4. GENOME-ENABLED PREDICTION
BAYES A, BAYES B, LASSO

Standard analysis (fixed **X**)

Genotypic value (signal from genome)

Assumption

$$y = f + e = X\beta + e$$

Bayesian or Frequentist?
(more later)

$$\beta|\sigma^2_\beta \sim N(0, I\sigma^2_\beta)$$

$$E(y|X,\beta) = X\beta$$

$$E(y|X) = 0$$

$$Var(y|X,\sigma^2_e,\sigma^2_\beta) = XX'\sigma^2_\beta + I\sigma^2_e$$

Example 1 (Ridge regression from Bayesian and frequentist points of view)
Suppose the conditional prior of the regressions has the form

$$\left[\begin{array}{c}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{array}\right]\Big|\sigma^2, H_{\boldsymbol{\beta}} \sim N\left(\left[\begin{array}{c}\mathbf{m}_1\\ \mathbf{m}_2\end{array}\right], \left[\begin{array}{cc}\mathbf{I}\frac{\sigma^2_{\beta_1}}{\sigma^2} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}\frac{\sigma^2_{\beta_2}}{\sigma^2}\end{array}\right]\sigma^2\right),$$

Frequentist: random effects model

so the two sets of coefficients are independent, a priori. Then the mean of the conditional posterior distribution of the regression coefficients, using (1.30) and (1.32), is

Frequentist: conditional distribution

$$\left[\begin{array}{c}\overline{\boldsymbol{\beta}}_1\\ \overline{\boldsymbol{\beta}}_2\end{array}\right] = \left[\begin{array}{cc}\mathbf{X}_1'\mathbf{X}_1 + \mathbf{I}\frac{\sigma^2}{\sigma^2_{\beta_1}} & \mathbf{X}_1'\mathbf{X}_2\\ \mathbf{X}_2'\mathbf{X}_1 & \mathbf{X}_2'\mathbf{X}_2 + \mathbf{I}\frac{\sigma^2}{\sigma^2_{\beta_2}}\end{array}\right]^{-1}\left[\begin{array}{c}\mathbf{X}_1'\mathbf{y} + \mathbf{m}_1\frac{\sigma^2}{\sigma^2_{\beta_1}}\\ \mathbf{X}_1'\mathbf{y} + \mathbf{m}_2\frac{\sigma^2}{\sigma^2_{\beta_2}}\end{array}\right].$$

When there is a single set of regression coefficients and when the prior mean is a null vector, this reduces to

$$\overline{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X} + \mathbf{I}k\right)^{-1}\mathbf{X}'\mathbf{y},$$

Frequentist: mean of conditional distribution (BLUP here)
Bayesian: mean of conditional posterior distribution

where

$$k = \frac{\sigma^2}{\sigma^2_{\beta}}$$

Frequentist: estimate var. comp by, e.g., REML
Bayesian: use posterior distributions

Prediction of marker effects: BLUP (iid marker effects)

$$\left[X'X + \frac{\sigma^2_e}{\sigma^2_\beta}I\right]\hat{\beta} = X'y$$

Assume inverse exists

$$\left[I + \frac{\sigma^2_e}{\sigma^2_\beta}(X'X)^{-1}\right]\hat{\beta} = (X'X)^{-1}X'y$$

$$\hat{\beta} = \left[I + \frac{\sigma^2_e}{\sigma^2_\beta}(X'X)^{-1}\right]^{-1}\tilde{\beta}_{\text{OLS}} \Rightarrow \text{SHRINKAGE}$$

Prediction of signal ($X\beta$) to phenotype

$$Var(X\beta|y) = XVar(\beta|y)X'$$

$$= X\left[I + \frac{\sigma^2_e}{\sigma^2_\beta}(X'X)^{-1}\right]^{-1}X'\sigma^2_e$$

Prediction of future record

$$y^* = X^*\beta + e^*$$

$$E(X^*\beta + e^*|y, X, X^*) = X^*E(\beta|y, X)$$
$$= X^*\left[I + \frac{\sigma_e^2}{\sigma_\beta^2}(X'X)^{-1} \right]^{-1} \tilde{\beta}_{\text{OLS}}$$

$$Var(X^*\beta + e^*|y, X, X^*) = X^*Var(\beta|y, X)X^* + I^*\sigma_e^2$$

1. Standard BLUP of signal (*f*)

X is fixed here

$$y = f + e = X\beta + e$$
$$f \sim N(0, Var(f)) \quad Var(f) = XX'Var(\beta)$$
$$Var(y|X) = XX'Var(\beta) + I\sigma_e^2$$
$$BLUP(f) = Cov(f, y')[XX'Var(\beta) + I\sigma_e^2]^{-1}y$$
$$= XX'Var(\beta)[XX'Var(\beta) + I\sigma_e^2]^{-1}y$$
$$= \left[I + (XX')^{-1}\frac{\sigma_e^2}{Var(\beta)} \right]^{-1} y$$

$$\left[I + (XX')^{-1}\frac{\sigma_e^2}{Var(\beta)} \right] BLUP(f) = y$$

2. Morph into genomic BLUP a la Van Raden

X is random here, but so is β

$$G = \frac{(X - E(X))(X - E(X))'}{2\sum_{j=1}^{p} p_j(1-p_j)} = \frac{X^*X^{*\prime}}{V_{M,HW}}$$

Center using allelic frequency information

$$\left[I + G^{-1}\frac{\sigma_e^2}{Var(\beta)/V_{M,HW}} \right] \hat{g} = y$$

IS THIS METAMORPHOSIS DONE CORRECTLY? I DO NOT THIK SO

GAUSSIAN PROCESS ANALYSIS (IID MARKER EFFECTS)

$$y = f + e = X\beta + e$$

$$\beta \sim N(0, I\sigma^2_\beta)$$

[Read Falconer and Mackay IQG]

$$X \sim F$$

[Genotypes vary at random: population Genetics]

$$E(y|X,\beta) = X\beta$$

$$E(y|\beta) = E_X E(y|X,\beta) = E(X)\beta$$

$$E(y) = E_\beta[E(X)\beta] = E(X)E(\beta) = 0$$

Big assumption

Are frequencies effect-dependent? Are effects frequency dependent?
TURELLI, ZHANG&HILL, MACKAY WITH MARKERS AND

$$Var(y) = Var(f) + Var(e) = Var(f) + I\sigma^2_e$$

$$\begin{aligned} Var(f) &= Var(X\beta) \\ &= E_X(Var(X\beta|X) + Var_X[E(X\beta|X)] \\ &= E_X[XVar(\beta)X']) + Var_X[XE(\beta)] \\ &= E_X[XX'\sigma^2_\beta] + Var_X(0) \\ &= \sigma^2_\beta E_X[XX'], \end{aligned}$$

Covariance matrix of signal

BP= "best predictor" (MULVN assumed)

$$\hat{f} = BP(f)$$

$$\left[\frac{1}{\sigma^2_e} I + Var^{-1}(f)\right]\hat{f} = \frac{1}{\sigma^2_e} y$$

$$\left[I + \frac{\sigma^2_e}{\sigma^2_\beta} E_X^{-1}[XX']\right]\hat{f} = y$$

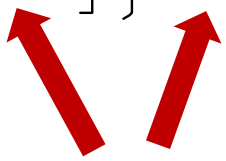
Looks like genomic BLUP (it is not)

$$E_X^{-1}[XX']\left[E_X[XX'] + \frac{\sigma^2_e}{\sigma^2_\beta} I\right]\hat{f} = y$$

$$\left[E_X[XX'] + \frac{\sigma^2_e}{\sigma^2_\beta} I\right]\hat{f} = E_X[XX']y$$

Under multivariate normality

$$
\begin{aligned}
Var(f|y) &= Var(f) - Cov(f,y)Var^{-1}(y)Cov'(f,y) \\
&= Var(f) - Var(f)[Var(f) + I\sigma_e^2]^{-1}Var(f) \\
&= \sigma_\beta^2 E_X[XX'] - \sigma_\beta^2 E_X[XX'][\sigma_\beta^2 E_X[XX'] + I\sigma_e^2]^{-1}\sigma_\beta^2 E_X[XX'] \\
&= \sigma_\beta^2 E_X[XX'] - \sigma_\beta^2 E_X[XX']\frac{E_X^{-1}[XX']}{\sigma_\beta^2}\left[I + \frac{\sigma_e^2}{\sigma_\beta^2}E_X[XX']\right]^{-1}\sigma_\beta^2 E_X[XX'] \\
&= \left\{I - \left[I + \frac{\sigma_e^2}{\sigma_\beta^2}E_X[XX']\right]^{-1}\right\}\sigma_\beta^2 E_X[XX'].
\end{aligned}
$$



Proper assessment of posterior uncertainty requires knowledge of the genotypic distribution

$$
X_{\text{ind,marker}} = \begin{bmatrix} x_{11} & . & x_{1p} \\ x_{21} & . & x_{2p} \\ . & . & . \\ x_{n1} & . & x_{np} \end{bmatrix}
$$

$$
XX' = \begin{bmatrix} x_{11} & . & x_{1p} \\ x_{21} & . & x_{2p} \\ . & . & . \\ x_{n1} & . & x_{np} \end{bmatrix} \begin{bmatrix} x_{11} & x_{21} & . & x_{n1} \\ . & . & . & . \\ x_{1p} & x_{2p} & . & x_{np} \end{bmatrix}
$$

$$
= \begin{bmatrix} \sum_{j=1}^{p} x_{1_j}^2 & \sum_{j=1}^{p} x_{1_j}x_{2_j} & & \sum_{j=1}^{p} x_{1_j}x_{n_j} \\ & \sum_{j=1}^{p} x_{2_j}^2 & & \\ & & . & \\ & & & \sum_{j=1}^{p} x_{n_j}^2 \end{bmatrix}
$$

Under HW

$$E\left(\sum_{j=1}^{p} x_{ij}^2\right) = \sum_{j=1}^{p} Var(x_{ij}) + \sum_{j=1}^{p} E^2(x_{ij})$$
$$= \sum_{j=1}^{p} 2p_j q_j + \sum_{j=1}^{p} (p_j - q_j)^2$$
$$= \sum_{j=1}^{p} (1 - 2p_j q_j) = p - \sum_{j=1}^{p} 2p_j q_j$$
$$E\left(\sum_{j=1}^{p} x_{1_j} x_{2_j}\right) = \sum_{j=1}^{p} Cov(x_{1_j} x_{2_j}) + \sum_{j=1}^{p} E(x_{1_j}) E(x_{2_j})$$
$$= \sum_{j=1}^{p} 2\phi_{ij} p_j q_j + \sum_{j=1}^{p} (p_j - q_j)^2$$
$$= \sum_{j=1}^{p} p_j^2 + q_j^2 - 2p_j q_j (1 - \phi)$$
$$Cov(x_{1_j} x_{2_j}) = p_j^2 + q_j^2 - 2p_j q_j (1 - \phi) - (p_j - q_j)^2$$
$$= 2pq\phi$$

-How to obtain sensible estimates? Is **XX**' a good estimate of E(**XX**')?
-Should we assume HW and use estimates of allelic frequencies and of φ (i,j) as if there were no selection, etc.?

Future record:

$$f^* = X^*\beta + e^*$$

uno $$E(f^*|f) = E(f^*) + Cov(X^*\beta, \beta X')Var^{-1}(f)f$$

dos $$E(f^*|y) = E_{f|y}E(f^*|f,y) = E_{f|y}E(f^*|f)$$

tres $$= E_{f|y}[Cov(X^*\beta, \beta X')Var^{-1}(f)f]$$

$$Cov(X^*\beta, \beta X') = Cov[E(X^*\beta, \beta X'|X^*, X)] + E[Cov(X^*\beta, \beta X')|X^*, X]$$
$$= \sigma_\beta^2 E[X^* X']$$

cuatro $$E(f^*|y) = E_{f|y}\left[\sigma_\beta^2 E[X^* X'](XX'\sigma_\beta^2 + I\sigma_e^2)^{-1} f\right]$$
$$= \sigma_\beta^2 E[X^* X'] E\left[(XX'\sigma_\beta^2 + I\sigma_e^2)^{-1}\right]\hat{f}$$

cinco DOES ANYBODY KNOW HOW TO COMPUTE THE ABOVE?
(CALCULATING THE PEV IS EVEN MORE INVOLVED)

Reliability: standard formulae

$$reliability = 1 - \frac{PEV}{Var(u)}$$

$$unreliability = 1 - reliability = \frac{PEV}{Var(u)}$$

$$PEV = Var(\hat{u} - u) = Var_{\hat{u}}[E(\hat{u} - u)|\hat{u}] + E_{\hat{u}}(Var[(\hat{u} - u)|\hat{u}])$$
$$= E_{\hat{u}}(Var[(\hat{u} - u)|\hat{u}])$$
$$= E_{\hat{u}}(Var[u|\hat{u}])$$

$$Var(u) = E_{\hat{u}}(Var[u|\hat{u}]) + Var_{\hat{u}}(E[u|\hat{u}])$$
$$E_{\hat{u}}(Var[u|\hat{u}]) \leq Var(u)$$
$$\text{Under MULVN} \Rightarrow E_{\hat{u}}(Var[u|\hat{u}]) = Var[u|\hat{u}]$$

IN SOME NON-GAUSSIAN MODELS, POSTERIOR VARIANCE CAN BE SOMETIMES LARGER THAN PRIOR VARIANCE, LEADING TO NEGATIVE RELIABILITY , AND POSITIVE UNRELIABILITY.

POINT 3: **WHAT WE CALL "RELIABILITY" IS VERY MUCH TAILORED FOR NORMAL DISTRIBUTIONS AND LINEAR MODELS**

IS MY MODEL "RIGHT"?

TAKING MODEL UNCERTAINTY INTO ACCOUNT BY MODEL AVERAGING

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

THE PUNCH LINE: VARIANCE OF PREDICTION ERRORS TAKING MODEL UNCERTAINTY INTO ACCOUNT

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

Average PEV

Variance among predictions from different models

CROSS-VALIDATION

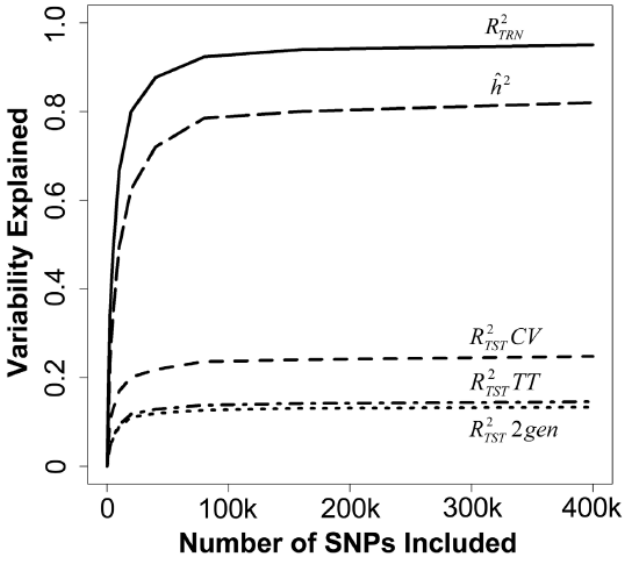
(take model uncertainty into account: never did this in the BLUP era)

➔A. Prediction and goodness of fit are different ball games: a model that fits well to training data may have atrocious predictive ability

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

THIS IS THE DISTRIBUTION THAT MATTERS AND NOT A MODEL DERIVED QUANTITY, THAT IGNORES UNCERTAINTY ABOUT THE MODEL!!!!!!

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



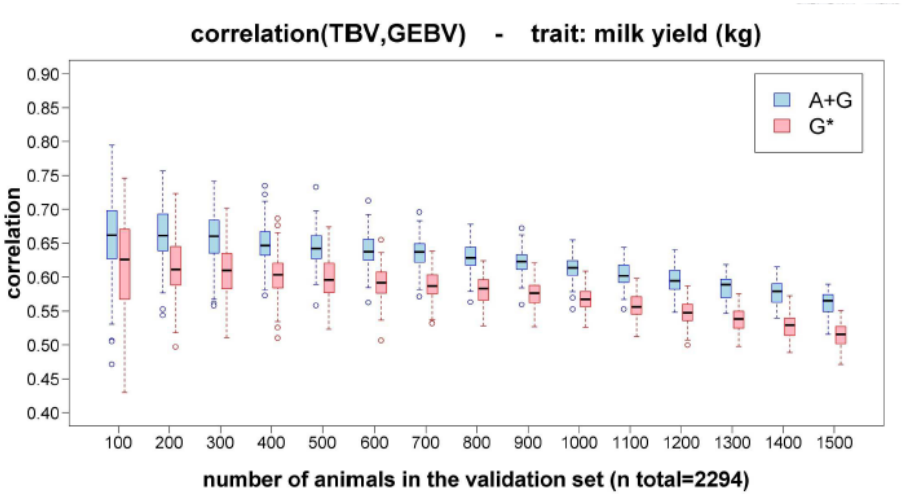
HUMAN STATURE: MAKOWSKY et al. , Plos Genetics 2011

REASONABLE BAYESIAN MODEL

- For any parameter, must be able to “kill” the prior asymptotically
- For any parameter, statistical distance between prior and posterior (and therefore conditional posterior) must go to infinity
- If this distance has a finite upper bound, it means that the prior is influential
- Must be able to reduce statistical entropy as conveyed by the prior by a sizable amount. If the reduction is tiny➔ prior very influential

CROSS-VALIDATION UNCERTAINTY

(Erbe et al. 2010)



THE CURSE OF THE BAYESIAN ALPHABET

Sarah Palin sings "To Russia with love, a view from my igloo"

Featuring

Halle Berry, as "A"

AND…

Kim-Jong Il, as "Bayes"

Scarlett Johansson as "B"

RECALL FROM EARLY PART OF COURSE

STATE OF KNOWLEDGE
(in a finite sample)

Minimum➔	Prior
Maximum➔	Conditional posterior
Intermediate➔	Marginal posterior

THE PROCESS OF DECONDITIONING CONSUMES INFORMATION ABOUT THE FOCAL POINT

Meaning: conditional posterior is the best world to live in

BAYES A + BAYES B

(as I understand them)

Linear model proposed by Meuwissen et al. (2001)

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

SCALAR

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

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

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

Additive effect of SNP j

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

$$\mu \sim uniform$$

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

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

$$\sigma_{b_j}^2 \sim \nu S^2 \chi_{\nu}^{-2} \text{ for all } j$$

Hyper-parameters, specified arbitrarily

BAYES A (Meuwissen et al., 2001)

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

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

Note: each SNP has a variance (think of a sire model in which each sire effect has a variance)

Marginal prior

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

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

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

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

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

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

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

Bayes B is Bayesianly "STRANGE"

Bayes B

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

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

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

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

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

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

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

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 $\boldsymbol{\theta} = [P_1, \ldots, P_K, \boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K, \sigma_1^2, \ldots, \sigma_K^2]'$, where the mixture proportions P_k are such that $\sum_{k=1}^K P_k = 1$, are

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

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

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

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 VARIANCE IN BAYES B!

BAYES A IS A SPECIAL CASE OF BAYES B (π=0)

Meaning: if Bayes A has a flaw, this carries to Bayes B

A Gibbs sampler for Bayes A

(element-wise sampling)

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

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

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**. One can increase the number of data 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

STATE OF KNOWLEDGE

Minimum➔ Prior
Maximum➔ Conditional posterior
Intermediate➔ Marginal posterior

For any parameter θ of the model, Bayesian learning must be such that the posterior coefficient of variation, that is $CV = \frac{\sqrt{Var(\theta|DATA)}}{E(\theta|DATA)}$, tends to 0, asymptotically. This does not happen in Bayes A or Bayes B for $\sigma^2_{a_k}$. In Bayes A, the prior coefficient of variation is

$$CV\left(\sigma^2_{a_k}\right) = \sqrt{\frac{2}{(\nu-4)}}$$

whereas the coefficient of variation of the fully conditional posterior distribution is

$$CV\left(\sigma^2_{a_k}|ELSE\right) = \sqrt{\frac{2}{(\nu-3)}},$$

so that $CV\left(\sigma^2_{a_k}|ELSE\right)/CV\left(\sigma^2_{a_k}\right) = \sqrt{1-\frac{1}{(\nu-3)}}$. This ratio goes to 1 rapidly as the degrees of freedom of the prior increase (meaning that the prior "dominates" inference), as illustrated in Figure 1. For example, if $\nu = 4.1, \nu = 5.1$ and $\nu = 6.1$, the ratio between the coefficients of

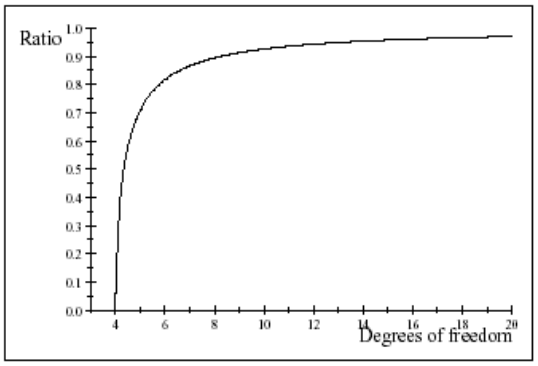


Figure 1. Ratio between coefficients of variation $CV\left(\sigma^2_{a_k}|ELSE\right)/CV\left(\sigma^2_{a_k}\right) = \sqrt{1-\frac{1}{(\nu-3)}}$ of the conditional posterior and prior distributions of the variance of the marker effect, as a function of the degrees of freedom ν of the prior.

For df>6, the relative variability of the posterior distribution of the variance of a SNP effect is essentially COPYING that of their prior distribution

ENTROPY CALCULATIONS

Prior entropy

Variance of marker effect (sorry, change of notation)

$$H\{[\sigma^2_{a_k}|\upsilon, S^2]\}$$

$$= -\int \log[p(\sigma^2_{a_k}|\upsilon, S^2)]p(\sigma^2_{a_k}|\upsilon, S^2)d\sigma^2_{a_k}$$

$$= -\frac{\nu}{2} - \log\left[\frac{\nu S^2}{2}\Gamma\left(\frac{\nu}{2}\right)\right] + \left(1+\frac{\nu}{2}\right)\frac{d}{d\left(\frac{\nu}{2}\right)}\log\Gamma\left(\frac{\nu}{2}\right).$$

Entropy of the conditional posterior

$$H\{[\sigma^2_{a_k}|ELSE]\}$$

$$= -\int \log[p(\sigma^2_{a_k}|ELSE)]p(\sigma^2_{a_k}|ELSE)d\sigma^2_{a_k}$$

$$= -\frac{\nu+1}{2} - \log\left[\left(\frac{\nu S^2 + a_k^2}{2}\right)\Gamma\left(\frac{\nu+1}{2}\right)\right] + \left(1+\frac{\nu+1}{2}\right)\frac{d}{d\left(\frac{\nu+1}{2}\right)}\log\Gamma\left(\frac{\nu+1}{2}\right).$$

Learning from data: reduces entropy (cannot calculate entropy of posterior)

Relative information gain

$$RIG = \frac{H\left\{\left[\sigma^2_{a_k}|\upsilon,S^2\right]\right\} - H\left\{\left[\sigma^2_{a_k}|ELSE\right]\right\}}{H\left\{\left[\sigma^2_{a_k}|\upsilon,S^2\right]\right\}}$$

For $a_k = 0,\ S = 1$ and $\nu = 100,\ RIG = 9.60 \times 10^{-3}$

For $a_k = 0,\ S = 1$ and $\nu = 10,\ RIG = 6.51 \times 10^{-2}$

For $a_k = 0,\ S = 1$ and $\nu = 4,\ RIG = 0.125$

Metaphorically: the prior is totalitarian in Bayes A (B)

STATISTICAL DISTANCE BETWEEN CONDITIONAL POSTERIOR AND PRIOR (KULLBACK-LEIBLER)

Specific distance at a given variance

$$KL\left[conditional, prior\right] = \int L\left(\upsilon, \upsilon + p, S^2, \mathbf{a}_m\right) p\left(\sigma^2_{a_k}|\upsilon, S^2\right) d\sigma^2_{a_k},$$

prior

where

$$L\left(\upsilon, \upsilon + p, S^2, \mathbf{a}_m, \sigma^2_{a_k}\right) = \log \frac{p\left(\sigma^2_{a_k}|\upsilon, S^2\right)}{p\left(\sigma^2_{a_k}|ELSE\right)},$$

➔IF KL IS LARGE, THEN LEARNING BEYOND THE PRIOR HAS TAKEN PLACE.

➔KL SHOULD GO TO INFINITY AS DATA ACCUMULATE IN ANY REASONABLE BAYESIAN MODEL

KULLBACK-LEIBLER DISTANCES BETWEEN CONDITIONAL POSTERIOR AND PRIOR

1) 7.33×10^{-2} for $\nu = 4, S = 1, p = 1$ and $a_k = 0$

2) 2.64×10^{-2} for $\nu = 10, S = 1, p = 1$ and $a_k = 0$

3) 2.52×10^{-3} for $\nu = 100, S = 1, p = 1$ and $a_k = 0$

If 10 markers are allowed to share the same variance, KL= 4.47
Relative to (1), KL distance increases 61 times…

Effect of the scale parameter of the prior

A pertinent question is whether or not the learned marker effect (i.e., a draw from its conditional posterior distribution) has an important impact on KL via modification of the scale parameter from S^2 into $\frac{\nu S^2 + a_k^2}{\nu+1}$. Let $c = \frac{a_k}{S}$ be the realized value of the marker effect in units of the "prior standard deviation" S, with $c = 0, 0.01, 0.5, 1$ and 2; the last two cases would be representative of markers with huge effects. The density of the conditional posterior distribution of $\sigma^2_{a_k}$ is then

$$p\left(\sigma^2_{a_k}|ELSE\right) = \frac{\left(\frac{(\nu+c^2)S^2}{2}\right)^{\frac{\nu+1}{2}}}{\Gamma\left(\frac{\nu+1}{2}\right)} \left(\sigma^2_{a_k}\right)^{-\frac{\nu+1+2}{2}} \exp\left(-\frac{(\nu+c^2)S^2}{2\sigma^2_{a_k}}\right).$$

The KL distance between the conditional posterior and the prior for each of these five situations, assuming $S = 1$ and $\nu = 4$ are: 1) $\mathrm{KL}(c = 0) = 7.33 \times 10^{-2}$, 2) $\mathrm{KL}(c = 0.01) = 7.32 \times 10^{-2}$, 3) $\mathrm{KL}(c = 0.5) = 4.67 \times 10^{-2}$, 4) $\mathrm{KL}(c = 1) = 1.54 \times 10^{-2}$ and 5) $\mathrm{KL}(c = 2) = 0.34$. Even though marker effects are drastically different, the conditional posteriors are not two different (in the KL sense) from each other, meaning that the exent of shrinkage in Bayes A (or B) continues to be dominated by the prior.

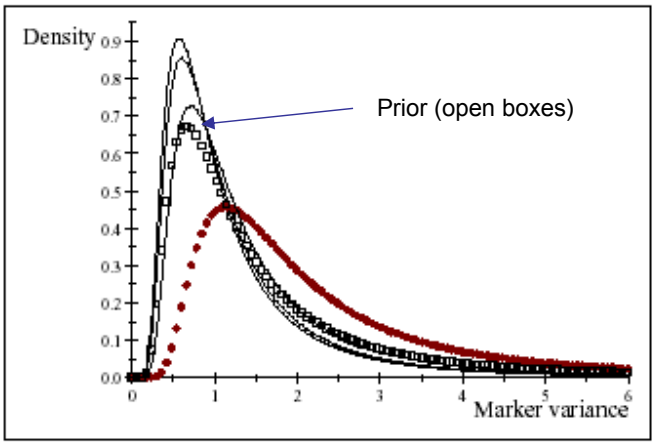


Figure 3. Effect of scale parameter on the conditional posterior distribution of the variance of the marker effect. Prior distribution marked with boxes; conditional posterior distribution for $c = 2$ (standardized marker effect) in dots. The other three distributions are barely distinguisable from the prior.

BAYES A (B)

- The prior always matters
- The effect of the prior is via the extent of shrinkage of marker effects
- The extent of shrinkage can be manipulated, with the data essentially providing no control
- Statistically greedy models (same will apply for any model assigning marker-specific variances)

SIMULATION

(never take a simulation too seriously)

RESURRECTION OF BAYES A

(If additive model holds, it may give sensible inferences about marker effects)

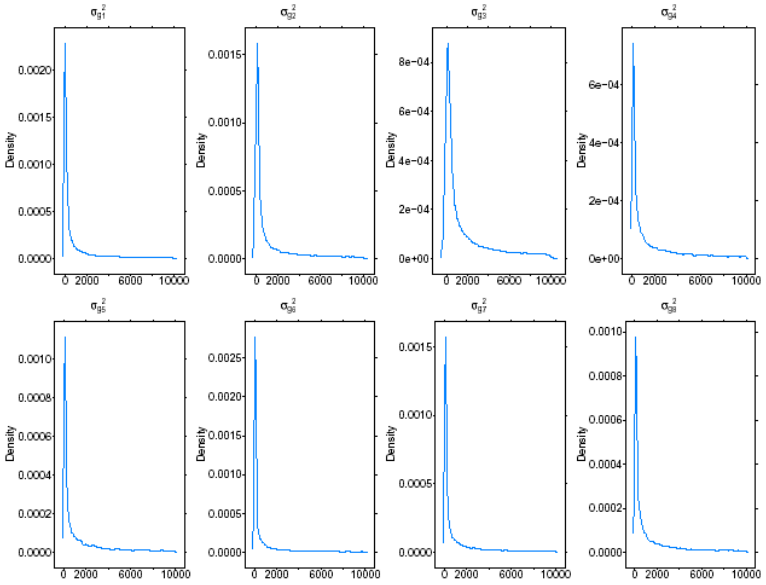
Description for slides 1, 2 and 3

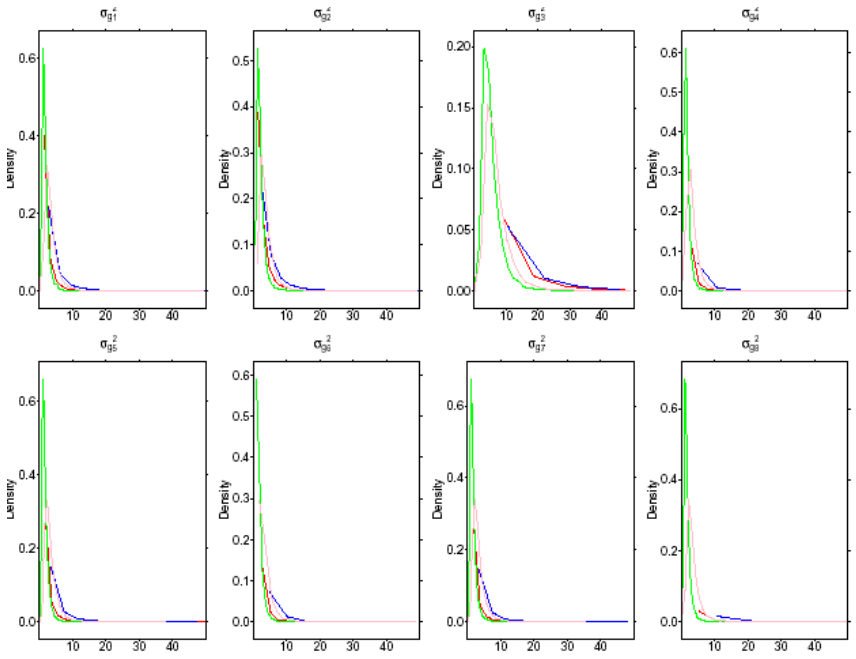
- **Bayes A was fitted on a simulated data of 50 observations.**
  - True linear relationship between response and SNP (x1 x2 x3) effects
    - $Y = w_1 + 2*w_2 + x_1 - 2*x_2 + 5*x_3 + \text{error}$ (~N(0, sd=1.2))
  - Model fitted:
    - $Y = W\beta + Xg + \text{error}$
    - W is incidence matrix for two nuisance parameters.
    - X is incidence matrix for SNP effects. Besides x1, x2 and x3, five additional irrelevant SNPs (x4 to x8) added. SNP value is allele copy numbers, i.e., 0, 1 or 2

- **Slide 1—Posterior distributions of SNP effects g_i (i = 1, 2,…,8) when using five different priors on σ_{gi}^2, scale determined by estimated residual variance (1.5)**
  - Black: σ_{gi} ~ unif(0,100)
  - Red: σ_{gi}^2 ~ scaled inverse χ^2 (df=4, scale=1.5)
  - Blue: σ_{gi}^2 ~ scaled inverse χ^2 (df=4, scale=3)
  - Green: σ_{gi}^2 ~ scaled inverse χ^2 (df=8, scale=1.5)
  - Pink: σ_{gi}^2 ~ scaled inverse χ^2 (df=8, scale=3)

- **Slides 2 and 3—Posterior distributions of 8 SNP specific variances under above five priors. Because the uniform prior leads to a very different posterior of SNP variance as compared to the other four priors, it was plotted separately (slide 2). Slide 3 is for the four scaled inverse chi-square priors, with same color representations in slide 1.**

POSTERIOR DISTRIBUTION OF SNP VARIANCES UNDER UNIFORM PRIOR

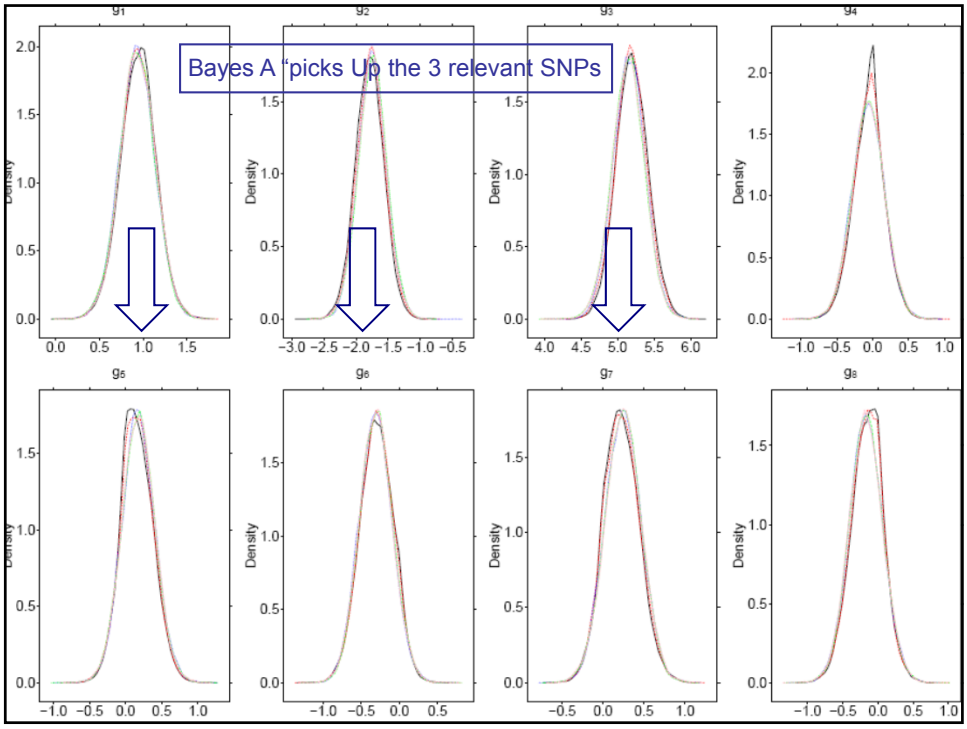




POSTERIOR DISTRIBUTION OF SNP VARIANCES UNDER FOUR S.INV. CHI-SQUARE PRIORS

THE GOOD NEWS

Posterior distribution of SNP effects



DEATH-RESURRECTION-DEATH

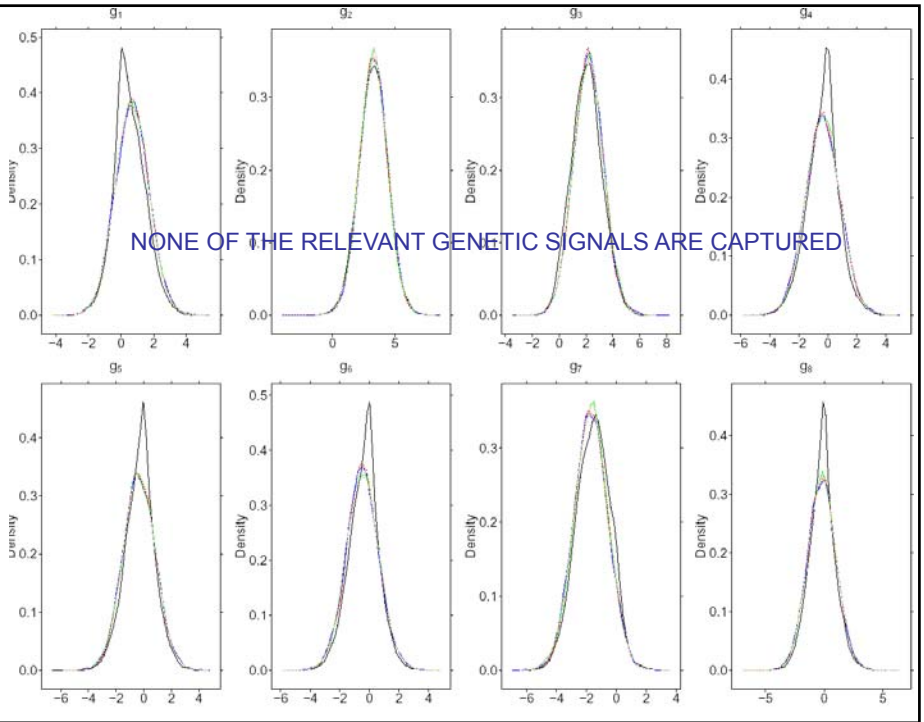
Bayes A may give a distorted picture if there is non-linearity or non-additivity

Description for slides 5, 6 and 7

- **Bayes A was fitted on a simulated data of 50 observations.**
  - True nonlinear relationship between response and SNP (x1 x2 x3) effects
    - $Y = w_1 + 2*w_2 + \exp(x_1)*\sin(x_2-0.5)*x_3^2 + \text{error}$ (~N(0, sd=0.25))
  - Model fitted:
    - $Y = W\beta + Xg + \text{error}$
    - W is incidence matrix for two nuisance parameters.
    - X is incidence matrix for SNP effects. Besides x1, x2 and x3, five additional irrelevant SNPs (x4 to x8) added. SNP value is allele copy numbers, i.e., 0, 1 or 2

- **Slide 5—Posterior distributions of SNP effects g_i (i = 1, 2,…,8) when using five different priors on σ_{gi}^2, scale determined by estimated residual variance (42)**
  - Black: σ_{gi}^2 ~ unif(0,100)
  - Red: σ_{gi}^2 ~ scaled inverse χ^2 (df=4, scale=42)
  - Blue: σ_{gi}^2 ~ scaled inverse χ^2 (df=4, scale=84)
  - Green: σ_{gi}^2 ~ scaled inverse χ^2 (df=8, scale=42)
  - Pink: σ_{gi}^2 ~ scaled inverse χ^2 (df=8, scale=84)

- **Slides 6 and 7—Posterior distributions of 8 SNP specific variances under the above five priors. Because the uniform prior leads to a very different posterior of SNP variance as compared to the other four priors, it was plotted separately (slide 6). Slide 7 is for the four scaled inverse chi-square priors, with same color representations in slide 5.**



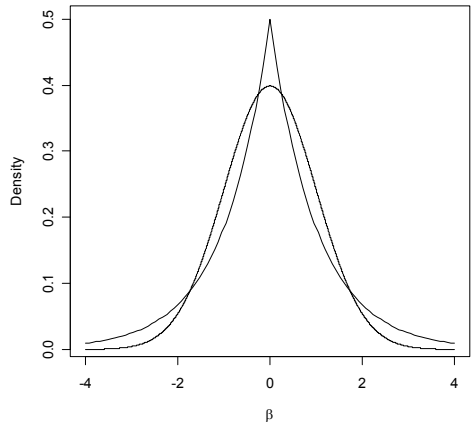
BAYES A vs. BAYES L

(Bayes L= Bayesian Lasso)

In the Bayesian Lasso, marker effects are assigned double exponential distributions

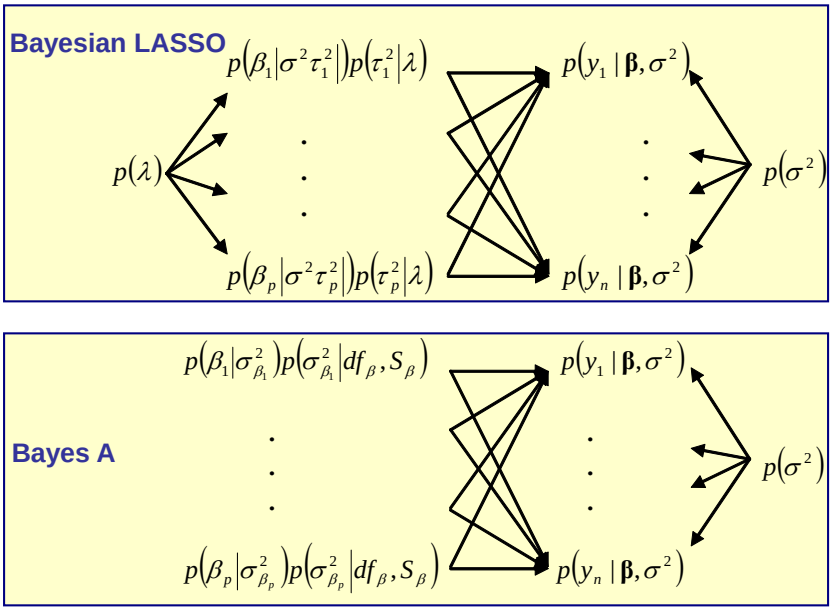
$$p(\boldsymbol{\beta}) = \prod_{j=1}^{p} \frac{1}{2} \lambda e^{-\lambda|\beta_j|}$$

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



Density of a Normal and of a Double-Exponential Distribution

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



Assume exponential distribution of variances

$$p\left(\sigma^2_{e_i}|\frac{\lambda^2}{2}\right) = \frac{\lambda^2}{2}\exp\left(-\frac{\lambda^2\sigma^2_{e_i}}{2}\right)$$

Mix (as in t-model)

$$p(y_i|\mu_i,\lambda) = \int_0^\infty N(y_i|\mu_i,\sigma^2_{e_i})\frac{\lambda^2}{2}\exp\left(-\frac{\lambda^2\sigma^2_{e_i}}{2}\right)d\sigma^2_e$$

$$= \frac{\lambda^2}{2\sqrt{2\pi}}\int_0^\infty(\sigma^2_{e_i})^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\left[\frac{(y_i-\mu_i)^2}{\sigma^2_{e_i}}+\lambda^2\sigma^2_{e_i}\right]\right\}d\sigma^2_{e_i};\quad i=1,2,\ldots,n.$$

Assume

$$\lambda^2|a,b \sim Gamma(a,b)$$

Implementation is as in a t-model but transform $\tau_{e_i}^2 = \frac{1}{\sigma_{e_i}^2}$

$$p(\tau_{e_i}^2|ELSE) \propto (\tau_{e_i}^2)^{-\frac{3}{2}} \exp\left\{-\frac{\lambda^2}{2\tau_{e_i}^2 \frac{\lambda^2}{(y_i-\mathbf{x}_i'\boldsymbol{\beta}-z_i'\mathbf{u})^2}}\left(\tau_{e_i}^2 - \sqrt{\frac{\lambda^2}{(y_i-\mathbf{x}_i'\boldsymbol{\beta}-z_i'\mathbf{u})^2}}\right)^2\right\}$$

Inverse Gaussian (Wald) distribution

$$E(\tau_{e_i}^2|ELSE) = \sqrt{\frac{\lambda^2}{(y_i-\mathbf{x}_i'\boldsymbol{\beta}-z_i'\mathbf{u})^2}}$$

$$Var(\tau_{e_i}^2|ELSE) = \frac{E^3(\tau_{e_i}^2|ELSE)}{\lambda^2} = \frac{\left(\frac{\lambda^2}{(y_i-\mathbf{x}_i'\boldsymbol{\beta}-z_i'\mathbf{u})^2}\right)^{\frac{3}{2}}}{\lambda^2}$$

ANOTHER SIMULATION

(never take simulation too seriously, although it is great for checking ideas and code)

DE LOS CAMPOS ET AL. (2009)

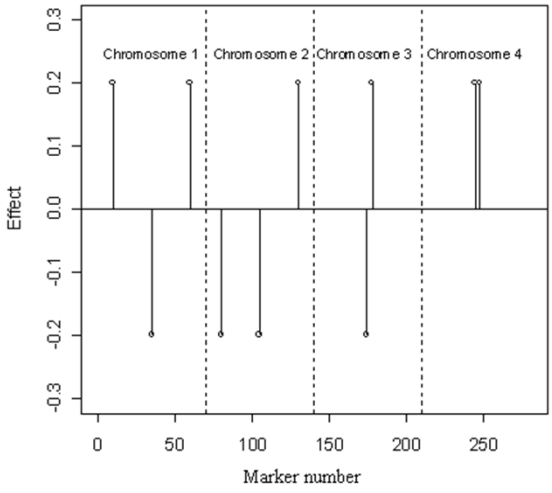
$$y_i = \sum_{j=1}^{280} x_{ij}\beta_j + \varepsilon_i \qquad i = 1,...,300.$$

280 markers. Residuals assumed N(0,1)

Pearson's correlation between marker genotypes (average across markersand 100 Monte-Carlo simulations) by scenario (X_0: low LD; X_1 high LD).

	Adjacency between markers			
Scenario	1	2	3	4
X_0	0.007	0.002	-0.002	0.013
X_1	0.722	0.567	0.450	0.356

Only 10 markers had effects➔ 270 had no effect on the trait simulated



Positions (chromosome and marker number) and effects of markers (there were 280 markers, with 270 with no effect)

NINE SPECIFICATIONS OF BAYES A

Prior df	Prior Scale		
	10^{-5}	10^{-3}	$5x10^{-2}$
0	(1)	(2)	(3)
½	(4)	(5)	(6)
1	(7)	(8)	(9)

PRIORS 1, 2, 3 ARE IMPROPER
PRIORS 7, 8, 9 WOULD LEAD TO CAUCHY PRIOR DISTRIBUTION OF MARKER EFFECTS IF SCALE WERE 1

Table 3. Posterior estimates of residual variance (σ^2) and correlation between the true and estimated value for several items (y , phenotypes; Xß , true genomic value ; ß , marker effects; all quantities averaged of 100 MC replicates).

	σ^2		$Corr(\mathbf{y}, \mathbf{X}\hat{\boldsymbol{\beta}})$		$Corr(\mathbf{X}\boldsymbol{\beta}, \mathbf{X}\hat{\boldsymbol{\beta}})$		$Corr(\boldsymbol{\beta}, \hat{\boldsymbol{\beta}})$	
	Mean 1/	SD 2/	Mean 3/	SD 2/	Mean 3/	SD 2/	Mean 3/	SD 2/
Low linkage disequilibrium between markers (X_0)								
Bayes A:								
(1)	0.518	0.062	0.839	0.027	0.580	0.063	0.102	0.048
(2)	0.941	0.089	0.577	0.028	0.721	0.092	0.200	0.022
(3)	1.074	0.105	0.496	0.032	0.701	0.106	0.199	0.020
(4)	0.394	0.053	0.895	0.022	0.531	0.060	0.079	0.051
(5)	0.824	0.077	0.652	0.025	0.699	0.079	0.183	0.028
(6)	0.950	0.089	0.578	0.027	0.722	0.088	0.201	0.021
(7)	0.173	0.053	0.966	0.015	0.455	0.057	0.042	0.043
(8)	0.575	0.056	0.813	0.019	0.606	0.066	0.116	0.044
(9)	0.710	0.066	0.728	0.020	0.659	0.072	0.152	0.037
BL	0.886	0.080	0.623	0.028	0.708	0.081	0.191	0.024

1/: Mean (across 100 MC replicates) of the posterior mean. 2/: Between-replicate standard deviation of the estimate. 3/: Mean (across MC replicates) of the correlation evaluated at the posterior mean of ß .

Table 3. Posterior estimates of residual variance (σ^2) and correlation between the true and estimated value for several items (y , phenotypes; Xß , true genomic value ; ß , marker effects; all quantities averaged of 100 MC replicates).

	σ^2		$Corr(\mathbf{y}, \mathbf{X}\hat{\beta})$		$Corr(\mathbf{X}\beta, \mathbf{X}\hat{\beta})$		$Corr(\beta, \hat{\beta})$	
	Mean[1/]	SD[2/]	Mean[3/]	SD[2/]	Mean[3/]	SD[2/]	Mean[3/]	SD[2/]
High linkage disequilibrium between markers (X_1)								
Bayes A:								
(1)	0.535	0.069	0.824	0.029	0.580	0.070	0.121	0.045
(2)	0.938	0.076	0.609	0.033	0.677	0.083	0.210	0.026
(3)	1.093	0.085	0.528	0.034	0.650	0.086	0.211	0.025
(4)	0.404	0.067	0.888	0.025	0.533	0.067	0.094	0.048
(5)	0.809	0.069	0.670	0.030	0.659	0.076	0.200	0.030
(6)	0.948	0.075	0.616	0.031	0.676	0.081	0.211	0.026
(7)	0.195	0.056	0.960	0.015	0.462	0.060	0.062	0.048
(8)	0.566	0.058	0.809	0.021	0.593	0.070	0.132	0.042
(9)	0.689	0.062	0.734	0.024	0.629	0.072	0.173	0.036
BL	1.004	0.088	0.610	0.042	0.668	0.079	0.211	0.025

1/: Mean (across 100 MC replicates) of the posterior mean. 2/: Between-replicate standard deviation of the estimate. 3/: Mean (across MC replicates) of the correlation evaluated at the posterior mean of ß .

Ability to uncover relevant genomic regions

- For each method and replicate, markers ranked on absolute values of posterior means
- For each effect, dummy variable created
- Dummy was 1 if marker (or any of its 4 flanking markers) ranked on top 20. O.W= 0
- Average over markers and replicates➔ Index of "retrieved regions"

Table 4. Fraction of retrieved regions by model (BL=Bayesian LASSO) and scenario of linkage disequilibrium (LD).

	Bayes A									BL
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
Low LD (X_0)	0.24	0.43	0.47	0.22	0.36	0.43	0.22	0.26	0.29	0.39
High LD (X_1)	0.21	0.34	0.33	0.19	0.31	0.34	0.18	0.22	0.26	0.34

Bayes A affected by priors:
➔Worse performance in Settings 1, 4 and 7
➔Bayes A (settings 2, 3 , 6) and Lasso almost doubled ability

Simple fixes of Bayes A

- Assign the same variance to all markers (trivial Bayesian regression problem)
- Assign the same variance to groups of markers (e.g., chromosomes or genomic regions): model comparison issue
- Assign non-informative priors to $\underline{S}$ and to the degrees of freedom $\underline{\nu}$

  ➔ can be done. Just an algorithmic matter

Issues and questions

- Bayes A can be "fixed", but may not the best thing to do. Open question…
- Bayes A, as is, may still have a good predictive (out of sample) behavior, even though it is not completely defensible
- Bayes B is Bayesianly ill-posed. If you do not believe me, check with local Bayesian statisticians…
- More reasonable: mixture at the level of the effects (not of the variances): I believe this is what the Dutch did (and Iowa people with beef cattle, mainly Fernando and Garrick)