Electronic Journal: Southwest Journal of Pure and Applied Mathematics
Internet: http://rattler.cameron.edu/swjpam.html
ISSN 1083-0464
Vol. No. 02, 1996, pp. 1–5.
Submitted: February 3, 1996. Published: July 8, 1996.

PREDICTIVE DISTRIBUTION FOR POLYTOMOUS LOGISTIC MODELS WITH NONLINEAR LINK FUNCTIONS

M.A. Tabatabai and K.P. Singh

Summary. This article is concerned with predictive distribution for polytomous logistic models with nonlinear link functions. Within a Bayesian framework, an approximation of the predictive distribution for determining probabilities is presented. We describe this method using nonlinear link functions. The generalized linear link function is simply a special case of nonlinear models. All of the generalized logistic models require nonlinear estimation procedures to evaluate maximum likelihood estimates.

(1991) A.M.S. Subject Classification Codes. 34.

Key Words and Phrases. polytomous logistic model,link function,predictive distribution

1. INTRODUCTION.

Statistical methods based on logistic models are routinely used in experiments with binary outcome data. It can also be used to model the association between a polytomous response variable and some explanatory variable of interest. This article's emphasis is on obtaining a predictive distribution for future observation based on previous data. Prediction analysis plays the key role in the real decision-making process implicit in most medical survival studies. Discussions of prediction analysis can be found in Geisser (1971, 1980, 1982) and Aitchison and Dunsmore (1975). Data with polytomous response variables are frequently encountered in social and biomedical sciences. The response may be recorded on an ordinal or nominal scale and the link function can be linear or nonlinear functions. Here we use a Bayesian formulation to make predictions for future individuals. Given data

$$D = \{(x_i, y_i),\ i = 1, \dots, n\}$$

and x_{n+1} for future individual, we can predict the distribution of unobserved response y_{n+1}. The future response is polytomous response variable. This usage has previously been applied by Percy (1993) to find predictive distribution for linear models, including its application to survival data. Heitjan (1991) applied nonlinear modeling to serial immunologic data from a multiple sclerosis clinical trial. Rosner (1989) presented a polychotomous logistic model to analyze ophthalmologic data. We focus on polytomous logistic models with nonlinear links.

2. THE POLYTOMOUS LOGISTIC MODEL.

Let Y represent the response variable and assume $Y \in \{0, 1, \dots, k\}$, and let a sample of size n be taken. When the response is categorical, the following logit model may be used to relate the response

$$ln(P_{ij}/P_{io}) = g_j(x_i; \beta_j),\ j = 1, \dots, k, \tag{2.1}$$

Cameron University, Lawton, Oklahoma
E-Mail Address: mohammat@cameron.edu
University of Alabama, Birmingham, Alabama

where $g_j(x_i;\beta_j)$ is a function of a parameter vector β_j and covariate vector x_i of dimension P, and

$$P_{ij} = P(y_i = j|x_i),\ i = 1, \ldots, n,\ j = 0, 1, \ldots, k \tag{2.2}$$

the conditional probabilities P_{ij} of each outcome category given the covariate vector are

$$P_{ij} = \exp[g_j(x_i;\beta_j)]/\left\{\sum_{l=0}^{k} \exp[g_l(x_i;\beta_l)]\right\},\ j = 0, 1, \ldots, k,\ i = 1, \ldots, n \tag{2.3}$$

where $g_0(x_i;\beta_0) \equiv 0, \sum_{j=0}^{k} P_{ij} = 1$, and $\beta_j = (\beta_{j1}, \beta_{j2}, \ldots, \beta_{jP_j})'$. To construct the likelihood function, we formulate $k+1$ binary variables coded as zero or 1 to indicate group membership of an observation. The variables are coded as follows:

If $y_i = j$, then the coded vector $y_{ij} = (y_{i0}, y_{i1}, \ldots y_{ih}, \ldots y_{ik})$ has $y_{ih} = 1$ if $h = j$ and 0 if $h \neq j$.

3. THE PROPOSED TECHNIQUE.

Let $y_1, y_2, \ldots, y_n$ be n independent random variable with joint probability mass function

$$\begin{aligned} f(y_1, y_2, \ldots, y_n|\, P_{1j}, P_{2j}, \ldots P_{nj},\ j = 0, 1, \ldots, k) = \\ \prod_{i=1}^{n} \{(P_{i0})^{y_{io}}(P_{i1})^{y_{i1}} \ldots (P_{ik})^{y_{ik}}\} \end{aligned} \tag{3.1}$$

where for each i, $\sum_{j=0}^{k} y_{ij} = 1$.

If we use (2.3), we obtain the joint distribution

$$\begin{aligned} f(y_1, y_2, \ldots, y_n|x_i,\ i = 1, \ldots, n;\ \beta_j,\ j = 1, \ldots, k) = \\ \prod_{i=1}^{n} \left\{\exp[\sum_{j=0}^{k} y_{ij} g_j(x_i,\beta_j)]/\sum_{j=0}^{k} \exp[g_j(x_i;\beta_j)]\right\} \end{aligned} \tag{3.2}$$

One major application of linear as well as nonlinear polytomous regression is to make forecasts. With these models, one can incorporate recent data

$$\{(y_i, x_i), i = 1, \ldots, n\} \text{ and } x_{n+1} \text{ into}$$

forecast for unobserved response y_{n+1}. To solve this problem and more importantly to estimate the predictive distribution we use Bayes theorem.

According to Bayes theorem, the predictive distribution

$$\begin{aligned} & f(y_{n+1}|x_1, x_2, \ldots, x_{n+1}; y_1, \ldots, y_n) \\ & = \int_{\beta} f(y_{n+1}|x_{n+1};\beta)\ f(\beta|x_1, x_2, \ldots, x_n) d\beta \end{aligned} \tag{3.3}$$

where $\beta = (\beta_1, \beta_2, \ldots, \beta_k)$ and $\dim(\beta_j) = P_j,\ j = 1, \ldots, k$ and $\sum_{j=1}^{k} P_j < n$.

To evaluate this integral we use an approximation similar to Percy's (1993) for binary linear logistic regression. Before proceeding, we must decide on a suitable prior distribution for the parameter matrix $\beta = (\beta_1, \beta_2, \ldots, \beta_k)$. We assume $\beta_1, \beta_2, \ldots, \beta_k$ are k independent random vectors with each having a prior distribution as multivariate normal. This allows considerable flexibility for practical applications. The parameter values for these priors are easily equated to the first two moments of β based on hyperprior knowledge.

Let

$$f(\beta_j) = (\det A_j)^{\frac{1}{2}}/(2\pi)^{\frac{P_j}{2}} \exp\left\{-\frac{1}{2}(\beta_j - b_j)' A_j(\beta_j - b_j)\right\} \ , j = 1, 2, \ldots, k \tag{3.4}$$

where A_j is symmetric positive definite matrix. Then the likelihood function is given by

$$L(\beta; x_1, x_2, \ldots, x_n; y_1, y_2, \ldots, y_n) = \prod_{i=1}^{n} \left\{ \exp[\sum_{j=0}^{k} y_{ij} g_j(x_i; \beta_j)/\sum_{j=0}^{k} \exp[g_j(x_i; \beta_j)] \right\} \tag{3.5}$$

Now, let $D(\beta)$ denote the posterior density. Then

$$\begin{aligned} D(\beta) &= f(\beta | x_1, x_2, \ldots, x_n; y_1, y_2, \ldots, y_n) \\ &\alpha L(\beta; x_1, x_2, \ldots, x_n; y_1, y_2, \ldots, y_n). \prod_{j=1}^{k} f(\beta_j) \\ &\alpha \prod_{i=1}^{n} \left\{ \exp[\sum_{j=0}^{k} y_{ij} g_j(x_i; \beta_j)/\sum_{j=0}^{k} \exp[g_j(x_i; \beta_j)] \right\}. \\ &\quad \exp\left\{-\frac{1}{2}\sum_{j=1}^{k}(\beta_j - b_j)' A_j(\beta_j - b_j)\right\} \end{aligned} \tag{3.6}$$

By taking natural logarithm we obtain

$$\begin{aligned} ln\ D(\beta) &= \sum_{j=0}^{k} \left\{ y_{ij} g_j(x_i; \beta_j) - ln(\sum_{j=0}^{k} \exp[g_j(x_i; \beta_j)]) \right\} \\ &\quad - \frac{1}{2}\sum_{j=1}^{k}(\beta_j - b_j)' A_j(\beta_j - b_j) \end{aligned} \tag{3.7}$$

Now, consider the vector $w(\beta)$ defined as

$$\begin{aligned} w(\beta) = \frac{\partial}{\partial \beta} ln\ D(\beta) &= \begin{bmatrix} w_1(\beta) \\ \vdots \\ w_k(\beta) \end{bmatrix} \\ &= \begin{bmatrix} \frac{\partial}{\partial \beta_1} ln\ D(\beta) \\ \vdots \\ \frac{\partial}{\partial \beta_k} ln\ D(\beta) \end{bmatrix} \end{aligned}$$

where

$$\begin{aligned} w_j(\beta) = \sum_{i=1}^{n} [y_{ij} - \exp(g_j(x_i; \beta_j)/\sum_{l=0}^{k} \exp[g_l(x_i; \beta_l)] \cdot \frac{\partial g_j}{\partial \beta_j}(x_i; \beta_j) \\ -A_j(\beta_j - b_j) \end{aligned}$$

and

$$\frac{\partial g_j}{\partial \beta_j}(x_i;\beta_j) = \begin{bmatrix} \frac{\partial g_j}{\partial \beta_{j1}}(x_i;\beta_j) \\ \vdots \\ \frac{\partial g_j}{\partial \beta_{jP_j}}(x_i,\beta_j) \end{bmatrix} \tag{3.8}$$

$j = 1, \ldots, k$.
Define

$$l(\beta) = \frac{\partial}{\partial \beta} w(\beta) = \begin{bmatrix} \frac{\partial}{\partial \beta_1} w_1(\beta) & \ldots & \frac{\partial}{\partial \beta_k} w_1(\beta) \\ \vdots & & \\ \frac{\partial}{\partial \beta_1} w_k(\beta) & \ldots & \frac{\partial}{\partial \beta_k} w_k(\beta) \end{bmatrix} \tag{3.9}$$

where $l(\beta)$ is the information matrix of order $(\sum_{j=1}^{k} P_j)$ by $(\sum_{j=1}^{k} P_j)$ which is adjusted for prior information and if $j = j'$, $j = 1, \ldots, k$, then

$$\frac{\partial}{\partial \beta_j} w_j(\beta) = \sum_{i=1}^{n} \{ \frac{\exp[2g_j(x_i;\beta_j)] - \exp[g_j(x_i;\beta_j)][\sum_{l=0}^{k} \exp[g_l(x_i;\beta_l)]}{[\sum_{l=0}^{k} \exp[g_l(x_i;\beta_l)]]^2} \cdot \left[\frac{\partial}{\partial \beta_j} g_j(x_i;\beta_j)\right] \cdot \left[\frac{\partial}{\partial \beta_j} g_j(x_i;\beta_j)\right]' + \left[y_{ij} - \frac{\exp[g_j(x_i;\beta_j)]}{\sum_{l=0}^{k} \exp[g_l(x_i;\beta_l)]} \right] \cdot \frac{\partial^2 g_j(x_i;\beta_j)}{\partial \beta_j \partial \beta'_{j'}} \} - A_j \tag{3.10}$$

But if $j \neq j', j = 1, \ldots, k, j' = 1, \ldots, k$, then we have

$$\frac{\partial}{\partial \beta_{j'}} w_j(\beta) = \sum_{i=1}^{n} \frac{\exp[g_j(x_i;\beta_j) + g_{j,}(x_i;\beta_{j'})]}{\left[\sum_{l=0}^{k} \exp[g_l(x_i;\beta_l)]\right]^2} \cdot \left(\frac{\partial}{\partial \beta_{j'}} g_{j'}(x_i;\beta_{j'})\right) \cdot \left(\frac{\partial}{\partial \beta_j} g_j(x_i;\beta_j)\right)' \tag{3.11}$$

The first order approximation is defined by

$$f(y_{n+1}|x_1, x_2, \ldots, x_{n+1}; y_1, y_2, \ldots, y_n) = f(y_{n+1}|x_{n+1}, \beta^*)$$

where β^* denotes the posterior mode, which can be found by solving the vector equation

$$w(\beta) = 0 \tag{3.12}$$

For this, an initial guess β^0 is made for β^* and the iterative algorithm is given by

$$\beta^{(l)} = \beta^{(l-1)} + [l(\beta^{(l-1)})]^{-1} w(\beta^{(l-1)}) \tag{3.13}$$

for $l = 1, 2, \ldots$ until convergence is achieved according to satisfaction of the criterion $\| \beta^{(l)} - \beta^{(l-1)} \| \leq \in$ for a chosen norm $\| \cdot \|$ and small positive value $\in$.

Having achieved convergence, β^* is taken to be $\beta^{(l)}$ having estimated $\beta = (\beta_1, \beta_2, \dots, \beta_k)$, our apparent predictive distribution for polytomous nonlinear logistic model is

$$f(y_{n+1}|x_1, x_2, \dots, x_{n+1};\, y_1, y_2, \dots, y_n)$$
$$= \exp\{\sum_{j=0}^{k}[y_{n+1,j}\, g_j(x_{n+1}; \beta_j)]\} / \sum_{j=0}^{k} \exp\{g_j(x_{n+1}; \beta_j)\}$$

It should be mentioned that the above method takes into account full prior knowledge which is available in most studies. If no prior information is available, set $A_j = 0$, and $b_j = 0$ for $j = 1, \dots, k$ in equation (2.6).

4. AN EXAMPLE.

As an example, consider the data appearing in Percy (1993) consisting of a sample of $n = 100$ patients who were admitted to intensive care units in the UK with acute respiratory and renal failure. The respose y is died (0) or survived (1) and covariates are age, shunt (a measure of lung damage), cieatinine (mg/100ml), albumin (g/100ml) and sepsis (0 = absent or 1 = present).

Percy used linear logistic model with linear link function ($g(x, \beta) = \beta_0 + \beta_1 x_1 + \dots + \beta_5 x_5$) with prior information $A = \text{diag}\ (1,\ 100,\ .01,\ 10000,\ 100,\ 1)$ and $\beta = (-1,\ -0.1,\ -10,\ 0.01,\ 0.1,\ 1)'$. For this data he obtained $\beta^* = (-1.2574,\ -0.052341,\ -7.8132,\ 0.0023812,\ 0.084193,\ 1.4057)'$ if the vector of covariates for the next patient admitted to an intensive care unit with acute respiratory and renal failure is

$$x_{101} = (1,\ 49,\ 0.21,\ 706,\ 38,\ 1)'$$

then Percy approximated the predictive distribution of y_{101} by

$$f(y_{101}|x-1, x-2, \dots, x_{101};\, y_1, y_2, \dots, y_{100}) = \begin{cases} 0.305 & \text{if } y_{101} = 0 \\ 0.695 & \text{if } y_{101} = 1. \end{cases}$$

References

J. Aitchison and I.R. Dunsmore, *Statistical Prediction Analysis*, Cambridge University Press, 1975.

S. Geisser, *The inferential use of predictive distributions*, Foundations of Statistical Inference (V.P. Godambe and D.A. Spratt, eds.), Holt, Rinehart and Winston, 1971, pp. 456–469.

S. Geisser, *The estimation of distribution functions and prediction of future observations*, Proceedings of the conference on Recent Developments in Statistical Methods and Applications, Taiwan: Academia Sinica (1980), 193–208.

S. Geisser, *Aspects of the predictive and estimative approaches in the determination of probabilities*, A Supplement to Biometrics **38, March** (1982).

D.F. Heitjan, *nonlinear modeling of serial immunologic data: a case study*, Journal of the American Statistical Association **20** (1991), no. 416.

D.F. Percy, *Prediction for generalized linear models*, Journal of Applied Statistics **20** (1993), no. 2.

B. Rosner, *Multivariate methods for clustered binary data with more than one level of nesting*, Journal of the American Statistical Association **84** (1989), no. 406.

Department of Mathematical Sciences, Cameron University, Lawton, Oklahoma

Department of Biostatistics, University of Alabama at Birmingham, Birmingham, Alabama