Non-Informative Priors

STA721 Linear Models Duke University

Merlise Clyde

September 29, 2019

Bayesian Estimation

Model

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

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

Difficulty with specifying hyperparameters in the Normal-Gamma prior in practice
Alternatives:

- Non-Informative Priors: Jeffreys' Priors
- g-prior $\mathsf{N}(0, \frac{g}{\phi}(\mathbf{X}^T\mathbf{X})^{-1})$
- Partitioned g-priors
- Zellner-Siow Cauchy Prior, mixtures and MCMC

Readings: Hoff Chapter 9

Non-Informative

What does it mean to be non-informative about $\boldsymbol{\beta}$ or about ϕ? Uniform distribution?

- Does it matter if we are non-informative about $\boldsymbol{\beta}$ versus $\boldsymbol{\mu}$
- Non-informative about ϕ?
- Non-informative about σ^2?
- Non-informative about σ?

These parameter spaces are unbounded so a uniform measure is not integrable on $\mathbb{R}^p$ or $\mathbb{R}^+$. Can these be justified?

Potential Problem with Uniform Measure

Take $p(\phi) \propto 1 d\phi$. What is $p(\sigma^2)$?

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

$$d\phi = [1/\sigma^2]^2 d\sigma^2$$

$$p(\sigma^2) = [1/\sigma^2]^2 d\sigma^2$$

Not uniform.

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

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

Fisher Information Matrix

Log Likelihood

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

$$\begin{aligned}
\frac{\partial^2 \log \mathcal{L}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} &= \left[\begin{array}{cc} -\phi(\mathbf{X}^T\mathbf{X}) & -(\mathbf{X}^T\mathbf{X})(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}) \\ -(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})^T(\mathbf{X}^T\mathbf{X}) & -\frac{n}{2}\frac{1}{\phi^2} \end{array}\right] \\
\mathsf{E}\left[\frac{\partial^2 \log \mathcal{L}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}\right] &= \left[\begin{array}{cc} -\phi(\mathbf{X}^T\mathbf{X}) & \mathbf{0}_p \\ \mathbf{0}_p^T & -\frac{n}{2}\frac{1}{\phi^2} \end{array}\right] \\
\mathcal{I}((\boldsymbol{\beta}, \phi)^T) &= \left[\begin{array}{cc} \phi(\mathbf{X}^T\mathbf{X}) & \mathbf{0}_p \\ \mathbf{0}_p^T & \frac{n}{2}\frac{1}{\phi^2} \end{array}\right]
\end{aligned}$$

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

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

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")
- $p_{IJ}(\boldsymbol{\beta}) \propto |\mathcal{I}(\boldsymbol{\beta})|^{1/2}$
- $p_{IJ}(\phi) \propto |\mathcal{I}(\phi)||^{1/2}$

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

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

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

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

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

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

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

Cannot represent anyone's prior beliefs, but used as a reference posterior