

Probability Model For An Ordered Probit Regression Model

Nan Wu
nanw@udel.edu

1 Ordered Probit Regression Model

The ordered probit regression is a regression model for ordinal dependent variables. Assume the observable dependent variable is y , which is a function of another unobserved, continuous latent variable y^* in the ordered probit model described as below.

$$y = \begin{cases} 1, & \text{if } y^* \leq c_1, \\ 2, & \text{if } c_1 < y^* \leq c_2, \\ \vdots & \\ K, & \text{if } c_{K-1} < y^* \end{cases}$$

And

$$\begin{aligned} y^* &= X\beta + \epsilon \\ \epsilon &\sim \text{Normal}(0, 1) \end{aligned}$$

where $y_n \in \{1, \dots, K\}$, $n = 1, \dots, N$; $y^* \in \mathbb{R}^N$; $X \in \mathbb{R}^{N \times P}$ is the predictor matrix, where N is the number of observations, and D is the number of predictors; $\beta \in \mathbb{R}^P$ is the coefficient vector; and $c \in \mathbb{R}^{K-1}$ is a sequence of cutpoints sorted so that $c_k < c_{k+1}$.

Therefore,

$$\begin{aligned} y^* | \beta &\sim \text{Normal}(X\beta, 1) \\ \Pr(y_n^* \leq c_k) &= \Phi\left(c_k - \sum_{p=1}^P X_{n,p}\beta_p\right), \quad n = 1, \dots, N \end{aligned}$$

where

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy \quad (1)$$

Assuming $c_0 = -\infty$, $c_K = \infty$, $\Phi(-\infty) = 0$, and $\Phi(\infty) = 1$, the probability of y_n is therefore computed as

$$\begin{aligned}\Pr(y_n = k) &= \Pr(c_{k-1} < y_n^* \leq c_k) \\ &= \Pr(y_n^* \leq c_k) - \Pr(y_n^* \leq c_{k-1}) \\ &= \Phi\left(c_k - \sum_{p=1}^P X_{n,p}\beta_p\right) - \Phi\left(c_{k-1} - \sum_{p=1}^P X_{n,p}\beta_p\right)\end{aligned}$$

where $n = 1, \dots, N$, and $k = 1, \dots, K$.

2 Probability Model

c and β are parameters in the ordered probit model. Stan allows us to use improper priors if we don't have any prior knowledge about the parameters. We can therefore start with a simple model by assuming improper priors for c and β .

$$\begin{aligned}c_k &\sim \text{Uniform}(-\infty, \infty), \quad c_k \leq c_{k+1} \\ \beta_p &\sim \text{Uniform}(-\infty, \infty) \\ k &= 1, \dots, K-1, \quad p = 1, \dots, P\end{aligned}$$

Putting it all together, the probability model for the ordered probit regression model is:

$$\begin{aligned}y_n &\sim \text{Categorical}(\mathbf{p}_n) \\ \mathbf{p}_n &= (p_{n1}, \dots, p_{nK}) \\ p_{nk} &= \Phi\left(c_k - \sum_{p=1}^P X_{n,p}\beta_p\right) - \Phi\left(c_{k-1} - \sum_{p=1}^P X_{n,p}\beta_p\right) \\ \beta_p &\sim \text{Uniform}(-\infty, \infty) \\ c_k &\sim \text{Uniform}(-\infty, \infty), \quad k = 1, \dots, K-1 \\ c_0 &= -\infty, \quad c_K = \infty \\ c_k &< c_{k+1}, \quad k = 1, \dots, K-1\end{aligned}$$