Gaussian Linear model: Conjugate Bayes

STA 215. Surya Tokdar

The Normal-Inverse-Chi-square distribution

Definition

The joint distribution of a random element $(W, V) \in \mathbb{R}^p \times \mathbb{R}_+$ is said to be a normal-inverse-chi-sqaure distribution if for some $m \in \mathbb{R}^p$, some $p \times p$ p.d. matrix K and some $r > 0$, $s > 0$,

1. $\frac{rs^2}{V} \sim \chi^2(r)$
2. $W|(V = v) \sim N_p(m, vK^{-1})$

We will denote this distribution by $N_p\chi^{-2}(m, K, r, s^2)$.

Pdf calculation

The conditional pdf of W given $V = v$ is:

$$f_{W|V}(w|v) = \text{const.} \times v^{-p/2} \exp\left\{ -\frac{(w-m)^T K(w-m)}{2v} \right\}$$

The pdf of V is

$$f_V(v) = \text{const.} \times v^{-(r+2)/2} \exp\left\{ -\frac{rs^2}{2v} \right\}$$

So the joint pdf of (W, V) is

$$f_{W,V}(w, v) = \text{const.} \times v^{-\frac{r+p+2}{2}} \exp\left\{ -\frac{(w-m)^T K(w-m) + rs^2}{2v} \right\}$$

Some important properties

If $(W, V) \sim N_p\chi^{-2}(m, K, r, s^2)$ then for any vector $a \in \mathbb{R}^p$,

$$(a^T W, V) \sim N_1\chi^{-2}\left(a^T m, \frac{1}{a^T K^{-1} a}, r, s^2\right)$$

and,

$$\frac{a^T W - a^T m}{s\sqrt{a^T K^{-1} a}} \sim t_r$$

Conjugate prior for Gaussian linear model

The prior

Consider the Gaussian linear model (with $\dim(z_i) = p$)

$$Y_i = z_i^T \beta + \epsilon_i, \quad \epsilon_i \overset{\text{IID}}{\sim} N(0, \sigma^2)$$

A conjugate prior family for $\theta = (\beta, \sigma^2)$ is given the normal-inverse-chi-square pdfs

$$N_p \chi^{-2}(m_0, K_0, r_0, s_0^2)$$

where m_0 ranges over all p dimensional vectors, K_0 ranges over all $p \times p$ positive definite matrices, r_0 and s_0 range over all positive numbers.

Conjugacy

To prove conjugacy, first note that the both in the likelihood

$$L_x(\beta, \sigma^2) \propto (\sigma^2)^{-\frac{n}{2}} \exp\left\{ -\frac{(y - Z\beta)^T (y - Z\beta)}{2\sigma^2} \right\}$$

and in the prior

$$\pi(\beta, \sigma^2) \propto (\sigma^2)^{-\frac{r_0+p+2}{2}} \exp\left\{ -\frac{(\beta - m_0)^T K_0 (\beta - m_0) + r_0 s_0^2}{2\sigma^2} \right\}$$

β appears through a quadratic term in the exponent. Therefore multiplying we get

$$\pi(\beta, \sigma^2 | x) \propto (\sigma^2)^{-\frac{r_n+p+2}{2}} \exp\left\{ -\frac{(\beta - m_n)^T K_n (\beta - m_n) + r_n s_n^2}{2\sigma^2} \right\},$$

for some m_n, K_n, r_n, s_n because adding two quadratic gives another quadratic.

So the posterior pdf is $N_p \chi^{-2}(m_n, K_n, r_n, s_n^2)$, and these four quantities can be identified as

1. $m_n = (K_0 + Z^T Z)^{-1}(K_0 m_0 + Z^T y)$
2. $K_n = K_0 + Z^T Z$
3. $r_n = r_0 + n$
4. $s_n^2 = \frac{r_0 s_0^2 + y^T y + m_0^T K m_0 - m_n^T K_n m_n}{r_0 + n}$.

When $n > p$, so that we can define $\hat{\beta}_{\text{LS}} = (Z^T Z)^{-1} Z^T y$ and $s_{y|z}^2 = \frac{1}{n-p} \sum_{i=1}^n (y_i - z_i^T \hat{\beta}_{\text{LS}})^2$, we can re-write: $m_n = (K_0 + Z^T Z)^{-1}(K_0 m_0 + Z^T Z \hat{\beta}_{\text{LS}})$ and $r_n s_n^2 = r_0 s_0^2 + (n-p) s_{y|z}^2 + (\hat{\beta}_{\text{LS}} - m_0)^T (K_0^{-1} + (Z^T Z)^{-1})^{-1} (\hat{\beta}_{\text{LS}} - m_0)$.

Important special case: normal data

For $X_i \overset{\text{IID}}{\sim} N(\mu, \sigma^2)$ we have $\hat{\beta}_{\text{LS}} = \bar{x}$, $Z^T Z = n$, $p = 1$ and $s_x^2 = \frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})^2$. Now $K_0 = k_0$ is a positive scalar. So, the posterior pdf of (μ, σ^2) is $N_1\chi^{-2}(m_n, k_n, r_n, s_n^2)$ where:

1. $m_n = \frac{k_0 m_0 + n\bar{x}}{k_0 + n}$
2. $k_n = k_0 + n$
3. $r_n = r_0 + n$
4. $s_n^2 = \frac{r_0 s_0^2 + (n-1)s_x^2 + \frac{k_0 n}{k_0 + n}(\bar{x} - m_0)^2}{r_0 + n}$

Posterior summaries

Clearly inference on σ^2 can be summarized based on the marginal posterior distribution of σ^2 given data. By definition of normal-inverse-chi-square, this marginal posterior can be written as $r_n s_n^2/\sigma^2 \sim \chi^2(r_n)$.

To draw inference on $\eta = a^T\beta$, for some $a \in \mathbb{R}^p$, we need the marginal posterior of η given data (and integrating out σ^2). By the special property of normal-inverse-chi-squares distributions mentioned before, the posterior distribution of

$$\frac{\eta - a^T m_n}{s_n\sqrt{a^T K_n^{-1} a}}$$

is $t(r_n)$. Hence a $(1-\alpha)$ posterior credible interval for η is

$$a^T m_n \mp s_n\sqrt{a^T K_n^{-1} a} \times z_{r_n}(\alpha).$$