OVERVIEW OF BAYESIAN INFERENCE: PART 1

Rev. Thomas Bayes

1702 London, England
1761 Tunbridge Wells, Kent, England

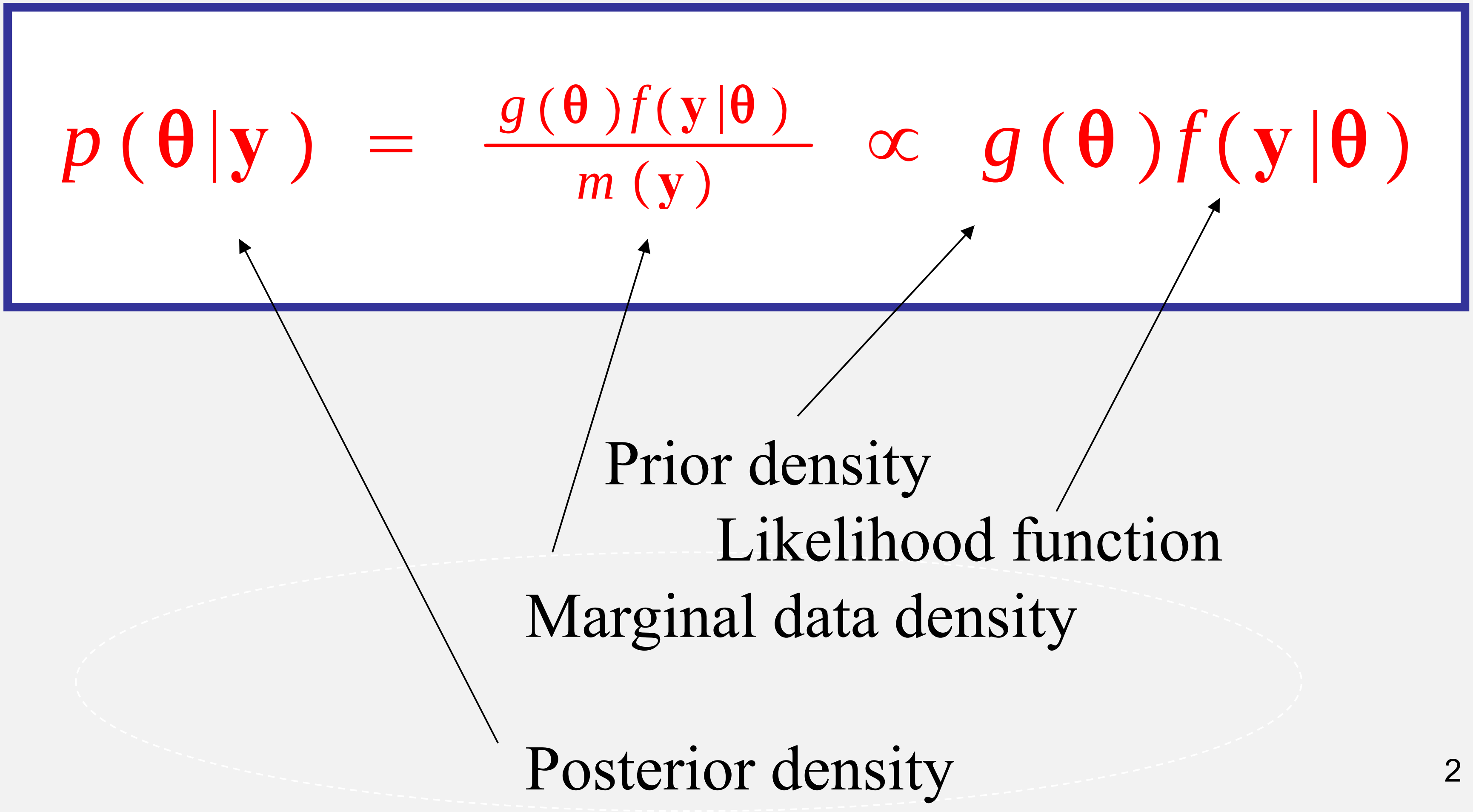
1763. "An essay towards solving a problem in the doctrine of chances". *Philosophical Transactions of the Royal Society of London* **53**, 370-418.

Pierre-Simon Laplace

1749 Beaumont-en-Auge, France
1827 Paris, France

1774. "Mémoire sur la probabilité des causes par les événements". *Savants étranges* **6**, 621-656. *Oeuvres* **8**, 27-65

BAYES THEOREM



$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{g(\boldsymbol{\theta})f(\mathbf{y}|\boldsymbol{\theta})}{m(\mathbf{y})} \propto g(\boldsymbol{\theta})f(\mathbf{y}|\boldsymbol{\theta})$$

Prior density

Likelihood function

Marginal data density

Posterior density

THE BAYESIAN APPROACH IN A NUTSHELL

- All unknowns in statistical system treated as random even if they are fixed entitities
- Randomness reflects (typically) **subjective** uncertainty
- Can include as unknowns:

➔ <u>The model </u>(distribution, functional form)

➔Its <u>parameters</u> (heritability, inbreeding coefficient, genetic relatedness, genetic effects, number of QTL loci, marker effects)

- Combine with what is known a priori with information from data: <u>Bayesian learning</u>
- Bayesian approach can also be used for:

PREDICTION of future observations

(without taking inference too seriously)

BASIC DEFINITIONS IN STATISTICAL THEORY

θ : a quantity to be inferred

$\tilde{\theta}(\mathbf{y})$: a statistic used to learn θ

IF θ is a fixed (frequentist) item: $\tilde{\theta}(\mathbf{y})$ is called "estimator"

IF θ is a random (frequentist) item : $\tilde{\theta}(\mathbf{y})$ is called "predictor"

IF $E_{y|\theta}\left[\tilde{\theta}(\mathbf{y})\right] = \theta$ the estimator is unbiased

IF $E_{y,\theta}\left[\tilde{\theta}(\mathbf{y})\right] = E_{\theta}(\theta)$ the predictor is unbiased

Bayesian wish to obtain the entire posterior : $[\theta|\mathbf{y}]$

Point estimates are posterior mean, mode and median : $E(\theta|\mathbf{y})$

If θ admits a frequentist distribution then: $E_{\mathbf{y}}[E_{\theta|\mathbf{y}}(\theta|\mathbf{y})] = E(\theta)$

$E(\theta|\mathbf{y})$: is unbiased with respect to $E(\theta)$ but not w.r.t θ

SOME BAYESIAN DISTRIBUTIONS

$$\begin{aligned}
p(\theta, y | hyper - parameters) &= \text{joint} \\
p(\theta | hyper - parameters) &= \text{prior} \\
p(y | hyper - parameters) &= \text{marginal of the data or "prior predictive"} \\
p(y_m, \theta, y | hyper - parameters) &= \text{augment joint with "missing" items} \\
y_m &= \text{missing (truly or future or latent)} \\
p(\theta | y, hyper - parameters) &= \text{posterior} \\
p(\theta_A | \theta_{-A}, y, hyper - parameters) &= \text{conditional posterior} \\
p(y_m | y) &= \text{"posterior predictive"}
\end{aligned}$$

EXAMPLE:
HOW DOES ONE FORMULA DOES IT ALL
: missing data or predictive distribution

$$
\begin{aligned}
& p(y_m|y, hyper-parameters) \\
&= \int_{\Re_\theta} \frac{p(y_m, \theta, y|hyper-parameters)}{p(y|hyper-parameters)} d\theta \\
&= \int_{\Re_\theta} \frac{p(y_m|\theta, y, hyper-parameters)p(y|\theta)p(\theta|hyper-parameters)}{p(y|hyper-parameters)} d\theta \\
&= \int_{\Re_\theta} \frac{p(y_m|\theta, y, hyper-parameters)p(\theta|y, hyper-parameters)p(y|hyper-parameters)}{p(y|hyper-parameters)} d\theta \\
&= \int_{\Re_\theta} p(y_m|\theta, y, hyper-parameters)p(\theta|y, hyper-parameters)d\theta
\end{aligned}
$$

IF y_m INDEPENDENT OF y GIVEN $\theta \Rightarrow$

$$
= \int_{\Re_\theta} p(y_m|\theta)p(\theta|y, hyper-parameters)d\theta
$$

IMPLICATION ON SAMPLING METHODS

Example: Infer mean of normal distribution with **known** variance

Sampling model

$y_i|\mu,\sigma^2 \sim N(\mu,\sigma^2)$ known σ^2 unknown μ

Assume prior : $\mu|\mu_0,\sigma^2 \sim N(\mu_0,\tau^2)$

Prior

Sample size : n

Posterior density

Posterior is: $\mu|\mu_0,\sigma^2,\mathbf{y} \sim N\left(\bar{\mu},\overline{\sigma^2}\right)$

$$\bar{\mu} = \left(\frac{1}{\tau^2}+\frac{n}{\sigma^2}\right)^{-1}\left(\frac{1}{\tau^2}\mu+\frac{n}{\sigma^2}\bar{y}\right)$$

μ_0

$$= \mu_0 + \frac{\tau^2}{\tau^2+\frac{\sigma^2}{n}}(\bar{y}-\mu_0)$$

$$= \mu_0 + \frac{n}{n+\frac{\sigma^2}{\tau^2}}(\bar{y}-\mu_0)$$

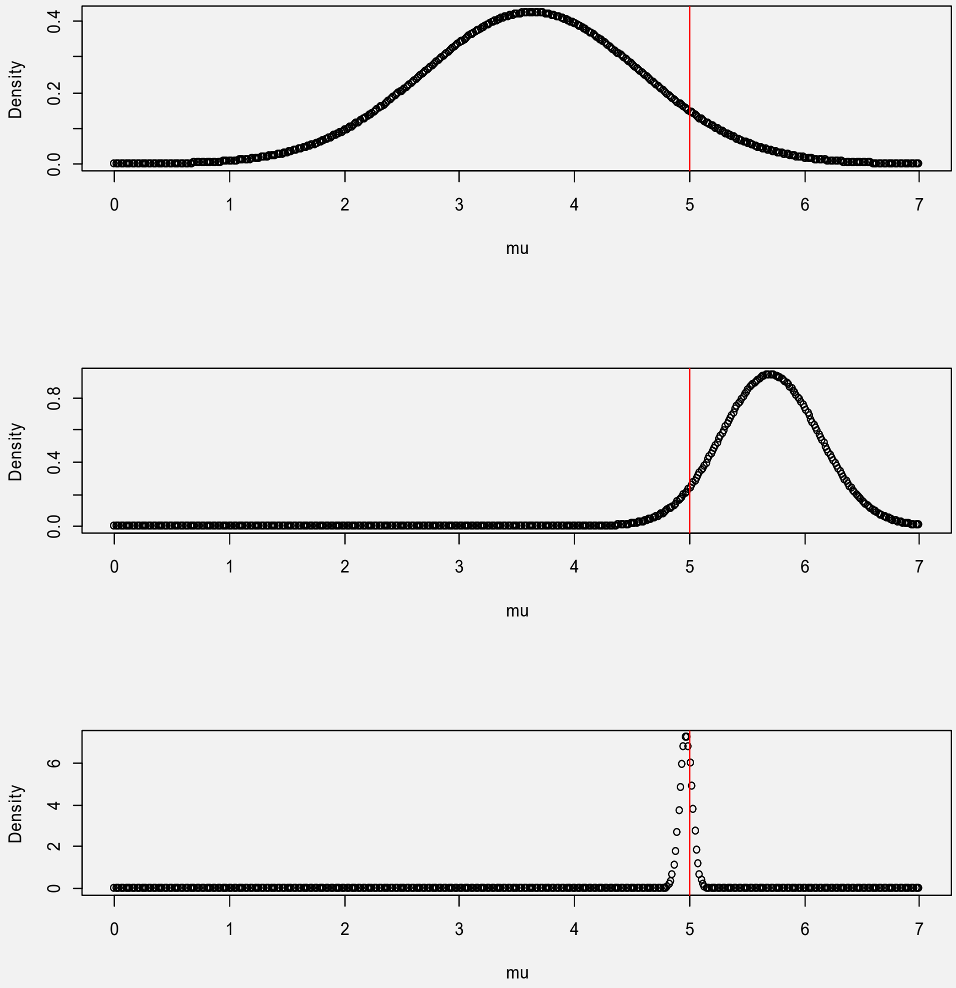
shrinkage

$$\overline{\sigma^2} = \left(\frac{1}{\tau^2}+\frac{n}{\sigma^2}\right)^{-1} = \left(n+\frac{\sigma^2}{\tau^2}\right)^{-1}\sigma_e^2$$

Note: normal data sampling model "conjugates: with normal prior producing normal posterior

```
###UNKNOWN MEAN KNOWN VARIANCE, NORMAL
DISTRIBUTION
#####DEFINITIONS
##Sampling model: normal
truemu<-5
sigma2<-9
#####SIMULATE 3 SAMPLES OF SIZE n=10,50,3000
samplesize<-c(10,50,3000)
ybar<-numeric(length(samplesize))
for (i in 1:length(samplesize)){
ybar[i]<-sum(rnorm(samplesize[i],truemu,sqrt(sigma2)))/samplesize[i]
}
Ybar
##Hyper-parameters of normal prior
mu0<-1
tau2<-81

###Calculate posterior means and variances
postmean<-numeric(length(samplesize))
postvar<-numeric(length(samplesize))
for (i in 1:length(samplesize)){
postmean[i]<-mu0+samplesize[i]*(ybar[i]-
mu0)/(samplesize[i]+sigma2/tau2)
postvar[i]<-sigma2/(samplesize[i]+sigma2/tau2)
}
postmean
postvar
###PLOT POSTERIORS. VERTICAL RED LINE IS truemu=5
muplot<-seq(0,7,length=500)
par(mfrow=c(3,1))
plot(muplot,dnorm(muplot,mean=postmean[1],sd=sqrt(postvar[1])),
xlab="mu",ylab="Density")
abline(v = 5,col="red")
plot(muplot,dnorm(muplot,mean=postmean[2],sd=sqrt(postvar[2])),
xlab="mu",ylab="Density")
abline(v = 5,col="red")
plot(muplot,dnorm(muplot,mean=postmean[3],sd=sqrt(postvar[3])),
xlab="mu",ylab="Density")
abline(v = 5,col="red")
par(mfrow=c(1,1))
```

AS SAMPLE SIZE INCREASES POSTERIOR DISTRIBUTION GETS SHARPER. WHEN n GOES TO INFINITY CONVERGES TO TRUE VALUE



BAYESIAN LEARNING TAKES PLACE

What is the posterior probability that the true mean is between 4 and 6 for each of the distributions?

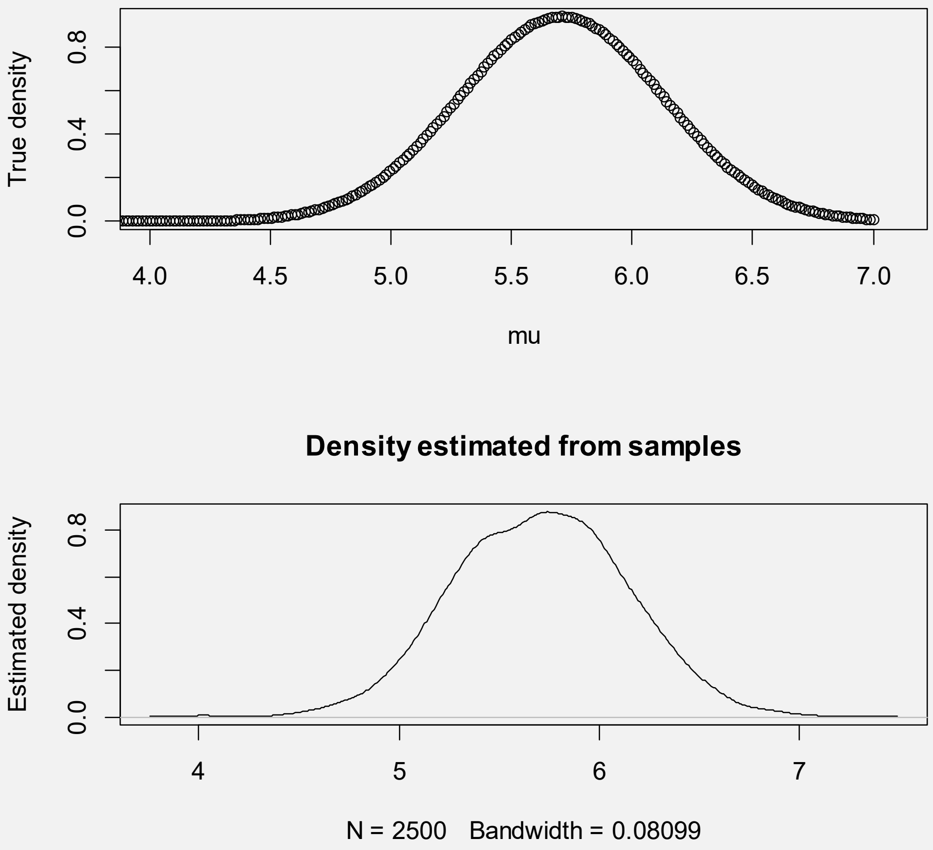
```
#####Posterior probabilities using normal CDF
for (i in 1:length(samplesize)){
U<-pnorm(6,mean=postmean[i],sd=sqrt(postvar[i]))
L<-pnorm(4,mean=postmean[i],sd=sqrt(postvar[i]))
prob46<-U-L
print(prob46)
}
[1] 0.3434397        n=10
[1] 0.75538          n=50
[1] 1                n=3000
```

Sampling from posterior distributions? Here, one needs to evaluate Monte Carlo error

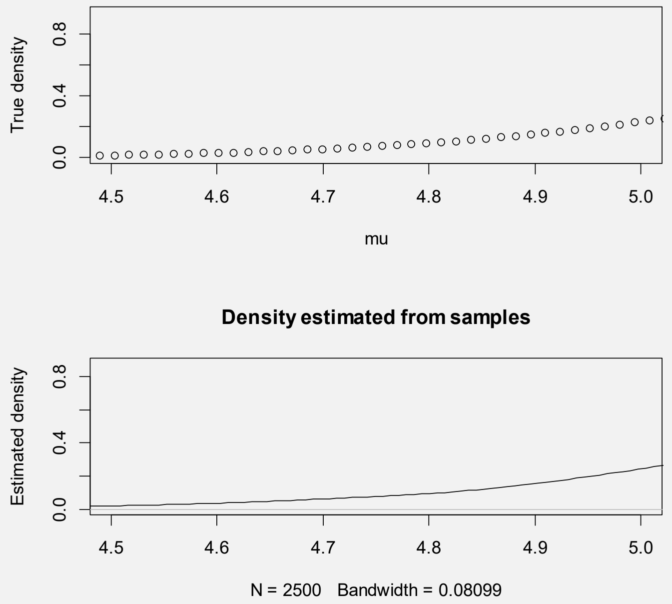
```
###Consider the case n=50 and take 2500 SAMPLES
samples50<-rnorm(2500,mean=postmean[2],sd=sqrt(postvar[2]))
###Estimate posterior mean from samples
 mean(samples50)
[1] 5.70487
####Calculate Monte Carlo error of estimating the posterior mean
mean(samples50)-postmean[2])
> mean(samples50)-postmean[2])
[1] -0.002031112
montecarlovar<-postvar[2]/length(samples50)
montecarlovar
[1] 7.184035e-05
montecarloSE<-sqrt(montecarlovar)
montecarloSE
[1] 0.008475869
###HAVE A VERY PRECISE ESTIMATE OF THE POSTERIOR MEAN!!!
###Estimate posterior variance from samples, and calculate MC error
var(samples50)-postvar[2]
[1] 0.005546713
```

Estimating the posterior density from the 2500 Monte Carlo samples?

```
####Estimating the density from samples

par(mfrow=c(2,1))
plot(muplot,dnorm(muplot,mean=postmean[2],sd=sqrt(postvar[2])),
xlab="mu",ylab="True density")
plot(density(samples50),main="Density estimated from samples",
ylab="Estimated density ")
```



ZOOM BETWEEN 4.5 AND 5



Example: Infer means of several ("exchangeable") normal distributions with **known** variance

$$Y_i|\theta_i,\sigma^2 \sim NIID(\theta_i,\sigma_e^2); i = 1,2,\ldots,p$$

$$\theta_i|\theta_0,\tau^2 \sim NIID(\theta_0,\tau^2); i = 1,2,\ldots,p$$

$$\theta_i|Y_i,\sigma_e^2,\theta_0,\tau^2 \sim N(\bar{\theta},V_\theta)$$

$$\bar{\theta} = \left[\frac{1}{\sigma_e^2}+\frac{1}{\tau^2}\right]^{-1}\left[\frac{1}{\sigma_e^2}Y_i+\frac{1}{\tau^2}\theta_0\right]$$

Weighted average: data-prior compromise

$$V_\theta = \left[\frac{1}{\sigma_e^2}+\frac{1}{\tau^2}\right]^{-1} = \sigma_e^2\left(1+\frac{\sigma_e^2}{\tau^2}\right)^{-1}$$

$$= \frac{\sigma_e^2\tau^2}{\tau^2+\sigma_e^2} = \tau^2\left(1-\frac{\tau^2}{\tau^2+\sigma_e^2}\right)$$

Relationship to quantitative genetic concepts

$$h^2 = \frac{\frac{1}{\sigma_e^2}}{\frac{1}{\sigma_e^2} + \frac{1}{\tau^2}} = \frac{\tau^2}{\tau^2 + \sigma_e^2} \quad \text{("Heritability")}$$

$$1 - h^2 = \frac{\sigma_e^2}{\tau^2 + \sigma_e^2} \quad \text{("anti-heritability")}$$

$$\begin{aligned}
\bar{\theta} &= \left[\frac{1}{\sigma_e^2} + \frac{1}{\tau^2}\right]^{-1}\left[\frac{1}{\sigma_e^2}Y_i + \frac{1}{\tau^2}\theta_0\right] \\
&= \frac{\sigma_e^2\tau^2}{\tau^2 + \sigma_e^2}\left[\frac{1}{\sigma_e^2}Y_i + \frac{1}{\tau^2}\theta_0\right] \\
&= \frac{\tau^2}{\tau^2 + \sigma_e^2}Y_i + \frac{\sigma_e^2}{\tau^2 + \sigma_e^2}\theta_0 \\
&= h^2 Y_i + (1 - h^2)\theta_0 = \theta_0 + h^2(Y_i - \theta_0)
\end{aligned}$$

"Estimated breeding value"

Example: model for allelic count with Beta prior

$$y_i \sim Bernoulli(\theta)$$ ← Probability of success

Assuming conditional independence → $$p(y_1, y_2, \ldots, y_N|\theta) \propto \theta^x(1-\theta)^{N-x}$$

Beta prior → $$p(\theta) \propto \theta^{a-1}(1-\theta)^{b-1}$$

$$\begin{aligned} p(\theta|\mathbf{y}) &\propto \theta^x(1-\theta)^{N-x}\theta^{a-1}(1-\theta)^{b-1} \\ &\propto \theta^{x+a-1}(1-\theta)^{N-x+b-1} \end{aligned} \quad \Rightarrow \quad \theta|\mathbf{y} \sim Beta(x+a, N-x+b)$$

	Prior	Posterior
Distribution	Beta	Beta
Mean	$\frac{a}{a+b}$	$\frac{x+a}{N+a+b}$
Variance	$\frac{ab}{(a+b)^2(a+b+1)}$	$\frac{(x+a)(N-x+b)}{(N+a+b)^2(N+a+b+1)}$

Example: inference of allelic frequencies using a Beta prior

```
###SET PARAMETERS OF SEVERAL BETA DISTRIBUTIONS
(PRIOR)

a<-seq(1,7,1)
b<-seq(1,7,1)

###CALCULATE MEANS OF PRIOR DISTRIBUTIONS

mean<-matrix(nrow=length(a),ncol=length(b))
for (i in 1:length(a)){
for (j in 1:length (b)){
mean[i,j]<-a[i]/(a[i]+b[j])
}
}

mean

### CALCULATE STRENGTH OF PRIOR EVIDENCE

strength<-matrix(nrow=length(a),ncol=length(b))
for (i in 1:length(a)){
for (j in 1:length (b)){
strength[i,j]<-a[i]+b[j]
}
}

strength
```

➔MEANS OF 49 DIFFERENT PRIOR DISTRIBUTIONS

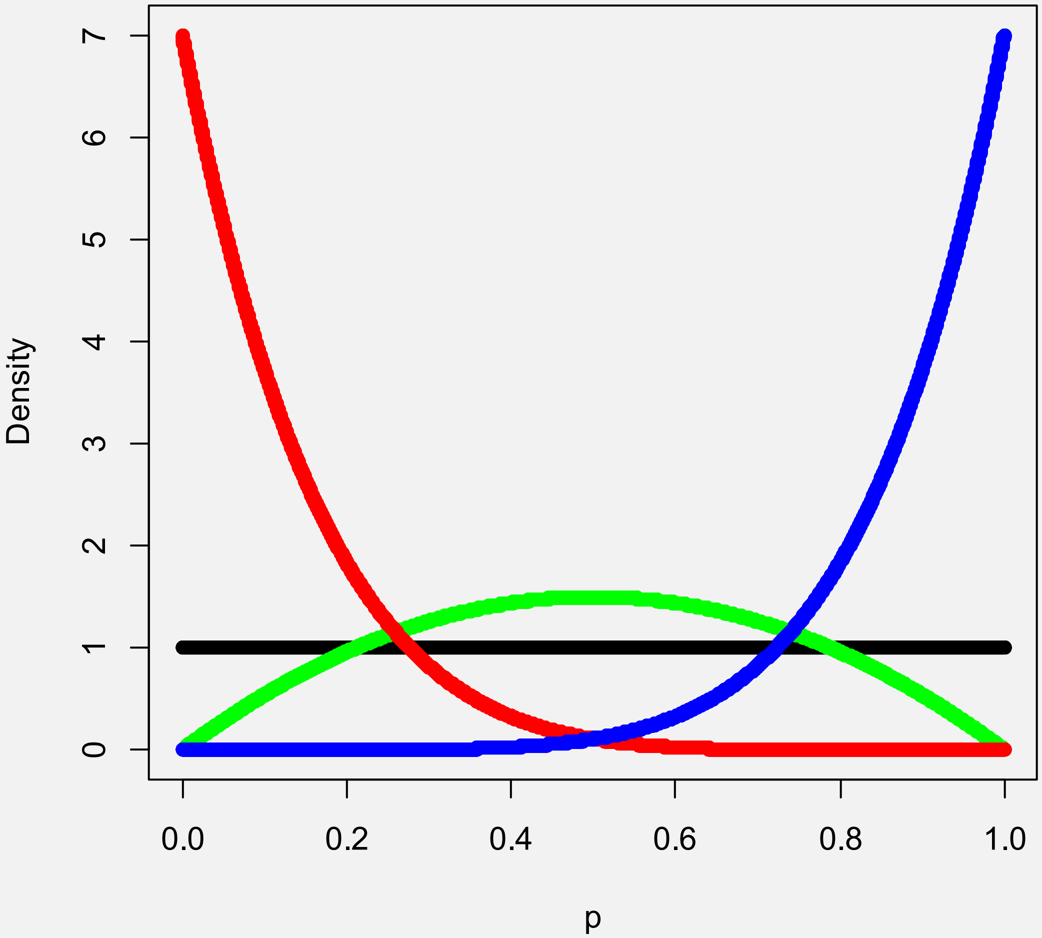
```
          [,1]      [,2]      [,3]      [,4]      [,5]      [,6]      [,7]
[1,] 0.5000000 0.3333333 0.2500000 0.2000000 0.1666667 0.1428571 0.1250000
[2,] 0.6666667 0.5000000 0.4000000 0.3333333 0.2857143 0.2500000 0.2222222
[3,] 0.7500000 0.6000000 0.5000000 0.4285714 0.3750000 0.3333333 0.3000000
[4,] 0.8000000 0.6666667 0.5714286 0.5000000 0.4444444 0.4000000 0.3636364
[5,] 0.8333333 0.7142857 0.6250000 0.5555556 0.5000000 0.4545455 0.4166667
[6,] 0.8571429 0.7500000 0.6666667 0.6000000 0.5454545 0.5000000 0.4615385
[7,] 0.8750000 0.7777778 0.7000000 0.6363636 0.5833333 0.5384615 0.5000000
```

➔“PRIOR SAMPLE SIZE” OF 49 DIFFERENT PRIOR DISTRIBUTIONS

```
     [,1] [,2] [,3] [,4] [,5] [,6] [,7]
[1,]    2    3    4    5    6    7    8
[2,]    3    4    5    6    7    8    9
[3,]    4    5    6    7    8    9   10
[4,]    5    6    7    8    9   10   11
[5,]    6    7    8    9   10   11   12
[6,]    7    8    9   10   11   12   13
[7,]    8    9   10   11   12   13   14
```

##FOUR PRIOR DENSITIES

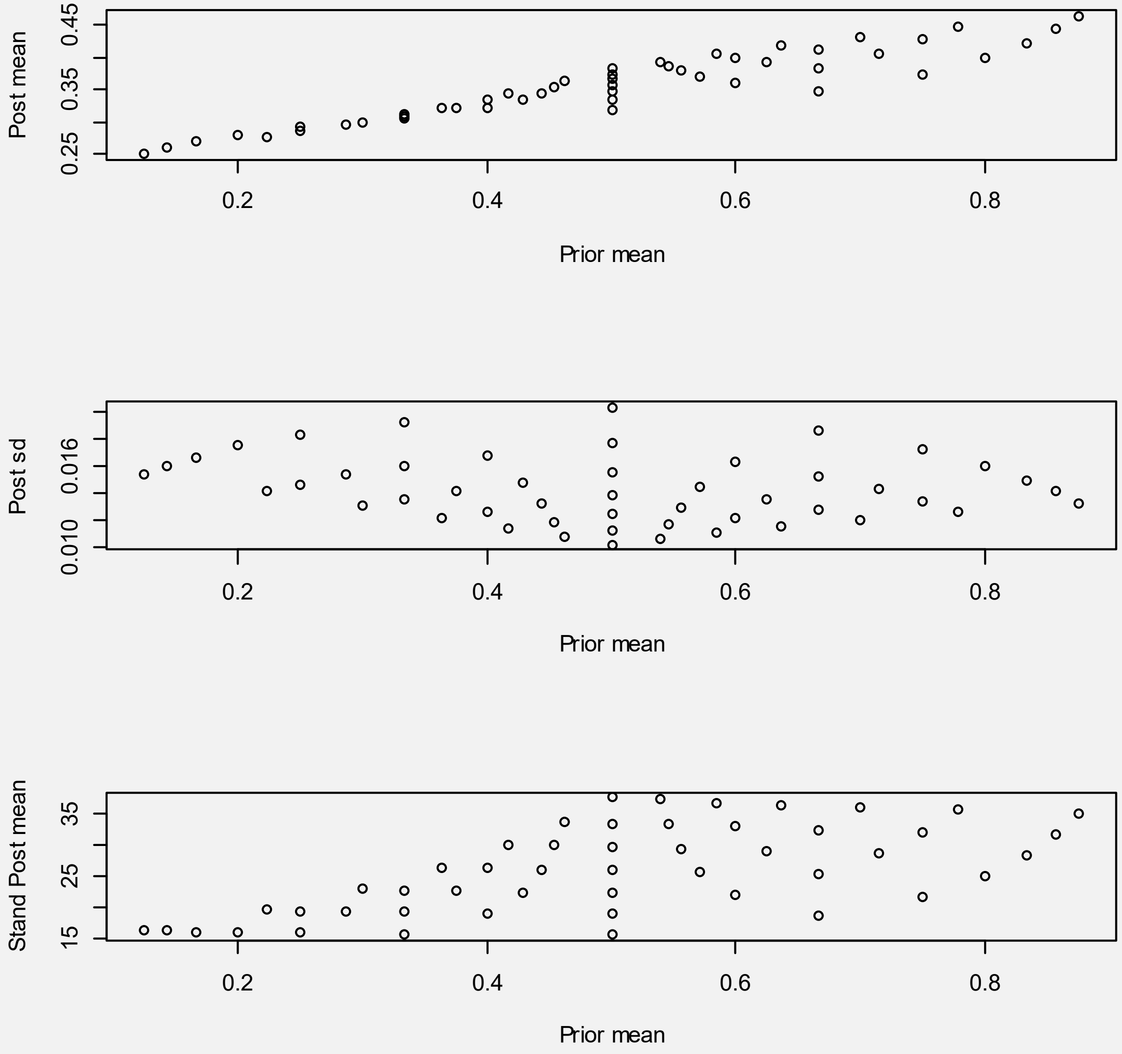
```
p<-seq(0,1,0.0001)
p11<-dbeta(p,shape1=1,shape2=1)
p22<-dbeta(p,shape1=2,shape2=2)
p17<-dbeta(p,shape1=1,shape2=7)
p71<-dbeta(p,shape1=7,shape2=1)

plot(p,p11, xlab="p",ylab="Density",ylim=c(0,7))
points(p,p22,col="green")
points(p,p17,col="red")
points(p,p71,col="blue")
```



POSTERIOR DISTRIBUTIONS; INCORPORATE INFORMATION FROM DATA

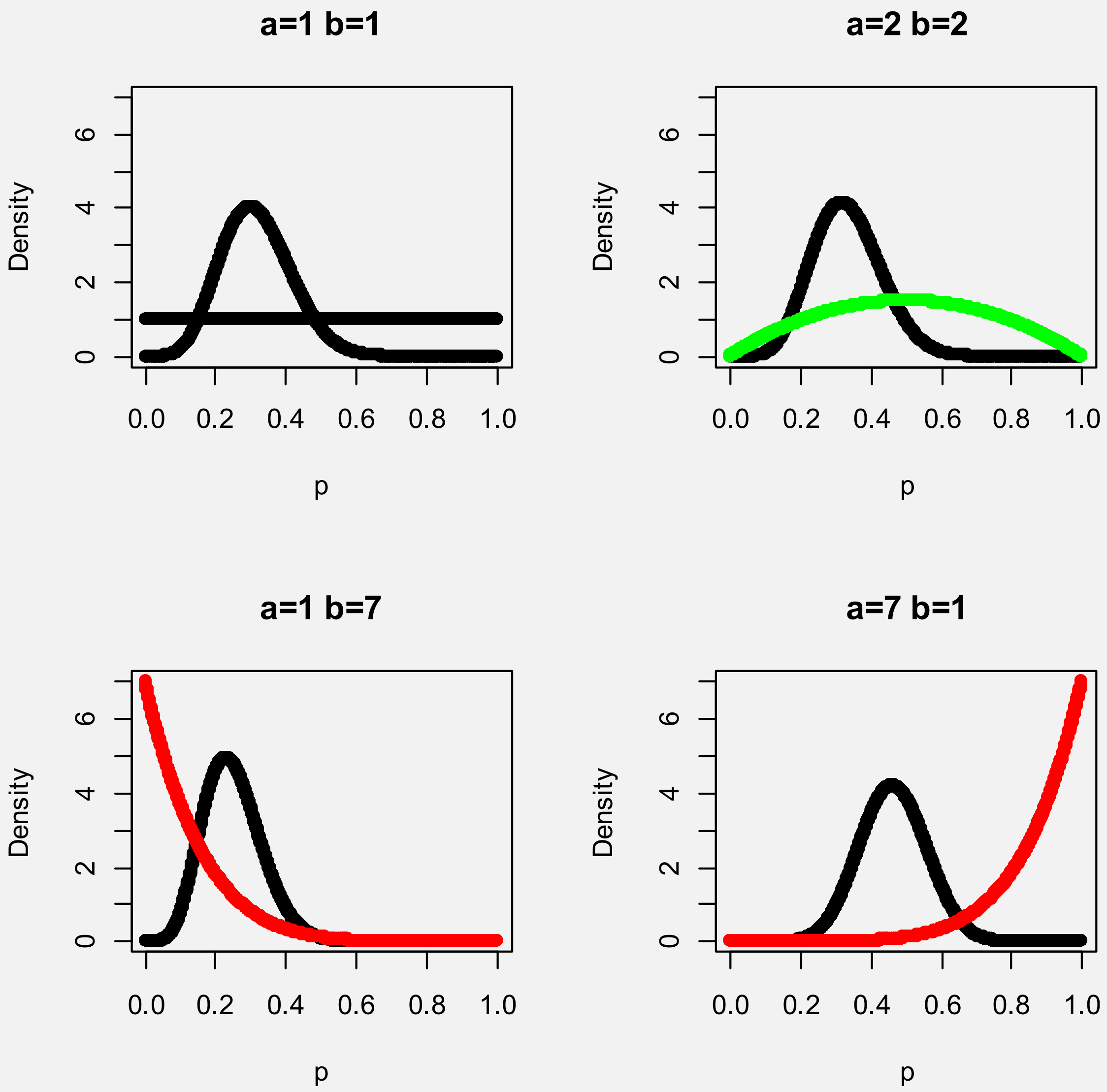
```
####POSTERIOR DENSITY: 20 ALLELES ARE SCORED
####6 ARE "A" AND 14 ARE "a"
####ASSUMING BINOMIAL SAMPLING POSTERIOR IS
####BETA(6+a,14+b)

posta<-a+6
postb<-b+14


postmean<-matrix(nrow=length(a),ncol=length(b))
for (i in 1:length(a)){
for (j in 1:length (b)){

postmean[i,j]<-posta[i]/(posta[i]+postb[j])

}
}

####POSTERIOR STANDARD DEVIATIONS

postsd<-matrix(nrow=length(a),ncol=length(b))
for (i in 1:length(a)){
for (j in 1:length (b)){

postsd[i,j]<-
posta[i]*postb[j]/((posta[i]*postb[j]))**2/(posta[i]+postb[j]+1)
postsd[i,j]<-sqrt(postsd[i,j])



###STANDARDIZED POSTERIOR MEANS

stpost<-postmean/postsd
```

```
####INFLUENCE OF PRIOR MEAN ON POSTERIOR
par(mfrow=c(3,1))
plot(mean,postmean,xlab="Prior mean",ylab="Post mean")
plot(mean,postsd,xlab="Prior mean",ylab="Post sd")
plot(mean,stpost, xlab="Prior mean",ylab="Stand Post mean")
par(mfrow=c(1,1))
```



```
###PRIOR AND POSTERIOR DENSITIES

p11<-dbeta(p,shape1=1,shape2=1)
p22<-dbeta(p,shape1=2,shape2=2)
p17<-dbeta(p,shape1=1,shape2=7)
p71<-dbeta(p,shape1=7,shape2=1)

p11POST<-dbeta(p,shape1=1+6,shape2=1+14)
p22POST<-dbeta(p,shape1=2+6,shape2=2+14)
p17POST<-dbeta(p,shape1=1+6,shape2=7+14)
p71POST<-dbeta(p,shape1=7+6,shape2=1+14)

par(mfrow=c(2,2))
plot(p,p11POST, xlab="p",ylab="Density",ylim=c(0,7),main="a=1
b=1")
points(p,p11)

plot(p,p22POST, xlab="p",ylab="Density",ylim=c(0,7),main="a=2
b=2")
points(p,p22,col="green")

plot(p,p17POST, xlab="p",ylab="Density",ylim=c(0,7),main="a=1
b=7")
points(p,p17,col="red")


plot(p,p71POST, xlab="p",ylab="Density",ylim=c(0,7),main="a=7
b=1")
points(p,p71,col="red")

par(mfrow=c(1,1))
```

THERE IS BAYESIAN LEARNING BUT HYPER-PARAMETERS STILL INFLUENTIAL



Mixtures of conjugate densities

Prior is a mixture

$$p(\theta) = \alpha p_1(\theta) + \beta p_2(\theta)$$

$$\alpha + \beta = 1; \quad 0 < \alpha < 1, 0 < \beta < 1$$

Posterior

$$p(\theta|\mathbf{y}) \propto l(\theta|\mathbf{y})[\alpha p_1(\theta) + \beta p_2(\theta)]$$

$$p(\theta|\mathbf{y}) = \frac{l(\theta|\mathbf{y})[\alpha p_1(\theta) + \beta p_2(\theta)]}{\int l(\theta|\mathbf{y})[\alpha p_1(\theta) + \beta p_2(\theta)]d\theta}$$

$$= \frac{\alpha l(\theta|\mathbf{y})p_1(\theta) + \beta l(\theta|\mathbf{y})p_2(\theta)}{\alpha \int l(\theta|\mathbf{y})p_1(\theta)d\theta + \beta \int l(\theta|\mathbf{y})p_2(\theta)d\theta}$$

$$= \frac{\alpha p_1(\theta|\mathbf{y}) \int l(\theta|\mathbf{y})p_1(\theta)d\theta + \beta p_2(\theta|\mathbf{y}) \int l(\theta|\mathbf{y})p_2(\theta)d\theta}{\alpha \int l(\theta|\mathbf{y})p_1(\theta)d\theta + \beta \int l(\theta|\mathbf{y})p_2(\theta)d\theta}$$

Mixture as well

$$= \boxed{\alpha' p_1(\theta|\mathbf{y}) + \beta' p_2(\theta|\mathbf{y})}$$

$$\alpha' = \frac{\alpha \int l(\theta|\mathbf{y})p_1(\theta)d\theta}{\alpha \int l(\theta|\mathbf{y})p_1(\theta)d\theta + \beta \int l(\theta|\mathbf{y})p_2(\theta)d\theta} \qquad \beta' = \frac{\beta \int l(\theta|\mathbf{y})p_2(\theta)d\theta}{\alpha \int l(\theta|\mathbf{y})p_1(\theta)d\theta + \beta \int l(\theta|\mathbf{y})p_2(\theta)d\theta}$$

$$\alpha' + \beta' = 1$$

Joint, Conditional and Marginal Posterior Distributions

- Put $\boldsymbol{\theta} = [\boldsymbol{\theta}_1', \boldsymbol{\theta}_2']'$ representing distinct features of models, (e.g., means and variances)
- Then, elicit a joint prior density

$$g(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = g(\boldsymbol{\theta}_1 | \boldsymbol{\theta}_2) g(\boldsymbol{\theta}_2) = g(\boldsymbol{\theta}_2 | \boldsymbol{\theta}_1) g(\boldsymbol{\theta}_1)$$

where $g(\boldsymbol{\theta}_1)$ is the marginal prior and $g(\boldsymbol{\theta}_2|\boldsymbol{\theta}_1)$ is a conditional prior

- Joint posterior density is

$$p(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 | \mathbf{y}) = \frac{L(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 | \mathbf{y}) g(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{\int\int L(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 | \mathbf{y}) g(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) d\boldsymbol{\theta}_1 \, d\boldsymbol{\theta}_2}$$

$$\propto L(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 | \mathbf{y}) g(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2),$$

- Must decide which is the object of inference
- Joint, conditional or marginal posterior probability statements?

Marginal posterior densities

- Obtained directly from probability calculus as:

$$p(\boldsymbol{\theta}_1|\mathbf{y}) = \int p(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2|\mathbf{y})\, d\boldsymbol{\theta}_2$$

$$p(\boldsymbol{\theta}_2|\mathbf{y}) = \int p(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2|\mathbf{y})\, d\boldsymbol{\theta}_1$$

- Additional marginalizing may be needed if $\boldsymbol{\theta}_1 = [\boldsymbol{\theta}'_{1A}, \boldsymbol{\theta}'_{1B}]'$

$$\begin{aligned} p(\boldsymbol{\theta}_{1A}|\mathbf{y}) &= \int\int p(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2|\mathbf{y})\, d\boldsymbol{\theta}_{1B}\, d\boldsymbol{\theta}_2 \\ &= \int p(\boldsymbol{\theta}_1|\mathbf{y})\, d\boldsymbol{\theta}_{1B}. \end{aligned}$$

Conditional posterior distributions

- By definition of conditional density:

$$p(\boldsymbol{\theta}_1 | \boldsymbol{\theta}_2, \mathbf{y}) = \frac{p(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 | \mathbf{y})}{p(\boldsymbol{\theta}_2 | \mathbf{y})}$$

- Here, one is interested in variation about $\boldsymbol{\theta}_1$ only

$$\begin{aligned} p(\boldsymbol{\theta}_1 | \boldsymbol{\theta}_2, \mathbf{y}) &\propto p(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 | \mathbf{y}) \\ &\propto L(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 | \mathbf{y}) p(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \\ &\propto L(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 | \mathbf{y}) p(\boldsymbol{\theta}_1 | \boldsymbol{\theta}_2) \\ &\propto L(\boldsymbol{\theta}_1 | \boldsymbol{\theta}_2, \mathbf{y}) p(\boldsymbol{\theta}_1 | \boldsymbol{\theta}_2). \end{aligned}$$

- Identifying conditional posterior distributions: important for implementing MCMC methods (sampling from posteriors)

EXAMPLE:

SOLVING A BAYESIAN REGRESSION PROBLEM WITH COMPOSITION SAMPLING (NO MCMC)

-Flat priors on regression coefficients

-Flat prior on log of residual variance

POSTERIOR DISTRIBUTION [A,B,C,D|DATA]
1) SAMPLE FROM A|DATA
2) SAMPLE FROM B|A, DATA
3) SAMPLE FROM C|A, B, DATA
4) SAMPLE FROM D|A, B, C, DATA

At the end, sample is from joint posterior

➔ Each coordinate is draw from marginal posterior

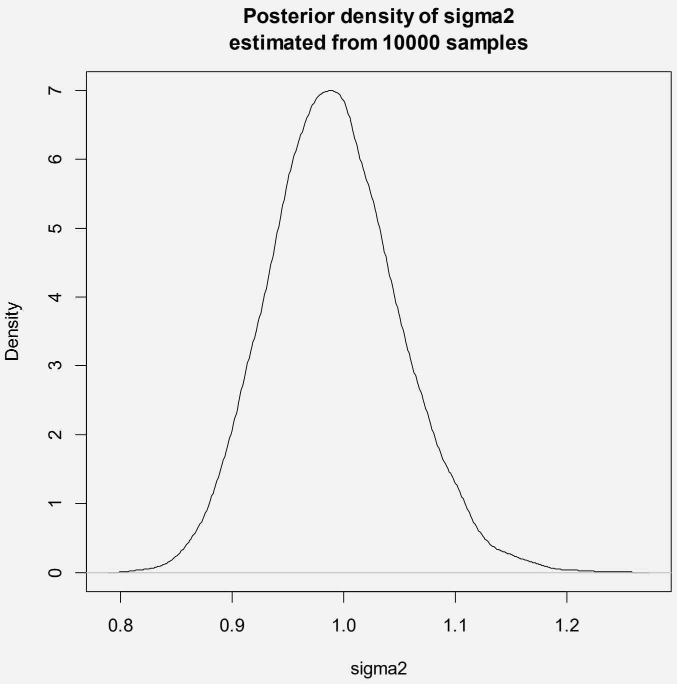
```
rm(list=ls(all=TRUE))
library(BGLR)
set.seed(1234567)

###LOAD DATA
data(wheat)
Y<-wheat.Y
X<-wheat.X
X<-X[,1:5]
y<-Y[,1]

######MODEL IS y=X[,1:5]*(b0.b1,b2,b3,b4,b5)'+e
####Compute betavar and s2 (can be done with OLS
package)

multreg<-lm(y~X)
summary(multreg)
```

Recall that posterior mean=OLS under flat priors

```
Coefficients:
              Estimate Std. Error  t value   Pr(>|t|)
(Intercept)   -0.43766   0.19305  -2.267   0.02374 *
XwPt.0538    -0.10782   0.08720  -1.237  0.21676
XwPt.8463     0.44167   0.16436   2.687   0.00741 **
XwPt.6348    -0.12168   0.08424  -1.444  0.14914
XwPt.9992     0.10498   0.33715   0.311   0.75562
XwPt.2838     0.10009   0.33713   0.297   0.76666
---
Residual standard error: 0.9939 on 593 degrees of freedom
Multiple R-squared:  0.02033, F-statistic: 2.462 on 5 and 593
DF,  p-value: 0.03198
```

1) **Sample from** $\sigma^2|\mathbf{y} \sim s^2(\nu = n - rank(\mathbf{X}))\frac{1}{\chi^2_\nu}$

```
####Sample from marginal posterior of sigma2: 10000 samples

nu<-n-multreg$rank
s2<-sum(multreg$residuals**2)/nu

samplessigma2<-s2*nu/rchisq(10000,nu)
summary(samplessigma2)

plot(density(samplessigma2),xlab="sigma2",main="Posterior density of
sigma2
estimated from 10000 samples")

##Estimate posterior probability that residual variance larger than 1
sigma2over1<-subset(samplessigma2,samplessigma2>1.0)
postprob<-length(sigma2over1)/length(samplessigma2)
Postprob
0.4247
```



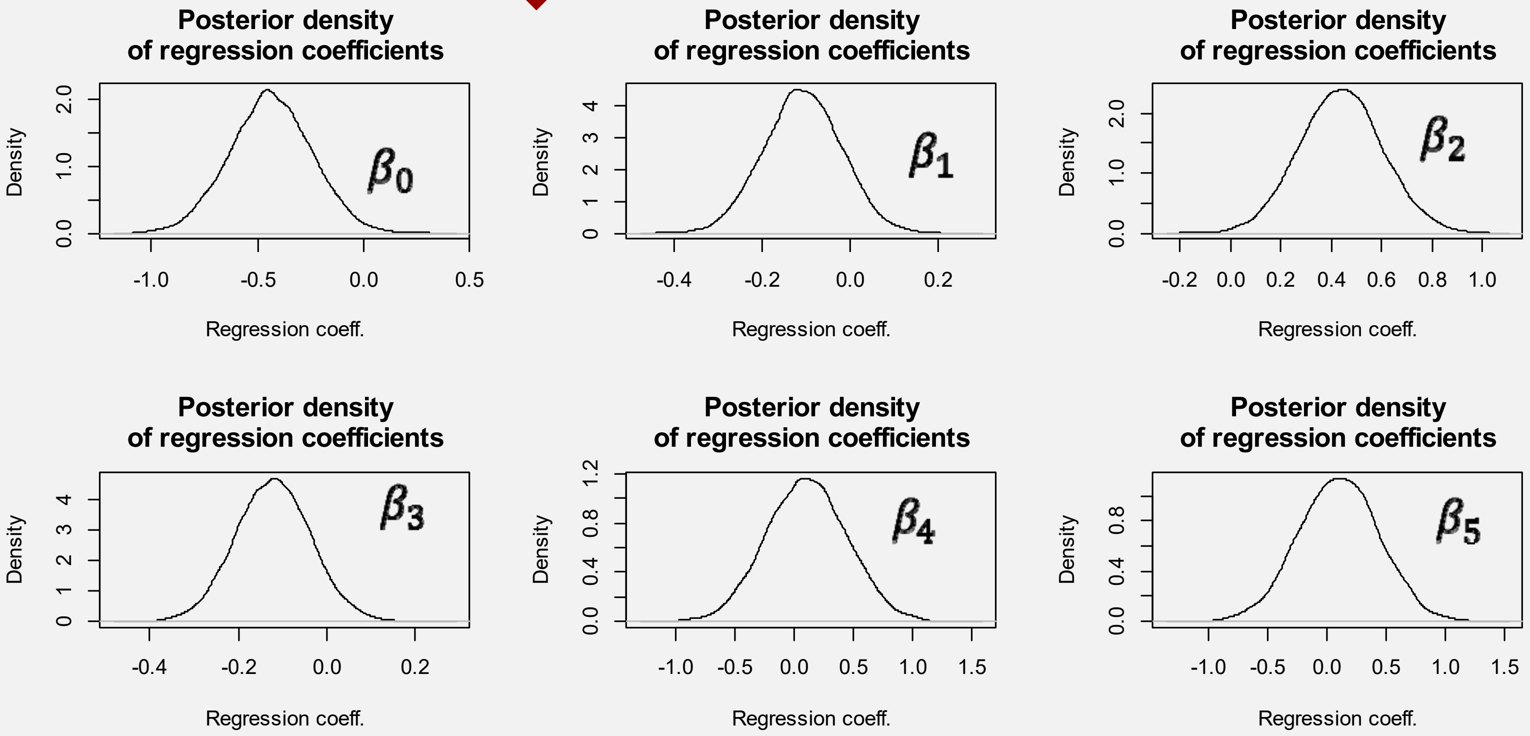
2) **Sample from** $\boldsymbol{\beta}|\sigma^2, \mathbf{y} \sim N\left(\bar{\boldsymbol{\beta}}, (\mathbf{X}'\mathbf{X})^{-1}\sigma^2\right)$

READ

```
###Samples from conditional posterior of betas given sigma2 are
###obtained for each sample of sigma2
###This is called composition sampling. At the end, this produces
###samples from the marginal posterior of betas

###

betasamples<-
matrix(nrow=ncol(X)+1,ncol=length(samplessigma2))
J<-rep(1,n)
Xnew<-cbind(J,X)
betabar<-multreg$coefficients
Sigmamat<-solve(crossprod(cbind(Xnew)))

for (j in 1:length(samplessigma2)){

betasamples[,j]<-mvrnorm(n = 1, betabar,
Sigmamat*samplessigma2[j], tol = 1e-6,
 empirical = FALSE)
}
```

```
####PLOT POSTERIOR DENSITIES OF REGRESSION COEFFICIENTS
par(mfrow=c(2,3))
for (i in 1:length(betabar)){
plot(density(betasamples[i,]),xlab="Regression coeff.",main="Posterior density
of regression coefficients")
}
par(mfrow=c(1,1))
```

```
####POSTERIOR PROBABILITY THAT BETA2 EXCEEDS
0.44167 (OLS=POSTERIOR MEAN)

beta2over<-subset(betasamples[3,],betasamples[3,]>0.44167)

postprobbeta2<-length(beta2over)/length(betasamples[3,])

Postprobbeta2
[1] 0.4972

####POSTERIOR PROBABILITY THAT BETA2 EXCEEDS 1.0

beta2over15<-subset(betasamples[3,],betasamples[3,]>1.0)

postprobbeta215<-length(beta2over15)/length(betasamples[3,])

postprobbeta215
[1] 2e-04
```



Prior Predictive Distributions

➔Sample n alleles at random from a population. At a single bi-allelic locus, an allele can be "A" or "a"

➔Observe "y" ot type "A" and n=y of type "a". Wish to estimate frequency of "A", which we call θ.

$$\Pr(y|\theta) = \binom{n}{y}\theta^{y}(1-\theta)^{n-y}$$

$$MLE(\theta) = \frac{y}{n}$$

$$Var[MLE(\theta)] = \frac{1}{n^2}Var(y) = \frac{n\theta(1-\theta)}{n^2} = \frac{\theta(1-\theta)}{n}$$

Model for allelic count with Beta prior

$y_i \sim Bernoulli(\theta)$ ← Probability of success

Assuming conditional independence → $p(y_1, y_2, \ldots, y_N|\theta) \propto \theta^x (1-\theta)^{N-x}$

Beta prior → $p(\theta) \propto \theta^{a-1}(1-\theta)^{b-1}$

$$\begin{aligned} p(\theta|\mathbf{y}) &\propto \theta^x (1-\theta)^{N-x} \theta^{a-1}(1-\theta)^{b-1} \\ &\propto \theta^{x+a-1}(1-\theta)^{N-x+b-1} \end{aligned} \quad \Rightarrow \quad \theta|\mathbf{y} \sim Beta(x+a, N-x+b)$$

	Prior	Posterior
Distribution	Beta	Beta
Mean	$\frac{a}{a+b}$	$\frac{x+a}{N+a+b}$
Variance	$\frac{ab}{(a+b)^2(a+b+1)}$	$\frac{(x+a)(N-x+b)}{(N+a+b)^2(N+a+b+1)}$

Binary Data: Predictive distribution

Future data

Future number of Bernoulli trials

$$p(\mathbf{y}_f|\mathbf{y}) = \int \binom{N_f}{x_f}\theta^{x_f}(1-\theta)^{N_f-x_f}\frac{\theta^{x+a-1}(1-\theta)^{N-x+b-1}}{B(x+a,N-x+b)}d\theta$$

Future number of “successes”

Posterior

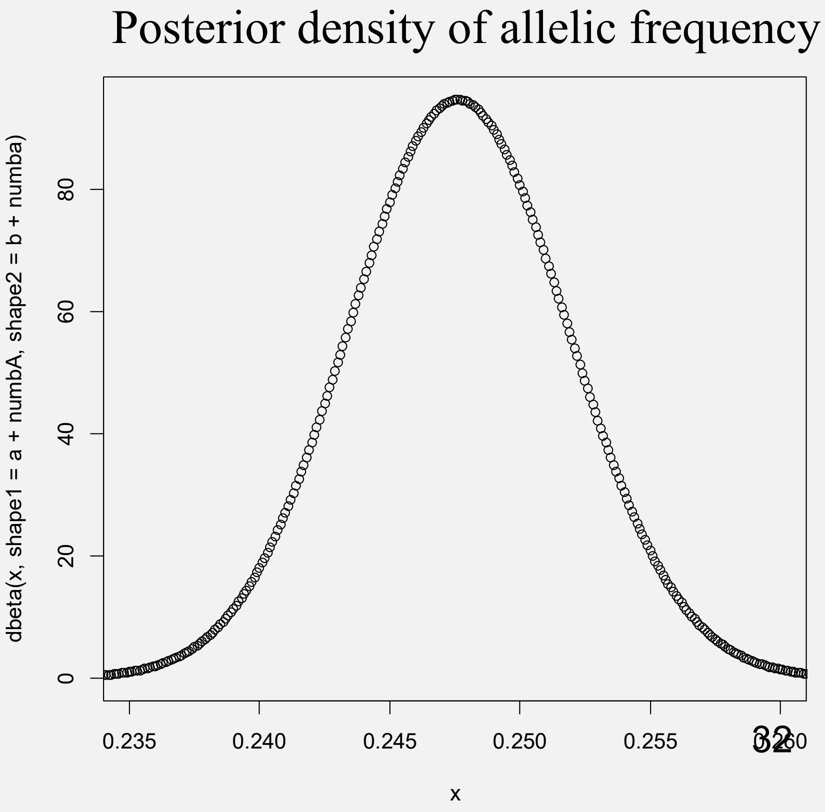
$$= \frac{\binom{N_f}{x_f}}{B(x+a,N-x+b)}\int \theta^{x_f+x+a-1}(1-\theta)^{N_f+N-x_f-x+b-1}d\theta$$

$$= \boxed{\binom{N_f}{x_f}\frac{B(x_f+x+a,N_f+N-x_f-x+b)}{B(x+a,N-x+b)}}$$

Beta-binomial distribution (discrete)

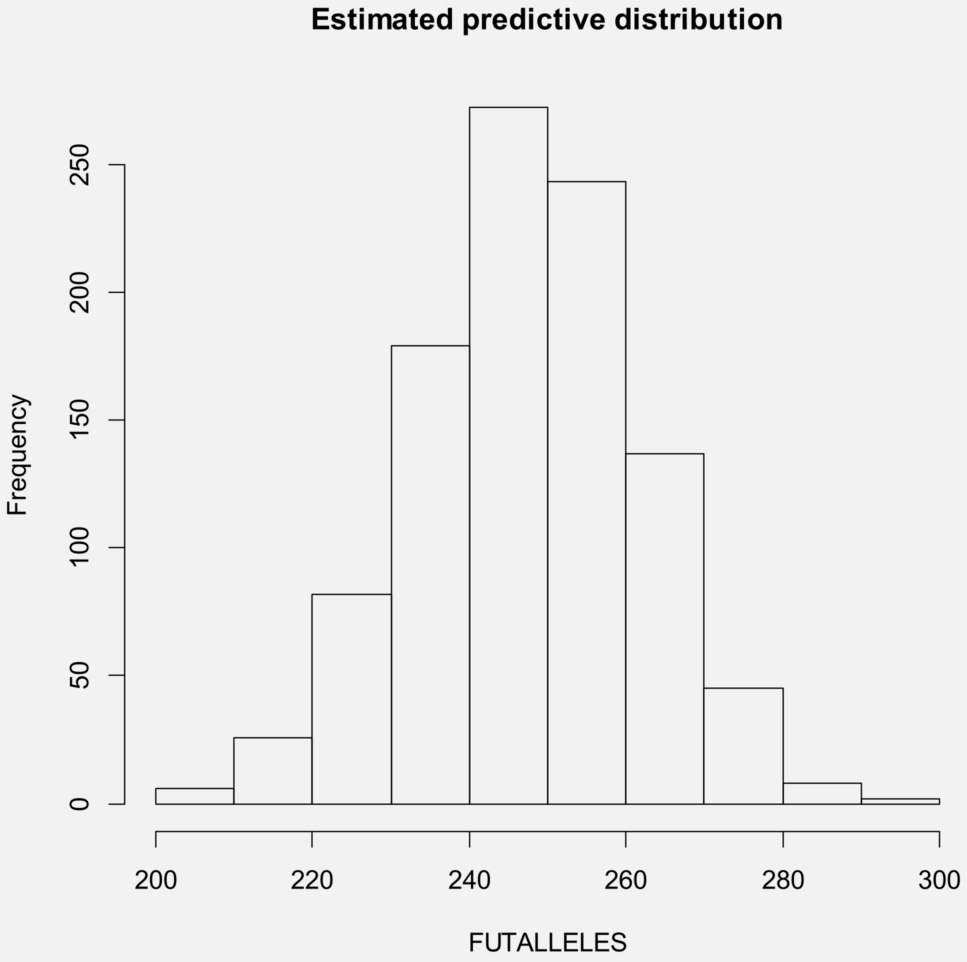
Example of predictive distribution:
beta-binomial model with composition sampling

```
#####SUPPOSE 10500 ALLELES ARE DRAWN FROM A POPULATION
WHERE Pr(A)=0.25
#####ASSUME EACH DRAW IS A BERNOULLY TRIAL
#####WISH TO ESTIMATE THE ALLELIC FREQUENCY

#####SIMULATE DATA

alleles<-numeric(10500)

alleles<-rbinom(10500,size=1,prob=0.25)

numbA<-sum(alleles)
numba<-10500-numbA

#####Prior Beta(a=1,b=1)
a<-1
b<-1

x<-seq(0,1,0.001)

####Posterior is Beta(a+numbA,b+numba)

postmean<-(a+numbA)/(a+b+10500)
postvar<-(a+numbA)*(b+numba)/(((a+b+10500)**2)*(a+b+10500+1))
postmean
Postvar
> postmean
[1] 0.2476671
> postvar
[1] 1.774047e-05
x<-seq(.15,0.40,0.0001)
plot(x,dbeta(x,shape1=a+numbA,shape2=b+numba),xlim=c(0.235,0.26))
```


Posterior density of allelic frequency

```
###I WISH TO PREDICT HOW MANY ALLELES "A" WILL BE
FOUND IN A FUTURE EXPERIMENT OF SIZE 1000
###THIS IS THE NUMBER OF FUTURE ALLELES COUNTED

###DERIVE PREDICTIVE DISTRIBUTION BY COMPOSITION
SAMPLING

FUTALLELES<-numeric(1000)
freq<-numeric(1000)

for (i in 1:1000){

###Sample from posterior of allelic frequency
freq[i]<-rbeta(1,a+numbA,b+numba)

###Sample future allele
FUTALLELES[i]<-rbinom(1,size=1000,prob=freq[i])
}

hist(FUTALLELES,main="Estimated predictive distribution")

summary(FUTALLELES)
mean(FUTALLELES)
var(FUTALLELES)
```

```
> summary(FUTALLELES)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    201     239     248     248     258     294
> mean(FUTALLELES)
[1] 248.023
> var(FUTALLELES)
[1] 205.0135
```



BAYESIAN MODEL COMPARISON: BAYES FACTOR

BAYES FACTOR

A PRIORI: MODEL A → $\Pr(M = A)$ MODEL B → $\Pr(M = B)$

A POSTERIORI

$$\Pr(M = A|\mathbf{y}, H_A) = \Pr(M = A)p(\mathbf{y}|M = A)/[\Pr(M = A)p(\mathbf{y}|M = A) + \Pr(M = B)p(\mathbf{y}|M = B)]$$

$$\Pr(M = B|\mathbf{y}, H_B) = \Pr(M = B)p(\mathbf{y}|M = M)/[\Pr(M = A)p(\mathbf{y}|M = A) + \Pr(M = B)p(\mathbf{y}|M = B)]$$

POSTERIOR ODDS RATIO

$$\frac{\Pr(M = A|\mathbf{y}, H_A)}{\Pr(M = B|\mathbf{y}, H_B)} = \frac{\Pr(M = A)}{\Pr(M = B)} \times \frac{p(\mathbf{y}|M = A)}{p(\mathbf{y}|M = B)} = \text{PRIOR ODDS RATIO} \times \text{BAYES FACTOR}$$

$$\text{BAYES FACTOR} = \frac{\text{POSTERIOR ODDS RATIO}}{\text{PRIOR ODDS RATIO}}$$

PRIOR PRED. DIST → $p(\mathbf{y}|M = A)$ AND $p(\mathbf{y}|M = B)$

Further

$$p(\mathbf{y}|M = A) = \int p(\mathbf{y}|\boldsymbol{\theta}_A)p(\boldsymbol{\theta}_A|H_A, M = A)$$

$$p(\mathbf{y}|M = B) = \int p(\mathbf{y}|\boldsymbol{\theta}_B)p(\boldsymbol{\theta}_B|H_B, M = B)$$

Bayes factor for 2 beta-binomial models

$$Model\ 1 = \Pr(y|n,a,b) = \int \binom{n}{y} \theta^{y}(1-\theta)^{n-y} \frac{\theta^{a-1}(1-\theta)^{b-1}}{B(a,b)} d\theta$$

$$= \binom{n}{y} \frac{B(y+a, n-y+b)}{B(a,b)}$$

$$Model\ 2 = \Pr(y|n,c,d) = \int \binom{n}{y} \theta^{y}(1-\theta)^{n-y} \frac{\theta^{c-1}(1-\theta)^{d-1}}{B(a,b)} d\theta$$

$B(c,d)$

$$= \binom{n}{y} \frac{B(y+c, n-y+d)}{B(c,d)}$$

$$BF_{12} = \frac{\dfrac{B(y+a,n-y+b)}{B(a,b)}}{\dfrac{B(y+c,n-y+d)}{B(c,d)}}$$

$Suppose\ n = 50, y = 1, a = 1, b = 2, c = 9, d = 1$

$$BF_{12} = \frac{\dfrac{B(y+a, n-y+b)}{B(a,b)}}{\dfrac{B(y+c, n-y+d)}{B(c,d)}}$$

```
> beta(2,51)
[1] 0.0003770739
> beta(1,2)
[1] 0.5
> beta(10,50)
[1] 1.591638e-12
> beta(9,1)
[1] 0.1111111
```

BAYES FACTOR

```
BF=beta(2,51)*beta(9,1)/beta(1,2)/beta(10,50)
> BF
[1] 52646519
> log(BF)
[1] 17.77911
```

THE SUPPORT IS OVERWHELMING IN FAVOR OF MODEL 1. WHY?

➔MODEL 1 CLAIMS A PRIORI “A” IS MORE RARE THAN “a”
➔THE DATA FAVORS THE HYPOTHESIS THAT “A” IS RARE
➔MODEL 2 ASSIGNS LITTLE PROBABILITY TO THE OBSERVED DATA

OVERVIEW OF BAYESIAN INFERENCE: PART 2

Rev. Thomas Bayes

1702 London, England
1761 Tunbridge Wells, Kent, England

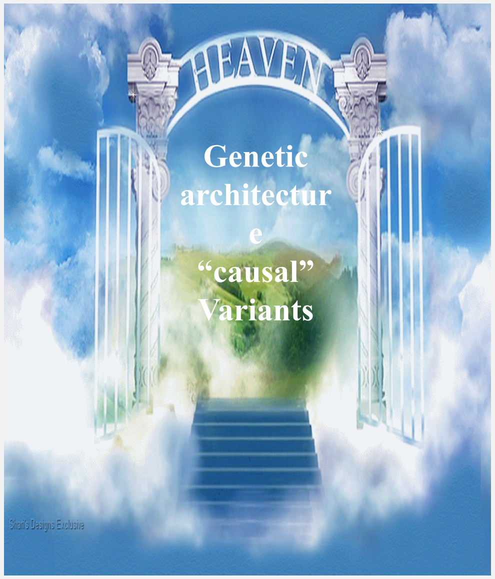
1763. “An essay towards solving a problem in the doctrine of chances”. *Philosophical Transactions of the Royal Society of London* **53**, 370-418.

Pierre-Simon Laplace

1749 Beaumont-en-Auge, France
1827 Paris, France

1774. "Mémoire sur la probabilité des causes par les événements". *Savants étranges* **6**, 621-656. *Oeuvres* **8**, 27-65

TO BAYES OR NOT TO BAYES?

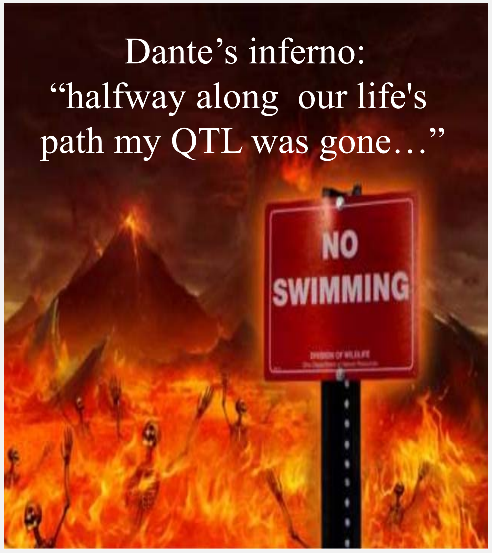


Genetic architecture
“causal”
Variants

PRIOR

DATA

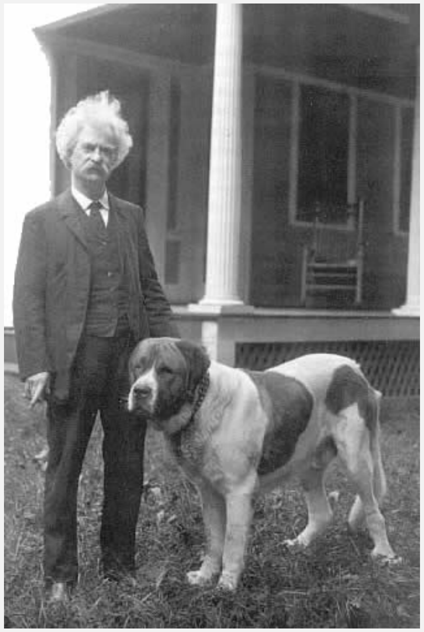
POSTERIOR



Dante’s inferno:
“halfway along our life's
path my QTL was gone…”

The Bayesian’s way out of hell in “small” samples

“It is rational to use prior knowledge, especially when evidence is scant”



***"It ain't what you don't know that gets you into trouble.
It's what you know for sure that just ain't so."***

(Mark Twain)

The Bayesian's way out of hell in "large" samples

- "Asymptotic domination" of prior by data for likelihood-identified unknowns, under regularity conditions (big issue for Bayesian mixtures)
- When samples are infinitely large, posterior distribution has variance 0: all mass on the "true" value
- Parameters with a lot of information from the data: prior matters little
- What if unknowns are NOT identified in the likelihood?

CAN ONE **ESTIMATE** MARKER EFFECTS FROM DATA WHEN n<<p WITHOUT BRINGING EXTERNAL INFORMATION?

$$y = X\beta + e;\ e \sim N(0, I\sigma_e^2)$$

$$L(\beta|y) \sim \exp\left[-\frac{(y - X\beta)'(y - X\beta)}{2\sigma_e^2}\right]$$

$$\log[L(\beta|y)] = -\frac{(y - X\beta)'(y - X\beta)}{2\sigma_e^2}$$

If p>>n

$$\text{Information}(\beta)_{p\times p} = -\frac{\partial^2}{\partial\beta\partial\beta'}\log[L(\beta|y)] = -\frac{\partial}{\partial\beta'}\left[-X'\frac{(y - X\beta)}{\sigma_e^2}\right] = \frac{X'X}{\sigma_e^2}$$

$$\text{Expected Information}(\beta)_{p\times p} = \frac{X'X}{\sigma_e^2} \Rightarrow \text{Does not have full-rank}$$

$\Rightarrow$ MARKER EFFECTS ARE NOT INDIVIDUALLY ESTIMABLE

VERY IMPORTANT TO KEEP IN MIND

Example: *n=4, p=5*

$$X = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 & 16 \\ 1 & 3 & 9 & 27 & 81 \end{bmatrix}$$

$$X^T X = \begin{bmatrix} 4 & 6 & 14 & 36 & 98 \\ 6 & 14 & 36 & 98 & 276 \\ 14 & 36 & 98 & 276 & 794 \\ 36 & 98 & 276 & 794 & 2316 \\ 98 & 276 & 794 & 2316 & 6818 \end{bmatrix}$$

$$|X^T X| = 0$$

LEAST-SQUARES DOES NOT ALLOW
ESTIMATION OF INDIVIDUAL REGRESSIONS
IN THE n<<p SITUATION

Example again: *n=4, p=5 BUT ADD **5** TO DIAGONALS OF X’X*

$$X = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 & 16 \\ 1 & 3 & 9 & 27 & 81 \end{bmatrix}$$

$$X^T X = \begin{bmatrix} 4 & 6 & 14 & 36 & 98 \\ 6 & 14 & 36 & 98 & 276 \\ 14 & 36 & 98 & 276 & 794 \\ 36 & 98 & 276 & 794 & 2316 \\ 98 & 276 & 794 & 2316 & 6818 \end{bmatrix} \qquad X^T X + I \times 5 = \begin{bmatrix} 9 & 6 & 14 & 36 & 98 \\ 6 & 19 & 36 & 98 & 276 \\ 14 & 36 & 103 & 276 & 794 \\ 36 & 98 & 276 & 799 & 2316 \\ 98 & 276 & 794 & 2316 & 6823 \end{bmatrix}$$

Determinant= 0

Determinant= 25782105

RIDGE REGRESSION (here we assume n>p)

Classical view: "estimator" of a fixed vector of regression coefficients

Can assess by cross-validation

$$
\begin{aligned}
\widehat{\beta}_{Ridge} &= [X'X + I\lambda]^{-1} X'y \\
&= \left[I + (X'X)^{-1}\lambda \right]^{-1} (X'X)^{-1} X'y \\
&= \left[I + (X'X)^{-1}\lambda \right]^{-1} \widehat{\beta}_{OLS} \\
E\left(\widehat{\beta}_{Ridge} | X\right) &= \left[I + (X'X)^{-1}\lambda \right]^{-1} E\left(\widehat{\beta}_{OLS}\right) \\
&= \left[I + (X'X)^{-1}\lambda \right]^{-1} \beta
\end{aligned}
$$

Shrinkage towards 0

Biased estimator but more precise

Classical view: "predictor" of a random vector of regression coefficients $\beta \sim (0, \sigma^2_\beta)$,

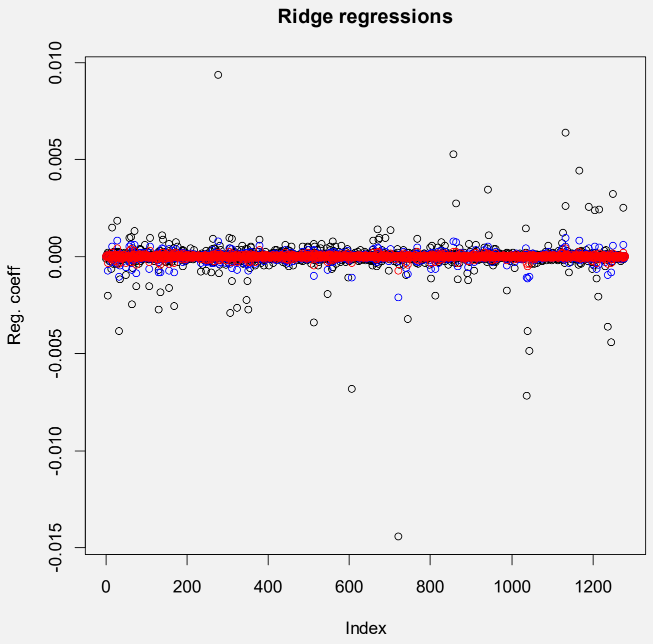
$\hat{\beta}$=BLUP , $E(\hat{\beta}) = E(\beta)$, taking $\lambda = \frac{\sigma^2_e}{\sigma^2_\beta}$ as known

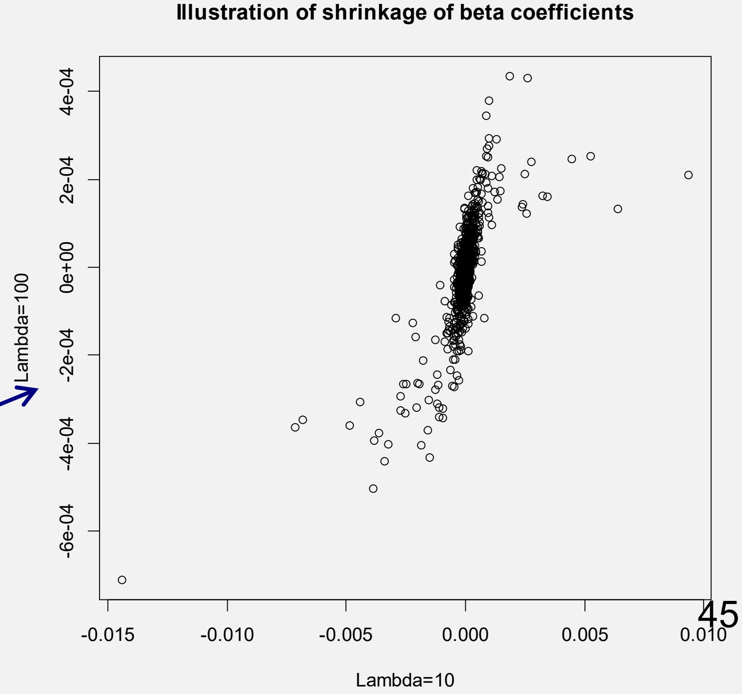
Bayesian view: Mean of posterior distribution of regressions under prior $\beta \sim (0, \sigma^2_\beta)$,

normal likelihood and known variance parameters

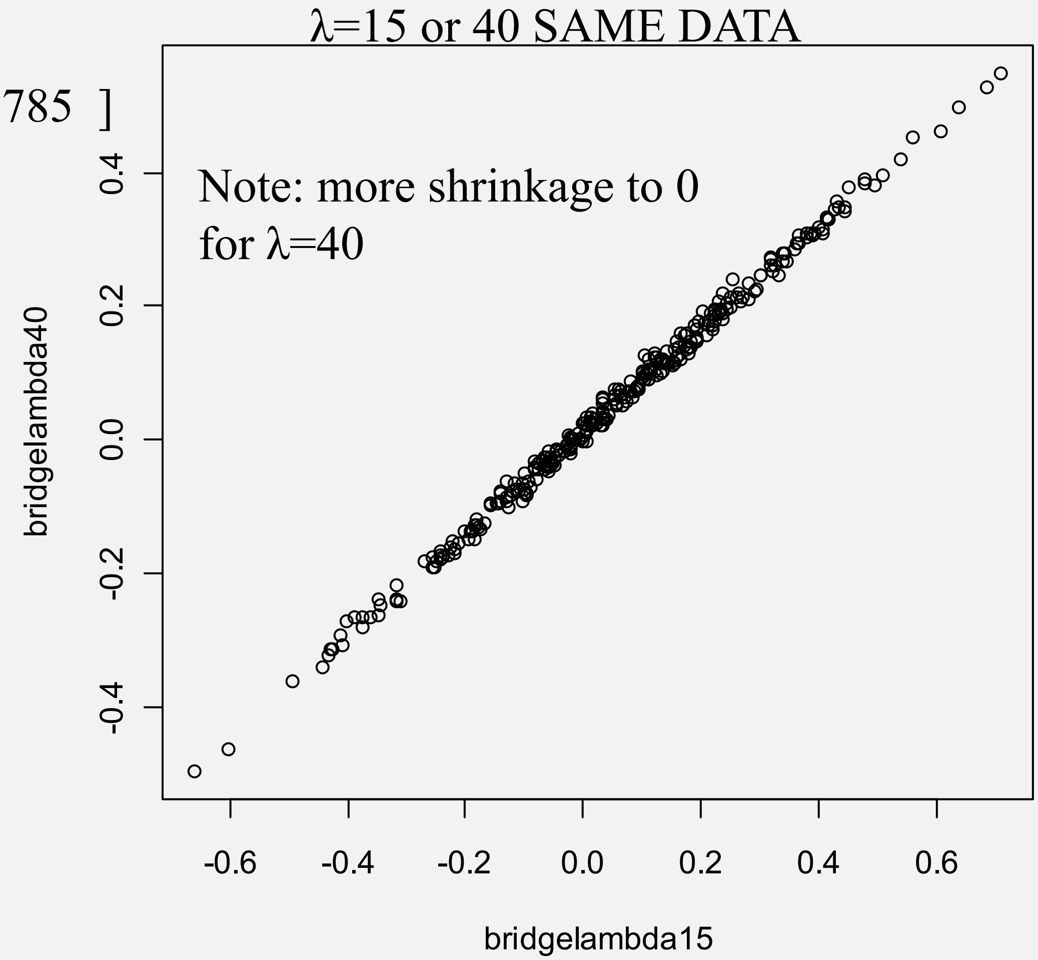
```
####RIDGE REGRESSION: SAME AS BLUP
####OR BAYESIAN CONDITIONAL POSTERIOR MEAN
####BUT DIFFERENT INTERPRETATION
rm(list=ls(all=TRUE))
library(MASS)
library(BGLR)
set.seed(1234567)

###LOAD DATA
data(wheat)
Y<-wheat.Y
X<-wheat.X

y<-Y[,1]
n<-nrow(X)
p<-ncol(X)

###RIDGE REGRESSION AS AN ESTIMATOR

lambda<-seq(10,100,10)
XPX<-crossprod(X)
XPy<-crossprod(X,y)

betaridge<-matrix(nrow=p,ncol=length(lambda))
for(j in 1:length(lambda)){
betaridge[,j]<-chol2inv(chol(XPX+lambda[j]*diag(p)))%*%XPy
}
plot(betaridge[,1],ylab="Reg. coeff",main="Ridge regressions")
points(betaridge[,5],col="blue")


plot(betaridge[,1],betaridge[,10],xlab="Lambda=10",ylab="Lamb
da=100", main="Illustration of shrinkage of beta coefficients")
```

As lambda grows, solutions closer to 0 (shrinkage) and are less variable





Example: n=30, p=300, binary markers with q=0.50
Effects sampled from N(0,0.5).
Residuals sampled from N(0,10).

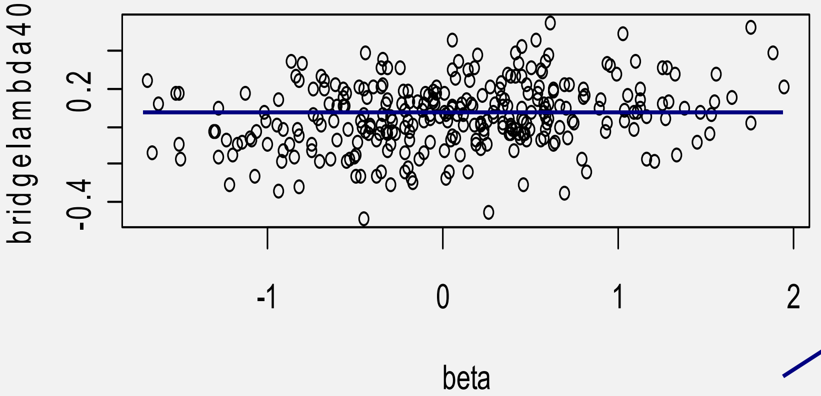


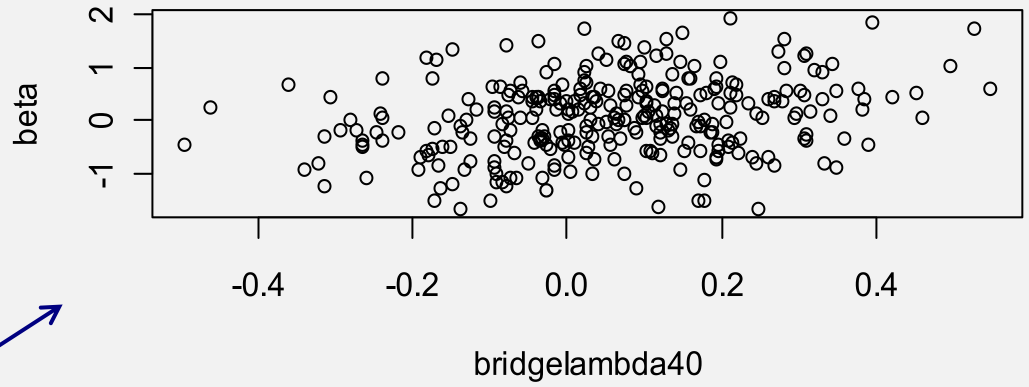
Note: more shrinkage to 0 for λ=40

[-0.49418, 0.54785]

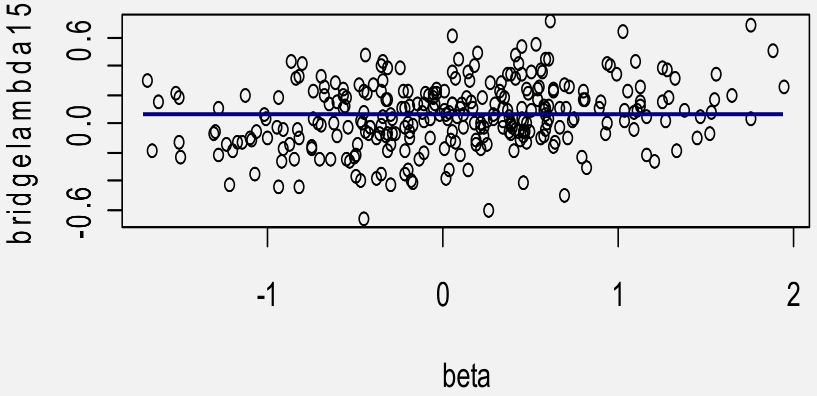
[-0.65852, 0.70828]

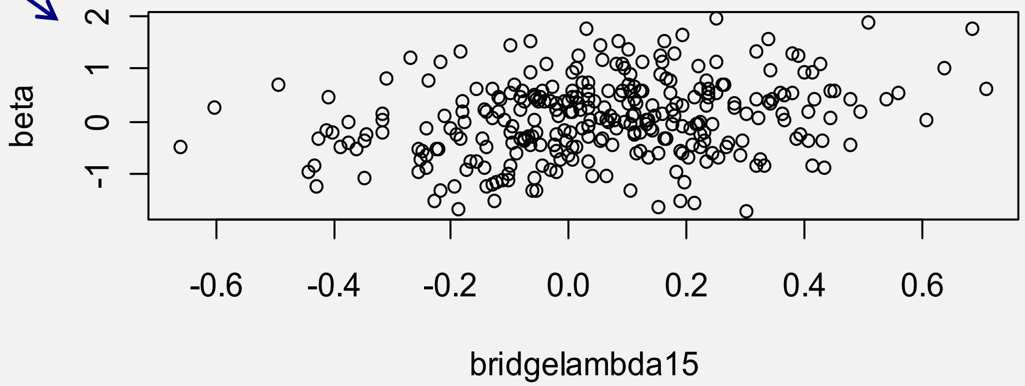
Interpretable as means of conditional distributions of true betas, given bridge estimates





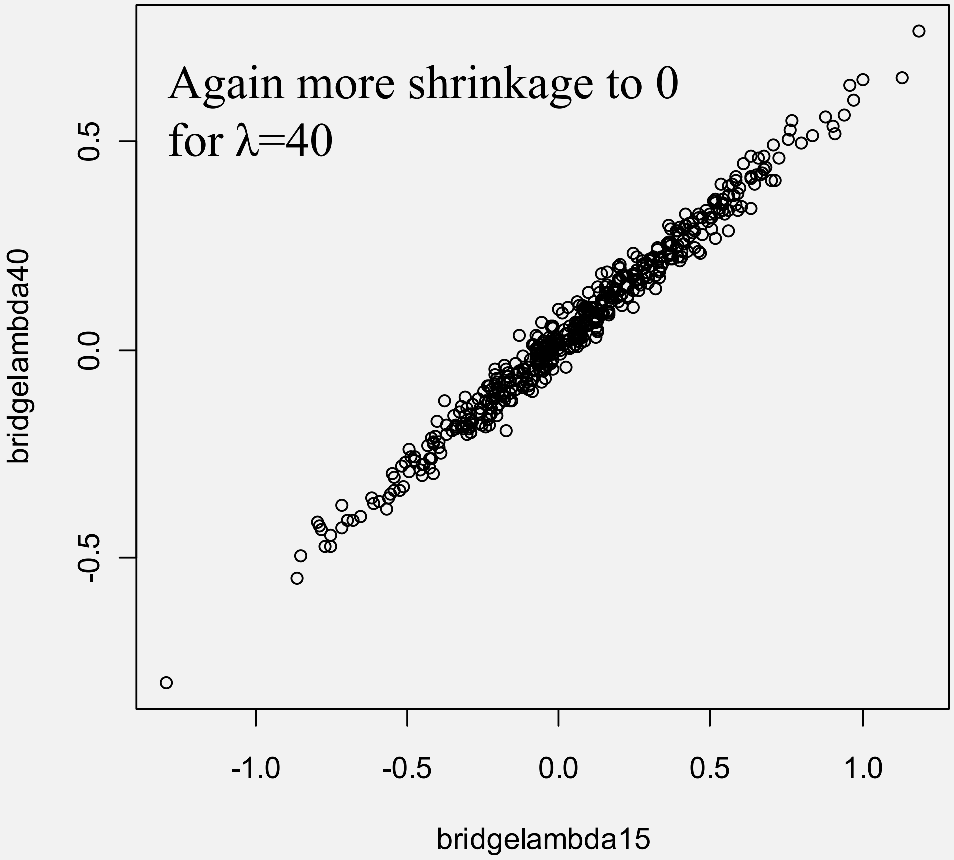
GIVEN THE MARKER ESTIMATE, WE HAVE UNCERTAINTY ABOUT TRUE EFFECT





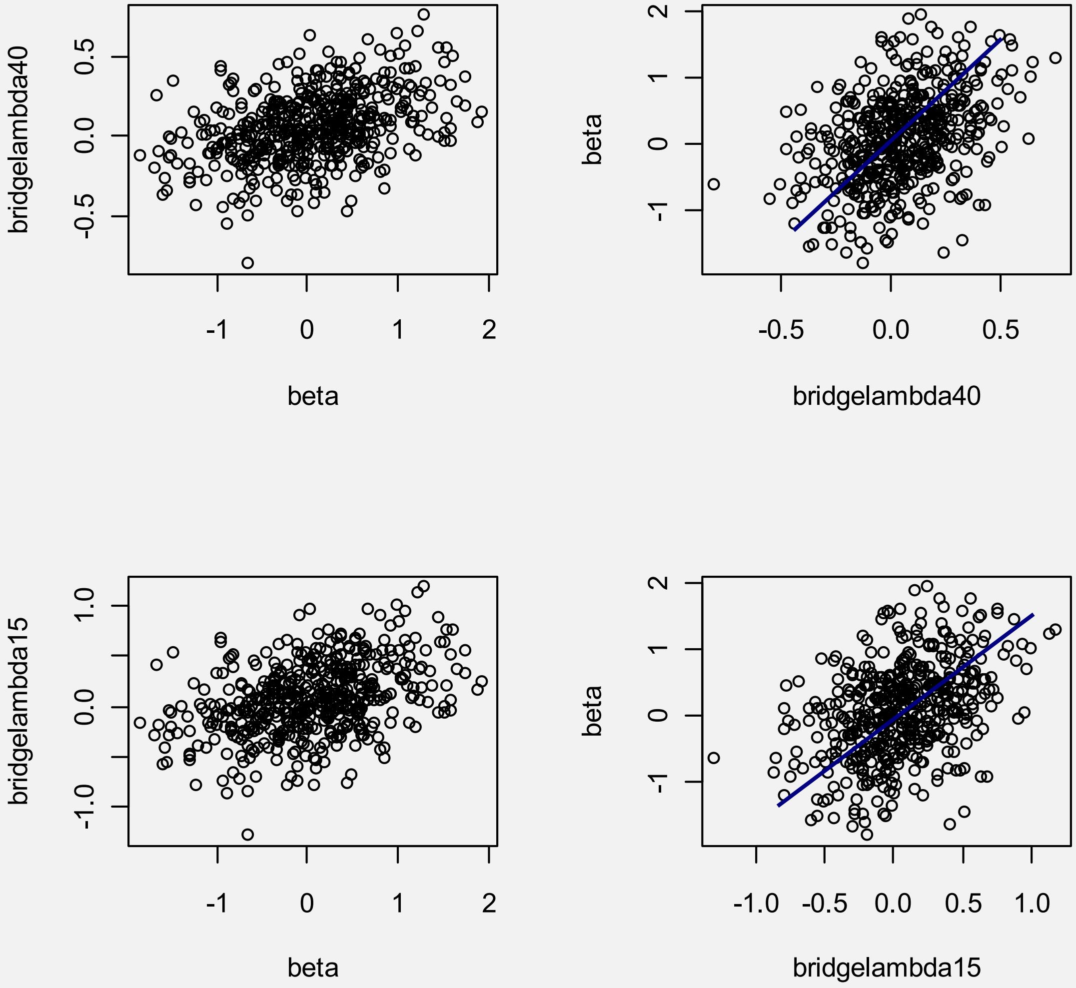
Example: n=200, p=500, binary markers with q=0.05 (“rare”)
Effects sampled from N(0,0.5).
Residuals sampled from N(0,10).

λ=15 or 40 SAME DATA



Example: n=200, p=500, binary markers with q=0.05 (“rare”)
Effects sampled from N(0,0.5).
Residuals sampled from N(0,10).

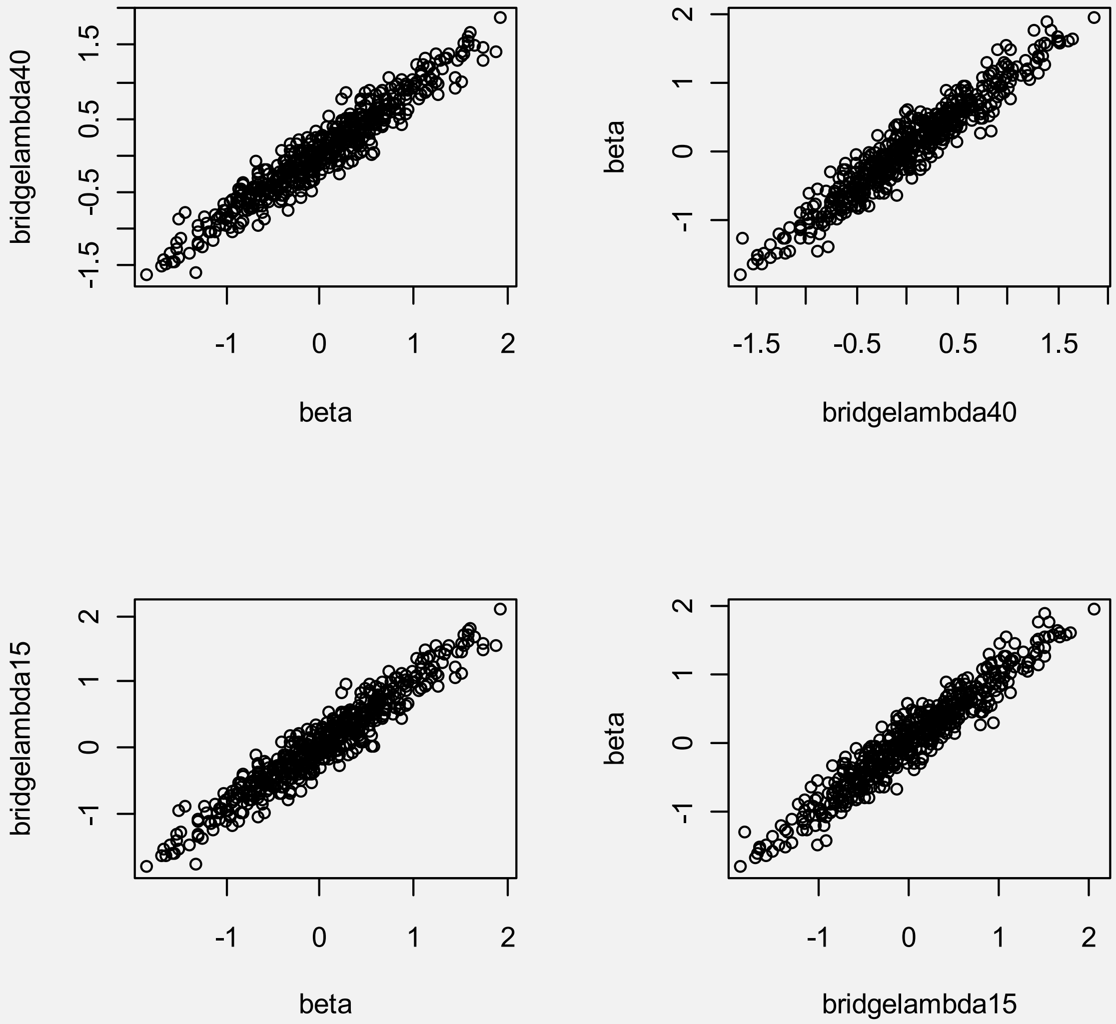
λ=15 or 40 SAME DATA



Can see “alignment”
but lots of uncertainty
about “true” marker effects

Example: n=5000, p=500, binary markers with q=0.05 (“rare”)
Effects sampled from N(0,0.5).
Residuals sampled from N(0,10).

λ=15 or 40 SAME DATA



ONLY IF
n>>p
ONE CAN
LEARN
EFECTS WELL

Hypothetical situation: trait affected by **20** additive QTL. Entire population has size **n=100** and can genotype and phenotype all. We buy **3** chips with an increasing number of markers (**100, 250, 700**) such that markers of Chip 1 are in 2, and so on.

Markers and QTL in weak LD.

```
rm(list=ls(all=TRUE))
library(MASS)
library(BGLR)
set.seed(1234567)

####SIMULATE 20 QTL AND PHENOTYPES N=100
####20 LOCI, FREQUENCIES Pr(A)=Pr(a)=0.50 AT EACH LOCUS
####Effects simulated drawing from 3 beta distributions
####and taking a linear combination

###SIMULATE QTL AND EFFECTS
Q20<-matrix(sample(0:2,prob=c(0.25,0.50,0.25),2000,replace=TRUE),nrow=100,ncol=20)
q20<-numeric(20)
q20<-4*rbeta(20,1,1)+1*rbeta(20,3,1)-10*rbeta(20,0.2,0.8)
g<-Q20%*%q20

##Evaluate additive genetic variance in population

###The genetic variance is approximately 75.
###Suppose h2=.35, then vare<-varg*(1-0.35)/0.35
###Simulate phenotypes

y<-g+rnorm(100,0,sqrt(varg*(1-0.35)/0.35))
h2sim<-varg/var(y)
h2sim
lambdatrue<-(1-h2sim)/h2sim
```

```
74.66115
0.3819273
1.618299
```

```
###CHIPS HAVE 100, 250, 700 MARKERS
###and allelic frequencies are 0.30 at each marker
###Markers are in LE with QTL and with themselves
###Any LD is due to finite sample size

X3<-matrix(sample(0:2,prob=c(0.09,0.42,0.49),70000,replace=TRUE),nrow=100,ncol=700)
X2<-X3[,1:250]
X1<-X3[,1:100]

###Define grid of lambda values
lambda<-c(2,5,10)
J<-rep(1,nrow(X1))

df<-matrix(nrow=3,ncol=length(lambda))
X1<-cbind(J,X1)
X2<-cbind(J,X2)
X3<-cbind(J,X3)
```

```
###Estimate marker effects for each chip at each lambda

XPX1<-crossprod(X1)
XPy1<-crossprod(X1,y)
XPX2<-crossprod(X2)
XPy2<-crossprod(X2,y)
XPX3<-crossprod(X3)
XPy3<-crossprod(X3,y)

betaridge1<-matrix(nrow=ncol(X1),ncol=length(lambda))
for(j in 1:length(lambda)){
C<-XPX1+diag(ncol(X1))*lambda[j]
C[1,1]<-C[1,1]-lambda[j]
Cinv<-chol2inv(chol(C))
betaridge1[,j]<-Cinv%*%XPy1
df[1,j]<-sum(diag(X1%*%Cinv%*%t(X1)))
betaridge1[,j]<-Cinv%*%XPy1
}

betaridge2<-matrix(nrow=ncol(X2),ncol=length(lambda))
for(j in 1:length(lambda)){
C<-XPX2+lambda[j]*diag(ncol(X2))
C[1,1]<-XPX2[1,1]
Cinv<-chol2inv(chol(C))
betaridge2[,j]<-Cinv%*%XPy2
df[2,j]<-sum(diag(X2%*%Cinv%*%t(X2)))
betaridge2[,j]<-Cinv%*%XPy2
}

betaridge3<-matrix(nrow=ncol(X3),ncol=length(lambda))
for(j in 1:length(lambda)){
C<-XPX3+lambda[j]*diag(ncol(X3))
C[1,1]<-XPX3[1,1]
Cinv<-chol2inv(chol(C))
betaridge3[,j]<-Cinv%*%XPy3
df[3,j]<-sum(diag(X3%*%Cinv%*%t(X3)))
betaridge3[,j]<-Cinv%*%XPy3
}
```

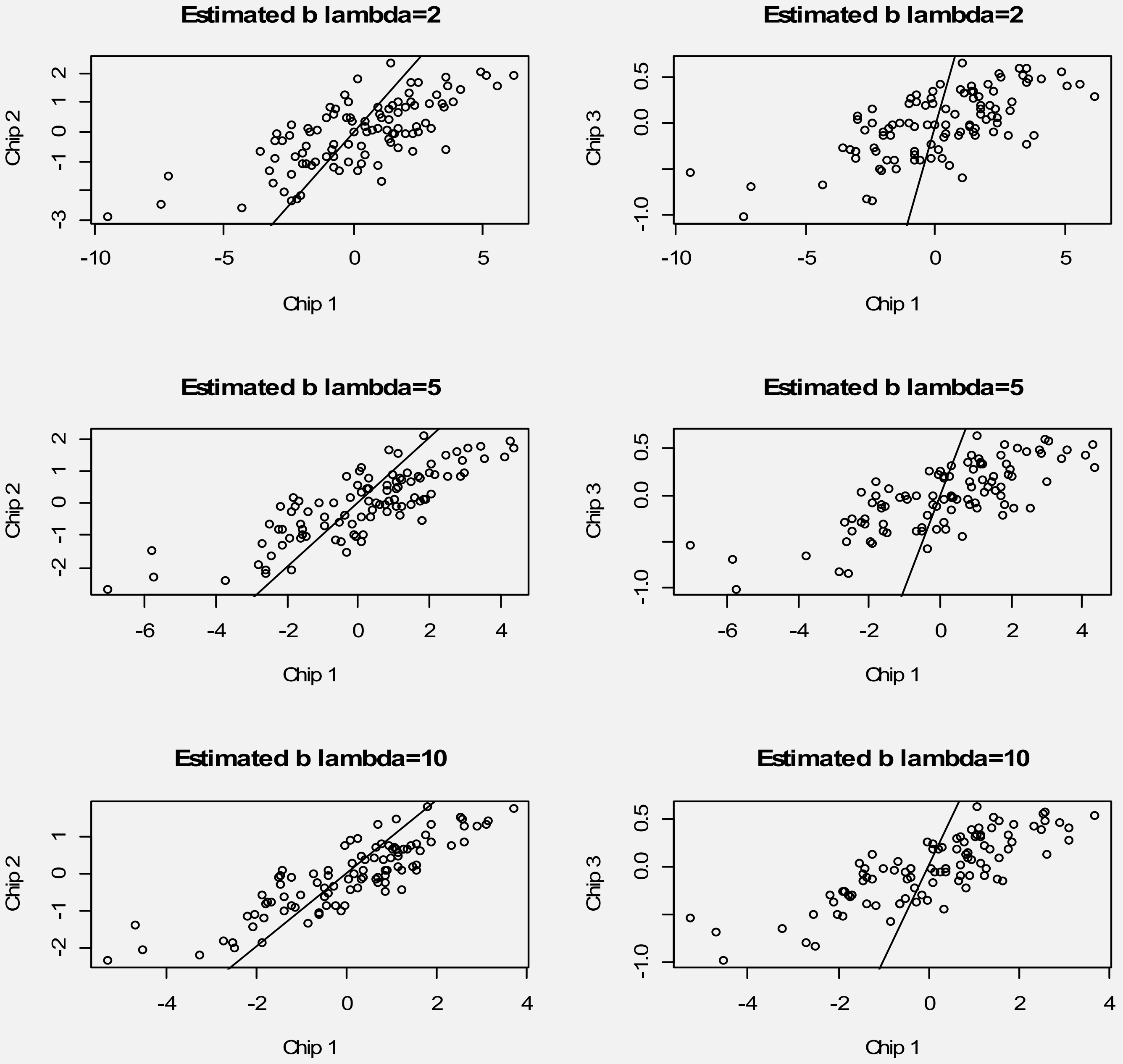
df=tr$(X(X'X + I * lambda)X'$

Intercept IS NOT regularized!!!!!

```
>Model  df
        [,1]          [,2]          [,3]
[1,] 81.06669 71.63596 62.39252   100 mark
[2,] 96.96482 92.92510 87.21130    250 mark
[3,] 99.22289 98.08331 96.25016    300 mark
```

```
Residual df= 100-Model df
          [,1]              [  ,2]           [,3]
[1,] 18.9333114 28.364042  37.607483
[2,]  3.0351821     7.074898  1 2.788700
[3,]  0.7771115      1.916690   3.749841
```

Chips 2 and 3 may overfit….


Estimated b lambda=2

Estimated b lambda=2

Estimated b lambda=5

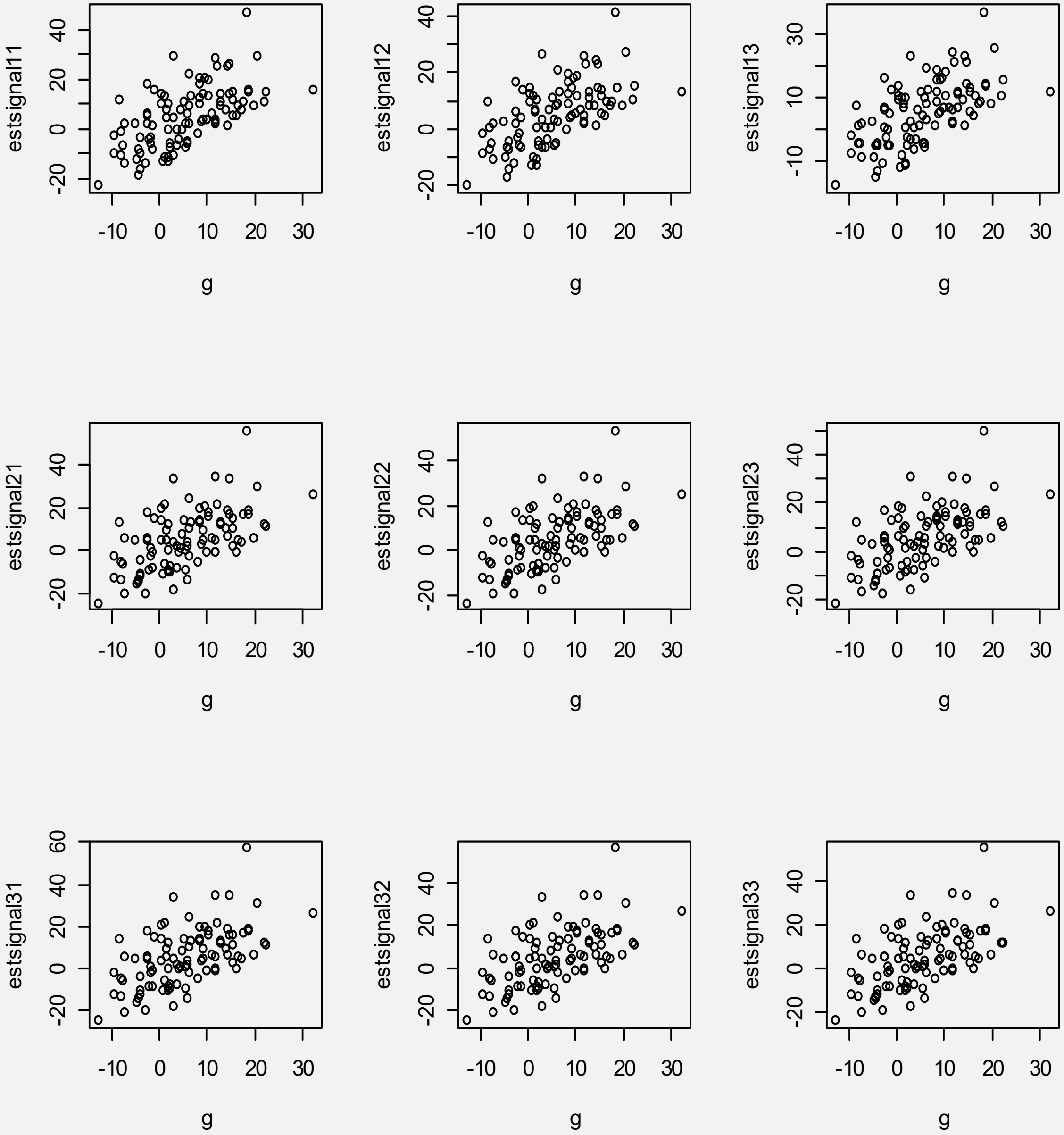
Estimated b lambda=5

Estimated b lambda=10

Estimated b lambda=10

```
#####HOW WELL DO MARKERS ESTIMATE TRUE SIGNAL?

estsignal11<-X1%*%betaridge1[,1]
estsignal12<-X1%*%betaridge1[,2]
estsignal13<-X1%*%betaridge1[,3]
estsignal21<-X2%*%betaridge2[,1]
estsignal22<-X2%*%betaridge2[,2]
estsignal23<-X2%*%betaridge2[,3]
estsignal31<-X3%*%betaridge3[,1]
estsignal32<-X3%*%betaridge3[,2]
estsignal33<-X3%*%betaridge3[,3]

par(mfrow=c(3,3))
plot(g,estsignal11)
plot(g,estsignal12)
plot(g,estsignal13)
plot(g,estsignal21)
plot(g,estsignal22)
plot(g,estsignal23)
plot(g,estsignal31)
plot(g,estsignal32)
plot(g,estsignal33)
par(mfrow=c(1,1))
```

RELATIONSHIP BETWEEN TRUE AND ESTIMATED SIGNALS



```
MATPRED<-cbind(g,estsignal11,estsignal12,estsignal13,
estsignal21,estsignal22,estsignal23,
estsignal31,estsignal32,estsignal33)

cors<-cor(MATPRED)

REL<-matrix(nrow=10,ncol=10)
for (i in 1:10){
for (j in 1:10){
REL[i,j]<-cors[i,j]**2
}
}

cors
REL
```

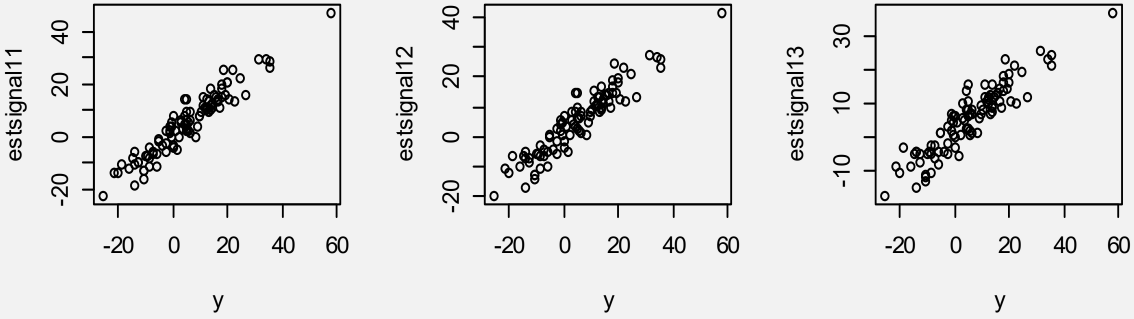
```
> cors
          [,1]      [,2]      [,3]      [,4]      [,5]      [,6]      [,7]
 [1,] 1.0000000 0.5987244 0.6019199 0.6044632 0.5770991 0.5792279 0.5819872
 [2,] 0.5987244 1.0000000 0.9967156 0.9881191 0.9540780 0.9562033 0.9577708
 [3,] 0.6019199 0.9967156 1.0000000 0.9971320 0.9383697 0.9413730 0.9441543
 [4,] 0.6044632 0.9881191 0.9971320 1.0000000 0.9225969 0.9263719 0.9302488
 [5,] 0.5770991 0.9540780 0.9383697 0.9225969 1.0000000 0.9995312 0.9975199
 [6,] 0.5792279 0.9562033 0.9413730 0.9263719 0.9995312 1.0000000 0.9992034
 [7,] 0.5819872 0.9577708 0.9441543 0.9302488 0.9975199 0.9992034 1.0000000
 [8,] 0.5755443 0.9519479 0.9356203 0.9193188 0.9997206 0.9985393 0.9956184
 [9,] 0.5758020 0.9521486 0.9358921 0.9196597 0.9997382 0.9986143 0.9957749
[10,] 0.5761826 0.9524287 0.9362869 0.9201663 0.9997220 0.9986901 0.9959814
          [,8]      [,9]     [,10]
 [1,] 0.5755443 0.5758020 0.5761826
 [2,] 0.9519479 0.9521486 0.9524287
 [3,] 0.9356203 0.9358921 0.9362869
 [4,] 0.9193188 0.9196597 0.9201663
 [5,] 0.9997206 0.9997382 0.9997220
 [6,] 0.9985393 0.9986143 0.9986901
 [7,] 0.9956184 0.9957749 0.9959814
 [8,] 1.0000000 0.9999894 0.9999281
 [9,] 0.9999894 1.0000000 0.9999727
[10,] 0.9999281 0.9999727 1.0000000
```

YELLOW: correlations with signal
GREEN: correlations between fitted signals

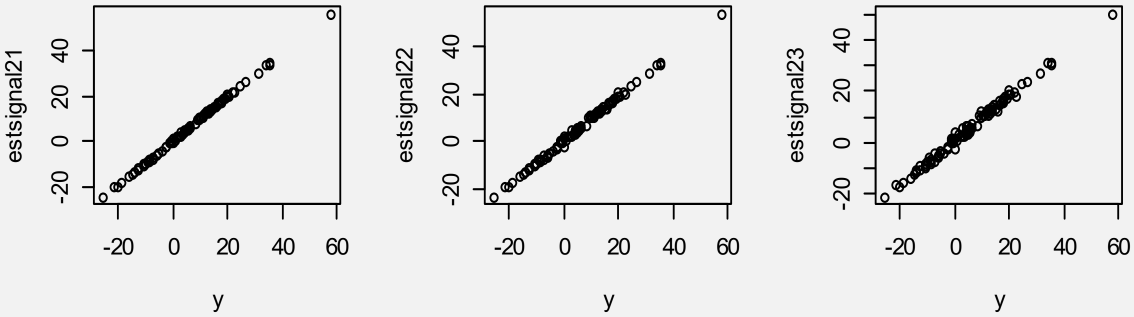
ASSOCIATION WITH PHENOTYPES

```
MATPREDY<-cbind(y,estsignal11,estsignal12,estsignal13,
estsignal21,estsignal22,estsignal23,
estsignal31,estsignal32,estsignal33)

par(mfrow=c(3,3))
plot(y,estsignal11)
plot(y,estsignal12)
plot(y,estsignal13)
plot(y,estsignal21)
plot(y,estsignal22)
plot(y,estsignal23)
plot(y,estsignal31)
plot(y,estsignal32)
plot(y,estsignal33)
par(mfrow=c(1,1))

corsy<-cor(MATPREDY)
RELY<-matrix(nrow=10,ncol=10)
for (i in 1:10){
for (j in 1:10){
RELY[i,j]<-corsy[i,j]**2
}
}
```

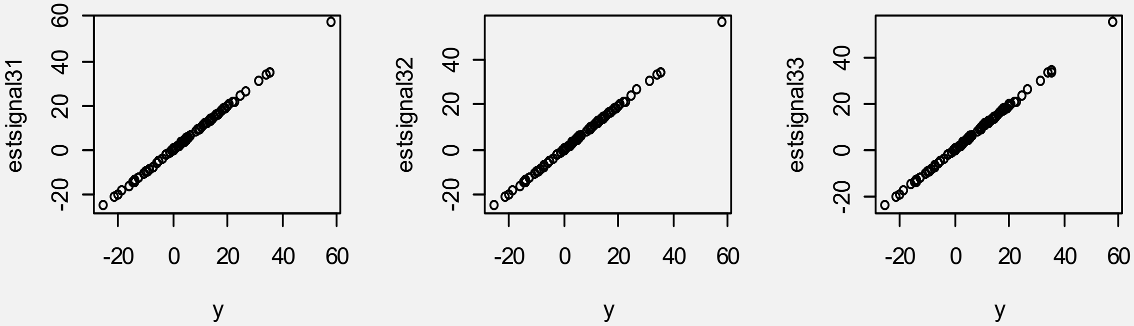
CHIP 1



CHIP 2



CHIP 3



```
> RELY
          [,1]      [,2]      [,3]      [,4]      [,5]      [,6]      [,7]
 [1,] 1.0000000 0.9059218 0.8750167 0.8446987 0.9993929 0.9969538 0.9910188
 [2,] 0.9059218 1.0000000 0.9934420 0.9763793 0.9102648 0.9143248 0.9173249
 [3,] 0.8750167 0.9934420 1.0000000 0.9942722 0.8805377 0.8861832 0.8914274
 [4,] 0.8446987 0.9763793 0.9942722 1.0000000 0.8511850 0.8581649 0.8653629
 [5,] 0.9993929 0.9102648 0.8805377 0.8511850 1.0000000 0.9990627 0.9950460
 [6,] 0.9969538 0.9143248 0.8861832 0.8581649 0.9990627 1.0000000 0.9984074
 [7,] 0.9910188 0.9173249 0.8914274 0.8653629 0.9950460 0.9984074 1.0000000
 [8,] 0.9999901 0.9062048 0.8753853 0.8451471 0.9994413 0.9970806 0.9912560
 [9,] 0.9999399 0.9065870 0.8758939 0.8457739 0.9994764 0.9972304 0.9915676
[10,] 0.9997708 0.9071203 0.8766332 0.8467060 0.9994441 0.9973819 0.9919790
          [,8]      [,9]     [,10]
 [1,] 0.9999901 0.9999399 0.9997708
 [2,] 0.9062048 0.9065870 0.9071203
 [3,] 0.8753853 0.8758939 0.8766332
 [4,] 0.8451471 0.8457739 0.8467060
 [5,] 0.9994413 0.9994764 0.9994441
 [6,] 0.9970806 0.9972304 0.9973819
 [7,] 0.9912560 0.9915676 0.9919790
 [8,] 1.0000000 0.9999788 0.9998562
 [9,] 0.9999788 1.0000000 0.9999455
[10,] 0.9998562 0.9999455 1.0000000
```

SEVERE OVER-FITTING WITH
CHIPS 2 AND 3.
SOLUTION? GENOTYPE MORE INDIVIDUALS OR USE LESS
GREEDY MODEL

Estimating a posterior expectation and variance from samples

Posterior Expectation: $E(\boldsymbol{\theta}|\mathbf{y}) = \int \boldsymbol{\theta} p(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta}$

May be posterior is unknown or integral impossible to compute

Samples available from $[\boldsymbol{\theta}|\mathbf{y}]$

$\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \ldots, \boldsymbol{\theta}^{(S)}$

Estimate integral as $\widehat{E}(\boldsymbol{\theta}|\mathbf{y}) = \frac{1}{S} \sum_{i=1}^{S} \boldsymbol{\theta}^{(i)}$

$$\text{Monte Carlo Error} = \widehat{E}(\boldsymbol{\theta}|\mathbf{y}) - E(\boldsymbol{\theta}|\mathbf{y})$$
$$= \frac{1}{S} \sum_{i=1}^{S} \boldsymbol{\theta}^{(i)} - E(\boldsymbol{\theta}|\mathbf{y})$$

Goes to *0* as
S tends to infinity

Monte Carlo Variance of estimate of posterior mean

Measures variability to be expected if repeated sampling (each time S samples drawn) is done from the posterior

$$Var\left(\text{Monte Carlo Error}\right) = Var_{\theta|\mathbf{y}}\left[\widehat{E}(\theta|\mathbf{y}) - E(\theta|\mathbf{y})\right]$$

$$\begin{aligned} Var\left(\text{Monte Carlo Error}\right) &= Var_{\theta|\mathbf{y}}\left[\frac{1}{S}\sum_{i=1}^{S}\theta^{(i)} - E(\theta|\mathbf{y})\right] \\ &= Var_{\theta|\mathbf{y}}\left[\frac{1}{S}\sum_{i=1}^{S}\theta^{(i)}\right] \end{aligned}$$

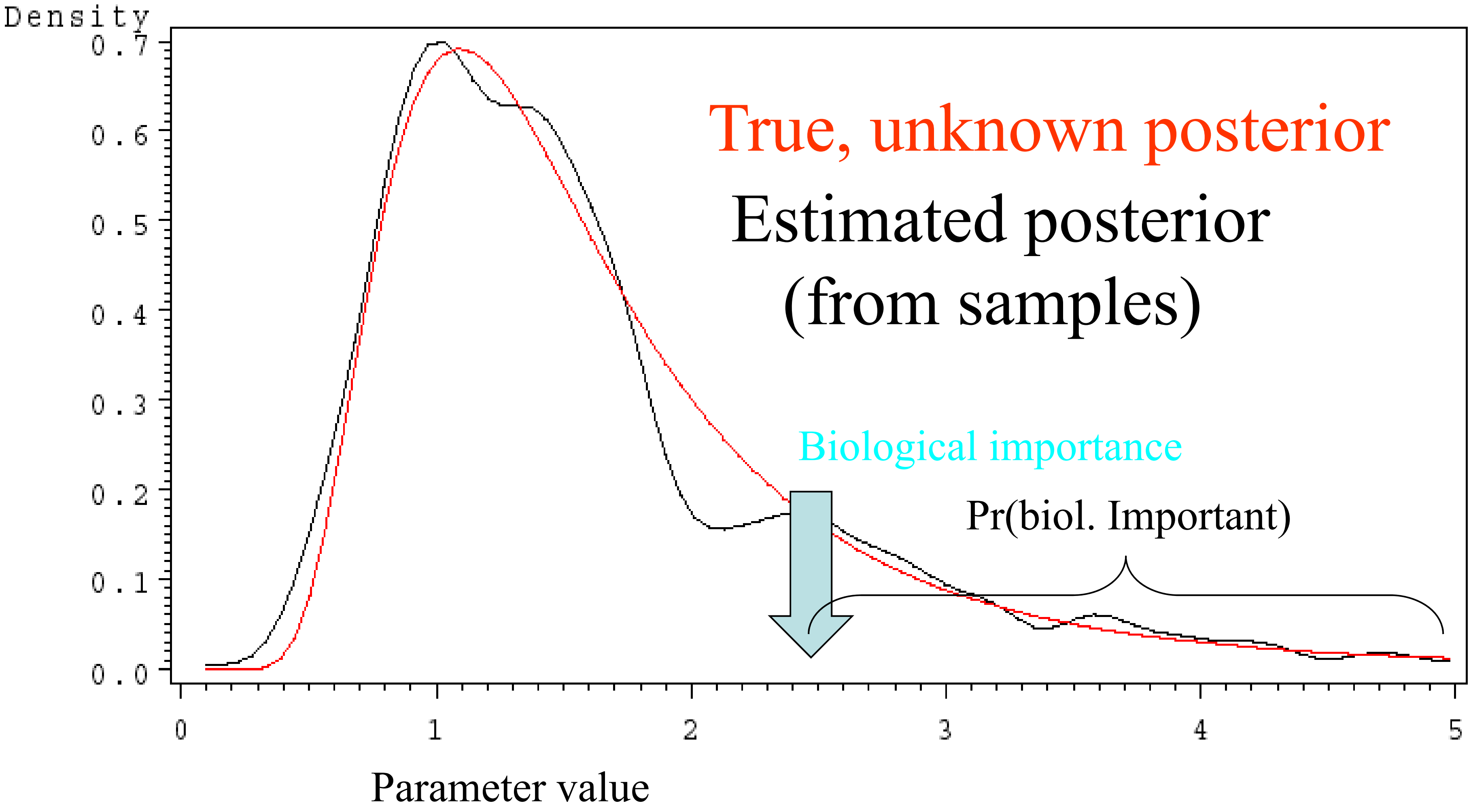
$$
\begin{aligned}
Var(\text{MCE}) &= \frac{1}{S^2}\left[\sum_{i=1}^{S} Var_{\boldsymbol{\theta}|\mathbf{y}}(\boldsymbol{\theta}^{(i)}) + 2\sum\sum\nolimits_{i<j} Cov_{\boldsymbol{\theta}|\mathbf{y}}(\boldsymbol{\theta}^{(i)}, \boldsymbol{\theta}^{(i)})\right] \\
&= \frac{1}{S^2}\left[\sum_{i=1}^{S} Var(\boldsymbol{\theta}|\mathbf{y}) + 2Var(\boldsymbol{\theta}|\mathbf{y})\sum\sum\nolimits_{i<j} \rho_{ij}\right] \\
&= \frac{Var(\boldsymbol{\theta}|\mathbf{y})}{S}\left(1 + \frac{2}{S}\sum\sum\nolimits_{i<j} \boxed{\rho_{ij}}\right)
\end{aligned}
$$

Null only if samples
are independent

- IF MARKOV CHAIN MONTE CARLO SAMPLING IS PRACTICED, SAMPLES ARE TYPICALLY SERIALLY CORRELATED

- IMPORTANT TO EVALUATE AUTO-CORRELATIONS IN MCMC, TO ASSES MONTE CARLO ERROR

Exact and estimated posterior densities (most of the time we will not be able to derive the posterior, but may be able to sample from it)



MCMC FOR BAYESIAN INFERENCE

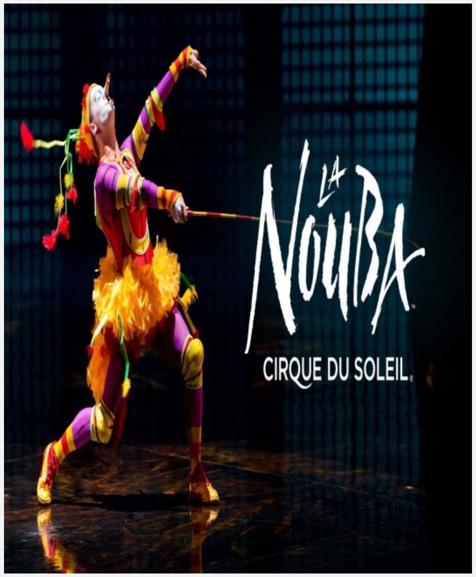
EXTRAVAGANT PRIORS coming out of the blue…

Great! No limits to imagination!

e.g., Bayes GRRR%##@*#

PRIOR

DATA



POSTERIOR

MARKOV CHAIN MONTE-CARLO

METROPOLIS-HASTINGS ALGORITHM

…and derivatives

1. FORM OF ALGORITHM: can be used to sample from any distribution, known or unknown

1. Generate candidate θ^* from proposal density $f(\theta^* | \theta^{[t-1]})$

2. Draw random number $U(0,1)$

3. Compute ratio

Posterior or conditional posterior

$$R = \frac{g(\theta^*)/f(\theta^* | \theta^{[t-1]})}{g(\theta^{[t-1]})/f(\theta^{[t-1]} | \theta^*)}$$

4. If $\begin{cases} U < \min(R,1) \text{ set } \theta^{[t]} = \theta^* \\ \theta^{[t]} = \theta^{[t-1]} \end{cases}$

Important: sample not rejected.
Chain value is just repeated

Integration constant is not needed

$$R = \frac{cp(y|\theta^*)p(\theta^*)/f(\theta^* | \theta^{[t-1]})}{cp(y|\theta^{[t-1]})p(\theta^{[t-1]})/f(\theta^{[t-1]} | \theta^*)}$$

$$= \boxed{\frac{p(y|\theta^*)p(\theta^*)/f(\theta^* | \theta^{[t-1]})}{p(y|\theta^{[t-1]})p(\theta^{[t-1]})/f(\theta^{[t-1]} | \theta^*)}}$$

GIBBS SAMPLING: MH USES CONDITIONAL POSTERIORS AS PROPOSAL (all proposal accepted)

Want to sample from joint posterior

[A,B,C|DATA]

Suppose sample j is

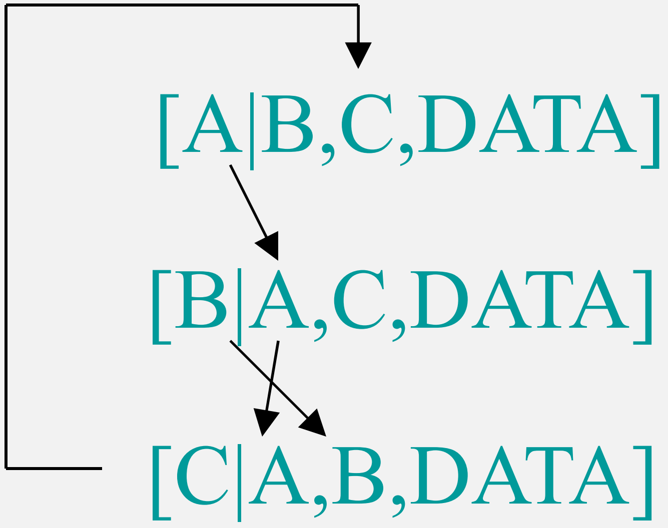
$[A^{(j)}, B^{(j)}, C^{(j)} | \text{DATA}]$

Each coordinate is a draw from marginal posterior

$[A^{(j)} | \text{DATA}]$ $[B^{(j)} | \text{DATA}]$ $[C^{(j)} | \text{DATA}]$

Gibbs sampling works as follows:

1) Form all fully conditional posteriors
2) Draw and update successively
3) Repeat a number of times without storing samples (burn-in)
4) Collect all subsequent samples, and thin them if needed for storage purposes



[A|B,C,DATA]

[B|A,C,DATA]

[C|A,B,DATA]

At the end of process:

j	A	B	C
1	$A^{(1)}$	$B^{(1)}$	$C^{(1)}$
2	$A^{(2)}$	$B^{(2)}$	$C^{(2)}$
.	.	.	.
t	$A^{(t)}$	$B^{(t)}$	$C^{(t)}$
t+1	$A^{(t+1)}$	$B^{(t+1)}$	$C^{(t+1)}$
.	.	.	.
t+m	$A^{(t+m)}$	$B^{(t+m)}$	$C^{(t+m)}$

Discard
first t samples
as burn-in

Keep subsequent
m samples for
Posterior analysis

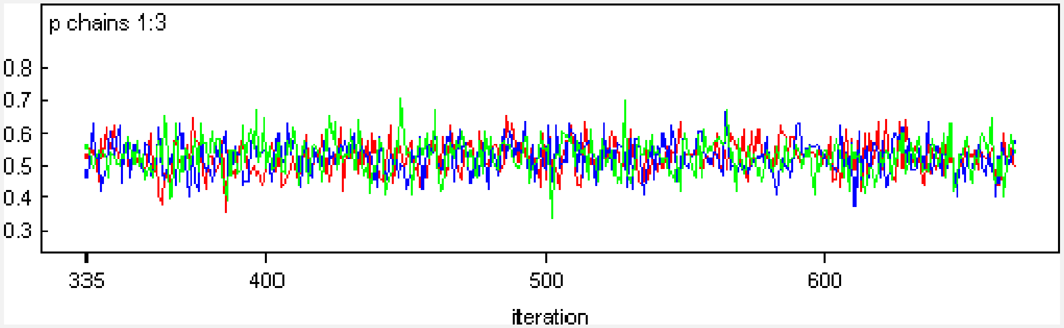
IMPORTANT POINTS:

- CONVERGENCE IS NOT TO A VALUE (E.G. MAXIMUM LIKELIHOOD ESTIMATE
- CONVERGENCE IS TO A DISTRIBUTION SO VALUES WILL CONTINUE FLUCTUATING AT RANDOM
- ‘MIXING” (CHAIN BEHAVIOR) MAY BE FAST OR SLOW
- DANGEROUS TO KEEP EARLY ITERATES OF THE MCMC. SHOULD BE CAREFUL
- THINGS YOU CAN DO (Gelman at al., 2014):

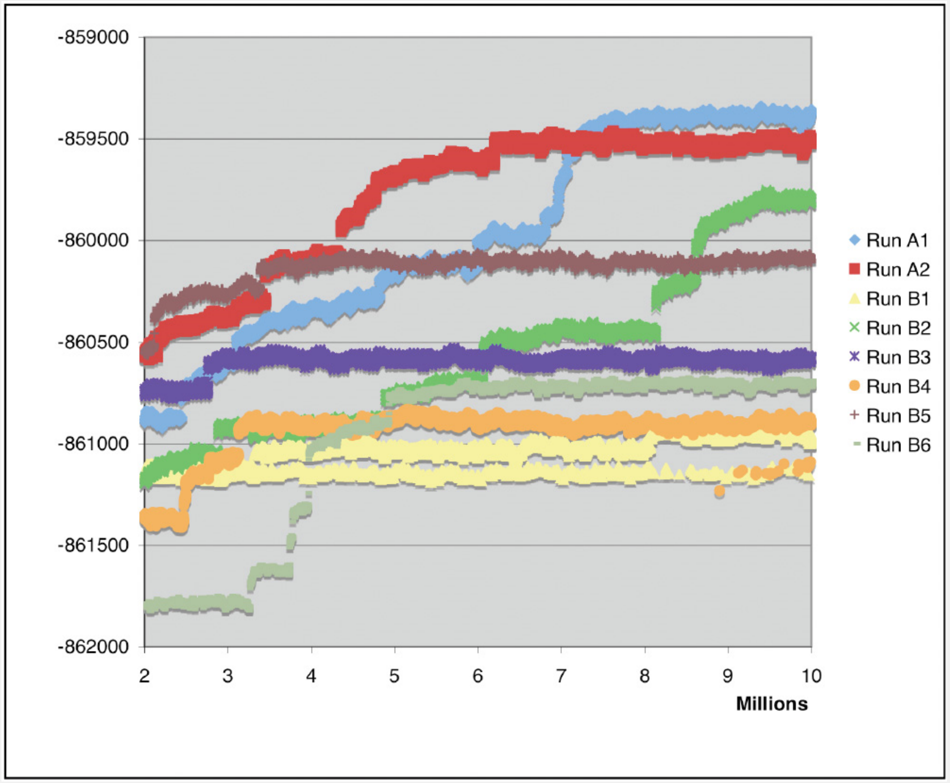
➔ MONITOR TRACE PLOTS OF CRITICAL PARAMETERS
➔ DISCARD EARLY ITERATES (CONSERVATIVE BURN-IN PERIOD
➔ RUN MULTIPLE CHAINS STARTING FROM OVERDISPERSED POINTS
➔ SPLIT SEQUENCES INTO PARTS AND COMPARE PARTS
➔ COMPARE BETWEEN AND WITHIN RUN VARIATIONS

USE PROGRAMS THAT EXAMINE CHAIN BEHAVIOR: BOA
http://cran.r-project.org/web/packages/boa/index.html

HOW DO TRACE PLOTS LOOK LIKE?



Three different chains are stationary, suggesting convergence to target posterior distribution



Several runs started from different places seems to converge to different places

EXAMPLE: MH FOR A GLIM (Carlin and Louis, 2000)

Number of flour beetles killed after exposure to carbon disulphide

Dosage	No. Killed	No. Exposed
$\underline{w_i}$	$\underline{y_i}$	$\underline{n_i}$
1.6097	6	59
1.7242	13	60
1.7552	18	62
1.7842	28	56
1.8113	52	63
1.8369	53	59
1.8610	61	62
1.8839	60	60

Generalized logit model

$$\Pr(death|w) = h(w) = \left[\frac{\exp(x)}{1+\exp(x)} \right]^{m_1}$$

$w_i =$ dose $i = 1,2,\ldots,k$

$x = \dfrac{w-\mu}{\sigma}$

$m_1 > 0$

Unknown parameters

Priors

$$m_1 \sim Gamma(a_0, b_0) \propto m_1^{a_0-1} \exp\left(-\frac{m_1}{b_0}\right) \qquad \mu \sim N(c_0, d_0)$$

$$\sigma^2 \sim Inverse\ Gamma(e_0, f_0) \propto (\sigma^2)^{-(e_0+1)} \exp\left(-\frac{1}{f_0 \sigma^2}\right)$$

Joint posterior

$$
\begin{aligned}
& p(\mu,\sigma^2,m_1|\mathbf{y},a_0,b_0,c_0,d_0,e_0,f_0) \\
\propto & \left\{\prod_{i=1}^{k}[h(w_i)]^{y_i}[1-h(w_i)]^{n_i-y_i}\right\}\exp\left[-\frac{(\mu-c_0)^2}{2d_0^2}\right] \\
& \times(\sigma^2)^{-(e_0+1)}\exp\left(-\frac{1}{f_0\sigma^2}\right)m_1^{a_0-1}\exp\left(-\frac{m_1}{b_0}\right) \\
\propto & \left\{\prod_{i=1}^{k}[h(w_i)]^{y_i}[1-h(w_i)]^{n_i-y_i}\right\}\frac{m_1^{a_0-1}}{(\sigma^2)^{(e_0+1)}}\exp\left[-\frac{(\mu-c_0)^2}{2d_0^2}-\frac{m_1}{b_0}-\frac{1}{f_0\sigma^2}\right]
\end{aligned}
$$

Joint posterior is not recognizable…Use Metropolis-Hastings

Hyper-parameters: a_0= .25, b_0=4, c_0=2, d_0=10, e_0=2.000004, f_0=1000

1) Metropolis-Hastings proposal distribution used

$$\begin{bmatrix} \theta_1^* \\ \theta_2^* \\ \theta_3^* \end{bmatrix} \sim N\left(\begin{bmatrix} \theta_1^{[t-1]} \\ \theta_2^{[t-1]} \\ \theta_3^{[t-1]} \end{bmatrix}, \mathbf{D} = \begin{bmatrix} .00012 & 0 & 0 \\ 0 & .033 & 0 \\ 0 & 0 & .10 \end{bmatrix} \right)$$

$$R = \frac{p(y|\boldsymbol{\theta}^*)p(\boldsymbol{\theta}^*)/f\left(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{[t-1]}\right)}{p\left(y|\boldsymbol{\theta}^{[t-1}\right)p\left(\boldsymbol{\theta}^{[t-1}\right)/f\left(\boldsymbol{\theta}^{[t-1]}|\boldsymbol{\theta}^*\right)}$$

$$f(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{[t-1]}) = \frac{1}{(2\pi)^3|\mathbf{D}|}\exp\left[-\frac{1}{2}(\boldsymbol{\theta}^* - \boldsymbol{\theta}^{[t-1]})'\mathbf{D}^{-1}(\boldsymbol{\theta}^* - \boldsymbol{\theta}^{[t-1]})\right]$$

$$f(\boldsymbol{\theta}^{[t-1]}|\boldsymbol{\theta}^*) = \frac{1}{(2\pi)^3|\mathbf{D}|}\exp\left[-\frac{1}{2}(\boldsymbol{\theta}^{[t-1]} - \boldsymbol{\theta}^*)'\mathbf{D}^{-1}(\boldsymbol{\theta}^{[t-1]} - \boldsymbol{\theta}^*)\right]$$

$$f(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{[t-1]}) = f(\boldsymbol{\theta}^{[t-1]}|\boldsymbol{\theta}^*)$$

Symmetric: use METROPOLIS RATIO

➔Three parallel chains run each with 10,000 iterations

➔ Burn-in= 2,000 in each chain

➔Histograms based on the (10,000-2,000)3= 24,000 sampled values

➔Autocorrelations and inter-correlations estimated from chain 2

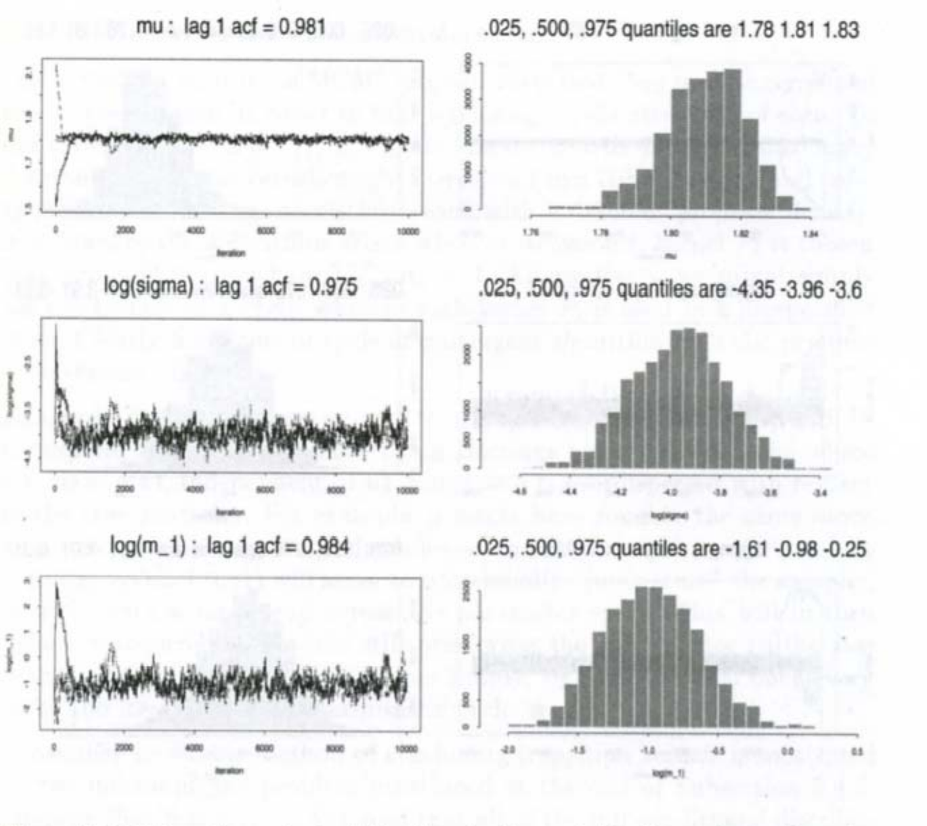


Figure 5.7 *Metropolis analysis of the flour beetle mortality data using a Gaussian proposal density with a diagonal* $\widetilde{\Sigma}$ *matrix. Monitoring plots use three parallel chains, and histograms use all samples following iteration 2000. Overall Metropolis acceptance rate: 13.5%.*

$$\begin{bmatrix} 1 & -0.78 & -0.94 \\ -0.78 & 1 & 0.89 \\ & & 1 \end{bmatrix}$$

-Chains mixed slowly (13.5% acceptance rate)

-High correlations between parameters:

-Makes sense to explore different proposal

2) Metropolis-Hastings proposal distribution used

➔From first algorithm ,estimate posterior covariance matrix as $\hat{\boldsymbol{\Sigma}} = \frac{1}{m}\sum_{j=1}^{m}(\boldsymbol{\theta}^{(j)} - \overline{\boldsymbol{\theta}})(\boldsymbol{\theta}^{(j)} - \overline{\boldsymbol{\theta}})'$

➔Use Gaussian proposal with covariance matrix (gave acceptance rate 27.3%)

$$\boldsymbol{\Psi} = 2\hat{\boldsymbol{\Sigma}} = \begin{bmatrix} 0.000292 & -003546 & -0.007856 \\ -003546 & 0.074733 & 0.117809 \\ -0.007856 & 0.117809 & 0.241551 \end{bmatrix}$$

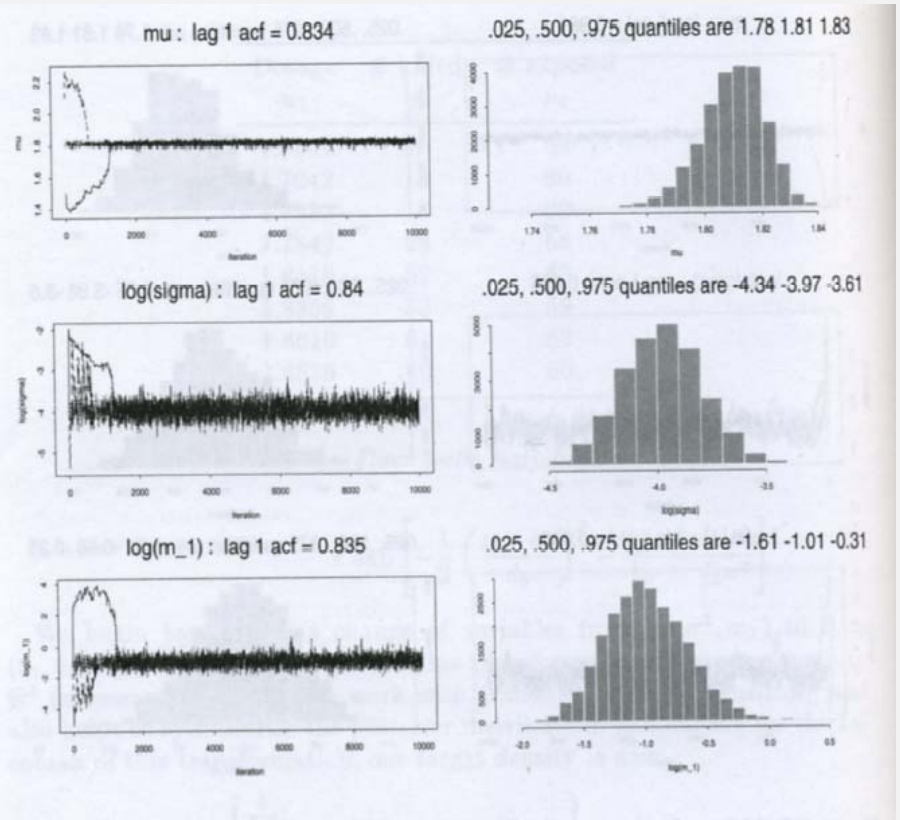


Figure 5.8 *Metropolis analysis of the flour beetle mortality data using a Gaussian proposal density with a nondiagonal* $\tilde{\Sigma}$ *matrix. Monitoring plots use three parallel chains, and histograms use all samples following iteration 2000. Overall Metropolis acceptance rate: 27.3%.*

EXAMPLE OF GIBBS SAMPLING

BAYESIAN REGRESSION MODEL WITH KNOWN VARIANCE RATIOS

$$\mathbf{y} = \mathbf{1}\beta_0 + \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{e}$$

Priors :

$$\rightarrow \beta_0 \propto \text{constant ("flat" prior)}$$

Conditional priors

$$\rightarrow \boldsymbol{\beta}_1 \sim N\left(0, \frac{\sigma_e^2}{\lambda_1}\right); \lambda_1 = \frac{\sigma_e^2}{\sigma_{\beta_1}^2} \text{ Known}$$

$$\rightarrow \boldsymbol{\beta}_2 \sim N\left(0, \frac{\sigma_e^2}{\lambda_2}\right); \lambda_2 = \frac{\sigma_e^2}{\sigma_{\beta_2}^2} \text{ Known}$$

$$\rightarrow \sigma_e^2 \sim \nu S^2 \chi_\nu^{-2}; \nu = 5, S^2 = \frac{\sigma_y^2}{2} \text{ Known}$$

Residual variance unknown

Conditional posteriors

(shown without derivation)

scalar

$$\beta_0|\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\sigma_e^2,\lambda_1,\lambda_2,v,S^2,\mathbf{y} = \beta_0|\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\sigma_e^2,\mathbf{y}$$

$$\sim N\left(\frac{1}{n}\mathbf{1}'(\mathbf{y}-\mathbf{1}\beta_0-\mathbf{X}_1\boldsymbol{\beta}_1-\mathbf{X}_2\boldsymbol{\beta}_2),\frac{\sigma_e^2}{n}\right)$$

p1 dimensions

$$\boldsymbol{\beta}_1|\beta_0,\boldsymbol{\beta}_2,\sigma_e^2,\lambda_1,\lambda_2,v,S^2,\mathbf{y} = \boldsymbol{\beta}_1|\boldsymbol{\beta}_2,\sigma_e^2,\lambda_1,\mathbf{y}$$

$$\sim N\left[(\mathbf{X}_1'\mathbf{X}_1+\mathbf{I}\lambda_1)^{-1}\mathbf{X}_1'(\mathbf{y}-\mathbf{1}\beta_0-\mathbf{X}_2\boldsymbol{\beta}_2),(\mathbf{X}_1'\mathbf{X}_1+\mathbf{I}\lambda_1)^{-1}\sigma_e^2\right]$$

p2 dimensions

$$\boldsymbol{\beta}_2|\beta_0,\boldsymbol{\beta}_1,\sigma_e^2,\lambda_1,\lambda_2,v,S^2,\mathbf{y} = \boldsymbol{\beta}_2|\boldsymbol{\beta}_1,\sigma_e^2,\lambda_2,\mathbf{y}$$

$$\sim N\left[(\mathbf{X}_2'\mathbf{X}_2+\mathbf{I}\lambda_2)^{-1}\mathbf{X}_2'(\mathbf{y}-\mathbf{1}\beta_0-\mathbf{X}_1\boldsymbol{\beta}_1),(\mathbf{X}_2'\mathbf{X}_2+\mathbf{I}\lambda_2)^{-1}\sigma_e^2\right]$$

scalar

$$\sigma_e^2|\beta_0,\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\lambda_1,\lambda_2,v,S^2,\mathbf{y}$$

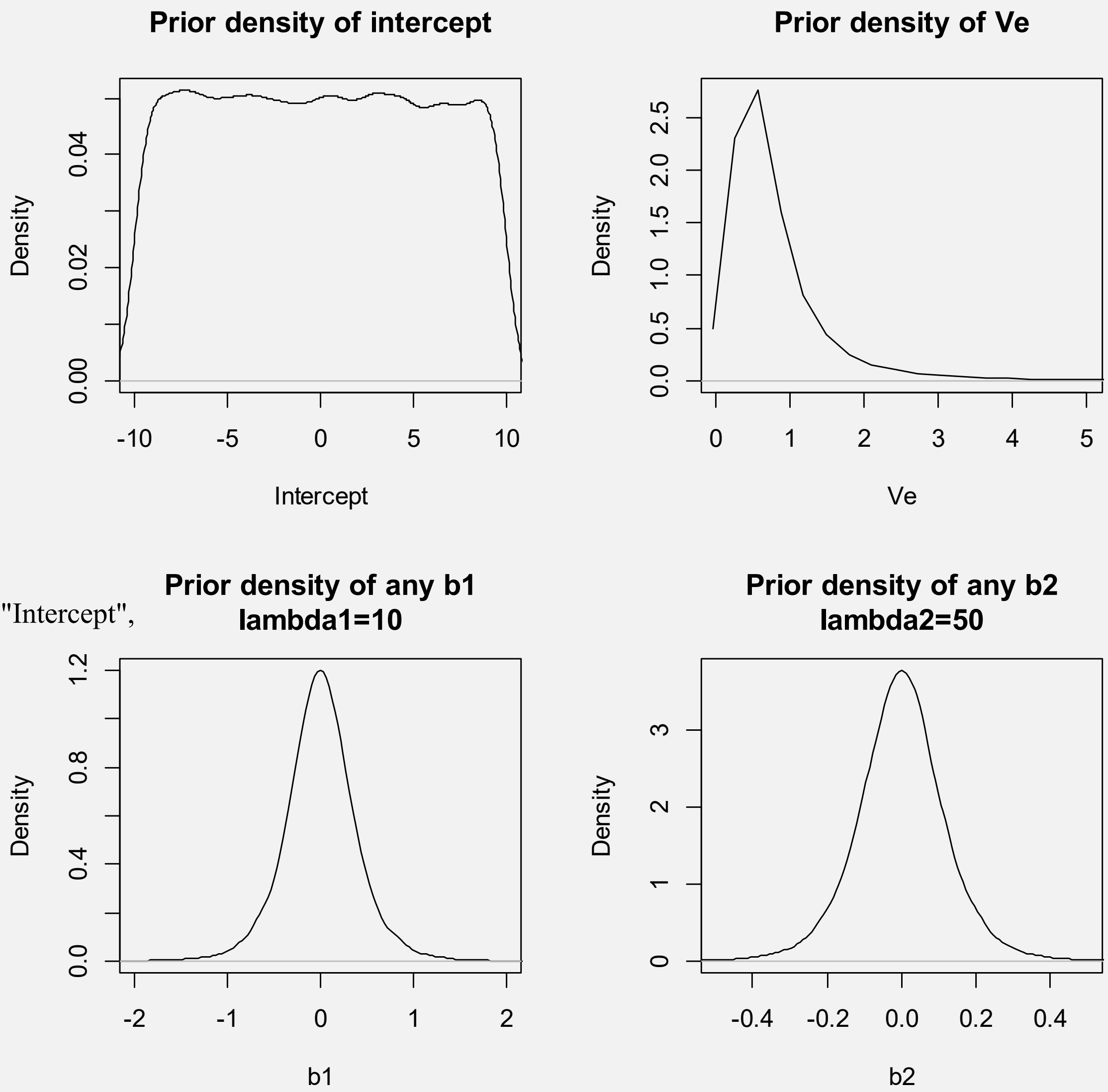
$$\sim (n+v+p_1+p_2)\frac{SSE+vS^2+\lambda_1\boldsymbol{\beta}_1'\boldsymbol{\beta}_1+\lambda_2\boldsymbol{\beta}_2'\boldsymbol{\beta}_2}{(n+v+p_1+p_2)}\chi^{-2}_{(n+v+p_1+p_2)}$$

$$SSE = (\mathbf{y}-\mathbf{1}\beta_0-\mathbf{X}_1\boldsymbol{\beta}_1-\mathbf{X}_2\boldsymbol{\beta}_2)'(\mathbf{y}-\mathbf{1}\beta_0-\mathbf{X}_1\boldsymbol{\beta}_1-\mathbf{X}_2\boldsymbol{\beta}_2)$$

Note offsets in regression coefficients

```
###USE OF GIBBS SAMPLER
###2 SETS OF MARKERS, KNOWN VARIANCE RATIOS
###UNKNOWN VE
###WHEAT DATA

rm(list=ls(all=TRUE))
library(MASS)
library(BGLR)
set.seed(1234567)

###LOAD DATA
data(wheat)
Y<-wheat.Y
X<-wheat.X
y<-Y[,1]
n<-length(y)

###TWO SETS OF MARKERS
###FIRST SET: 1:9
###SECOND SET: 1001:1006

X1<-X[,1:9]
X2<-X[,1001:1006]

p1<-ncol(X1)
p2<-ncol(X2)

###Prior distribution of marker effects set 1: N(0,ve/lambda1)
###Prior distribution of marker effects set 2: N(0,ve/lambda2)
###Assign arbitrary scaled inverted chi-square prior to ve. Scale=Var(y)/2,
###df=5

lambda1<-5
lambda2<-50
scale<-var(y)/2
nu<-5
```

```
####PLOT PRIOR DISTRIBUTIONS
####THE PRIORS ARE ESTIMATED EMPIRICALLY
####BY DRAWING 50,000 SAMPLES FROM PRIORS
####NOTE THAT PRIORS ON MARKERS DEPEND ON VE
####SO MUST SAMPLE FROM MARGINAL PRIORS

priorve<-nu*scale/rchisq(50000,nu)

b0prior<-runif(50000,-10,10)
b1prior<-numeric(50000)
b2prior<-numeric(50000)

for (i in 1:length(priorve)){


b1prior[i]<-rnorm(1,0,sqrt(priorve[i]/lambda1))
b2prior[i]<-rnorm(1,0,sqrt(priorve[i]/lambda2))
}


par(mfrow=c(2,2))
plot(density(b0prior),main="Prior density of intercept",xlab="Intercept",
xlim=c(-10,10))

plot(density(priorve),main="Prior density of Ve",xlab="Ve",
xlim=c(0,5))

plot(density(b1prior),main="Prior density of any b1
lambda1=10",xlab="b1",
xlim=c(-0.20,0.20))

plot(density(b2prior),main="Prior density of any b2
lambda2=50",xlab="b2",
xlim=c(-0.10,0.10))

par(mfrow=c(1,1))
```



```
###FORM J VECTOR AND INCIDENCE MATRIX
J<-rep(1,n)
Xmodel<-cbind(1,X1,X2)

###SET NUMBER OF GIBBS SAMPLES AND BURN IN
###SHOULD USE MORE ITERATIONS IN SERIOUS ANALYSIS

NITER<-2000
BURN<-500
POSTERIORSAMPLES<-NITER-BURN

###DEFINE CHAIN AND OBJECTS STORING SAMPLES

nparams<-1+p1+p2+1
nparams

[1] 17

b0samp<-numeric(NITER)
b1samp<-matrix(nrow=NITER,ncol=p1)
b2samp<-matrix(nrow=NITER,ncol=p2)
vesamp<-numeric(NITER)

###CHAIN STARTING VALUES (ITERATION 1)

b0samp[1]<-0.0001
b1samp[1,]<-rep(0.0001,p1)
b2samp[1,]<-rep(0.0001,p2)
vesamp[1]<-var(y)/2


###FORM MATRICES OF SUM OF SQUARES AND PRODUCTS OF
###COLUMNS IN Xmod


XPX<-crossprod(Xmodel)
XPX0<-crossprod(Xmodel[,1])
XPX1<-crossprod(X1)
XPX2<-crossprod(X2)
```

```
###GIBBS SAMPLING (CONDITIONAL POSTERIORS
ARE ####PROPOSALS)
###NOTE THE OFFSETS and IMMEDIATE UPDATING

for (i in 2:NITER){
```

Sample intercept

```
mean0<-sum(y-X1%*%b1samp[i-1,]-X2%*%b2samp[i-
1,])/XPX0
var0<-vesamp[i-1]/XPX0

b0samp[i]<-rnorm(1,mean0,sqrt(var0))
```

Sample b1 vector

```
Sigma1<-chol2inv(chol(XPX1+lambda1*diag(p1)))
mean1<-Sigma1%*%t(X1)%*%(y-J*b0samp[i]-
X2%*%b2samp[i-1,])

b1samp[i,]<-mvrnorm(1,mean1,Sigma1*vesamp[i-1])
```

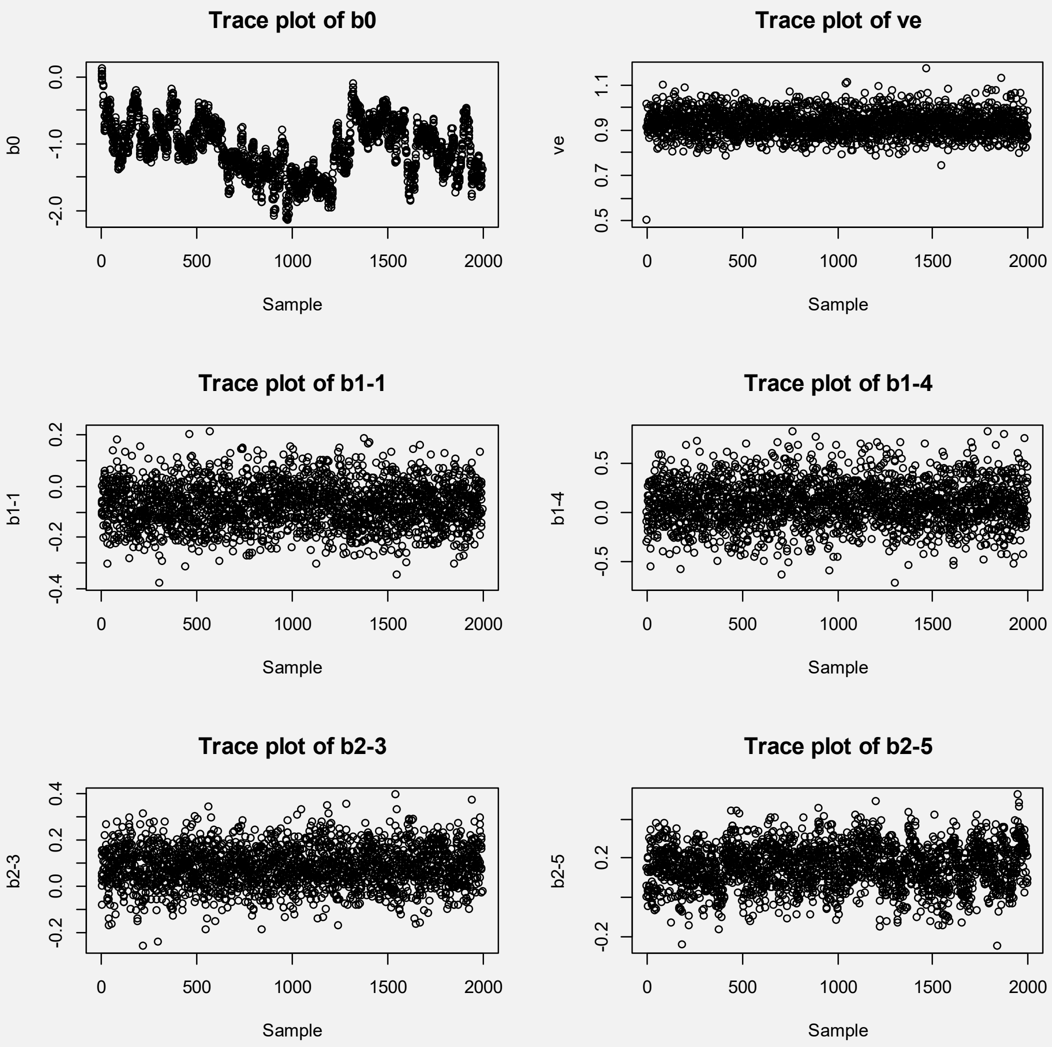
Sample b2 vector

```
Sigma2<-chol2inv(chol(XPX2+lambda2*diag(p2)))
mean2<-Sigma2%*%t(X2)%*%(y-J*b0samp[i]-
X1%*%b1samp[i-1,])

b2samp[i,]<-mvrnorm(1,mean2,Sigma2*vesamp[i-1])
```

Sample ve

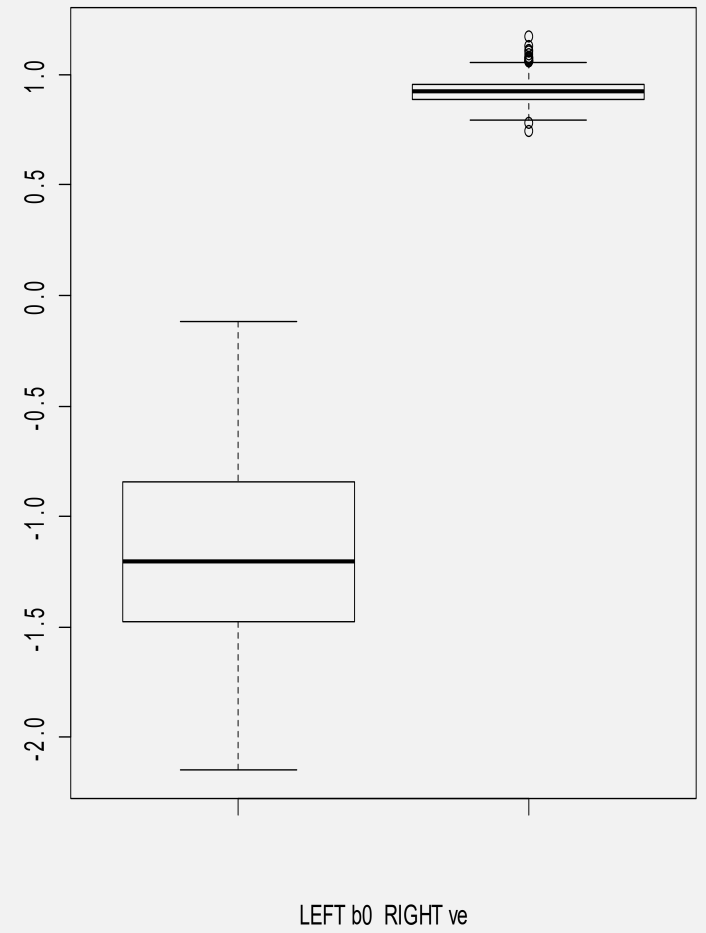
```
numodel<-nu+n+p1+p2
ssr<-crossprod(y-J*b0samp[i]-X1%*%b1samp[i,]-
X2%*%b2samp[i,])

vesamp[i]<-numodel*ssr/(numodel*rchisq(1,numodel))
}
```

```
####Look at some trace plots

par(mfrow=c(3,2))
plot(b0samp,ylab="b0",xlab="Sample",main="Trace plot of b0")
plot(vesamp,ylab="ve",xlab="Sample",main="Trace plot of ve")
plot(b1samp[,1],ylab="b1-1",xlab="Sample",main="Trace plot of b1-1")
plot(b1samp[,4],ylab="b1-4",xlab="Sample",main="Trace plot of b1-4")
plot(b2samp[,3],ylab="b2-3",xlab="Sample",main="Trace plot of b2-3")
plot(b2samp[,5],ylab="b2-5",xlab="Sample",main="Trace plot of b2-5")
par(mfrow=c(1,1))
```

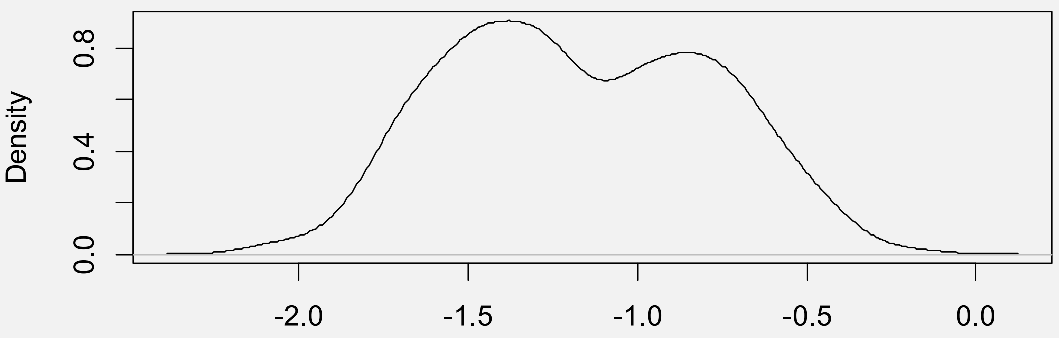


Intercept not mixing well

```
boxplot(b0samp[BURN+1:NITER],vesamp[BURN+1:NITER],xlab="LEFT b0  RIGHT ve")
par(mfrow=c(2,1))
plot(density(b0samp[501:NITER]),xlab="b0")
plot(density(vesamp[501:NITER]),xlab="ve")
par(mfrow=c(2,1))
```

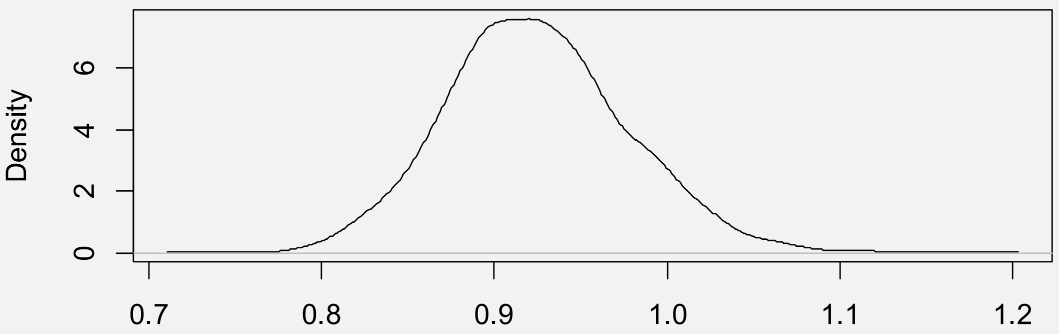


Post. density of b0



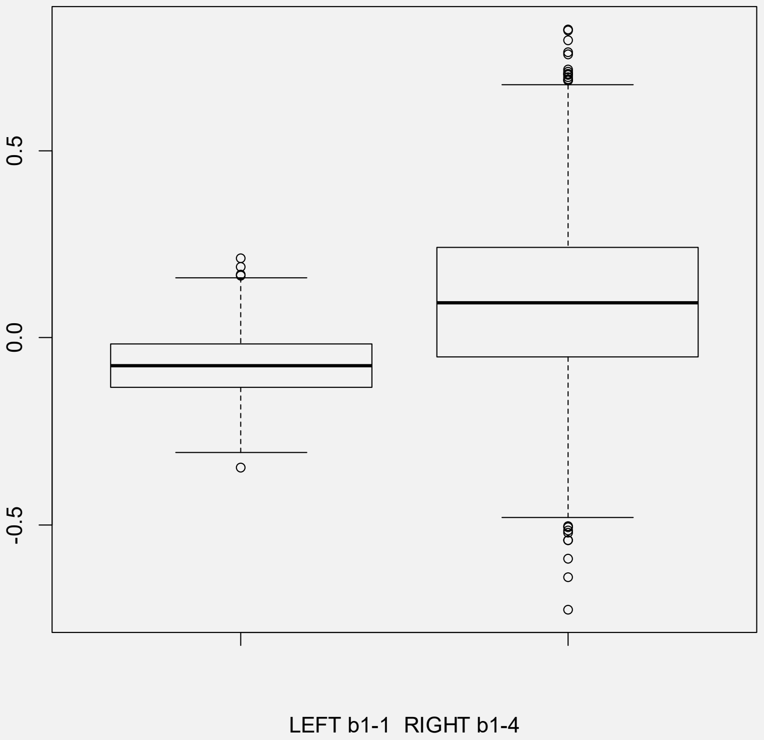
```
> summary(b0samp)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-2.1460 -1.3860 -1.0690 -1.0860 -0.7890  0.1143
```

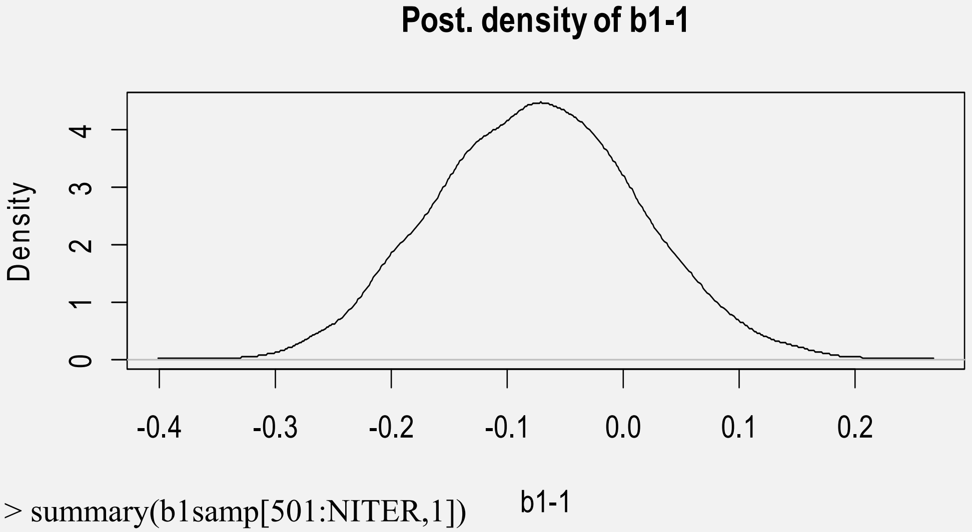
post. density of ve



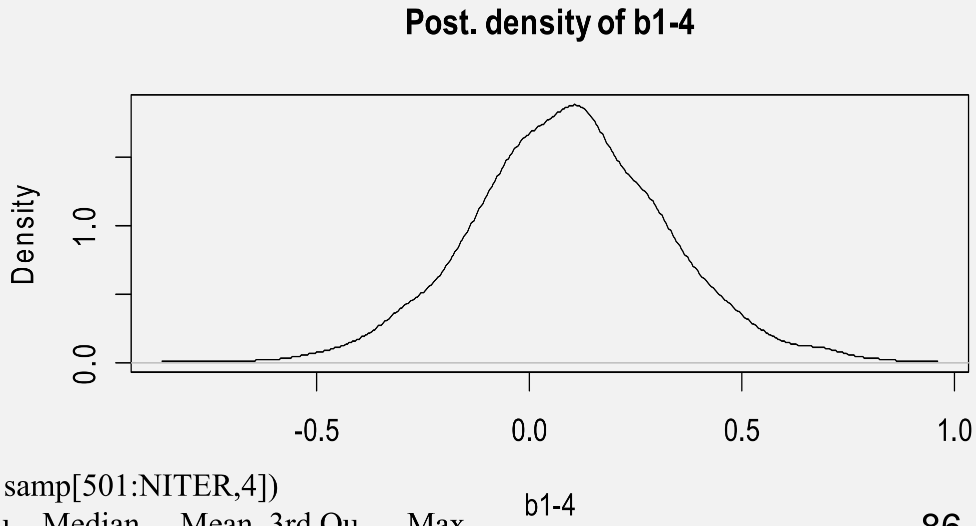
```
> summary(vesamp)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.5000  0.8891  0.9238  0.9258  0.9599  1.1710
```

```
boxplot(b1samp[501:NITER,1],b1samp[501:NITER,4],xlab="LEFT b1-1  RIGHT b1-4")
par(mfrow=c(2,1))
plot(density(b1samp[501:NITER,1]),xlab="b1-1")
plot(density(b1samp[501:NITER,4]),xlab="b1-4")
par(mfrow=c(2,1))
summary(b1samp[501:NITER,1])
summary(b1samp[501:NITER,4])
```

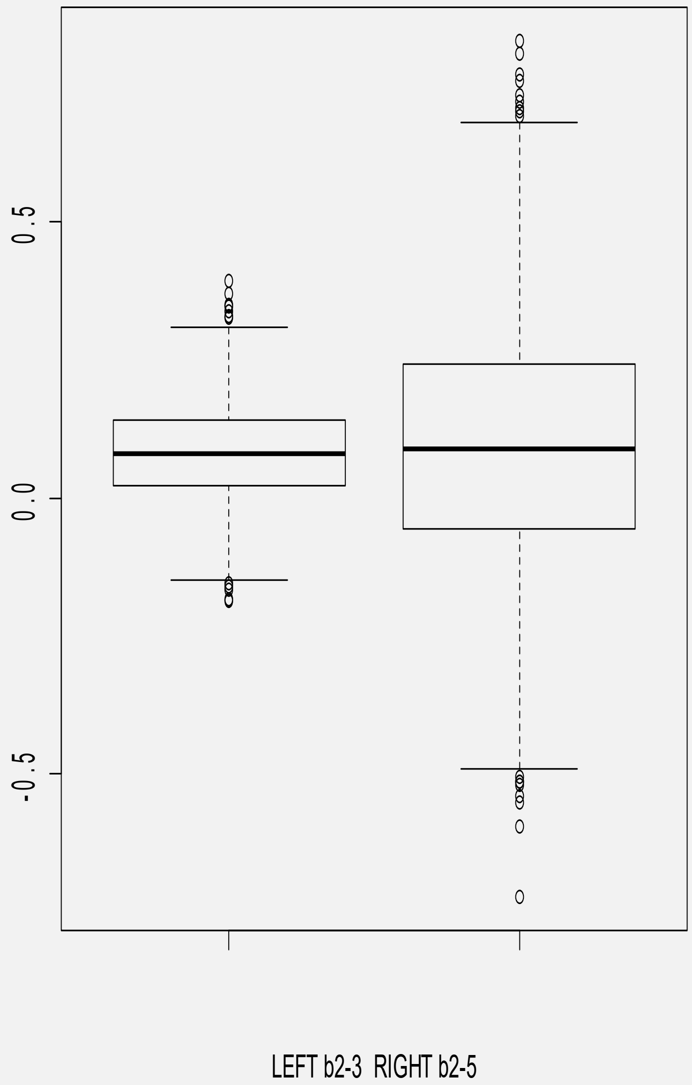


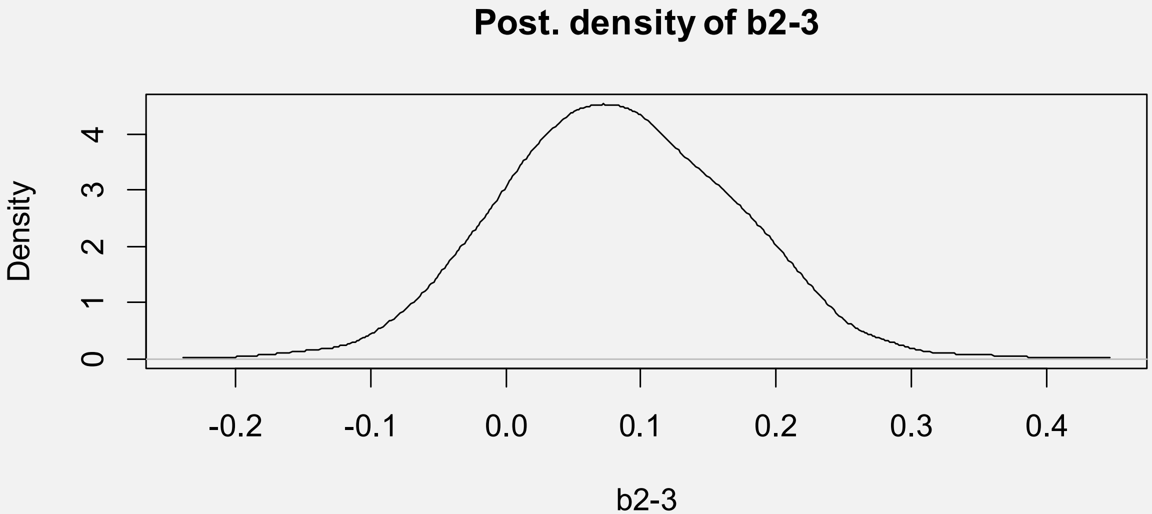


```
> summary(b1samp[501:NITER,1])
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
-0.34670 -0.13320 -0.07413 -0.07356 -0.01530  0.21360
```

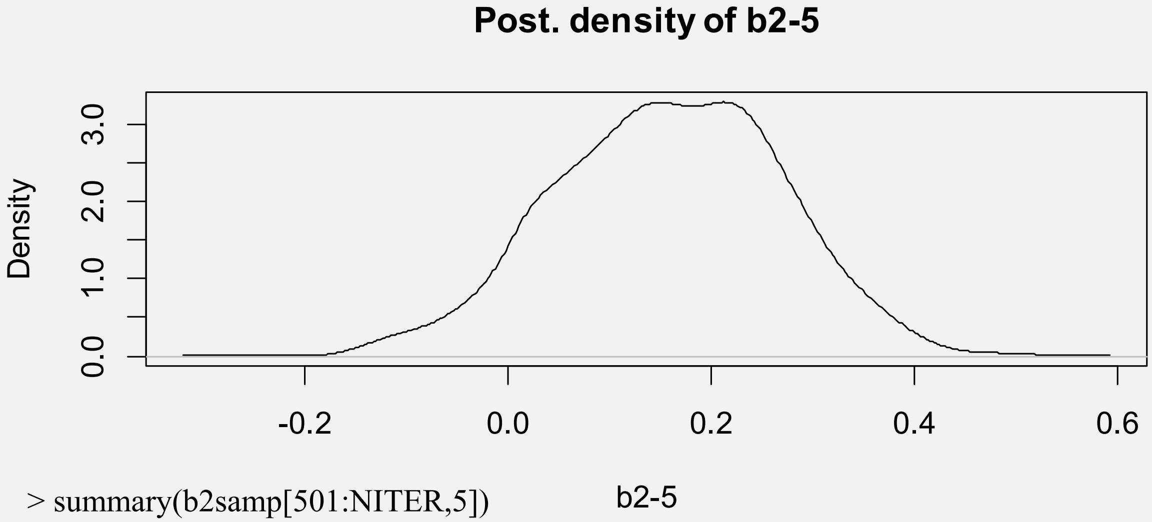


```
> summary(b1samp[501:NITER,4])
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
-0.72570 -0.05094  0.09404  0.09486  0.24280  0.82330
```




Post. density of b2-3

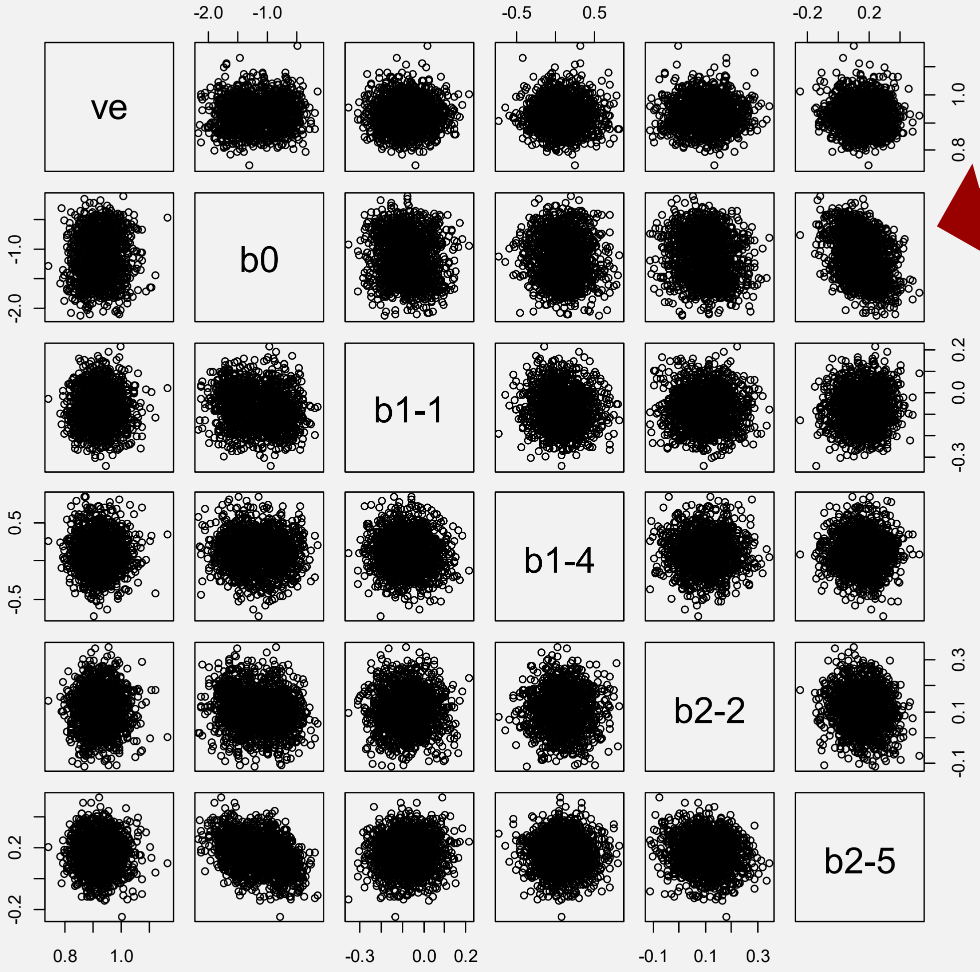
```
> summary(b2samp[501:NITER,3])
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
-0.18520  0.02295  0.08040  0.08276  0.14090  0.39330
```


Post. density of b2-5

```
> summary(b2samp[501:NITER,5])
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
-0.25080  0.07924  0.16060  0.15780  0.23710  0.52320
```

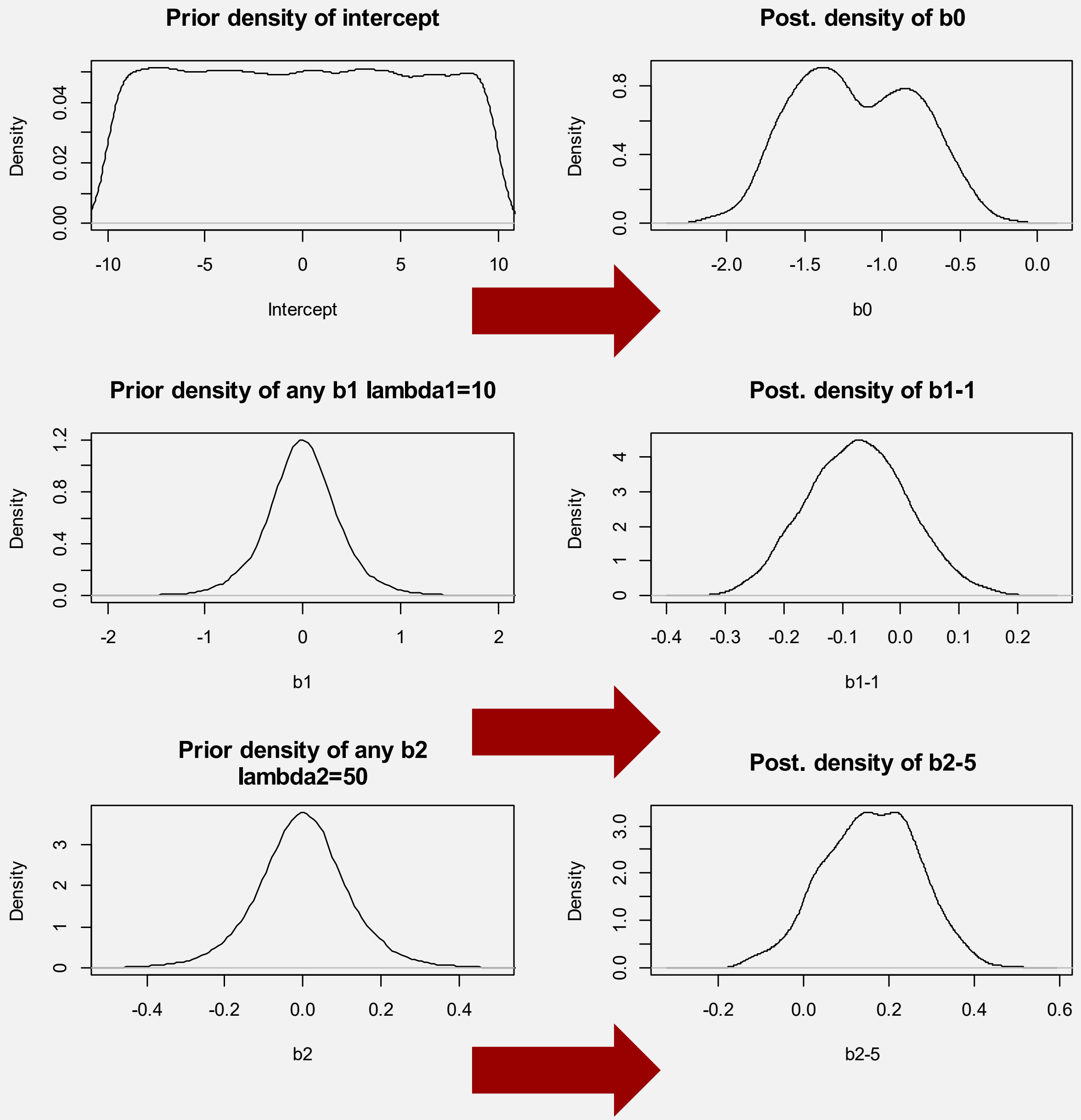
```
cor(BSAMPLES)
pairs(BSAMPLES,c("ve","b0","b1-1","b1-4","b2-2","b2-5"))
 [,1]       [,2]        [,3]        [,4]        [,5]
[1,]  1.000000000  0.07683002 -0.003356222 -0.018833592 -0.009192548
[2,]  0.076830016  1.00000000 -0.117588201 -0.066627510 -0.160606803
[3,] -0.003356222 -0.11758820  1.000000000 -0.064608782 -0.002809269
[4,] -0.018833592 -0.06662751 -0.064608782  1.000000000  0.012582993
[5,] -0.009192548 -0.16060680 -0.002809269  0.012582993  1.000000000
[6,] -0.046339531 -0.35995083  0.033435178 -0.007491403 -0.120513905
            [,6]
[1,] -0.046339531
[2,] -0.359950825
[3,]  0.033435178
[4,] -0.007491403
[5,] -0.120513905
```

POSTERIOR INTER-CORRELATION



Explanation:
86% of columns
of marker 1006
(b2-5) are 1!!!
J is vector of 1’s!!

Bayesian learning: from prior to posterior



BAYESIAN IDENTIFICATION

Let a Bayesian linear model consist of location parameters $\boldsymbol{\theta}_{\mathrm{A}}$ and $\boldsymbol{\theta}_{\mathrm{B}}$

likelihood $p(\mathbf{y} \mid \boldsymbol{\theta}_{\mathrm{A}}, \boldsymbol{\theta}_{\mathrm{B}})$

If the conditional posterior density of $\boldsymbol{\theta}_{\mathrm{B}}$ is such that

$$p(\theta_{\mathrm{B}} | \theta_{\mathrm{A}}, \mathbf{y}) = p(\theta_{\mathrm{B}} | \theta_{\mathrm{A}}),$$

$\boldsymbol{\theta}_{\mathrm{B}}$ is not identifiable

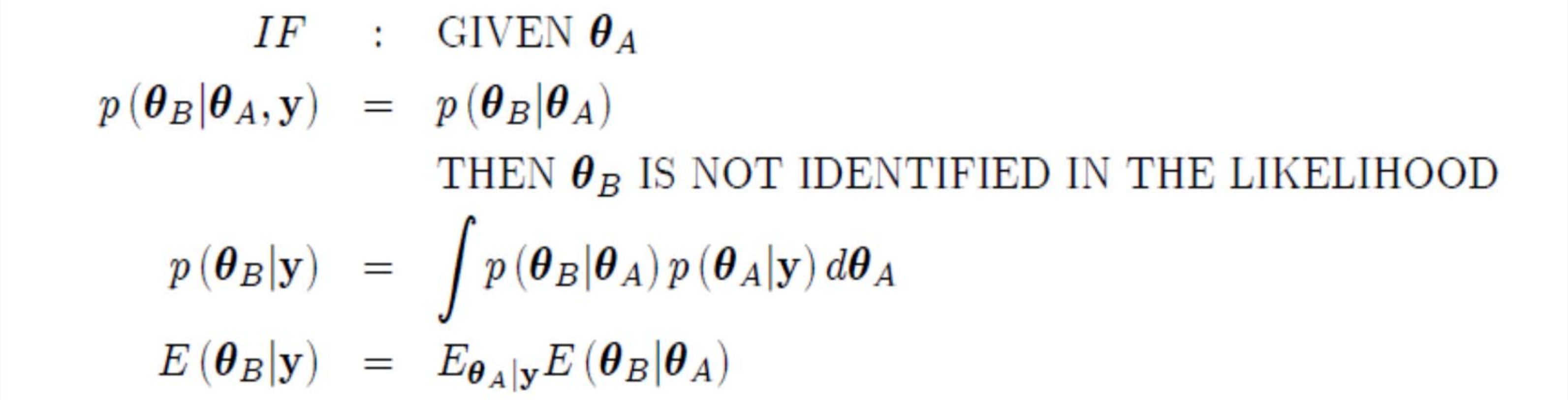
observation of data does not increase knowledge about $\boldsymbol{\theta}_{\mathrm{B}}$

beyond what is conveyed by the conditional prior $p(\boldsymbol{\theta}_{\mathrm{B}} \mid \boldsymbol{\theta}_{\mathrm{A}})$

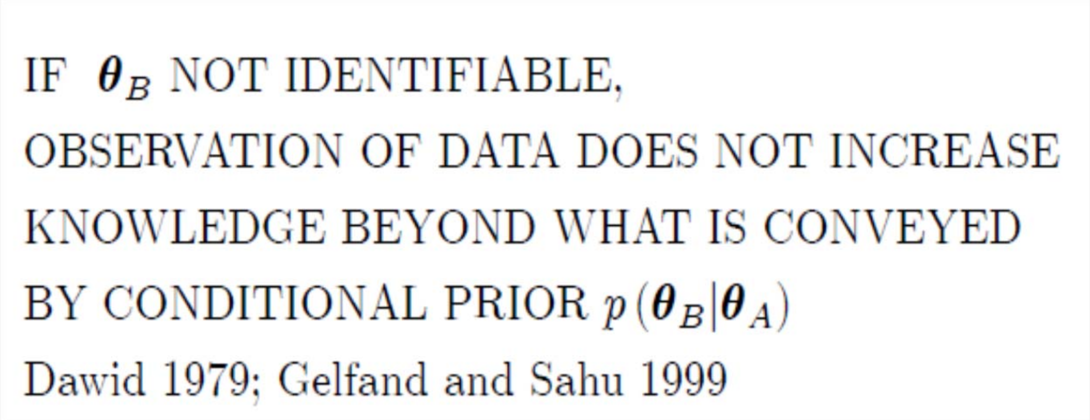
(Dawid 1979; Gelfand and Sahu 1999)

THE PRECEDING IS JUST AN EXAMPLE OF HOW A PRIOR MAY STRONGLY INFLUENCE INFERENCE IN THE n<p SITUATION.

THE GENERAL SETTING IS



$$\begin{aligned}
IF \quad &: \quad \text{GIVEN } \boldsymbol{\theta}_A \\
p\left(\boldsymbol{\theta}_B|\boldsymbol{\theta}_A,\mathbf{y}\right) &= p\left(\boldsymbol{\theta}_B|\boldsymbol{\theta}_A\right) \\
& \quad \text{THEN } \boldsymbol{\theta}_B \text{ IS NOT IDENTIFIED IN THE LIKELIHOOD} \\
p\left(\boldsymbol{\theta}_B|\mathbf{y}\right) &= \int p\left(\boldsymbol{\theta}_B|\boldsymbol{\theta}_A\right) p\left(\boldsymbol{\theta}_A|\mathbf{y}\right) d\boldsymbol{\theta}_A \\
E\left(\boldsymbol{\theta}_B|\mathbf{y}\right) &= E_{\boldsymbol{\theta}_A|\mathbf{y}} E\left(\boldsymbol{\theta}_B|\boldsymbol{\theta}_A\right)
\end{aligned}$$



IF $\boldsymbol{\theta}_B$ NOT IDENTIFIABLE,

OBSERVATION OF DATA DOES NOT INCREASE

KNOWLEDGE BEYOND WHAT IS CONVEYED

BY CONDITIONAL PRIOR $p\left(\boldsymbol{\theta}_B|\boldsymbol{\theta}_A\right)$

Dawid 1979; Gelfand and Sahu 1999

$$\begin{aligned} p(\theta_{\mathrm{B}}|\mathbf{y}) &= \int p(\theta_{\mathrm{B}}|\theta_{\mathrm{A}},\mathbf{y})p(\theta_{\mathrm{A}}|\mathbf{y})d\theta_{\mathrm{A}} = \int p(\theta_{\mathrm{B}}|\theta_{\mathrm{A}})p(\theta_{\mathrm{A}}|\mathbf{y})d\theta_{\mathrm{A}} \\ &= E_{p(\theta_{\mathrm{A}}|\mathbf{y})}[p(\theta_{\mathrm{B}}|\theta_{\mathrm{A}})]. \end{aligned}$$

If $\boldsymbol{\theta}_{\mathrm{B}}$ is any marker effect, say β_j, the preceding becomes

$$p\left(\beta_j|\mathbf{y}\right) = \int \left[p\left(\beta_j|\mathbf{X}_1\beta_1\right)\right]p(\mathbf{X}_1\beta_1|\mathbf{y})d(\mathbf{X}_1\beta_1).$$

In conclusion, for any letter of the alphabet and for any prior distribution adopted, any inference made about genetic architecture always depends on the form of $p(\beta_j|\mathrm{X}_1\boldsymbol{\beta}_1)$ or, more generally, of $p(\boldsymbol{\theta}_{\mathrm{B}}|\boldsymbol{\theta}_{\mathrm{A}})$, and these densities depend on the prior adopted, but not on the data. Proper Bayesian learning takes place for $\mathrm{X}_1\boldsymbol{\beta}_1$ only.

Example: under-identified Bayesian regression model

$$X\beta = [\mathbf{X}_1 \quad \mathbf{X}_2]\begin{bmatrix}\beta_1\\ \beta_2\end{bmatrix}$$

Note that θ_A is the signal

$$\begin{bmatrix}\theta_{\mathrm{A}}\\ \theta_{\mathrm{B}}\end{bmatrix} = \begin{bmatrix}\mathbf{X}_1 & \mathbf{X}_2\\ 0 & \mathbf{I}_{(p-n)\times(p-n)}\end{bmatrix}\begin{bmatrix}\beta_1\\ \beta_2\end{bmatrix}$$

$$\boldsymbol{\beta}_2 = \boldsymbol{\theta}_B \text{ and } \beta_1 = \mathbf{X}_1^{-1}(\boldsymbol{\theta}_{\mathrm{A}} - \mathbf{X}_2\boldsymbol{\theta}_{\mathrm{B}})$$

$$\mathbf{y} = \theta_{\mathrm{A}} + \mathbf{e},$$

implying that the data contain information about $\boldsymbol{\theta}_{\mathrm{A}}$ but not about $\boldsymbol{\theta}_{\mathrm{B}}$

$$p(\theta_{\mathrm{A}}, \theta_{\mathrm{B}}|\mathbf{y}) \propto p(\mathbf{y}|\theta_{\mathrm{A}}, \theta_{\mathrm{B}})p(\theta_{\mathrm{B}}|\theta_{\mathrm{A}})p(\theta_{\mathrm{A}}) \propto p(\mathbf{y}|\theta_{\mathrm{A}})p(\theta_{\mathrm{B}}|\theta_{\mathrm{A}})p(\theta_{\mathrm{A}}),$$

$$p(\theta_{\mathrm{B}}|\theta_{\mathrm{A}}, \mathbf{y}) = p(\theta_{\mathrm{B}}|\theta_{\mathrm{A}})$$

EXAMPLE (BAYESIAN TREATMENT OF BLUP WITH KNOWN VARIANCE COMPONENTS)

I :parameterization on markers

$$\begin{aligned} \mathbf{y}_{n\times 1} &= \mathbf{X}\boldsymbol{\beta}_{p\times 1} + \mathbf{e}; n < p \\ \mathbf{e}|\sigma_e^2 &\sim N\left(\mathbf{0}, \mathbf{I}\sigma_e^2\right) \\ \boldsymbol{\beta}|\sigma_\beta^2 &\sim N\left(\mathbf{0}, \mathbf{I}\sigma_\beta^2\right) \end{aligned}$$

II:parameterization on signal

$$\begin{aligned} \mathbf{y}_{n\times 1} &= \mathbf{g} + \mathbf{e}; \\ \mathbf{e}|\sigma_e^2 &\sim N\left(\mathbf{0}, \mathbf{I}\sigma_e^2\right) \\ &\mathbf{g}|\sigma_\beta^2 \sim N\left(\mathbf{0}, \mathbf{X}\mathbf{X}'\sigma_\beta^2\right) \end{aligned}$$

Can construct an augmented posterior distribution

Joint prior distribution

$$\begin{pmatrix}\mathbf{g}\\ \boldsymbol{\beta}\end{pmatrix}|\sigma_\beta^2 \sim N\left(\begin{pmatrix}\mathbf{0}\\ \mathbf{0}\end{pmatrix},\begin{pmatrix}\mathbf{XX}'\sigma_\beta^2 & \mathbf{X}\sigma_\beta^2\\ \mathbf{X}'\sigma_\beta^2 & \mathbf{I}\sigma_\beta^2\end{pmatrix}\right)$$

$$\begin{aligned}\boldsymbol{\beta}|\mathbf{g},\sigma_\beta^2 &\sim N\left(\mathbf{X}'\sigma_\beta^2\left(\mathbf{XX}'\sigma_\beta^2\right)^{-1}\mathbf{g},\mathbf{I}\sigma_\beta^2-\mathbf{X}'\sigma_\beta^2\left(\mathbf{XX}'\sigma_\beta^2\right)^{-1}\mathbf{X}\sigma_\beta^2\right)\\ &= N\left(\mathbf{X}'\left(\mathbf{XX}'\right)^{-1}\mathbf{g},\left[\mathbf{I}-\mathbf{X}'\left(\mathbf{XX}'\right)^{-1}\mathbf{X}\right]\sigma_\beta^2\right)\end{aligned}$$

Augmented posterior distribution

$$\begin{aligned}p\left(\boldsymbol{\beta},\mathbf{g}|\mathbf{y},\sigma_\beta^2\right) &= p\left(\boldsymbol{\beta}|\mathbf{g},\mathbf{y},\sigma_\beta^2\right)p\left(\mathbf{g}|\mathbf{y},\sigma_\beta^2\right)\\ &= p\left(\boldsymbol{\beta}|\mathbf{g},\sigma_\beta^2\right)p\left(\mathbf{g}|\mathbf{y},\sigma_\beta^2\right)\end{aligned}$$

Marginal posterior distribution

$$\begin{aligned}p\left(\boldsymbol{\beta}|\mathbf{y},\sigma_\beta^2\right) &= \int p\left(\boldsymbol{\beta}|\mathbf{g},\sigma_\beta^2\right)p\left(\mathbf{y}|\mathbf{g},\sigma_\beta^2\right)p\left(\mathbf{g}|\sigma_\beta^2\right)d\mathbf{g}\\ &= E_{\mathbf{g}|\mathbf{y},\sigma_\beta^2}\left[p\left(\boldsymbol{\beta}|\mathbf{g},\sigma_\beta^2\right)\right]\end{aligned}$$

MEANING: **WHAT I LEARN ABOUT MARKERS DEPENDS ON WHAT I LEARN ABOUT SIGNAL, AND ON THE CONDITIONAL PRIOR OF MARKERS GIVEN SIGNAL**

Expected value

$$
\begin{aligned}
&\Longrightarrow\; E\left(\boldsymbol{\beta}|\mathbf{y},\sigma^2_\beta\right) = E_{\mathbf{g}|\mathbf{y},\sigma^2_\beta}\left[E\left(\boldsymbol{\beta}|\mathbf{g},\sigma^2_\beta\right)\right] \\
&= E_{\mathbf{g}|\mathbf{y},\sigma^2_\beta}\left[\mathbf{X}'\sigma^2_\beta\left(\mathbf{X}\mathbf{X}'\sigma^2_\beta\right)^{-1}\mathbf{g}\right] \\
&= \mathbf{X}'\left(\mathbf{X}\mathbf{X}'\right)^{-1}E_{\mathbf{g}|\mathbf{y},\sigma^2_\beta}\left(\mathbf{g}\right) = \mathbf{X}'\left(\mathbf{X}\mathbf{X}'\right)^{-1}\widehat{\mathbf{g}} \\
\widehat{\mathbf{g}} &= GBLUP
\end{aligned}
$$

Variance

$$
\begin{aligned}
&\Longrightarrow\; Var\left(\boldsymbol{\beta}|\mathbf{y},\sigma^2_\beta\right) = Var\left(\boldsymbol{\beta}\right) - Cov\left(\boldsymbol{\beta},\mathbf{y}'\right)Var^{-1}\left(\mathbf{y}\right)Cov\left(\mathbf{y},\boldsymbol{\beta}'\right) \\
&= \mathbf{I}\sigma^2_\beta - \sigma^2_\beta\mathbf{X}'\left(\mathbf{X}\mathbf{X}'\sigma^2_\beta + \mathbf{I}\sigma^2_e\right)^{-1}\mathbf{X}\sigma^2_\beta \\
&= \left[\mathbf{I} - \mathbf{X}'\left(\mathbf{X}\mathbf{X}' + \mathbf{I}\frac{\sigma^2_e}{\sigma^2_\beta}\right)^{-1}\mathbf{X}\right]\sigma^2_\beta \\
&= \left[\mathbf{I} - \mathbf{X}'\left(\mathbf{X}\mathbf{X}'\right)^{-1}\left(\mathbf{I} + \left(\mathbf{X}\mathbf{X}'\right)^{-1}\frac{\sigma^2_e}{\sigma^2_\beta}\right)^{-1}\mathbf{X}\right]\sigma^2_\beta \\
&= \left(\mathbf{X}'\mathbf{X} + \mathbf{I}\frac{\sigma^2_e}{\sigma^2_\beta}\right)^{-1}\sigma^2_e \text{ [This part shown without proof]}
\end{aligned}
$$

MARKERS ARE NOT IDENTIFIED IN GBLUP MODEL (n<<p)

Suppose random effects model

$$
\begin{aligned}
\mathbf{y} &= \mathbf{g}+\mathbf{e} \\
\mathbf{g}_{n\times 1} &= \mathbf{X}_{n\times p}\boldsymbol{\beta}_{p\times 1} \quad n<p \\
\boldsymbol{\beta} &\sim \mathbf{N}\left(\mathbf{0},\mathbf{I}\sigma^2_\beta\right) \implies \mathbf{g} \sim \mathbf{N}\left(\mathbf{0},\mathbf{XX}'\sigma^2_\beta\right) \text{ Let } \mathbf{XX}'=\mathbf{G} \\
Cov\left(\boldsymbol{\beta},\mathbf{g}'\right) &= \sigma^2_\beta\mathbf{X}'
\end{aligned}
$$

PART 1

$$
\begin{aligned}
E(\beta|\mathrm{g}) &= Cov\left(\beta,\mathrm{g}'\right)Var^{-1}(\mathrm{g}) = \mathrm{X}'\mathrm{G}^{-1}\mathrm{g} \\
Var(\beta|\mathrm{g}) &= Var(\beta) - Cov\left(\beta,\mathrm{g}'\right)Var^{-1}(\mathrm{g})\,Cov\left(\mathrm{g},\beta'\right) \\
&= \mathrm{I}\sigma^2_\beta - \sigma^2_\beta\mathrm{X}'\left(\sigma^2_\beta\mathrm{XX}'\right)^{-1}\mathrm{X}\sigma^2_\beta \\
&= \left(\mathrm{I}-\mathrm{X}'\mathrm{G}^{-1}\mathrm{X}\right)\sigma^2_\beta
\end{aligned}
$$

$$
\begin{aligned}
E(\boldsymbol{\beta}|\mathbf{y}) &= E_{\mathbf{g}|\mathbf{y}}E(\boldsymbol{\beta}|\mathbf{g}) = \mathbf{X}'\mathbf{G}^{-1}E\left(\mathbf{g}|\mathbf{y}\right) \\
Var(\boldsymbol{\beta}|\mathbf{y}) &= Var_{\mathbf{g}|\mathbf{y}}\left[E\left(\boldsymbol{\beta}|\mathbf{g}\right)\right] + E_{\mathbf{g}|\mathbf{y}}\left[Var\left(\boldsymbol{\beta}|\mathbf{g}\right)\right] \\
&= Var_{\mathbf{g}|\mathbf{y}}\left[\mathbf{X}'\mathbf{G}^{-1}\mathbf{g}\right] + Var\left(\boldsymbol{\beta}|\mathbf{g}\right) \\
&= \mathbf{X}'\mathbf{G}^{-1}Var(\mathbf{g}|\mathbf{y})\mathbf{G}^{-1}\mathbf{X} + Var\left(\boldsymbol{\beta}|\mathbf{g}\right) \\
&\quad \textbf{IF ELEMENTS OF g CLONED AND} \\
&\quad \textbf{MEASURED INFINITE TIMES} \\
Var(\mathbf{g}|\mathbf{y}) &\rightarrow \mathbf{0} \text{ SO } Var(\boldsymbol{\beta}|\mathbf{y}) \rightarrow Var\left(\boldsymbol{\beta}|\mathbf{g}\right) = \\
\text{Means} &: \text{ we learn about } \boldsymbol{\beta} \text{ from data through } \mathbf{g}
\end{aligned}
$$

$$
\begin{aligned}
&\quad \textbf{IF ELEMENTS OF g CLONED AND} \\
&\quad \textbf{MEASURED INFINITE NUMBER OF TIMES} \\
Var(\mathbf{g}|\mathbf{y}) &\rightarrow \mathbf{0} \\
\text{SO } Var(\boldsymbol{\beta}|\mathbf{y}) &\rightarrow Var\left(\boldsymbol{\beta}|\mathbf{g}\right) \\
\text{Meaning} &: \text{ we learn about data from } \boldsymbol{\beta} \text{ through } \mathbf{g} \\
&\quad \text{up to a point (upper bound on Bayesian learning)} \\
\text{"Smallest" uncertainty: is } &\quad Var\left(\boldsymbol{\beta}|\mathbf{g}\right) = \left(\mathbf{I}-\mathbf{X}'\mathbf{G}^{-1}\mathbf{X}\right)\sigma^2_\beta
\end{aligned}
$$

Invert problem

PART 2

$$
\begin{aligned}
E(\mathbf{g}|\boldsymbol{\beta}) &= \mathbf{X}\boldsymbol{\beta} \\
Var\,(\mathbf{g}|\boldsymbol{\beta}) &= \mathbf{0} \\
E(\mathbf{g}|\mathbf{y}) &= E_{\boldsymbol{\beta}|\mathbf{y}}E(\mathbf{g}|\boldsymbol{\beta}) = \mathbf{X}E\left(\boldsymbol{\beta}|\mathbf{y}\right) \\
Var(\mathbf{g}|\mathbf{y}) &= Var_{\boldsymbol{\beta}|\mathbf{y}}\left[E(\mathbf{g}|\boldsymbol{\beta})\right] + E_{\boldsymbol{\beta}|\mathbf{y}}\left[Var\,(\mathbf{g}|\boldsymbol{\beta})\right] \\
&= Var_{\boldsymbol{\beta}|\mathbf{y}}\left[\mathbf{X}\boldsymbol{\beta}\right] \\
&= \mathbf{X}Var\left(\boldsymbol{\beta}|\mathbf{y}\right)\mathbf{X}'
\end{aligned}
$$

TWO CONCLUSIONS

1) IF $Var(\mathbf{g}|\mathbf{y}) \to \mathbf{0}$ THEN $Var(\boldsymbol{\beta}|\mathbf{y}) \to Var\,(\boldsymbol{\beta}|\mathbf{g})$

2) IF $Var(\boldsymbol{\beta}|\mathbf{y}) \to \mathbf{0}$ THEN $Var(\mathbf{g}|\mathbf{y}) \to \mathbf{0}$

$Var(\boldsymbol{\beta}|\mathbf{y}) \to \mathbf{0}$ ONLY FOR LIKELIHOOD-IDENTIFIED PARAMETERS

THE LUCKY CASE OF THE GENETIC ARCHITECT THAT HAD THE "CAUSAL" VARIANTS IN THE PANEL AND KEPT ADDING "JUNK" MARKERS

$$\begin{aligned}
\mathbf{y} &= \mathbf{X}_c\boldsymbol{\beta}_c + \mathbf{X}_p\boldsymbol{\beta}_p + \mathbf{e} \\
\mathbf{e}|\sigma_e^2 &\sim N\left(\mathbf{0}, \mathbf{I}\sigma_e^2\right) \\
\boldsymbol{\beta}_c|\sigma_c^2 &\sim N\left(\mathbf{0}, \mathbf{I}\sigma_c^2\right) \\
\boldsymbol{\beta}_p|\sigma_c^2 &\sim N\left(\mathbf{0}, \mathbf{I}\sigma_p^2\right) \\
rank\left(\mathbf{X}_c\right) &\leq \min(c, n)
\end{aligned}$$

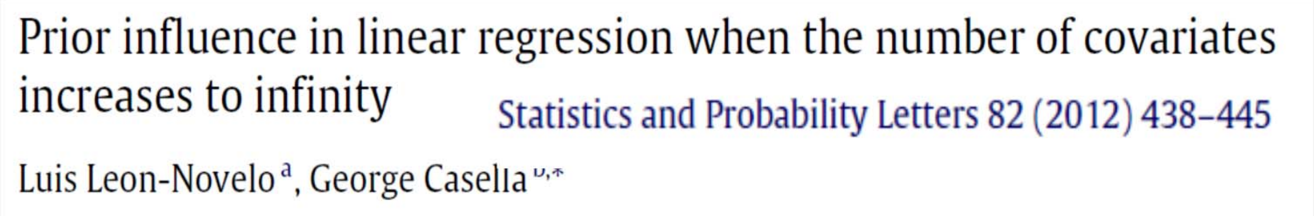
$$\sigma_p^2 \quad \rightarrow \quad 0 \text{ as the number of markers increases}$$

RECOMBINETTI Gmbh (SALE!!): infinite no. markers for $1, so our GARCHITECT bo

$$\text{Let } \mathbf{G}^{\infty} = \lim_{p\rightarrow\infty} \mathbf{G}\left(=\frac{1}{p}\mathbf{X}_p\mathbf{X}_p'\right)$$

Limiting behavior of $\frac{p\sigma_p^2}{\sigma_e^2}$ controls $\left[\boldsymbol{\beta}_c|\mathbf{y},\sigma_e^2,\sigma_c^2\right]$

$$if\ \frac{p\sigma_p^2}{\sigma_e^2} \;\rightarrow\; \infty \left\{ \begin{array}{c} E\left(\boldsymbol{\beta}_c|\mathbf{y},\sigma_e^2,\sigma_c^2\right)=\mathbf{0}=E\left(\boldsymbol{\beta}_c|\sigma_c^2\right) \\ Var\left(\boldsymbol{\beta}_c|\mathbf{y},\sigma_e^2,\sigma_c^2\right)=\mathbf{I}_c\sigma_c^2=Var\left(\boldsymbol{\beta}_c|\sigma_c^2\right) \\ \left[\boldsymbol{\beta}_c|\mathbf{y},\sigma_e^2,\sigma_c^2\right]=\left[\boldsymbol{\beta}_c|\sigma_c^2\right]=\textbf{prior distribution} \end{array} \right.$$



Prior influence in linear regression when the number of covariates increases to infinity

Statistics and Probability Letters 82 (2012) 438–445

Luis Leon-Novelo [a], George Casella [b,*]

TAKE HOME MESSAGE

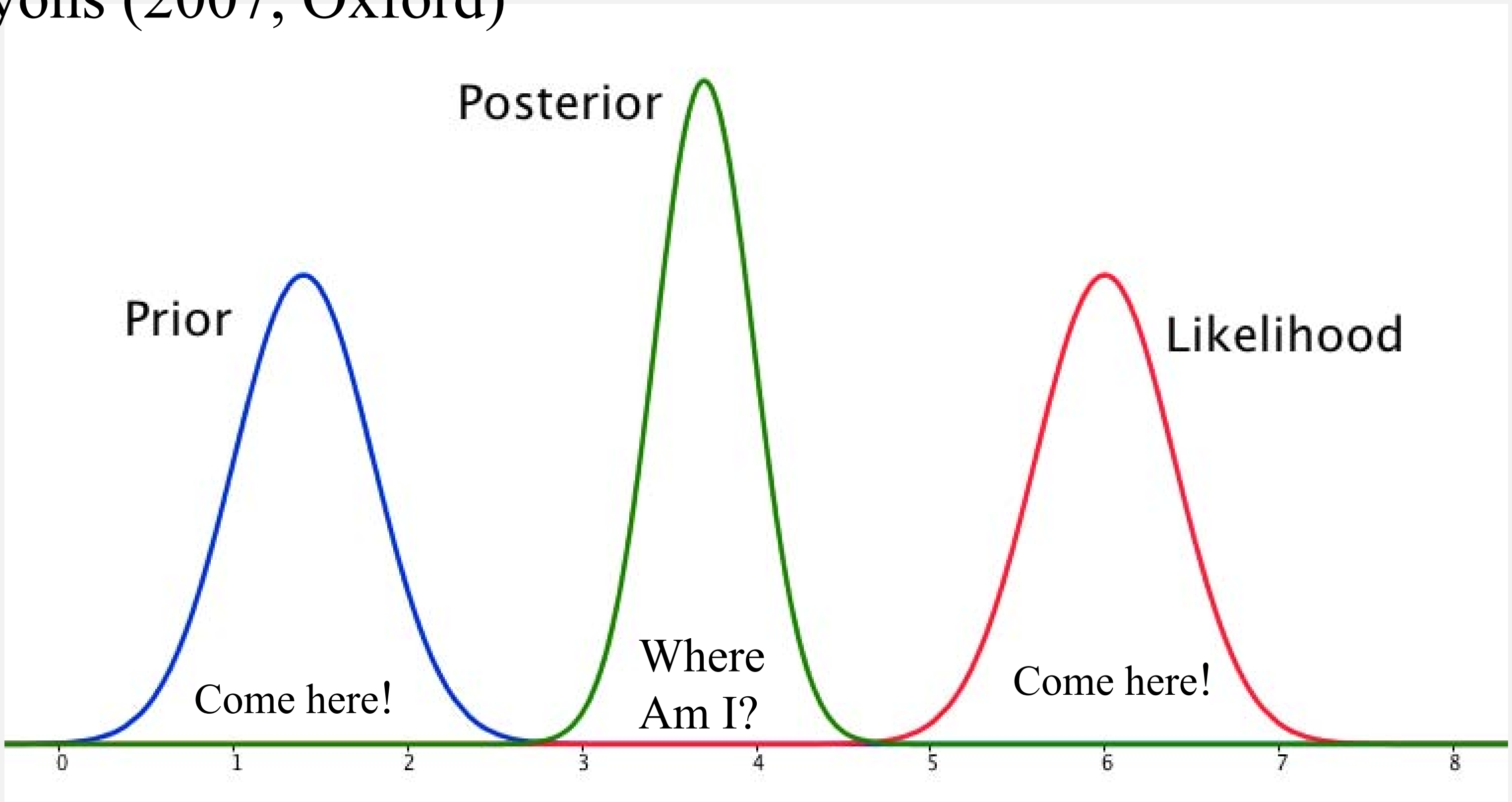
IN THE n<<p SETTING,

THE MORE EXTRAVAGANT THE PRIOR IS,

THE MORE LIKELY YOUR INFERENCES WILL GET SCREWED UP.

DIGRESSION ON BAYESIAN LEARNING

“Bayesians address the question everyone is interested in,
by using assumptions no-one believes”
Lyons (2007, Oxford)



Quest for “good priors ? **“Objective” Bayesians search for bad priors**

Want to learn, i.e., get away from prior

IMPORTANT: MEASURING EXTENT OF BAYESIAN LEARN

(only one serious attempt in QG literature)

Lehermeier et al. (2013, SAGMB)

BR: Bayes ridge
BL: Bayesian Lasso
BA: Bayes A
BB: Bayes B

HELLINGER DISTANCE BETWEEN PRIOR AND POSTERIOR

IF H close to 0 prior is influential

$$H = \sqrt{1 - \int \sqrt{p(\theta|y)\,p(\theta)}d\theta}$$

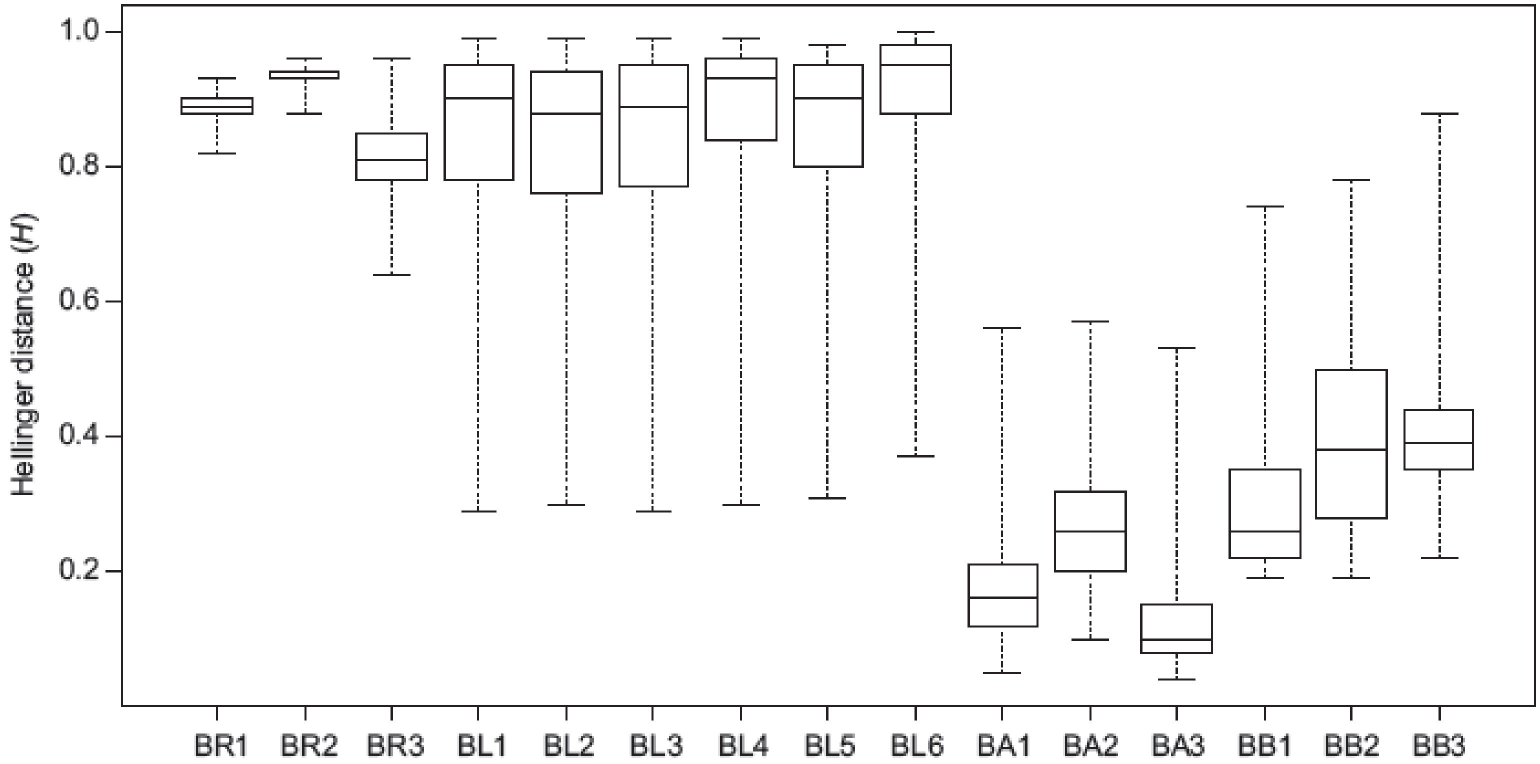
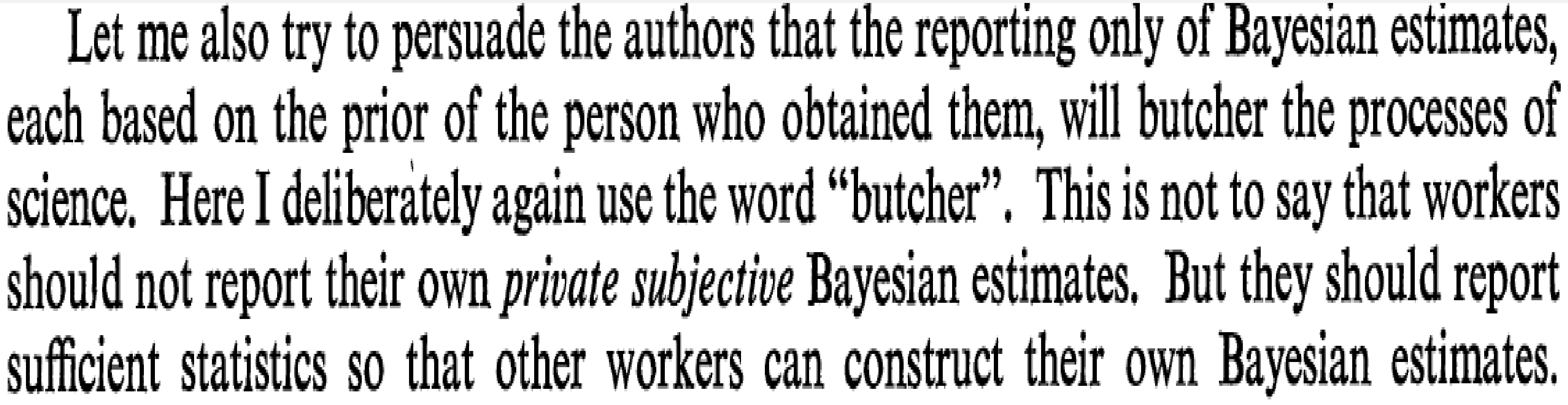


Figure 2 Distribution of Hellinger distance (H) between the marginal prior and posterior densities of marker effects $\boldsymbol{\beta}$ from different model scenarios, calculated with simulated dataset maizeA. Each boxplot displays the distribution of Hellinger distances of the 1117 marker effects out of each model.

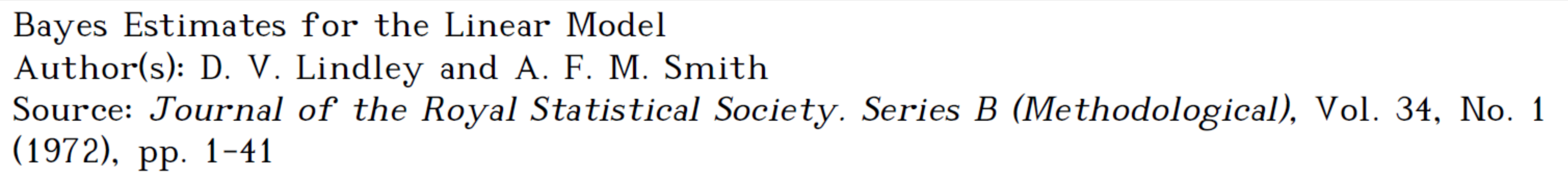
WHAT TO DO IF A REVIEWER OF A PAPER WROTE …



Let me also try to persuade the authors that the reporting only of Bayesian estimates, each based on the prior of the person who obtained them, will butcher the processes of science. Here I deliberately again use the word "butcher". This is not to say that workers should not report their own *private subjective* Bayesian estimates. But they should report sufficient statistics so that other workers can construct their own Bayesian estimates.

RECOMMENDED ACTION: DO NOT PLACE DIRTY LAUNDRY OR SMOKING GUN IN SUPPLEMENTAL FILES

WRITTEN BY: Oscar Kempthorne (1972), discussing:



Bayes Estimates for the Linear Model
Author(s): D. V. Lindley and A. F. M. Smith
Source: *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 34, No. 1 (1972), pp. 1–41

OK with Indira Gandhi