Choice of Prior Distributions

STA721 Linear Models Duke University

Merlise Clyde

September 23, 2019

Bayesian Estimation

Model

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

precision $\phi = 1/\sigma^2$

Bayesian Estimation

Model

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

precision $\phi = 1/\sigma^2$

Normal-Gamma Conjugate prior $\mathrm{NG}(\mathbf{b}_0, \Phi_0, \mathbf{v}_0, \mathrm{SS}_0)$

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

Bayesian Estimation

Model

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

precision $\phi = 1/\sigma^2$

Normal-Gamma Conjugate prior $\mathrm{NG}(\mathbf{b}_0, \Phi_0, \mathbf{v}_0, \mathrm{SS}_0)$

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

Posterior Distribution

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

Bayesian Estimation

Model

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

precision $\phi = 1/\sigma^2$

Normal-Gamma Conjugate prior $\mathrm{NG}(\mathbf{b}_0, \Phi_0, \mathbf{v}_0, \mathrm{SS}_0)$

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

Posterior Distribution

$$\begin{aligned}
\beta \mid \phi, \mathbf{Y} &\sim \mathrm{N}(\mathbf{b}_n, (\phi\Phi_n)^{-1}) \\
\phi \mid \mathbf{Y} &\sim \mathbf{G}\left(\frac{\nu_n}{2}, \frac{\nu_n\hat{\sigma}_n^2}{2}\right)
\end{aligned}$$

Marginal Distribution from Normal–Gamma

Theorem

Let $\boldsymbol{\theta} \mid \phi \sim N(m, \frac{1}{\phi}\Sigma)$ *and* $\phi \sim \mathbf{G}(\nu/2, \nu\hat{\sigma}^2/2)$. *Then* $\mathbf{t}$ $(p \times 1)$ *has a p dimensional multivariate t distribution*

$$\boldsymbol{\theta} \sim t_\nu(m, \hat{\sigma}^2\Sigma)$$

with density

$$p(\boldsymbol{\theta}) \propto \left[1 + \frac{1}{\nu}\frac{(\boldsymbol{\theta} - m)^T\Sigma^{-1}(\boldsymbol{\theta} - m)}{\hat{\sigma}^2}\right]^{-\frac{p+\nu}{2}}$$

Derivation

Marginal density $p(\boldsymbol{\theta}) = \int p(\boldsymbol{\theta} \mid \phi) p(\phi) \, d\phi$

Derivation

Marginal density $p(\boldsymbol{\theta}) = \int p(\boldsymbol{\theta} \mid \phi) p(\phi) \, d\phi$

$$p(\boldsymbol{\theta}) \quad \propto \quad \int |\Sigma/\phi|^{-1/2} e^{-\frac{\phi}{2}(\boldsymbol{\theta}-m)^T \Sigma^{-1}(\boldsymbol{\theta}-m)} \phi^{\nu/2-1} e^{-\phi \frac{\nu \hat{\sigma}^2}{2}} \, d\phi$$

Derivation

Marginal density $p(\boldsymbol{\theta}) = \int p(\boldsymbol{\theta} \mid \phi) p(\phi)\, d\phi$

$$
\begin{aligned}
p(\boldsymbol{\theta}) &\propto \int |\Sigma/\phi|^{-1/2} e^{-\frac{\phi}{2}(\boldsymbol{\theta}-m)^T\Sigma^{-1}(\boldsymbol{\theta}-m)} \phi^{\nu/2-1} e^{-\phi\frac{\nu\hat{\sigma}^2}{2}}\, d\phi \\
&\propto \int \phi^{p/2} \phi^{\nu/2-1} e^{-\phi\frac{(\boldsymbol{\theta}-m)^T\Sigma^{-1}(\boldsymbol{\theta}-m)+\nu\hat{\sigma}^2}{2}}\, d\phi
\end{aligned}
$$

Derivation

Marginal density $p(\boldsymbol{\theta}) = \int p(\boldsymbol{\theta} \mid \phi) p(\phi)\, d\phi$

$$
\begin{aligned}
p(\boldsymbol{\theta}) &\propto \int |\Sigma/\phi|^{-1/2} e^{-\frac{\phi}{2}(\boldsymbol{\theta}-m)^T \Sigma^{-1}(\boldsymbol{\theta}-m)} \phi^{\nu/2-1} e^{-\phi \frac{\nu \hat{\sigma}^2}{2}}\, d\phi \\
&\propto \int \phi^{p/2} \phi^{\nu/2-1} e^{-\phi \frac{(\boldsymbol{\theta}-m)^T \Sigma^{-1}(\boldsymbol{\theta}-m)+\nu\hat{\sigma}^2}{2}}\, d\phi \\
&\propto \int \phi^{\frac{p+\nu}{2}-1} e^{-\phi \frac{(\boldsymbol{\theta}-m)^T \Sigma^{-1}(\boldsymbol{\theta}-m)+\nu\hat{\sigma}^2}{2}}\, d\phi
\end{aligned}
$$

Derivation

Marginal density $p(\boldsymbol{\theta}) = \int p(\boldsymbol{\theta} \mid \phi) p(\phi)\, d\phi$

$$
\begin{aligned}
p(\boldsymbol{\theta}) &\propto \int |\Sigma/\phi|^{-1/2} e^{-\frac{\phi}{2}(\boldsymbol{\theta}-m)^T\Sigma^{-1}(\boldsymbol{\theta}-m)} \phi^{\nu/2-1} e^{-\phi\frac{\nu\hat{\sigma}^2}{2}}\, d\phi \\
&\propto \int \phi^{p/2}\phi^{\nu/2-1} e^{-\phi\frac{(\boldsymbol{\theta}-m)^T\Sigma^{-1}(\boldsymbol{\theta}-m)+\nu\hat{\sigma}^2}{2}}\, d\phi \\
&\propto \int \phi^{\frac{p+\nu}{2}-1} e^{-\phi\frac{(\boldsymbol{\theta}-m)^T\Sigma^{-1}(\boldsymbol{\theta}-m)+\nu\hat{\sigma}^2}{2}}\, d\phi \\
&= \Gamma((p+\nu)/2)\left(\frac{(\boldsymbol{\theta}-m)^T\Sigma^{-1}(\boldsymbol{\theta}-m)+\nu\hat{\sigma}^2}{2}\right)^{-\frac{p+\nu}{2}}
\end{aligned}
$$

Derivation

Marginal density $p(\boldsymbol{\theta}) = \int p(\boldsymbol{\theta} \mid \phi) p(\phi) \, d\phi$

$$
\begin{aligned}
p(\boldsymbol{\theta}) &\propto \int |\Sigma/\phi|^{-1/2} e^{-\frac{\phi}{2}(\boldsymbol{\theta}-m)^T\Sigma^{-1}(\boldsymbol{\theta}-m)} \phi^{\nu/2-1} e^{-\phi\frac{\nu\hat{\sigma}^2}{2}} \, d\phi \\
&\propto \int \phi^{p/2}\phi^{\nu/2-1} e^{-\phi\frac{(\boldsymbol{\theta}-m)^T\Sigma^{-1}(\boldsymbol{\theta}-m)+\nu\hat{\sigma}^2}{2}} \, d\phi \\
&\propto \int \phi^{\frac{p+\nu}{2}-1} e^{-\phi\frac{(\boldsymbol{\theta}-m)^T\Sigma^{-1}(\boldsymbol{\theta}-m)+\nu\hat{\sigma}^2}{2}} \, d\phi \\
&= \Gamma((p+\nu)/2)\left(\frac{(\boldsymbol{\theta}-m)^T\Sigma^{-1}(\boldsymbol{\theta}-m)+\nu\hat{\sigma}^2}{2}\right)^{-\frac{p+\nu}{2}} \\
&\propto \left((\boldsymbol{\theta}-m)^T\Sigma^{-1}(\boldsymbol{\theta}-m)+\nu\hat{\sigma}^2\right)^{-\frac{p+\nu}{2}}
\end{aligned}
$$

Derivation

Marginal density $p(\boldsymbol{\theta}) = \int p(\boldsymbol{\theta} \mid \phi) p(\phi)\, d\phi$

$$\begin{aligned}
p(\boldsymbol{\theta}) &\propto \int |\Sigma/\phi|^{-1/2} e^{-\frac{\phi}{2}(\boldsymbol{\theta}-m)^T\Sigma^{-1}(\boldsymbol{\theta}-m)} \phi^{\nu/2-1} e^{-\phi\frac{\nu\hat{\sigma}^2}{2}}\, d\phi \\
&\propto \int \phi^{p/2}\phi^{\nu/2-1} e^{-\phi\frac{(\boldsymbol{\theta}-m)^T\Sigma^{-1}(\boldsymbol{\theta}-m)+\nu\hat{\sigma}^2}{2}}\, d\phi \\
&\propto \int \phi^{\frac{p+\nu}{2}-1} e^{-\phi\frac{(\boldsymbol{\theta}-m)^T\Sigma^{-1}(\boldsymbol{\theta}-m)+\nu\hat{\sigma}^2}{2}}\, d\phi \\
&= \Gamma((p+\nu)/2)\left(\frac{(\boldsymbol{\theta}-m)^T\Sigma^{-1}(\boldsymbol{\theta}-m)+\nu\hat{\sigma}^2}{2}\right)^{-\frac{p+\nu}{2}} \\
&\propto \left((\boldsymbol{\theta}-m)^T\Sigma^{-1}(\boldsymbol{\theta}-m)+\nu\hat{\sigma}^2\right)^{-\frac{p+\nu}{2}} \\
&\propto \left(1+\frac{1}{\nu}\frac{(\boldsymbol{\theta}-m)^T\Sigma^{-1}(\boldsymbol{\theta}-m)}{\hat{\sigma}^2}\right)^{-\frac{p+\nu}{2}}
\end{aligned}$$

Marginal Posterior Distribution of $\boldsymbol{\beta}$

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

Marginal Posterior Distribution of $\boldsymbol{\beta}$

$$
\begin{aligned}
\boldsymbol{\beta} \mid \phi, \mathbf{Y} &\sim \mathsf{N}(\mathbf{b}_n, \phi^{-1}\Phi_n^{-1}) \\
\phi \mid \mathbf{Y} &\sim \mathbf{G}\left(\frac{\nu_n}{2}, \frac{\mathsf{SS}_n}{2}\right)
\end{aligned}
$$

Marginal Posterior Distribution of $\boldsymbol{\beta}$

$$
\begin{aligned}
\boldsymbol{\beta} \mid \phi, \mathbf{Y} &\sim \mathsf{N}(\mathbf{b}_n, \phi^{-1}\Phi_n^{-1}) \\
\phi \mid \mathbf{Y} &\sim \mathbf{G}\left(\frac{\nu_n}{2}, \frac{\mathsf{SS}_n}{2}\right)
\end{aligned}
$$

Let $\hat{\sigma}^2 = \mathsf{SS}_n/\nu_n$ (Bayesian MSE)

Marginal Posterior Distribution of $\boldsymbol{\beta}$

$$\begin{aligned} \boldsymbol{\beta} \mid \phi, \mathbf{Y} &\sim \mathsf{N}(\mathbf{b}_n, \phi^{-1}\Phi_n^{-1}) \\ \phi \mid \mathbf{Y} &\sim \mathbf{G}\left(\frac{\nu_n}{2}, \frac{\mathsf{SS}_n}{2}\right) \end{aligned}$$

Let $\hat{\sigma}^2 = \mathsf{SS}_n/\nu_n$ (Bayesian MSE)
Then the marginal posterior distribution of $\boldsymbol{\beta}$ is

$$\boldsymbol{\beta} \mid \mathbf{Y} \sim t_{\nu_n}(\mathbf{b}_n, \hat{\sigma}^2\Phi_n^{-1})$$

Marginal Posterior Distribution of $\boldsymbol{\beta}$

$$\begin{aligned}
\boldsymbol{\beta} \mid \phi, \mathbf{Y} &\sim \mathsf{N}(\mathbf{b}_n, \phi^{-1}\Phi_n^{-1}) \\
\phi \mid \mathbf{Y} &\sim \mathbf{G}\left(\frac{\nu_n}{2}, \frac{\mathsf{SS}_n}{2}\right)
\end{aligned}$$

Let $\hat{\sigma}^2 = \mathsf{SS}_n/\nu_n$ (Bayesian MSE)
Then the marginal posterior distribution of $\boldsymbol{\beta}$ is

$$\boldsymbol{\beta} \mid \mathbf{Y} \sim t_{\nu_n}(\mathbf{b}_n, \hat{\sigma}^2\Phi_n^{-1})$$

Any linear combination $\lambda^T\boldsymbol{\beta}$

$$\lambda^T\boldsymbol{\beta} \mid \mathbf{Y} \sim t_{\nu_n}(\lambda^T\mathbf{b}_n, \hat{\sigma}^2\lambda^T\Phi_n^{-1}\lambda)$$

has a univariate t distribution with ν_n degrees of freedom

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

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

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

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

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

Conjugate Priors

Definition

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

Conjugate Priors

Definition

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

Advantages:

Conjugate Priors

Definition

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

Advantages:

- Closed form distributions for most quantities; bypass MCMC for calculations

Conjugate Priors

Definition

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

Advantages:

- Closed form distributions for most quantities; bypass MCMC for calculations
- Simple updating in terms of sufficient statistics "weighted average"

Conjugate Priors

Definition

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

Advantages:

- Closed form distributions for most quantities; bypass MCMC for calculations
- Simple updating in terms of sufficient statistics "weighted average"
- Interpretation as prior samples - prior sample size

Conjugate Priors

Definition

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

Advantages:

- Closed form distributions for most quantities; bypass MCMC for calculations
- Simple updating in terms of sufficient statistics "weighted average"
- Interpretation as prior samples - prior sample size
- Elicitation of prior through imaginary or historical data

Conjugate Priors

Definition

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

Advantages:

- Closed form distributions for most quantities; bypass MCMC for calculations
- Simple updating in terms of sufficient statistics "weighted average"
- Interpretation as prior samples - prior sample size
- Elicitation of prior through imaginary or historical data
- limiting "non-proper" form recovers MLEs

Conjugate Priors

Definition

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

Advantages:

- Closed form distributions for most quantities; bypass MCMC for calculations
- Simple updating in terms of sufficient statistics "weighted average"
- 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 $\boldsymbol{\beta} \mid \phi \sim \mathsf{N}(\hat{\boldsymbol{\beta}}, n(\mathbf{X}^T\mathbf{X})^{-1}/\phi)$

Unit Information Prior

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

- Fisher Information is $\phi\mathbf{X}^T\mathbf{X}$ based on a sample of n observations

Unit Information Prior

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

- Fisher Information is $\phi\mathbf{X}^T\mathbf{X}$ based on a sample of n observations
- Inverse Fisher information is covariance matrix of MLE

Unit Information Prior

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

- 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 $\phi\mathbf{X}^T\mathbf{X}/n$

Unit Information Prior

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

- 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

Unit Information Prior

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

- 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

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

Unit Information Prior

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

- 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

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

- Posterior Distribution

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

Unit Information Prior

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

- 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

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

- Posterior Distribution

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

Cannot represent real prior beliefs; double use of data but has the "right" behaviour.

Zellner's g-prior

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

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) $\boldsymbol{\beta} \mid \phi \sim \mathsf{N}(\mathbf{b}_0, g(\mathbf{X}^T\mathbf{X})^{-1}/\phi)$

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

- Invariance: Require posterior of $\mathbf{X}\beta$ equal the posterior of $\mathbf{XH}\alpha$

Zellner's g-prior

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

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

- Invariance: Require posterior of $\mathbf{X}\beta$ equal the posterior of $\mathbf{X}\mathbf{H}\alpha$ ($\mathbf{a}_0 = \mathbf{H}^{-1}\mathbf{b}_0$) (take $\mathbf{b}_0 = \mathbf{0}$)
- Choice of g?

Zellner's g-prior

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

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

- Invariance: Require posterior of $\mathbf{X}\boldsymbol{\beta}$ equal the posterior of $\mathbf{XH}\boldsymbol{\alpha}$ ($\mathbf{a}_0 = \mathbf{H}^{-1}\mathbf{b}_0$) (take $\mathbf{b}_0 = \mathbf{0}$)
- Choice of g?
- $\frac{g}{1+g}$ weight given to the data

Zellner's g-prior

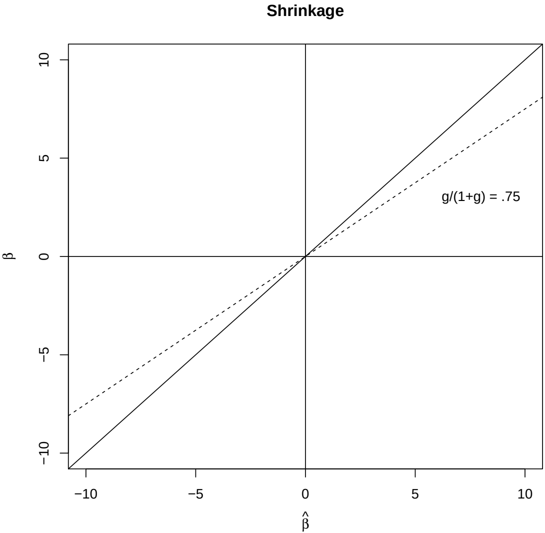
Zellner's g-prior(s) $\boldsymbol{\beta} \mid \phi \sim \mathsf{N}(\mathbf{b}_0, g(\mathbf{X}^T\mathbf{X})^{-1}/\phi)$

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

- Invariance: Require posterior of $\mathbf{X}\boldsymbol{\beta}$ equal the posterior of $\mathbf{X}\mathbf{H}\boldsymbol{\alpha}$ ($\mathbf{a}_0 = \mathbf{H}^{-1}\mathbf{b}_0$) (take $\mathbf{b}_0 = \mathbf{0}$)
- Choice of g?
- $\frac{g}{1+g}$ weight given to the data
- Fixed g effect does not vanish as $n \to \infty$
- Use $g = n$ or place a prior distribution on g

Shrinkage

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



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

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

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

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

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

Fisher Information Matrix

Log Likelihood

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

Fisher Information Matrix

Log Likelihood

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

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

Fisher Information Matrix

Log Likelihood

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

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

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

Fisher Information Matrix

Log Likelihood

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

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

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

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

Jeffreys Prior

Jeffreys Prior

$$p_J(\boldsymbol{\beta}, \phi) \quad \propto \quad |\mathcal{I}((\boldsymbol{\beta}, \phi)^T)|^{1/2}$$

Jeffreys Prior

Jeffreys Prior

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

Jeffreys Prior

Jeffreys Prior

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

Jeffreys Prior

Jeffreys Prior

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

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 $\int\int p_J(\boldsymbol{\beta}, \phi)\, d\boldsymbol{\beta}\, d\phi$ not finite

Formal Bayes Posterior

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

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

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

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

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

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

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

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

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

Limiting case of Conjugate prior with $\mathbf{b}_0 = 0$, $\Phi = \mathbf{0}$, $\nu_0 = 0$ and $\mathrm{SS}_0 = 0$
Posterior does not depend on dimension p;

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

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

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

Posterior does not depend on dimension p;

Jeffreys did not recommend using this

Independent Jeffreys Prior

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

Independent Jeffreys Prior

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

Independent Jeffreys Prior

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

Independent Jeffreys Prior

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

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

Independent Jeffreys Prior

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

$$\mathcal{I}((\boldsymbol{\beta}, \phi)^T) = \left[\begin{array}{cc} \phi(\mathbf{X}^T\mathbf{X}) & \mathbf{0}_p \\ \mathbf{0}_p^T & \frac{n}{2}\frac{1}{\phi^2} \end{array} \right]$$

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

Independent Jeffreys Prior

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

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

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

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

Independent Jeffreys Prior

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

$$\mathcal{I}((\boldsymbol{\beta}, \phi)^T) = \left[\begin{array}{cc} \phi(\mathbf{X}^T\mathbf{X}) & \mathbf{0}_p \\ \mathbf{0}_p^T & \frac{n}{2}\frac{1}{\phi^2} \end{array} \right]$$

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

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

Independent Jeffreys Prior is

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

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

Formal Posterior Distribution

With Independent Jeffreys Prior

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

Formal Posterior Distribution

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

Formal Posterior Distribution

With Independent Jeffreys Prior

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

Formal Posterior Distribution

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

Formal Posterior Distribution

With Independent Jeffreys Prior

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

Formal Posterior Distribution

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

Formal Posterior Distribution

With Independent Jeffreys Prior

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

Formal Posterior Distribution

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

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

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

Disadvantages of Conjugate Priors

Disadvantages:

Disadvantages of Conjugate Priors

Disadvantages:

- Results may have be sensitive to prior “outliers” due to linear updating

Disadvantages of Conjugate Priors

Disadvantages:

- Results may have be sensitive to prior “outliers” due to linear updating
- Cannot capture all possible prior beliefs

Disadvantages of Conjugate Priors

Disadvantages:

- Results may have be sensitive to prior “outliers” due to linear updating
- Cannot capture all possible prior beliefs
- Mixtures of Conjugate Priors