

# Probability Model For An Ordered Logistic Regression Model

Nan Wu  
nanw@udel.edu

## 1 Ordered Logistic Regression Model

The ordered logistic 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 logit 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

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

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  $P$  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,

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

where

$$\frac{1}{1 + \exp\left(-\left(c_k - \sum_{p=1}^P X_{n,p}\beta_p\right)\right)} = \text{logit}^{-1}\left(c_k - \sum_{p=1}^P X_{n,p}\beta_p\right) \quad (1)$$

Assuming  $c_0 = -\infty$ ,  $c_K = \infty$ ,  $\text{logit}^{-1}(-\infty) = 0$ , and  $\text{logit}^{-1}(\infty) = 1$ , the probability of  $y_n$  is therefore computed as

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

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

## 2 Probability Model

$c$  and  $\beta$  are parameters in the ordered logit 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  $\beta$ .

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

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

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