Chapter 10: Multiple Regression: Bayesian Inference

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

$$\begin{aligned}\pi(\boldsymbol{\theta}|\boldsymbol{y}) &= \frac{f(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int f(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}} \\ &= \frac{1}{c(\boldsymbol{y})}f(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}),\end{aligned}$$

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

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

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

which leads to a conditional inference formula.

2 A Bayesian Multiple Linear Regression Model

For convenience, one often parameterizes Bayesian models using precision (τ) rather than variance (σ^2). With this reparameterization, one often assumes

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

where α and δ 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,* $\pi(\boldsymbol{\beta}, \tau|\boldsymbol{y})$ *is of the same form as* $\pi(\boldsymbol{\beta}, \tau)$*.*

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

$$\boldsymbol{\phi}_* = (\boldsymbol{V}^{-1} + \boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{V}^{-1}\boldsymbol{\phi} + \boldsymbol{X}'\boldsymbol{y}), \tag{2}$$

and

$$\boldsymbol{W}_* = \left(\frac{(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\phi})'(\boldsymbol{I} + \boldsymbol{X}\boldsymbol{V}\boldsymbol{X}')^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\phi}) + 2\delta}{n + 2\alpha} \right) (\boldsymbol{V}^{-1} + \boldsymbol{X}'\boldsymbol{X})^{-1}. \tag{3}$$

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

3 Inference in Bayesian Multiple Linear Regression

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

$$\frac{\boldsymbol{a}'\boldsymbol{\beta} - \boldsymbol{a}'\boldsymbol{\phi}_*}{\boldsymbol{a}'\boldsymbol{W}_*\boldsymbol{a}} \sim t(n + 2\alpha),$$

and, as an important special case,

$$\frac{\beta_i - \phi_{*i}}{w_{*ii}} \sim t(n + 2\alpha),$$

where ϕ_{*i} is the ith element of $\boldsymbol{\phi}_*$ and w_{*ii} is the ith diagonal element of $\boldsymbol{W}_*$. Thus a Bayesian point estimate of β_i is its posterior mean ϕ_{*i} and a $100(1 - \omega)\%$ Bayesian credible interval for β_i is

$$\phi_{*i} + t_{\omega/2, n+2\alpha} w_{*ii}.$$

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

$$P\left(t(n + 2\alpha) > \frac{\beta_{i0} - \phi_{*i}}{w_{*ii}} \right).$$

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 $\boldsymbol{\phi} = 0$, let $\boldsymbol{V}$ be a diagonal matrix with all diagonal elements equal to a large constant (say, 10^6), and let α and δ both be equal to a small constant (say, 10^{-6}). In this case, $\boldsymbol{V}^{-1}$ is close to 0, and so $\boldsymbol{\phi}_*$, and the Bayesian point estimate of $\boldsymbol{\beta}$ in (2) is approximately equal to

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

Also, since $(\boldsymbol{I} + \boldsymbol{X}\boldsymbol{V}\boldsymbol{X}')^{-1} = \boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{V})^{-1}\boldsymbol{X}'$, the covariance matrix $\boldsymbol{W}_*$ approaches

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

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

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

which is known as the ridge estimator.

Bayesian Point and Interval Estimation of σ^2 A possible Bayesian point estimator of σ^2 is the mean of the marginal inverse gamma density given in Theorem **??**:

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

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

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

$$\frac{(\boldsymbol{y}'\boldsymbol{y} - \boldsymbol{\phi}_*\boldsymbol{V}_*^{-1}\boldsymbol{\phi}_*)/2}{n/2 - 1} = \frac{\boldsymbol{y}'\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}}{n-2} = \frac{n-k-1}{n-2}s^2.$$

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 $\boldsymbol{\beta}|\tau, \boldsymbol{y}$ is $N_{k+1}(\boldsymbol{\phi}_*, \tau^{-1}\boldsymbol{V}_*$, and the conditional distribution of $\tau|\boldsymbol{\beta}, \boldsymbol{y} \sim Gamma((\alpha_{**} + k + 3)/2, [(\boldsymbol{\beta} - \boldsymbol{\phi}_*)'\boldsymbol{V}_*^{-1}(\boldsymbol{\beta} - \boldsymbol{\phi}_*) + \delta_{**}]/2)$, where $\alpha_{**} = 2\alpha - 2 + n$, and $\delta_{**} = -\boldsymbol{\phi}_*'\boldsymbol{V}_*^{-1}\boldsymbol{\phi}_* + \boldsymbol{\phi}'\boldsymbol{V}^{-1}\boldsymbol{\phi} + \boldsymbol{y}'\boldsymbol{y} + \delta_*$. Then the Gibbs sampler can be used for simulating from the posterior distribution.