

# MCMC: Gibbs Sampler

# Building Complex Models: Conditional Sampling

- Consider a more complex model

$$p(b, \alpha, \beta | y) \propto p(y | b) p(b | \alpha, \beta) p(\alpha) p(\beta)$$

- When sampling for each parameter iteratively, only need to consider distributions that have that parameter

$$p(b | \alpha, \beta, y) \propto p(y | b) p(b | \alpha, \beta)$$

$$p(\alpha | b, \beta, y) \propto p(b | \alpha, \beta) p(\alpha)$$

$$p(\beta | b, \alpha, y) \propto p(b | \alpha, \beta) p(\beta)$$

# In one MCMC step

$$p(b^{(g+1)} | \alpha^{(g)}, \beta^{(g)}, y) \propto p(y | b^{(g)}) p(b^{(g)} | \alpha^{(g)}, \beta^{(g)})$$

$$p(\alpha^{(g+1)} | b^{(g+1)}, \beta^{(g)}, y) \propto p(b^{(g+1)} | \alpha^{(g)}, \beta^{(g)}) p(\alpha^{(g)})$$

$$p(\beta^{(g+1)} | b^{(g+1)}, \alpha^{(g+1)}, y) \propto p(b^{(g+1)} | \alpha^{(g+1)}, \beta^{(g)}) p(\beta^{(g)})$$

# Gibbs Sampling

- If we can solve a conditional distribution analytically
  - We can sample from that conditional directly
  - Even if we can't solve for full joint distribution analytically
- Example

$$p(\mu, \sigma^2 | \vec{y}) \propto N(\vec{y} | \mu, \sigma^2) N(\mu | \mu_0, V_0) IG(\sigma^2 | \alpha, \beta)$$

$$p(\mu^{(g+1)} | \sigma^{2(g)}, \vec{y}) \propto N(\vec{y} | \mu^{(g)}, \sigma^{2(g)}) N(\mu^{(g)} | \mu_0, V_0)$$

$$= N \left( \frac{\left( \frac{n}{\sigma^{2(g)}} \bar{y} + \frac{\mu_0}{\tau^2} \right)}{\left( \frac{n}{\sigma^{2(g)}} + \frac{1}{\tau^2} \right)}, \frac{1}{\left( \frac{1}{\sigma^{2(g)}} + \frac{1}{\tau^2} \right)} \right)$$

$$p(\sigma^{2(g+1)} | \mu^{(g+1)}, \vec{y}) \propto N(\vec{y} | \mu^{(g+1)}, \sigma^{2(g)}) IG(\sigma^{2(g)} | \alpha, \beta)$$

$$= IG \left( \alpha + \frac{n}{2}, \beta + \frac{1}{2} \sum (y_i - \mu^{(g+1)})^2 \right)$$

# Trade-offs of Gibbs Sampling

- 100% acceptance rate
- Quicker mixing, quicker convergence
- No need to specify a Jump distribution
- No need to tune a Jump variance
- Requires that we know the analytical solution for the conditional
  - Conjugate distributions
- Can mix different samplers within a MCMC

# Bayesian Regression via Gibbs Sampling

- Consider the standard regression model

$$y_i = b_0 + b_1 x_{i,1} + b_2 x_{i,2} + \dots + b_n x_{i,n} + \epsilon_i$$

$$\epsilon_i \sim N(0, \sigma^2)$$

- $Y$  = response variable
- $X_j$  = covariates
- $\eta_j$  = parameters

# Matrix version

- As noted earlier this

$$y_i = b_0 + b_1 x_{i,1} + b_2 x_{i,2} + \cdots + b_n x_{i,n}$$

can be expressed as

$$y_i = X_i \vec{b}$$

- The full data set can be expressed as

$$\vec{y} = X \vec{b}$$



# Quick review of matrix math

- Vector

$$\vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

3 x 1

- Transpose

$$\vec{b}^T = [b_1 \quad b_2 \quad b_3]$$

1 x 3

- Vector x vector

$$\vec{b}^T \times \vec{b} = [b_1 \quad b_2 \quad b_3] \times \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = b_1^2 + b_2^2 + b_3^2$$

1 x 3                      3 x 1                      1 x 1

See Appendix C for more details

# Quick review of matrix math

- Vector x vector

$$\vec{b} \times \vec{b}^T = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \times [b_1 \quad b_2 \quad b_3] = \begin{bmatrix} b_1^2 & b_1 b_2 & b_1 b_3 \\ b_2 b_1 & b_2^2 & b_2 b_3 \\ b_3 b_1 & b_3 b_2 & b_3^2 \end{bmatrix}$$

$3 \times 1$                        $1 \times 3$                                        $3 \times 3$

- Vector x matrix

$$X \vec{b} = \vec{\mu}$$

$n \times 3$     $3 \times 1$     $n \times 1$

- Matrix x Matrix

$$X X^T = Z$$

$n \times 3$     $3 \times n$     $n \times n$

- Matrix x Matrix

$$X^T X = \begin{bmatrix} \sum x_{i,1}^2 & \sum x_{i,1}x_{i,2} & \sum x_{i,1}x_{i,3} \\ \sum x_{i,2}x_{i,1} & \sum x_{i,2}^2 & \sum x_{i,2}x_{i,3} \\ \sum x_{i,3}x_{i,1} & \sum x_{i,3}x_{i,2} & \sum x_{i,3}^2 \end{bmatrix}$$

3 x n   n x 3

3 x 3

- Identity Matrix

$$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- Matrix Inversion

$$X^{-1} X = X X^{-1} = I$$

# Transpose Relationships

$$(A^T)^T = A$$

$$(A + B)^T = A^T + B^T$$

$$(AB)^T = B^T A^T$$

# ...back to regression

$$\vec{y} = \mathbf{X} \vec{b} + \epsilon$$

Process model

$$\vec{\epsilon} \sim N_n(\mathbf{0}, \mathbf{\Sigma})$$

Data model

$$\vec{\epsilon} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I})$$

## Likelihood

$$p(\vec{y} \mid \vec{b}, \sigma^2, \mathbf{X}) = N_n(\vec{y} \mid \mathbf{X} \vec{b}, \sigma^2 \mathbf{I})$$

## Priors

$$p(\vec{b}) = N_p(\vec{b} \mid \vec{b}_0, \mathbf{V}_b)$$

Parameter model

$$p(\sigma^2) = IG(\sigma^2 \mid s_1, s_2)$$

# General pattern

- Write down likelihood
  - Data model
  - Process model
- Write down priors for the free parameters in the likelihood
  - Parameter model
- Solve for posterior distribution of model parameters
  - Analytical
  - Numerical



## Posterior

$$p(\vec{b}, \sigma^2 | \mathbf{X}, \vec{y}) \propto N_n(\vec{y} | \mathbf{X}\vec{b}, \sigma^2 \mathbf{I}) \times \\ N_p(\vec{b} | \vec{b}_0, \mathbf{V}_b) IG(\sigma^2 | s_1, s_2)$$

## Sample in terms of conditionals

$$p(\vec{b} | \sigma^2, \mathbf{X}, \vec{y}) \propto N_n(\vec{y} | \mathbf{X}\vec{b}, \sigma^2 \mathbf{I}) N_p(\vec{b} | \vec{b}_0, \mathbf{V}_b)$$

$$p(\sigma^2 | \vec{b}, \mathbf{X}, \vec{y}) \propto N_n(\vec{y} | \mathbf{X}\vec{b}, \sigma^2 \mathbf{I}) IG(\sigma^2 | s_1, s_2)$$

# Regression Parameters

$$p(\vec{b} \mid \sigma^2, \mathbf{X}, \vec{y}) \propto N_n(\vec{y} \mid \mathbf{X}\vec{b}, \sigma^2 \mathbf{I}) N_p(\vec{b} \mid \vec{b}_0, \mathbf{V}_b)$$

Will have a posterior that is multivariate Normal...

$$N_p(\vec{b} \mid \mathbf{V}\mathbf{v}, \mathbf{V}) = \frac{1}{(2\pi)^{p/2} |\mathbf{V}|^{1/2}} \exp\left[-\frac{1}{2} (\vec{b} - \mathbf{V}\mathbf{v})^T \mathbf{V}^{-1} (\vec{b} - \mathbf{V}\mathbf{v})\right]$$

What are  $\mathbf{V}$  and  $\mathbf{v}$ ??



$$(\vec{b} - V\mathbf{v})^T V^{-1} (\vec{b} - V\mathbf{v})$$

$$\vec{b}^T V^{-1} \vec{b} - \vec{b}^T V^{-1} V \mathbf{v} - \mathbf{v}^T V V^{-1} \vec{b} + \mathbf{v}^T V V^{-1} V \mathbf{v}$$

$$\vec{b}^T V^{-1} \vec{b} - \vec{b}^T \mathbf{v} - \mathbf{v}^T \vec{b} + \mathbf{v}^T V \mathbf{v}$$

$$\boxed{\vec{b}^T V^{-1} \vec{b}} - \boxed{2 \vec{b}^T \mathbf{v}} + \mathbf{v}^T V \mathbf{v}$$

$$(\vec{y} - Xb)^T (\sigma^2 I)^{-1} (\vec{y} - Xb) + (\vec{b} - b_0)^T V_b^{-1} (\vec{b} - b_0)$$

$$\begin{aligned} & \sigma^{-2} \vec{y}^T \vec{y} - \sigma^{-2} \vec{y}^T Xb - \sigma^{-2} b^T X^T \vec{y} + \sigma^{-2} b^T X^T Xb \\ & + \vec{b}^T V_b^{-1} \vec{b} - \vec{b}^T V_b^{-1} \vec{b}_0 - b_0^T V_b^{-1} \vec{b} + b_0^T V_b^{-1} b_0 \end{aligned}$$

$$\vec{b}^T V_b^{-1} \vec{b}$$

$$\sigma^{-2} b^T X^T Xb + \vec{b}^T V_b^{-1} \vec{b} \longrightarrow V^{-1} = \sigma^{-2} X^T X + V_b^{-1}$$

$$-2\vec{b}^T v$$

$$-\sigma^{-2} \vec{y}^T Xb - \sigma^{-2} b^T X^T \vec{y} - \vec{b}^T V_b^{-1} \vec{b}_0 - b_0^T V_b^{-1} \vec{b}$$

$$\longrightarrow v = \sigma^{-2} X^T \vec{y} + V_b^{-1} \vec{b}_0$$

## Variance

$$p(\sigma^2 | \vec{b}, \mathbf{X}, \vec{y}) \propto N_n(\vec{y} | \mathbf{X}\vec{b}, \sigma^2 \mathbf{I}) IG(\sigma^2 | s_1, s_2)$$

$$\propto \frac{1}{\sigma^{n/2}} \exp\left[\frac{1}{2\sigma^2} (\vec{y} - \mathbf{X}\vec{b})^T (\vec{y} - \mathbf{X}\vec{b})\right] (\sigma^2)^{-(s_1+1)} \exp\left[-\frac{s_2}{\sigma^2}\right]$$

Will have a posterior that is Inverse Gamma...

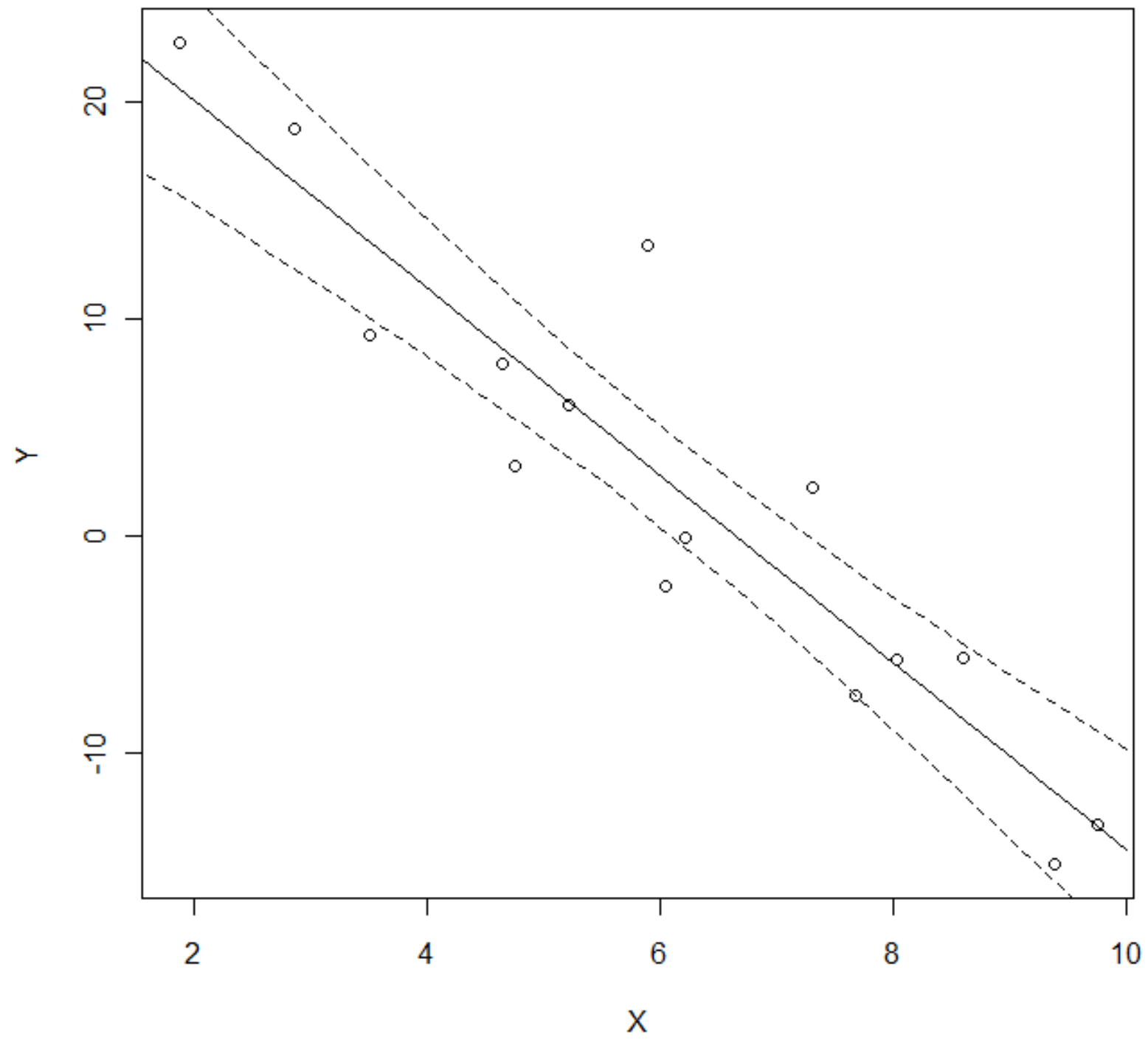
$$IG(\sigma^2 | u_1, u_2) \propto (\sigma^2)^{-(u_1+1)} \exp\left[-\frac{u_2}{\sigma^2}\right]$$

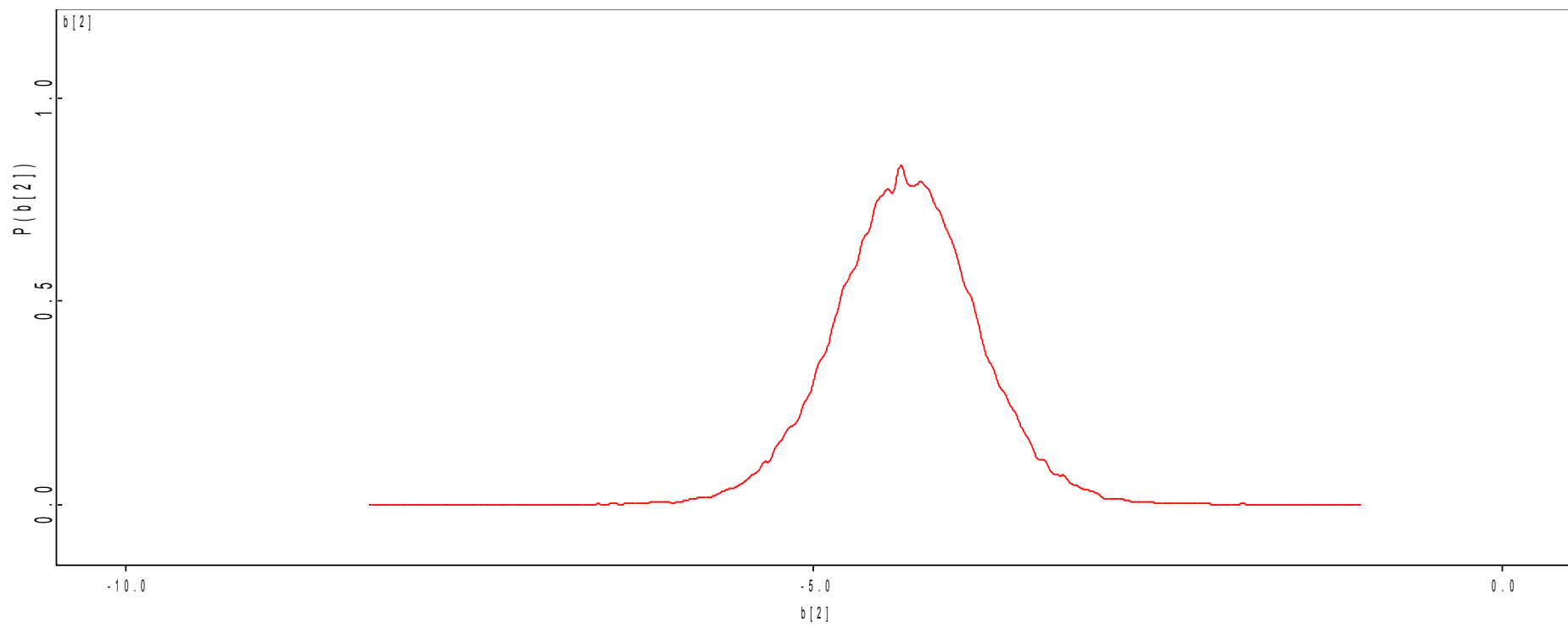
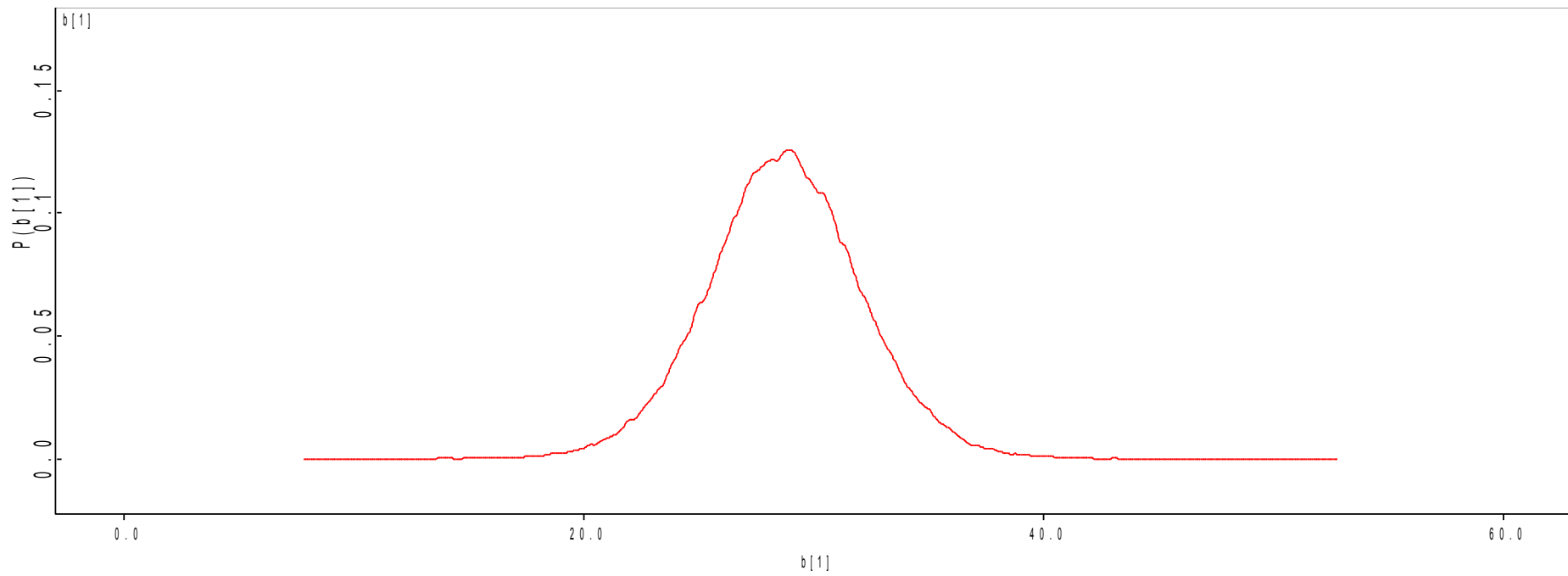
$$u_1 = s_1 + \frac{n}{2}$$

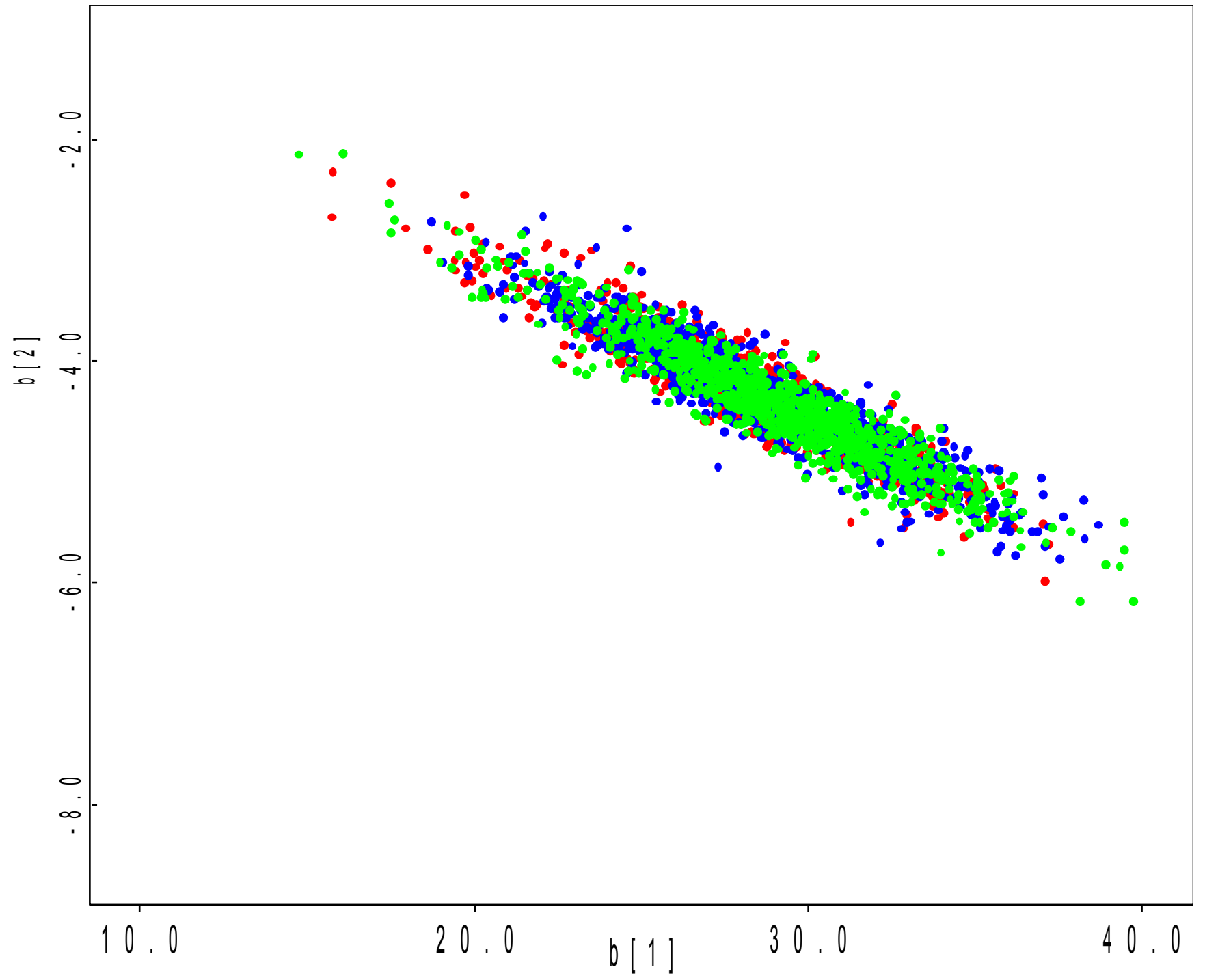
$$u_2 = s_2 + \frac{1}{2} (\vec{y} - \mathbf{X}\vec{b})^T (\vec{y} - \mathbf{X}\vec{b})$$

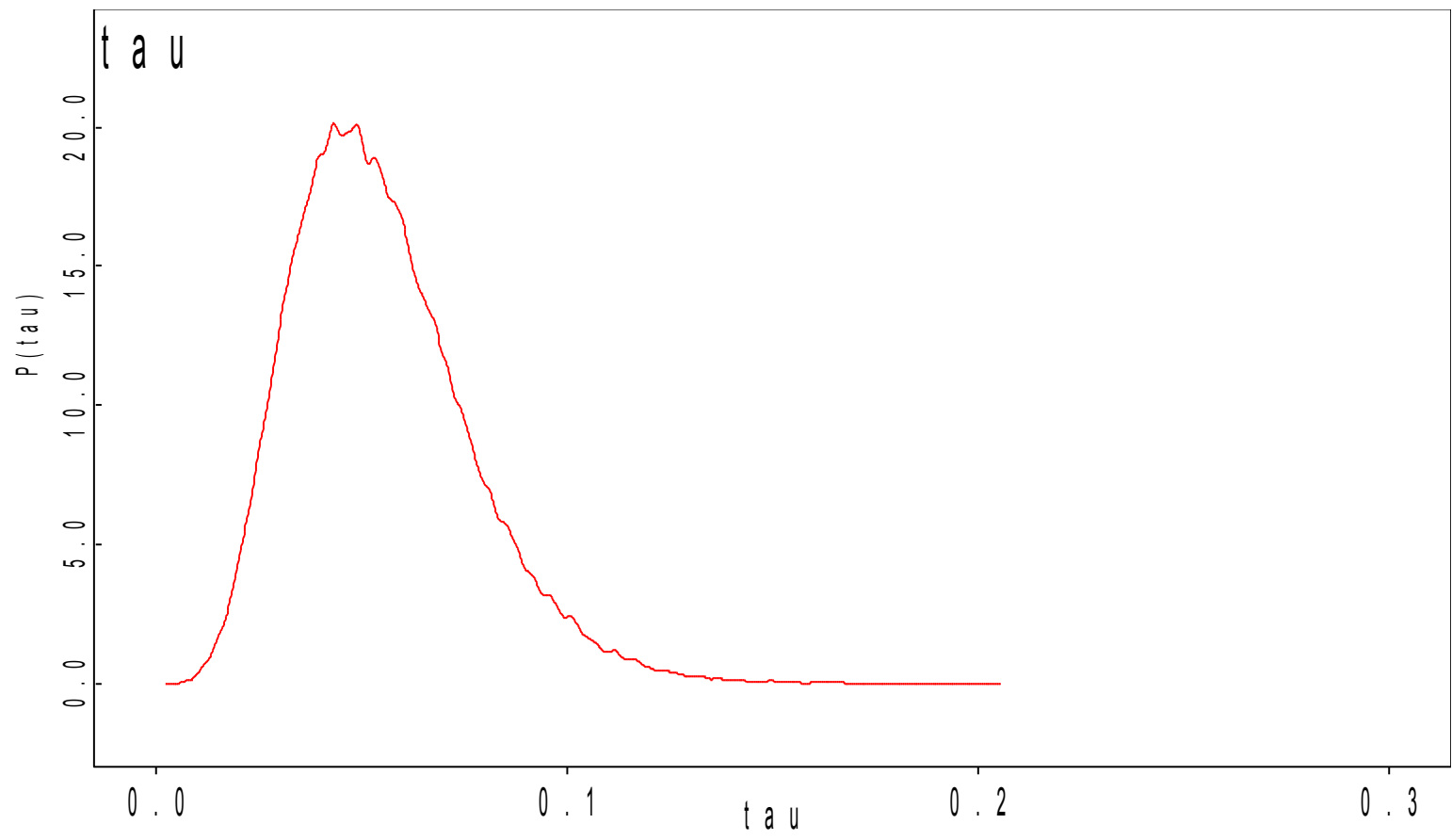
# BUGS implementation

```
model{  
  
##prior distributions  
  b[1] ~ dnorm(0,1.0E-5)  
  b[2] ~ dnorm(0,1.0E-5)  
  tau ~ dgamma(0.1,0.001)  
  
##likelihood  
  for(i in 1:n){  
    mu[i] <- b[1] + x[i] * b[2]  
    y[i] ~ dnorm(mu[i],tau)  
  }  
  
}
```

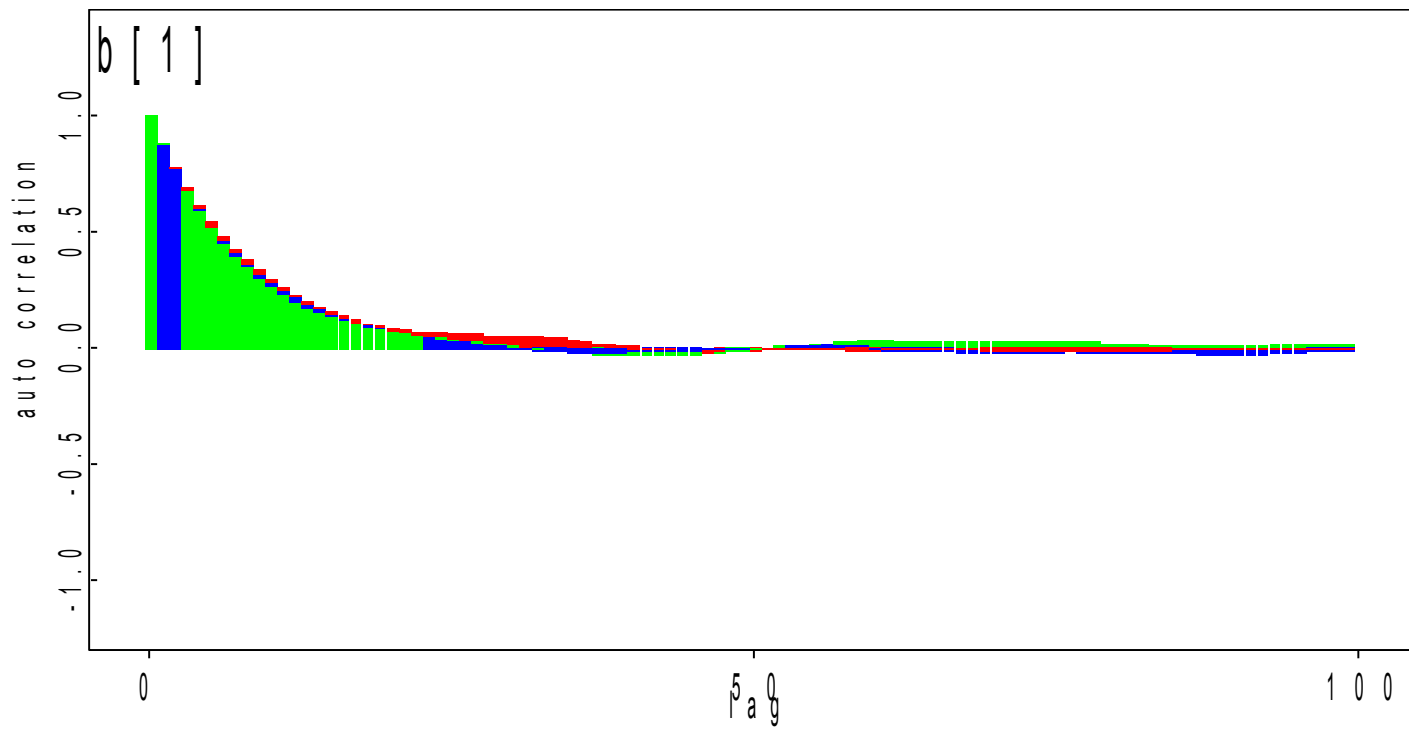












# Centering the data

```
model{  
  
  ##prior distributions  
  b[1] ~ dnorm(0,1.0E-5)  
  b[2] ~ dnorm(0,1.0E-5)  
  tau ~ dgamma(0.1,0.001)  
  mX <- mean(x[])  
  
  ##likelihood  
  for(i in 1:n){  
    mu[i] <- b[1] + (x[i]-mX) * b[2]  
    y[i] ~ dnorm(mu[i],tau)  
  }  
  
}
```

