GENOME-ENABLED PREDICTION WITH THE BAYESIAN ALPHABET

BAYES A, BAYES B, Bayes C-pi,
BAYESIAN LASSO, BAYES R
(many other methods. These are prototypical)

Gianola, D., G. de los Campos, W. G. Hill, E. Manfredi, and R. Fernando (2009) Additive genetic variability and the Bayesian alphabet. Genetics 183: 347-363.

Gianola, D. (2013) Priors in whole-genome regression: the Bayesian alphabet returns. Genetics 194: 573-596

Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps

T. H. E. Meuwissen,* B. J. Hayes† and M. E. Goddard†,‡

Genetics 157: 1819–1829 (April 2001)

ABSTRACT

Recent advances in molecular genetic techniques will make dense marker maps available and genotyping many individuals for these markers feasible. Here we attempted to estimate the effects of ~50,000 marker haplotypes simultaneously from a limited number of phenotypic records. A genome of 1000 cM was simulated with a marker spacing of 1 cM. The markers surrounding every 1-cM region were combined into marker haplotypes. Due to finite population size (N_e = 100), the marker haplotypes were in linkage disequilibrium with the QTL located between the markers. Using least squares, all haplotype effects could not be estimated simultaneously. When only the biggest effects were included, they were overestimated and the accuracy of predicting genetic values of the offspring of the recorded animals was only 0.32. Best linear unbiased prediction of haplotype effects assumed equal variances associated to each 1-cM chromosomal segment, which yielded an accuracy of 0.73, although this assumption was far from true. Bayesian methods that assumed a prior distribution of the variance associated with each chromosome segment increased this accuracy to 0.85, even when the prior was not correct. It was concluded that selection on genetic values predicted from markers could substantially increase the rate of genetic gain in animals and plants, especially if combined with reproductive techniques to shorten the generation interval.

CRUCIAL CONTRIBUTIONS OF THE PAPER

- Use all markers in a linear regression model as opposed to just a few ones
- Shift attention from a doubtful emphasis on QTL (now superseded by the GWAS obsession) search to genome-enabled prediction
- Hint at the possibility of obtaining earlier and more accurate predictions of genetic values
- Use of cross-validation: something that had not received much emphasis before (Utz et al. 2000 in plant breeding)

WE ARRIVED TOO LATE….

On Marker-Assisted Prediction of Genetic Value: Beyond the Ridge

Daniel Gianola,*[,1] Miguel Perez-Enciso[†] and Miguel A. Toro[‡]

Genetics 163: 347–365 (January 2003)

ABSTRACT

Marked-assisted genetic improvement of agricultural species exploits statistical dependencies in the joint distribution of marker genotypes and quantitative traits. An issue is how molecular (*e.g.*, dense marker maps) and phenotypic information (*e.g.*, some measure of yield in plants) is to be used for predicting the genetic value of candidates for selection. Multiple regression, selection index techniques, best linear unbiased prediction, and ridge regression of phenotypes on marker genotypes have been suggested, as well as more elaborate methods. Here, phenotype-marker associations are modeled hierarchically via multilevel models including chromosomal effects, a spatial covariance of marked effects within chromosomes, background genetic variability, and family heterogeneity. Lorenz curves and Gini coefficients are suggested for assessing the inequality of the contribution of different marked effects to genetic variability. Classical and Bayesian methods are presented. The Bayesian approach includes a Markov chain Monte Carlo implementation. The generality and flexibility of the Bayesian method is illustrated when a Lorenz curve is to be inferred.

Genome-enabled prediction using linear regression models

STARTED IN ANIMAL BREEDING BUT PLANT BREEDERS EMBRACED IT WITH JOY!!

Meuwissen, Hayes and Goddard (2001):
"Genomic selection"
Perhaps better terms:
"Genome-enabled selection"
"Genome-assisted selection"
PREFER NOT TO USE "Genomic prediction"

Prototypical linear regression model (no nuisance parameters)

SNP effects combined additively

Effect of chromosomal segment alleles, haplotypes

$$\mathbf{y} = \mu \mathbf{1}_n + \Sigma_i \mathbf{X}_i \mathbf{g}_i + \mathbf{e},$$

$$\frac{\partial \mathbf{y}}{\partial \mathbf{g}_i} = \mathbf{X'}_i$$

Linear model in parameters

QUESTIONS:
ABANDON QTLS, PEDIGREES, KNOWN GENES?
"Whole-genome prediction" (whole or part? Variable selection?)

THE DESIGN OF METHODS FOR PREDICTION IS SOMEWHAT "DIFFERENT" THAN FOR INFERENCE

IS MODEL *M* "RIGHT"? θ CAN BE A FUTURE PHENOTYPE

TAKING MODEL UNCERTAINTY INTO ACCOUNT BY MODEL AVERAGING

$$p(\theta|y) = \sum p(\theta|y,M)p(M|y)$$

$$= \int p(\theta|y,M)p(M|y)dM$$

FEW MODELS

MANY MODELS

Posterior after Accounting for model uncertaintiy

Posterior distribution or density for model *M*

Posterior probability or density of model

VARIANCE OF PREDICTAND TAKING MODEL UNCERTAINTY INTO ACCOUNT

$$Var(\theta|y) = E_M[Var(\theta|y,M)] + Var[E_M[\theta|y,M]]$$

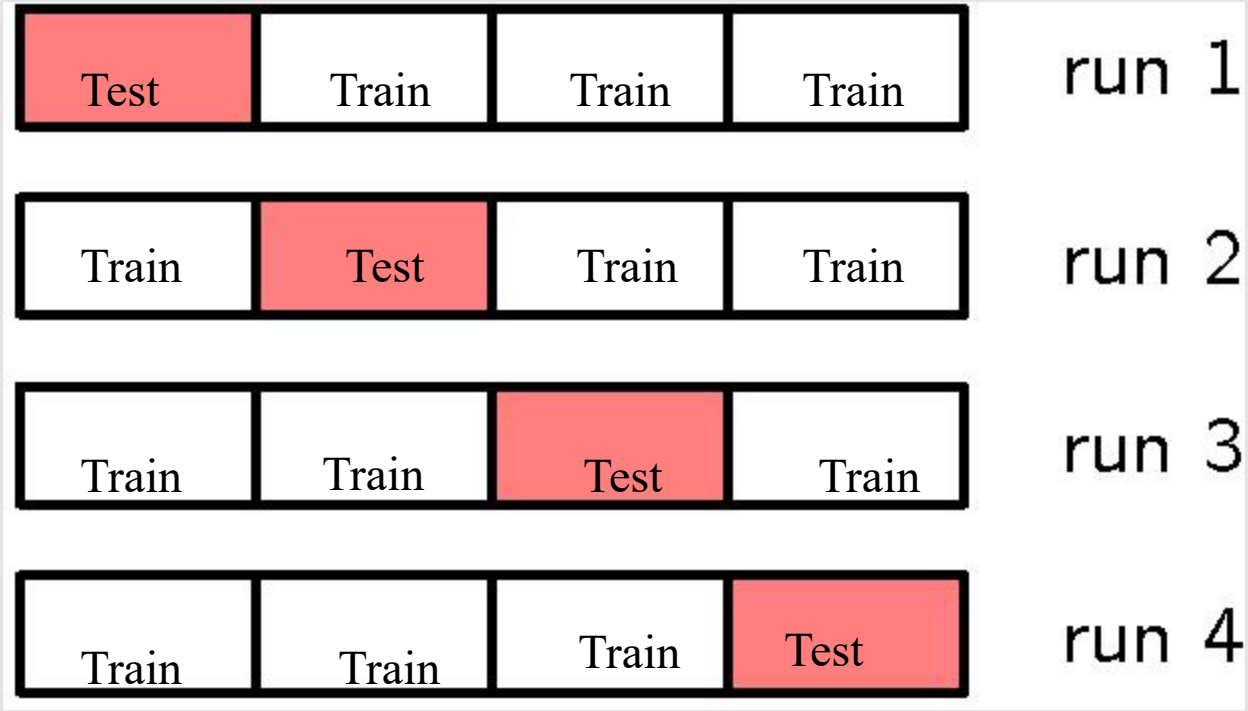
Average "prediction Error" variance

Variance among predictions from different models

CROSS-VALIDATION (CV)

- Data available (genomic, phenotypic)
- Data generated according to unknown process
- Split into training (fitting)- testing (predictand) sets
- Fitting process describes current data (model is typically wrong). Sample may be idiosyncratic
- Use training process to make statement about yet-to-be observed data (testing set)
- Prediction error (conditional and unconditional): point estimate is obtained
- Distribution of prediction errors (conditional or unconditional): interval estimate. For this, CV must be replicated

ILLUSTRATION OF A 4-FOLD CROSS-VALIDATION (red: testing set; white: training set)



ALGORITHM:

1: Choose a loss function L (e.g., mean squared error between predicted and observed outcome
2: Choose a set of training-testing splits (K=4)
3: Choose a set of regularization parameter values $\tau_1, \tau_2, \ldots, \tau_A$
4: **for** a=1 to A **do**
5: **for** k=1 to K **do**
6: train model and find "best" parameter estimates corresponding to τ_a
7: **end for**
8: $L(\tau_a) = \frac{\sum_1^K L_k(\tau_a)}{K}$
9: **end for**
10: $\tau_{opt} = \arg min_{\tau_A}(L(\tau_a))$

Important

This is a single-realization from the CV distribution. Repeat many times! (data structure issues here)

CROSS-VALIDATION

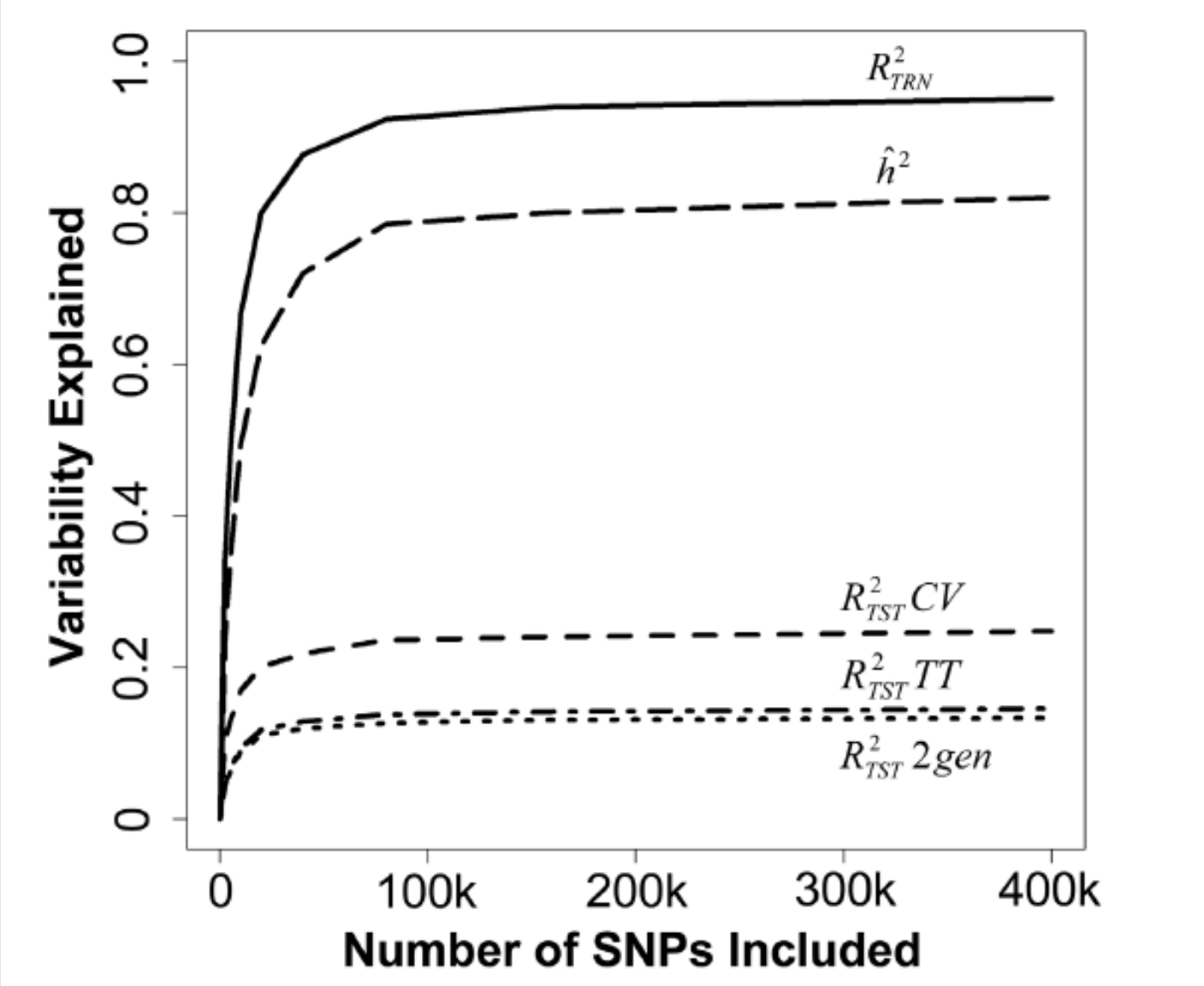
seldom done in animal breeding in the pre-genomic era. Often absent in GWAS and medical studies)

➔A. Prediction and goodness of fit are different: a model that fits well to training data may predict badly. A mechanistically poor model can give better predictions that "fancier" models

➔B. Any cross-validation scheme (e.g., k-folds) has a cross-validation distribution

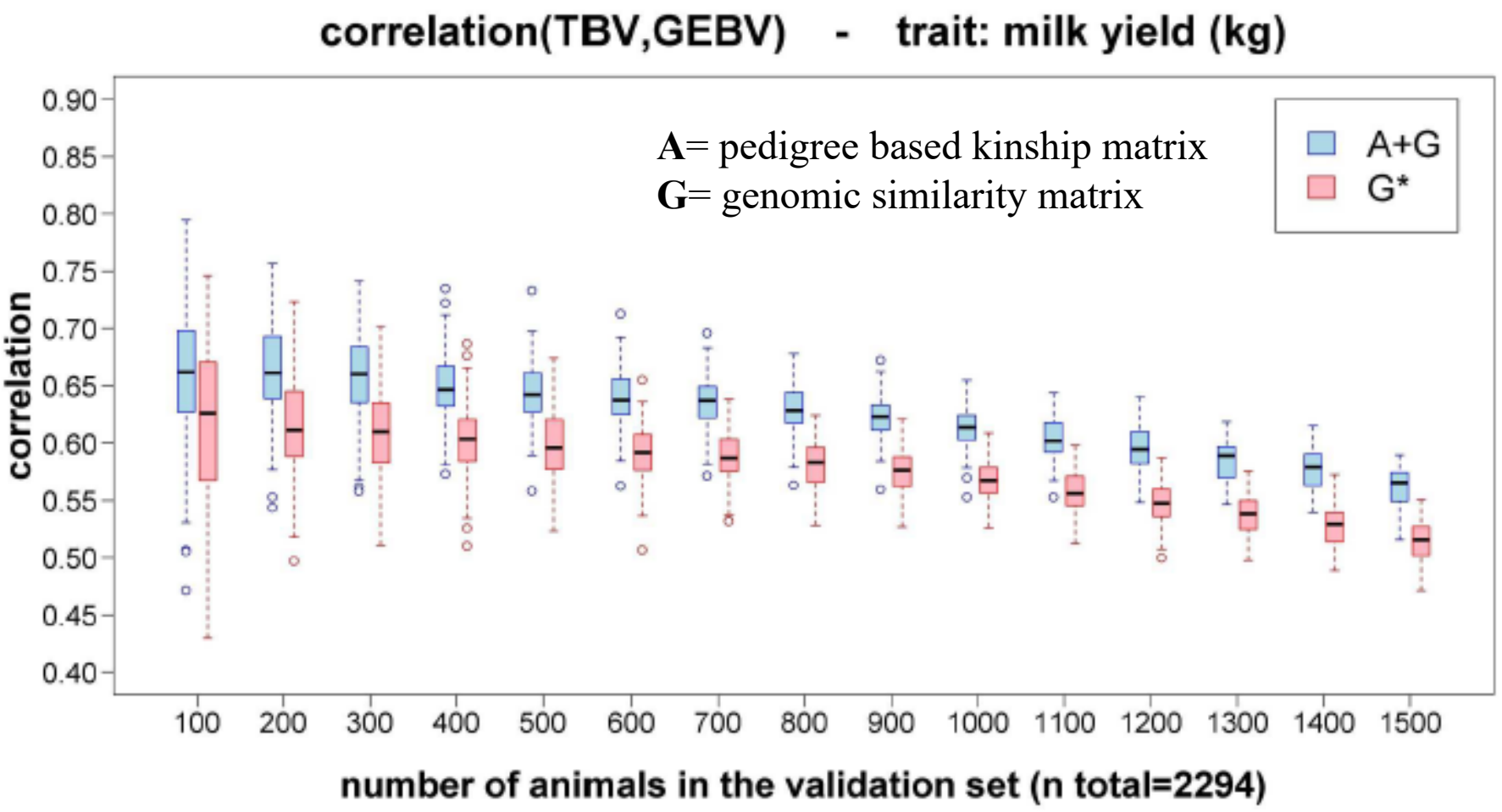
THIS IS THE DISTRIBUTION THAT MATTERS AND NOT THAT BASED ON THEORETICAL CONSIDERATIONS FROM SOME MODEL

GOODNESS OF FIT (TRAINING= TRN) vs. **PREDICTIVE ABILITY** (TESTING= TST)



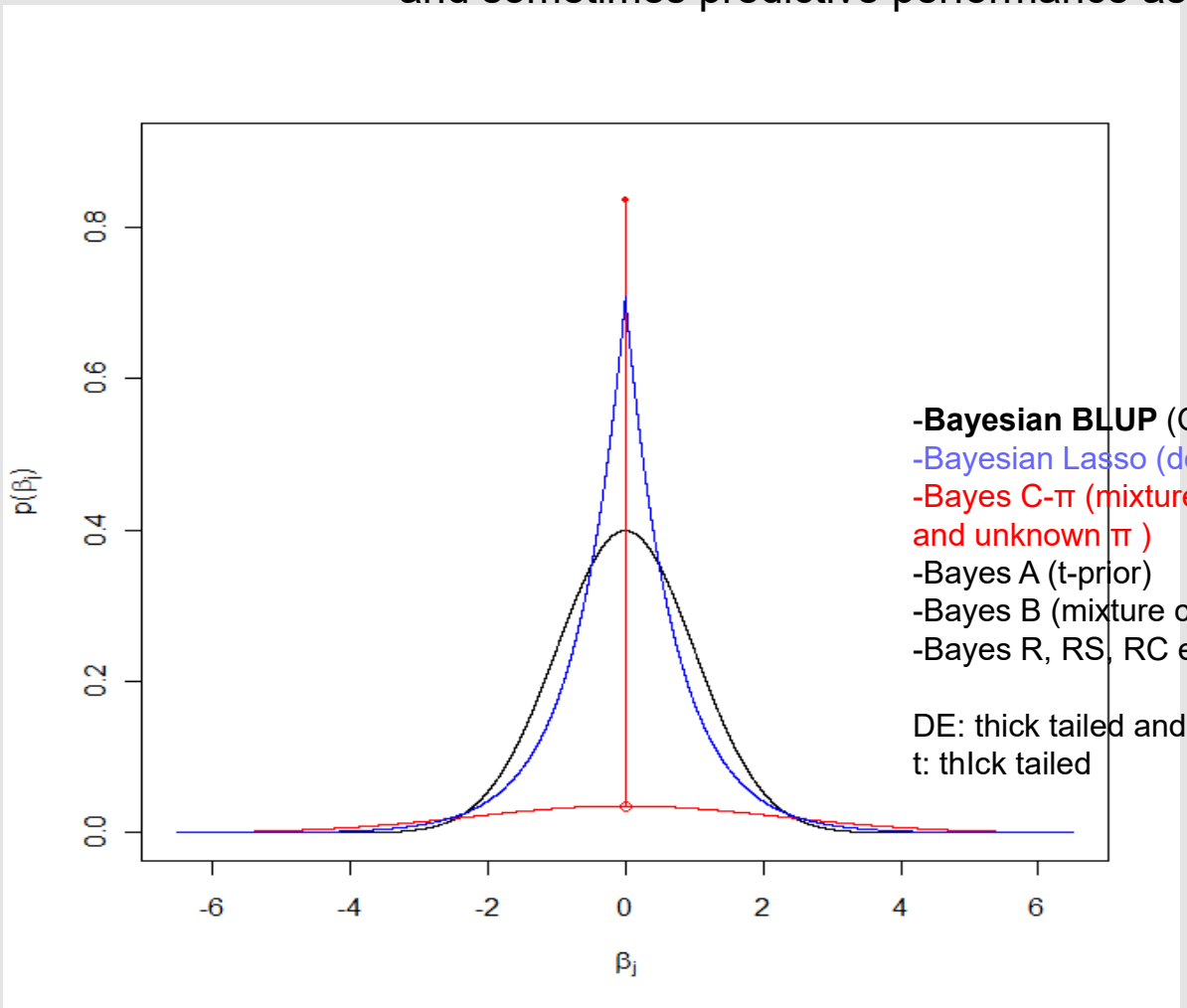
HUMAN STATURE: MAKOWSKY et al. , Plos Genetics 2011

CROSS-VALIDATION UNCERTAINTY AND IMPACT OF LAYOUT:
2294 dairy bulls with progeny tests ("TBV")
(Erbe et al. 2010)



"THE BAYESIAN ALPHABET"

Series of Bayesian multiple linear regression models
that just differ in the **conditional** prior adopted at some level of the hierarchical model. All such priors and hyper-parameters INFLUENCE inference whenever n<<p
and sometimes predictive performance as well



EXAMPLES

-**Bayesian BLUP** (Gaussian conditional prior)
-Bayesian Lasso (double exponential conditional prior)
-Bayes C-π (mixture conditional prior with Gaussian component and unknown π)
-Bayes A (t-prior)
-Bayes B (mixture conditional prior with t- component and know π
-Bayes R, RS, RC even greedier mixtures

DE: thick tailed and leptokurtic
t: thIck tailed

Density of a standard normal random variable (black), of a double-exponential random variable (blue) and of a random variable following a mixture density with a mass point at zero (with probability 0.8) and a Gaussian process with probability 0.2. All variables with zero mean and variance equal to one.

IMPORTANT ISSUE HERE:

The Bayesian statement about marker effects is the marginal and not the conditional prior!

What are your Bayesian beliefs about marker effects?

HIERARCHICAL MODEL: SUPPOSE θ_1 ARE MARKER EFFECTS

MODEL INVOLVES: $\theta_1, \theta_2, \theta_3, \theta_4$

$$p(\theta_1, \theta_2, \theta_3, \theta_4 | H)$$

$$= p(\theta_1 | \theta_2, \theta_3, \theta_4, H) p(\theta_2 | \theta_3, \theta_4, H) p(\theta_3 | \theta_4, H) p(\theta_4, H)$$

Your Bayesian beliefs about marker effects depends on the values of θ2, θ3 and θ4

Since you are also uncertain about θ2, θ3 and θ4, your beliefs about markers are conveyed by the marginal prior

$$p(\theta_1 | H)$$

$$= \int\int\int p(\theta_1 | \theta_2, \theta_3, \theta_4, H) p(\theta_2 | \theta_3, \theta_4, H) p(\theta_3 | \theta_4, H) p(\theta_4, H) d\theta_2 d\theta_3 d\theta_4$$

THE CURSE OF THE BAYESIAN ALPHABET

Genetics: Early Online, published on July 9, 2014 as 10.1534/genetics.114.164442

Genome-wide Regression & Prediction with the BGLR statistical package

Paulino Pérez
Socio Economía Estadística e Informática,
Colegio de Postgraduados, México
perpdgo@colpos.mx

Gustavo de los Campos
Department of Biostatistics, Section on Statistical Genetics
University of Alabama at Birmingham, USA
gcampos@uab.edu

CAN FIT MOST MEMBERS OF THE "ALPHABET WITH THIS R PACKAGE

Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps

T. H. E. Meuwissen,* B. J. Hayes† and M. E. Goddard†,‡

Genetics 157: 1819–1829 (April 2001)

BAYES A + BAYES B
(both pose the same data-generating model)

(as I understand these two methods)

Linear model proposed by Meuwissen et al. (2001)

Code for genotype of SNP j: x= -1,0,1

Additive substitution effect of SNP j

SCALAR

$$y_i = \mu + \sum_{j=1}^{p} x_{ij} b_j + e_i,$$

$$i = 1, 2, \ldots, n; \quad n << p$$

$$y_i | \mu, \mathbf{x}_i, \mathbf{b}, \sigma_e^2 \sim N\left(\mu + \sum_{j=1}^{p} x_{ij} b_j, \sigma_e^2 \right)$$

MATRIX

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{X}\mathbf{b} + \mathbf{e},$$

$$\mathbf{y} | \mu, \mathbf{X}, \mathbf{b} \sim N(\mathbf{1}\mu + \mathbf{X}\mathbf{b}, \mathbf{I}\sigma_e^2)$$

The priors for Bayes A

Note: this is a **conditional** prior
Not true that each marker has different prior
(as claimed in Meuwissen et al. (2001))

$$\mu \sim uniform$$

$$\sigma_e^2 | \nu_e, S_e^2 \sim \nu_e S_e^2 \chi_{\nu_e}^{-2}$$

$$b_j | \sigma_{b_j}^2 \sim N\left(0, \sigma_{b_j}^2\right); j = 1,2,\ldots,p$$

$$\sigma_{b_j}^2 | \upsilon, S^2 \sim \nu_b S_b^2 \chi_{\nu_b}^{-2}; j = 1,2,\ldots,p$$

Hyper-parameters, specified arbitrarily

MARGINAL PRIOR OF MARKER EFFECTS IN BAYES A (Meuwissen et al., 2001)

$$b_j|\sigma_j^2 \sim N\left(0,\sigma_j^2\right) \qquad j=1,2,\ldots,p$$

$$\sigma_j^2|\upsilon,S^2 \sim \upsilon S^2\,\chi_\upsilon^{-2}$$

Marginal prior

$$p(b_j|\upsilon,S^2) = \int_0^\infty N\left(0,\sigma_j^2\right)p(\upsilon S^2\chi_\upsilon^{-2})d\sigma_j^2$$

Kernel of Gaussian

$$\int_0^\infty \left(\sigma_j^2\right)^{-\frac{1}{2}}\exp\left(-\frac{b_j^2}{\sigma_j^2}\right)\left(\sigma_j^2\right)^{-\left(\frac{\upsilon+2}{2}\right)}\exp\left[-\frac{\upsilon S^2}{\sigma_j^2}\right]d\sigma_j^2$$

These hyper-parameters will control the extent of shrinkage. Question: does their influence vanish asymptotically?

$$\propto \int_0^\infty \left(\sigma_j^2\right)^{-\frac{1+\upsilon+2}{2}}\exp\left(-\frac{b_j^2+\upsilon S^2}{\sigma_j^2}\right)d\sigma_j^2$$

$$\propto \Gamma\left(\frac{1+\upsilon}{2}\right)\left(b_j^2+\upsilon S^2\right)^{-\frac{\upsilon+1}{2}}$$

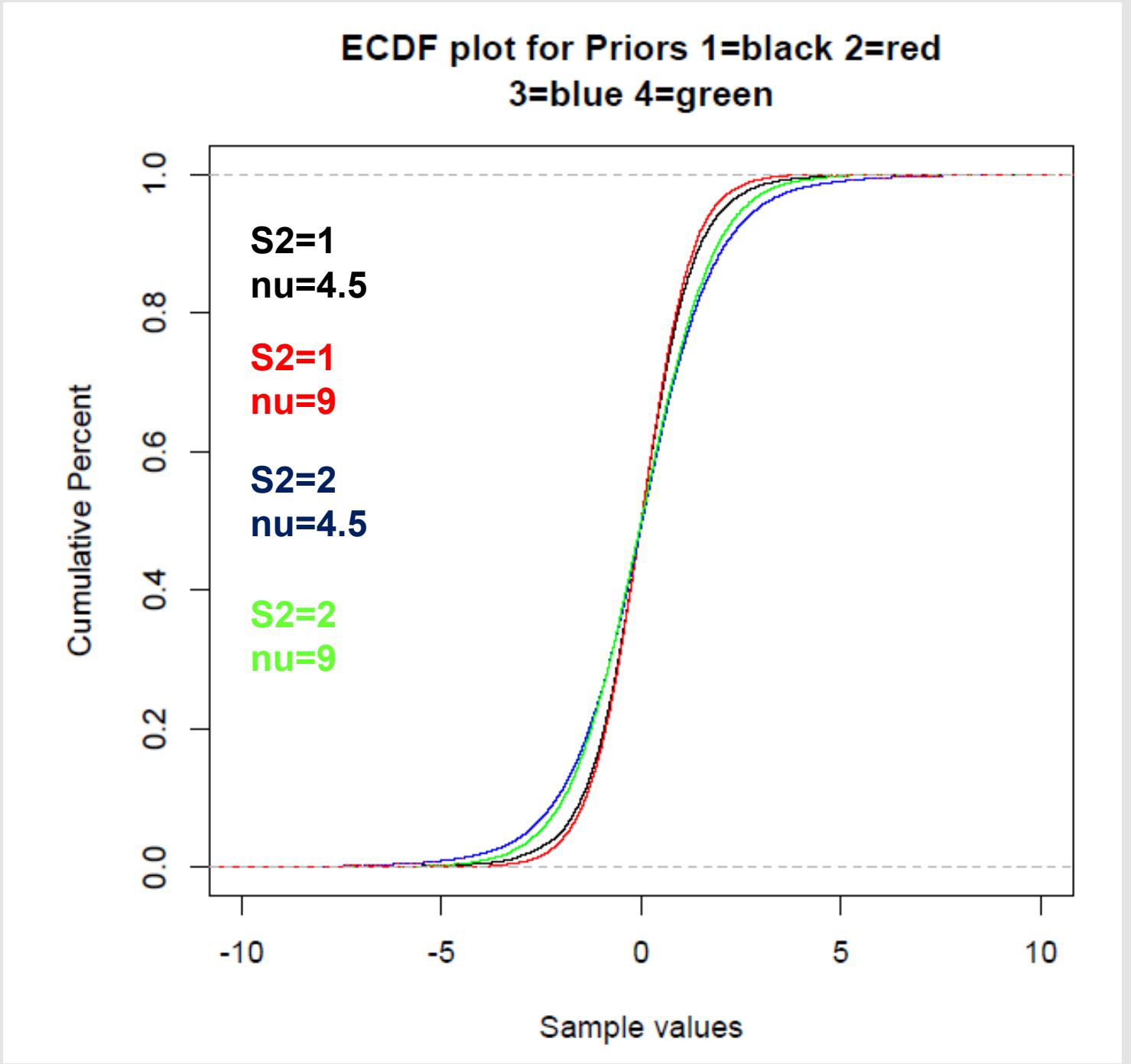
$$\propto \left(1+\frac{b_j^2}{\upsilon S^2}\right)^{-\frac{\upsilon+1}{2}} \Rightarrow \boxed{t(0,\upsilon,S^2)}$$

The prior of a marker effect is a *t*-distribution with known scale and df

MARGINALLY: IN BAYES A ALL MARKERS HAVE THE SAME VARIANCE

```
###FOUR PRIORS FOR BAYES A

###PRIOR 1 SCALE INVERTED CHI-SQUARE S2=1, nu=4.5
###PRIOR 2 SCALE INVERTED CHI-SUQARE S2=1, nu=9
###PRIOR 3 SCALE INVERTED CHI-SQUARE S2=2, nu=4.5
###PRIOR 4 SCALE INVERTED CHI-SQUARE S2=2, NU=9
###IN ALL CASES, CONDITIONAL PRIOR FOR b is N(0,var)

######
S2=1
nu=4.5
var<-numeric(25000)
var<-rchisq(25000,nu)
var<-nu*S2/var

b1<-numeric(length(var))

for (i in 1:length(var)){
b1[i]<-rnorm(1,0,sqrt(var[i]))
}
#######
S2=1
nu=9

var<-rchisq(25000,nu)
var<-nu*S2/var

b2<-numeric(length(var))

for (i in 1:length(var)){
b2[i]<-rnorm(1,0,sqrt(var[i]))
}
```

```
###FOUR PRIORS FOR BAYES A

###PRIOR 1 SCALE INVERTED CHI-SQUARE S2=1, nu=4.5
###PRIOR 2 SCALE INVERTED CHI-SUQARE S2=1, nu=9
###PRIOR 3 SCALE INVERTED CHI-SQUARE S2=2, nu=4.5
###PRIOR 4 SCALE INVERTED CHI-SQUARE S2=2, nu=9
###IN ALL CASES, CONDITIONAL PRIOR FOR b is N(0,var)

#######
S2=2
nu=4.5

var<-rchisq(25000,nu)
var<-nu*S2/var

b3<-numeric(length(var))

for (i in 1:length(var)){
b3[i]<-rnorm(1,0,sqrt(var[i]))
}
#######
S2=2
nu=9

var<-rchisq(25000,nu)
var<-nu*S2/var

b4<-numeric(length(var))

for (i in 1:length(var)){
b4[i]<-rnorm(1,0,sqrt(var[i]))
}
```


ECDF plot for Priors 1=black 2=red
3=blue 4=green
S2=1
nu=4.5
S2=1
nu=9
S2=2
nu=4.5
S2=2
nu=9
Cumulative Percent
0.0
0.2
0.4
0.6
0.8
1.0
-10
-5
0
5
10
Sample values

A Gibbs sampler for Bayes A
(see BGLR)

(element-wise sampling)

Note: the form of the implementation is just an algorithmic matter: it is immaterial with respect to model-related issues

Sampling the mean

$$\mu|ELSE \sim N\left[\frac{1}{n} \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{p} x_{ij} b_j \right), \frac{\sigma_e^2}{n} \right]$$

Flat prior for the mean (or for the fixed effects) is not influential

Sampling the residual variance

$$\sigma_e^2 | ELSE \sim n\left(1 + \frac{\nu_e}{n}\right) \left(\frac{\sum_{i=1}^{n} \left(y_i - \mu - \sum_{j=1}^{p} x_{ij} b_j \right)^2 + \nu_e S_e^2}{n + \nu_e} \right) \chi^{-2}_{\nu_e + n}$$

Goes to *n*

The prior can be "killed" simply by increasing sample size

This will dominate the weighted average as *n* increases

Sampling the marker effects

$$b_j|ELSE \sim N\left[\frac{\sum_{i=1}^{n} x_{ij}\left(y_i - \mu - \sum_{j'=1}^{p} x_{ij}b_j\right)}{\sum_{i=1}^{n} x_{ij}^2 + \frac{\sigma_e^2}{\sigma_{b_j}^2}}, \frac{\sigma_e^2}{\sum_{i=1}^{n} x_{ij}^2 + \frac{\sigma_e^2}{\sigma_{b_j}^2}}\right]$$

$$j = 1, 2, \ldots, p$$

Kill the prior simply by increasing sample size. The effect of the shrinkage ratio vanishes

$$\sum_{i=1}^{n} x_{ij}^2 + \frac{\sigma_e^2}{\sigma_{b_j}^2} \rightarrow \sum_{i=1}^{n} x_{ij}^2$$

Sampling the variance of the marker effects

$$\sigma^2_{b_j}|ELSE \sim \nu\left(1+\frac{1}{\nu}\right)\left(\frac{b_j^2+\nu S^2}{1+\nu}\right)\chi^{-2}_{\nu+1}$$

Typically very small

Prior df: very influential

$$= \nu\left(1+\frac{1}{\nu}\right)S^2\left(\left[\frac{\left(\frac{b_j}{S}\right)^2+\nu}{1+\nu}\right]\right)\chi^{-2}_{\nu+1}$$

$$j = 1,2,\ldots,p$$

- **Prior cannot be killed here**. Can increase number of data points or of markers *ad nauseum* and gain only one degree of freedom, **always**
- Recall that, in the conditional posterior, all other parameters are known (i.e., they are assigned values)
- Since one must de-condition, actually the true posterior moves "less than one degree of freedom" away from the prior
- Meuwissen et al. (2001) mistakenly thought that learning about $\sigma^2_{b_j}$ would inform about "genetic variance architecture"

A WAY OF LOOKING AT SHRINKAGE OF MARKER EFFECTS IN BAYES A

$$y_i|\boldsymbol{\beta},\sigma_e^2 \sim N\left(y_i|\mathbf{x}_i'\boldsymbol{\beta},\sigma_e^2\right), i=1,2,...,n;\ \beta_j|S_\beta^2,\nu \sim IID\ t\left(\beta_j|0,S_\beta^2,\nu\right),\ j=1,2,...,p,$$

where $\mathbf{x}_i'$ is the i^{th} row of $\mathbf{X}$. Conditionally on σ_e^2, S_β^2 and ν, the joint posterior density is

$$p\left(\boldsymbol{\beta}|S_\beta^2,\nu,\sigma_e^2,\mathbf{y}\right) \propto \prod_{i=1}^{n}\exp\left[-\frac{1}{2\sigma_e^2}\left(y_i-\mathbf{x}_i'\boldsymbol{\beta}\right)^2\right]\prod_{j=1}^{p}\left[1+\frac{\beta_j^2}{S_\beta^2\nu}\right]^{-\frac{1+\nu}{2}}. \quad (7)$$

Using results presented in the Appendix ("Mode of the conditional posterior distribution in Bayes A"), an iterative scheme for locating a mode of (7) is given by

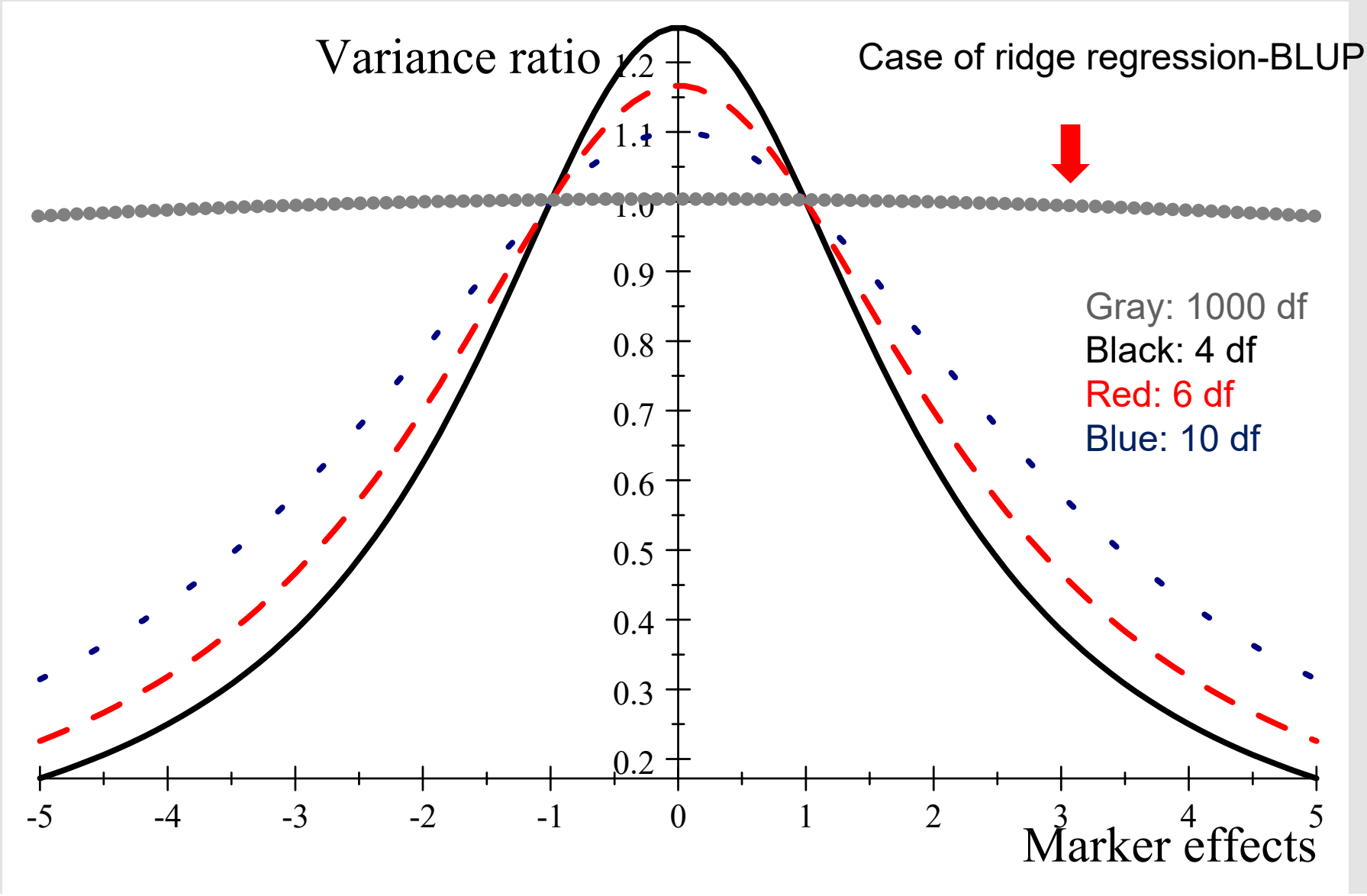
Gianola (2013)

$$\boldsymbol{\beta}^{[t+1]} = \left(\mathbf{X}'\mathbf{X}+\mathbf{W}_\beta^{[t]}\right)^{-1}\mathbf{X}'\mathbf{y} = \left(\mathbf{X}'\mathbf{X}+\mathbf{W}_\beta^{[t]}\right)^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta}^{(0)} \quad (8)$$

with successive updating; here, $\mathbf{W}_\beta^{[t]} = Diag\left\{\frac{\sigma_e^2}{S_\beta^2}\frac{\left(1+\frac{1}{\nu}\right)}{\left(1+\frac{\beta_j^{2[t]}}{S_\beta^2\nu}\right)}\right\}$ is a diagonal matrix. If this converges, it will do so to one of perhaps many stationary points, as it is known that $t-$regression models may produce multi-modal log-posterior surfaces, especially if ν is small (McLachlan and Krishnan 1997). Hence, iteration

Markers with tiny effects are shrunk strongly)

Impact of marker effect and of degrees of freedom (df) on shrinkage towards prior distribution in Bayes A: the larger the value in the y-axis, the stronger the shrinkage towards 0. df=4: solid line; df=6: dashed line; df=10: dotted line; df=1000: thick line, almost horizontal .



Bayes B as formulated originally

Bayes B assumptions
(Bayesianly strange)

$$b_j|\sigma_j^2 \sim \begin{cases} \text{point mass at some constant } k \text{ if } \sigma_j^2 = 0 \\ N\left(0, \sigma_j^2\right) \text{ if } \sigma_j^2 > 0 \end{cases}$$

$$\sigma_j^2|\pi = \begin{cases} 0 \text{ with probability } \pi \\ \upsilon S^2 \chi_{\nu}^{-2} \text{ with probability } 1-\pi \end{cases}$$

3. Recall: if a prior variance is 0, this means complete certainty

1. Meuwissen takes the constant = 0

2. Meuwissen assumes π is known, e.g., 0.95

Joint density:

$$p\left(b_j, \sigma_j^2 | \pi\right) = \begin{cases} b_j = k \text{ and } \sigma_j^2 = 0 \text{ with probability } \pi \\ N\left(0, \sigma_j^2\right) p(\upsilon S^2 \chi_\upsilon^{-2}) \text{ with probability } 1 - \pi \end{cases}$$

Marginal prior

$$p(b_j | \pi) = \begin{cases} b_j = k \text{ with probability } \pi \\ \int_0^\infty N\left(0, \sigma_j^2\right) p(\upsilon S^2 \chi_\upsilon^{-2}) d\sigma_j^2 \text{ with probability } 1 - \pi \end{cases}$$

Further

$$\int_0^\infty \left(\sigma_j^2\right)^{-\frac{1}{2}} \exp\left(-\frac{b_j^2}{\sigma_j^2}\right)\left(\sigma_j^2\right)^{-\left(\frac{\upsilon+2}{2}\right)} \exp\left[-\frac{\upsilon S^2}{\sigma_j^2}\right] d\sigma_j^2$$

$$= \int_0^\infty \left(\sigma_j^2\right)^{-\frac{1+\upsilon+2}{2}} \exp\left(-\frac{b_j^2 + \upsilon S^2}{\sigma_j^2}\right) d\sigma_j^2$$

$$= \Gamma\left(\frac{1+\upsilon}{2}\right)\left(b_j^2 + \upsilon S^2\right)^{-\frac{\upsilon+1}{2}}$$

$$\propto \left(1 + \frac{b_j^2}{\upsilon S^2}\right)^{-\frac{\upsilon+1}{2}} \Rightarrow t(0, \upsilon, S^2)$$

Then:

PRIOR = MIXTURE OF A POINT MASS AND OF A *t*-DISTRIBUTION. BAYES B PUTS THE MASS AT 0 (IF NOT 0, THIS GETS ABSORBED INTO THE GENERAL MEAN)

$$p(b_j|\pi) = \begin{cases} b_j = k \text{ with probability } \pi \\ t(0, \upsilon, S^2) \text{ with probability } 1 - \pi \end{cases}$$

MARGINALLY: ALL MARKERS HAVE THE SAME DISTRIBUTION

Mean and variance of a mixture (e.g., Gianola et al. 2006, Genetics)

The first and second moments, and the variance of a finite mixture of K Gaussian distributions, with parameters $\boldsymbol{\theta} = [P_1, \ldots, P_K, \mu_1, \ldots, \mu_K, \sigma_1^2, \ldots, \sigma_K^2]'$, where the mixture proportions P_k are such that $\sum_{k=1}^{K} P_k = 1$, are

$$E(y|\boldsymbol{\theta}) = \int y \left[\sum_{k=1}^{K} P_k N(y|\mu_k, \sigma_k^2) \right] dy = \sum_{k=1}^{K} P_k \mu_k, \quad \text{(A1)}$$

$$E(y^2|\boldsymbol{\theta}) = \int y^2 \left[\sum_{k=1}^{K} P_k N(y|\mu_k, \sigma_k^2) \right] dy = \sum_{k=1}^{K} P_k (\mu_k^2 + \sigma_k^2),$$

$$\text{Var}(y|\boldsymbol{\theta}) = \sum_{k=1}^{K} P_k \sigma_k^2 + \sum_{k=1}^{K} P_k \mu_k^2 - \left(\sum_{k=1}^{K} P_k \mu_k \right)^2.$$

In Bayes B:

$$E(b_j|\pi) = \pi k + (1-\pi)0 = \pi k$$

$$\Rightarrow 0 \text{ if } k = 0$$

$$Var(b_j|\pi) = \pi \times 0 + (1-\pi)\frac{S^2\nu}{\nu-2} + \pi k^2 + (1-\pi)0^2 - (\pi k)^2$$

$$= (1-\pi)\frac{S^2\nu}{\nu-2} + \pi k^2(1-\pi)$$

$$= (1-\pi)\frac{S^2\nu}{\nu-2} \text{ if } k = 0$$

ALL MARKERS HAVE THE SAME PRIOR MEAN AND VARIANCE IN BAYES B!

###FOUR PRIORS FOR BAYES B
###IN ALL CASES PI=0.95
###PRIOR 1 SCALE INVERTED CHI-SQUARE S2=1, nu=4.5
###PRIOR 2 SCALE INVERTED CHI-SUQARE S2=1, nu=9
###PRIOR 3 SCALE INVERTED CHI-SQUARE S2=2, nu=4.5
###PRIOR 4 SCALE INVERTED CHI-SQUARE S2=2, nu=9
###IN ALL CASES, CONDITIONAL PRIOR FOR b is N(0,var)

####BELOW WE GENERATE SAMPLES FROM T-DISTRIBUTION
####BUT WE SET THE MARKERS TO 0 IF U(0,1)<=0.95
pie=0.95

S2=1
nu=4.5

```
var<-numeric(25000)
var<-rchisq(25000,nu)
var<-nu*S2/var
b1<-numeric(length(var))
for (i in 1:length(var)){
U=runif(1,0,1)
if (U<=pie) b1[i]=0 else
b1[i]<-rnorm(1,0,sqrt(var[i]))
}
```

#####
S2=1
nu=9

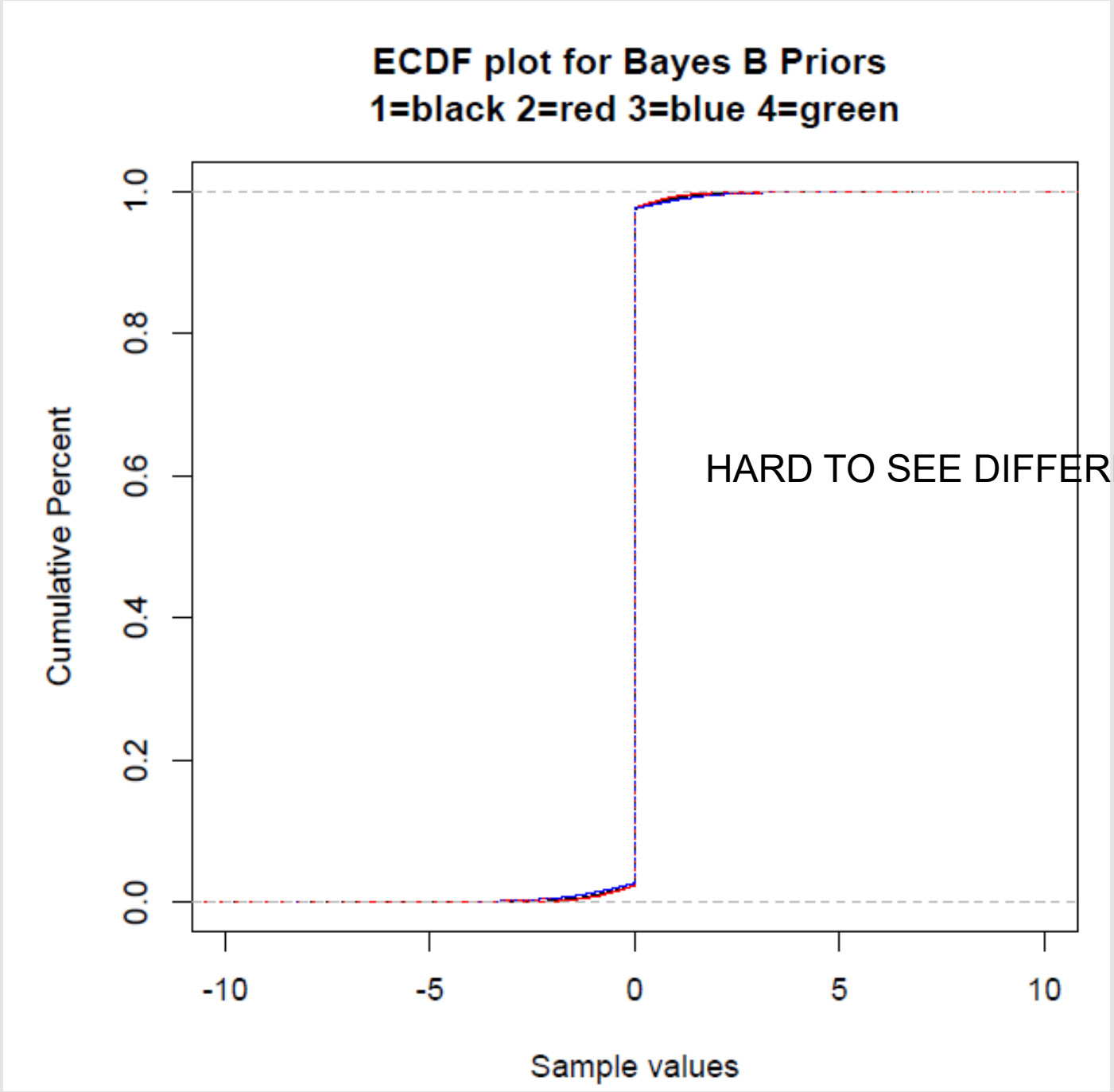
```
var<-rchisq(25000,nu)
var<-nu*S2/var
b2<-numeric(length(var))
for (i in 1:length(var)){
U=runif(1,0,1)
if (U<=pie) b2[i]=0 else
b2[i]<-rnorm(1,0,sqrt(var[i]))
}
```

#######
S2=2
nu=4.5

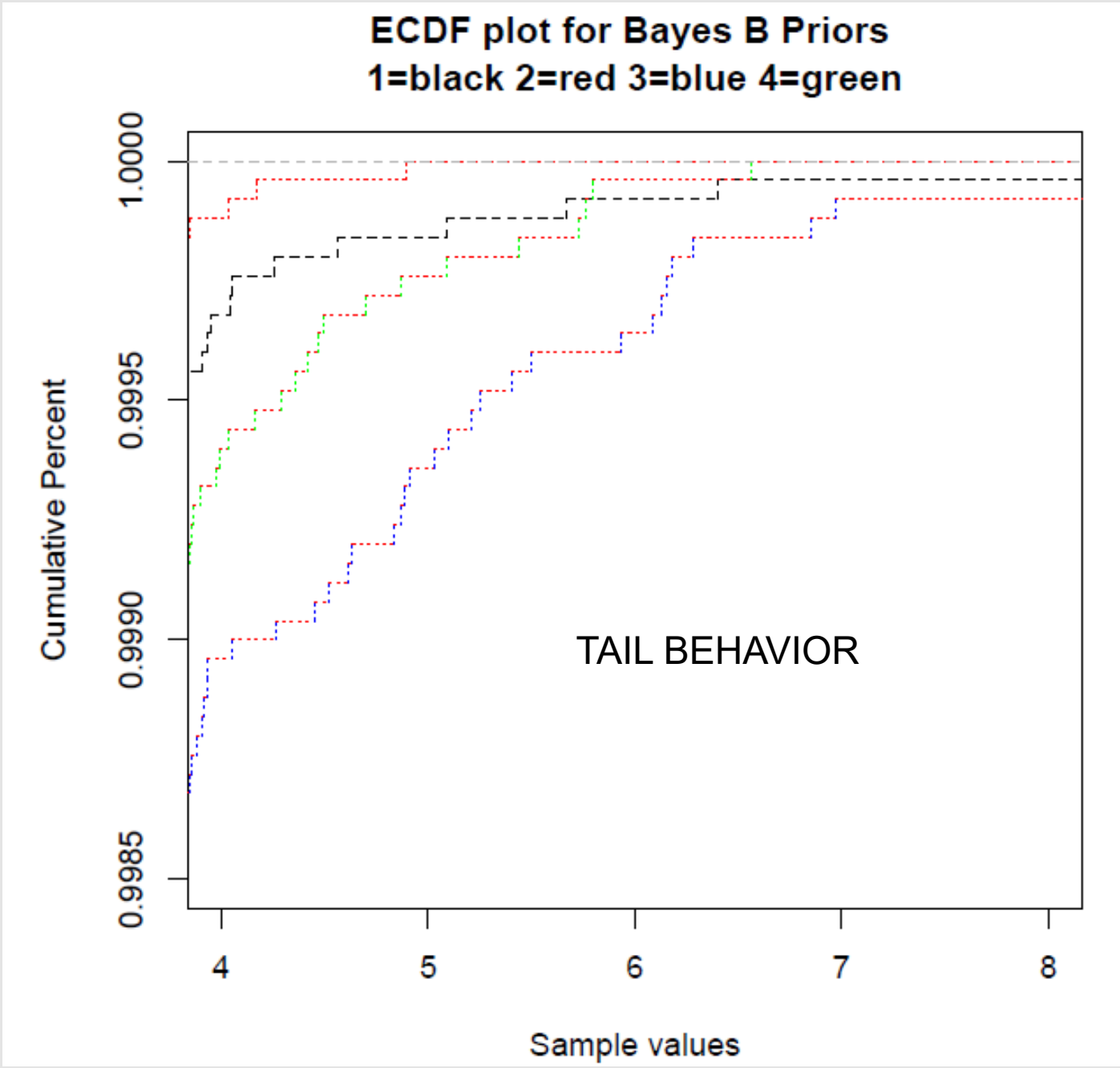
```
var<-rchisq(25000,nu)
var<-nu*S2/var

b3<-numeric(length(var))

for (i in 1:length(var)){
U=runif(1,0,1)
if (U<=pie) b3[i]=0 else
b3[i]<-rnorm(1,0,sqrt(var[i]))
}
```

#######
S2=2
nu=9

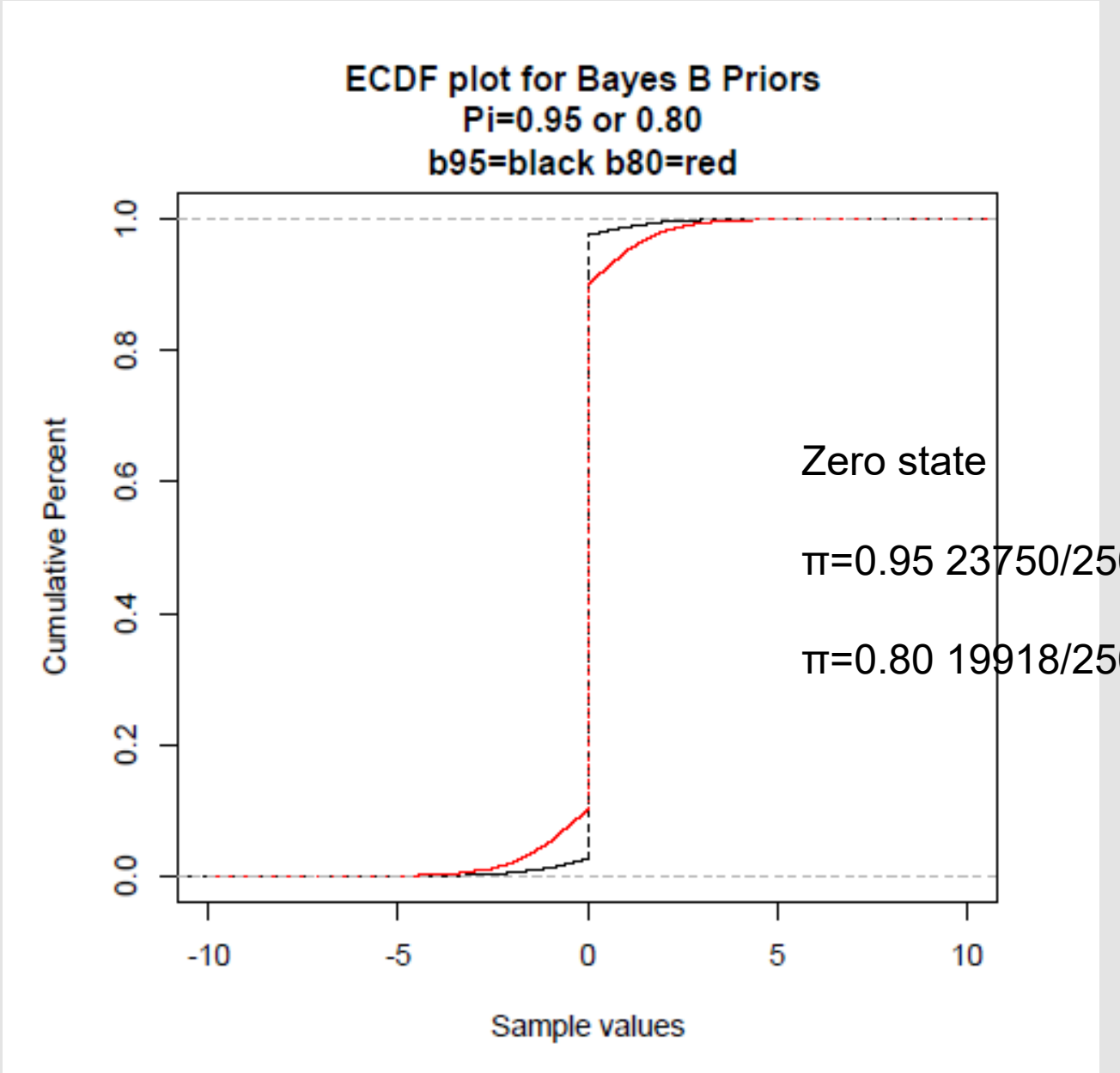
```
var<-rchisq(25000,nu)
var<-nu*S2/var

b4<-numeric(length(var))

for (i in 1:length(var)){
U=runif(1,0,1)
if (U<=pie) b4[i]=0 else
b4[i]<-rnorm(1,0,sqrt(var[i]))
}
```



HARD TO SEE DIFFERENCES…


ECDF plot for Bayes B Priors
1=black 2=red 3=blue 4=green
Cumulative Percent
1.0000
0.9995
0.9990
0.9985
4
5
6
7
8
Sample values
TAIL BEHAVIOR

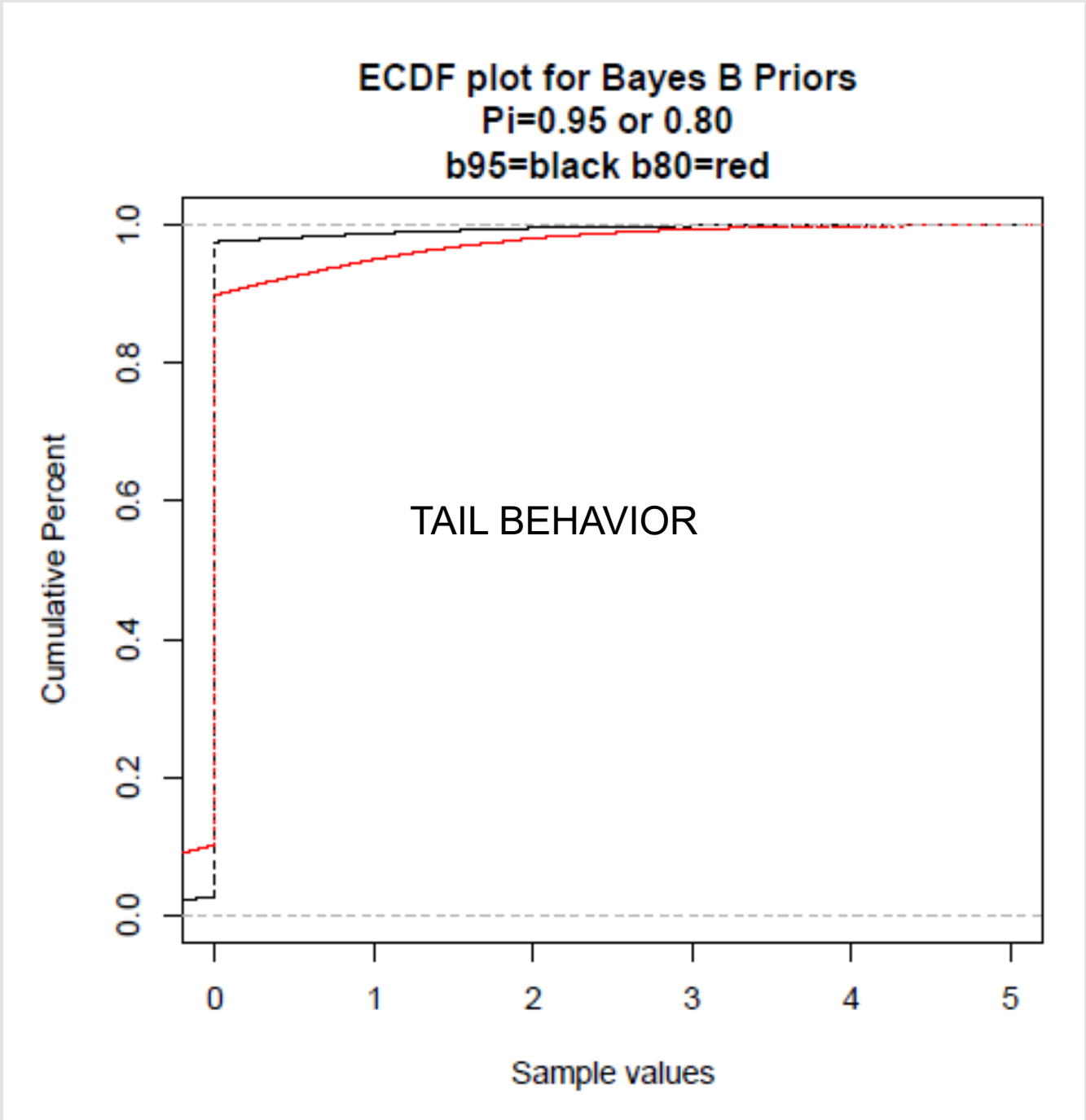
IMPACT OF π=0.95 OR 0.80



Zero state

π=0.95 23750/25000

π=0.80 19918/25000


ECDF plot for Bayes B Priors
Pi=0.95 or 0.80
b95=black b80=red
TAIL BEHAVIOR
Cumulative Percent
0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4
5
Sample values

Bayes C and C-π (code in BGLR)

Habier *et al. BMC Bioinformatics* 2011, **12**:186
http://www.biomedcentral.com/1471-2105/12/186

RESEARCH ARTICLE Open Access

Extension of the bayesian alphabet for genomic selection

David Habier[1*], Rohan L Fernando[1], Kadir Kizilkaya[1,2] and Dorian J Garrick[2,3]

In Bayes B, given the variance of the marker effect and , it is postulated that:

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

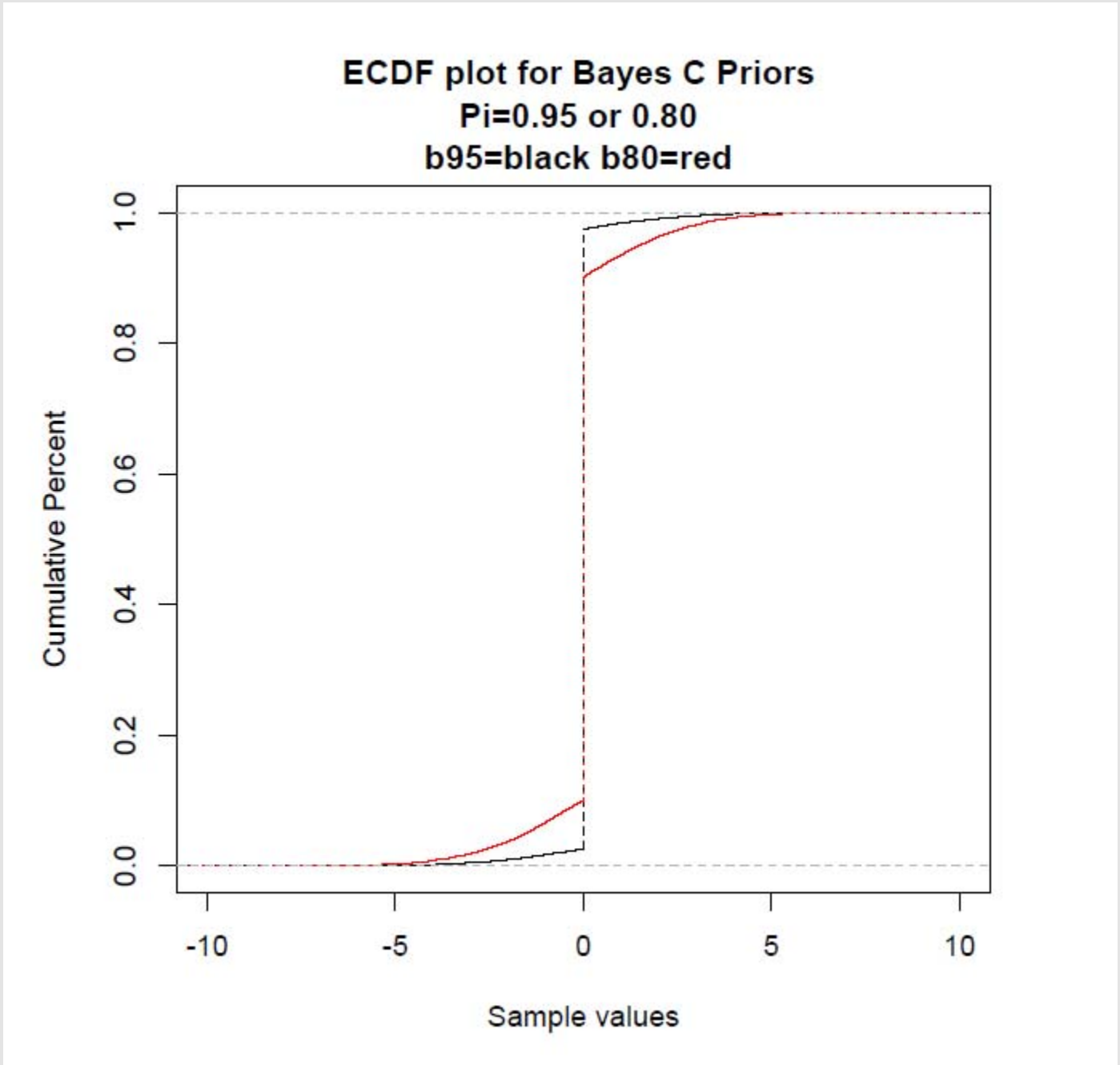
In BayesCπ, it is postulated that:

In BayesCπ, $\sigma^2_{a_k} = \sigma^2_a$, i.e., the priors of all SNP effects have a common variance, which has a scaled inverse chi-square prior with parameters $\nu_a = 4.2$ and S^2_a, where S^2_a is derived as for BayesA and BayesB. As a result, the effect of a SNP fitted with probability (1-π) comes from a mixture of multivariate student's t-distributions, $t(0,\ \nu_a,\ \mathbf{I}S^2_a)$. Fo 's are

Further, it is assumed that the mixture probability π follows a U(0,1) prior distribution

```
###TWO PRIORS FOR BAYES C (AT FIXED PI)
###PI=0.95 and PI=0.80
###PRIOR for non-zero state SCALE INVERTED CHI-SQUARE
S2=2, NU=9
###CONDITIONAL PRIOR FOR b is N(0,var)

####BELOW WE GENERATE SAMPLES FROM
MULTIVARIATE T-DISTRIBUTION

pie=0.95

S2=2
nu=9

##ONLY ONE DRAW FROM SCALED INVERTED CHI-
SQUARE
var<-rchisq(1,nu)
var<-nu*S2/var

b95<-numeric(25000)

for (i in 1:length(b95)){
U=runif(1,0,1)
if (U<=pie) b95[i]=0 else
b95[i]<-rnorm(1,0,sqrt(var))
}

pie=0.80
b80<-numeric(25000)

for (i in 1:length(b80)){
U=runif(1,0,1)
if (U<=pie) b80[i]=0 else
b80[i]<-rnorm(1,0,sqrt(var))
}
```

```
#########
plot(ecdf(b95), verticals=TRUE,
do.p=FALSE,
main="ECDF plot for Bayes C Priors
Pi=0.95 or 0.80
b95=black b80=red ",
 xlim=c(-10,10),
xlab="Sample values",
ylab="Cumulative Percent",lty="dashed")

lines(ecdf(b80), verticals=TRUE,
do.p=FALSE,
col.h="red", col.v="red",lty="dotted")
```


ECDF plot for Bayes C Priors
Pi=0.95 or 0.80
b95=black b80=red
Cumulative Percent
0.0
0.2
0.4
0.6
0.8
1.0
-10
-5
0
5
10
Sample values

BAYESIAN LASSO (code in BGLR)

(Bayes L= double exponential distribution)

$$
\begin{aligned}
p(b|\mu,\lambda) &= \frac{\lambda}{2}\exp\left[-\lambda|b-\mu|\right] \\
&= \frac{\lambda}{2}\begin{cases} \exp\left[-\lambda(b-\mu)\right] & \text{if } b \geq \mu \\ \exp\left[-\lambda(\mu-b)\right] & \text{if } b < \mu \end{cases} \\
E(b|\mu,\lambda) &= \mu \qquad Var(b|\mu,\lambda) = \frac{2}{\lambda^2}
\end{aligned}
$$

In Bayesian Lasso, marker effects are assigned DE conditional (given a parameter λ treated as unknown) prior distributions

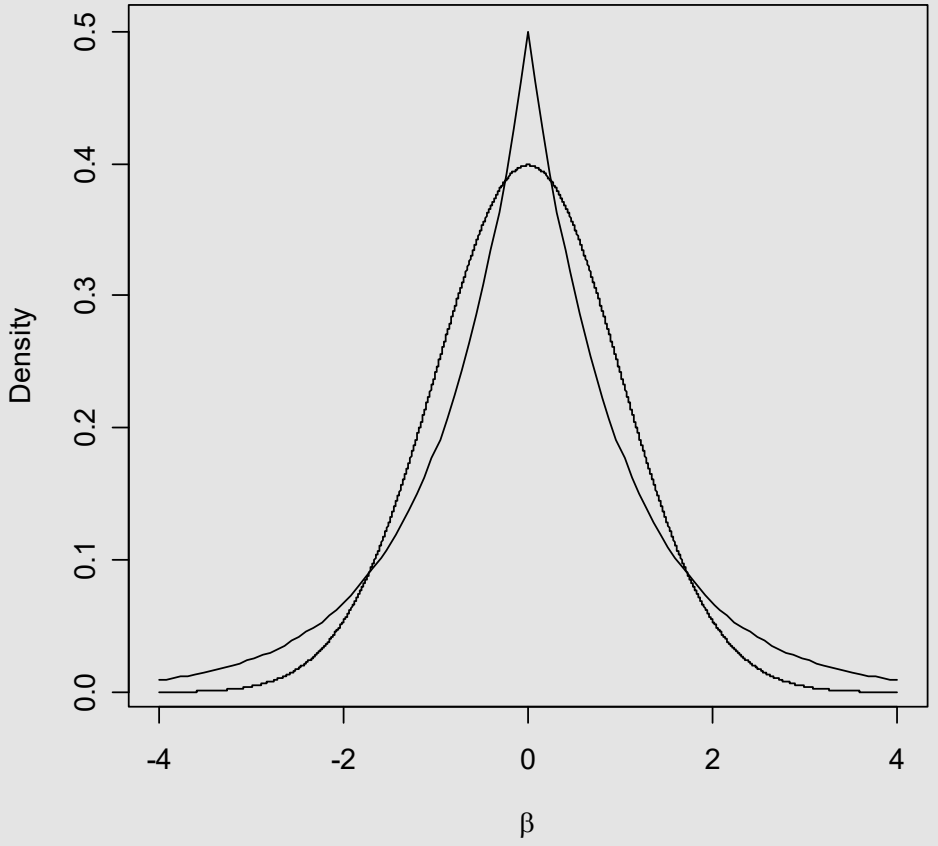
$$p(\boldsymbol{\beta} \mid \lambda) = \prod_{j=1}^{p} \frac{1}{2} \lambda e^{-\lambda |\beta_j|}$$

EACH MARKER HAS THE SAME D.E DISTRIBUTION: NO HETEROGENEOUS VARIANCE OVER MARKERS EITHER

Marker effects are even more strongly shrunk to 0

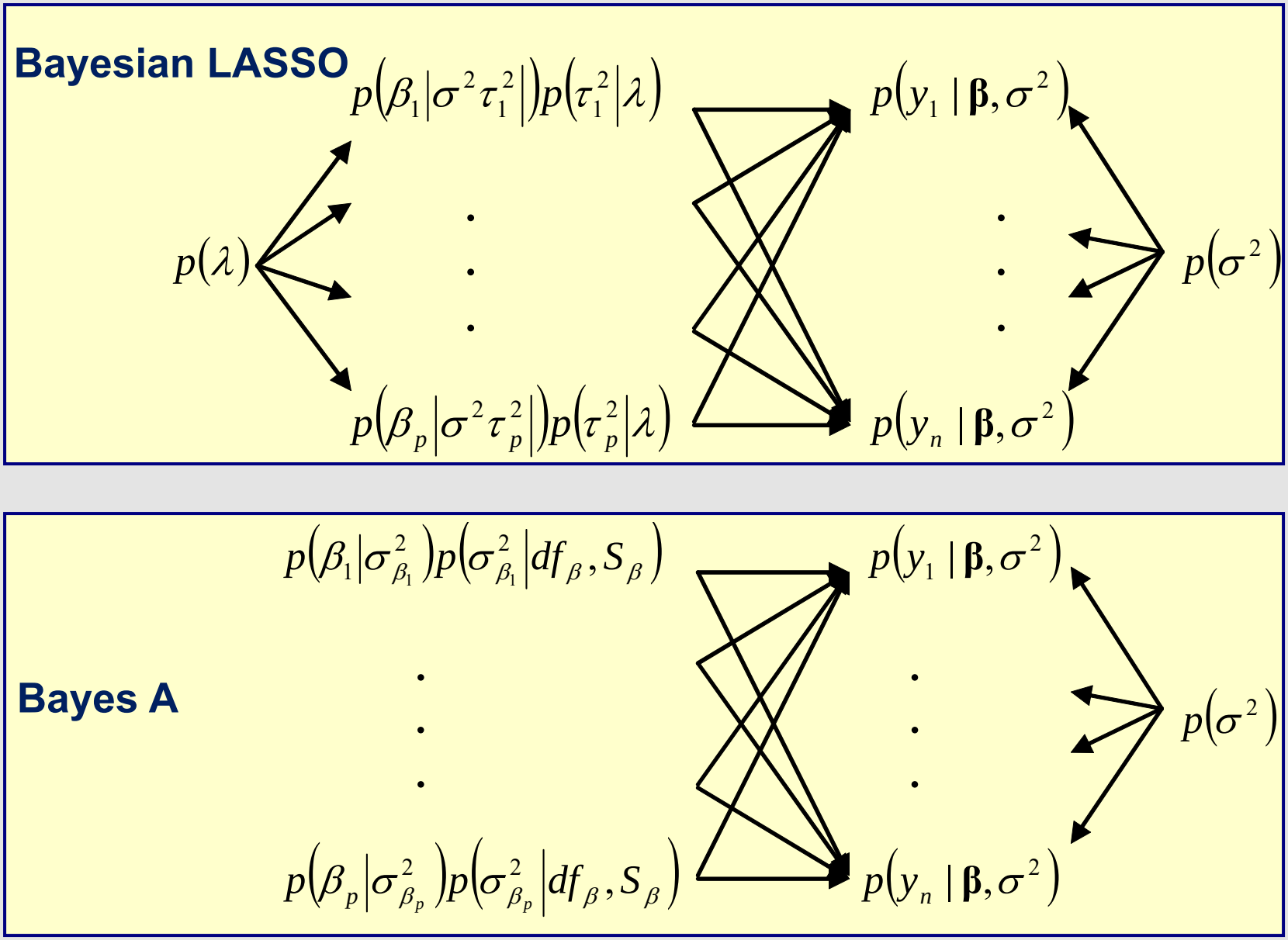
➔ Produces a "quasi-sparse" model

➔ Prior of lambda must be chosen carefully. Avoid flat prior



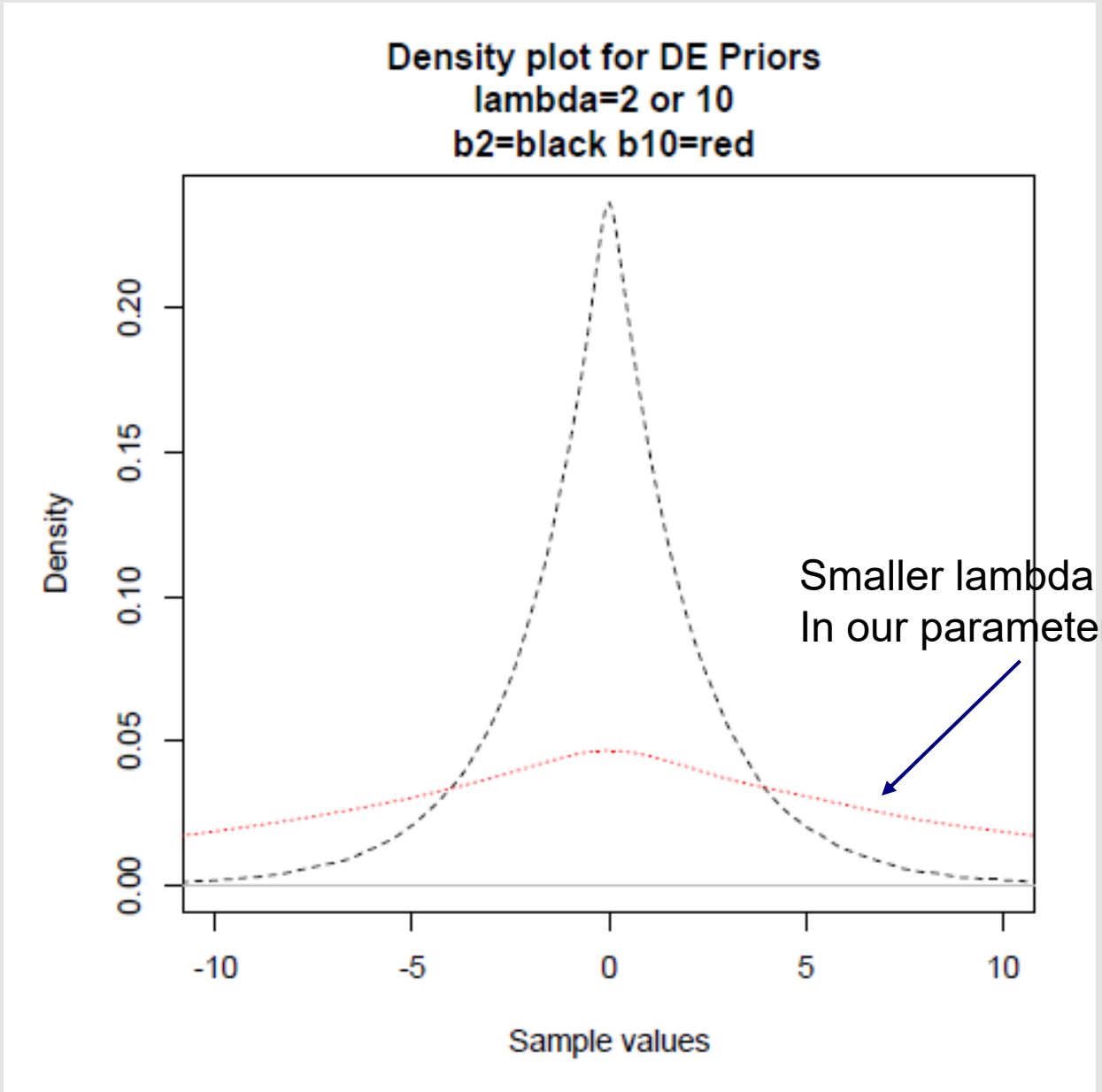
Density of a Normal and of a Double-Exponential Distribution

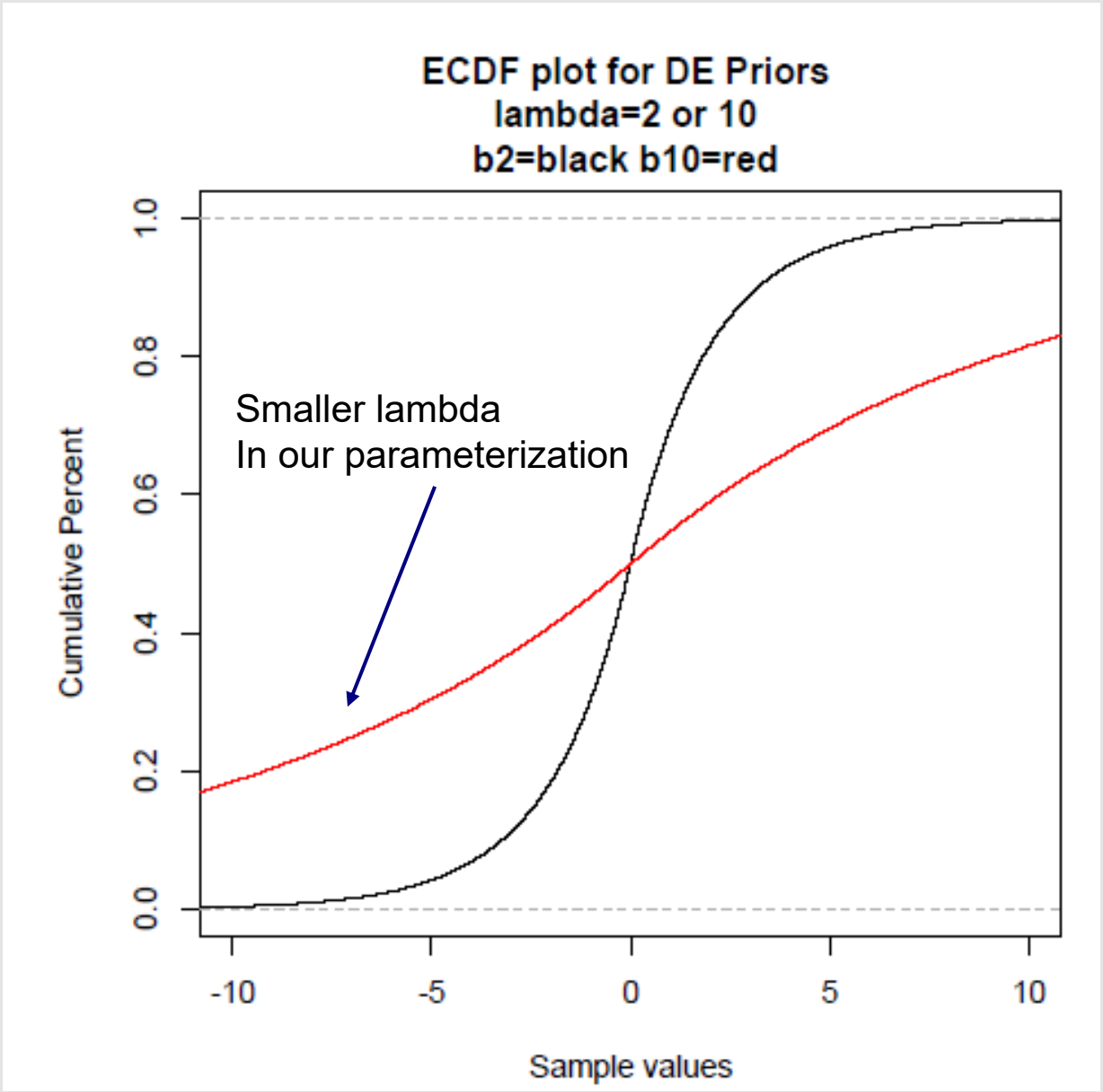
Graphical Representation of the hierarchical structures of the Bayesian LASSO and Bayes A



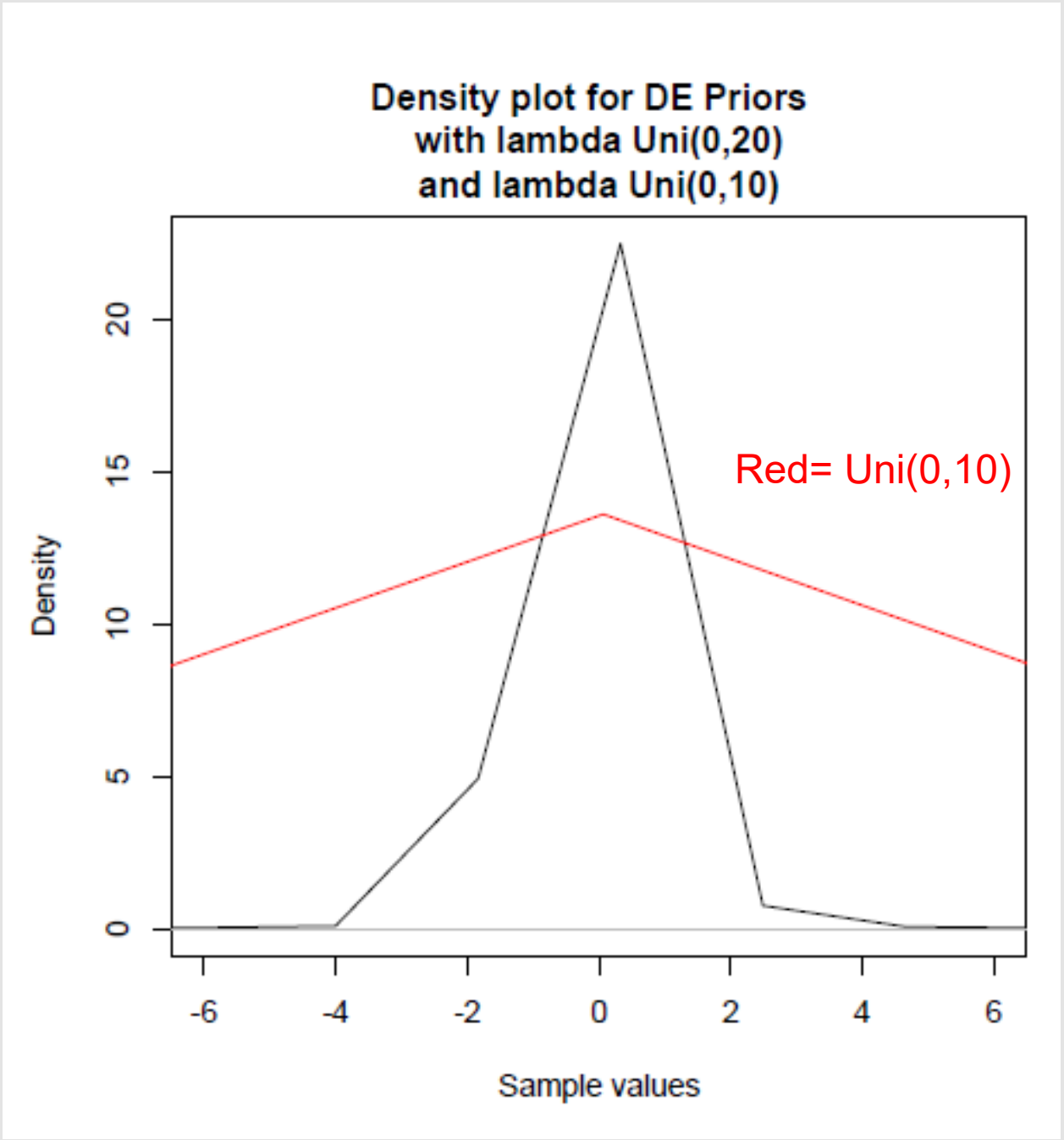
```
#Simulating marker effects from a double exponential (Laplace)
#distribution with parameter lambda
#NOTE: in R, lambda is the reciprocal of the lambda described
earlier

#Enter lambda and S=sample size
lambda<-2
S<-250000

#Simulate S exponential numbers with parameter 1/lambda [due to
parameterization in R]

expnums<-rexp(S,1/lambda)

#Simulate S signs from a fair coin
signs<-sign(runif(S)-0.5)

#simulate Laplace numbers
b2<-signs*expnums

#Enter lambda and S=sample size

lambda<-10
S<-250000

#Simulate S exponential numbers with parameter 1/lambda [due to
parameterization in R]

expnums<-rexp(S,1/lambda)

#Simulate S signs from a fair coin

signs<-sign(runif(S)-0.5)

#simulate Laplace numbers

b10<-signs*expnums
```

```
#########

plot(density(b2),
main="Density plot for DE Priors
lambda=2 or 10
b2=black b10=red ",
 xlim=c(-10,10),
xlab="Sample values",
ylab="Density",lty="dashed")

lines(density(b10),
col="red",lty="dotted")



####
plot(ecdf(b2), verticals=TRUE, do.p=FALSE,
main="ECDF plot for DE Priors
lambda=2 or 10
b2=black b10=red ",
 xlim=c(-10,10),
xlab="Sample values",
ylab="Cumulative Percent",lty="dashed")

lines(ecdf(b10), verticals=TRUE, do.p=FALSE,
col.h="red", col.v="red",lty="dotted")
```


Density plot for DE Priors
lambda=2 or 10
b2=black b10=red
Density
0.00
0.05
0.10
0.15
0.20
-10
-5
0
5
10
Sample values
Smaller lambda
In our parameterization


ECDF plot for DE Priors
lambda=2 or 10
b2=black b10=red
Smaller lambda
In our parameterization
Cumulative Percent
0.0
0.2
0.4
0.6
0.8
1.0
-10
-5
0
5
10
Sample values

```
################################################

#Simulate marginal of marker effects when lambda
#follows a Uniform(LO,UP) distribution
#enter S   LO and UP

S<-25000
LO<-0
UP<-20

lambda<-numeric(S)
expnumb<-numeric(S)
buni<-numeric(S)
sign<-numeric(S)


#simulate b given lambda as
#expnumbs<-rexp(S,lambda)
#signs<-sign(runif(S)-0.5)
#x<-signs*expnums

for (i in 1:S) {
lambda[i]<-runif(1,LO,UP)
expnumb[i]<-rexp(1,lambda[i])
sign[i]<-sign(runif(1)-0.5)
buni[i]<-sign[i]*expnumb[i]  }
#################################

S<-25000
LO<-0
UP<-10

lambda<-numeric(S)
expnumb<-numeric(S)
buni10<-numeric(S)
sign<-numeric(S)


#simulate b given lambda as
#expnumbs<-rexp(S,lambda)
#signs<-sign(runif(S)-0.5)
#x<-signs*expnums

for (i in 1:S) {
lambda[i]<-runif(1,LO,UP)
expnumb[i]<-rexp(1,lambda[i])
sign[i]<-sign(runif(1)-0.5)
buni10[i]<-sign[i]*expnumb[i]  }
```


Density plot for DE Priors
with lambda Uni(0,20)
and lambda Uni(0,10)
Red= Uni(0,10)
Density
0
5
10
15
20
-6
-4
-2
0
2
4
6
Sample values

EXAMINING SHRINKAGE IN THE BAYESIAN LASSO

Mode of the conditional posterior distribution in Bayes L. As a side note, consider what happens if it is not ignored that $\mathbf{W}_{\beta}^{-1} = Diag\left\{\frac{1}{|\beta_j|}\right\}$ is a random matrix, contrary to what was done by Tibshirani (1996) in a modal representation of Bayes L. Recalling that $|\beta_j| = \frac{\beta_j^2}{|\beta_j|}$ and that $\frac{d|x|}{dx} = sign(x)$

$$\frac{\partial}{\partial \beta_j}|\beta_j| = \frac{\partial}{\partial \beta_j}\left(\frac{\beta_j^2}{|\beta_j|}\right) = \frac{2\beta_j}{|\beta_j|} - \frac{\beta_j^2}{|\beta_j|^2}sign(\beta_j) = \frac{2\beta_j}{|\beta_j|} - sign(\beta_j).$$

Differentiating (11) with respect to $\boldsymbol{\beta}$

$$\begin{aligned}\frac{\partial L\left(\boldsymbol{\beta}|\mathbf{y},\lambda,\sigma_e^2\right)}{\partial \boldsymbol{\beta}} &= -\frac{\frac{\partial}{\partial \boldsymbol{\beta}}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)'\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right) + \sigma_e^2\lambda\frac{\partial}{\partial \boldsymbol{\beta}}\sum_{j=1}^{p}|\beta_j|}{2\sigma_e^2} \\ &= -\frac{1}{2\sigma_e^2}\left[-2\mathbf{X}'\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right) + 2\sigma_e^2\lambda\mathbf{W}_{\beta}^{-1}\boldsymbol{\beta} - \sigma_e^2\lambda \mathbf{s}_{\beta}\right],\end{aligned}$$

where $\mathbf{s}_{\beta}$ is a vector containing the signs of the elements of $\boldsymbol{\beta}$. Here, the first-order condition would lead to the iteration

$$\left(\mathbf{X}'\mathbf{X} + \sigma_e^2\lambda\mathbf{W}_{\beta^{[t]}}^{-1}\right)\boldsymbol{\beta}^{[t+1]} = \mathbf{X}'\mathbf{y} + \frac{\sigma_e^2\lambda}{2}\mathbf{s}_{\beta}^{[t]}.$$

Sensitivity to Prior Specification in Bayesian Genome-based Prediction Models

CHRISTINA LEHERMEIER, VALENTIN WIMMER, THERESA ALBRECHT,

HANS-JÜRGEN AUINGER

Plant Breeding, Technische Universität München, 85354 Freising, Germany

DANIEL GIANOLA

Department of Animal Sciences, University of Wisconsin-Madison, Wisconsin, United States of America

CHRIS-CAROLIN SCHÖN*

Plant Breeding, Technische Universität München, 85354 Freising, Germany

chris.schoen@wzw.tum.de

VOLKER J. SCHMID*

Department of Statistics, Ludwig-Maximilians-Universität München, 80539 München, Germany

volker.schmid@stat.uni-muenchen.de

Citation Information: Statistical Applications in Genetics and Molecular Biology. Volume 0, Issue 0, Pages 1–17, ISSN (Online) 1544-6115, ISSN (Print) 2194-6302, DOI: 10.1515/sagmb-2012-0042, April 2013

Table 1. Summary of the simulated and experimental datasets for grain dry matter yield (GDY) and grain dry matter content (GDC). Represented are the number of polymorphic single nucleotide polymorphism markers (no. SNP), number of polymorphic quantitative trait loci in the simulated datasets (no. QTL), number of lines (n) and the trait heritability (h^2). U represents an unknown number of QTL.

	no. SNP	no. QTL	n	h^2
Simulated datasets				
maizeA	1117	500	1250	0.46
maizeB	7425	369	1250	0.64
Experimental dataset				
GDY	11646	U	698	0.74
GDC	11646	U	698	0.94

4.1 *Hellinger distance*

The Hellinger distance $H(f,g)$ (Le Cam, 1986), which is also used in Roos and Held (2011) to evaluate the sensitivity of models with respect to the choice of prior distribtions, measures the distance between two densities f and g:

$$H(f,g) = \sqrt{\frac{1}{2}\int_{-\infty}^{\infty}\left(\sqrt{f(u)} - \sqrt{g(u)}\right)^2 du}. \tag{4.12}$$

$H(f,g)$ is a symmetric measure, which takes its maximum value of 1, if the density f assigns probability 0 to every data point to which g assigns a positive value, and vice versa. The minimum value of H is 0, if $f = g$. of

Simulated data sets

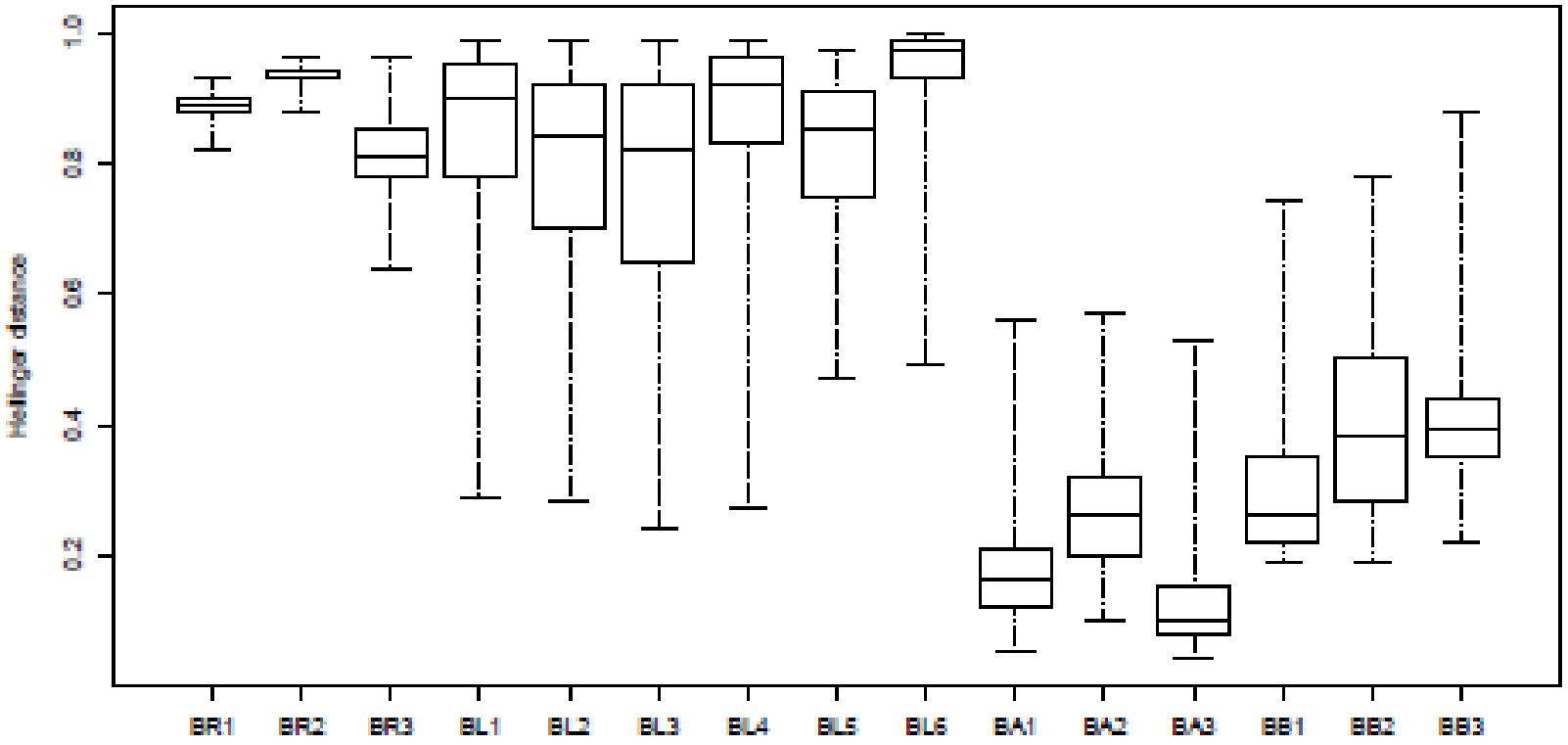


Fig. 2. Distribution of Hellinger distance between the marginal prior and the posterior densities of marker effects β from different model scenarios, calculated with simulated dataset maizeA. Each boxplot displays the distribution of Hellinger distances of the 1117 marker effects out of each model.

BAYES R

J. Dairy Sci. 95:4114–4129
http://dx.doi.org/10.3168/jds.2011-5019

Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels

M. Erbe,*[1] B. J. Hayes,†‡§[1,2] L. K. Matukumalli,# S. Goswami,‖ P. J. Bowman,†‡ C. M. Reich,†‡ B. A. Mason,†‡ and M. E. Goddard†¶
*Department of Animal Sciences, Animal Breeding and Genetics Group, Georg-August-University Göttingen, 37075 Göttingen, Germany
†Biosciences Research Division, Department of Primary Industries, Bundoora, Victoria 3083, Australia
‡Dairy Futures Cooperative Research Centre, Victoria 3083, Australia
§La Trobe University, Bundoora, Victoria 3086, Australia
#Bovine Functional Genomics Laboratory, USDA, Beltsville, MD 20705
‖Bioinformatics and Computational Biology, George Mason University, Manassas, VA 20110
¶Faculty of Land and Food Resources, University of Melbourne, Parkville, Victoria, 3010, Australia

ABSTRACT

Achieving accurate genomic estimated breeding values for dairy cattle requires a very large reference population of genotyped and phenotyped individuals. Assembling such reference populations has been achieved for breeds such as Holstein, but is challenging for breeds with fewer individuals. An alternative is to use a multibreed reference population, such that smaller breeds gain some advantage in accuracy of genomic estimated breeding values (GEBV) from information from larger breeds. However, this requires that marker-quantitative trait loci associations persist across breeds. Here, we assessed the gain in accuracy of GEBV in Jersey cattle as a result of using a combined Holstein and Jersey reference population, with either 39,745 or 624,213 single nucleotide polymorphism (SNP) markers. The surrogate used for accuracy was the correlation of GEBV with daughter trait deviations in a validation population. Two methods were used to predict breeding values, either a genomic BLUP (GBLUP_mod), or a new method, BayesR, which used a mixture of normal distributions as the prior for SNP effects, including one distribution that set SNP effects to zero. The GBLUP_mod method scaled both the genomic relationship matrix and the additive relationship matrix to a base at the time the breeds diverged, and regressed the genomic relationship matrix to account for sampling errors in estimating relationship coefficients due to a finite number of markers, before combining the 2 matrices. Although these modifications did result in less biased breeding values for Jerseys compared with an unmodified genomic relationship matrix, BayesR gave the highest accuracies of GEBV for the 3 traits investigated (milk yield, fat yield, and protein yield), with an average increase in accuracy compared with GBLUP_mod across the 3 traits of 0.05 for both Jerseys and Holsteins. The advantage was limited for either Jerseys or Holsteins in using 624,213 SNP rather than 39,745 SNP (0.01 for Holsteins and 0.03 for Jerseys, averaged across traits). Even this limited and nonsignificant advantage was only observed when BayesR was used. An alternative panel, which extracted the SNP in the transcribed part of the bovine genome from the 624,213 SNP panel (to give 58,532 SNP), performed better, with an increase in accuracy of 0.03 for Jerseys across traits. This panel captures much of the increased genomic content of the 624,213 SNP panel, with the advantage of a greatly reduced number of SNP effects to estimate. Taken together, using this panel, a combined breed reference and using BayesR rather than GBLUP_mod increased the accuracy of GEBV in Jerseys from 0.43 to 0.52, averaged across the 3 traits.
Key words: genomic selection, multiple breeds

INTRODUCTION

To accurately predict genomic breeding values for selection candidates with no phenotype of their own, a very large reference population of genotyped and phenotyped individuals is required to derive the prediction equation (Goddard, 2009; VanRaden et al., 2009; Brøndum et al., 2011). Although this has been achieved

4.5 Bayes R

Erbe et al. (2012) presented this method as follows. Bayes R starts the hierarchical model with (1) and poses a mixture of four zero-mean normal distributions as a conditional prior for a specific SNP effect:

WHY THIS AND NOT SOMETHING ELSE?

$$\begin{aligned} & p\left(\beta|\sigma^2_{\beta_1}=0,\sigma^2_{\beta_2}=10^{-4}\sigma^2_g,\sigma^2_{\beta_3}=10^{-3}\sigma^2_g,\sigma^2_{\beta_4}=10^{-2}\sigma^2_g,\pi_1,\pi_2,\pi_3,\pi_4\right) \\ = \; & \pi_1\times 0+\pi_2 N\left(\beta|0,10^{-4}\sigma^2_g\right)+\pi_3 N\left(\beta|0,10^{-3}\sigma^2_g\right)+\pi_4 N\left(\beta|0,10^{-2}\sigma^2_g\right). \end{aligned} \tag{23}$$

Here, if the SNP effect is generated from the first component of the mixture (with probability π_1) it will be 0 with complete certainty; if drawn from the second component it will have a normal distribution with null mean and variance $\sigma^2_{\beta_2}=10^{-4}\sigma^2_g$, and so on. In Bayes R, $\sigma^2_g=r^2\sigma^2$ is the "assumed genetic variance", r^2 is the "assumed reliability" and σ^2 is the variance of the target trait. Presumably, the assumption about r^2 is either model derived or based on prior cross-validation information, which is good Bayesian behavior, normatively. Makowsky et al. (2011) gave evidence that what one assumes about genetic variance from inference in training data is not recovered in cross-validation.

E(β) CLEARLY 0 FOR ALL MARKERS

$$Var\left(\beta|\boldsymbol{\pi}\right)=\left(\pi_2\times10^{-4}+\pi_3\times10^{-3}+\pi_4\times10^{-2}\right)\sigma_g^2.$$

Further,

$$Var\left(\beta\right)=E_\pi\left[Var\left(\beta|\boldsymbol{\pi}\right)\right]+Var_\pi\left[E\left(\beta|\boldsymbol{\pi}\right)\right]=E_\pi\left[Var\left(\beta|\boldsymbol{\pi}\right)\right].$$

Erbe et al. (2012) used a Dirichlet distribution with parameter vector $\boldsymbol{\alpha}=(\alpha_1,\alpha_2,\alpha_3,\alpha_4)'$ as prior for the elements of $\boldsymbol{\pi}$, so that

$$Var\left(\beta|\boldsymbol{\alpha}\right)=E_\pi\left[Var\left(\beta_j|\boldsymbol{\pi}\right)\right]=\frac{\left(10^{-4}\alpha_2+10^{-3}\alpha_3+10^{-2}\alpha_4\right)}{\alpha_1+\alpha_2+\alpha_3+\alpha_4}\sigma_g^2. \tag{24}$$

In particular, Erbe et al. (2012) took $\alpha_1=\alpha_2=\alpha_3=\alpha_4=1$, producing a uniform distribution on $\boldsymbol{\pi}$. It follows that all SNPs have the same marginal prior distribution, with null mean, and variance

$$Var\left(\beta_j|\boldsymbol{\alpha}\right)=\frac{r^2\sigma^2}{400}\left(1+\frac{1}{10}+\frac{1}{100}\right)=\frac{111}{4\times10^4}r^2\sigma^2.$$

Variance same for all markers

This suggests that a simple ridge-regression BLUP obtained by solving

$$\left[\mathbf{X}'\mathbf{X}+\frac{\sigma_e^2\left(\alpha_1+\alpha_2+\alpha_3+\alpha_4\right)}{r^2\sigma^2\left(10^{-4}\alpha_2+10^{-3}\alpha_3+10^{-2}\alpha_4\right)}\right]\widehat{\boldsymbol{\beta}}=\mathbf{X}'\mathbf{y},$$

may deliver predictive abilities that are similar to those of Bayes R, except that it would differ with respect to Bayes R on how marker effects are shrunk.

SHRINKAGE IN BAYES R

Insight on how shrinkage takes place in Bayes R is gained by inspecting the joint posterior density of all marker effects, given r^2, σ^2 and $\boldsymbol{\pi}$. Here

$$p\left(\boldsymbol{\beta}|\mathbf{y},\boldsymbol{\pi},r^2,\sigma^2\right)$$
$$\propto \exp\left(-\frac{\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)'\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)}{2\sigma_e^2}\right)\prod_{j=1}^{p}\left[\pi_1\times 0+\pi_2 N\left(\beta_j|0,\sigma_2^2\right)+\pi_3 N\left(\beta_j|0,\sigma_3^2\right)+\pi_4 N\left(\beta_j|0,\sigma_4^2\right)\right], \quad (25)$$

where $\sigma_2^2=r^2\sigma^2 10^{-4}, \sigma_3^2=r^2\sigma^2\ 10^{-3}$ and $\sigma_4^2=r^2\sigma^2\ 10^{-2}$ (these values can be modified *a piacere*). Taking derivatives of the log-posterior with respect to $\boldsymbol{\beta}$ gives (apart from an additive constant)

$$\frac{\partial}{\partial\boldsymbol{\beta}}\log\left[p\left(\boldsymbol{\beta}|\mathbf{y},\boldsymbol{\pi},r^2,\sigma^2\right)\right]=\frac{1}{\sigma_e^2}\left(\mathbf{X}'\mathbf{y}-\mathbf{X}'\mathbf{X}\boldsymbol{\beta}\right)+\left\{\frac{\sum_{i=2}^{4}\pi_i\frac{d}{d\beta_j}\phi_i\left(\beta_j|0,\sigma_i^2\right)}{\pi_2\phi_2\left(\beta_j|0,\sigma_2^2\right)+\pi_3\phi_3\left(\beta_j|0,\sigma_3^2\right)+\pi_4\phi_4\left(\beta_j|0,\sigma_4^2\right)}\right\}, \quad (26)$$

where $\{.\}$ denotes a $p\times 1$ vector. Above, $\phi_i\left(\beta_i|0,\sigma_i^2\right)$ $(i=2,3,4)$ is the density of β_j under the normal distribution corresponding to component i of the mixture, with

$$\frac{d}{d\beta_j}\phi_i\left(\beta_j|0,\sigma_i^2\right)=-\frac{\phi_i\left(\beta_j|0,\sigma_i^2\right)}{\sigma_i^2}\beta_j.$$

Employing the preceding expression in equation (26) yields

$$\frac{\partial}{\partial \boldsymbol{\beta}} \log\left[p\left(\boldsymbol{\beta}|\boldsymbol{\pi}, r^2, \sigma^2\right)\right] = \left(\frac{1}{\sigma_e^2}\right)\left(\mathbf{X}'\mathbf{y} - \mathbf{X}'\mathbf{X}\boldsymbol{\beta}\right) - \left\{\frac{\sum_{i=2}^{4} \pi_i \frac{\phi_i\left(\beta_j|0,\sigma_i^2\right)}{\sigma_i^2}}{\sum_{i=2}^{4} \pi_i \phi_i\left(\beta_j|0,\sigma_i^2\right)} \beta_j\right\}.$$

Setting this to zero and rearranging leads to iteration

$$\boldsymbol{\beta}^{[t+1]} = \left(\mathbf{X}'\mathbf{X} + \Omega_{\beta^{[t]}}\right)^{-1} \mathbf{X}'\mathbf{y},$$

where $\Omega_{\beta^{[t]}}$ is a $p \times p$ diagonal matrix with typical element

$$\Omega_{jj,\boldsymbol{\beta}}^{[t]} = \sigma_e^2 \frac{\sum_{i=2}^{4} \pi_i \phi_i\left(\beta_j^{[t]}|0,\sigma_i^2\right) \frac{1}{\sigma_i^2}}{\sum_{i=2}^{4} \pi_i \phi_i\left(\beta_j^{[t]}|0,\sigma_i^2\right)} = \sum_{i=2}^{4} \pi_{ij}'^{[t]} \frac{\sigma_e^2}{\sigma_i^2},$$

where

$$\pi_{ij}'^{[t]}\left(\beta_j\right) = \frac{\pi_i \phi_i\left(\beta_j^{[t]}|0,\sigma_i^2\right)}{\sum_{i=2}^{4} \pi_i \phi_i\left(\beta_j^{[t]}|0,\sigma_i^2\right)}, i = 1, 2, ..., 4 \text{ and } \; j = 1, 2, ..., p.$$

PLOS | GENETICS

RESEARCH ARTICLE

Simultaneous Discovery, Estimation and Prediction Analysis of Complex Traits Using a Bayesian Mixture Model

Gerhard Moser[1]*, Sang Hong Lee[1], Ben J. Hayes[2,3], Michael E. Goddard[2,4], Naomi R. Wray[1], Peter M. Visscher[1,5]

Abstract

Gene discovery, estimation of heritability captured by SNP arrays, inference on genetic architecture and prediction analyses of complex traits are usually performed using different statistical models and methods, leading to inefficiency and loss of power. Here we use a Bayesian mixture model that simultaneously allows variant discovery, estimation of genetic variance explained by all variants and prediction of unobserved phenotypes in new samples. We apply the method to simulated data of quantitative traits and Welcome Trust Case Control Consortium (WTCCC) data on disease and show that it provides accurate estimates of SNP-based heritability, produces unbiased estimators of risk in new samples, and that it can estimate genetic architecture by partitioning variation across hundreds to thousands of SNPs. We estimated that, depending on the trait, 2,633 to 9,411 SNPs explain all of the SNP-based heritability in the WTCCC diseases. The majority of those SNPs (>96%) had small effects, confirming a substantial polygenic component to common diseases. The proportion of the SNP-based variance explained by large effects (each SNP explaining 1% of the variance) varied markedly between diseases, ranging from almost zero for bipolar disorder to 72% for type 1 diabetes. Prediction analyses demonstrate that for diseases with major loci, such as type 1 diabetes and rheumatoid arthritis, Bayesian methods outperform profile scoring or mixed model approaches.

Kemper et al. Genetics Selection Evolution (2015) 47:29
DOI 10.1186/s12711-014-0074-4

GSE Genetics Selection Evolution

RESEARCH Open Access

Improved precision of QTL mapping using a nonlinear Bayesian method in a multi-breed population leads to greater accuracy of across-breed genomic predictions

Kathryn E Kemper[1*], Coralie M Reich[2], Philip J Bowman[2], Christy J vander Jagt[2], Amanda J Chamberlain[2], Brett A Mason[2], Benjamin J Hayes[2,3,4] and Michael E Goddard[1,2]

Abstract

Background: Genomic selection is increasingly widely practised, particularly in dairy cattle. However, the accuracy of current predictions using GBLUP (genomic best linear unbiased prediction) decays rapidly across generations, and also as selection candidates become less related to the reference population. This is likely caused by the effects of causative mutations being dispersed across many SNPs (single nucleotide polymorphisms) that span large genomic intervals. In this paper, we hypothesise that the use of a nonlinear method (BayesR), combined with a multi-breed (Holstein/Jersey) reference population will map causative mutations with more precision than GBLUP and this, in turn, will increase the accuracy of genomic predictions for selection candidates that are less related to the reference animals.

Results: BayesR improved the across-breed prediction accuracy for Australian Red dairy cattle for five milk yield and composition traits by an average of 7% over the GBLUP approach (Australian Red animals were not included in the reference population). Using the multi-breed reference population with BayesR improved accuracy of prediction in Australian Red cattle by 2 – 5% compared to using BayesR with a single breed reference population. Inclusion of 8478 Holstein and 3917 Jersey cows in the reference population improved accuracy of predictions for these breeds by 4 and 5%. However, predictions for Holstein and Jersey cattle were similar using within-breed and multi-breed reference populations. We propose that the improvement in across-breed prediction achieved by BayesR with the multi-breed reference population is due to more precise mapping of quantitative trait loci (QTL), which was demonstrated for several regions. New candidate genes with functional links to milk synthesis were identified using differential gene expression in the mammary gland.

Conclusions: QTL detection and genomic prediction are usually considered independently but persistence of genomic prediction accuracies across breeds requires accurate estimation of QTL effects. We show that accuracy of across-breed genomic predictions was higher with BayesR than with GBLUP and that BayesR mapped QTL more precisely. Further improvements of across-breed accuracy of genomic predictions and QTL mapping could be achieved by increasing the size of the reference population, including more breeds, and possibly by exploiting pleiotropic effects to improve mapping efficiency for QTL with small effects.

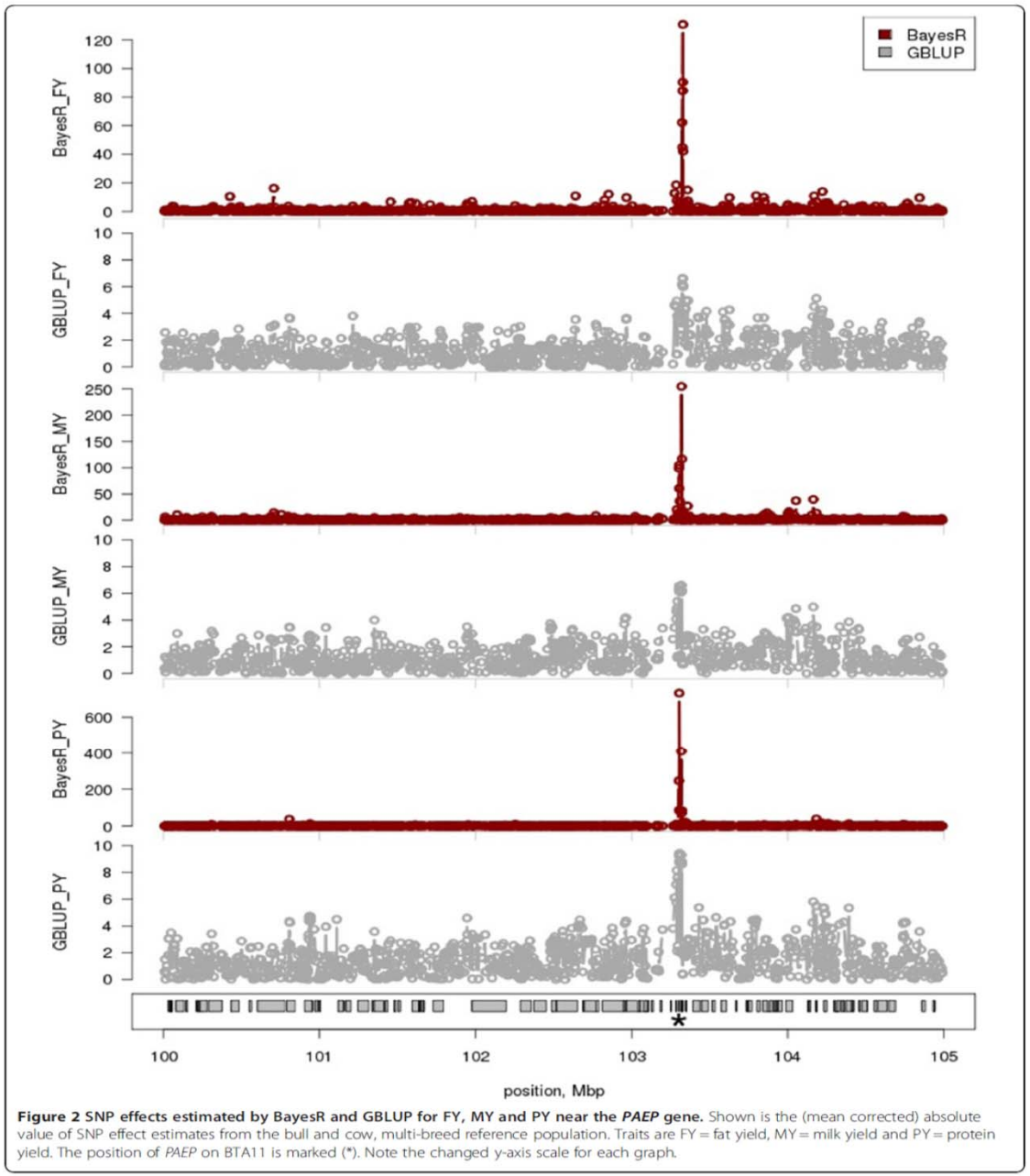


Figure 2 SNP effects estimated by BayesR and GBLUP for FY, MY and PY near the *PAEP* gene. Shown is the (mean corrected) absolute value of SNP effect estimates from the bull and cow, multi-breed reference population. Traits are FY = fat yield, MY = milk yield and PY = protein yield. The position of *PAEP* on BTA11 is marked (*). Note the changed y-axis scale for each graph.

RECALL:

If two priors lead to two different posteriors, informatiois scant.

NOTE:

Bayes R“flattens” to 0, most marker effects? What About the quest for Small-effect variants?

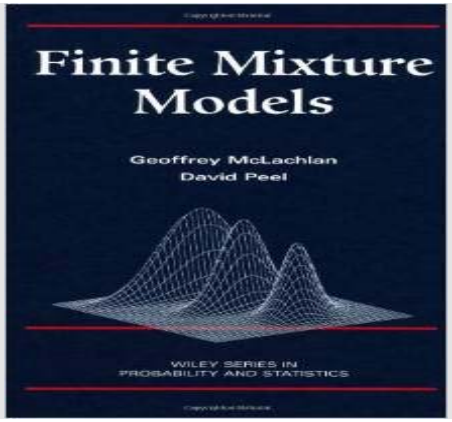
REMARK:

No model comparisons at all. For example, does Bayes R receive more support than Bayes A?

DO YOU LIKE MIXTURES?

If there are variants with
large effects (e.g., HDL)
feature selection makes sense

[redacted] McLachlan and Peel (2000) give a warning: estimation of the parameters of a mixture (Ψ) on the basis of data are meaningful only if Ψ is likelihood identifiable. In Bayes RS (apart from nuisance effects and the residual variance) the number of unknown parameters is $2p + 4S$. Here, $2p$ comes from the fact that each marker is assigned a distinct variance; the $4S$ comes from the fact that there are S segments each having four segment-specific mixing probabilities $\pi_s (s = 1, 2, \ldots, S)$. Unfortunately, $n \ll\ll < 2p + 4S$, and this creates a huge identification deficit relative to the information content in a sample of size n. In



[redacted]

[redacted] In a Bayesian context, there is the additional issue (occurring even when $n > p$) called label switching, leading Celeux *et al.* (2000) to write: "Although somewhat presumptuous, we consider that almost the entirety of Markov chain Monte Carlo samplers for mixture models has failed to converge!" In view of these pitfalls, one wonders what meaningful mechanistic sense can be extracted from these richly parameterized specifications intended to inform about genetic architecture.

Celeux, G., M. Hurn, and C. Robert, 2000 Computational and inferential difficulties with mixture posterior distributions. J. Am. Stat. Assoc. 95: 957–979.

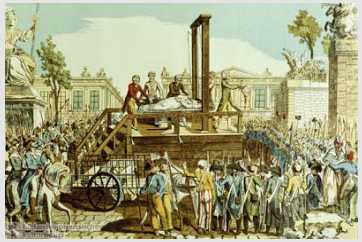
WILD IDEAS:
(open the umbrella…)

- Some friends in animal breeding promoting mixtures with zillions of markers are getting wild (for discovery purposes)
- Most mixtures fail to converge, plus there is severe under-identification problem: will never discover small effect variants
- Mixtures can be dangerous (like mixing drinks!)

Take home message about the alphabet:
(for n<<p)

ABSTRACT Whole-genome enabled prediction of complex traits has received enormous attention in animal and plant breeding and is making inroads into human and even *Drosophila* genetics. The term "Bayesian alphabet" denotes a growing number of letters of the alphabet used to denote various Bayesian linear regressions that differ in the priors adopted, while sharing the same sampling model. We explore the role of the prior distribution in whole-genome regression models for dissecting complex traits in what is now a standard situation with genomic data where the number of unknown parameters (p) typically exceeds sample size (n). Members of the alphabet aim to confront this overparameterization in various manners, but it is shown here that the prior is always influential, unless $n \gg p$. This happens because parameters are not likelihood identified, so Bayesian learning is imperfect. Since inferences are not devoid of the influence of the prior, claims about genetic architecture from these methods should be taken with caution. However, all such procedures may deliver reasonable predictions of complex traits, provided that some parameters ("tuning knobs") are assessed via a properly conducted cross-validation. It is concluded that members of the alphabet have a room in whole-genome prediction of phenotypes, but have somewhat doubtful inferential value, at least when sample size is such that $n \ll p$.

Inferences about "genetic architecture":



Prediction of overall genetic signal: