Bayesian Model Choice

Hoff Chapter 9, Liang et al 2007, Hoeting et al (1999), Clyde & George (2004) Statistical Science

October 31, 2022

Posterior Distribution

With a Normal-Gamma Prior $\mathrm{NG}(\mathrm{b}_0, \Phi_0, \nu_0, \mathrm{SS}_0)$, the posterior is $\mathrm{NG}(\mathrm{b}_n, \Phi_n, \nu_n, \mathrm{SS}_n))$: with

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

Posterior Distribution

With a Normal-Gamma Prior $\mathrm{NG}(b_0, \Phi_0, \nu_0, \mathrm{SS}_0)$, the posterior is $\mathrm{NG}(b_n, \Phi_n, \nu_n, \mathrm{SS}_n))$: with

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

and hyper-parameters

$$
\Phi_n = X^T X + \Phi_0
$$

Posterior Distribution

With a Normal-Gamma Prior $\mathsf{NG}(\mathsf{b}_0, \Phi_0, \nu_0, \mathsf{SS}_0)$, the posterior is $\mathsf{NG}(\mathsf{b}_n, \Phi_n, \nu_n, \mathsf{SS}_n))$: with

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

and hyper-parameters

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

Posterior Distribution

With a Normal-Gamma Prior $\mathrm{NG}(\mathrm{b}_0, \Phi_0, \nu_0, \mathrm{SS}_0)$, the posterior is $\mathrm{NG}(\mathrm{b}_n, \Phi_n, \nu_n, \mathrm{SS}_n))$: with

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

and hyper-parameters

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

Posterior Distribution

With a Normal-Gamma Prior $\mathrm{NG}(\mathrm{b}_0, \Phi_0, \nu_0, \mathrm{SS}_0)$, the posterior is $\mathrm{NG}(\mathrm{b}_n, \Phi_n, \nu_n, \mathrm{SS}_n))$: with

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

and hyper-parameters

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

Marginal Distribution from Normal–Gamma

Theorem

Let $\boldsymbol{\theta} \mid \phi \sim N(m, \frac{1}{\phi}\Sigma)$ *and* $\phi \sim \mathsf{G}(\nu/2, \nu\hat{\sigma}^2/2)$. *Then* $\boldsymbol{\theta}$ $(p \times 1)$ *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 $p(\boldsymbol{\theta}) = \int p(\boldsymbol{\theta} \mid \phi) p(\phi) \, d\phi$

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

Derivation

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

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

Derivation

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

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

Derivation

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

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

Derivation

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Any linear combination $\mathrm{x}^T\boldsymbol{\beta}$

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

has a univariate t distribution with v_n degrees of freedom

Predictive Distribution

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

Predictive Distribution

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

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

Predictive Distribution

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

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

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

Predictive Distribution

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

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

Predictive Distribution

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

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

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

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

Predictive Distribution

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

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

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

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

Predictive Distribution

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

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

Alternative Derivation

Conditional Distribution:

$$f(Y^* \mid Y) = \frac{f(Y^*, Y)}{f(Y)}$$

Alternative Derivation

Conditional Distribution:

$$
\begin{aligned}
f(Y^* \mid Y) &= \frac{f(Y^*, Y)}{f(Y)} \\
&= \frac{\int\int f(Y^*, Y \mid \boldsymbol{\beta}, \phi) p(\boldsymbol{\beta}, \phi) d\boldsymbol{\beta}\, d\phi}{f(Y)}
\end{aligned}
$$

Alternative Derivation

Conditional Distribution:

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

Alternative Derivation

Conditional Distribution:

$$\begin{aligned} f(Y^* \mid Y) &= \frac{f(Y^*, Y)}{f(Y)} \\ &= \frac{\int\int f(Y^*, Y \mid \boldsymbol{\beta}, \phi) p(\boldsymbol{\beta}, \phi) d\boldsymbol{\beta}\, d\phi}{f(Y)} \\ &= \frac{\int\int f(Y^* \mid \boldsymbol{\beta}, \phi) f(Y \mid \boldsymbol{\beta}, \phi) p(\boldsymbol{\beta}, \phi) d\boldsymbol{\beta}\, d\phi}{f(Y)} \\ &= \int\int f(Y^* \mid \boldsymbol{\beta}, \phi) p(\boldsymbol{\beta}, \phi \mid Y) d\boldsymbol{\beta}\, d\phi \end{aligned}$$

Alternative Derivation

Conditional Distribution:

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

Requires completing the square/quadratic!

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 $p(\boldsymbol{\theta}) \in \mathcal{P}$, $p(\boldsymbol{\theta} \mid \mathrm{Y}) \in \mathcal{P}$.

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 $p(\boldsymbol{\theta}) \in \mathcal{P}$, $p(\boldsymbol{\theta} \mid \mathrm{Y}) \in \mathcal{P}$.

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 $p(\boldsymbol{\theta}) \in \mathcal{P}$, $p(\boldsymbol{\theta} \mid \mathrm{Y}) \in \mathcal{P}$.

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 $p(\boldsymbol{\theta}) \in \mathcal{P}$, $p(\boldsymbol{\theta} \mid \mathrm{Y}) \in \mathcal{P}$.

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 $p(\boldsymbol{\theta}) \in \mathcal{P}$, $p(\boldsymbol{\theta} \mid \mathrm{Y}) \in \mathcal{P}$.

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 $p(\boldsymbol{\theta}) \in \mathcal{P}$, $p(\boldsymbol{\theta} \mid \mathrm{Y}) \in \mathcal{P}$.

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 $p(\boldsymbol{\theta}) \in \mathcal{P}$, $p(\boldsymbol{\theta} \mid \mathrm{Y}) \in \mathcal{P}$.

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 $p(\boldsymbol{\theta}) \in \mathcal{P}$, $p(\boldsymbol{\theta} \mid \mathrm{Y}) \in \mathcal{P}$.

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?

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

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

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

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

Fisher Information Matrix

Log Likelihood

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

Fisher Information Matrix

Log Likelihood

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

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

Fisher Information Matrix

Log Likelihood

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

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

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

Fisher Information Matrix

Log Likelihood

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

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

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

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

Jeffreys Prior

Jeffreys Prior

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

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

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

Formal Bayes Posterior

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

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

Formal Bayes Posterior

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

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

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

Formal Bayes Posterior

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

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

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

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

Formal Bayes Posterior

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

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

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

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

Formal Bayes Posterior

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

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

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

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

Jeffreys did not recommend using this Posterior does not depend on dimension p

Independent Jeffreys Prior

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

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

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

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

Independent Jeffreys Prior

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

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

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

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

Independent Jeffreys Prior

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

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

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

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

Independent Jeffreys Prior 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

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

Formal Posterior Distribution

With Independent Jeffreys Prior

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

Formal Posterior Distribution

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

Formal Posterior Distribution

With Independent Jeffreys Prior

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

Formal Posterior Distribution

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

Formal Posterior Distribution

With Independent Jeffreys Prior

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

Formal Posterior Distribution

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

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

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

Zellner's g-prior

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

Zellner's g-prior

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

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

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 $\mathbf{X}\boldsymbol{\beta}$ equal the posterior of $\mathbf{X}\mathbf{H}\boldsymbol{\alpha}$

Zellner's g-prior

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

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

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

Zellner's g-prior

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

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

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

Zellner's g-prior

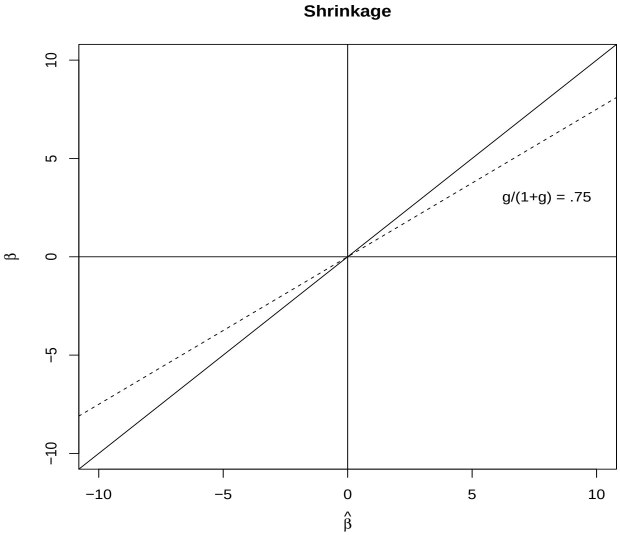
Zellner's g-prior(s) $\boldsymbol{\beta} \mid \phi \sim \mathsf{N}(\mathsf{b}_0, g(\mathsf{X}^T\mathsf{X})^{-1}/\phi)$

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

- Invariance: Require posterior of $\mathsf{X}\boldsymbol{\beta}$ equal the posterior of $\mathsf{XH}\boldsymbol{\alpha}$ ($\mathsf{a}_0 = \mathsf{H}^{-1}\mathsf{b}_0$) ($\mathsf{b}_0 = 0$)
- 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 diistribution on g

Shrinkage

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



Ridge Regression

- If $X^T X$ is nearly singular, certain elements of $\boldsymbol{\beta}$ or (linear combinations of $\boldsymbol{\beta}$) may have huge variances under the g-prior (or flat prior) as the MLEs are highly unstable!
- **Ridge regression** protects against the explosion of variances and ill-conditioning with the conjugate prior:

$$\boldsymbol{\beta} \mid \phi \sim \mathrm{N}(0, \frac{1}{\phi\lambda} \mathrm{I}_p)$$

- Posterior for $\boldsymbol{\beta}$ (conjugate case)

$$\boldsymbol{\beta} \mid \phi, \lambda, Y \sim \mathrm{N}\left((\lambda \mathrm{I}_p + X^T X)^{-1} X^T Y, \frac{1}{\phi} (\lambda \mathrm{I}_p + X^T X)^{-1} \right)$$

- induces shrinkage as well!

Model Choice ?

- Redundant variables lead to unstable estimates

Model Choice ?

- Redundant variables lead to unstable estimates
- Some variables may not be relevant ($\beta_j = 0$)

Model Choice ?

- Redundant variables lead to unstable estimates
- Some variables may not be relevant ($\beta_j = 0$)
- Can we infer a "good" model from the data?
- Expand model hierarchically to introduce another latent variable $\boldsymbol{\gamma}$ that encodes models $\mathcal{M}_\gamma$ $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots \gamma_p)^T$ where

$$
\begin{aligned}
\gamma_j = 0 &\Leftrightarrow \beta_j = 0 \\
\gamma_j = 1 &\Leftrightarrow \beta_j \neq 0
\end{aligned}
$$

- Find Bayes factors and posterior probabilities of models $\mathcal{M}_\gamma$
- 2^p models!

Zellner's g-prior

Centered model:

$$Y = 1_n \alpha + X^c \beta + \epsilon$$

where X^c is the centered design matrix where all variables have had their mean subtracted

Zellner's g-prior

Centered model:

$$Y = 1_n \alpha + X^c \beta + \epsilon$$

where X^c is the centered design matrix where all variables have had their mean subtracted

- $p(\alpha, \phi) \propto 1/\phi$

Zellner's g-prior

Centered model:

$$\mathsf{Y} = 1_n\alpha + \mathsf{X}^c\boldsymbol{\beta} + \epsilon$$

where X^c is the centered design matrix where all variables have had their mean subtracted

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

Zellner's g-prior

Centered model:

$$Y = 1_n\alpha + X^c\boldsymbol{\beta} + \epsilon$$

where X^c is the centered design matrix where all variables have had their mean subtracted

- $p(\alpha, \phi) \propto 1/\phi$
- $\boldsymbol{\beta}_\gamma \mid \alpha, \phi, \boldsymbol{\gamma} \sim N(0, g\phi^{-1}(X_{\boldsymbol{\gamma}}^{c\prime}X_{\boldsymbol{\gamma}}^c)^{-1})$

which leads to marginal likelihood of $\boldsymbol{\gamma}$ that is proportional to

$$p(Y \mid \gamma) = C(1+g)^{\frac{n-p-1}{2}}(1 + g(1 - R_\gamma^2))^{-\frac{(n-1)}{2}}$$

where R^2 is the usual coefficient of determination for model $\mathcal{M}_\gamma$.

Zellner's g-prior

Centered model:

$$\mathrm{Y} = 1_n\alpha + \mathrm{X}^c\boldsymbol{\beta} + \epsilon$$

where X^c is the centered design matrix where all variables have had their mean subtracted

- $p(\alpha, \phi) \propto 1/\phi$
- $\boldsymbol{\beta}_\gamma \mid \alpha, \phi, \boldsymbol{\gamma} \sim \mathrm{N}(0, g\phi^{-1}(\mathrm{X}_{\boldsymbol{\gamma}}^{c\,\prime}\mathrm{X}_{\boldsymbol{\gamma}}^c)^{-1})$

which leads to marginal likelihood of $\boldsymbol{\gamma}$ that is proportional to

$$p(\mathrm{Y} \mid \boldsymbol{\gamma}) = C(1+g)^{\frac{n-p-1}{2}}(1+g(1-R_\gamma^2))^{-\frac{(n-1)}{2}}$$

where R^2 is the usual coefficient of determination for model $\mathcal{M}_\gamma$. Trade-off of model complexity versus goodness of fit

Lastly, assign distribution to space of models

Sketch

- Integrate out β_γ using sums of normals

Sketch

- Integrate out β_γ using sums of normals
- Find inverse of $\mathrm{I}_n + g\mathrm{P}_{\mathrm{X}_\gamma}$ (properties of projections)

Sketch

- Integrate out $\boldsymbol{\beta}_\gamma$ using sums of normals
- Find inverse of $\mathsf{I}_n + g\mathsf{P}_{\mathsf{X}_\gamma}$ (properties of projections)
- Find determinant of $\phi(\mathsf{I}_n + g\mathsf{P}_{\mathsf{X}_\gamma})$

Sketch

- Integrate out $\boldsymbol{\beta}_\gamma$ using sums of normals
- Find inverse of $\mathbf{I}_n + g\mathbf{P}_{\mathbf{X}_\gamma}$ (properties of projections)
- Find determinant of $\phi(\mathbf{I}_n + g\mathbf{P}_{\mathbf{X}_\gamma})$
- Integrate out intercept (normal)

Sketch

- Integrate out β_γ using sums of normals
- Find inverse of $\mathsf{I}_n + g\mathsf{P}_{\mathsf{X}_\gamma}$ (properties of projections)
- Find determinant of $\phi(\mathsf{I}_n + g\mathsf{P}_{\mathsf{X}_\gamma})$
- Integrate out intercept (normal)
- Integrate out ϕ (gamma)

Sketch

- Integrate out $\boldsymbol{\beta}_\gamma$ using sums of normals
- Find inverse of $\mathrm{I}_n + g\mathrm{P}_{\mathrm{X}_\gamma}$ (properties of projections)
- Find determinant of $\phi(\mathrm{I}_n + g\mathrm{P}_{\mathrm{X}_\gamma})$
- Integrate out intercept (normal)
- Integrate out ϕ (gamma)
- algebra to simplify quadratic forms to R^2_γ

Or integrate α, $\boldsymbol{\beta}_\gamma$ and ϕ (complete the square!)

Priors on Model Space

$p(\mathcal{M}_\gamma) \Leftrightarrow p(\gamma)$

- $p(\gamma_j = 1) = .5 \Rightarrow P(\mathcal{M}_\gamma) = .5^p$ Uniform on space of models

Priors on Model Space

$p(\mathcal{M}_\gamma) \Leftrightarrow p(\gamma)$

- $p(\gamma_j = 1) = .5 \Rightarrow P(\mathcal{M}_\gamma) = .5^p$ Uniform on space of models
  $p_\gamma \sim \text{Bin}(p, .5)$
- $\gamma_j \mid \pi \overset{\text{iid}}{\sim} \text{Ber}(\pi)$ and $\pi \sim \text{Beta}(a, b)$ then $p_\gamma \sim \text{BB}_p(a, b)$

$$p(p_\gamma \mid p, a, b) = \frac{\Gamma(p+1)\Gamma(p_\gamma + a)\Gamma(p - p_\gamma + b)\Gamma(a+b)}{\Gamma(p_\gamma + 1)\Gamma(p - p_\gamma + 1)\Gamma(p + a + b)\Gamma(a)\Gamma(b)}$$

- $p_\gamma \sim \text{BB}_p(1, 1) \sim \text{Unif}(0, p)$

Posterior Probabilities of Models

- Calculate analytically under enumeration

$$p(\mathcal{M}_\gamma \mid \mathsf{Y}) = \frac{p(\mathsf{Y} \mid \gamma)p(\gamma)}{\sum_{\gamma' \in \Gamma} p(\mathsf{Y} \mid \gamma')p(\gamma')}$$

  Express as a function of Bayes factors and prior odds!
- Use MCMC over Γ - Gibbs, Metropolis Hastings
- Do we need to run MCMC over γ, $\boldsymbol{\beta}_\gamma$, α, and ϕ?

Inference/Decisions ?