Gibbs Sampling in Linear Models #1

Econ 690

Purdue University

Outline

1. Conditional Posterior Distributions for Regression Parameters in the Linear Model [Lindley and Smith (1972, JRSSB)]
2. Gibbs Sampling in the Classical Linear Regression Model
   - Application to our Wage Data
3. Gibbs Sampling in a Linear Regression Model with Groupwise Heteroscedasticity

- Before discussing applications of Gibbs sampling in several different linear models, we must first prove an important result that will assist us in deriving needed conditional posterior distributions.
- This result is relevant for the derivation of posterior conditional distributions for regression parameters, and makes use of our completion of the square formulae presented earlier.
- We will use this result *all the time* in this and subsequent lectures.
- I'm not kidding. ALL THE TIME!
- (So try and get familiar with it).

Deriving $\beta|\Sigma, y$ in the LRM

Consider the regression model (where y is $n \times 1$)

-

under the proper prior

-

(and some independent prior on Σ, $p(\Sigma)$ which we will not explicitly model).

We will show that

-

where

-

and

-

First, note (under prior independence)

-

where $p(\Sigma)$ denotes the prior for Σ.

Since the conditional posterior distribution $p(\beta|\Sigma, y)$ is proportional to the joint above (why?), we can ignore all terms which enter multiplicatively and involve only Σ or y (or both).

It follows that

$$p(\beta|\Sigma, y) \propto \exp\left(-\frac{1}{2}\left[(y - X\beta)'\Sigma^{-1}(y - X\beta) + (\beta - \mu_\beta)'V_\beta^{-1}(\beta - \mu_\beta)\right]\right).$$

This is almost in the form where we can apply our previous completion of the square result. However, before doing this, we must get a quadratic form for β in the "likelihood function part" of this expression.

To this end, let us define

-

- Note that β^* is just the familiar GLS estimator.
- Also note that, given Σ and y, β^* is effectively "known" for purposes of deriving $\beta|\Sigma, y$ (i.e., we can and will condition on it).

For the sake of space, let

-

and note

-

The reason why this middle term was zero on the previous page follows since

-

Using this last result, we can replace the quadratic form

$$(y - X\beta)'\Sigma^{-1}(y - X\beta)$$

with

$$\hat{u}'\Sigma^{-1}\hat{u} + (\beta - \beta^*)'X'\Sigma^{-1}X(\beta - \beta^*)$$

to yield

-

We can apply our completion of the square formula to the above to get:

$$p(\beta|\Sigma, y) \propto \exp\left(-\frac{1}{2}(\beta - \overline{\beta})'\overline{V}_\beta(\beta - \overline{\beta})\right)$$

where

$$\begin{aligned}
\overline{V}_\beta &= (X'\Sigma^{-1}X + V_\beta^{-1}) \\
\overline{\beta} &= \overline{V}_\beta^{-1}(X'\Sigma^{-1}X\beta^* + V_\beta^{-1}\mu_\beta) \\
&= \overline{V}_\beta^{-1}(X'\Sigma^{-1}y + V_\beta^{-1}\mu_\beta)
\end{aligned}$$

Defining

$$D_\beta = \overline{V}_\beta^{-1} \quad \text{and} \quad d_\beta = X'\Sigma^{-1}y + V_\beta^{-1},$$

it follows that

$$\beta|\Sigma, y \sim N(D_\beta d_\beta, D_\beta),$$

as claimed.

Gibbs Sampling in the Classical Linear Model

Consider the regression model

$$y \sim N(X\beta, \sigma^2 I_n)$$

under the priors

$$\beta \sim N(\mu_\beta, V_\beta), \quad \sigma^2 \sim IG(a, b).$$

We seek to show how the Gibbs sampler can be employed to fit this model (noting, of course, that we do not need to resort to this algorithm in such a simplified model, but introduce it here as a starting point.)

- To implement the sampler, we need to derive two things:
  1.
  2.

The first of these can be obtained by applying our previous theorem (with $\Sigma = \sigma^2 I_n$). Specifically, we obtain:

$$\beta|\sigma^2, y \sim N(D_\beta d_\beta, D_\beta),$$

where

$$D_\beta = \left(X'X/\sigma^2 + V_\beta^{-1}\right)^{-1}, \quad d_\beta = X'y/\sigma^2 + V_\beta^{-1}\mu_\beta.$$

As for the posterior conditional for σ^2, note

$$p(\beta, \sigma^2 | y) \quad \propto \quad p(\beta)p(\sigma^2)p(y|\beta, \sigma^2).$$

Since $p(\sigma^2|\beta, y)$ is proportional to the joint posterior above, it follows that

-

The density on the last page is easily recognized as the kernel of an

-

density.

Thus, to implement the Gibbs sampler in the linear regression model, we can proceed as follows.

1. Given a current value of σ^2 :
2. Calculate $D_\beta = D_\beta(\sigma^2)$ and $d_\beta = d_\beta(\sigma^2)$.
3. Draw $\tilde{\beta}$ from a $N[D_\beta(\sigma^2)d_\beta(\sigma^2), D_\beta(\sigma^2)]$ distribution.
4. Draw $\tilde{\sigma}^2$ from an
$$IG\left(\frac{n}{2}+a, \left[b^{-1}+\frac{1}{2}(y-X\tilde{\beta})'(y-X\tilde{\beta})\right]^{-1}\right)$$
distribution.
5. Repeat this process many times, updating the posterior conditionals at each iteration to condition on the most recent simulations produced in the chain.
6. Discard an early set of parameter simulations as the burn-in period.
7. Use the subsequent draws to compute posterior features of interest.

Application to Wage Data

- We apply this Gibbs sampler to our wage-education data set.
- We seek to compare posterior features estimated via the Gibbs output to those we previously derived analytically in our linear regression model lecture notes.
- The following results are obtained under the prior

$$\beta \sim N(0, 4I_2), \quad \sigma^2 \sim IG[3, (1/[2 * .2])].$$

- We obtain 5,000 simulations, and discard the first 100 as the burn-in period.

Posterior Calculations Using the Gibbs Sampler

	β_0	β_1	σ^2
$E(\cdot\|y)$	1.18	.091 [.091]	.267
$Std(\cdot\|y)$	.087	.0063 [.0066]	.0011

In addition, we calculate

$$\Pr(\beta_1 < .10|y) = .911.$$

The numbers in square brackets were our analytical results based on a diffuse prior. Thus, we are able to match these analytical results almost exactly with 4,900 post-convergence simulations.

Gibbs with Groupwise Heteroscedasticity

Suppose we generalize the regression model in the previous exercise so that

$$y = X\beta + \epsilon,$$

where

$$E(\epsilon\epsilon'|X) \equiv \Sigma = \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \sigma_1^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_2^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \sigma_2^2 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_J^2 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_J^2 \end{bmatrix}.$$

That is, we relax the homoscedasticity assumption with one of groupwise heteroscedasticity - where we presume there are J different groups with identical variance parameters within each group, but different parameters across groups.

Let n_j represent the number of observations belonging to group J. Given priors of the form

-

show how the Gibbs sampler can be used to fit this model.

- It is important to recognize that even in this "simple" model, proceeding analytically, as we did in the homoscedastic regression model, will prove to be very difficult.
- As will soon become clear, however, estimation of this model via the Gibbs sampler involves only a simple extension of our earlier algorithm for the "classical" linear regression model.

Since we have pre-sorted the data into blocks by type of variance parameter, define:

y_j as the $n_j \times 1$ outcome vector and X_j the $n_j \times k$ covariate matrix, respectively, for group j, $j = 1, 2, \ldots, J$.

Thus,

$$\mathrm{Var}(y_j|X_j) = \sigma_j^2 I_{n_j}, \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_J \end{bmatrix}, \quad X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_J \end{bmatrix}.$$

The likelihood function for this sample of data can be written as:

-

Combining this likelihood with our priors (which we won't write out explicitly), we obtain:

-

To implement the Gibbs sampler, we need to derive the posterior conditionals:

1. $$p(\beta|\sigma_1^2, \sigma_2^2, \ldots, \sigma_J^2, y) \equiv p(\beta|\Sigma, y)$$

2. $$p(\sigma_1^2|\beta, \sigma_2^2, \ldots, \sigma_J^2, y)$$

3. $$\vdots$$

4. $$p(\sigma_J^2|\beta, \sigma_1^2, \ldots, \sigma_{J-1}^2, y).$$

The first posterior conditional, again, can be obtained by a direct application of the Lindley and Smith (1972) result. Specifically, since $y \sim N(X\beta, \Sigma)$ and $\beta \sim N(\mu_\beta, V_\beta)$, it follows that

$$\beta | \{\sigma_j^2\}_{j=1}^J, y \sim N(D_\beta d_\beta, D_\beta)$$

where

-

and

-

Note that Σ will need to be updated at each iteration of the algorithm when sampling β.

As for the complete posterior for each σ_j^2, note

-

(with the "$-j$" subscript denoting all parameters other than j).

Staring at the expression for the joint posterior, we see that all terms involving σ_{-j}^2 are separable from those involving σ_j^2 and so the posterior conditional of interest is of the form

-

In this form, it is recognized that

-

Thus, to implement the sampler, we would first need to specify an initial condition for β or $\{\sigma_j^2\}_{j=1}^J$. What might be reasonable values for these?

Let's say we set $\beta = \tilde{\beta}$ initially. Then, we

1. Sample $\tilde{\sigma}_1^2$ from an

$$IG\left(\frac{n_1}{2} + a_1, \left[b_1^{-1} + (1/2)(y_1 - X_1\tilde{\beta})'(y_1 - X_1\tilde{\beta})\right]^{-1}\right).$$

2.

$$\vdots$$

3. Sample $\tilde{\sigma}_J^2$ from an

$$IG\left(\frac{n_J}{2} + a_J, \left[b_J^{-1} + (1/2)(y_J - X_J\tilde{\beta})'(y_J - X_J\tilde{\beta})\right]^{-1}\right).$$

4. Sample a new $\tilde{\beta}$ from a $N(D_\beta d_\beta, D_\beta)$ density, noting that Σ must be updated appropriately.