The Korean Journal of Applied Statistics (2010) DOI: 10.5351/KJAS.2010.23.6.1057
23*(6), 1057–1065*

Notes on the Goodness-of-Fit Tests for the Ordinal Response Model

Kwang Mo Jeong[1] · Hyun Yung Lee[2]

[1]Department of Statistics, Pusan National University
[2]Department of Mathematics Education, Silla University

(Received August 2010; accepted October 2010)

Abstract

In this paper we discuss some cautionary notes in using the Pearson chi-squared test statistic for the goodness-of-fit of the ordinal response model. If a model includes continuous type explanatory variables, the resulting table from the fit of a model is not a regular one in the sense that the cell boundaries are not fixed but randomly determined by some other criteria. The chi-squared statistic from this kind of table does not have a limiting chi-square distribution in general and we need to be very cautious of the use of a chi-squared type goodness-of-fit test.

We also study the limiting distribution of the chi-squared type statistic for testing the goodness-of-fit of cumulative logit models with ordinal responses. The regularity conditions necessary to the limiting distribution will be reformulated in the framework of the cumulative logit model by modifying those of Moore and Spruill (1975). Due to the complex limiting distribution, a parametric bootstrap testing procedure is a good alternative and we explained the suggested method through a practical example of an ordinal response dataset.

Keywords: Ordinal response data, cumulative logit model, goodness-of-fit test, ordinal scores, random table, limiting distribution, parametric bootstrap.

1. Introduction

The chi-squared statistic is usually used for testing the goodness-of-fit of a model, for example the independence of two variables in a dataset given by contingency table. In the generalized linear model (that includes the loglinear model and logistic regression model) the Pearson chi-squared statistic and the deviance statistic are given by the common statistical packages. However, these measures are sometimes inappropriate for assessing the goodness-of-fit of the logistic regression model having continuous type explanatory variables because of the sparseness of the resulting table. The Hosmer-Lemeshow test for the goodness-of-fit of binary logistic regression model is a popular one suggested by Hosmer and Lemeshow (1980). This test statistic has the form of a Pearson chi-squared statistic which is computed from a $g \times 2$ random table, where g is the levels of subgroups determined by the predicted probabilities. As studied by Hosmer and Lemeshow (1980), this kind of Pearson chi-squared statistic based on the random cell boundaries depends on the estimated

[2]Corresponding author: Full time Lecturer, Department of Mathematics Education, Silla University, Gwaebeop-dong, Sasang-gu, Busan 617-736, Korea. E-mail: hylee@silla.ac.kr

parameters and hence it does not have the regular limiting chi-squared distribution. Through an extensive Monte Carlo study Hosmer and Lemeshow (1980) showed that the test statistic can be approximated by chi-squared distribution with $g-2$ degrees of freedom. Graubard *et al.* (1997) proposed an alternative grouping strategy for establishing a table using deciles of risk for the Hosmer-Lemeshow goodness-of-fit test. Other goodness-of-fit tests for logistic regression models have been proposed by other researchers such as Cox (1958), Tsiatis (1980), Brown (1982), Su and Wei (1991), Osius and Rojek (1992), Pigeon and Heyse (1999 a,b).

In this paper we focus our attention to the goodness-of-fit of an ordinal response model which can be an extension of a logistic regression model. The ordinal response data has become increasingly common in many areas such as biomedical and health sciences. Pulkstenis and Robinson (2004) already proposed a chi-squared type statistic by forming a table using the ordinal scores and the patterns of categorical covariates. However, the requirement of both types of categorical and continuous covariates in the model is a weakness of the test by Pulkstenis and Robinson (2004) even though it can simply be approximated by a chi-square distribution with appropriate degrees of freedom. Because the ordinal response model has r response categories we consider a linear combination of the response probabilities which is the so called ordinal scores by Lipsitz *et al.* (1996). We partition the whole subjects into g subgroups using the these ordinal scores, which results in a $g \times r$ cross-classified table. We routinely compute the chi-squared statistic from this $g \times r$ table to perform the goodness-of-fit test for the assumed model.

A discussion on the limiting distribution of the chi-squared type statistic from the random table was first introduced by Chernoff and Lehmann (1954), and later by other researchers such as Moore (1971), Moore and Spruill (1975). Because its limiting distribution is generally represented by a weighted sum of independent chi-squared random variates with one degrees of freedom, it is a formidable task to use the limiting distribution directly. It leads us to consider alternative approaches such as a bootstrap testing procedure to find the critical point of the test statistic. In the recent work by Jeong *et al.* (2005) a bootstrap testing procedure was proposed to test the independence of two-way ordinal tables in the respect of a power increase. Jeong and Lee (2009) also proposed the chi-squared type test statistic and a bootstrapping procedure to check the goodness-of-fit with misspecified link functions through a Monte Carlo study. But the limiting distribution of the proposed test statistic has not been studied in the previous work by Jeong and Lee (2009). This paper adresses the misuse of a chi-squared type statistic with a special emphasis on the limiting distribution and its necessary conditions.

In Section 2, we briefly review the cumulative logit models for ordinal response data. In Section 3, through an example we comment on the misuse of the usual chi-squared test when a model includes continuous type explanatory variables. The proposed test statistic and its limiting distribution will be discussed in Section 4. We modify the regularity conditions of Moore and Spruill (1975) in the framework of the cumulative logit model to find the limiting distribution of the test statistic. Finally, we summarize with comments on using chi-squared type tests for the goodness-of-fit of a model.

2. Cumulative Logit Models

Let $\mathbf{x} = (x_1, \ldots, x_p)'$ be a vector of explanatory variables. An explanatory variable is interchangeably called a predictor or covariate variable. Each x_i can be continuous, categorical, or a mix of both types. Sometimes the predictor variable is called a covariate. We denote the response variable by Y taking the integer values of 1 through r with respective probabilities $\pi_k(\mathbf{x})$, $k = 1, 2, \ldots, r$, satisfying $\sum_{k=1}^{r} \pi_k(\mathbf{x}) = 1$. Given a dataset consists of n observations $(y_i, \mathbf{x}_i)$, $i = 1, 2, \ldots, n$, we let $P(Y \leq j|\mathbf{x}_i)$ be a cumulative probability of ordinal responses smaller than or equal to j given $\mathbf{x}_i$, which is defined by

$$P(Y \leq j|\mathbf{x}_i) = \sum_{k=1}^{j} \pi_k(\mathbf{x}_i), \tag{2.1}$$

where $j = 1, \ldots, r-1$. From (2.1) each response probability $\pi_k(\mathbf{x}_i)$ can recursively be found as

$$\pi_k(\mathbf{x}_i) = P(Y \leq k|\mathbf{x}_i) - P(Y \leq k-1|\mathbf{x}_i), \quad k = 2, \ldots, r \tag{2.2}$$

with $\pi_1(\mathbf{x}_i) = P(Y \leq 1|\mathbf{x}_i)$. The cumulative logit model is defined as

$$\log\left[\frac{P(Y \leq j|\mathbf{x})}{1 - P(Y \leq j|\mathbf{x})}\right] = \alpha_j + \boldsymbol{\beta}_j'\mathbf{x}, \quad j = 1, \cdots, r. \tag{2.3}$$

The left hand side of (2.3) is simply written as logit$[P(Y \leq j|\mathbf{x})]$. In particular when the $\boldsymbol{\beta}_j's$ are all equal to a common $\boldsymbol{\beta}$, where $\boldsymbol{\beta}$ is a vector of $\beta_1, \ldots, \beta_p$ we obtain the so called proportional odds model of the form

$$\text{logit}\left[P(Y \leq j|\mathbf{x})\right] = \alpha_j + \boldsymbol{\beta}'\mathbf{x}. \tag{2.4}$$

We note that the model (2.4) has the same effects $\boldsymbol{\beta}$ for each logit but each cumulative logit has its own intercept α_j satisfying $\alpha_1 \leq \cdots \leq \alpha_{r-1}$. The proportional odds model constrains the $r-1$ response curves to have the same shape. From (2.4) we find the cumulative probability in terms of α_k and $\boldsymbol{\beta}$ as

$$P(Y \leq k|\mathbf{x}_i) = \frac{\exp(\alpha_k + \boldsymbol{\beta}'\mathbf{x}_i)}{1 + \exp(\alpha_k + \boldsymbol{\beta}'\mathbf{x}_i)}. \tag{2.5}$$

Using the relationship (2.2) and (2.5) we note that the response probability $\pi_k(\mathbf{x}_i)$ is a function of unknown parameters $\boldsymbol{\theta} = (\alpha_1, \ldots, \alpha_{r-1}, \beta_1, \ldots, \beta_p)'$, which are routinely estimated by the maximum likelihood method.

3. Chi-Squared Type Test Statistics

Suppose that a dataset is given in the form of $l \times r$ table, where l and r denote the levels of the row and column variable, respectively. The goodness-of-fit of a model (for example the independence of row and column variables) can be tested using the regular Pearson chi-squared statistic. The Pearson chi-squared statistic is given by

$$X^2 = \sum_{j=1}^{l}\sum_{k=1}^{r} \frac{(n_{jk} - E_{jk})^2}{E_{jk}}, \tag{3.1}$$

where n_{ij} and E_{jk} denote the observed frequency and the expected value of (j,k) cell under the assumed model, respectively. If there are unknown parameters that should be estimated then we also need to estimate E_{jk}. Hereafter we denote the estimated expected value as $\hat{E}_{jk}$. The X^2 is known to have a limiting chi-squared distribution with appropriate degrees of freedom depending on the model to be tested.

Table 3.1. Dataset of mental impairment status

Life events	Mental status				Total
	1	2	3	4	
0	1	0	1	0	2
1	3	2	0	0	5
2	2	1	0	1	4
3	3	3	2	0	8
4	1	0	2	2	5
5	0	3	0	1	4
6	0	1	1	0	2
7	1	0	0	1	2
8	0	1	0	3	4
9	1	1	1	1	4
Total	12	12	7	9	40

3.1. An example of ordinal response data

The dataset of Table 3.1 given in Goodman (1979) and Agresti (2002) has been obtained from a study of mental health for adult residents of Alachua County, Florida. Mental impairment can be regarded as an ordinal response variable y with four categories; well, mild, moderate, and impaired. In Table 3.1 mental status 1 denotes well and 2 denotes mild. Mental status is related to two kinds of covariates, the one is life events and the other is socioeconomic status. The life events index is a composite measure of the number and severity of important life events that happened to the subject within the past 3 years. The socioeconomic status is measured as high or low. We note that Table 3.1 is very sparse as we see so many cells of zero count. Sometimes the test statistic X^2 gives missed result because of the inappropriateness of chi-square approximation.

Here we use only the life events index as a covariate in a model to be simple. Assume a cumulative proportional odds model for this dataset

$$\log\left[\frac{P(Y \leq k)}{1 - P(Y \leq k)}\right] = \alpha_k + \beta x, \quad k = 1, 2, 3, \tag{3.2}$$

where x is the life events index. The usual goodness-of-fit statistics such as the deviance or the Pearson chi-squared statistics are inappropriate for assessing the model fit of the assumed model. When we fit the model using PROC GENMOD in SAS it gives $X^2 = 25.75$ with 26 degrees of freedom. According to the chi-square distribution the P-value is about 0.4769. However, we doubt this result because the contingency table is so sparse due to the continuous type covariate variable and it will be compared with other tests in a later section.

As a remedy to the regular chi-squared test many researchers have studied modified test statistics in the presence of continuous type covariates. Among others, we simply introduce the test by Pulkstenis and Robinson (2004). The test statistic is applied to a table reconstructed using ordinal scores. The ordinal score which was originally defined by Lipsitz *et al.* (1996) is given by

$$s(\mathbf{x}_i) = \pi_1(\mathbf{x}_i) + 2\pi_2(\mathbf{x}_i) + \cdots + r\pi_r(\mathbf{x}_i). \tag{3.3}$$

In particular, when $r = 2$ the ordinal score $s(\mathbf{x}_i)$ is equivalent to the binary response probability $\pi_2(\mathbf{x}_i)$. Since each $0 \leq \pi_k(\mathbf{x}_i) \leq 1$ with $\sum_{k=1}^{r} \pi_k(\mathbf{x}_i) = 1$ the ordinal scores satisfy the relationship $1 \leq s(\mathbf{x}_i) \leq r$ for each observation. When both types of covariates are included in the model

Pulkstenis and Robinson (2004) suggested a statistic of the form

$$X^2_{PR} = \sum_{j=1}^{I}\sum_{h=1}^{2}\sum_{k=1}^{r}\frac{(n_{jhk}-\hat{E}_{jhk})^2}{\hat{E}_{jhk}}, \tag{3.4}$$

where I denote the number of categorical covariates patterns, and the index h means the subpartition using the ordinal scores which will explained in the next section. Pulkstenis and Robinson (2004) recommend a limiting chi-square distribution with $(2I-1)(r-1)-q-1$ degrees of freedom, where q is the number of categorical covariates which compose I covariates patterns. We note that $2I$ reflects the doubles of I by the contribution of continuous variables. This statistic is analogous to the regular Pearson chi-squared test for a model containing an additional indicator variable which has resulted from continuous variables dichotomized at its median of ordinal scores. As commented by Kuss (2002) the requirement of both types of categorical and continuous covariates in the model seems to be a weakness of this test owing to the construction principle of the random table.

4. Proposed Goodness-of-Fit Test

4.1. Test statistic on the random table

When a model contains at least one continuous covariate the chi-squared statistic X^2 cannot directly be obtained from such a table in Table 3.1. In order to find the goodness-of-fit statistic we need an artificial $g \times r$ table randomly determined by some criterion. Here we use g instead of l to denote the number of subgroups which are not fixed in advance as in Table 3.1. The whole observations are partitioned according to the following steps. We first sort the ordinal scores defined in (2.3) in an increasing order and then partition the whole observations into g groups having approximately equal numbers of elements. The first group consists of $[n/g]$ number of subjects having the ordinal scores from the smallest, where $[a]$ denotes the largest integer smaller than or equal to a. The second group is similarly formed by the subjects having the next $[n/g]$ smallest ordinal scores. In this fashion the subgroup boundaries $[c_j, c_{j+1})$ of row variable W are determined with $c_1 = 0$ and $c_{g+1} = \infty$, and we finally obtain a random table given by Table 4.1.

An observation $(y_i, \mathbf{x}_i)$ with the ordinal score $s(\mathbf{x}_i)$ and $y_i = k$ belongs to the $(j,k)^{th}$ cell of Table 4.1 provided $c_j \le s(\mathbf{x}_i) < c_{j+1}$. When the assumed model is fitted on the dataset we use the estimated parameters in finding the ordinal scores $s(\mathbf{x}_i)$ and the expected value $\hat{E}_{jk}$. The $\hat{E}_{jk}$ is obtained by summing the estimated probabilities $\hat{\pi}_k(\mathbf{x}_i)$ for subjects belonging to the $(j,k)^{th}$ cell. That is,

$$\hat{E}_{jk} = \sum_{\hat{s}(\mathbf{x}_i)\in[c_j, c_{j+1})} \hat{\pi}_k(\mathbf{x}_i), \tag{4.1}$$

where the summation is taken over all subjects with $c_j \le s(\mathbf{x}_i) < c_{j+1}$ and $y_i = k$.

The proposed test statistic based on Table 4.1 is of the form

$$T_n = \sum_{j=1}^{g}\sum_{k=1}^{r}\frac{(n_{jk}-\hat{E}_{jk})^2}{\hat{E}_{jk}}, \tag{4.2}$$

where the $\hat{E}_{jk}$ is given by (4.1).

The proposed statistic T_n is not guaranteed to have an approximate chi-square distribution because T_n has not been obtained from the regular contingency on which the X^2 of (3.1) is computed. Thus

Table 4.1. Grouping structure by ordinal scores

Subgroups	Response levels					Total
	1	$\cdots$	k	$\cdots$	r	
1	n_{11}	$\cdots$	n_{1k}	$\cdots$	n_{1r}	n_1
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
j	n_{j1}	$\cdots$	n_{jk}	$\cdots$	n_{jr}	n_j
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
g	n_{g1}	$\cdots$	n_{gk}	$\cdots$	n_{gr}	n_g

we encounter a difficulty in applying the proposed test statistic in assessing the goodness-of-fit of the ordinal response model. Among many authors who discussed the approximate distribution of the chi-squared type statistic on random tables, Moore and Spruill (1975) has been a good reference.

Bull (1994) suggested that the Pearson chi-squared statistic on the random cell tables can be approximated by a chi-squared distribution with $g(r-1)-2$ degrees of freedom, which coincides with the Hosmer-Lemeshow test when $r = 2$.

4.2. Regularity conditions for the limiting distribution

In this section we reformulate the regularity conditions of Moore and Spruill (1975) in the framework of the cumulative logit model to find the limiting distribution of the test statistic T_n in (4.2). According to the result of Moore and Spruill (1975), the Pearson chi-squared type statistic T_n seems to have a value between a chi-squared variate with $df = g(r-1)-m$ and a chi-squared variate with $df = g(r-1)$. We discuss the necessary conditions to the limiting distribution of T_n in the setup of cumulative logit model by referring to those of Moore and Spruill (1975). We suggest that the goodness-of-fit test statistic T_n has the following limiting distribution given by

$$\chi^2(g(r-1)-m) + \sum_{k_1}^{k_2} \lambda_j \chi_j^2(1), \tag{4.3}$$

where the first term $\chi^2(g(r-1)-m)$ is a chi-squared distribution with $g(r-1)-m$ degrees of freedom, and the second term is a linear combination of independent chi-square variates $\chi_j^2(1)'s$ each having one degree of freedom with $k_1 = g(r-1)-(m-1)$ and $k_2 = g(r-1)$. The $\lambda_j's$ in (4.3) satisfy $0 \leq \lambda_j < 1$. As we see in the Example mentioned before, when only categorical covariates are included in the model the Pearson type X^2 statistic in (3.1) has an approximate chi-squared distribution with $g(r-1)-m$ degrees of freedom, which coincides with the first term of (4.3).

Now we briefly discuss on the regularity conditions of Moore and Spruill (1975) in the framework of our model but we skip detailed derivation because of complex notations and lengthy proof. Firstly, we specify the distribution function $F(y, \mathbf{x}|\boldsymbol{\theta})$ and its probability density (or mass) $f(y, \mathbf{x}|\boldsymbol{\theta})$ for the given dataset $(y_i, \mathbf{x}_i)$, $i = 1, 2, \ldots, n$. In the regression model including GLM's the response variable y_i is random but the covariate $\mathbf{x}_i$ is fixed by design. But it is more desirable to assume a joint distribution $F(y, \mathbf{x}|\boldsymbol{\theta})$ to investigate the regularity conditions of Moore and Spruill (1975). The density $f(y, \mathbf{x}|\boldsymbol{\theta})$ depends on $\boldsymbol{\theta}$ but we sometimes omit its dependence on $\boldsymbol{\theta}$ to simplify the notation. The density $f(y, \mathbf{x})$ can be written as $f(y|\mathbf{x})v(\mathbf{x})$, where $v(\mathbf{x})$ is the marginal distribution of covariate $\mathbf{x}$. The conditional density $f(y_i|\mathbf{x}_i)$ is given by the response probabilities $\pi_k(\mathbf{x}_i)$ satisfying the relationship of GLM in (2.4). On the other hand we may let $v(\mathbf{x}_i) = 1/n$ for each $\mathbf{x}_i$ in the design points of the regression model. Then for the observation $(y_i, \mathbf{x}_i)$ with $y_i = k$ the joint density can be written as

$$f(y_i, \mathbf{x}_i) = \pi_k(\mathbf{x}_i) \times \frac{1}{n}. \tag{4.4}$$

The probability $p_{jk}(\boldsymbol{\theta}|\mathbf{x})$ of an observation belonging to $(j,k)^{th}$ cell of Table 4.1 can be represented by the relationship

$$\begin{aligned} p_{jk}(\boldsymbol{\theta}|\mathbf{x}) &= P\left[W = j, Y = k|\mathbf{x}\right] \\ &= P\left[s(\mathbf{x}) \in [c_j, c_{j+1}), Y = k|\mathbf{x}\right], \end{aligned} \tag{4.5}$$

where W denotes an artificial row variable partitioning the whole into g subgroups. In terms of the joint density in (4.4) with $y_i = k$

$$p_{jk}(\boldsymbol{\theta}) = \int_{s(\mathbf{x}) \in [c_j, c_{j+1})} f(k, \mathbf{x}) d\mathbf{x} = \frac{1}{n} \sum_{s(\mathbf{x}_i) \in [c_j, c_{j+1})} \pi_k(\mathbf{x}_i), \tag{4.6}$$

where $\pi_k(\mathbf{x}_i)$ depends on $\boldsymbol{\theta}$ but we omit for simplification. We note that $\sum_j^g \sum_k^r p_{jk}(\boldsymbol{\theta}|\mathbf{x}) = 1$. The relationship in (4.6) is an extension of the binary logistic regression model studied by Hosmer and Lemeshow (1980).

Among many of the regularity conditions the following two conditions on the asymptotic properties of MLE hold as discussed in the Appendix of Agresti (2002). Firstly, the MLE $\hat{\boldsymbol{\theta}}$ satisfies

$$\hat{\boldsymbol{\theta}} = \boldsymbol{\theta} + O_p\left(n^{-\frac{1}{2}}\right) \tag{4.7}$$

and

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) = n^{-\frac{1}{2}} \sum_{i=1}^{n} h(y_i, \mathbf{x}_i|\boldsymbol{\theta}) + o_p(1). \tag{4.8}$$

The $h(y_i, \mathbf{x}_i|\boldsymbol{\theta})$ in (4.8) is given by $J^{-1}[\partial \log f(y_i, \mathbf{x}_i|\boldsymbol{\theta})/\partial\boldsymbol{\theta}]$, where J is an information matrix for the density $f(y, \mathbf{x})$. To simplify the notation hereafter we denote p_{jk} by p_σ using a single index σ instead of index jk. The p_σ is continuous function of $\boldsymbol{\theta}$ and also continuously differentiable in $\boldsymbol{\theta}$ since it can be represented in terms of $\pi_k(\mathbf{x})$. Let B be a $g \times r$ matrix having the $p_\sigma^{-1/2} \partial p_\sigma / \partial \boldsymbol{\theta}_\tau$ as a $(\sigma, \tau)^{th}$ element. Furthermore the $F(y, \mathbf{x}|\boldsymbol{\theta})$ is need to be continuous at every vertexes of Table 4.1 with the $\log f(y, \mathbf{x}|\boldsymbol{\theta})$ differentiable in $\boldsymbol{\theta}$ satisfying the relationship

$$\frac{\partial}{\partial \boldsymbol{\theta}} F(y, \mathbf{x}|\boldsymbol{\theta}) = \int \frac{\partial}{\partial \boldsymbol{\theta}} f(y, \mathbf{x}|\boldsymbol{\theta}) dy d\mathbf{x}. \tag{4.9}$$

Finally we assume that the rank of a matrix having $\partial p_\sigma(\boldsymbol{\theta})/\partial\theta_\tau$ as an $(\sigma, \tau)^{th}$ element is m. The other regularity conditions of Moore and Spruill (1975) can be routinely checked. In particular, if the matrix $J - B'B$ is positive definite then all $0 < \lambda_j < 1$.

4.3. Bootstrapping the test statistic

In general, the null distribution of the proposed test statistic T_n is difficult to obtain. This also holds for its limiting null distribution. In this setting, the bootstrap methods provide an attractive approach to determine a critical point for the test, or to approximate the P-value of the observed value of the proposed test statistic. The bootstrap based test will be a good alternative to the complex limiting distribution mentioned of the previous section. For the given dataset $(y_i, \mathbf{x}_i)$, $i = 1, 2, \ldots, n$, we proceed the parametric bootstrapping procedure in the following steps.

Table 4.2. Random table formed by ordinal scores when $g = 3$

Subgroup	Mental status				Total
	1	2	3	4	
1	6	3	1	1	11
2	4	6	4	3	17
3	2	3	2	5	12
Total	12	12	7	8	40

Step 1: Fit the assumed model to find the estimated $\hat{\pi}_k(\mathbf{x}_i)$ and the ordinal scores $\hat{s}(\mathbf{x}_i)$.

Step 2: Calculate the value of T_n from the $g \times r$ table given in Table 4.1.

Step 3: Generate the ordinal responses y_i^* from the multinomial distribution having the estimated probabilities $\hat{\pi}_k(\mathbf{x}_i)$ to form a bootstrap sample $(y_i^*, \mathbf{x}_i)$, $i = 1, 2, \ldots, n$.

Step 4: Similarly to Step 2 we calculate the statistic T_n^* based on the bootstrap sample of Step 3.

Step 5: Step 3 and Step 4 are repeated B times to find the empirical significance probability of the statistic T_n.

We explain the suggested bootstrap testing procedure through the dataset of Section 3.1. We assume the same model in (3.2). If we take $g = 3$ then we can form a random table such as Table 4.2 using the ordinal scores.

The proposed test statistic T_n provides a P-value of 0.97 when we take $B = 1{,}000$. On the other hand, the value $T_n = 1.20$ provides P-value $= 0.991$ when we use the chi-square approximation with $df = 7$ which has been suggested by Bull (1994). These results are quite different from the P-value 0.4769 of the regular Pearson X^2 test statistic with $df = 26$ which has been applied to the Table 3.1. The test by Pulkstenis and Robinson (2004) cannot be applied because there are no categorical covariates. Finally, we comment that the levels of Life events index (x) are grouped as subgroup 1, 2, and 3 according to $x \leq 2$, $3 \leq x \leq 5$ and $6 \leq x$, respectively.

5. Concluding Remarks

The chi-squared statistic has been widely used for testing goodness-of-fit of generalized linear models such as a log-linear model and logistic model. However, we should be cautious of the use of this test when it is based on the random table which is determined by other partitioning rules. The chi-squared statistic does not have a limiting chi-square distribution in general. In the goodness-of-fit of binary logistic regression model the test by Hosmer and Lemeshow (1980) has been a popular test that can be applied when continuous type explanatory variables are included in a model.

In this paper we focus our attention to the ordinal response data in which the outcome response is an ordinal variable. To form a table from which the chi-squared statistic is computed we require a criteria partitioning the whole observations into several subgroups. As an extension of predicted probabilities in binary logistic regression we use the ordinal scores as a partition measure which are linear combinations of response probabilities. We proposed a grouping method and the chi-squared statistic for testing goodness-of-fit of cumulative logit model. We also discussed the limiting distribution and the parametric bootstrap to apply the suggested test. The suggested method has been explained through an example of an ordinal response dataset.

Finally, we comment on the use of a chi-squared test for the goodness-of-fit of a model containing continuous type explanatory variables. The resulting table from the fit of a model is not a regular one in the sense that the cell boundaries are not fixed but randomly determined by some other criteria. The chi-squared statistic from this kind of table does not have a limiting chi-distribution in general, and we need to be cautious in the routine use of a chi-squared type goodness-of-fit test.

References

Agresti, A. (2002). *Categorical Data Analysis*, 2nd ed., Wiley, New York.

Brown, C. C. (1982). On a goodness-of-fit test for the logistic model based on score statistics, *Communications in Statistics, Theory and Methods*, **11**, 1087–1105.

Bull, S. (1994). *Analysis of Attitudes toward Workplace Smoking Restrictions*, In N. Lange, L. Ryan, D. Billard, L. Conquest, and J. Greenhouse (eds.): Case Studies in Biometry. Wiley, New York, 249–271.

Chernoff, H. and Lehmann, E. L. (1954). The use of maximum likelihood estimates in χ^2 tests for goodness of fit, *The Annals of Mathematical Statistics*, **25**, 579–586.

Cox, D. R. (1958). Two further applications of a model for binary responses, *Biometrika*, **45**, 562–565.

Goodman, L. A. (1979). Simple models for the analysis of association in cross-classifications having ordered categories, *Journal of the American Statistical Association*, **74**, 537–552.

Graubard, B. I., Korn, E. L. and Midthune, D. (1997). Testing goodness-of-fit for logistic regression with survey data, In *Proceedings of the Section on Survey Research Methods, American Statistical Association*.

Hosmer, D. W. and Lemeshow, S. (1980). Goodness-of-fit tests for the multiple logistic regression model, *Communications in Statistics, Theory and Methods*, **9**, 1043–1069.

Jeong, H. C., Jhun, M. and Kim, D. (2005). Bootstrap test for independence in two-way ordinal contingency tables, *Computational Statistics & Data Analysis*, **48**, 623–631.

Jeong, K. M. and Lee, H. Y. (2009). Goodness-of fit tests for the ordinal response models with misspecified links, *Communications of the Korean Statistical Society*, **16**, 697–705.

Kuss, O. (2002). Global goodness-of-fit tests in logistic regression with sparse data, *Statistics in Medicine*, **21**, 3789–3801.

Lipsitz, S. R., Fitzmaurice, G. M. and Molenberghs, G. (1996). Goodness-of-fit tests for ordinal response regression models, *Applied Statistics*, **45**, 175–190.

Moore, D. S. (1971). A chi-square statistic with random cell boundaries, *The Annals of Mathematical Statistics*, **42**, 147–156.

Moore, D. S. and Spruill, M. C. (1975). Unified large sample theory of general chi-squared statistics for tests of fit, *The Annals of Statistics*, **3**, 599–616.

Osius, G. and Rojek, D. (1992). Normal goodness-of-fit tests for multinomial models with large degrees of freedom, *Journal of the American Statistical Association*, **87**, 1145–1152.

Pigeon, J. G. and Heyse, J. F. (1999a). An improved goodness-of-fit statistic for probability prediction models, *Biometrical Journal*, **41**, 71–82.

Pigeon, J. G. and Heyse, J. F. (1999b). A cautionary note about assessing the fit of logistic regression models, *Journal of Applied Statistics*, **26**, 847–853.

Pulkstenis, E. and Robinson, T. J. (2004). Goodness-of-fit tests for ordinal response regression models, *Statistics in Medicine*, **23**, 999–1014.

Su, J. Q. and Wei, L. J. (1991). A lack-of-fit test for the mean function in a generalized linear model, *Journal of the American Statistical Association*, **86**, 420–426.

Tsiatis, A. A. (1980). A note on a goodness-of-fit test for the logistic regression model, *Biometrika*, **67**, 250–251.