A Latent Scale Linear Model for Ordered Categorical Responses

Uesaka, Hiroyuki
Department of Information Systems, Interdisciplinary Graduate School of Engineering Sciences, Kyushu University

Asano, Choichiro
Department of Information Systems, Interdisciplinary Graduate School of Engineering Sciences, Kyushu University

https://doi.org/10.15017/17615

出版情報：九州大学大学院総合理工学報告. 6 (2), pp.259-265, 1985-01-01. 九州大学大学院総合理工学研究科

A Latent Scale Linear Model for Ordered Categorical Responses

Hiroyuki UESAKA and Chooichiro ASANO

(September 29, 1984)

The present paper is related to a contingency table with an ordered categorical response. The ordered categories are assumed to be manifestations of an unobservable quantitative response or trait, and the linear models for location parameters of the underlying distributions are considered. This model is referred to as the latent scale linear model. It is stressed that latent scales and families of distributions are substantial to the models and their properties are discussed. Several models proposed earlier are reconsidered in the framework of the latent scale linear model. The model is illustrated through two examples.

1. Introduction

The analysis of a contingency table with an ordered categorical response is of great interest in various fields, and recently several models based on transformations of probabilities have been proposed, e.g. McCullagh [17], Goodman [12]. On the other hand, models based on a latent continuous response have been considered in analysis of stimulus-response relationship, e.g. Ashford [4], Gurland, Lee and Dahm [13], Farewell [10]. Similar models have been considered for data obtained by successive categories or rating scale methods in psychometric contexts, e.g. Saffir [19], Edwards and Thurstone [9], Bock [5], Andrich [2].

In this paper, the ordered categories are assumed to be manifestations of an unobservable quantitative response or trait, and linear models are considered for location parameters of the underlying distributions. This model will be referred to as latent scale linear model. In section 2, a latent scale linear model is defined and some considerations are given on situations to which such a model applies. In section 3, the families of latent distributions and their properties are discussed. Estimation and testing statistical hypotheses are briefly summarized in section 4. In section 5, two data sets are analysed and implication of the model is discussed. Additional comments are given in the final section.

Department of Information Systems, Interdisciplinary Graduate School of Engineering Sciences, Kyushu University

2. The latent scale linear model

2.1 Examples of ordered categorical responses

Let us illustrate three typical situations where the ordered categories are thought as manifestations of latent variables.

(i) Biological stimulus-response analysis. A subject, exposed to a stimulus intensity x, has a latent continuous response $z(x)$ and is manifestly observed as one of the ordered categories, say the j-th category, in case of $\tau_{j-1} \leq z(x) < \tau_j$, $j=1, \cdots, J$, where $\tau_1, \cdots, \tau_{J-1}$ are unknown threshold values and $\tau_0=-\infty$ and $\tau_J=\infty$. Generally, the latent response for the stimulus intensity x varies depending upon uncontrollable internal and external variations and is assumed to have a distribution $F_x(z)=F(z-\mu(x))$ on a suitably chosen latent scale, where $\mu(x)$ is a function of x. Such a model has been considered by Ashford [4] and Gurland, Lee and Dahm [13] on the basis of normal or logistic distribution, and can be extended to situations with multiple stimuli using multivariable function $\mu(x)$.

(ii) Psychometric analysis on scaling. In psychometric experiments, measures on an ordered categorical scale such as preferences, attitudes or opinions of subjects are thought to be manifestations of latent continuous varibles. It is further assumed that when the latent variable of a subject falls in an interval $[\tau_{j-1}, \tau_j)$, the subject manifestly responds to the j-th category. The successive intervals $[\tau_{j-1}, \tau_j)$, $j=1,\cdots,J$ are constant over the subjects and the distribution of the latent variable under a condition x is $F(z-\mu(x))$. Though it is assumed (e.g. Saffir [19], Edwards and Thurstone [9]) that the latent variable is normally distributed, this assumption should be critically checked and if this is not the case, other distribution should be tried.

(iii) Sensory analysis on rating experiments. When a subject is rated on an ordinal scale with J categories, the subject is located at a point on a latent continuum, and classified into one of the successive categories.

2.2 The general latent scale linear model

Suppose an $I\times J$ contingency table $\{n_{ij}\}$ is defined by I samples of the respective sizes $n_1,\cdots,n_I$ and jointly J ordered response categories. Let $\{p_{ij}\}$ be the expected proportions, and denote cumulative probabilities by

$$q_{ij}=p_{i1}+\cdots+p_{ij},\ j=1,\cdots,J,\ i=1,\cdots,I. \quad (1)$$

Now let us define the latent scale linear model. Suppose that the observed categories $\{z^*_{i\alpha}\}$ are manifestations of latent continuous variables $\{z_{i\alpha}\}$. Let $F(z)$ be a continuous distribution function whose support is the whole real axis and let $\tau_1,\cdots,\tau_{J-1}$ be unknown constants, satisfying a constraint

$$-\infty=\tau_0<\tau_1<\cdots<\tau_{J-1}<\tau_J=+\infty. \quad (2)$$

Then assuming that $z_{i\alpha}$, $\alpha=1,\cdots,n_i$ are independently and identically distributed in $F(z-\mu_i)$, $i=1,\cdots,I$, we have

$$\Pr.(z^*_{i\alpha}\leq j)=q_{ij}=F(\tau_j-\mu_i),\quad j=1,\cdots,J-1,\ i=1,\cdots,I, \quad (3)$$

where $\mu_1,\cdots,\mu_I$ are unknown location parameters. Consider a linear model

$$\mu_i=\boldsymbol{x}_i\boldsymbol{\beta},\ \ i=1,\cdots,I, \quad (4)$$

where $\boldsymbol{\beta}$ is a p-dimensional unknown parameter vector of $\beta_1,\cdots,\beta_p$ and $\boldsymbol{X}=(\boldsymbol{x}_1'\cdots,\boldsymbol{x}_I')'$ is an $I\times p$ known matrix. For the uniquness of parameters, we assume that

$$\boldsymbol{Xa}\neq\boldsymbol{1}_I,\ \text{for any}\ \boldsymbol{a}\neq\boldsymbol{0}, \quad (5)$$

where $\boldsymbol{1}_I$ is the p-dimensional vector of $1's$. Thus the transformation of cumulative probability $F^{-1}(q_{ij})$ is represented as an additive model of row parameter μ_i and column parameter τ_j.

3. The latent distributions

The above model implies that if the scale of the latent variable is suitably chosen, the differences in distributions of the given I populations are ascribable to those of location parameters. But since the latent scale is unknown,

Table 1. Typical distributions of latent variables.

Name	Definition
Normal	$F(z-\mu)=\int_{-\infty}^{z-\mu}\frac{1}{\sqrt{2\pi}}\exp(-x^2/2)dx$
Logistic	$F(z-\mu)=\frac{1}{1+\exp\{-(z-\mu)\}}$
Double exponential of the first kind	$F(z-\mu)=\exp[-\exp\{-(z-\mu)\}]$
Double exponential of the second kind	$F(z-\mu)=1-\exp[-\exp\{(z-\mu)\}]$
Generalized logistic	$F(z-\mu;\nu)=\left\{\frac{1}{1+\exp[-(z-\mu)]}\right\}^{\nu}$
Reversed generalized logistic	$F(z-\mu;\nu)=1-\left\{\frac{1}{1+\exp(z-\mu)}\right\}^{\nu}$
Log gamma	$F(z-\mu;\nu)=\int_{-\infty}^{z-\mu}\frac{1}{\Gamma(\nu)}\exp\{\nu x-\exp(x)\}dx$
Reversed log gamma	$F(z-\mu;\nu)=\int_{-\infty}^{z-\mu}\frac{1}{\Gamma(\nu)}\exp\{-\nu x-\exp(-x)\}dx$

it is to be estimated. This also means that the distributional form is to be estimated. It is true that the normal or logistic distribution is common in such situations, but these two distributions sometimes fail to fit to data and skewed distributions have come into use. In this section, we will discuss properties of latent scales and distributions. Though there are a lot of families of distributions and any distribution whose support is the whole real axis may be used, several useful distributions are known. These are given in Table 1.

First let us consider the effect of transformation of the latent scale.

(a) Suppose for I distributions $G(y,\beta_i)$, $i=1,\cdots,I$, there exists a monotonic transformation $z=z(y)$, which is independent of β_i's, and that $G(y,\beta_i)=F(z-\mu_i)$, $i=1,\cdots,I$ holds, then we can treat $F(z-\mu_i)$'s instead of $G(y,\beta_i)$'s. Thus typical positive distributions such as the exponential, Weibull, (generalized) gamma and second extreme value distributions are transformed to log-gamma distribution, F_{ν}^{1}, Pareto are transformed to the generalized logistic distribution by logarithmic transformations.

The next proposition is a direct consequence of the property (a).

(b) When the latent distributions $G_i(y,\beta)$, $i=1,\cdots,I$ are expressed by the Cox's proportional hazard model (Cox [8]), the latent scale linear model is given on the basis of the double exponential distribution of the second kind. In fact if

$$1-G_i(y;\beta)=\exp\{-\exp(\boldsymbol{x}_i\boldsymbol{\beta})\Lambda_0(y)\},\quad i=1,\cdots,I,$$

then the transformed variable $z=\log[\Lambda_0(y)]$ has a double exponential distribution of the second kind.

(c) The categories on a bipolar scale, which often appears in attitude measurements, can be ordered in two directions. The family of a symmetric distribution is invariant under reverse of category order, but those of asymmetric distributions change from $F(z-\mu)$ to

$1-F(-(z-\mu))$. The reversed family will be called by heading "reverse" to the name of the original family.

Extension and generalization of a family: Any location family of distributions $F(z-\mu)$, $-\infty<\mu<+\infty$ is extended by $\{F(z-\mu)\}^{\nu}$ or $1-\{1-F(z-\mu)\}^{\nu}$, where ν is any positive value. These distributions are the extended families of maxima and minima of random samples from $F(z-\mu)$. This extension can be applied to any distribution but for the double exponential distributions of the first and second kind. Let us give some considerations on latent distributions.

(d) The generalized logistic distribution is the generalized distribution of maxima of logistic samples. This distribution is obtained by compounding double exponential distribution of the first kind and log-gamma distribution, and is asymptotically equivalent to the double exponential distribution of the first kind as ν tends to infinity. Also this is a special case of the generalized F defined by Prentice [18]. The reversed generalized logistic distribution is the generalized distribution of minima of logistic samples. This distribution is obtained by compounding double exponential distribution of the second kind and log-gamma distribution. Farewell [10] has obtained this distribution in a different form. This is asymptotically equivalent to the double exponential distribution of the second kind as ν tends to infinity.

(e) The (reversed) log-gamma distribution is an extended family of the double exponential distribution of the second (first) kind.

Relationship to the other proposed model:

(f) The proportional odds model of McCullagh [17] is given on the basis of the logistic distribution.

(g) The asymmetric power transformation of probabilities proposed by Aranda-Ordaz [3] is equivalent to the generalized (reversed) logistic model. In fact, the model

$$\log\{(1-q_{ij})^{-1/\nu}-1\}=\tau_j-\boldsymbol{x}_i\boldsymbol{\beta},\quad \nu>0, \tag{6}$$

is equivalent to that based on the reversed generalized logistic distribution. Similarly the model

$$\log\{(q_{ij})^{-1/\nu}-1\}=-(\tau_j-\boldsymbol{x}_i\boldsymbol{\beta}),\quad \nu>0 \tag{7}$$

is equivalent to that defined by the generalized logistic distribution. When ν tends to infinity (6) and (7) approach to the double exponential distribution of the first and second kind, respectively.

4. Parameter estimation and testing statistical hypotheses

The parameters may be estimated by several methods, e.g. the method of maximum likelihood. Testing statistical hypotheses will be consequently performed on the basis of likelihood ratio and Wald criterion. These are well known and omitted in the present paper.

5. Illustrative examples

We will give two illustrative examples, the data have been analysed by several authors. The present analysis will give deeper insight into the data.

5.1 Mental health study

A 6×4 table on mental health status (four response categories) by socioeconomic status (six categories) was analysed by Haberman [14] using decomposition of interaction terms of a log-linear model. The data is given in table 5 of Haberman [14]. Let

$$X_{(1)}=\begin{pmatrix} -5 & 5 & -5 & 1 & -1 \\ -3 & -1 & 7 & -3 & 5 \\ -1 & -4 & 4 & 2 & -10 \\ 1 & -4 & -4 & 2 & 10 \\ 3 & -1 & -7 & -3 & -5 \\ 5 & 5 & 5 & 1 & 1 \end{pmatrix}.$$

The model defined by $X=X_{(1)}$ was fitted to every family in Table 1. For assessing goodness of fit of the model, the likelihood ratio chi-square statistic

$$G=2\sum n_{ij}\log\left(n_{ij}/n_i\,\hat{p}_{ij}\right)$$

was computed for each family, where $\{\hat{p}_{ij}\}$ are maximum likelihood (ML) estimates of $\{p_{ij}\}$. The goodness of fit statistics, with 10 degrees of freedom, are given below.

Goodness of fit of the four parametric families in Table 1.

Distribution	Normal	Logistic	Double exponential First	Double exponential Second
G with 10 df	5.591	7.827	10.396	8.159

The extended families showed little improvement in fitness, so omitted from the table. Thus the normal distribution was adopted. The estimated β_j's, their estimated standard deviations and Wald stastitic for testing H_j: $\beta_j=0$ are given in Table 2. We can conclude that the terms of order over 2 are unnecessary. The goodness of fit statistics G for the model of order up to 1 and 2 were $G=8.717$ and $G=6.219$, respectively. The difference is 2.498, which is chi-squared statistic with one degree of freedom for testing H_2. Thus the model of order 1 was accepted. The estimated parameters and their variance-covariance matrix are given below.

Parameter	Estimate	Variance-covariance matrix ($\times 10^4$)			
τ_1	−0.910	12.962	4.741	3.116	−0.192
τ_2	0.119	4.741	9.594	6.360	−0.008
τ_3	0.732	3.116	6.360	11.622	0.092
β	0.051	−0.192	−0.008	0.092	0.674

We can conclude that the scale values of the six socioeconomic groups are equally spaced in order of socioeconomic status, and the mental health status gets better according to the socioeconomic status.

Haberman [14] has adopted the model having linear by linear interaction effect. The resultant G was 9.73 with 14 degrees of freedom. Our model defined by the linear trend for the main effects of socioeconomic status gave smaller G value than that of Haberman. Further our model may be interpreted more easily than that of Haberman.

5.2 Study on severity of cancer

Table 3 gives the frequency distributions of severity of cancer, Hirotsu [15, 16]. Hirotsu [16] analysed the data by cumulative chi-square method, and found that ten occupations consisted of two homogenous groups. The differences

Table 2. Estimates of parameters and their estimated standard deviations and Wald statistics for testing hypotheses on the parameter values.

Structure parameter	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$
Estimate	0.0523	0.0111	−0.0013	0.0038	−0.0028
SD	0.0083	0.0072	0.0050	0.0123	0.0038
X_W^2	39.705	2.376	0.068	0.095	0.366

Table 3. Frequency table of severity of cancer and the estimated location parameters and their SD's obtained in the reversed generalized logistic model and $X_{(1)}$.

Occupation	Mild	Moderate	Sever	$\hat{\beta}_i$	SD $(\hat{\beta}_i)$
A	148	444	86	0	0
B	111	352	49	−0.019	0.337
C	645	1911	328	−0.105	0.246
D	165	771	119	0.835	0.274
E	383	1829	311	0.909	0.247
F	96	290	47	−0.062	0.355
G	98	330	58	0.200	0.337
H	199	870	155	0.763	0.257
I	59	199	30	0.147	0.399
J	262	1320	236	1.027	0.252

between groups are due to difference in the proportions of the first category. Yanagimoto and Shimizu [20] analysed the same data using the discrete proportional hazard model of Cox [8] and found the model defined below fitted well ($G=8.90$ and 17 degrees of freedom).

$$(a_1, \cdots, a_{10})'=\tau_1\mathbf{1}-X_{(2)}\beta_1 \quad (8)$$
$$(b_1, \cdots, b_{10})'=\tau_2\mathbf{1}$$

where

$$a_i=\log\frac{q_{i1}}{1-q_{i1}},\ b_i=\log\frac{q_{i2}}{1-q_{i2}},\ i=1, \cdots, 10,$$

and $X_{(2)}=(1,1,1,0,0,1,1,0,1,0)'$.

On the other hand, we analyse the same data by latent scale linear model. First the model defined by $X_{(1)}=[0_9, I_9]'$ was fitted. The goodness of fit for several distributions is summarized as follows:

Distribution	Goodness of fit statistic	Degrees of freedom	Shape parameter
Normal	33.065	9	
Logistic	29.819	9	
Double exponential (first)	9.616	9	
Double exponential (second)	66.621	9	
Reversed generalized logistic	5.248	8	0.10
Reversed log-gamma*	6.568	8	0.50

*Not searched enough

The results indicate that the latent distributions are highly positively skewed and proportional hazard model is unsatisfied. The reversed generalized logistic distribution gave the smallest G value of 5.248 with 8 degrees of freedom. The double exponential distribution of the first kind fitted fairly well to the data but there is appreciable difference in the likelihoods between the two distributions. Thus the reversed generalized logistic distribution is adopted. The estimated parameter values of β_i's and their estimated SD's based on the reversed generalized logistic distribution are given in the last two columns of Table 3. The β_i is the difference in location parameter between the first and the $(i+1)$-th occupations, $i=1, \cdots, 9$. These values suggest that the occupations consist of two homogeneous groups (A, B, C, F, G, I) and (D, E, H, J), which is the same as obtained by Hirotsu [15, 16] and Yanagimoto and Shimizu [20]. Following this grouping, we fitted the model $X_{(2)}$. The goodness of fit statistic was $G=8.867$ with 16 degrees of freedom. The maximum value of sample-wise components of G was 1.75, thus the model fits well to every occupation. The estimated parameters and their estimated variance-covariance matrix are as follows:

Parameter	Estimate	Variance-covariance matrix $(\times 10^4)$		
τ_1	2.680	74.788	72.767	74.776
τ_2	24.213	72.767	81.450	77.130
β	1.017	74.776	77.130	122.146

6. Further discussions

The present paper discussed a latent scale linear model for a contigency table defined by factor variables and an ordered categorical response. The model was originally proposed on the basis of normal assumption of the latent distributions as a natural extension of probit model for quantal responses, Ashford [4]. In fact when the number of categories is two, our model based on normal or logistic distribution reduces to the linear model for probit, Finney [11], or linear logistic model, Cox [7]. The generalized logistic model reduces to that of Prentice [18].

Three generalizations of latent scale linear model may be considered. The first is transformation of quantitative regressor variable instead of higher order polynomials as is discussed by Box and Tidwell [6] in normal regression problem. The second is nonlinear model which includes scale parameters, Mc Cullagh [17]. The third is ordinal regression situation with ordered categorical response, Anderson and Philips [1].

Our model is also applicable to scaling problem in the method of successive intervals by assuming the homogeneity of scale parameters and deleting normal assumption on the latent distributions. The internal consistency check is given by assessing goodness of fit of the model.

REFERENCES

1) Anderson, J. A. and Philips, P. R.: Regression, discrimination and measurement models for ordered categorical variables. Appl. Statist. 30 (1981) 22–31.
2) Andrich, D.: A model for contingency table having an ordered response classification. Biometrics 35 (1979) 403–415.
3) Aranda-Ordaz, F. J.: On two families of transformations to additivity for binary response data. Biometrika 68 (1981) 357–363.
4) Ashford, J. R.: An approach to the analysis of data for semi-quantal responses in biological assay. Biometrics 15 (1959) 573–581.
5) Bock, R. D.: Multivariate Statistical Methods in Behavioral Research. (1975) McGraw-Hill. New York.
6) Box, G. E. P. and Tidwell, P. W.: Transformations of the independent variables. Technometrics 4 (1962) 531–550.
7) Cox, D. R.: Analysis of Binary Data. (1970) Chapman-Hall.
8) Cox, D. R.: Regression models and life-tables (with discussion). Jour. Roy. Statist. Soc. B 34 (1972) 187–220.
9) Edwards, A. L. and Thurstone, L. L.: An internal consistency check for scale values determined by the method of successive intervals. Psychometrika 17 (1952) 169–180.
10) Farewell, V. T.: A note on regression analysis of ordinal data with variability of classification. Biometrika 69 (1982) 533–538.
11) Finney, D. J. : Probit Analysis. (1971) Cambridge University Press.
12) Goodman, L. A.: The analysis of dependence in cross-classifications having ordered categories, using log-linear models for frequencies and log-linear models for odds. Biometrics 39 (1983) 149–160.
13) Gurland, J., Lee, I. and Dahm, P. A.: Polychotomous quantal response in biological assay. Biometrics 16 (1960) 382–397.
14) Haberman, S. J.: Log-linear models for frequency tables with ordered classifications. Biometrics 30 (1974) 589–600.
15) Hirotsu, C.: Multiple comparisons and grouping rows in a two-way contingency table. Rep. Statist. Appl. Res., Union Jap. Sci. Eng. 25 (1978) 1-12.
16) Hirotsu, C.: Defining the pattern of association in two-way contingency tables. Biometrika 70 (1983) 579–589.
17) McCullagh, P.: Regression models for ordinal data (with discussion). Jour. Roy. Statist. Soc. B 42 (1980) 109–142.
18) Prentice, R. L.: A generalization of the probit and logit models for dose response curves. Biometrics 32 (1976) 761–768.
19) Saffir, M. A. : A comparative study of scales constructed by three psychophysical methods. Psychometrika 2 (1937) 179–191.
20) Yanagimoto, T. and Shimizu, E.: Application of the proportional hazard model to two-dimensional contingency tables. Japanese Jour. Appl. Statist. 12 (1983) 17–29.