Probit Regression

Yingbo Li

Clemson University

MATH 9810

Latent Variable Interpretation of Binary Response Models

- Suppose latent variable, U, has a continuous cdf $F(u; x)$
- Binary response $Y = \mathbf{1}(U > 0)$, where $\mathbf{1}(A)$ is an 0-1 indicator of event A:
$$\theta = P(Y = 1 \mid X) = 1 - F(0; X)$$
- Since U is not directly observed, there is no loss of generality in taking the cut-point to be 0.
- In addition, we can take the standard deviation of U (or some other measure of dispersion) to be 1, without loss of generality.

Probit Models

- For example, if $U_i \sim \mathsf{N}(\beta X_i, 1)$, then

$$\theta_i = P(Y_i = 1 \mid X_i) = \Phi(\beta X_i),$$

  where $\Phi(\cdot)$ is the cdf of standard normal distribution.
- The relation is linearized by the inverse normal transformation

$$\Phi^{-1}(\theta_i) = \beta X_i$$

- We have regarded the cutoff value of U as fixed and the mean of U to be changing with X. Alternatively, one could assume that the distribution of U is fixed and allow the critical value to vary with X.

Bayesian Analysis: Prior, Likelihood & Posterior

$$P(Y_i = 1 \mid X_i) = \Phi(\beta X_i),$$

- Prior: $\beta \sim \mathsf{N}(\beta_0, \sigma_0^2)$
- Likelihood:

$$p(\mathbf{Y} \mid \beta, \mathbf{X}) = \prod_{i=1}^{n} \Phi(\beta X_i)^{Y_i} \left\{1 - \Phi(\beta X_i)\right\}^{1-Y_i}$$

- Posterior:

$$p(\beta \mid \mathbf{Y}, \mathbf{X}) \propto p(\mathbf{Y} \mid \beta, \mathbf{X})\; p(\beta)$$

- No closed form available for normalizing constant.

Posterior Computation using Data Augmentation

- Full conditional posterior distributions needed for Gibbs sampling are not automatically available
- However, we can rely on a very useful data augmentation trick proposed by Albert and Chib (1993):
  - Augment observed data $\{Y_i, X_i\}$ with latent variable U_i.
  - Probit model can be expressed in hierarchical form as follows:

$$Y_i = \mathbf{1}(U_i > 0)$$
$$U_i \sim \mathsf{N}(\beta X_i, 1)$$

- Marginalizing out U_i we obtain $P(Y_i = 1 \mid X_i) = \Phi(\beta X_i)$.

Gibbs Sampling Steps

- Gibbs sampling relies on alternately sampling from full conditional posterior distributions of unknown parameters
- After data augmentation, unknowns include latent data $U_1, \ldots, U_n$ and regression parameters β.
- Full conditional posterior distributions:
  - $p(U_i \mid \mathbf{Y}, \mathbf{X}, \beta) = \mathsf{N}(\beta X_i, 1)$ truncated below by zero if $Y_i = 1$ and above by zero if $Y_i = 0$.
  - $p(\beta \mid \mathbf{Y}, \mathbf{X}, \mathbf{U}) = \mathsf{N}(\beta_n, \sigma_n^2)$, where $\sigma_n^2 = (1/\sigma_0^2 + \sum_i X_i^2)^{-1}$, and $\beta_n = \sigma_n^2(\beta_0/\sigma_0^2 + \sum_i U_i X_i)$