Prior Distributions for GLMs

We'll assume that $\phi = 1$. A commonly used noninformative prior for $\boldsymbol{\beta}$ is the uniform prior, i.e., $\pi(\boldsymbol{\beta}) \propto 1$. In this case

$$\pi(\boldsymbol{\beta}|\boldsymbol{y}) \propto f(\boldsymbol{y}|\boldsymbol{\beta})$$

and the posterior mode is the MLE. We also have

$$\pi(\boldsymbol{\beta}|\boldsymbol{y}) \approx N(\widehat{\boldsymbol{\beta}}, I(\widehat{\boldsymbol{\beta}})^{-1}),$$

where

$$I(\widehat{\boldsymbol{\beta}}) = \boldsymbol{X}^T \boldsymbol{\Delta} \boldsymbol{V} \boldsymbol{\Delta} \boldsymbol{X} \Big|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}}.$$

Another noninformative prior for $\boldsymbol{\beta}$ is the Jeffreys prior, which is

$$\pi(\boldsymbol{\beta}) \propto |\boldsymbol{X}^T \boldsymbol{\Delta} \boldsymbol{V} \boldsymbol{\Delta} \boldsymbol{X}|^{1/2}.$$

Note that both $\boldsymbol{\Delta}$ and $\boldsymbol{V}$ depend on $\boldsymbol{\beta}$. So, in general, the Jeffreys prior for GLMs is quite different than the uniform prior.

For the binomial model, Jeffreys prior is proper for any link function. For the normal, Poisson, and gamma GLMs, the Jeffreys prior is improper for every link function.

Ibrahim and Laud (1991, *JASA*):

> article on Jeffreys prior for GLMs and its implications on the posterior

Informative priors for $\boldsymbol{\beta}$

Most commonly used informative prior for $\boldsymbol{\beta}$ is a normal prior, i.e.,

$$\boldsymbol{\beta} \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0).$$

Can also use a (modified) likelihood as a prior if historical data are available. Let the historical data be $(n_0, \boldsymbol{y}_0, \boldsymbol{X}_0)$. Then we may take

$$\pi(\boldsymbol{\beta}) \propto \exp\left\{ a_0 \sum_{i=1}^{n_0} \left[y_{0i}\theta(\boldsymbol{x}_{0i}^T\boldsymbol{\beta}) - b\left(\theta(\boldsymbol{x}_{0i}^T\boldsymbol{\beta})\right)\right]\right\},$$

where a_0 is a specified hyperparameter. This prior is computationally challenging, but can be approximated by a normal prior.

For n_0 large,

$$\pi(\boldsymbol{\beta}|a_0) \approx N(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\Sigma}}/a_0),$$

where $\tilde{\boldsymbol{\beta}}$ is the MLE of $\boldsymbol{\beta}$ based on the historical data,

$$\tilde{\boldsymbol{\Sigma}} = (\boldsymbol{X}_0^T \boldsymbol{\Delta}_0 \boldsymbol{V}_0 \boldsymbol{\Delta}_0 \boldsymbol{X}_0)^{-1} \Big|_{\boldsymbol{\beta}=\tilde{\boldsymbol{\beta}}},$$

and $\boldsymbol{\Delta}_0$ and $\boldsymbol{V}_0$ are $n_0 \times n_0$ versions of $\boldsymbol{\Delta}$ and $\boldsymbol{V}$ that use covariates in $\boldsymbol{X}_0$, the historical covariate matrix.

Asymptotic Normality of Posterior

For large n

$$\pi(\boldsymbol{\beta}|\boldsymbol{y}) \approx N(\boldsymbol{\beta}^*, \boldsymbol{\Sigma}^*),$$

where $\boldsymbol{\beta}^*$ is the mode of the posterior and

$$\boldsymbol{\Sigma}^* = \left(-\frac{\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T} \log \pi^*(\boldsymbol{\beta}|\boldsymbol{y}) \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*} \right)^{-1},$$

where $\pi^*(\boldsymbol{\beta}|\boldsymbol{y})$ is the unnormalized posterior, i.e.,

$$\pi^*(\boldsymbol{\beta}|\boldsymbol{y}) = f(\boldsymbol{y}|\boldsymbol{\beta})\pi(\boldsymbol{\beta}).$$

The posterior mode is obtained as the solution of

$$\frac{\partial}{\partial \boldsymbol{\beta}} \log \pi^*(\boldsymbol{\beta}|\boldsymbol{y}) = \mathbf{0}.$$

Remarks

1. Obviously

$$-\frac{\partial^2}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \log \pi^*(\boldsymbol{\beta}|\boldsymbol{y}) = \boldsymbol{X}^T \boldsymbol{\Delta} \boldsymbol{V} \boldsymbol{\Delta} \boldsymbol{X} - \frac{\partial^2}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \log \pi(\boldsymbol{\beta}).$$

   (a) If $\pi(\boldsymbol{\beta}) \propto 1$, then

$$\frac{\partial^2}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \log \pi(\boldsymbol{\beta}) = \mathbf{0}.$$

(b) If $\boldsymbol{\beta} \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$, then

$$-\frac{\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T} \log \pi(\boldsymbol{\beta}) = \boldsymbol{\Sigma}_0^{-1},$$

and so

$$-\frac{\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T} \log \pi^*(\boldsymbol{\beta}|\boldsymbol{y}) = \boldsymbol{X}^T\boldsymbol{\Delta}\boldsymbol{V}\boldsymbol{\Delta}\boldsymbol{X}+\boldsymbol{\Sigma}_0^{-1}.$$

2.(a) If $\pi(\boldsymbol{\beta}) \propto 1$, then $\boldsymbol{\beta}^* = \widehat{\boldsymbol{\beta}} =$ MLE of $\boldsymbol{\beta}$.

(b) If $\boldsymbol{\beta} \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$, then

$$\frac{\partial}{\partial\boldsymbol{\beta}} \log \pi^*(\boldsymbol{\beta}|\boldsymbol{y}) = \boldsymbol{0}$$

is equivalent to

$$\boldsymbol{X}^T\boldsymbol{\Delta}\boldsymbol{S} - \boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu}_0) = \boldsymbol{0}.$$