STAT6494: Adv Topics Bayesian Statistics

Inyoung Kim

Outline

Prior Distributions for GLM

- Let us first consider classes of priors for exponential family models.
- Let us assume $\phi = 1$ thoughout, since this is the case for the binomial, possion, and exponential distributions.
- The exponential family density can be

$$p(y|\theta) = exp\{y\theta - b(\theta) + c(y)\} \tag{1}$$

- We can develop a class of conjugate priors for the exponential family. We are led to the following theorem.

Prior Distributions for GLM

Theorem:
Suppose $p(y|\theta)$ is given by (1). Suppose we consider a prior on θ of the form

$$\pi(\theta) \propto exp\{a_0(\theta y_0 - b(\theta))\} \tag{2}$$

where (a_0, y_0) are specified hyperparameters. Then $\pi(\theta)$ is a conjugate prior for θ, which indexes the exponential family $p(y|\theta)$ (Note: Conjugate priors are proper priors by definition)

Noninformative Priors for β

- The **uniform prior**, i.e., $\pi(\beta) \propto 1$ is always an attractive noninformative prior for β, since it yields a posterior with parameters **that match frequentist point estimates**.
- Thus with $\pi \propto 1$, the posterior of β is $p(\beta|y) \propto L(\beta)$, and in this case the posteior model of β **equals** the maximum likelihood estimates of β. Morever, by the **Bayesian central limit theorem**,

  $$\beta|y \longrightarrow N(\hat{\beta}, I^{-1}(\hat{\beta}))$$

  as $n \longrightarrow \infty$, where

  $$I(\beta) = X' \Delta V \Delta X|_{\beta=\hat{\beta}}$$

  and $\hat{\beta}$ is the MLE of β
- Thus $\pi(\beta) \propto 1$ has lots of desirable properties as a noninformative prior for β

Noninformative Priors for β

- Another noninformative prior for β is Jeffrey's prior, given by $\pi(\beta) \propto |I(\beta)|^{1/2}$
- For GLMS's $I(\beta) \propto |X' \Delta V \Delta X|^{1/2}$
- Note that Δ and V depend on β. That is, $\Delta \equiv \Delta(\beta)$, $V \equiv V(\beta)$.
- We see that Jeffrey's prior for GLM's is, in general, **quite** different from the **uniform** prior.
- It can be showen that for the **binomial model**, Jeffrey's prior for β is **proper** for any link function.
- For the normal, Poisson, and gamma regression models, Jeffrey's prior for β is an **improper** prior for any link function.

Noninformative Priors for β

- Note that for the usual normal model,

$$\begin{aligned} Y &\sim N_n(X\beta, \sigma^2 I) \\ V &= \sigma^2 I, \Delta = I \end{aligned}$$

  If σ^2 is assumed known, then

$$\pi(\beta) \propto |\sigma^2(X'X)|^{1/2} = constant$$

  so that Jeffrey's prior is a uniform prior for β in this case.

- For the binomial, Poission, and gamma regression models, Jeffrey's prior for β is **NOT** a uniform prior.
- An article characterizing Jeffreys's prior for GLM's and its implications on the posteior can be found in Ibrahim and Laud(1991, JASA)

Informative Priors for β

- The most common type of informative prior for β in GLM's is a **normal prior**, i.e., $\beta \sim N(\mu_0, \Sigma_0)$
- If a previous study with historical data $D_0 = (n_0, y_0, X_0)$, then we can take

$$\begin{aligned} \pi(\beta|a_0) &\propto [L(\beta|y_0)]^{a_0} \\ &= exp[a_0 \sum_{i=1}^{n_0} \{y_{0i}\theta(x_{0i}^{'}\beta) - b(\theta(x_{0i}^{'}\beta))\}] \end{aligned} \quad (3)$$

  where $y_0 = (y_{01}, ..., y_{0n_0})'$ is the $n_0 \times 1$ vector of response for the historical data, X_0 is an $n_0 \times p$ matrix of covariates with ith row $x_{0i}^{'}$, n_0 is the sample size for the historical data, and a_0 is a specified hyperparameter.
- Note that even when a_0 is fixed, this prior does NOT have a closed form β
- This is a computationally challenginig prior to work with, in general.

Informative Priors for β

- We can approximate the prior in (3) by a normal distribution. That is, as $n_0 \longrightarrow \infty$, one can show that

$$\pi(\beta|a_0) \longrightarrow N_p(\tilde{\beta}, a_0^{-1}\tilde{\Sigma})$$

  where $\tilde{\beta}$ is the MLE of β based on the historical data $D_0(n_0, y_0, X_0)$ and $\tilde{\Sigma} = (X_0' \Delta_0' V_0' \Delta_0')^{-1}|_{\beta=\tilde{\beta}}$ where X_0 is the $n_0 \times p$ covariates matrix for the histrorical data Δ_0 and V_0 are Δ and V using the covariates from the historical data.

- The approximation in (4) is computationally easier to wotk with than the prior in (3)

Informative Priors for β

- We can establish the Bayesian central limit theorem for GLM with an arbitrary prior $\pi(\beta)$ for β
- Theorem: Suppose $L(\beta)$ is the likelihood function of any GLM based on n observations, and suppose $\pi(\beta)$ is the prior for β. Then as $n \rightarrow \infty$,

$$\beta|y \rightarrow N_p(\beta^*, \Sigma^*)$$

where β^* is the posterior mode of β obtained by solving

$$\frac{\partial}{\partial \beta}[logp^*(\beta|y)] = 0$$

and

$$\Sigma^* = (-\frac{\partial^2}{\partial \beta \partial \beta'}[logp^*(\beta|y)]|_{\beta=\beta^*})^{-1}$$

where $p^*(\beta|y) = L(\beta)\pi(\beta)$ is the unnormalized posterior density of β

Property of the Posterior Distribution for GLM's

- If $\pi(\beta) \propto 1$ and the MLE of β exists for the GLM, then $p(\beta|y)$ is proper.
- With Jeffreys's prior, i.e. $\pi(\beta) \propto |X'\Delta V \Delta X|^{1/2}$, it can be shown that $p(\beta|y)$ is proper under some very general conditions.
- Note that Jeffreys's prior is improper for the Poission, gamma, and normal models. See Ibrahim and Laud (1991, JASA) for more details.
- If $\pi(\beta)$ is proper, then of course $p(\beta|y)$ is always proper for any GLM.