# BESSiE
# A program for Best Linear Unbiased Prediction and Bayesian analysis of linear mixed models including large scale genomic markers

**Vinzent Boerner**

Animal Genetics and Breeding Unit (AGBU), University of New England
Armidale, 2351, NSW, Australia

February 4, 2016

# Overview

1. Bayesian estimation of marker effects

2. About BESSiE

## The linear model for estimating effects of genetic marker

$$y = Xb + Zu + KMg + e$$

- $y$ => observations
- $b$ => effects of fixed factors
- $u$ => effects of a random polygenic factor
- $g$ => random marker effects
- $e$ => random residuals
- $X,\ Z,\ K$ => indcidence matrices linking effects to observations
- $M$ => marker matrix (dimensions "N-animals$\times$N-marker")

Note that $Mg$ yields genomic breeding values for every animal with a marker genotype.

## We are interested in:

$$p(b, u, g, \sigma_a^2, D, \sigma_e^2|y)$$

- $\sigma_a^2$=polygenic variance, $\sigma_e^2$=residual variance, $D$=diagonal matrix with marker variance, elements of $D$ may vary.
- We could infer about $b$, $u$, $g$, $\sigma_a^2$, $D$ and $\sigma_e^2$ by sampling directly from this distribution $\rightarrow$ usually impossible.

## Using the Bayesian paradigm:

$$p(b, u, g, \sigma_a^2, D, \sigma_e^2|y) \propto (y|b, u, g, \sigma_e^2)p(b)p(u|\sigma_a^2)p(\sigma_a^2)p(g|D)p(D)p(\sigma_e^2)$$

- $p(b, u, g, \sigma_a^2, D, \sigma_e^2|y) \rightarrow$ joint posterior distribution
- $(y|b, u, g, \sigma_e^2) \rightarrow$ likelihood of the data
- $p(b)$, $p(\sigma_a^2)$, $p(\sigma_e^2)$, $p(D) \rightarrow$ unconditional prior distributions
- $p(u|A\sigma_a^2)$, $p(g|D) \rightarrow$ conditional prior distributions
- prior distributions need to be defined (known) to make the Bayesian paradigm work

## Prior distribution

$$
\begin{aligned}
y|b,u,g,\sigma_e^2 &\sim N(Xb+Zu+KMg, I\sigma_e^2) && \textit{normal}\\
b &\sim \textit{constant}\\
u|A,\sigma_a^2 &\sim N(0, A\sigma_a^2) && \textit{normal}\\
g_i|D_i &\sim N(0, D_i) && \textit{normal}\\
\sigma_a^2 &\sim \nu_a S_a^2 \chi_{\nu_a}^{-2} && \textit{inverse chi}-\textit{square}\\
\sigma_e^2 &\sim \nu_e S_e^2 \chi_{\nu_e}^{-2} && \textit{inverse chi}-\textit{square}\\
D_i &\sim \nu_i S_i^2 \chi_{\nu_i}^{-2} && \textit{inverse chi}-\textit{square}
\end{aligned}
$$

# Fully conditional posterior distributions I

- simplify the joint posterior by forming a sequence of fully conditional posteriors assuming (pretending) that some parameters are know (assign starting values)
- fully conditional posteriors have usually a simpler form that the joint posterior distribution $\rightarrow$ sample directly

**assume that everything is known except $\Theta_i$, $\Theta = [b, u]'$**

$$\begin{aligned} p(\Theta_i|\sigma_a^2, \Theta_{j,j\neq i}, g, D, \sigma_e^2, y) &\propto p(y|b, u, g, \sigma_e^2)p(b)p(u|\sigma_a^2) \\ &\sim N(\hat{\Theta}_i', C_{i,i}^{-1}\sigma_e^2) \end{aligned}$$

Note that $C_{i,i}^{-1}$ is the diagonal element of the MME coefficient matrix. $\hat{\Theta}_i$ is obtained by solving the MME for $\Theta_i$ assuming that all other parameters are known

# Fully conditional posterior distributions II

assume that everything is known except $\sigma_a^2$

$$\begin{aligned} p(\sigma_a^2|u,b,g,D,\sigma_e^2,y) &\propto p(u|\sigma_a^2)p(\sigma_a^2) \\ &\sim \tilde{\nu}_a\tilde{S}_a^2\chi_{\tilde{\nu}_a}^{-2},\ \tilde{S}_a^2 = \frac{a'A^{-1}a+\nu_a S_a^2}{q+\nu_a} \end{aligned}$$

Note that $S_a^2$ and $\nu_a$ are so called "hyper-parameters" which represent prior knowledge. For example it can be a variance obtained in a different trial with $\nu_a$ degrees of freedom. Now you already see what we are doing in the fraction above: $\nu_a S_a^2$ calculates the sum of squares of that trial this it added to our sum of squares $a'A^{-1}a$. Then this total sum of squares is divided by the total degrees of freedom, which is our degrees of freedom $q$ and the degrees of freedom from the different trial $\nu_a$. Try to imagine how $\nu_a$ can dominate our results!!

# Fully conditional posterior distributions III

## assume that everything is know except $g_i$

$$\begin{aligned} p(g_i|b,u,g_{i+1,N},\sigma_a^2,D,\sigma_e^2,y) &\propto p(y|b,u,g,\sigma_e^2)p(g_i|D_i) \\ &\sim N(\hat{g}_i, C_{i,i}^{-1}\sigma_e^2) \end{aligned}$$

Note that $C_{i,i}$ is the diagonal element of the MME coefficient matrix at row/column of $g_i$.

## assume that everything is know except $D_i$

$$\begin{aligned} p(D_i|b,u,g,\sigma_a^2,D_{i+1,N},\sigma_e^2,y) &\propto p(g_i|D_i)p(D_i) \\ &\sim \tilde{\nu}_g\tilde{S}_g^2\chi_{\tilde{\nu}_g}^{-2},\ \tilde{S}_g^2 = \frac{g_i g_i + \nu_g S_g^2}{1+\nu_g} \end{aligned}$$

# Fully conditional posterior distributions IV

**assume that everything is know except $\sigma_e^2$**

$$\begin{aligned} p(\sigma_e^2|u,b,g,D,\sigma_e^2,y) &\propto p(y|b,u,g,\sigma_e^2)p(\sigma_e^2) \\ &\sim \tilde{\nu}_e\tilde{S}_e^2\chi_{\tilde{\nu}_e}^{-2},\ \tilde{S}_e^2 = \frac{e'e+\nu_e S_e^2}{q+\nu_e} \end{aligned}$$

See above for an explanation of $S_e^2$ and $\nu_e$.

# Gibbs sampling (Markov Chain Monte Carlo technique) I

## The problem

- Sampling from $p(x_i|x_{j,j=1..N,j\neq i})$ may not yield unbiased results because the outcome of sampling $x_i$, and therefore a parameter calculated from this samples (e.g. $\hat{x}_i$) may change if $x_j$ changes.

## The solution

- Precondition: all conditional posteriors can be defined
- Sample successively through the chain of conditional posteriors and replace old parameters by the sampled one.

# Gibbs sampling (Markov Chain Monte Carlo technique) II

## Example

$$y = Xb + Zu + KMg + e$$

The MME is then:

$$\begin{pmatrix} X'X & X'Z & X'KM \\ Z'X & Z'Z + A^{-1}\sigma_a^2 & Z'KM \\ M'K'X & M'K'Z & M'K'KM + D \end{pmatrix} \begin{pmatrix} \Theta_b \\ \Theta_u \\ \Theta_g \end{pmatrix} = \begin{pmatrix} X'y \\ Z'y \\ M'K'y \end{pmatrix}$$

$$C \qquad\qquad \Theta = R$$

# Gibbs sampling (Markov Chain Monte Carlo technique) III

## Example (continued)

- assign starting value to all elements in $\Theta$, $\sigma_a^2$, $\sigma_e^2$ and all elements in $D$
- for i in 1:length($\Theta$)
  1. cancel $\Theta_i$ by $\Theta_i = 0$
  2. calculate $\hat{\Theta}_i = \frac{R_i - C_{i,:}\Theta}{C_{i,i}}$
  3. draw a new $\Theta_i$ from $N(\hat{\Theta}_i, C_{i,i}^{-1}\sigma_e^2)$
- repeat iterating over $\Theta$ until convergence

- intermediate steps
  - if $\Theta_u$ is finished
    - calculate a $\hat{S}_a^2$ by $\Theta_u A^{-1} \Theta_u$
    - draw a new $\sigma_a^2$ from $\chi^{-2}(\hat{S}_a^2 + S_a^2\nu_a, \nu_a + n_u)$
  - when starting with $\Theta_g$
    - draw a new $\sigma_{g_i}^2$ from $\chi^{-2}(\hat{S}_{g_i}^2 + S_g^2\nu_g, \nu_g + 1)$
    - calculate a $\hat{S}_{g_i}^2$ by $\Theta_i^2$
    - Note that we draw for every single marker an own variance
  - when all elements of $\Theta$ are processed
    - calculate a $\hat{S}_e^2$ by $(y - Xb - Zu - KMg)'(y - Xb - Zu - KMg)$
    - draw a new $\sigma_e^2$ from $\chi^{-2}(\hat{S}_e^2 + S_e^2\nu_e, \nu_e + n_y)$

## Conclusion

- Gibb sampling $\rightarrow$ Markov Chain Monte Carlo Method
  - explores the joint space of all parameters in the model by sampling from conditional distributions
  - provides estimates for all parameters
  - parameters are more reliable than REML (likelihood surface)
  - it takes some time

# The Bayesian "Alphabet" I

## Naming background (convention??)

- founding publication about estimating marker effects via Markov Chain Monte Carlo $\rightarrow$ Meuwissen et. al 2001
- called their algorithm "BayesA" and "BayesB"
- science full of followers, subsequent developments $\rightarrow$ "BayesC", "BayesC$\pi$", "BayesD" and "BayesR"
- questions:
  - will "BayesZ" be the final invention??
  - what if we run out of letters??

# The Bayesian "Alphabet" II

## Differences

- Recall $p(g|D)$ and $p(D)$
- In full $p(D_{i,i}) = p(D_{i,i}|\nu, S^2)$ $\rightarrow$ conditional
  - making $p(g_i|D_{i,i})$ unconditional of $D_{i,i}$ yields the unconditional prior of $g_i$ $\rightarrow$ different for the different algorithms
- diagonal elements of $D$ are from different distributions

## BayesA

- all marker have an effect
- unconditional prior $\rightarrow$ t-distribution
- $D_{i,i}$ is drawn from inverse chi-square
- that's what we did in the example

# The Bayesian "Alphabet" III

## BayesB

- marker have no effect with probability $\pi$
- $\pi \rightarrow$ user defined
- unconditional prior for marker with effect $\rightarrow$ t distribution
- generating $\sigma^2_{g_i}$ from inverse chi-square

## BayesC$\pi$

- marker have no effect with probability $\pi$
- $\pi$ is sampled from $\beta$ distribution after all $g_i$ have been processed
- unconditional prior for marker with effect $\rightarrow$ t distribution
- $D_{i,i} = \sigma^2_g$.
- $\sigma^2_g$ is generated once from inverse chi-square after all $g_i$ have been processed

# The Bayesian "Alphabet" IV

## BayesR

- unconditional prior of marker is a mixture of normal distributions
  - $N(0, \sigma_1^2)$, $N(0, \sigma_j^2)$, ....., $N(0, \sigma_n^2)$, where $\sigma_1^2$=0
  - probability assigned to every distribution $\pi_1, .....\pi_n$, $\sum_j^n \pi_j = 1$
- for every single $g_i$
  - calculate $\epsilon_j$=p(y|$\sigma_j^2$)$\pi_j$ for all $j$
  - calculate $\phi_j = \frac{\epsilon_j}{\sum_j^n \epsilon_j}$
  - calculate $\Phi_j$
  - draw a uniform random number $\tau$ between zero and 1
  - assign that variance of distribution $j$ to $D_{i,i}$ where $\Phi_{j-1} < \tau < \Phi_{j+1}$
- after all $g_i$ have been processed
  - count the number of marker in each distribution $(c_1, ....., c_n)$
  - draw $\pi$ from a Dirichlet distribution $D(c_1, ...., c_n, K)$ where $K$ is prior knowledge about values in $c_1, ...., c_n$

# BESSiE I

## What is it:

a program for Best Linear Unbiased Prediction (BLUP) and Bayesian (MCMC) analysis of linear mixed models including genetic markers

## program algorithms

### mode BLUP

- "normal" BLUP
- GBLUP (replace $A^{-1}$ by $G^{-1}$)
- SNP BLUP (replace $A^{-1}$ by $D^{-1}$, diagonal elements in $D$ are $\sigma_a^2/N_{marker}$)
- single step BLUP (replace $A^{-1}$ by $H^{-1}$)

### mode GIBBS (Gibbs sampling)

- "normal" models without "Bayesian alphabet"
- BayesA
- BayesB
- BayesC$\pi$
- BayesR

# BESSiE II

## The global model in BESSiE

$$\begin{pmatrix} y_1 \\ . \\ y_n \end{pmatrix} = \begin{pmatrix} X_1 & 0 & 0 \\ 0 & . & 0 \\ 0 & 0 & X_n \end{pmatrix} \begin{pmatrix} b_1 \\ . \\ b_n \end{pmatrix} +$$

$$\begin{pmatrix} Z_{1,1} & . & Z_{1,k} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & . & . & . & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & Z_{n,1} & . & Z_{n,k} \end{pmatrix} \begin{pmatrix} u_{1,1} \\ . \\ u_{1,k} \\ . \\ . \\ . \\ u_{n,1} \\ . \\ u_{n,k} \end{pmatrix} +$$

$$+ \begin{pmatrix} K_1 M & 0 & 0 \\ 0 & . & 0 \\ 0 & 0 & K_n M \end{pmatrix} \begin{pmatrix} g_1 \\ . \\ g_2 \end{pmatrix} + \begin{pmatrix} e_1 \\ . \\ e_2 \end{pmatrix}$$

# BESSiE III

## possible random factors in the model

- NRM $\sim N([0,..,0]'\Sigma \otimes A)$ ($A \rightarrow$ pedigree derived relationship matrix)
- GRM $\sim N([0,..,0]'\Sigma \otimes G)$ ($G \rightarrow$ marker based relationship matrix)
- Single step $\sim N([0,..,0]'\Sigma \otimes H)$ ($H \rightarrow$ combination of $A$ and $G$ if some individual are not genotyped)
- IDE $\sim N([0,..,0]'\Sigma \otimes I)$
- external $\sim N([0,..,0]'\Sigma \otimes K)$ ($K \rightarrow$ a user defined matrix)
- genetic groups
- SNP $\sim N([0,..,0]'\Sigma \otimes I)$
  - $\Sigma \rightarrow D$ ($D$ is diagonal, its elements are derived via "Bayesian Alphabet" or a fraction of the total genetic variance("SNP_BLUP"))

# BESSiE IV

## possible fixed factors in the model

- dummy (mean, contemporary group etc.)
- co-variable (age, weight etc., polynomial user-defined (e.g. $age^1$+$age^2$ ....))
- genetic groups

## possible phenotypes

- continuous
- binary (0,1)
- categorical (0,1,2,....,k)
- every combination of these phenotypes
- weighted observations (e.g. breeding values)

# BESSiE V

## output

- default:
  - Logfile only
- to be switched on:
  - sampled/solved factor level solutions (e.g. marker effects, animal effects) and/or their means (asii, binary)
  - sampled variances for random factors (e.g. additive genetic variance) or their means (asii, binary)
  - sampled marker variances
  - distribution counter (BayesR, BayesC$\pi$)
  - distribution probabilities (BayesR, BayesC$\pi$)

# BESSiE VI

## What else:

- no limits
  - unlimited number phenotypes
  - unlimited number traits
  - unlimited number of factors
  - unlimited number of marker

Its just a matter of time!!!