A Regression Model for Polytomous Data and Its Application

ALEXANDER ANDRONOV 1, NADEZDA KOLMAKOVA 2, IRINA YATSKIV 3

Transport and Telecommunication Institute
1. Lomonosova Str., Riga, LV-1019
LATVIA
1 lora@mailbox.riga.lv, 2 nk@tsi.lv, 3 ivl@tsi.lv http://www.tsi.lv

Abstract: A nonlinear regression model with polytomous response is considered. Unknown parameters are estimated using maximum likelihood method. Corresponding information matrix is presented. Gotten results are used for an evaluation of transport service quality.

Key-Words: regression, categories, the maximum likelihood method, information matrix, quality of service

1 Introduction

A considered problem was arisen as a result of collaboration with Riga Coach Terminal. The last one being a leader in the area of the passenger bus transportation services in Latvia provides the international, intercity and regional trips. Recent studies of the role of buses and coaches seem to confirm the already excellent safety, environmental and social record of bus and coach transport. In Latvia this mode of transport is in competition with railway (and also private cars) that’s why the quality of services is very important from all points of view [4].

Many authors considered the different approaches for measuring service quality (Cronin, Parasuraman, Morgan, Penja) [3, 8]. One of the most used approaches considers quality as a function from several particular attributes - variables and the key step consists in definition of weight of each attribute-factor. The problems of the service quality provided by a coach terminal have been considered by the paper’s authors several times. Suppose that random sample with size denoted \( n \) from population of users involves estimates of overall quality of service – \( y_i \), \( (i = 1, \ldots, n) \) and estimates of attributes (particular quality indexes), which define quality of service – \( x_{ik} \) for \( k \) concrete attributes \( (i=1, \ldots, n; j=1, \ldots, k) \). Assume that these estimates are made on the basis (0-5) scale. In previous researches [8] the theory of linear composite indicator constructing and statistical methods are being used, namely linear regression model with constraints on parameters’ sign and value. The model constructed for a scalar quality indicator, allows estimating influence of particular quality indicators on the overall quality estimation and to simplify monitoring of quality indicators. But the models presented in paper [10] were based on a number of assumptions. Most critical one is that the overall estimations are continuous variables when in estimation and to simplify monitoring of quality indicators.

In the given work authors offer the new improved approach which helps to estimate not only parameters (weights) of particular attributes of quality, but also intervals of categories to which values of the overall indicator of quality belong.

The described case is called in the literature as polytomous response or polytomous data [6]. McCullagh and Nelder wrote: “Often the categories are defined in a qualitative or non-numerical way”. Our responses are numerical but considered approach allows using it for qualitative or non-numerical cases.

Our approach is based on regression theory [6, 7]. The next Section contains problem setting. Section 3 is devoted to the maximal likelihood method. An information matrix for unknown parameters is calculated in Section 4. The numerical example is considered in Section 5. Some concluding remarks are given in section 6. Appendix contains some technical calculations.

2 Problem setting

We describe the considered problem following McCullagh and Nelder [6], chapter 5.

The response \( Y_i \) of a concrete individual or item \( i \) \( (i = 1, \ldots, n) \) is one of the fixed set of possible values, let \( \{1, 2, \ldots, k\} \). These values are called categories. In our paper we suppose that the categories are ordered: the category \( j \) is “better” than \( i \) if \( i < j \). The response probabilities \( \pi = (\pi_1, \pi_2, \ldots, \pi_k) \) are function of vector of
covariates or explanatory variables \( x_i = (x_{i,1}, x_{i,2}, ..., x_{i,d}) \) associated with the \( i \)-th individual: \( \pi = \pi(x_i) \). We have at our disposal the matrix of the covariates \( X \) and a vector of responses \( Y \):

\[
X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}.
\]

Our aim is to suggest the relationship between the response probability \( \pi = \pi(x_i) \) and the explanatory variables \( x_i = (x_{i,1}, x_{i,2}, ..., x_{i,d}) \). For that aim we use an unobserved continuous random variable \( Z_i \) for the \( i \)-th individual:

\[
Z_i = \sum_{j=1}^{d} \beta_j x_{ij} + \zeta_i = x_i \beta + \zeta_i,
\]

where \( \{\zeta_i\} \) are independent and identically normally distributed random variables with zero expectation and unknown variance \( \sigma^2 \), \( \{\beta_j\} \) are unknown regression coefficients, \( (\beta_1, ..., \beta_d)^T = \beta \).

Further, we introduce unknown parameters \( \theta_1 < \theta_2 < ... < \theta_{k-1} \). If the unobserved variable \( Z_i \) lies in the interval \( (\theta_{j-1}, \theta_j) \) then \( y_i = j \) is recorded. Here \( j = 1, ..., k, \theta_0 = -\infty, \theta_k = \infty \).

We want to get maximal likelihood estimates of the unknown parameters \( \beta = (\beta_1, ..., \beta_d)^T \), \( \theta = (\theta_1, ..., \theta_{k-1})^T \) and \( \sigma \). For that we have \( n \) observations with fixed values \( \{Y_i\} \) and \( x_i = (x_{i,1}, ..., x_{i,d}) \).

3 The likelihood function and parameter estimation

Note that

\[
\pi_j(x_i) = P(Y_i = j) = P(\theta_{j-1} < Z_i \leq \theta_j) = \Phi\left(\frac{\theta_j - x_i \beta}{\sigma}\right) - \Phi\left(\frac{\theta_{j-1} - x_i \beta}{\sigma}\right)
\]

(2)

We see that the unknown parameters \( \beta = (\beta_1, ..., \beta_d) \), \( \theta = (\theta_1, ..., \theta_{k-1}) \) and \( \sigma \) can be estimated up to constant factor. Therefore we use united parameters \( \tilde{\beta} = (\tilde{\beta}_1, ..., \tilde{\beta}_d) \) = \( \beta / \sigma \) and \( \tilde{\theta} = (\tilde{\theta}_1, ..., \tilde{\theta}_{k-1}) = \theta / \sigma \).

Note that situation takes place often in the econometrics [1, 2].

To rewrite down the corresponding log-likelihood function, we rearrange our observation (individuals) as follows: at first observations of the first category are written, then the second category and so on. Let \( n_j \) be an observation size for the \( j \)-th category, \( n_0 = 0, n_1 + ... + n_k = n, N_0 = 0 \), \( N_j = N_{j-1} + n_j, j = 1, ..., k-1; N_k = n \).

Then the log-likelihood function can be written as

\[
\ell(\tilde{\beta}, \tilde{\theta}; y) = \sum_{j=0}^{k} \sum_{i=N_{j-1}+1}^{N_j} \log(\Phi(\tilde{\theta}_j - \tilde{\beta} x_i) - \Phi(\tilde{\theta}_{j-1} - \tilde{\beta} x_i))
\]

(3)

where \( \Phi \) is the cumulative standard normal distribution function, \( \tilde{\theta}_k = \infty \).

Following the usual technique, let us derive the derivatives of the log-likelihood function with respect to \( \tilde{\theta}_j \) for \( j = 1, ..., k-1 \) are

\[
\frac{d\ell}{d\tilde{\theta}_j} = \sum_{i=N_{j-1}+1}^{N_j} (\Phi(\tilde{\theta}_j - \tilde{\beta} x_i) - \Phi(\tilde{\theta}_{j-1} - \tilde{\beta} x_i)) \times
\]

\[
\frac{1}{\sqrt{2\pi\sigma} \exp\left(-\frac{1}{2}(\tilde{\theta}_j - \tilde{\beta} x_i)^2\right)}
\]

(4)

\[
\frac{1}{\sqrt{2\pi\sigma} \exp\left(-\frac{1}{2}(\tilde{\theta}_j - \tilde{\beta} x_i)^2\right)}
\]

The derivatives of the log-likelihood function with respect to \( \tilde{\beta} = (\tilde{\beta}_1, ..., \tilde{\beta}_d) \) :

\[
\frac{d\ell}{d\tilde{\beta}} = -\sum_{j=1}^{k} \sum_{i=N_{j-1}+1}^{N_j} (\Phi(\tilde{\theta}_j - \tilde{\beta} x_i) - \Phi(\tilde{\theta}_{j-1} - \tilde{\beta} x_i)) \times
\]

\[
\frac{1}{\sqrt{2\pi\sigma} \exp\left(-\frac{1}{2}(\tilde{\theta}_j - \tilde{\beta} x_i)^2\right)}
\]

\[
\exp\left(-\frac{1}{2}(\tilde{\theta}_j - \tilde{\beta} x_i)^2\right) \right) x_i.
\]

(5)
Now we should solve the maximal likelihood equations
\[
\frac{\partial l}{\partial \beta} = 0; \quad \frac{\partial l}{\partial \theta_j} = 0, \quad j = 1, \ldots, k - 1.
\] (6)

Our experiences show that a solution is gotten easily by using standard computer programs. In other case it is necessary using Taylor expansion for a solution of the maximal likelihood equations (6). Corresponding expressions are presented in the Appendix.

If parameter estimates \( \tilde{\theta}^* = (\tilde{\theta}_1^*, \ldots, \tilde{\theta}_{k-1}^*)^T \) and \( \tilde{\beta}^* = (\tilde{\beta}_1^*, \ldots, \tilde{\beta}_d^*)^T \) are known, it is possible to estimate the probability of interest (2):
\[
\pi^*_j(x_i) = P^*\{Y_i = j\} = 
= P(\tilde{\theta}^*_{j-1} < Z_i \leq \tilde{\theta}^*_j) = 
\Phi(\tilde{\theta}^*_j - x_i, \tilde{\beta}^*) - \Phi(\tilde{\theta}^*_{j-1} - x_i, \tilde{\beta}^*)
\] (7)

4 Information matrix and asymptotic distribution of the estimate

Now we consider information matrix of unknown parameters [6, 7]:
\[
I = -\frac{1}{n} E \left[ \frac{\partial^2 l}{\partial \beta \partial \beta^T} l(\tilde{\beta}, \tilde{\theta}), \frac{\partial^2 l}{\partial \theta_j \partial \theta_k} l(\tilde{\beta}, \tilde{\theta}) \right]
\] (8)

where a dimension of matrix \( I \) is \((d + k - 1)\times(d + k - 1)\), and dimensions of its submatrices are \(d\times d\), \(d\times(k - I)\), \((k - I)\times d\) and \((k - I)\times(k - I)\).

From (4) we have for \( j = 1, \ldots, k - 1 \):
\[
\frac{\partial^2 l}{\partial \theta_j^2} = -\sum_{i=1}^N \left( \Phi(\tilde{\theta}_{j+1} - \tilde{\beta}x_i) - \Phi(\tilde{\theta}_j - \tilde{\beta}x_i) \right)^2 \times 
\frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} (\tilde{\theta}_j - \tilde{\beta}x_i)^2 \right\} + 
\frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} (\tilde{\theta}_{j-1} - \tilde{\beta}x_i)^2 \right\} \times 
\end{align}
\[
\left\{ \Phi(\tilde{\theta}_{j+1} - \tilde{\beta}x_i) - \Phi(\tilde{\theta}_{j-1} - \tilde{\beta}x_i) \right\}^{1\times} 
\times \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} (\tilde{\theta}_{j-1} - \tilde{\beta}x_i)^2 \right\} x_i x_i^T.
\]
The questionnaire included 7 groups of questions concerning the following groups of quality particular attributes:

- accessibility (availability);
- information;
- comfort;
- safety;
- infrastructure and environment.

Also the overall quality of service was evaluated. As well as particular attributes of quality the overall quality service was estimated on the scale 0-5. In total 44 questionnaires have been returned but some questions remained without the answer in two questionnaires (42 responses totally).

According to questionnaire results we have 42 responses thus the response \( Y_i \) of a concrete expert \( I \) \((i = 1, ..., 42)\) is one of a fixed set \(\{1, 2, 3, 4, 5\}\) of possible values, \( k = 5 \). Totally there were 22 particular attributes \((X)\) of quality distributed among 7 groups, \( x_i = (x_{i1}, x_{i2}, ..., x_{id}) \), \( d = 22 \) is number of factors.

In authors’ previous researches which are considered in [10] one of the goals was to reduce dimension of the problem. With the purpose of comparing reduced and full factor set let’s firstly describe our problem with not all 22 factors together, but the most significant factors according to previous investigations [10]. Least Squares Method (LSE) for a classical linear regression model was applied and because of partial quality attributes could correlate between each other, that’s why the stepwise regression model definition procedure was used (Forward Stepwise in SPSS package). Due to stepwise procedure 5 factors were selected as significant ones – \( X3 \) (Ticket booking), \( X8 \) (Punctuality), \( X11 \) (Customer trust to terminal employees), \( X13 \) (Requirements to employees), \( X22 \) (Infrastructure). Thus \( d=1,...,5 \).

Build-in optimization block in Mathcad (functions Given and Find) was used to find optimal values of estimated parameters \( \beta = (\beta_1, ..., \beta_5)^T \), \( \theta = (\theta_1, ..., \theta_4)^T \). In program we set \( \theta_0 = -3 \), \( \theta_5 = 8 \). The initial values of the parameters are \( \beta = (0.01 \ 0.01 \ 0.01 \ 0.01 \ 0.01)^T \) and \( \theta = (1 \ 2 \ 3 \ 4)^T \).

After optimization part we get the following estimations of unknown parameters:

\[ \beta = (0.358 \ -0.203 \ 0.776 \ -0.24 \ 0.65)^T \quad \text{and} \quad \theta = (1.534 \ 2.268 \ 3.929 \ 6.84)^T. \]

The optimum value of likelihood function is \( l(\hat{\beta}, \hat{\theta}) = -21.125 \).

Checking gives correct results because all partial derivatives are equal to zero.

Table 1 contains values of expert estimates and predicted estimates according to formula (7).
Table 1
Real and predicted estimates

<table>
<thead>
<tr>
<th>Expert Number, i</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert estimates</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Predicted estimates</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Expert Number, i</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
<th>22</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert estimates</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Predicted estimates</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Expert Number, i</th>
<th>23</th>
<th>24</th>
<th>25</th>
<th>26</th>
<th>27</th>
<th>28</th>
<th>29</th>
<th>30</th>
<th>31</th>
<th>32</th>
<th>33</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert estimates</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Predicted estimates</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Expert Number, i</th>
<th>34</th>
<th>35</th>
<th>36</th>
<th>37</th>
<th>38</th>
<th>39</th>
<th>40</th>
<th>41</th>
<th>42</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert estimates</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Predicted estimates</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4</td>
</tr>
</tbody>
</table>

As we can see the predicted values are close to real response vector $Y$. The considerable quantity of deviations is observed in categories with small amount of observations. It can be connected with sample imperfection.

Suggested approach allows using in model all 22 particular attributes at the same time, namely let $\beta = (\beta_1, ..., \beta_{22})^T$.

Applying previously described procedure we get the following estimations of unknown parameters:

$\beta_1 = 0.567, \beta_2 = 0.269, \beta_3 = 0.101, \beta_4 = -0.174,$

$\beta_5 = -0.543, \beta_6 = 0.567, \beta_7 = -0.28, \beta_8 = 1.55,$

$\beta_9 = -1.481, \beta_{10} = -1.112, \beta_{11} = 0.606, \beta_{12} = 0.113,$

$\beta_{13} = 0.084, \beta_{14} = 0.177, \beta_{15} = 0.31, \beta_{16} = -0.641,$

$\beta_{17} = 0.253, \beta_{18} = -0.227, \beta_{19} = 1.198, \beta_{20} = -0.826,$

$\beta_{21} = 0.13, \beta_{22} = 0.642; \theta = \{-2.091, -1.219, 1.675, 6.473\}.$

Now we have only two observations (in category $y_i = 5$) where the estimated value is not equal to the observed value. It says about an advantage of the suggested approach.

A practical result of the given investigation that it will allow coach terminal management to estimate service quality on a quantitative basis not reducing dimension of the problem.

Our example shows that developed model for an integrated quality indicator allows not only to compare and reveal the significant categories of qualities influencing the general indicator, but also to predict customer choice.

6 Conclusion
In this paper, we have considered a nonlinear regression model when the response of an individual in a study is restricted to one of a fixed set of possible values (categories). Unknown parameters of the model are estimated by the maximal likelihood methods. Expressions for the information matrix have been got. Numerical example concerns passenger evaluation of service quality in Riga Coach Terminal. The obtained results demonstrate good possibilities of the suggested approach.

Acknowledgement
The article is written with the financial assistance of European Social Fund. Project Nr. 2009/0159/1DP/1.1.2.1.2/09/IPIA/VIAA/006 (The Support in Realization of the Doctoral Programme “Telematics and Logistics” of the Transport and Telecommunication Institute)

Appendix
Below the Taylor expansions of derivatives (4) and (5) are presented for a solution of the maximal likelihood equations (6).

We begin with the derivative (4). If we use three members of Taylor expansion then for $j = 2, ..., k - 2$

$\Phi(\hat{d}_{j-1} - \hat{b}x_i) = \Phi(\hat{d}_{j-1} - \hat{d}) + \frac{\partial}{\partial \hat{d}_{j-1}} \Phi(\hat{d}_{j-1} - \hat{d}) + \frac{1}{2} (\hat{d}_{j-1} - \hat{d})^2 \frac{\partial^2}{\partial \hat{d}_{j-1}^2} \Phi(\hat{d}_{j-1} - \hat{d}) + o((\hat{d}_{j-1} - \hat{d})^3) = 0$

$+ \frac{1}{2} (\hat{d}_{j-1} - \hat{d})^2 \frac{\partial^2}{\partial \hat{d}_{j}^2} \Phi(\hat{d}_{j} - \hat{b}x_i) + o((\hat{d}_{j-1} - \hat{d})^3) = \Phi(\hat{d}_{j} - \hat{b}x_i) + \tau_j + \frac{1}{2} (\hat{d}_{j-1} - \hat{d})^2 \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} (\hat{d}_{j} - \hat{b}x_i)^2\right) - \frac{1}{2} (\hat{d}_{j-1} - \hat{d})^2 \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} (\hat{d}_{j} - \hat{b}x_i)^2\right) \times \exp\left(-\frac{1}{2} (\hat{d}_{j-1} - \hat{d})^2\right) + o((\hat{d}_{j-1} - \hat{d})^3) \right)

(9)

A substitution into (4) gives for $j = 2, ..., k - 2$
\[ \frac{\partial l}{\partial \theta_j} = \sum_{i=N+1}^{n} \left( \left( \tilde{\theta}_j - \tilde{\theta}_i \right) \left( \frac{1}{2} \tilde{\theta}_j - \tilde{\theta}_i \right) (\tilde{\theta}_j - \beta x_i) - 1 \right)^{-1} \]

\[ \sum_{i=N+1}^{n} \left( \left( \tilde{\theta}_j - \tilde{\theta}_i \right) \left( \frac{1}{2} \tilde{\theta}_j - \tilde{\theta}_i \right) \right)^{-1} = 0. \]

Furthermore

\[ \frac{\partial l}{\partial \theta_i} = \sum_{i=N+1}^{n} \Phi(\tilde{\theta}_i - \tilde{\beta} x_i)^{-1} \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{1}{2} (\tilde{\theta}_i - \tilde{\beta} x_i)^2 \right) \]

\[ - \sum_{i=N+1}^{n} \left( \tilde{\theta}_j - \tilde{\theta}_i \right) \left( \frac{1}{2} \tilde{\theta}_j - \tilde{\theta}_i \right) (\tilde{\theta}_j - \beta x_i) - \sum_{i=N+1}^{n} \left( 1 - \Phi(\tilde{\theta}_i - \tilde{\beta} x_i)^{-1} \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{1}{2} (\tilde{\theta}_i - \tilde{\beta} x_i)^2 \right) \right) x_i = 0. \]

Now we consider the derivative (5). The substitution (9) gives

\[ \frac{\partial l}{\partial \beta} = \sum_{i=1}^{N_i} \phi(\tilde{\theta}_i - \tilde{\beta} x_i)^{-1} \frac{1}{\sqrt{2\pi}} \times \]

\[ \left( - \exp \left( -\frac{1}{2} (\tilde{\theta}_i - \tilde{\beta} x_i)^2 \right) \right) x_i + \sum_{i=N+1}^{n} \left( 1 - \Phi(\tilde{\theta}_i - \tilde{\beta} x_i)^{-1} \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{1}{2} (\tilde{\theta}_i - \tilde{\beta} x_i)^2 \right) \right) x_i + \sum_{j=1}^{k-2} \sum_{i=N+1}^{n} \left( \tilde{\theta}_j - \tilde{\theta}_i \right) (1 - \exp \left( -\frac{1}{2} (\tilde{\theta}_j - \beta x_i, (\tilde{\theta}_j - \tilde{\beta} x_i)^2 + (\tilde{\beta} x_i)^2 \right) \right) x_i = 0. \]

References:


