Priors and Intro to Bayesian Variable Selection

Hoff Chapter 9, Mixtures of g-Priors Liang et al JASA

October 16, 2019

Outline

- Conjugate Priors in Bayesian Regression
- Model Selection

Posterior Distribution

Prior Distribution Normal-Gamma

$$\boldsymbol{\beta} \mid \phi \quad \sim \quad \mathsf{N}(\mathbf{b}_0, (\phi\Phi_0)^{-1})$$

Posterior Distribution

Prior Distribution Normal-Gamma

$$\begin{aligned} \boldsymbol{\beta} \mid \phi &\sim \mathsf{N}(\mathbf{b}_0, (\phi \Phi_0)^{-1}) \\ \phi &\sim \mathbf{G}\left(\frac{\nu_0}{2}, \frac{\nu_0 \hat{\sigma}_0^2}{2}\right) \end{aligned}$$

Posterior Distribution

Prior Distribution Normal-Gamma

$$\begin{aligned} \boldsymbol{\beta} \mid \phi &\sim \mathsf{N}(\mathbf{b}_0, (\phi \Phi_0)^{-1}) \\ \phi &\sim \mathbf{G}(\frac{\nu_0}{2}, \frac{\nu_0 \hat{\sigma}_0^2}{2}) \end{aligned}$$

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

Posterior Distribution

Prior Distribution Normal-Gamma

$$
\begin{aligned}
\boldsymbol{\beta} \mid \phi &\sim \mathsf{N}(\mathbf{b}_0, (\phi \Phi_0)^{-1}) \\
\phi &\sim \mathbf{G}\left(\frac{\nu_0}{2}, \frac{\nu_0 \hat{\sigma}_0^2}{2}\right)
\end{aligned}
$$

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

Posterior Distribution

Prior Distribution Normal-Gamma

$$
\begin{aligned}
\boldsymbol{\beta} \mid \phi &\sim \mathrm{N}(\mathbf{b}_0, (\phi\Phi_0)^{-1}) \\
\phi &\sim \mathbf{G}\left(\frac{\nu_0}{2}, \frac{\nu_0\hat{\sigma}_0^2}{2}\right)
\end{aligned}
$$

$$
\begin{aligned}
\Phi_n &= \mathbf{X}^T\mathbf{X} + \Phi_0 \\
\mathbf{b}_n &= \Phi_n^{-1}(\mathbf{X}^T\mathbf{X}\hat{\beta} + \Phi_0\mathbf{b}_0) \\
\mathrm{SSE}_n &= \mathrm{SSE} + \mathrm{SS}_0 + \hat{\beta}^T\mathbf{X}^T\mathbf{X}\hat{\beta} + \mathbf{b}_0^T\Phi_0\mathbf{b}_0 - \mathbf{b}_n^T\Phi_n\mathbf{b}_n
\end{aligned}
$$

Posterior Distribution

Prior Distribution Normal-Gamma

$$\begin{aligned}
\boldsymbol{\beta} \mid \phi &\sim \mathrm{N}(\mathbf{b}_0, (\phi\Phi_0)^{-1}) \\
\phi &\sim \mathbf{G}(\frac{\nu_0}{2}, \frac{\nu_0\hat{\sigma}_0^2}{2})
\end{aligned}$$

$$\begin{aligned}
\Phi_n &= \mathbf{X}^T\mathbf{X} + \Phi_0 \\
\mathbf{b}_n &= \Phi_n^{-1}(\mathbf{X}^T\mathbf{X}\hat{\beta} + \Phi_0\mathbf{b}_0) \\
\mathrm{SSE}_n &= \mathrm{SSE} + \mathrm{SS}_0 + \hat{\beta}^T\mathbf{X}^T\mathbf{X}\hat{\beta} + \mathbf{b}_0^T\Phi_0\mathbf{b}_0 - \mathbf{b}_n^T\Phi_n\mathbf{b}_n \\
\nu_n &= n + \nu_0
\end{aligned}$$

Posterior Distribution

Prior Distribution Normal-Gamma

$$
\begin{aligned}
\boldsymbol{\beta} \mid \phi &\sim \mathrm{N}(\mathbf{b}_0, (\phi \Phi_0)^{-1}) \\
\phi &\sim \mathbf{G}\left(\frac{\nu_0}{2}, \frac{\nu_0 \hat{\sigma}_0^2}{2}\right)
\end{aligned}
$$

$$
\begin{aligned}
\Phi_n &= \mathbf{X}^T \mathbf{X} + \Phi_0 \\
\mathbf{b}_n &= \Phi_n^{-1}(\mathbf{X}^T \mathbf{X} \hat{\beta} + \Phi_0 \mathbf{b}_0) \\
\mathrm{SSE}_n &= \mathrm{SSE} + \mathrm{SS}_0 + \hat{\beta}^T \mathbf{X}^T \mathbf{X} \hat{\beta} + \mathbf{b}_0^T \Phi_0 \mathbf{b}_0 - \mathbf{b}_n^T \Phi_n \mathbf{b}_n \\
\nu_n &= n + \nu_0 \\
\hat{\sigma}_n^2 &= \mathrm{SSE}_n / \nu_n
\end{aligned}
$$

Posterior Distribution

Prior Distribution Normal-Gamma

$$\begin{aligned}
\boldsymbol{\beta} \mid \phi &\sim \mathrm{N}(\mathbf{b}_0, (\phi\Phi_0)^{-1}) \\
\phi &\sim \mathbf{G}\left(\frac{\nu_0}{2}, \frac{\nu_0\hat{\sigma}_0^2}{2}\right)
\end{aligned}$$

$$\begin{aligned}
\Phi_n &= \mathbf{X}^T\mathbf{X} + \Phi_0 \\
\mathbf{b}_n &= \Phi_n^{-1}(\mathbf{X}^T\mathbf{X}\hat{\beta} + \Phi_0\mathbf{b}_0) \\
\mathrm{SSE}_n &= \mathrm{SSE} + \mathrm{SS}_0 + \hat{\beta}^T\mathbf{X}^T\mathbf{X}\hat{\beta} + \mathbf{b}_0^T\Phi_0\mathbf{b}_0 - \mathbf{b}_n^T\Phi_n\mathbf{b}_n \\
\nu_n &= n + \nu_0 \\
\hat{\sigma}_n^2 &= \mathrm{SSE}_n/\nu_n
\end{aligned}$$

Posterior Distribution Normal-Gamma

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

Posterior Distribution

Prior Distribution Normal-Gamma

$$
\begin{aligned}
\boldsymbol{\beta} \mid \phi &\sim \mathsf{N}(\mathbf{b}_0, (\phi\Phi_0)^{-1}) \\
\phi &\sim \mathbf{G}(\frac{\nu_0}{2}, \frac{\nu_0\hat{\sigma}_0^2}{2})
\end{aligned}
$$

$$
\begin{aligned}
\Phi_n &= \mathbf{X}^T\mathbf{X} + \Phi_0 \\
\mathbf{b}_n &= \Phi_n^{-1}(\mathbf{X}^T\mathbf{X}\hat{\beta} + \Phi_0\mathbf{b}_0) \\
\mathsf{SSE}_n &= \mathsf{SSE} + \mathsf{SS}_0 + \hat{\beta}^T\mathbf{X}^T\mathbf{X}\hat{\beta} + \mathbf{b}_0^T\Phi_0\mathbf{b}_0 - \mathbf{b}_n^T\Phi_n\mathbf{b}_n \\
\nu_n &= n + \nu_0 \\
\hat{\sigma}_n^2 &= \mathsf{SSE}_n/\nu_n
\end{aligned}
$$

Posterior Distribution Normal-Gamma

$$
\begin{aligned}
\boldsymbol{\beta} \mid \phi, \mathbf{Y} &\sim \mathsf{N}(\mathbf{b}_n, (\phi\Phi_n)^{-1}) \\
\phi \mid \mathbf{Y} &\sim \mathbf{G}(\frac{n + \nu_0}{2}, \frac{\hat{\sigma}_n^2}{2})
\end{aligned}
$$

Predictive Distribution

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

Predictive Distribution

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

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

Predictive Distribution

Suppose $\mathbf{Y}^* \mid \boldsymbol{\beta}, \phi \sim \mathsf{N}(\mathbf{X}^* \boldsymbol{\beta}, \mathbf{I}/\phi)$ 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

Predictive Distribution

Suppose $\mathbf{Y}^* \mid \boldsymbol{\beta}, \phi \sim \mathrm{N}(\mathbf{X}^*\boldsymbol{\beta}, \mathbf{I}/\phi)$ 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

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

Predictive Distribution

Suppose $\mathbf{Y}^* \mid \boldsymbol{\beta}, \phi \sim \mathrm{N}(\mathbf{X}^*\boldsymbol{\beta}, \mathbf{I}/\phi)$ 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

$$\begin{aligned}
\mathbf{X}^*\boldsymbol{\beta} + \boldsymbol{\epsilon}^* \mid \phi, \mathbf{Y} &\sim \mathrm{N}(\mathbf{X}^*\mathbf{b}_n, (\mathbf{X}^*\Phi_n^{-1}\mathbf{X}^{*T} + \mathbf{I})/\phi) \\
\mathbf{Y}^* \mid \phi, \mathbf{Y} &\sim \mathrm{N}(\mathbf{X}^*\mathbf{b}_n, (\mathbf{X}^*\Phi_n^{-1}\mathbf{X}^{*T} + \mathbf{I})/\phi)
\end{aligned}$$

Predictive Distribution

Suppose $\mathbf{Y}^* \mid \boldsymbol{\beta}, \phi \sim \mathsf{N}(\mathbf{X}^*\boldsymbol{\beta}, \mathbf{I}/\phi)$ 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

$$\begin{aligned}
\mathbf{X}^*\boldsymbol{\beta} + \boldsymbol{\epsilon}^* \mid \phi, \mathbf{Y} &\sim \mathsf{N}(\mathbf{X}^*\mathbf{b}_n, (\mathbf{X}^*\Phi_n^{-1}\mathbf{X}^{*T} + \mathbf{I})/\phi) \\
\mathbf{Y}^* \mid \phi, \mathbf{Y} &\sim \mathsf{N}(\mathbf{X}^*\mathbf{b}_n, (\mathbf{X}^*\Phi_n^{-1}\mathbf{X}^{*T} + \mathbf{I})/\phi) \\
\phi \mid \mathbf{Y} &\sim \mathbf{G}\left(\frac{\nu_n}{2}, \frac{\hat{\sigma}^2\nu_n}{2}\right)
\end{aligned}$$

Predictive Distribution

Suppose $\mathbf{Y}^* \mid \boldsymbol{\beta}, \phi \sim \mathsf{N}(\mathbf{X}^*\boldsymbol{\beta}, \mathbf{I}/\phi)$ 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

$$
\begin{aligned}
\mathbf{X}^*\boldsymbol{\beta} + \boldsymbol{\epsilon}^* \mid \phi, \mathbf{Y} &\sim \mathsf{N}(\mathbf{X}^*\mathbf{b}_n, (\mathbf{X}^*\Phi_n^{-1}\mathbf{X}^{*T} + \mathbf{I})/\phi) \\
\mathbf{Y}^* \mid \phi, \mathbf{Y} &\sim \mathsf{N}(\mathbf{X}^*\mathbf{b}_n, (\mathbf{X}^*\Phi_n^{-1}\mathbf{X}^{*T} + \mathbf{I})/\phi) \\
\phi \mid \mathbf{Y} &\sim \mathbf{G}\left(\frac{\nu_n}{2}, \frac{\hat{\sigma}^2\nu_n}{2}\right) \\
\mathbf{Y}^* \mid \mathbf{Y} &\sim t_{\nu_n}(\mathbf{X}^*\mathbf{b}_n, \hat{\sigma}_n^2(\mathbf{I} + \mathbf{X}^*\Phi_n^{-1}\mathbf{X}^T))
\end{aligned}
$$

Jeffreys Prior

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

Jeffreys Prior

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

- Treat $\boldsymbol{\beta}$ and ϕ independently ("orthogonal parameterization")

Jeffreys Prior

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

- Treat $\boldsymbol{\beta}$ and ϕ independently ("orthogonal parameterization")
- $p_{IJ}(\boldsymbol{\beta}) \propto |\mathcal{J}(\boldsymbol{\beta})|^{1/2}$

Jeffreys Prior

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

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

Jeffreys Prior

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

- Treat $\boldsymbol{\beta}$ and ϕ independently ("orthogonal parameterization")
- $p_{IJ}(\boldsymbol{\beta}) \propto |\mathcal{J}(\boldsymbol{\beta})|^{1/2}$
- $p_{IJ}(\phi) \propto |\mathcal{J}(\phi)||^{1/2}$

where $\mathcal{J}(\tilde{\boldsymbol{\beta}})$ is the Expected Fisher Information matrix

Jeffreys Prior

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

- Treat $\boldsymbol{\beta}$ and ϕ independently ("orthogonal parameterization")
- $p_{IJ}(\boldsymbol{\beta}) \propto |\mathcal{J}(\boldsymbol{\beta})|^{1/2}$
- $p_{IJ}(\phi) \propto |\mathcal{J}(\phi)\|^{1/2}$

where $\mathcal{J}(\tilde{\boldsymbol{\beta}})$ is the Expected Fisher Information matrix

$$\mathcal{J}(\tilde{\boldsymbol{\beta}}) = -\mathrm{E}[\left[\frac{\partial^2 \log(\mathcal{L}(\tilde{\boldsymbol{\beta}}))}{\partial\theta_i \partial\theta_j}\right]]$$

Fisher Information Matrix

Log Likelihood

$$\log(\mathcal{L}(\boldsymbol{\beta}, \phi)) = \frac{n}{2}\log(\phi) - \frac{\phi}{2}\text{SSE} - \frac{\phi}{2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T(\mathbf{X}^T\mathbf{X})(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})$$

Fisher Information Matrix

Log Likelihood

$$\log(\mathcal{L}(\boldsymbol{\beta}, \phi)) = \frac{n}{2}\log(\phi) - \frac{\phi}{2}\text{SSE} - \frac{\phi}{2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T(\mathbf{X}^T\mathbf{X})(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})$$

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

Fisher Information Matrix

Log Likelihood

$$\log(\mathcal{L}(\boldsymbol{\beta}, \phi)) = \frac{n}{2}\log(\phi) - \frac{\phi}{2}\text{SSE} - \frac{\phi}{2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T(\mathbf{X}^T\mathbf{X})(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})$$

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

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

Fisher Information Matrix

Log Likelihood

$$\log(\mathcal{L}(\boldsymbol{\beta}, \phi)) = \frac{n}{2}\log(\phi) - \frac{\phi}{2}\text{SSE} - \frac{\phi}{2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T(\mathbf{X}^T\mathbf{X})(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})$$

$$\begin{aligned}
\frac{\partial^2 \log \mathcal{L}}{\partial \tilde{\boldsymbol{\beta}} \partial \tilde{\boldsymbol{\beta}}^T} &= \begin{bmatrix} -\phi(\mathbf{X}^T\mathbf{X}) & -(\mathbf{X}^T\mathbf{X})(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \\ -(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T(\mathbf{X}^T\mathbf{X}) & -\frac{n}{2}\frac{1}{\phi^2} \end{bmatrix} \\
\mathrm{E}[\frac{\partial^2 \log \mathcal{L}}{\partial \tilde{\boldsymbol{\beta}} \partial \tilde{\boldsymbol{\beta}}^T}] &= \begin{bmatrix} -\phi(\mathbf{X}^T\mathbf{X}) & \mathbf{0}_p \\ \mathbf{0}_p^T & -\frac{n}{2}\frac{1}{\phi^2} \end{bmatrix} \\
\mathcal{I}((\boldsymbol{\beta}, \phi)^T) &= \begin{bmatrix} \phi(\mathbf{X}^T\mathbf{X}) & \mathbf{0}_p \\ \mathbf{0}_p^T & \frac{n}{2}\frac{1}{\phi^2} \end{bmatrix}
\end{aligned}$$

Independent Jeffreys Prior "Reference Prior"

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

Independent Jeffreys Prior "Reference Prior"

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

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

Independent Jeffreys Prior "Reference Prior"

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

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

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

$\hat{\Sigma}$

$\det |(X^T X)^{-1}|$

$f(x) \geq c$

$f(x)\,dx$

Independent Jeffreys Prior "Reference Prior"

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

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

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

Independent Jeffreys Prior is

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

Formal Posterior Distribution

With Independent Jeffreys Prior

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

Formal Posterior Distribution

With Independent Jeffreys Prior

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

Formal Posterior Distribution (Show!)

Formal Posterior Distribution

With Independent Jeffreys Prior

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

Formal Posterior Distribution (Show!)

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

Formal Posterior Distribution

With Independent Jeffreys Prior

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

Formal Posterior Distribution (Show!)

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

Formal Posterior Distribution

With Independent Jeffreys Prior

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

Formal Posterior Distribution (Show!)

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

Formal Posterior Distribution

With Independent Jeffreys Prior

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

Formal Posterior Distribution (Show!)

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

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

$$\frac{\beta_j - \hat{\boldsymbol{\beta}}_j}{\sqrt{\hat{\sigma}^2[(\mathbf{X}^T\mathbf{X})^{-1}]_{jj}}} \sim t_{n-p}$$

Formal Posterior Distribution

With Independent Jeffreys Prior

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

Formal Posterior Distribution (Show!)

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

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

$$\frac{\beta_j - \hat{\boldsymbol{\beta}}_j}{\sqrt{\hat{\sigma}^2[(\mathbf{X}^T\mathbf{X})^{-1}]_{jj}}} \sim t_{n-p}$$

US Air Example

```
## Error in library(HH): there is no package called
'HH'
## Warning in data(usair): data set 'usair' not found
## Error in colnames(usair) = c("SO2", "temp",
"firms", "popn", "wind", "precip", : object 'usair'
not found
## Error in select(usair, c(temp, firms, popn, wind,
precip, rain, SO2)): object 'usair' not found
## Error in ggpairs(usair): object 'usair' not found
```

Credible Intervals

Reference analysis

```
Error in is.data.frame(data): object 'usair' not
found
Error in confint(poll.lm): object 'poll.lm' not found
```

Credible Intervals

Reference analysis

```
Error in is.data.frame(data): object 'usair' not
found
Error in confint(poll.lm): object 'poll.lm' not found
```

Many intervals contain 0!

Credible Intervals

Reference analysis

```
Error in is.data.frame(data): object 'usair' not
found
Error in confint(poll.lm): object 'poll.lm' not found
```

Many intervals contain 0!
Multi-collinearity

Credible Intervals

Reference analysis

```
Error in is.data.frame(data): object 'usair' not
found
Error in confint(poll.lm): object 'poll.lm' not found
```

Many intervals contain 0!
Multi-collinearity
Jeffreys prior cannot be used for variable selection (more later)

Zellner's g-prior

rescales axes (sphering)

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

$g = n$

JP

why g is a knob

prior prec. on resid.

all data $\longleftrightarrow$ all prior

Zellner's g-prior

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

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

Zellner's g-prior

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

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

- Zellner proposed informative choice for the prior mean

Zellner's g-prior

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

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

- Zellner proposed informative choice for the prior mean
- Invariance under linear transformations of X and Y

Zellner's g-prior

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

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

- 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)

Zellner's g-prior

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

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

- 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)
- $\frac{g}{1+g}$ weight given to the data

Zellner's g-prior

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

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

- 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)
- $\frac{g}{1+g}$ weight given to the data
- Choice of g?

Zellner's g-prior

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

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

- 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)
- $\frac{g}{1+g}$ weight given to the data
- Choice of g?
- Same g for intercept and other coefficients

Zellner's g-prior II

Centered model: $\mathbf{Y} = \mathbf{1}_n\alpha + \mathbf{X}_c\boldsymbol{\beta} + \epsilon$
where $\mathbf{X}_c$ is the centered design matrix where all variables have had their mean subtracted

Zellner's g-prior II

moved from $X \to X_c$

Centered model: $\mathbf{Y} = \mathbf{1}_n \alpha + \mathbf{X}_c \boldsymbol{\beta} + \epsilon$

where $\mathbf{X}_c$ is the centered design matrix where all variables have had their mean subtracted

$p(\phi) \propto 1/\phi$

Zellner's g-prior II

Centered model: $\mathbf{Y} = \mathbf{1}_n \alpha + \mathbf{X}_c \boldsymbol{\beta} + \epsilon$
where $\mathbf{X}_c$ is the centered design matrix where all variables have had their mean subtracted
$p(\phi) \propto 1/\phi$ $p(\alpha \mid \phi) \propto 1$

Zellner's g-prior II

Centered model: $\mathbf{Y} = \mathbf{1}_n\alpha + \mathbf{X}_c\boldsymbol{\beta} + \epsilon$
where $\mathbf{X}_c$ is the centered design matrix where all variables have had their mean subtracted
$p(\phi) \propto 1/\phi$ $p(\alpha \mid \phi) \propto 1$ $\boldsymbol{\beta} \mid \alpha, \phi, g \sim \mathsf{N}(0, g\phi^{-1}(\mathbf{X}_c{}'\mathbf{X}_c)^{-1})$

Zellner's g-prior II

pre- conditioning ?

Centered model: $\mathbf{Y} = \mathbf{1}_n\alpha + \mathbf{X}_c\boldsymbol{\beta} + \epsilon$
where $\mathbf{X}_c$ is the centered design matrix where all variables have had their mean subtracted
$p(\phi) \propto 1/\phi \;\; p(\alpha \mid \phi) \propto 1 \;\; \boldsymbol{\beta} \mid \alpha, \phi, g \sim \mathrm{N}(0, g\phi^{-1}(\mathbf{X}_c{}'\mathbf{X}_c)^{-1})$ $(X_c^T X_c)^{-1}$

$$\alpha \mid \mathbf{Y}, \phi \sim N(\bar{y}, 1/(\phi n))$$

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

$$\phi \mid \mathbf{Y} \sim \text{Gamma}\left(\frac{n-1}{2}, \frac{\text{SSE} + \frac{1}{1+g}\hat{\boldsymbol{\beta}}^T(\mathbf{X}_c^T\mathbf{X}_c)\hat{\boldsymbol{\beta}}}{2}\right)$$

M^{-1}

$$\boldsymbol{\beta} \mid \mathbf{Y} \sim t\left(n-1, \frac{g}{1+g}\hat{\boldsymbol{\beta}}, \hat{\sigma}^2_n\frac{g}{1+g}(\mathbf{X}_c^T\mathbf{X}_c)^{-1}\right)$$

$$\hat{\sigma}^2_n = \frac{\text{SSE} + \frac{1}{1+g}\hat{\boldsymbol{\beta}}^T(\mathbf{X}_c^T\mathbf{X}_c)\hat{\boldsymbol{\beta}}}{n-1}$$

$M = X_c^T X_c$

very ellipsoidal massive then po... $\to$ $\lambda_{max}/\lambda_{min}(M)$ is

Poorly conditioned

joint posterior draws of beta's

```
## Error in nrow(usair): object 'usair' not found
## Error in ncol(usair): object 'usair' not found
## Error in coef(poll.lm): object 'poll.lm' not found
## Error in eval(expr, envir, enclos): object 'bhat'
not found
## Error in is.data.frame(x): object 'usair' not
found
## Error in eval(expr, envir, enclos): object
'poll.lm' not found
## Error in eval(expr, envir, enclos): object 'totss'
not found
## Error in eval(expr, envir, enclos): object 'sse'
not found
## Error in model.matrix(poll.lm): object 'poll.lm'
not found
## Error in eval(expr, envir, enclos): object
'sigma2.bayes' not found
## Error in rmvt(5000, delta = bhat, type =
```

Bayesian Variable Selection

- Avoid the use of redundant variables (problems with interpretations)

Bayesian Variable Selection

- Avoid the use of redundant variables (problems with interpretations)
- Inclusion of un-necessary terms yields less precise estimates, particularly if explanatory variables are highly correlated with each other

Bayesian Variable Selection

- Avoid the use of redundant variables (problems with interpretations)
- Inclusion of un-necessary terms yields less precise estimates, particularly if explanatory variables are highly correlated with each other
- reduced MSE: reduced variance but possibly higher bias

Bayesian Variable Selection

- Avoid the use of redundant variables (problems with interpretations)
- Inclusion of un-necessary terms yields less precise estimates, particularly if explanatory variables are highly correlated with each other
- reduced MSE: reduced variance but possibly higher bias
- it is too “expensive” to use all variables

Bayesian Model Choice

- Models for the variable selection problem are based on a subset of the $\mathbf{X}_1, \ldots \mathbf{X}_p$ variables

Bayesian Model Choice

- Models for the variable selection problem are based on a subset of the $\mathbf{X}_1, \ldots \mathbf{X}_p$ variables
- Encode models with a vector $\boldsymbol{\gamma} = (\gamma_1, \ldots \gamma_p)$ where $\gamma_j \in \{0, 1\}$ is an indicator for whether variable $\mathbf{X}_j$ should be included in the model $\mathcal{M}_\gamma$. $\gamma_j = 0 \Leftrightarrow \beta_j = 0$

Bayesian Model Choice

- Models for the variable selection problem are based on a subset of the $\mathbf{X}_1, \ldots \mathbf{X}_p$ variables
- Encode models with a vector $\boldsymbol{\gamma} = (\gamma_1, \ldots \gamma_p)$ where $\gamma_j \in \{0, 1\}$ is an indicator for whether variable $\mathbf{X}_j$ should be included in the model $\mathcal{M}_\gamma$. $\gamma_j = 0 \Leftrightarrow \beta_j = 0$
- Each value of $\boldsymbol{\gamma}$ represents one of the 2^p models.

Bayesian Model Choice

- Models for the variable selection problem are based on a subset of the $\mathbf{X}_1, \ldots \mathbf{X}_p$ variables
- Encode models with a vector $\boldsymbol{\gamma} = (\gamma_1, \ldots \gamma_p)$ where $\gamma_j \in \{0, 1\}$ is an indicator for whether variable $\mathbf{X}_j$ should be included in the model $\mathcal{M}_\gamma$. $\gamma_j = 0 \Leftrightarrow \beta_j = 0$
- Each value of $\boldsymbol{\gamma}$ represents one of the 2^p models.
- Under model $\mathcal{M}_\gamma$:

$$\mathbf{Y} \mid \alpha, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\gamma} \sim \mathsf{N}(\mathbf{1}\alpha + \mathbf{X}_{\boldsymbol{\gamma}}\boldsymbol{\beta}_\gamma, \sigma^2\mathbf{I})$$

Where $\mathbf{X}_\gamma$ is design matrix using the columns in $\mathbf{X}$ where $\gamma_j = 1$ and $\boldsymbol{\beta}_\gamma$ is the subset of $\boldsymbol{\beta}$ that are non-zero.

Posterior Probabilities of Models

- Posterior model probabilities

$$p(\mathcal{M}_j \mid \mathbf{Y}) = \frac{p(\mathbf{Y} \mid \mathcal{M}_j)p(\mathcal{M}_j)}{\sum_j p(\mathbf{Y} \mid \mathcal{M}_j)p(\mathcal{M}_j)}$$

Posterior Probabilities of Models

- Posterior model probabilities

$$p(\mathcal{M}_j \mid \mathbf{Y}) = \frac{p(\mathbf{Y} \mid \mathcal{M}_j)p(\mathcal{M}_j)}{\sum_j p(\mathbf{Y} \mid \mathcal{M}_j)p(\mathcal{M}_j)}$$

Marginal likelihod of a model is proportional to

$$p(\mathbf{Y} \mid \mathcal{M}_\gamma) = \iint p(\mathbf{Y} \mid \boldsymbol{\beta}_\gamma, \sigma^2)p(\boldsymbol{\beta}_\gamma \mid \boldsymbol{\gamma}, \sigma^2)p(\sigma^2 \mid \boldsymbol{\gamma})d\boldsymbol{\beta}\, d\sigma^2$$

Posterior Probabilities of Models

- Posterior model probabilities

$$p(\mathcal{M}_j \mid \mathbf{Y}) = \frac{p(\mathbf{Y} \mid \mathcal{M}_j)p(\mathcal{M}_j)}{\sum_j p(\mathbf{Y} \mid \mathcal{M}_j)p(\mathcal{M}_j)}$$

  Marginal likelihod of a model is proportional to

$$p(\mathbf{Y} \mid \mathcal{M}_\gamma) = \iint p(\mathbf{Y} \mid \boldsymbol{\beta}_\gamma, \sigma^2)p(\boldsymbol{\beta}_\gamma \mid \boldsymbol{\gamma}, \sigma^2)p(\sigma^2 \mid \boldsymbol{\gamma})d\boldsymbol{\beta}\, d\sigma^2$$

- Bayes Factor $BF[i:j]$

$$\frac{P(\mathcal{M}_i \mid \mathbf{Y})}{P(\mathcal{M}_j \mid \mathbf{Y})} = \frac{p(\mathbf{Y} \mid \mathcal{M}_i)}{p(\mathbf{Y} \mid \mathcal{M}_j)} \times \frac{P(\mathcal{M}_i)}{P(\mathcal{M}_j)}$$

  Posterior Odds $=$ Bayes Factor $\times$ Prior odds

Posterior Probabilities of Models

- Posterior model probabilities

$$p(\mathcal{M}_j \mid \mathbf{Y}) = \frac{p(\mathbf{Y} \mid \mathcal{M}_j)p(\mathcal{M}_j)}{\sum_j p(\mathbf{Y} \mid \mathcal{M}_j)p(\mathcal{M}_j)}$$

  Marginal likelihod of a model is proportional to

$$p(\mathbf{Y} \mid \mathcal{M}_\gamma) = \int\int p(\mathbf{Y} \mid \boldsymbol{\beta}_\gamma, \sigma^2)p(\boldsymbol{\beta}_\gamma \mid \boldsymbol{\gamma}, \sigma^2)p(\sigma^2 \mid \boldsymbol{\gamma})d\boldsymbol{\beta}\, d\sigma^2$$

- Bayes Factor $BF[i:j]$

$$\frac{P(\mathcal{M}_i \mid \mathbf{Y})}{P(\mathcal{M}_j \mid \mathbf{Y})} = \frac{p(\mathbf{Y} \mid \mathcal{M}_i)}{p(\mathbf{Y} \mid \mathcal{M}_j)} \times \frac{P(\mathcal{M}_i)}{P(\mathcal{M}_j)}$$

  Posterior Odds = Bayes Factor × Prior odds

- Probability $\beta_j \neq 0$: $\sum_{\mathcal{M}_j:\beta_j \neq 0} p(\mathcal{M}_j \mid \mathbf{Y})$ (marginal posterior inclusion probability)

Posterior Probabilities of Models

- Posterior model probabilities

$$p(\mathcal{M}_j \mid \mathbf{Y}) = \frac{p(\mathbf{Y} \mid \mathcal{M}_j) p(\mathcal{M}_j)}{\sum_j p(\mathbf{Y} \mid \mathcal{M}_j) p(\mathcal{M}_j)}$$

  Marginal likelihod of a model is proportional to

$$p(\mathbf{Y} \mid \mathcal{M}_\gamma) = \iint p(\mathbf{Y} \mid \boldsymbol{\beta}_\gamma, \sigma^2) p(\boldsymbol{\beta}_\gamma \mid \boldsymbol{\gamma}, \sigma^2) p(\sigma^2 \mid \boldsymbol{\gamma}) d\boldsymbol{\beta}\, d\sigma^2$$

- Bayes Factor $BF[i:j]$

$$\frac{P(\mathcal{M}_i \mid \mathbf{Y})}{P(\mathcal{M}_j \mid \mathbf{Y})} = \frac{p(\mathbf{Y} \mid \mathcal{M}_i)}{p(\mathbf{Y} \mid \mathcal{M}_j)} \times \frac{P(\mathcal{M}_i)}{P(\mathcal{M}_j)}$$

  Posterior Odds $=$ Bayes Factor $\times$ Prior odds

- Probability $\beta_j \neq 0$: $\sum_{\mathcal{M}_j : \beta_j \neq 0} p(\mathcal{M}_j \mid \mathbf{Y})$ (marginal posterior inclusion probability)

Prior Distributions

- Bayesian Model choice requires proper prior distributions on regression coefficients (exception parameters that are included in all models)

Prior Distributions

- Bayesian Model choice requires proper prior distributions on regression coefficients (exception parameters that are included in all models)
- Vague but proper priors may lead to paradoxes!

Prior Distributions

- Bayesian Model choice requires proper prior distributions on regression coefficients (exception parameters that are included in all models)
- Vague but proper priors may lead to paradoxes!
- Conjugate Normal-Gammas lead to closed form expressions for marginal likelihoods, Zellner's g-prior is one of the most popular.

Zellner's g-prior within Models

Centered model:

$$\mathbf{Y} = \mathbf{1}_n\alpha + \mathbf{X}^c_{\boldsymbol{\gamma}}\boldsymbol{\beta}_\gamma + \epsilon$$

- Common parameters

$$p(\alpha, \phi) \propto \phi^{-1}$$

Zellner's g-prior within Models

Centered model:

$$\mathbf{Y} = \mathbf{1}_n \alpha + \mathbf{X}^c_{\boldsymbol{\gamma}} \boldsymbol{\beta}_\gamma + \epsilon$$

- Common parameters

$$p(\alpha, \phi) \propto \phi^{-1}$$

- Model Specific parameters

$$\boldsymbol{\beta}_\gamma \mid \alpha, \phi, \boldsymbol{\gamma} \sim \mathsf{N}(0, g\phi^{-1}(\mathbf{X}^{c\,\prime}_{\boldsymbol{\gamma}} \mathbf{X}^c_{\boldsymbol{\gamma}})^{-1})$$

Zellner's g-prior within Models

Centered model:

$$\mathbf{Y} = \mathbf{1}_n\alpha + \mathbf{X}^c_{\boldsymbol{\gamma}}\boldsymbol{\beta}_\gamma + \epsilon$$

- Common parameters

$$p(\alpha, \phi) \propto \phi^{-1}$$

- Model Specific parameters

$$\boldsymbol{\beta}_\gamma \mid \alpha, \phi, \boldsymbol{\gamma} \sim \mathsf{N}(0, g\phi^{-1}(\mathbf{X}^{c\,\prime}_{\boldsymbol{\gamma}}\mathbf{X}^c_{\boldsymbol{\gamma}})^{-1})$$

- Marginal likelihood of $\mathcal{M}_\gamma$ is proportional to

$$p(\mathbf{Y} \mid \mathcal{M}_\gamma) = C(1+g)^{\frac{n-p-1}{2}}(1+g(1-R^2_\gamma))^{-\frac{(n-1)}{2}}$$

where R^2_γ is the usual R^2 for model $\mathcal{M}_\gamma$ and C is a constant that is $p(\mathbf{Y} \mid \mathcal{M}_0)$ (model with intercept alone)

Zellner's g-prior within Models

Centered model:

$$\mathbf{Y} = \mathbf{1}_n \alpha + \mathbf{X}^c_{\boldsymbol{\gamma}} \boldsymbol{\beta}_\gamma + \epsilon$$

- Common parameters

$$p(\alpha, \phi) \propto \phi^{-1}$$

- Model Specific parameters

$$\boldsymbol{\beta}_\gamma \mid \alpha, \phi, \boldsymbol{\gamma} \sim \mathsf{N}(0, g\phi^{-1}({\mathbf{X}^c_{\boldsymbol{\gamma}}}'\mathbf{X}^c_{\boldsymbol{\gamma}})^{-1})$$

- Marginal likelihood of $\mathcal{M}_\gamma$ is proportional to

$$p(\mathbf{Y} \mid \mathcal{M}_\gamma) = C(1+g)^{\frac{n-p-1}{2}}(1 + g(1 - R^2_\gamma))^{-\frac{(n-1)}{2}}$$

  where R^2_γ is the usual R^2 for model $\mathcal{M}_\gamma$ and C is a constant that is $p(\mathbf{Y} \mid \mathcal{M}_0)$ (model with intercept alone)
- uniform distribution over space of models $p(\mathcal{M}_\gamma) = 1/(2^p)$

USair Data: Enumeration of All Models

```
library(devtools)

## Loading required package: usethis

suppressMessages(install_github("merliseclyde/BAS"))  # cu
library(BAS)
poll.bma = bas.lm(log(SO2) ~ temp + log(firms) +
                             log(popn) + wind +
                             precip+ rain,
                  data=usair,
                  prior="g-prior",
                  alpha=41,     # g = n
                  n.models=2^7,# enumerate (can omit)
                  modelprior=uniform(),
                  method="deterministic")    # fast enumera

## Error in is.data.frame(data): object 'usair' not
found
```

residual plot)

```
plot(poll.bma, which=1)

## Error in plot(poll.bma, which = 1): object
'poll.bma' not found
```

Model Complexity)

```
plot(poll.bma, which=3)

## Error in plot(poll.bma, which = 3): object
'poll.bma' not found
```

Inclusion Probabilities)

```r
plot(poll.bma, which=4)

## Error in plot(poll.bma, which = 4): object
'poll.bma' not found
```

Model Space

```r
summary(poll.bma)

## Error in summary(poll.bma): object 'poll.bma' not
found
```

Summary

```
image(poll.bma)

Error in image(poll.bma): object 'poll.bma' not found
```

Coefficients

```
beta = coef(poll.bma, n.models=1)

## Error in coef(poll.bma, n.models = 1): object
'poll.bma' not found

beta

## function (a, b)
## .Internal(beta(a, b))
## <bytecode: 0x7f87eeb0f0b0>
## <environment: namespace:base>
```

Coefficients

```
par(mfrow=c(2,2));  plot(beta, subset=c(3, 6))
```

```
## Error in x(x): argument "b" is missing, with no
default
```

Bayesian Confidence Intervals

```r
confint(beta)

## Error: object of type 'closure' is not subsettable
```