Extended Poisson Regression Model for Analyzing Ordered Categorical Response Data

Megu OHTAKI, *Department of Environmetrics and Biometrics, RIRBM, Hiroshima University, Japan*,
E-mail: ohtaki@hiroshima-u.ac.jp

Key words: Categorical data, Ordinal response, Penalized Likelihood, Poisson process, Poisson regression

Abstract:

Ordinal categorical response data are often encountered in bio-medical, social and behavioral sciences. For analyzing such kind of data the adjacent categories logit model with the maximum likelihood method [1],[2] or the conditional likelihood method [3] for estimating unknown parameters have been commonly used. It was also shown that use of a penalized likelihood is effective for fitting the adjacent categories logit model.[4] The formula of adjacent categories logits model is so simple that it is easy to use in data analysis; however, the structure may seem somewhat peculiar from a mechanistic view point. We consider an alternative method that is based on the classical theory of Poisson process [5] with introducing a quantity of latent score as a notion of virtual time, which is easier to interpret.

Assume that Y is an ordered categorical response variable taking non-negative integer values from 0 to K, and that $x = (1,\ x_1, \cdots, x_p)^T$ is a vector of covariates. To explore relationship between the ordinal categorical response variable Y and the covariates x, we borrow the theory of Poisson process with substituting the score $e^{\beta^T x}$ for time t. Then the conditional probability of Y given x, $Pr(Y = y|\ x) = G(y|\ x, \theta)$, is expressed as

$$G(y|\ x, \theta) = \begin{cases} 1 - F_1(e^{\beta^T x}|\lambda), & \text{if } y = 0, \\ F_y(e^{\beta^T x}|\lambda) - F_{y+1}(e^{\beta^T x}|\lambda), & \text{if } y \in \{1, \cdots, K-1\}, \\ F_K(e^{\beta^T x}|\lambda), & \text{if } y = K, \end{cases}$$

where $\theta = (\lambda^T, \beta^T)^T,\ \beta = (\beta_0, \beta_1, \cdots, \beta_p)^T,\ \lambda = (\lambda_1, \cdots, \lambda_{K-1})^T$ with $\prod_{j=1}^{K} \lambda_j = 1$ and

$$F_k(t|\lambda) = 1 - \sum_{j=1}^{k} e^{-\lambda_j t} \prod_{\ell \in I_j^{(k)}} \frac{\lambda_\ell}{\lambda_\ell - \lambda_j}, \qquad (\text{if } \lambda_i \neq \lambda_j\ (i \neq j)\ \text{ holds,})$$

where $I_j^{(k)} = \{i|\ i \neq j,\ i = 1, \cdots, k\},\ j = 1, \cdots, k,\ k = 1, \cdots, K$.

Specially when $\lambda_1 = \cdots = \lambda_K (= 1)$ holds, we obtain that $G(y|x, \theta) = \frac{(e^{\beta^T x})^y}{y!} e^{-e^{\beta^T x}},\ \ y = 1, \cdots, K-1$. Thus this model can be regarded as an extension of the classical Poisson regression model for count variable to ordered categorical response variable. BAN estimate of the parameters is obtained from asymptotic theory with a penelized likelihood. The performance of the proposed method is illustrated through some simulation studies.

References

[1] Agresti, A. (1984). Analysis of Ordinal Categorical Data, New York: John Wiley.

[2] Fienberg, S.E. (1980). The Analysis of Cross-Classified Categorical Data, Cambridge, MA: The MIT Press.

[3] Hirji, K.F. (1992). Computing Exact Distribution for Polytomous Response Data, *JASA*, 87, 487-492.

[4] Ohtaki, M., Kuribayashi, K. and Goto, M. (2005). Bias Adjusted Penalized Maximum Likelihood Estimates in the Adjacent Categories Logit Model, submitted to *Biometrics*.

[5] Chiang, C.L. (1979). Survival and stages of disease, *Mathematical Biosciences*, 43, 159-171.