Gibbs Sampling in Latent Variable Models #1

Econ 690

Purdue University

February 8, 2010

Outline

- Many microeconometric applications (including binary, discrete choice, tobit and generalized tobit analyses) involve the use of latent data.
- This latent data is unobserved by the econometrician, but the observed choices economic agents make typically impose some type of truncation or ordering among the latent variables.

- In this lecture we show how the Gibbs sampler can be used to fit a variety of common microeconomic models involving the use of latent data.
- In particular, we review how data augmentation [see, e.g., Tanner and Wong (1987), Chib (1992) and Albert and Chib (1993)] can be used to simplify the computations in these models.
- Importantly, we recognize that many popular models in econometrics are essentially linear regression models on suitably defined latent data.
- Thus, conditioned on the latent data's values, we can apply all of our previous techniques to sample parameters in the framework of a linear regression model.

- To review the idea behind data augmentation in general terms, suppose that we are primarily interested in characterizing the joint posterior

$$p(\theta|y).$$

- An alternative to analyzing this posterior directly involves working with a seemingly more completed posterior $p(\theta, z|y)$ for some latent data z.
- Although the addition of the variables z might seem to needlessly complicate the estimation exercise, it often proves to be computationally convenient.
- To implement the Gibbs sampler here, we need to obtain the posterior conditionals $p(\theta|z, y)$ and $p(z|\theta, y)$. Often, the posterior conditional $p(\theta|z, y)$ will take a convenient form, thus making it possible to fit the model using the Gibbs sampler.

Probit Model

Consider the following latent variable representation of the probit model:

-

The value of the binary variable y_i is observed, as are the values of the explanatory variables x_i. The latent data z_i, however, are unobserved.

For this model, we seek to:

- (a) Derive the likelihood function.
- (b) Using a prior for β of the form $\beta \sim N(\mu_\beta, V_\beta)$, derive the *augmented* joint posterior $p(\beta, z|y)$.
- (c) Verify that marginalized over z, the joint posterior for the parameters β is exactly the same as the posterior you would obtain without introducing any latent variables to the model.
- (d)Discuss how the Gibbs sampler can be employed to fit the model.

(a) Note that

$$
\begin{aligned}
\Pr(y_i = 1|x_i, \beta) &= \Pr(x_i\beta + \epsilon_i > 0) \\
&= \Pr(\epsilon_i > -x_i\beta) \\
&= 1 - \Phi(-x_i\beta) \\
&= \Phi(x_i\beta).
\end{aligned}
$$

Similarly,

$$
\Pr(y_i = 0|x_i, \beta) = 1 - \Phi(x_i\beta).
$$

Because of the assumed independence across observations, the likelihood function is obtained as:

$$
L(\beta) = \prod_{i=1}^{n} \Phi(x_i\beta)^{y_i}[1 - \Phi(x_i\beta)]^{1-y_i}.
$$

(b) To derive the *augmented* joint posterior, note that

-

implying that

-

The term $p(\beta)$ is simply our prior, while $p(y, z|\beta)$ represents the complete or augmented data density.

To characterize this density in more detail, note

-

Immediately, from our latent variable representation, we know

-

As for the conditional for y given z and β, note that when $z_i > 0$ then y_i must equal one, while when $z_i \leq 0$, the y_i must equal zero.

In other words, the sign of z_i perfectly predicts the value of y. Hence, we can write

-

with I denoting the indicator function taking on the value one if the statement in the parentheses is true, and is otherwise zero.

This previous expression simply states that when z_i is positive, then y_i is one with probability one, and conversely, when z_i is negative, then y_i is zero with probability one.
Putting the pieces together, we obtain the augmented data density $p(y, z|\beta)$. We combine this with our prior to obtain

-

(c) To show the equality of these quantities, we integrate our last expression for the joint posterior over z to obtain:

-

- Note that this is exactly the prior times the likelihood obtained from part (a) which can be obtained without explicitly introducing any latent variables.
- What is important to note is that the posterior of β is unchanged by the addition of the latent variables.
- So, augmenting the posterior with z will not change any inference regarding β, though it does make the problem computationally easier, as described in the solution to the next question.

(d) With the addition of the latent data, the complete conditionals are easily obtained.

In particular, the complete conditional for β *given* z and the data y follows directly from standard results from the linear regression model

-

where

-

As for the complete conditional for z, first note that the independence across observations implies that each z_i can be drawn independently.

We also note that

-

Thus,

-

where the notation $\text{TN}_{[a,b]}(\mu, \sigma^2)$ denotes a Normal distribution with mean μ and variance σ^2 truncated to the interval $[a, b]$.

- These derivations show that one can implement the Gibbs sampler by drawing $\beta|z, y$ from a multivariate Normal distribution and then drawing each $z_i|\beta, y,\ i = 1, 2, \cdots, n$ independently from its conditional posterior distribution.
- To generate draws from the Truncated Normal one can use the inverse transform method as described in previous lectures.
- I am supplying a MATLAB program, "truncnorm3.m" which generates draws from a truncated normal distribution (and uses the method of inversion).

Probit: Application

In an paper that has been used to illustrate the binary choice model in different econometrics texts [e.g., Gujarati (2003)], Fair (1978) analyzed the decision to have an extramarital affair.

We take a version of this data and fit a binary choice model to examine factors that are related to the decision to have an affair.

We specify a probit model where the decision to have an extramarital affair depends on seven variables:

1. An intercept (CONS)
2. A male dummy (MALE)
3. Number of years married (YS-MARRIED)
4. A dummy variable if the respondent has children from the marriage (KIDS)
5. A dummy for classifying one's self as "religious" (RELIGIOUS)
6. Years of schooling completed (ED)
7. A final dummy variable denoting if the person views the marriage as happier than an average marriage (HAPPY)

For our prior for the 7×1 parameter vector β, we specify

$$\beta \sim N(0, 10^2 I_7),$$

so that the prior is quite noninformative and centered at zero. We run the Gibbs sampler for 2000 iterations, and discard the first 500 as the burn-in.

Table 14.1: Coefficient and Marginal Effect Posterior Means and Standard Deviations from Probit Model Using Fair's (1978) Data

	Coefficient		Marginal Effect	
	Mean	Std. Dev	Mean	Std. Dev
CONS	-.726	(.417)	—-	—-
MALE	.154	(.131)	.047	(.040)
YS-MARRIED	.029	(.013)	.009	(.004)
KIDS	.256	(.159)	.073	(.045)
RELIGIOUS	-.514	(.124)	-.150	(.034)
ED	.005	(.026)	.001	(.008)
HAPPY	-.514	(.125)	-.167	(.042)

To illustrate how quantities of interest other than regression coefficients could be calculated, we reported posterior means and standard deviations associated with marginal effects from the probit model.

For a continuous explanatory variable we note:

-

For given x, the marginal effect on the previous slide is a function of the regression parameters β.

Thus, the posterior distribution of the marginal effect can be obtained by calculating and collecting the values $\{\beta_j^{(i)}\phi(x\beta^{(i)})\}_{i=1}^{1500}$, where $\beta^{(i)}$ represents the i^{th} post-convergence draw obtained from the Gibbs sampler.

When x_j is binary, we calculate the marginal effect as the difference between the Normal c.d.f.'s when the binary indicator is set to 1 and then 0.

A Panel Probit Model

To illustrate how our results for nonlinear and hierarchical models can be combined, consider a panel probit model of the form:

-

The observed binary responses y_{it} are generated according to:

-

The random effects $\{\alpha_i\}$ are drawn from a common distribution:

$$\alpha_i \stackrel{iid}{\sim} N(\alpha, \sigma_\alpha^2).$$

Suppose that you employ priors of the following forms:

$$\begin{aligned} \beta &\sim N(\mu_\beta, V_\beta) \\ \alpha &\sim N(\mu_\alpha, V_\alpha) \\ \sigma_\alpha^2 &\sim IG(a, b). \end{aligned}$$

We seek to show a Gibbs sampler can be employed to fit this panel probit model.

Like the probit without a panel structure, we will use data augmentation to fit the model.

Specifically, we will work with an augmented posterior distribution of the form

-

Derivation of the complete posterior conditionals then follows similarly to those derived in this lecture for the probit and a previous lecture for linear hierarchical models.

Specifically, we obtain

$$z_{it}|\alpha, \{\alpha_i\}, \sigma_\alpha^2, \beta, y \overset{ind}{\sim} \begin{cases} TN_{(-\infty,0]}(\alpha_i + x_{it}\beta, 1) & \text{if } y_{it} = 0 \\ TN_{(0,\infty)}(\alpha_i + x_{it}\beta, 1) & \text{if } y_{it} = 1. \end{cases}$$

-

where

-

-

where

-

X and z have been stacked appropriately and

$$\overline{\overline{\alpha}} \equiv [\alpha_1 \iota_T' \ \cdots \ \alpha_n \iota_T']'$$

with ι_T denoting a $T \times 1$ vector of ones.

$$\alpha|\beta, \{\alpha_i\}, \sigma_\alpha^2, z, y \sim N(D_\alpha d_\alpha, D_\alpha)$$

where

$$D_\alpha = (n/\sigma_\alpha^2 + V_\alpha^{-1})^{-1}, \quad d_\alpha = \sum_i \alpha_i/\sigma_\alpha^2 + V_\alpha^{-1}\mu_\alpha.$$

Finally,

$$\sigma_\alpha^2|\alpha, \{\alpha_i\}, \beta, z, y \sim IG\left(\frac{n}{2} + a, [b^{-1} + .5\sum_i(\alpha_i - \alpha)^2]^{-1}\right).$$

Fitting the model involves cycling through these conditional posterior distributions.

Of course, blocking steps can and should be used to improve the mixing of the posterior simulator

The Tobit Model

The tobit model specifies a mixed discrete-continuous distribution for a censored outcome variable y.

In most applications of tobit models, values of y are observed provided y is positive, while we simultaneously see a clustering of y values at zero.

Formally, we can write the tobit specification in terms of a latent variable model:

-

and

-

For this model, we seek to

1. (a) Write down the likelihood function for the tobit model.
2. (b) Describe how data augmentation can be used in conjunction with the Gibbs sampler to carry out a Bayesian analysis of this model.

(a) The likelihood function breaks into two parts.

For the set of observations censored at zero, the contribution to the likelihood is:

$$\Pr(y_i = 0 | x_i, \beta, \sigma^2) = \Pr(\epsilon_i \leq -x_i\beta) = 1 - \Phi(x_i\beta/\sigma).$$

Similarly, when $y_i > 0$, the contribution to the likelihood is

$$\phi(y_i; x_i\beta, \sigma^2) = \frac{1}{\sigma}\phi\left[\left(\frac{y_i - x_i\beta}{\sigma}\right)\right],$$

with $\phi(\cdot)$ denoting the *standard* Normal density function. Hence,

$$L(\beta, \sigma^2) = \prod_{i:y_i=0}\left[1 - \Phi\left(\frac{x_i\beta}{\sigma}\right)\right]\prod_{i:y_i>0}\frac{1}{\sigma}\phi\left[\left(\frac{y_i - x_i\beta}{\sigma}\right)\right].$$

(b) Chib (1992, Journal of Econometrics) was the original Bayesian tobit paper. Before discussing how the model can be fit using the Gibbs sampler, we note that the following prior specifications are employed:

$$\begin{aligned} \beta &\sim N(\mu_\beta, V_\beta) \\ \sigma^2 &\sim IG(a, b). \end{aligned}$$

Following Chib (1992), we augment the joint posterior with the latent data z_i.

Thus, we will work with

-

It is also important to recognize that when $y_i > 0$, z_i is observed (and equals y_i), while when $y_i = 0$, we know that z_i is truncated from above at zero.

To this end, let D_i be a binary variable that equals one if the observation is censored ($y_i = 0$) and equals zero otherwise. We define

-

so that y_z just takes the value y for the uncensored observations, and for the censored observations, y_z will take the value of the latent data z.

Conditioned on z, this makes it easy to draw from the β conditional.

We obtain the following complete posterior conditionals for the standard tobit model:

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

where

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

$$\sigma^2|\beta, z, y \sim IG\left(n/2 + a, \left(b^{-1} + (1/2)\sum_{i=1}^{n}(y_{z_i} - x_i\beta)^2\right)^{-1}\right).$$

Finally,

-

A Gibbs sampler involves cycling through these three sets of conditional posterior distributions.

Tobit Model: Application

We illustrate the use of the tobit model with a simple data set providing information on the number of weeks worked in 1990 by a sample of 2,477 married women.

The data are taken from the National Longitudinal Survey of Youth (NLSY).

Importantly, we recognize that this data set has a two-sided censoring problem, wherein the number of weeks worked is clustered both at zero and at 52 (full-time employment).

We seek to account for this two-sided censoring in our model specification.

For our priors, we use the same forms as those presented for the general tobit model and set

$$\mu_\beta = 0, \quad V_\beta = 10^2 I_5, \quad a = 3, \quad b = (1/40).$$

In our data, approximately 45 percent of the sample reports working 52 hours per week and 17 percent of the sample reports working 0 weeks per year.
To account for this feature of our data, we specify a slightly generalized form of the tobit model:

$$z_i = x_i\beta + \epsilon_i, \qquad \epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$$

and

-

Equivalently,

-

In terms of the Gibbs sampler, the only difference between the algorithm required for this model and the one for the standard tobit is that z_i must be truncated from below at 52 for all of those observations with $y_i = 52$.

One possible programming expedient in this regard is to sample a latent variable truncated from above at zero (say z_i^1) and one truncated from below at 52 (say z_i^2) for *every* observation in the sample.

Then, let D_i^1 be a dummy variable if $y_i = 0$ and D_i^2 be a dummy if $y_i = 52$. We can then construct our complete data vector y_z as:

$$y_z = D^1 z^1 + D^2 z^2 + (1 - D^1 - D^2)y.$$

With this construction, the Gibbs sampler then proceeds identically to the previous case.

Table 14.4: Coefficient and Marginal Effect Posterior Means and Standard Deviations from Tobit Model of Weeks Worked by Married Women

	Coefficient		Marginal Effect	
	Mean	Std. Dev.	Mean	Std. Dev
CONST	29.65	(6.06)	—-	—-
AFQT	.107	(.047)	.043	(.019)
SPOUSE-INC	-.245	(.053)	-.098	(.021)
KIDS	-26.65	(2.44)	-10.61	.971
ED	2.94	(.513)	1.17	(.202)
σ	44.35	(1.20)	—-	—-

- The sampler was run for 5,500 iterations and the first 500 were discarded as the burn-in.
- For a tobit with two-sided censoring, one can show:
  -
- Our results suggest that married women with low test scores, few years of education, and children in the home are likely to work fewer weeks during the year.