

Bayesian Estimation in Linear Models & Choice of Prior Distributions

STA521 Predictive Models Duke University

Merlise Clyde

February 28, 2017

Bayesian Estimation

Model $\mathbf{Y} = \mathbf{X}\beta + \epsilon$ with $\epsilon \sim N(\mathbf{0}_n, \sigma^2 \mathbf{I}_n)$ is equivalent to

$$\mathbf{Y} \sim N(\mathbf{X}\beta, \mathbf{I}_n/\phi)$$

$\phi = 1/\sigma^2$ is the *precision*.

In the Bayesian paradigm describe uncertainty about unknown parameters using probability distributions

- ▶ Prior Distribution $p(\beta, \phi)$ describes uncertainty about parameters prior to seeing the data
- ▶ Posterior Distribution $p(\beta, \phi \mid \mathbf{Y})$ describes uncertainty about the parameters after updating beliefs given the observed data
- ▶ updating rule is based on Bayes Theorem

$$p(\beta, \phi \mid \mathbf{Y}) \propto \mathcal{L}(\beta, \phi)p(\beta, \phi)$$

reweight prior beliefs by likelihood of parameters under observed data

Prior Distributions

Factor joint prior distribution

$$p(\boldsymbol{\beta}, \phi) = p(\boldsymbol{\beta} \mid \phi)p(\phi)$$

Convenient choice is to take “Conjugate” family

- ▶ $\boldsymbol{\beta} \mid \phi \sim \mathbf{N}(\mathbf{b}_0, \Phi_0^{-1}/\phi)$ where \mathbf{b}_0 is the prior mean and Φ_0^{-1}/ϕ is the prior covariance of $\boldsymbol{\beta}$

$$p(\boldsymbol{\beta} \mid \phi) = \frac{1}{(2\pi)^{p/2}} |\phi \Phi_0|^{1/2} e^{-\frac{1}{2}\phi(\boldsymbol{\beta} - \mathbf{b}_0)^T \Phi_0(\boldsymbol{\beta} - \mathbf{b}_0)}$$

- ▶ $\phi \sim \mathbf{G}(\nu_0/2, SS_0/2)$ with $E(\sigma^2) = SS_0/(\nu_0 - 2)$

$$p(\phi) = \frac{1}{\Gamma(\nu_0/2)} \left(\frac{SS_0}{2} \right)^{\nu_0/2} \phi^{\nu_0/2-1} e^{-\phi SS_0/2}$$

- ▶ $(\boldsymbol{\beta}, \phi)^T \sim \mathbf{NG}(\mathbf{b}_0, \Phi_0, \nu_0, SS_0)$
- ▶ Conjugate “Normal-Gamma” family implies

$$(\boldsymbol{\beta}, \phi)^T \mid \mathbf{Y} \sim \mathbf{NG}(\mathbf{b}_n, \Phi_n, \nu_n, SS_n)$$

Finding the Posterior Distribution

Express Likelihood: $\mathcal{L}(\beta, \phi) \propto \phi^{n/2} e^{-\phi \frac{\text{SSE}}{2}} e^{-\frac{\phi}{2} (\beta - \hat{\beta})^T (\mathbf{X}^T \mathbf{X}) (\beta - \hat{\beta})}$

$$p(\beta, \phi \mid \mathbf{Y}) \propto \phi^{\frac{n+p+\nu_0}{2}-1} e^{-\frac{\phi}{2} (\text{SSE} + \text{SS}_0)} \\ e^{-\frac{\phi}{2} (\beta - \hat{\beta})^T (\mathbf{X}^T \mathbf{X}) (\beta - \hat{\beta})} e^{-\frac{\phi}{2} (\beta - \mathbf{b}_0)^T \Phi (\beta - \mathbf{b}_0)}$$

Finding the Posterior Distribution

$$p(\boldsymbol{\beta}, \phi \mid \mathbf{Y}) \propto \phi^{\frac{n+p+\nu_0}{2}-1} e^{-\frac{\phi}{2}(\text{SSE}+\text{SS}_0)} \\ e^{-\frac{\phi}{2}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})^T(\mathbf{X}^T\mathbf{X})(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})} e^{-\frac{\phi}{2}(\boldsymbol{\beta}-\mathbf{b}_0)^T\boldsymbol{\Phi}(\boldsymbol{\beta}-\mathbf{b}_0)}$$

Quadratic in Normal

$$\exp \left\{ -\frac{\phi}{2}(\boldsymbol{\beta} - \mathbf{b})^T \boldsymbol{\Phi}(\boldsymbol{\beta} - \mathbf{b}) \right\} = \exp \left\{ -\frac{\phi}{2}(\boldsymbol{\beta}^T \boldsymbol{\Phi} \boldsymbol{\beta} - 2\boldsymbol{\beta}^T \boldsymbol{\Phi} \mathbf{b} + \mathbf{b}^T \boldsymbol{\Phi} \mathbf{b}) \right\}$$

- ▶ Expand quadratics and regroup terms
- ▶ Read off posterior precision from Quadratic in $\boldsymbol{\beta}$
- ▶ Read off posterior mean from Linear term in $\boldsymbol{\beta}$
- ▶ will need to complete the quadratic in the posterior mean

Expand and Regroup

Quadratic in Normal

$$\exp \left\{ -\frac{\phi}{2}(\boldsymbol{\beta} - \mathbf{b})^T \boldsymbol{\Phi}(\boldsymbol{\beta} - \mathbf{b}) \right\} = \exp \left\{ -\frac{\phi}{2}(\boldsymbol{\beta}^T \boldsymbol{\Phi} \boldsymbol{\beta} - 2\boldsymbol{\beta}^T \boldsymbol{\Phi} \mathbf{b} + \mathbf{b}^T \boldsymbol{\Phi} \mathbf{b}) \right\}$$

$$\begin{aligned} p(\boldsymbol{\beta}, \phi \mid \mathbf{Y}) &\propto \phi^{\frac{n+p+\nu_0}{2}-1} e^{-\frac{\phi}{2}(\text{SSE} + \text{SS}_0)} \\ &\quad e^{-\frac{\phi}{2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T (\mathbf{X}^T \mathbf{X})(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})} e^{-\frac{\phi}{2}(\boldsymbol{\beta} - \mathbf{b}_0)^T \boldsymbol{\Phi}_0(\boldsymbol{\beta} - \mathbf{b}_0)} \\ &= \phi^{\frac{n+p+\nu_0}{2}-1} e^{-\frac{\phi}{2}(\text{SSE} + \text{SS}_0)} \\ &\quad e^{-\frac{\phi}{2}(\boldsymbol{\beta}^T (\mathbf{X}^T \mathbf{X} + \boldsymbol{\Phi}_0) \boldsymbol{\beta})} \\ &\quad e^{-\frac{\phi}{2}(-2\boldsymbol{\beta}^T (\mathbf{X}^T \mathbf{X} \hat{\boldsymbol{\beta}} + \boldsymbol{\Phi}_0 \mathbf{b}_0))} \\ &\quad e^{-\frac{\phi}{2}(\hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{X} \hat{\boldsymbol{\beta}} + \mathbf{b}_0^T \boldsymbol{\Phi}_0 \mathbf{b}_0)} \end{aligned}$$

Identify Hyperparameters and Complete the Quadratic

Quadratic in Normal

$$\exp \left\{ -\frac{\phi}{2} (\boldsymbol{\beta} - \mathbf{b})^T \Phi (\boldsymbol{\beta} - \mathbf{b}) \right\} = \exp \left\{ -\frac{\phi}{2} (\boldsymbol{\beta}^T \Phi \boldsymbol{\beta} - 2\boldsymbol{\beta}^T \Phi \mathbf{b} + \mathbf{b}^T \Phi \mathbf{b}) \right\}$$

Let $\Phi_n = \mathbf{X}^T \mathbf{X} + \Phi_0$

$$\begin{aligned} p(\boldsymbol{\beta}, \phi \mid \mathbf{Y}) &\propto \phi^{\frac{n+p+\nu_0}{2}-1} e^{-\frac{\phi}{2}(\text{SSE}+\text{SS}_0)} \\ &\quad e^{-\frac{\phi}{2}(\boldsymbol{\beta}^T (\mathbf{X}^T \mathbf{X} + \Phi_0) \boldsymbol{\beta})} \\ &\quad e^{-\frac{\phi}{2}(-2\boldsymbol{\beta}^T \Phi_n \Phi_n^{-1} (\mathbf{X}^T \mathbf{X} \hat{\boldsymbol{\beta}} + \Phi_0 \mathbf{b}_0))} \\ &\quad e^{-\frac{\phi}{2}(\mathbf{b}_n^T \Phi_n \mathbf{b}_n - \mathbf{b}_n^T \Phi_0 \mathbf{b}_n)} \\ &\quad e^{-\frac{\phi}{2}(\hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{X} \hat{\boldsymbol{\beta}} + \mathbf{b}_0^T \Phi_0 \mathbf{b}_0)} \\ &= \phi^{\frac{n+p+\nu_0}{2}-1} e^{-\frac{\phi}{2}(\text{SSE}+\text{SS}_0 + \hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{X} \hat{\boldsymbol{\beta}} + \mathbf{b}_0^T \Phi_0 \mathbf{b}_0 - \mathbf{b}_n^T \Phi_n \mathbf{b}_n)} \\ &\quad e^{-\frac{\phi}{2}(\boldsymbol{\beta}^T (\Phi_n) \boldsymbol{\beta})} \\ &\quad e^{-\frac{\phi}{2}(-2\boldsymbol{\beta}^T \Phi_n \Phi_n^{-1} (\mathbf{X}^T \mathbf{X} \hat{\boldsymbol{\beta}} + \Phi_0 \mathbf{b}_0))} \\ &\quad e^{-\frac{\phi}{2}(\mathbf{b}_n^T \Phi_n \mathbf{b}_n)} \end{aligned}$$

Posterior Distribution

$$p(\boldsymbol{\beta}, \phi \mid \mathbf{Y}) \propto \phi^{\frac{n+\nu_0}{2}-1} e^{-\frac{\phi}{2}(\text{SSE} + \text{SS}_0 + \hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{X} \hat{\boldsymbol{\beta}} + \mathbf{b}_0^T \Phi_0 \mathbf{b}_0 - \mathbf{b}_n^T \Phi_n \mathbf{b}_n)} \\ \phi^{\frac{p}{2}} e^{-\frac{\phi}{2}(\boldsymbol{\beta} - \mathbf{b}_n)^T \Phi_n (\boldsymbol{\beta} - \mathbf{b}_n)}$$

$$\Phi_n = \mathbf{X}^T \mathbf{X} + \Phi_0$$

$$\mathbf{b}_n = \Phi_n^{-1}(\mathbf{X}^T \mathbf{X} \hat{\boldsymbol{\beta}} + \Phi_0 \mathbf{b}_0)$$

Posterior Distribution

$$\boldsymbol{\beta} \mid \phi, \mathbf{Y} \sim \mathbf{N}(\mathbf{b}_n, (\phi \Phi_n)^{-1})$$

$$\phi \mid \mathbf{Y} \sim \mathbf{G}\left(\frac{n + \nu_0}{2}, \frac{\text{SSE} + \text{SS}_0 + \hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{X} \hat{\boldsymbol{\beta}} + \mathbf{b}_0^T \Phi_0 \mathbf{b}_0 - \mathbf{b}_n^T \Phi_n \mathbf{b}_n}{2}\right)$$

Predictive Distribution

Suppose $\mathbf{Y}^* \mid \boldsymbol{\beta}, \phi \sim \mathcal{N}(\mathbf{X}^* \boldsymbol{\beta}, \mathbf{I}/\phi)$ and is conditionally independent of \mathbf{Y} given $\boldsymbol{\beta}$ and ϕ

What is the predictive distribution of $\mathbf{Y}^* \mid \mathbf{Y}$?

$\mathbf{Y}^* = \mathbf{X}^* \boldsymbol{\beta} + \boldsymbol{\epsilon}^*$ and $\boldsymbol{\epsilon}^*$ is independent of \mathbf{Y} given ϕ

$$\mathbf{X}^* \boldsymbol{\beta} + \boldsymbol{\epsilon}^* \mid \phi, \mathbf{Y} \sim \mathcal{N}(\mathbf{X}^* \mathbf{b}_n, (\mathbf{X}^* \boldsymbol{\Phi}_n^{-1} \mathbf{X}^{*T} + \mathbf{I})/\phi)$$

$$\mathbf{Y}^* \mid \phi, \mathbf{Y} \sim \mathcal{N}(\mathbf{X}^* \mathbf{b}_n, (\mathbf{X}^* \boldsymbol{\Phi}_n^{-1} \mathbf{X}^{*T} + \mathbf{I})/\phi)$$

$$\phi \mid \mathbf{Y} \sim \mathbf{G}\left(\frac{\nu_n}{2}, \frac{\hat{\sigma}^2 \nu_n}{2}\right)$$

$$\mathbf{Y}^* \mid \mathbf{Y} \sim t_{\nu_n}(\mathbf{X}^* \mathbf{b}_n, \hat{\sigma}^2 (\mathbf{I} + \mathbf{X}^* \boldsymbol{\Phi}_n^{-1} \mathbf{X}^T))$$

Alternative Derivation

Conditional Distribution:

$$\begin{aligned} f(\mathbf{Y}^* | \mathbf{Y}) &= \frac{f(\mathbf{Y}^*, \mathbf{Y})}{f(\mathbf{Y})} \\ &= \frac{\iint f(\mathbf{Y}^*, \mathbf{Y} | \beta, \phi) p(\beta, \phi) d\beta d\phi}{f(\mathbf{Y})} \\ &= \frac{\iint f(\mathbf{Y}^* | \beta, \phi) f(\mathbf{Y} | \beta, \phi) p(\beta, \phi) d\beta d\phi}{f(\mathbf{Y})} \\ &= \iint f(\mathbf{Y}^* | \beta, \phi) p(\beta, \phi | \mathbf{Y}) d\beta d\phi \end{aligned}$$

Conjugate Priors

Definition

A class of prior distributions \mathcal{P} for θ is conjugate for a sampling model $p(y \mid \theta)$ if for every $p(\theta) \in \mathcal{P}$, $p(\theta \mid \mathbf{Y}) \in \mathcal{P}$.

Advantages:

- ▶ Closed form distributions for most quantities; bypass Monte Carlo for calculations
- ▶ Simple updating in terms of sufficient statistics “weighted average” - useful with big data
- ▶ Interpretation as prior samples - prior sample size
- ▶ Elicitation of prior through imaginary or historical data
- ▶ limiting “non-proper” form recovers MLEs

Choice of conjugate prior?

Unit Information Prior

Unit information prior $\beta \mid \phi \sim N(\hat{\beta}, n(\mathbf{X}^T \mathbf{X})^{-1} / \phi)$

- ▶ Fisher Information is $\phi \mathbf{X}^T \mathbf{X}$ based on a sample of n observations
- ▶ Inverse Fisher information is covariance matrix of MLE
- ▶ “average information” in one observation is $\phi \mathbf{X}^T \mathbf{X} / n$
- ▶ center prior at MLE and base covariance on the information in “1” observation
- ▶ Posterior mean

$$\frac{n}{1+n} \hat{\beta} + \frac{1}{1+n} \hat{\beta} = \hat{\beta}$$

- ▶ Posterior Distribution

$$\beta \mid \mathbf{Y}, \phi \sim N \left(\hat{\beta}, \frac{n}{1+n} (\mathbf{X}^T \mathbf{X})^{-1} \phi^{-1} \right)$$

Cannot represent real prior beliefs; double use of data

Zellner's g -prior

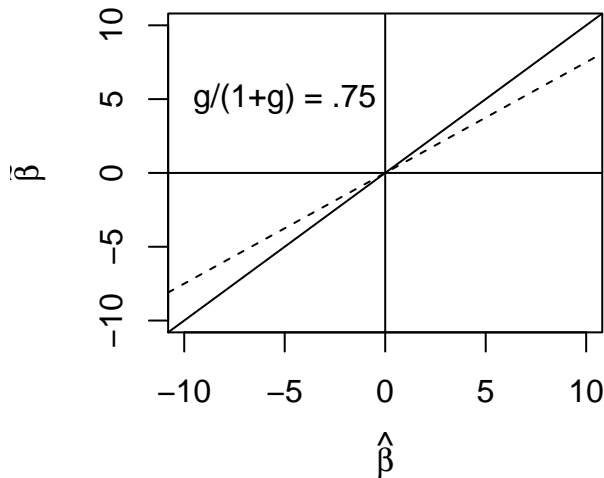
Zellner's g -prior(s) $\beta \mid \phi \sim N(\mathbf{b}_0, g(\mathbf{X}^T \mathbf{X})^{-1} / \phi)$

$$\beta \mid \mathbf{Y}, \phi \sim N \left(\frac{g}{1+g} \hat{\beta} + \frac{1}{1+g} \mathbf{b}_0, \frac{g}{1+g} (\mathbf{X}^T \mathbf{X})^{-1} \phi^{-1} \right)$$

- ▶ Zellner proposed informative choice for the prior mean
- ▶ Invariance under linear transformations of X and Y
- ▶ Avoids extra inverses beyond those in obtaining OLS estimates (computational)
- ▶ Choice of g ?
- ▶ $\frac{g}{1+g}$ weight given to the data

Shrinkage

Posterior mean under g -prior with $\mathbf{b}_0 = 0$ $\tilde{\beta} = \frac{g}{1+g}\hat{\beta}$



Jeffreys Prior

Jeffreys proposed a default procedure so that resulting prior would be invariant to model parameterization

$$p(\boldsymbol{\theta}) \propto |\mathcal{I}(\boldsymbol{\theta})|^{1/2}$$

where $\mathcal{I}(\boldsymbol{\theta})$ is the Expected Fisher Information matrix

$$\mathcal{I}(\boldsymbol{\theta}) = -\mathbb{E}\left[\frac{\partial^2 \log(\mathcal{L}(\boldsymbol{\theta}))}{\partial \theta_i \partial \theta_j}\right]$$

Fisher Information Matrix

Log Likelihood

$$\log(\mathcal{L}(\boldsymbol{\beta}, \phi)) = \frac{n}{2} \log(\phi) - \frac{\phi}{2} \|(\mathbf{I} - \mathbf{P}_{\mathbf{x}}) \mathbf{Y}\|^2 - \frac{\phi}{2} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T (\mathbf{X}^T \mathbf{X}) (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})$$

$$\frac{\partial^2 \log \mathcal{L}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} = \begin{bmatrix} -\phi(\mathbf{X}^T \mathbf{X}) & -(\mathbf{X}^T \mathbf{X})(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \\ -(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T (\mathbf{X}^T \mathbf{X}) & -\frac{n}{2} \frac{1}{\phi^2} \end{bmatrix}$$

$$\mathbb{E}\left[\frac{\partial^2 \log \mathcal{L}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}\right] = \begin{bmatrix} -\phi(\mathbf{X}^T \mathbf{X}) & \mathbf{0}_p \\ \mathbf{0}_p^T & -\frac{n}{2} \frac{1}{\phi^2} \end{bmatrix}$$

$$\mathcal{I}((\boldsymbol{\beta}, \phi)^T) = \begin{bmatrix} \phi(\mathbf{X}^T \mathbf{X}) & \mathbf{0}_p \\ \mathbf{0}_p^T & \frac{n}{2} \frac{1}{\phi^2} \end{bmatrix}$$

Jeffreys Prior

Jeffreys Prior

$$\begin{aligned} p_J(\boldsymbol{\beta}, \phi) &\propto |\mathcal{I}((\boldsymbol{\beta}, \phi)^T)|^{1/2} \\ &= |\phi(\mathbf{X}^T \mathbf{X})|^{1/2} \left(\frac{n}{2} \frac{1}{\phi^2} \right)^{1/2} \\ &\propto \phi^{p/2-1} |\mathbf{X}^T \mathbf{X}|^{1/2} \\ &\propto \phi^{p/2-1} \end{aligned}$$

Improper prior $\iint p_J(\boldsymbol{\beta}, \phi) d\boldsymbol{\beta} d\phi$ not finite

Formal Bayes Posterior

$$p(\boldsymbol{\beta}, \phi \mid \mathbf{Y}) \propto p(\mathbf{Y} \mid \boldsymbol{\beta}, \phi) \phi^{p/2-1}$$

if this is integrable, then renormalize to obtain formal posterior distribution

$$\begin{aligned}\boldsymbol{\beta} \mid \phi, \mathbf{Y} &\sim \mathbf{N}(\hat{\boldsymbol{\beta}}, (\mathbf{X}^T \mathbf{X})^{-1} \phi^{-1}) \\ \phi \mid \mathbf{Y} &\sim \mathbf{G}(n/2, \|\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2/2)\end{aligned}$$

Limiting case of Conjugate prior with $\mathbf{b}_0 = \mathbf{0}$, $\Phi = \mathbf{0}$, $\nu_0 = 0$ and $SS_0 = 0$

Posterior for ϕ does not depend on dimension p ;

Jeffreys did not recommend using this

Independent Jeffreys Prior "Reference Prior"

- ▶ Treat β and ϕ separately ("orthogonal parameterization")
- ▶ $p_{IJ}(\beta) \propto |\mathcal{I}(\beta)|^{1/2}$
- ▶ $p_{IJ}(\phi) \propto |\mathcal{I}(\phi)|^{1/2}$

$$\mathcal{I}((\beta, \phi)^T) = \begin{bmatrix} \phi(\mathbf{X}^T \mathbf{X}) & \mathbf{0}_p \\ \mathbf{0}_p^T & \frac{n}{2} \frac{1}{\phi^2} \end{bmatrix}$$

$$p_{IJ}(\beta) \propto |\phi \mathbf{X}^T \mathbf{X}|^{1/2} \propto 1$$

$$p_{IJ}(\phi) \propto \phi^{-1}$$

Independent Jeffreys Prior is

$$p_{IJ}(\beta, \phi) \propto p_{IJ}(\beta)p_{IJ}(\phi) = \phi^{-1}$$

Formal Posterior Distribution

With Independent Jeffreys Prior

$$p_{IJ}(\beta, \phi) \propto p_{IJ}(\beta)p_{IJ}(\phi) = \phi^{-1}$$

Formal Posterior Distribution (Show!)

$$\begin{aligned}\beta \mid \phi, \mathbf{Y} &\sim \mathbf{N}(\hat{\beta}, (\mathbf{X}^T \mathbf{X})^{-1} \phi^{-1}) \\ \phi \mid \mathbf{Y} &\sim \mathbf{G}((n-p)/2, \|\mathbf{Y} - \mathbf{X}\hat{\beta}\|^2/2) \\ \beta \mid \mathbf{Y} &\sim t_{n-p}(\hat{\beta}, \hat{\sigma}^2 (\mathbf{X}^T \mathbf{X})^{-1})\end{aligned}$$

Bayesian Credible Sets $p(\beta \in C_\alpha) = 1 - \alpha$ correspond to frequentist Confidence Regions

$$\frac{\lambda^T \beta - \lambda^T \hat{\beta}}{\sqrt{\hat{\sigma}^2 \lambda^T (\mathbf{X}^T \mathbf{X})^{-1} \lambda}} \sim t_{n-p}$$

Disadvantages of Conjugate Priors

Disadvantages:

- ▶ Results may have be sensitive to the prior mean which may appear as an “outlier”
- ▶ Cannot capture all possible prior beliefs
- ▶ Mixtures of Conjugate Priors

Mixtures of Conjugate Priors

Theorem (Diaconis & Ylvisaker 1985)

Given a sampling model $p(y \mid \theta)$ from an exponential family, any prior distribution can be expressed as a mixture of conjugate prior distributions

- ▶ Prior $p(\theta) = \int p(\theta \mid \omega)p(\omega) d\omega$
- ▶ Posterior

$$\begin{aligned} p(\theta \mid \mathbf{Y}) &\propto \int p(\mathbf{Y} \mid \theta)p(\theta \mid \omega)p(\omega) d\omega \\ &\propto \int \frac{p(\mathbf{Y} \mid \theta)p(\theta \mid \omega)}{p(\mathbf{Y} \mid \omega)} p(\mathbf{Y} \mid \omega)p(\omega) d\omega \\ &\propto \int p(\theta \mid \mathbf{Y}, \omega)p(\mathbf{Y} \mid \omega)p(\omega) d\omega \\ p(\theta \mid \mathbf{Y}) &= \frac{\int p(\theta \mid \mathbf{Y}, \omega)p(\mathbf{Y} \mid \omega)p(\omega) d\omega}{\int p(\mathbf{Y} \mid \omega)p(\omega) d\omega} \end{aligned}$$

Zellner-Siow Cauchy Prior

- ▶ Model $\mathbf{Y} \sim N(\mathbf{1}\beta_0 + \mathbf{X}\boldsymbol{\beta}, \mathbf{I}_n\phi)$ and assume $\mathbf{X}^T\mathbf{1} = 0$ case where X is centered and has mean 0
- ▶ Conditional Zellner g prior on $\boldsymbol{\beta}$

$$\boldsymbol{\beta} \mid g, \phi, \beta_0 \sim N(0, g(\mathbf{X}^T\mathbf{X})^{-1}/\phi)$$

- ▶ Independent Jeffrey's prior on β_0, ϕ

$$p(\beta_0, \phi \mid g) \propto 1/\phi$$

- ▶ Gamma prior: $1/g \sim G(1/2, n/2) \Rightarrow$ Cauchy prior distribution

$$\boldsymbol{\beta} \mid \phi, \beta_0 \sim C(0, (\mathbf{X}^T\mathbf{X}/n)^{-1}/\phi)$$

- ▶ Posterior Distributions

$$\beta_0 \mid \mathbf{Y}, \boldsymbol{\beta}, \phi, g \sim N(\bar{Y}, \frac{1}{n\phi})$$

$$\boldsymbol{\beta} \mid \mathbf{Y}, \phi, g \sim N(\frac{g}{1+g}\hat{\boldsymbol{\beta}}, \frac{g}{1+g}\frac{1}{\phi}(\mathbf{X}^T\mathbf{X})^{-1})$$