as:

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

  - Note that $\boldsymbol{\alpha}$ is common to all i
  - Thus, the variance of g_i comes from $\boldsymbol{x}'_i$ being random
- So, σ^2_α is not the genetic variance at a locus
- If locus j is randomly sampled from all the loci available for analysis:
  - Then, α_j will be a random variable
  - $\sigma^2_\alpha = \text{Var}(\alpha_j)$

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

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

Relationship of σ^2_α to genetic variance

Rearranging the previous 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 $\sigma^2_\alpha = \frac{\sum_j \alpha_j^2}{k}$ gives

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

and,

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

Blocked Gibbs sampler

- Let $\boldsymbol{\theta}' = [\mu, \boldsymbol{\alpha}']$
- Can show that $(\boldsymbol{\theta}|\boldsymbol{y}, \sigma_e^2) \sim \mathrm{N}(\hat{\boldsymbol{\theta}}, \boldsymbol{C}^{-1}\sigma_e^2)$
- $$\hat{\boldsymbol{\theta}} = \boldsymbol{C}^{-1}\boldsymbol{W}'\boldsymbol{y}; \quad \boldsymbol{W} = [\boldsymbol{1}, \boldsymbol{X}]$$
- $$\boldsymbol{C} = \begin{bmatrix} \boldsymbol{1}'\boldsymbol{1} & \boldsymbol{1}'\boldsymbol{X} \\ \boldsymbol{X}'\boldsymbol{1} & \boldsymbol{X}'\boldsymbol{X} + \boldsymbol{I}\frac{\sigma_e^2}{\sigma_\alpha^2} \end{bmatrix}$$
- Blocked Gibbs sampler
  - García-Cortés and Sorensen (1996, GSE 28:121-126)
  - *Likelihood, Bayesian and MCMC Methods* $\cdots$ (LBMMQG, Sorensen and Gianola, 2002)

Full conditionals for single-site Gibbs

- $(\mu|\boldsymbol{y}, \boldsymbol{\alpha}, \sigma_e^2) \sim \mathrm{N}(\frac{\mathbf{1}'(\boldsymbol{y}-\boldsymbol{X\alpha})}{n}, \frac{\sigma_e^2}{n})$
- $(\alpha_j|\boldsymbol{y}, \mu, \boldsymbol{\alpha}_{j\_}, \sigma_e^2) \sim \mathrm{N}(\hat{\alpha}_j, \frac{\sigma_e^2}{c_j})$
  - $$\hat{\alpha}_j = \frac{\boldsymbol{x}_j'\boldsymbol{w}}{c_j}$$
  - $$\boldsymbol{w} = \boldsymbol{y} - \mathbf{1}\mu - \sum_{j' \neq j} \boldsymbol{x}_{j'}\alpha_{j'}$$
  - $$c_j = (\boldsymbol{x}_j'\boldsymbol{x}_j + \frac{\sigma_e^2}{\sigma_\alpha^2})$$
- $(\sigma_e^2|\boldsymbol{y}, \mu, \boldsymbol{\alpha}) \sim [(\boldsymbol{y} - \boldsymbol{W\theta})'(\boldsymbol{y} - \boldsymbol{W\theta}) + \nu_e S_e^2]\chi^{-2}_{(\nu_e+n)}$

Derive: full conditional for α_j

From Bayes' Theorem,

$$f(\alpha_j|\boldsymbol{y},\mu,\boldsymbol{\alpha}_{j\_},\sigma_e^2) = \frac{f(\alpha_j,\boldsymbol{y},\mu,\boldsymbol{\alpha}_{j\_},\sigma_e^2)}{f(\boldsymbol{y},\mu,\boldsymbol{\alpha}_{j\_},\sigma_e^2)}$$

$$\propto f(\boldsymbol{y}|\alpha_j,\mu,\boldsymbol{\alpha}_{j\_},\sigma_e^2)f(\alpha_j)f(\mu,\boldsymbol{\alpha}_{j\_},\sigma_e^2)$$

$$\propto (\sigma_e^2)^{-n/2}\exp\{-\frac{(\boldsymbol{w}-\boldsymbol{x}_j\alpha_j)'(\boldsymbol{w}-\boldsymbol{x}_j\alpha_j)}{2\sigma_e^2}\}(\sigma_\alpha^2)^{-1/2}\exp\{-\frac{\alpha_j^2}{2\sigma_\alpha^2}\}$$

where

$$\boldsymbol{w} = \boldsymbol{y} - \mathbf{1}\mu - \sum_{j\neq j'}\boldsymbol{x}_{j'}\alpha_{j'}$$

Derive: full conditional for α_j

The exponential terms in the joint density can be written as:

$$-\frac{1}{2\sigma_e^2}\{\boldsymbol{w}'\boldsymbol{w} - 2\boldsymbol{x}_j'\boldsymbol{w}\alpha_j + [\boldsymbol{x}_j'\boldsymbol{x}_j + \frac{\sigma_e^2}{\sigma_\alpha^2}]\alpha_j^2\}$$

Completing the square in this expression with respect to α_j gives

$$-\frac{1}{2\sigma_e^2}\{c_j(\alpha_j - \hat{\alpha}_j)^2 + \boldsymbol{w}'\boldsymbol{w} - c_j\hat{\alpha}_j{}^2\}$$

where

$$\hat{\alpha}_j = \frac{\boldsymbol{x}_j\boldsymbol{w}}{c_j}$$

So,

$$f(\alpha_j|\boldsymbol{y}, \mu, \boldsymbol{\alpha}_{j\_}, \sigma_e^2) \propto \exp\{-\frac{(\alpha_j - \hat{\alpha}_j)^2}{2\frac{\sigma_e^2}{c_j}}\}$$

Full conditional for σ_e^2

From Bayes' theorem,

$$f(\sigma_e^2|\boldsymbol{y},\mu,\boldsymbol{\alpha}) = \frac{f(\sigma_e^2,\boldsymbol{y},\mu,\boldsymbol{\alpha})}{f(\boldsymbol{y},\mu,\boldsymbol{\alpha})}$$

$$\propto f(\boldsymbol{y}|\sigma_e^2,\mu,\boldsymbol{\alpha})f(\sigma_e^2)f(\mu,\boldsymbol{\alpha})$$

where

$$f(\boldsymbol{y}|\sigma_e^2,\mu,\boldsymbol{\alpha}) \propto (\sigma_e^2)^{-n/2}\exp\{-\frac{(\boldsymbol{w}-\boldsymbol{x}_j\alpha_j)'(\boldsymbol{w}-\boldsymbol{x}_j\alpha_j)}{2\sigma_e^2}\}$$

and

$$f(\sigma_e^2) = \frac{(S_e^2\nu_e/2)^{\nu_e/2}}{\Gamma(\nu/2)}(\sigma_e^2)^{-(2+\nu_e)/2}\exp(-\frac{\nu_e S_e^2}{2\sigma_e^2})$$

Full conditional for σ_e^2

So,

$$f(\sigma_e^2|\boldsymbol{y},\mu,\boldsymbol{\alpha}) \propto (\sigma_e^2)^{-(2+n+\nu_e)/2} \exp(-\frac{SSE+\nu_e S_e^2}{2\sigma_e^2})$$

where

$$SSE = (\boldsymbol{w}-\boldsymbol{x}_j\alpha_j)'(\boldsymbol{w}-\boldsymbol{x}_j\alpha_j)$$

So,

$$f(\sigma_e^2|\boldsymbol{y},\mu,\boldsymbol{\alpha}) \sim \tilde{\nu}_e\tilde{S}_e^2\chi^{-2}_{\tilde{\nu}_e}$$

where

$$\tilde{\nu}_e = n+\nu_e; \quad \tilde{S}_e^2 = \frac{SSE+\nu_e S_e^2}{\tilde{\nu}_e}$$

Alternative view of Normal prior

Consider fixed linear model:

$$\boldsymbol{y} = \mathbf{1}\mu + \boldsymbol{X}\boldsymbol{\alpha} + \boldsymbol{e}$$

This can be also written as

$$\boldsymbol{y} = \begin{bmatrix} \mathbf{1} & \boldsymbol{X} \end{bmatrix} \begin{bmatrix} \mu \\ \boldsymbol{\alpha} \end{bmatrix} + \boldsymbol{e}$$

Suppose we observe for each locus:

$$y_j^* = \alpha_j + \epsilon_j$$

Least Squares with Additional Data

Fixed linear model with the additional data:

$$\begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{y}^* \end{bmatrix} = \begin{bmatrix} \mathbf{1} & \boldsymbol{X} \\ \mathbf{0} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \mu \\ \boldsymbol{\alpha} \end{bmatrix} + \begin{bmatrix} \boldsymbol{e} \\ \boldsymbol{\epsilon} \end{bmatrix}$$

OLS Equations:

$$\begin{bmatrix} \mathbf{1}' & \mathbf{0}' \\ \boldsymbol{X}' & \boldsymbol{I}' \end{bmatrix} \begin{bmatrix} \boldsymbol{I}_n \frac{1}{\sigma_e^2} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_k \frac{1}{\sigma_\epsilon^2} \end{bmatrix} \begin{bmatrix} \mathbf{1} & \boldsymbol{X} \\ \mathbf{0} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{\boldsymbol{\alpha}} \end{bmatrix} = \begin{bmatrix} \mathbf{1}' & \mathbf{0}' \\ \boldsymbol{X}' & \boldsymbol{I}' \end{bmatrix} \begin{bmatrix} \boldsymbol{I}_n \frac{1}{\sigma_e^2} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_k \frac{1}{\sigma_\epsilon^2} \end{bmatrix} \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{y}^* \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{1}'\mathbf{1} & \mathbf{1}'\boldsymbol{X} \\ \boldsymbol{X}'\mathbf{1} & \boldsymbol{X}'\boldsymbol{X} + \boldsymbol{I}\frac{\sigma_e^2}{\sigma_\epsilon^2} \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{\boldsymbol{\alpha}} \end{bmatrix} = \begin{bmatrix} \mathbf{1}'\boldsymbol{y} \\ \boldsymbol{X}'\boldsymbol{y} + \boldsymbol{y}^* \frac{\sigma_e^2}{\sigma_\epsilon^2} \end{bmatrix}$$

Univariate-t

Prior:

$$(\alpha_j|\sigma_j^2) \sim \mathrm{N}(0, \sigma_j^2)$$

$$\sigma_j^2 \sim \nu_\alpha S^2_{\nu_\alpha} \chi^{-2}_{\nu_\alpha}$$

Can show that the unconditional distribution for α_j is

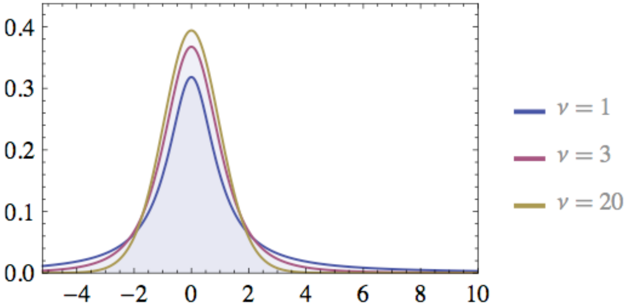
$$\alpha_j \sim (\mathrm{iid}) t(0, S^2_{\nu_\alpha}, \nu_\alpha)$$

(Sorensen and Gianola, 2002, LBMMQG pages 28,60)

This is Bayes-A (Meuwissen et al., 2001; Genetics 157:1819-1829)

Univariate-t

Plots of PDF for typical parameters:



Generated by Wolfram|Alpha (www.wolframalpha.com)

Full conditional for single-site Gibbs

Full conditionals are the same as in the "Normal" model for μ, α_j, and σ_e^2. Let

$$\boldsymbol{\xi} = [\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2]$$

Full conditional conditional for σ_j^2:

$$f(\sigma_j^2 | \boldsymbol{y}, \mu, \boldsymbol{\alpha}, \boldsymbol{\xi}_{j\_}, \sigma_e^2) \propto f(\boldsymbol{y}, \mu, \boldsymbol{\alpha}, \boldsymbol{\xi}, \sigma_e^2)$$

$$\propto f(\boldsymbol{y}|\mu, \boldsymbol{\alpha}, \boldsymbol{\xi}, \sigma_e^2) f(\alpha_j|\sigma_j^2) f(\sigma_j^2) f(\mu, \boldsymbol{\alpha}_{j\_}, \boldsymbol{\xi}_{j\_} \sigma_e^2)$$

$$\propto (\sigma_j^2)^{-1/2} \exp\{-\frac{\alpha_j^2}{2\sigma_j^2}\} (\sigma_j^2)^{-(2+\nu_\alpha)/2} \exp\{\frac{\nu_\alpha S_\alpha^2}{2\sigma_j^2}\}$$

$$\propto (\sigma_j^2)^{-(2+\nu_\alpha+1)/2} \exp\{\frac{\alpha_j^2 + \nu_\alpha S_\alpha^2}{2\sigma_j^2}\}$$

Full conditional for σ^2_α

So,

$$(\sigma^2_\alpha | \boldsymbol{y}, \mu, \boldsymbol{\alpha}, \boldsymbol{\xi}\_, \sigma^2_e) \sim \tilde{\nu}_\alpha \tilde{S}^2_\alpha \chi^{-2}_{\nu_\alpha}$$

where

$$\tilde{\nu}_\alpha = \nu_\alpha + 1$$

and

$$\tilde{S}^2_\alpha = \frac{\alpha^2_j + \nu_\alpha S^2_\alpha}{\tilde{\nu}_\alpha}$$

Multivariate-t

Prior:

$$(\alpha_j|\sigma^2_\alpha) \sim \text{(iid)N}(0, \sigma^2_\alpha)$$

$$\sigma^2_\alpha \sim \nu_\alpha S^2_{\nu_\alpha}\chi^{-2}_{\nu_\alpha}$$

Can show that the unconditional distribution for $\boldsymbol{\alpha}$ is

$$\boldsymbol{\alpha} \sim \text{multivariate-}t(\mathbf{0}, \boldsymbol{I}S^2_{\nu_\alpha}, \nu_\alpha)$$

(Sorensen and Gianola, 2002, LBMMQG page 60)

We will see later that this is Bayes-C with $\pi = 0$.

Full conditional for σ_α^2

We will see later that

$$(\sigma_\alpha^2|\boldsymbol{y}, \mu, \boldsymbol{\alpha}, \sigma_e^2) \sim \tilde{\nu}_\alpha \tilde{S}_\alpha^2 \chi_{\nu_\alpha}^{-2}$$

where

$$\tilde{\nu}_\alpha = \nu_\alpha + k$$

and

$$\tilde{S}_\alpha^2 = \frac{\boldsymbol{\alpha}'\boldsymbol{\alpha} + \nu_\alpha S_\alpha^2}{\tilde{\nu}_\alpha}$$

Spike and univariate-t

Prior:

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

and

$$(\sigma_j^2|\nu_\alpha, S_\alpha^2) \sim \nu_\alpha S_\alpha^2 \chi_{\nu_\alpha}^{-2}$$

Thus,

$$(\alpha_j|\pi)(\text{iid})\begin{cases}\sim \text{univariate-}t(0,S_\alpha^2,\nu_\alpha) & \text{probability } (1-\pi),\\ = 0 & \text{probability } \pi\end{cases}$$

This is Bayes-B (Meuwissen et al., 2001; Genetics 157:1819-1829)

Notation for sampling from mixture

The indicator variable δ_j is defined as

$$\delta_j = 1 \Rightarrow (\alpha_j|\sigma_j^2) \sim \mathrm{N}(0, \sigma_j^2)$$

and

$$\delta_j = 0 \Rightarrow (\alpha_j|\sigma_j^2) = 0$$

Sampling strategy in MHG (2001)

- Sampling σ_e^2 and μ are as under the Normal prior.
- MHG proposed to use a Metropolis-Hastings sampler to draw samples for σ_j^2 and α_j jointly from their full-conditional distribution.
- First, σ_j^2 is sampled from

$$f(\sigma_j^2 | \boldsymbol{y}, \mu, \boldsymbol{\alpha}_{j\_}, \boldsymbol{\xi}_{\_}, \sigma_e^2)$$

- Then, α_j is sampled from its full-conditional, which is identical to that under the Normal prior

Sampling σ_j^2

The prior for σ_j^2 is used as the proposal. In this case, the MH acceptance probability becomes

$$\alpha = \min(1, \frac{f(\boldsymbol{y}|\sigma^2_{can}, \boldsymbol{\theta}_{j\_})}{f(\boldsymbol{y}|\sigma^2_j, \boldsymbol{\theta}_{j\_})})$$

where σ^2_{can} is used to denote the candidate value for σ_j^2, and $\boldsymbol{\theta}_{j\_}$ all the other parameters. It can be shown that, α_j depends on $\boldsymbol{y}$ only through $r_j = \boldsymbol{x}_j'\boldsymbol{w}$ (look here). Thus

$$f(\boldsymbol{y}|\sigma_j^2, \boldsymbol{\theta}_{j\_}) \propto f(r_j|\sigma_j^2, \boldsymbol{\theta}_{j\_})$$

"Likelihood" for σ_j^2

Recall that

$$\boldsymbol{w} = \boldsymbol{y} - \mathbf{1}\mu - \sum_{j' \neq j} \boldsymbol{x}_{j'}\alpha_{j'} = \boldsymbol{x}_j\alpha_j + \boldsymbol{e}$$

Then,

$$\mathrm{E}(\boldsymbol{w}|\sigma_j^2, \boldsymbol{\theta}_{j\_}) = \mathbf{0}$$

When $\delta = 1$:

$$\mathrm{Var}(\boldsymbol{w}|\delta_j = 1, \sigma_j^2, \boldsymbol{\theta}_{j\_}) = \boldsymbol{x}_j\boldsymbol{x}_j'\sigma_j^2 + \boldsymbol{I}\sigma_e^2$$

and $\delta = 0$:

$$\mathrm{Var}(\boldsymbol{w}|\delta_j = 0, \sigma_j^2, \boldsymbol{\theta}_{j\_}) = \boldsymbol{I}\sigma_e^2$$

"Likelihood" for σ_j^2

So,

$$E(r_j|\sigma_j^2, \boldsymbol{\theta}_{j\_}) = 0$$

and

$$\text{Var}(r_j|\delta_j = 1, \sigma_j^2, \boldsymbol{\theta}_{j\_}) = (\boldsymbol{x}_j'\boldsymbol{x}_j)^2\sigma_j^2 + \boldsymbol{x}_j'\boldsymbol{x}_j\sigma_e^2 = v_1$$

$$\text{Var}(r_j|\delta_j = 0, \sigma_j^2, \boldsymbol{\theta}_{j\_}) = \boldsymbol{x}_j'\boldsymbol{x}_j\sigma_e^2 = v_0$$

So,

$$f(r_j|\delta_j, \sigma_j^2, \boldsymbol{\theta}_{j\_}) \propto (v_\delta)^{-1/2} \exp\{-\frac{r_j^2}{2v_\delta}\}$$

MH acceptance probability when prior is used as proposal

Suppose we want to sample θ from $f(\theta|\boldsymbol{y})$ using the MH with its prior as proposal. Then, the MH acceptance probability becomes:

$$\alpha = min(1, \frac{f(\theta_{can}|\boldsymbol{y})f(\theta^{t-1})}{f(\theta^{t-1}|\boldsymbol{y})f(\theta_{can})}$$

where $f(\theta)$ is the prior for θ. Using Bayes' theorem, the target density can be written as:

$$f(\theta|\boldsymbol{y}) = f(\boldsymbol{y}|\theta)f(\theta)$$

Then, the acceptance probability becomes

$$\alpha = min(1, \frac{f(\boldsymbol{y}|\theta_{can})f(\theta_{can})f(\theta^{t-1})}{f(\boldsymbol{y}|\theta^{t-1})f(\theta^{t-1})f(\theta_{can})}$$

Alternative algorithm for spike and univariate-t

Rather than use the prior as the proposal for sampling σ_j^2, we

- sample $\delta_j = 1$ with probability 0.5
- when $\delta = 1$, sample σ_j^2 from a scaled inverse chi-squared distribution with
  - scale parameter = $\sigma_j^{2(t-1)}/2$ and 4 degrees of freedom when $\sigma_j^{2(t-1)} > 0$, and
  - scale parameter = S_α^2 and 4 degrees of freedom when $\sigma_j^{2(t-1)} = 0$

Multivariate-t mixture

Prior:

$$(\alpha_j|\pi, \sigma^2_\alpha) \begin{cases} \sim \mathrm{N}(0, \sigma^2_\alpha) & \text{probability } (1-\pi), \\ = 0 & \text{probability } \pi \end{cases}$$

and

$$(\sigma^2_\alpha|\nu_\alpha, S^2_\alpha) \sim \nu_\alpha S^2_\alpha \chi^{-2}_{\nu_\alpha}$$

Further,

$$\pi \sim \mathrm{Uniform}(0, 1)$$

- The α_j variables with their corresponding $\delta_j = 1$ will follow a multivariate-t distribution.
- This is what we have called Bayes-Cπ

Full conditionals for single-site Gibbs

Full-conditional distributions for μ, $\boldsymbol{\alpha}$, and σ_e^2 are as with the Normal prior.
Full-conditional for δ_j:

$$\Pr(\delta_j|\boldsymbol{y}, \mu, \boldsymbol{\alpha}_{-j}, \boldsymbol{\delta}_{-j}, \sigma_\alpha^2, \sigma_e^2, \pi) = \Pr(\delta_j|r_j, \boldsymbol{\theta}_{j\_})$$

$$\Pr(\delta_j|r_j, \boldsymbol{\theta}_{j\_}) = \frac{f(\delta_j, r_j|\boldsymbol{\theta}_{j\_})}{f(r_j|\boldsymbol{\theta}_{j\_})}$$

$$= \frac{f(r_j|\delta_j, \boldsymbol{\theta}_{j\_}) \Pr(\delta_j|\pi)}{f(r_j|\delta_j = 0, \boldsymbol{\theta}_{j\_})\pi + f(r_j|\delta_j = 1, \boldsymbol{\theta}_{j\_})(1 - \pi)}$$

Full conditional for σ^2_α

This can be written as

$$f(\sigma^2_\alpha | \boldsymbol{y}, \mu, \boldsymbol{\alpha}, \boldsymbol{\delta}, \sigma^2_e) \propto f(\boldsymbol{y} | \sigma^2_\alpha, \mu, \boldsymbol{\alpha}, \boldsymbol{\delta}, \sigma^2_e) f(\sigma^2_\alpha, \mu, \boldsymbol{\alpha}, \boldsymbol{\delta}, \sigma^2_e)$$

But, can see that

$$f(\boldsymbol{y} | \sigma^2_\alpha, \mu, \boldsymbol{\alpha}, \boldsymbol{\delta}, \sigma^2_e) \propto f(\boldsymbol{y} | \mu, \boldsymbol{\alpha}, \boldsymbol{\delta}, \sigma^2_e)$$

So,

$$f(\sigma^2_\alpha | \boldsymbol{y}, \mu, \boldsymbol{\alpha}, \boldsymbol{\delta}, \sigma^2_e) \propto f(\sigma^2_\alpha, \mu, \boldsymbol{\alpha}, \boldsymbol{\delta}, \sigma^2_e)$$

Note that σ^2_α appears only in $f(\boldsymbol{\alpha} | \sigma^2_\alpha)$ and $f(\sigma^2_\alpha)$:

$$f(\boldsymbol{\alpha} | \sigma^2_\alpha) \propto (\sigma^2_\alpha)^{-k/2} \exp\{-\frac{\boldsymbol{\alpha}'\boldsymbol{\alpha}}{2\sigma^2_\alpha}\}$$

and

$$f(\sigma^2_\alpha) \propto (\sigma^2_\alpha)^{-(\nu_\alpha + 2)/2} \exp\{\frac{\nu_\alpha S^2_\alpha}{2\sigma^2_\alpha}\}$$

Full conditional for σ_α^2

Combining these two densities gives:

$$f(\sigma_\alpha^2|\boldsymbol{y},\mu,\boldsymbol{\alpha},\boldsymbol{\delta},\sigma_e^2) \propto (\sigma_\alpha^2)^{-(k+\nu_\alpha+2)/2}\exp\{\frac{\boldsymbol{\alpha}'\boldsymbol{\alpha}+\nu_\alpha S_\alpha^2}{2\sigma_\alpha^2}\}$$

So,

$$(\sigma_\alpha^2|\boldsymbol{y},\mu,\boldsymbol{\alpha},\boldsymbol{\delta},\sigma_e^2) \sim \tilde{\nu}_\alpha \tilde{S}_\alpha^2 \chi_{\tilde{\nu}_\alpha}^{-2}$$

where

$$\tilde{\nu}_\alpha = k+\nu_\alpha$$

and

$$\tilde{S}_\alpha^2 = \frac{\boldsymbol{\alpha}'\boldsymbol{\alpha}+\nu_\alpha S_\alpha^2}{\tilde{\nu}_\alpha}$$

Hyper parameter: S_α^2

If σ^2 is distributed as a scaled, inverse chi-square random variable with scale parameter S^2 and degrees of freedom ν

$$E(\sigma^2) = \frac{\nu S^2}{\nu - 2}$$

Recall that under some assumptions

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

So, we take

$$S_\alpha^2 = \frac{(\nu_\alpha - 2) V_a}{\nu_\alpha k(1-\pi) 2\overline{pq}}$$

Full conditional for π

Using Bayes' theorem,

$$f(\pi|\boldsymbol{\delta}, \mu, \boldsymbol{\alpha}, \sigma_\alpha^2, \sigma_e^2, \boldsymbol{y}) \propto f(\boldsymbol{y}|\pi, \boldsymbol{\delta}, \mu, \boldsymbol{\alpha}, \sigma_\alpha^2, \sigma_e^2) f(\pi, \boldsymbol{\delta}, \mu, \boldsymbol{\alpha}, \sigma_\alpha^2, \sigma_e^2)$$

But,

- Conditional on $\boldsymbol{\delta}$ the likelihood is free of π
- Further, π only appears in probability of the vector of bernoulli variables: $\boldsymbol{\delta}$

Thus,

$$f(\pi|\boldsymbol{\delta}, \mu, \boldsymbol{\alpha}, \sigma_\alpha^2, \sigma_e^2, \boldsymbol{y}) = \pi^{(k-m)}(1-\pi)^m$$

where $m = \boldsymbol{\delta}'\boldsymbol{\delta}$, and k is the number of markers. Thus, π is sampled from a beta distribution with $a = k - m + 1$ and $b = m + 1$.

Simulation I

- 2000 unlinked loci in LE
- 10 of these are QTL: $\pi = 0.995$
- $h^2 = 0.5$
- Locus effects estimated from 250 individuals

Results for Bayes-B

Correlations between true and predicted additive genotypic values estimated from 32 replications

π	S^2	Correlation
0.995	0.2	0.91 (0.009)
0.8	0.2	0.86 (0.009)
0.0	0.2	0.80 (0.013)
0.995	2.0	0.90 (0.007)
0.8	2.0	0.77 (0.009)
0.0	2.0	0.35 (0.022)

Simulation II

- 2000 unlinked loci with Q loci having effect on trait
- N is the size of training data set
- Heritability = 0.5
- Validation in an independent data set with 1000 individuals
- Bayes-B and Bayes-Cπ with $\pi = 0.5$

Results

Results from 15 replications

				Corr($g, \hat{g}$)	
N	Q	π	$\hat{\pi}$	Bayes-Cπ	Bayes-B
2000	10	0.995	0.994	0.995	0.937
2000	200	0.90	0.899	0.866	0.834
2000	1900	0.05	0.202	0.613	0.571
4000	1900	0.05	0.096	0.763	0.722

Simulation II

- Genotypes: 50k SNPs from 1086 Purebred Angus animals, ISU
- Phenotypes:
  - QTL simulated from 50 randomly sampled SNPs
  - substitution effect sampled from N(0,σ^2_α)
  - $\sigma^2_\alpha = \frac{\sigma^2_g}{50 2\bar{p}\bar{q}}$
  - $h^2 = 0.25$
- QTL were included in the marker panel
- Marker effects were estimated for 50k SNPs

Validation

- Genotypes: 50k SNPs from 984 crossbred animals, CMP
- Additive genetic merit (g_i) computed from the 50 QTL
- Additive genetic merit predicted ($\hat{g}_i$) using estimated effects for 50k SNP panel

Results

Correlations between g_i and $\hat{g}_i$ estimated from 3 replications

	Correlation	
π	Bayes-B	Bayes-C
0.999	0.86	0.86
0.25	0.70	0.26

BayesCπ:

- $\hat{\pi} = 0.999$
- Correlation = 0.86