Bayesian Econometrics: Computational Statistics

Linear regression: Normal-Inverse Gamma model

Andrés Ramírez Hassan

Universidad Eafit
Departamento de Economía

March 15, 2021

Outline

The linear regression

The likelihood

$y_i = \beta_1 x_{i1} + \beta_1 x_{i2} + \ldots + \beta_1 x_{iK} + \mu_i = \mathrm{x}_i'\beta + \mu_i$ where $\mu_i \overset{i.i.d}{\sim} \mathcal{N}(0, \sigma^2)$ is an stochastic error such that $\mathrm{x}_i \perp \mu_i$.
Writing this model in matrix form we get $\mathrm{y} = \mathrm{X}\beta + \mu$ such that $\mu \sim \mathcal{N}(0, \sigma^2\mathrm{I})$ which implies that $\mathrm{y} \sim \mathcal{N}(\mathrm{X}\beta, \sigma^2\mathrm{I})$. So the likelihood function is

$$\begin{aligned} L(\mathrm{y}|\beta, \sigma^2, \mathrm{X}) &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)\right\} \\ &\propto (\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)\right\} \quad (1) \end{aligned}$$

The linear regression

Prior distribution

The conjugate priors for the parameters are:

$$\beta|\sigma^2 \sim \mathcal{N}(\beta_0, \sigma^2 B_0) \qquad (2)$$

$$\sigma^2 \sim \mathcal{IG}(\alpha_0/2, \delta_0/2) \qquad (3)$$

The linear regression

The posterior distributions

This means posterior distributions for β and σ^2 of the form

$$\beta|\sigma^2, \mathrm{y}, \mathrm{X} \sim \mathcal{N}(\beta^*, \sigma^2 B) \tag{4}$$
$$\sigma^2|\mathrm{y}, \mathrm{X} \sim \mathcal{IG}(\alpha^*/2, \delta^*/2) \tag{5}$$

where $B = (B_0^{-1} + X'X)^{-1}$, $W = (B_0^{-1} + X'X)^{-1}X'X$,

$$\beta^* = B(B_0^{-1}\beta_0 + X'y) = (I - W)\beta_0 + W\hat{\beta}, \tag{6}$$

$\alpha^* = \alpha_0 + n$ and

$$\begin{aligned}\delta^* &= \delta_0 + y'y + \beta_0' B_0^{-1}\beta_0 - \beta^{*\prime} B^{-1}\beta^* \\ &= \delta_0 + (n-k)\hat{\sigma}^2_{LSE} + (\hat{\beta} - \beta_0)'[(X'X)^{-1} + B_0]^{-1}(\hat{\beta} - \beta_0)\end{aligned}$$

The linear regression

Posterior marginal distribution for location parameters

$$\beta|y \sim t_k(\alpha^*, \beta^*, H),$$

where $H = \delta^*/\alpha^* B$.

$$\frac{\beta_j - \beta_j^*}{(h^{jj})^{1/2}}|\text{y}, \text{X} \sim t_{\alpha^*}$$

where h^{jj} is the jth diagonal element of H^{-1} and β_j^* is the jth element of β^*.

The linear regression

The marginal likelihood

$$p(\mathrm{y}) = \int_0^\infty \int_\beta \pi(\beta \mid \sigma^2 \mathrm{B}_0, \beta_0) \pi(\sigma^2 \mid \alpha_0/2, \delta_0/2) L(\mathrm{y}|\beta, \sigma^2, \mathrm{X}) \, d\sigma^2 d\beta$$

$$= t\left(\mathrm{X}\beta_0, \frac{\alpha_0(I + \mathrm{X}\mathrm{B}_0\mathrm{X}')}{\delta_0}, \delta_0\right)$$

The linear regression

The predictive distribution

$$y_0 \mid y, X, X_0 \sim t\left(X_0\beta^*, \frac{\alpha^*(I + X_0 B X_0')}{\delta^*}, \delta^*\right) \tag{7}$$

Gibbs Sampler

The Gibbs algorithm proposes the following transition kernel for two parameter blocks

$$p(\theta_1, \theta_2) = \pi(\theta_{22}|\theta_{21})\pi(\theta_{21}|\theta_{12})$$

where $\theta_1 = (\theta_{11}, \theta_{12})$ and $\theta_2 = (\theta_{21}, \theta_{22})$. We can see that in order for the Gibbs sampler to be of use, we must first obtain the conditional distributions of each parameter block in terms of the others.

Gibbs Sampler

Proof that the Gibbs kernel leads to the invariant distribution:

$$\begin{aligned}
\pi(\theta_2) &= \int \pi(\theta_1) p(\theta_1, \theta_2) d\theta_1 \\
&= \int \pi(\theta_{11}, \theta_{12}) \pi(\theta_{22}|\theta_{21}) \pi(\theta_{21}|\theta_{12}) d\theta_{11} d\theta_{12} \\
&= \pi(\theta_{22}|\theta_{21}) \int \pi(\theta_{21}|\theta_{12}) \pi(\theta_{11}, \theta_{12}) d\theta_{11} d\theta_{12} \\
&= \pi(\theta_{22}|\theta_{21}) \int \pi(\theta_{21}|\theta_{12}) \pi(\theta_{12}) d\theta_{12} \\
&= \pi(\theta_{22}|\theta_{21}) \int \pi(\theta_{21}, \theta_{12}) d\theta_{12} \\
&= \pi(\theta_{22}|\theta_{21}) \pi(\theta_{21}) = \pi(\theta_{22}, \theta_{21}) = \pi(\theta_2)
\end{aligned}$$

Gibbs Sampler

A word of caution on the careless use of the Gibbs sampler algorithm:

Caution

Even when the conditional distributions $\pi(\theta_{21}|\theta_{12})$ and $\pi(\theta_{22}|\theta_{21})$ are well defined and can be simulated from, the joint distribution $\pi(\theta_2)$ may not correspond to any proper distribution. This is specially true when using improper priors, so care is to be taken! (See Robert & Casella, 2004, section 10.4.3)

Algorithm

For two parameter blocks

1. Choose a starting value $\theta_2^{(0)}$.
2. At the first iteration, draw

$$\theta_1^{(1)} \text{ from } \pi(\theta_1|\theta_2^{(0)}),$$
$$\theta_2^{(1)} \text{ from } \pi(\theta_2|\theta_1^{(1)}).$$

3. At the gth iteration, draw

$$\theta_1^{(g)} \text{ from } \pi(\theta_1|\theta_2^{(g-1)}),$$
$$\theta_2^{(g)} \text{ from } \pi(\theta_2|\theta_1^{(g)}).$$

Algorithm

For d parameter blocks

1. Choose starting values $\theta_2^{(0)}, \dots, \theta_d^{(0)}$.
2. At the gth iteration, draw

$$\begin{aligned} \theta_1^{(g)} &\text{ from } \pi(\theta_1 | \theta_2^{(g-1)}, \dots, \theta_d^{(g-1)}), \\ \theta_2^{(g)} &\text{ from } \pi(\theta_2 | \theta_1^{(g)}, \theta_3^{(g-1)}, \dots, \theta_d^{(g-1)}), \\ &\vdots \\ \theta_d^{(g)} &\text{ from } \pi(\theta_d | \theta_1^{(g)}, \dots, \theta_{d-1}^{(g)}). \end{aligned}$$

Simulation Exercise

Initial setting for the simulation:

- $N = 1000$
- $\beta = (1.5, -3.5, 2)'$
- $x_1 \sim \mathcal{N}_N(0, 1)\,, x_2 \sim \mathcal{B}_N(0.5)\,, X = (1, x_1, x_2)$
- $y = X\beta + \mu\,, \quad \mu \sim \mathcal{N}_N(0, 1)$

Simulation Exercise

The Gibbs algorithm for this simulation is therefore

1. Choose a starting value $\sigma^{2(0)}$.
2. At the gth iteration, draw

$$\beta(g) \text{ from } \mathcal{N}_3(\beta^*, \sigma^{2(g)} B),$$
$$\sigma^{2(g)} \text{ from } \mathcal{IG}(\alpha^*/2, \delta^*/2).$$