Chapter 10: Multiple Regression: Bayesian Inference

This chapter considers Bayesian estimation and prediction for the multiple linear regression model in which $x$ variables are fixed constants.

1 Elements of Bayesian Statistical Inference

Let $f(y|\theta)$ denote the joint likelihood function of $y_1, y_2, \ldots, y_n$, and let $p(\theta)$ denote the prior distribution imposed on $\theta$. Then the posterior distribution of $\theta$ is given by

$$
\pi(\theta|y) = \frac{f(y|\theta)p(\theta)}{\int f(y|\theta)p(\theta)d\theta} = \frac{1}{c(y)}f(y|\theta)p(\theta),
$$

where $c(y) = \int f(y|\theta)p(\theta)d\theta$ is the normalizing constant of the posterior distribution.

Bayesian inference for the model is always based on the posterior distribution $\pi(\theta|y)$. For example, let $q(y_0|\theta)$ denote a prediction function for $y_0$ given $\theta$. Then

$$
r(y_0|y) = \int q(y_0|\theta)\pi(\theta|y)d\theta,
$$

which leads to a conditional inference formula.

2 A Bayesian Multiple Linear Regression Model

For convenience, one often parameterizes Bayesian models using precision ($\tau$) rather than variance ($\sigma^2$). With this reparameterization, one often assumes

$$
\begin{align*}
    y|\beta, \tau &\sim N_n(X\beta, \frac{1}{\tau}I), \\
    \beta|\tau &\sim N_{k+1}(\phi, \frac{1}{\tau}V), \\
    \tau &\sim Gamma(\alpha, \delta),
\end{align*}
$$

where $\alpha$ and $\delta$ are prior-hyperparameters.
Theorem 2.1. Consider the Bayesian multiple regression model, for which the prior distributions are as specified in (1). Then the joint prior distribution is conjugate, that is, \( \pi(\beta, \tau|y) \) is of the same form as \( \pi(\beta, \tau) \).

Theorem 2.2. Consider the Bayesian multiple regression model, for which the prior distributions are as specified in (1). The marginal posterior distribution \( \pi(\beta|y) \) is a multivariate \( \text{t} \) distribution with parameters \( (n+2\alpha, \phi_*, W_*) \), where

\[
\phi_* = (V^{-1} + X'X)^{-1}(V^{-1}\phi + X'y),
\]

and

\[
W_* = \left( \frac{(y - X\phi)'(I + XV X')^{-1}(y - X\phi) + 2\delta}{n + 2\alpha} \right) (V^{-1} + X'X)^{-1}.
\]

Theorem 2.3. Consider the Bayesian multiple regression model, for which the prior distributions are as specified in (1). The marginal posterior distribution \( \pi(\tau|y) \) is a gamma distribution with parameters \( \alpha + n/2 \) and \( (-\phi_0 V_*^{-1} \phi + \phi' V_*^{-1} \phi + y'y + 2\delta)/2 \), where \( V_* = (V^{-1} + X'X)^{-1} \) and \( \phi_* = V_*(V^{-1}\phi + X'y) \).

3 Inference in Bayesian Multiple Linear Regression

Point Estimate and Credible Interval A convenient property of the multivariate \( \text{t} \)-distribution is that linear functions of the random vector follow the (univariate) \( \text{t} \)-distribution. Thus, given \( y \),

\[
\frac{a'\beta - a'\phi_*}{a'W_*a} \sim t(n + 2\alpha),
\]

and, as an important special case,

\[
\frac{\beta_i - \phi_{si}}{w_{sii}} \sim t(n + 2\alpha),
\]

where \( \phi_{si} \) is the \( i \)th element of \( \phi_* \) and \( w_{sii} \) is the \( i \)th diagonal element of \( W_* \). Thus a Bayesian point estimate of \( \beta_i \) is its posterior mean \( \phi_{si} \) and a \( 100(1-\omega)\% \) Bayesian credible interval for \( \beta_i \) is

\[
\phi_{si} + t_{\omega/2, n+2\alpha} w_{sii}.
\]

Hypothesis Test For example, to test the hypothesis test \( \beta_i > \beta_{i0} \), we can calculate the probability

\[
P \left( t(n + 2\alpha) > \frac{\beta_{i0} - \phi_{si}}{w_{sii}} \right).
\]

The larger the probability is, the more credible is the hypothesis.
Special cases of Inference  First, we consider the use of a diffuse prior. Let $\phi = 0$, let $V$ be a diagonal matrix with all diagonal elements equal to a large constant (say, $10^6$), and let $\alpha$ and $\delta$ both be equal to a small constant (say, $10^{-6}$). In this case, $V^{-1}$ is close to 0, and so $\phi^*$, and the Bayesian point estimate of $\beta$ in (2) is approximately equal to

$$(X'X)^{-1}X'y.$$ 

Also, since $(I + XVX')^{-1} = I - X(X'X + V)^{-1}X'$, the covariance matrix $W_*$ approaches

$$W_* = \frac{y'(I - X(X'X)^{-1}X')y}{n} (X'X)^{-1} = \frac{n - 1}{n} s^2 (X'X)^{-1}.$$ 

The second special case of inference is the case in which $\phi = 0$ and $V$ is a diagonal matrix with a constant on the diagonal. Thus $V = aI$, where $a$ is a positive number, and the Bayesian estimator of $\beta$ becomes

$$(X'X + \frac{1}{a}I)^{-1}X'y,$$ 

which is known as the ridge estimator.

Bayesian Point and Interval Estimation of $\sigma^2$ A possible Bayesian point estimator of $\sigma^2$ is the mean of the marginal inverse gamma density given in Theorem 7:

$$\frac{(-\phi'V_*^{-1}\phi_* + \phi'V^{-1}\phi + y'y + 2\delta)/2}{\alpha + n/2 - 1},$$

and a 100$(1 - \omega)$% Bayesian credible interval for $\sigma^2$ is given by the $1 - \omega/2$ and $\omega/2$ quantiles of the inverse gamma distribution.

Consider a special case that $\alpha$ and $\delta$ are both close to 0, $\phi = 0$, and $V$ is a diagonal matrix with all diagonal elements equal to a large constant. Then the Bayesian point estimator of $\sigma^2$ is approximately

$$\frac{(y'y - \phi_*V_*^{-1}\phi_*)/2}{n/2 - 1} = \frac{y'y - y'X(X'X)^{-1}X'y}{n - 2} = \frac{n - k - 1}{n - 2} s^2.$$ 

4  Bayesian Inference through MCMC Simulations

Consider the Bayesian multiple regression model, for which the prior distributions are as specified in (1). Then the conditional distribution of $\beta|\tau$, $y$ is $N_{k+1}(\phi_*, \tau^{-1}V_*)$, and the conditional distribution of $\tau|\beta, y \sim Gamma((\alpha_{ss} + k + 3)/2, ([\beta - \phi_*/V_*^{-1}(\beta - \phi_*) + \delta_{ss}]/2)$, where $\alpha_{ss} = 2\alpha - 2 + n$, and $\delta_{ss} = -\phi'_V^{-1}\phi_* + \phi'V^{-1}\phi + y'y + \delta_*$. Then the Gibbs sampler can be used for simulating from the posterior distribution.