INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA

UMI®
800-521-0600

UNIFIED ORDINAL REGRESSION: MODEL ASSESSMENT AND SEMIPARAMETRIC ANALYSIS

BY

LIMIN FU

B.S., Beijing Institute of Technology, 1991
M.A., Eastern Illinois University, 1995

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Statistics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2000

Urbana, Illinois

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

THE GRADUATE COLLEGE

DECEMBER 1999

(date)

WE HEREBY RECOMMEND THAT THE THESIS BY

LIMIN FU

ENTITLED UNIFIED ORDINAL REGRESSION:

MODEL ASSESSMENT AND SEMIPARAMETRIC ANALYSIS

BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR

THE DEGREE OF DOCTOR OF PHILOSOPHY

Douglas G. Simpson

Director of Thesis Research

Adam T. Martinsek

Head of Department

Committee on Final Examination†

Douglas G. Simpson

Chairperson

John J. Marden

†Required for doctor's degree but not for master's.

O-517

ABSTRACT

Various forms of logit model have been employed in the regression analysis of ordinal response data. In the first part of the thesis we adapt the theory and methodology of generalized estimating equations (GEE) and a binary coding of the ordinal response so that the various forms of logit model can be handled in a unified fashion. We develop Rao-type generalized score tests for model assessment within this framework.

In the second part of the thesis we propose a class of latent structure models to the analysis of longitudinal ordinal data. We assume the observed ordinal scale is a manifestation of a latent continuous variable categorized by a set of unknown threshold values. The dependence among the repeated measurements for a subject is modeled through the latent continuous variable. Monte Carlo EM (MCEM) is applied to obtain the maximum likelihood estimates.

In the third part of the thesis we extend the generalized additive models proposed by Hastie and Tibshirani (1984) to multivariate data and propose a class of multivariate generalized additive models, which model the correlation structure of the components of a multivariate observation as well as the marginal means.

In the last part of the thesis we discuss the methods for handling censored data for both categorical and continuous responses. We develop EM algorithms for censored data in general, and also develop a weighted least squares algorithm for cumulative link models for ordinal responses.

To my parents, and my lovely and beautiful fiancee, Annie

ACKNOWLEDGEMENTS

I wish to express my deepest gratitude to my thesis advisor, Professor Douglas Simpson, for suggesting the research problems and for his guidance and direction throughout the entire research. Without his encouragement and generous support, this thesis would not have been possible.

I also would like to thank the members of my thesis committee (Professor Xuming He, Professor Peter Imrey, and Professor John Marden) for their comments and helpful suggestions.

I am so deeply indebted to my family for what I am today. My parents deserve my greatest appreciation for their love and support over years. I am so lucky to have been born into this my family.

Finally, to my fiancee, Annie, for her love and support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Introduction

1.1 Ordinal Response Data

It is widely recognized that the types of data as well as the class of problems that a statistician is likely to encounter vary greatly with the field of research. For example, the overwhelming proportion of data in the physical sciences is essentially quantitative although possibly measured on an arbitrary scale. In the social sciences and biological sciences, qualitative data are more common. These qualitative measurements, whether subjective or objective, usually take values in a limited set of categories which may be on a ordinal or on a purely nominal scale.

Ordinal scales are pervasive in the social sciences, in particular for measuring attitudes and opinions on various issues and status of various types. Ordinal scales are also commonly occur in such diverse fields as marketing (e.g., ordinal preference scales or resource scales) and medical and public health disciplines (e.g., for variables describing amount of exposure to a potentially harmful substance, stages of a disease, degree of recovery from an illness, and severity of an injury). In all fields ordinal scales often result when discrete measurement is used with inherently continuous variables such as age, education, and degree of prejudice.

This thesis is devoted to the statistical analysis and modeling of ordinal response data. The methodological and theoretical research summarized here is motivated by our

experience with ordinal data in consulting projects and interdisciplinary collaborations with scientists. In the remainder of Chapter 1 we provide historical background on methods for ordinal data, focusing primarily on regression models for ordinal response data, and we provide an overview of the research summarized in this thesis.

1.1.1 Historical Note

There are a variety of models proposed for ordinal data. Many of these may be formulated as logit models. Such models describe effects of a set of explanatory variables on a response probability through logit transformations. Ordinal logit models are generalizations of logit models for binary responses. To reflect the ordinal nature of the responses, different types of logits for ordered response categories, such as cumulative logits, continuation ratio logits, and adjacent category logits, are used to build models. Agresti (1990) and Clogg and Shihadeh (1994) discussed these models in detail. Currently cumulative logit models appear to be the most popular models in applications. The article by McCullagh (1980) is a good one for motivating their use. They have the appealing feature of the existence of an underlying continuous and perhaps unobservable random variable. Motivation for the continuation logit models is that the results of fitting models for separate logits are independent. Thompson (1977) proposed a model having these logits for the analysis of discrete survival-time data. He showed that when the lengths of the time intervals approach zero, his model converges to Cox's (1972) proportional hazard model for survival data. Models using adjacent category logits are presented by Goodman (1983). His models are equivalent to ordinal loglinear models. As generalizations of the proportional odds models, Agresti (1990) described another class models, the cumulative link models, which represent the cumulative probabilities of the response via a strictly monotone function F^{-1} from $(0, 1)$ on the real line. All the models of this type share the property that the categories can be thought of as contiguous intervals on some continuous scale. They differ in their assumptions concerning the distributions of the latent variable. McCullagh (1980) discussed several links, such as logit link, probit link, and complementary log-log link.

Recently, considerable progress has been made on the methodology for the analysis of ordinal longitudinal data. The defining characteristic of a longitudinal study is that individuals are measured repeatedly through time, so that statistical inference must recognize the likely correlation structure in the data. A widely used approach for many problems is the generalized estimating equations (GEE) method proposed by Liang and Zeger (1986) and Prentice (1988). Several authors have adopted the GEE approach to proportional odds models for clustered ordinal responses (Clayton (1992), Gang et al. (1993), Miller et al. (1994), Heagerty and Zeger (1996)). Simpson et al. (1996a) developed the methodology for regression analysis of interval censored and clustered ordinal data. Another approach is to use random effects models, in which the regression coefficients measure the more direct influence of explanatory variables on the responses for heterogeneous individuals. This approach has been discussed by Albert and Chib (1993), and by Xie, Simpson and Carroll (2000). Modeling the marginal expectation and treating the correlation as a nuisance may be less appropriate when the time course of the outcome for each subject is of primary interest or when the correlation itself has scientific relevance. Transitional models describe the conditional distribution of present response as an explicit function of past responses. Diggle, Liang and Zeger (1996) presented a transitional model based on cumulative logits.

Non-parametric models provide a flexible tool for understanding covariate effects. They can be used in a data analytic fashion to model and to test hypotheses about covariates. They can also be viewed as diagnostics for identifying functional form. The fitted functions can used to inspire parsimonious reparameterization of variables. Generalized additive proportional odds regression was proposed by Hastie and Tibshirani (1987 and 1990). Yee and Wild (1996) proposed vector generalized additive models, which include proportional odds models as a special case. These are all in the framework of the generalized additive models (GAM) introduced by Hastie and Tibshirani (1984 and 1986).

1.1.2 Regression Models for Ordinal Responses

When response categories have a natural ordering, models for this kind of data should utilize that ordering. There are different ways we can incorporate the ordering in the model. This section introduces different types of regression models for ordered response categories.

1.1.2.1 Logit Models for Ordinal Responses

One way to model response variables having more than two categories is to use generalized logit models, which are generalizations of logit models for binary responses. We incorporate the ordering directly in the way we construct logits. There are three types of logits for ordinal data.

(a) Proportional Odds Models

Let Y be a ordinal response variable takes value in $0, 1, 2, \ldots, S$, and x be a vector of covariates associated with Y. The proportional odds models are defined as

$$\text{logit}\{Pr(Y \geq s \mid x)\} = \alpha_s + x'\beta, \qquad s = 1, \ldots, S. \tag{1.1}$$

This model assumes a variable's effect on the odds of response at or above category s is the same for all s. It satisfies

$$\begin{aligned} \text{logit}\{Pr(Y \geq s \mid x_1)\} - \text{logit}\{Pr(Y \geq s \mid x_2)\} &= \log\left[\frac{Pr(Y \geq s \mid x_1)Pr(Y < s \mid x_2)}{Pr(Y \geq s \mid x_2)Pr(Y < s \mid x_1)}\right] \\ &= (x_1 - x_2)'\beta. \end{aligned}$$

Its interpretation is that the odds of making response $\geq s$ are $\exp[(x_1 - x_2)'\beta]$ times higher at $x = x_1$ than at $x = x_2$. Models of the form (1.1) describe strict stochastic ordering, i.e. either

$$Pr(Y \geq s \mid x_1) > Pr(Y \geq s \mid x_2) \qquad \text{for all } s$$

or

$$Pr(Y \geq s \mid x_1) < Pr(Y \geq s \mid x_2) \qquad \text{for all } s$$

according as $(x_1 - x_2)'\beta > 0$ or $(x_1 - x_2)'\beta < 0$.

Motivation for this kind of model is provided by an appeal to the existence of an underlying continuous and perhaps unobserved random variable.

(b) Continuation Ratio Models

The continuation ratio models provide an alternative way of constructing logit for ordinal response, which have the form

$$\text{logit}\{Pr(Y = s \mid Y \geq s, x)\} = \alpha_s + x'\beta_s, \qquad s = 1, \ldots, S. \tag{1.2}$$

The marginal probabilities can be obtained by recursion.

$$\begin{aligned} Pr(Y \leq s \mid x) &= \prod_{t=s+1}^{S} \{1 - Pr(Y = t \mid Y \leq t, x)\} \\ &= \prod_{t=s+1}^{S} \left\{1 - \frac{\exp(\alpha_t + x'\beta_t)}{1 + \exp(\alpha_t + x'\beta_t)}\right\}. \end{aligned}$$

Model (1.2) is often used in the context of survival analysis. We can imagine the ordered categories as failure times in increasing order. as noted above, Thompson (1977) proposed a model having these logits for the analysis of discrete survival time data, and showed that the model converges to Cox's (1972) proportional hazards model when the length of the time intervals approach zero. If we replace the logit link function with the complementary log-log link function, the discrete proportional hazard model is obtained (see Kalbfleisch and Prentice (1980), Chapter 2).

(c) Adjacent Category Odds Models

Another way to use ordered response categories is by forming logits of adjacent categories. The model is given by

$$\log\left\{\frac{Pr(Y=s \mid x)}{Pr(Y=s-1 \mid x)}\right\} = \alpha_s + x'\beta_s, \qquad s = 1,\ldots,S. \tag{1.3}$$

The adjacent category odds model is simply a reparameterized reference category logit model.

1.1.2.2 Cumulative Link Models

An equivalent form of the proportional odds models is given by

$$\text{logit}\{Pr(Y \leq s \mid x)\} = \alpha_s - x'\beta, \qquad s = 1,\ldots,S.$$

This model assumes that effects of x are the same for each cutpoint. This assumption holds if there is a linear regression for an underlying continuous response having standardized logistic distribution. $x'\beta$ is the location parameter, and $\alpha_1,\ldots,\alpha_S$ is a set of cutpoints which categorize the underlying continuous response into $S+1$ ordered categories. A generalization of the proportional odds model is to use other monotone transformations. Let F denote the CDF of a continuous random variable having positive density over the entire real line. The F^{-1}, so called link function, is a strictly monotone function from $(0,1)$ onto the real line. The cumulative link model has the form

$$F^{-1}\{Pr(Y \leq s \mid x)\} = \alpha_s - x'\beta,$$

or, equivalently

$$Pr(Y \leq s \mid x) = F(\alpha_s - x'\beta,) \tag{1.4}$$

McCullagh (1980) discussed several cumulative link models. The logit link, $F^{-1}(u) = \log[u/(1-u)]$, gives the proportional odds models. The standard normal CDF $F = \Phi$

gives the threshold probit models, a generalization of the binary probit model to ordinal data. The complementary log-log link, $F^{-1}(u) = \log[-\log(1-u)]$ is appropriate when the underlying distribution follows an exponential or extreme-value distribution. Artchison and Silvey (1957), Bock and Jones (1968, Chapter 8), and Gurland et al. (1960) have used cumulative probit models. Prentice and Gloeckler (1978) used the complementary log-log model to analyze grouped survival data. Farewill (1982) generalized it to allow for variation among the sample in the values regarded as category boundaries for the underlying scale. Genter and Farewell (1985) introduced a generalized link function that permit comparison of fits provided by probit, complementary log-log, and other links.

1.2 Censored Data

In many situations, some responses we observe may only contain partial information. Information for a categorical variable may be known only to fall into a subset of the categories instead of the exact category, or information for a continuous variable may be only known to lie in an interval of real line. Data of this type are called censored data. When referring to the censored data mechanism, we use the terminology of Little and Rubin (1987). A censoring process is said to be *missing completely at random* (MCAR), if the censoring is independent of both observed and unobserved data, and *missing at random* (MAR) if conditional on the observed data, the censoring is independent of the unobserved data. A process that is neither MCAR nor MAR is called informative.

Previous work on censored categorical data has largely been in the context of survey data and partially classified contingency table. Including Hartley (1958), Blumenthal (1968), Koch, Imrey, and Reinfurt (1972), Chen and Feinberg (1976), and Shipp, Howe, Watson, and Hogg (1991). The methodology for analysis of censored ordinal data has been discussed by Simpson et al. (1996), and Xie, Simpson and Carroll (2000). Various authors, including Baker and Laird (1988), Chambers and Welsh (1993), and Molenberghs and Goetghebeur (1997), have used the EM algorithm to maximize the observed data likelihood via the complete data likelihood for missing data problems.

For censored continuous variables, the main focus of research is on survival analysis. For right censored data, parametric and non-parametric methods are available (Lawless 1982, Kalbfleisch and Prentice 1980). Interval censoring is another mechanism of censoring, which needs special treatment. Turnbull (1976) studied the empirical distribution function with arbitrarily grouped, censored and truncated data. Finkelstein (1986) proposed a method for fitting the proportional hazard model for interval censored data. Odell, Anderson, and D'Agostino (1992) applied a Weibull-based accelerated failure time model for interval censored data. Kim (1997) discussed analyzing interval censored failure time data using a loglinear model.

1.3 Overview

In Chapter 2, we adapt the theory and methodology of generalized estimating equations and a binary coding of the ordinal response so that the various logit models can be handled in a unified fashion. This approach is also well suited for alternatives to logit models such as ordinal probit analysis. We develop Rao-type generalized score tests for model assessment within this framework. Generalized score tests and graphical tools for assessing goodness of fit are discussed in the context of two examples.

In Chapter 3, we propose a class of latent structure models for analyzing correlated ordinal data. In many situations it is reasonable to assume that the observed ordinal scale is a manifestation of a latent continuous variable categorized by a set of unknown threshold values. Various correlation structure can be modeled via the underlying continuous random variable. We discuss two special cases, the random effects models and the autoregressive model. The Monte Carlo EM algorithm, with importance sampling, is applied to get maximum likelihood estimates.

In Chapter 4, we extend the generalized additive models to include a class of multivariate regression models. We call the resulting models "multivariate generalized additive models" (MGAMs). In addition to modeling the marginal means of multivariate observations through the regression parameters, we are also able to model the correlation

structure of the multivariate observations. The class of models includes the multiple logistic regression models for nominal responses and the ordinal regression discussed in Section 1. We use regression splines to fit non-parametric multivariate regression models.

In Chapter 5, we discuss how to handle censored data. Censoring is defined in general for both categorical and continuous responses. We develop an EM algorithm for censored categorical data. A weighted least squares algorithm is proposed for a class of cumulative link models. An EM algorithm is developed for right censored, left censored and interval censored continuous data.

Chapter 2

Unified ordinal regression analysis via generalized estimating equations and generalized score tests

2.1 Introduction

Ordinal data are common in social science research and increasingly common in other areas such as the biological sciences. There are different strategies for modeling ordinal response data. Commonly used models include the proportional odds model (McCullagh 1980), the adjacent category odds model (see, for example, Clogg and Shihadeh 1994), and the continuation-ratios model (see Feinberg 1980 and Agresti 1990). Simpson et al (1996) discussed the maximum likelihood estimation and marginal analysis subject to interval censoring for these three types of models. Heagerty and Zeger (1996) proposed a GEE approach to the analysis of clustered ordinal data using the proportional odds model. Maximum likelihood estimations (MLE) are commonly used to get parameter estimates, but direct maximization of the likelihood is somewhat complicated when different model forms are used. One of our goals is to develop a unified approach for a broad

class of ordinal regression models. We anticipate that the approach taken here will also facilitate the analysis of correlated ordinal data.

Begg and Gray (1984) studied the technique of individualized logistic regression for calculating polychotomous logistic regression parameters. They fit separate logistic regressions to binary indicators for the different response categories. In a simulation study they found that the resulting estimators have reasonably high efficiency in comparison with maximum likelihood. Ordinal models usually impose constraints on the parameters. For example, the proportional odds model assumes the same slopes across all response levels. For ordinal data, all the models mentioned above can be viewed as a set of logistic regressions at different levels. Each ordinal response contributes to these logistic regressions according to its observed value and the model form, which can be represented by two vectors of binary indicator variables, respectively. Clayton (1992) discussed fitting these logistic regressions simultaneously for the proportional odds model.

We propose a general class of models for ordinal response data, of which the different logit models mentioned above are all special cases. The general formulation is based on modeling a set of unconditional or conditional probabilities of the response at different levels. For obtaining the parameter estimates, we consider simultaneous logistic regression (SLR), an extension of individualized logistic regression that allows consistent estimation of the parameters and provides valid large sample inferences. This method can be easily applied to different models using standard logistic regression software. Another advantage is that one can enforce linear constraints on the parameters conveniently by forming the corresponding design matrix. For example, one can use this method to fit proportional odds model, partial proportional odds model (Peterson and Harrell 1990), and unrestricted models in a unified fashion. It is clear that the collection of indicator variables generated from a given ordinal response are correlated with each other. In order to perform valid inferences these correlated binary response variables can be treated as if they are clustered in a generalized estimating equation analysis.

We also consider efficient logistic regression (ELR), a further refinement, in which we incorporate the correlations between the indicator variables that code the ordinal

response. This is an adaptation of the efficient GEE estimation of Liang and Zeger (1986). We establish equivalence results of the ELR with MLE in special cases, and conjecture that this equivalence holds for a broad class of uncensored ordinal regression models.

Another advantage of the SLR and ELR approaches is that they lend themselves to the development of generalized score tests for composite hypotheses about the model. Like the efficient score tests proposed by Rao (1947), generalized score tests require only the null parameter estimates, and they are invariant to full-rank differentiable transformations of the parameters. Boos (1992) developed generalized score tests in the context of GEE.

The SLR and ELR approaches are developed in Sections 2.2 and 2.3. Section 2.2 describes the coding of the ordinal response into a collection of indicator variables. Section 2.3 develops marginal models and unbiased estimating equations for the binary coded ordinal regression models. Section 2.4 evaluates the asymptotic relative efficiencies of the parameter estimates using both approaches. In Section 2.5 we introduce generalized score tests for SLR and ELR. These score tests can be applied to test proportionality in proportional odds models and other hypotheses about the model. Finally, we present two examples to illustrate our methodology. Different ordinal regression models are considered to analyze the two datasets.

2.2 Binary Coding of Ordinal Responses

Ordinal regression models are designed to model the probability distribution of the ordinal score, Y_i, as a function of covariate information represented by a vector, x_i. We assume without loss of generality that Y_i takes values in $\{0, 1, 2, \ldots, S\}$. It is sufficient to model the probabilities for $\{1, 2, \ldots, S\}$, because the probability of 0 is obtained by subtraction. We consider the following general class of models:

$$Pr(Y_i \in T_s \mid Y_i \in A_s) = \mathcal{H}(\alpha_s + x_i^T \beta_s), \quad s = 1, 2, \ldots, S;$$

where A_s is the *active set*, T_s is the *target set*, $\mathcal{H}$ is a cumulative distribution function, and it is assumed that $T_s \subset A_s$. This class of models includes well-known forms such as the proportional odds model, ordinal threshold probit regression, adjacent category logit regression, and ordinal regression based on continuation ratios.

In this general framework we introduce binary codings of the events $(Y_i \in A_s)$ and $(Y_i \in T_s)$. This binary coding facilitates a unified approach to parameter estimation and inference. Assume that $\mathbf{Y} = (Y_1, \ldots, Y_n)^T$ represents a vector of ordinal measurements for n cases, Y_i represents the i^{th} observation of $\mathbf{Y}$, taking on the values $s = 0, 1, \ldots, S \geq 1$. We represent Y_i by two vectors of binary indicator variables $Y_i^* = (Y_{i1}^*, \ldots, Y_{iS}^*)^T$ and $W_i^* = (W_{i1}^*, \ldots, W_{iS}^*)^T$. We use Y_{is}^* to specify the contribution of Y_i in the logistic regression at level s, and W_{is}^* to specify whether we would include Y_{is}^* in the regression at level s. The values of the indicator variables of an ordinal response are determined by its observed value and the model form. Let x_i denote the covariate associated with Y_i. Denote by θ the vector of all unknown parameters. Three examples are used to illustrate this coding method.

The proportional odds model, developed by McCullagh (1980), assumes parallel effects for different levels. In the proportional odds models for the marginal means, it is assumed that

$$\text{logit}\{Pr(Y_i \geq s)\} = \alpha_s + x_i^T \beta, \quad s = 1, ..., S.$$

We can naturally represent the ordinal measure Y_i through a vector of cumulative indicator variables

$$Y_{is}^* = \begin{cases} 1 & \text{if } Y_i \geq s \\ 0 & \text{if } Y_i < s \end{cases} \qquad \text{and} \quad W_{is}^* = 1, \quad s = 1, \ldots, S,$$

An alternative to the proportion odds model, which can model nonparallel effects, is the continuation ratio model given by

$$\text{logit}\{Pr(Y_i = s \mid Y_i \leq s)\} = \alpha_s + x_i^T \beta_s.$$

If we were to perform an individual logistic regression at level s, we would consider all the responses whose scores are less than or equal to s. This is the active set at level s. So for Y_i at level s, there are three possible outcomes: being 1 or 0 in the s^{th} active set, or excluded from the s^{th} "active" set. So we can defined the indicator variables of Y_i by

$$Y_{is}^{\star} = \begin{cases} 1 & \text{if } Y_i = s \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad W_{is}^{\star} = \begin{cases} 1 & \text{if } Y_i \leq s \\ 0 & \text{otherwise} \end{cases} \quad s = 1, \ldots, S.$$

Another possibility is to model the adjacent category odds in the log-linear model (see Clogg and Shihadeh (1994)).

$$\log\left\{\frac{Pr(Y_i = s)}{Pr(Y_i = s-1)}\right\} = \alpha_s + x_i^T \beta_s.$$

This model is parametrically equivalent, that is $\alpha_s = \tilde{\alpha}_s$ and $\beta_s = \tilde{\beta}_s$, to the conditional model

$$\text{logit}\{Pr(Y_i = s \mid s-1 \leq Y_i \leq s)\} = \tilde{\alpha}_s + x_i^T \tilde{\beta}_s.$$

since $\text{logit}\{Pr(Y_i = s \mid s-1 \leq Y_i \leq s)\} = \log\{Pr(Y_i = s)/Pr(Y_i = s-1)\}$. Similar to the continuation ratio model, we can form the "active" set at each level, and define the indicator variables as

$$Y_{is}^{\star} = \begin{cases} 1 & \text{if } Y_i = s \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad W_{is}^{\star} = \begin{cases} 1 & \text{if } Y_i = s \text{ or } Y_i = s-1 \\ 0 & \text{otherwise} \end{cases} \quad s = 1, \ldots, S.$$

We illustrate the binary coding for all three forms in the case where the response takes on the four possible values $s = 0, 1, 2, 3$. Table 2.1 lists the values of the indicator variables for all the possible responses for the three types of logit models.

The three types of models described above can be treated in a unified fashion by specifying the conditional link function model: $E(Y_{is}^{\star} \mid W_{is}^{\star} = 1) = \mathcal{H}(\alpha_s + x_i^T \beta_s)$. We make the convention that $W_{is}^{\star} = 0$ implies $Y_{is}^{\star} = 0$, i.e., the binary response is defined to be zero outside of the active set. We then have the full specification of the conditional

model,

$$P_{is}^{\star} := E(Y_{is}^{\star} \mid W_{is}^{\star}) = W_{is}^{\star}\mathcal{H}(\alpha_s + x_i^T\beta_s). \tag{2.1}$$

It is convenient to express (2.1) as

$$P_i^{\star} = \mathbf{W}_i\mathcal{H}(X_i\theta), \tag{2.2}$$

where $P_i^{\star} = (P_{i1}^{\star}, \ldots, P_{iS}^{\star})^T$, $\mathbf{W}_i = \text{diag}(W_{i1}^{\star}, \ldots, W_{iS}^{\star})$,

$$\mathbf{X}_i^T = \begin{pmatrix} e_1 & e_2 & \cdots & e_S \\ e_1 \otimes x_i & e_2 \otimes x_i & \cdots & e_S \otimes x_i \end{pmatrix} \quad \text{with} \quad e_j = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}_{S\times 1} \leftarrow j^{th} \text{ position},$$

and $\theta = (\alpha_1, \ldots, \alpha_S, \beta_{11}, \ldots, \beta_{1p}, \ldots, \beta_{S1}, \ldots, \beta_{Sp},)^T$. A common modeling assumption is that the slope parameters for the explanatory variables are constant across severity levels. Under this parallel slopes assumption, e.g., in the proportional odds model, the pseudo-design matrix has the simplified form,

$$\mathbf{X}_i^T = \begin{pmatrix} e_1 & e_2 & \cdots & e_S \\ x_i & x_i & \cdots & x_i \end{pmatrix}.$$

Observe that the S binary responses generated by observation i appear as if they were separate responses in a binary regression with augmented design matrix. In performing inferences we adjust for their correlation using GEE theory as described by Diggle et al. (1994, Chapter 8). The marginal model corresponding to (2.1) is given by

$$E(Y_{is}^{\star}) = E(W_{is}^{\star})\mathcal{H}(\alpha_s + x_i^T\beta_s). \tag{2.3}$$

Table 2.1: Indicator variables for ordinal responses

	Proportional odds model			Continuation ratio model			Adjacent category odds model		
Y_i	$Y_{i1}^\star$	$Y_{i2}^\star$	$Y_{i3}^\star$	$Y_{i1}^\star$	$Y_{i2}^\star$	$Y_{i3}^\star$	$Y_{i1}^\star$	$Y_{i2}^\star$	$Y_{i3}^\star$
0	0	0	0	0	0	0	0	0	0
1	1	0	0	1	0	0	1	0	0
2	1	1	0	0	1	0	0	1	0
3	1	1	1	0	0	1	0	0	1
Y_i	$W_{i1}^\star$	$W_{i2}^\star$	$W_{i3}^\star$	$W_{i1}^\star$	$W_{i2}^\star$	$W_{i3}^\star$	$W_{i1}^\star$	$W_{i2}^\star$	$W_{i3}^\star$
0	1	1	1	1	1	1	1	0	0
1	1	1	1	1	1	1	1	1	0
2	1	1	1	0	1	1	0	1	1
3	1	1	1	0	0	1	0	0	1

In a fully parametric approach, the term $E(W_{is}^\star)$ may have a complicated dependence on on the parameters. In the development that follows we avoid this complication through the use of conditionally unbiased estimating equations.

2.3 SLR and ELR Estimation

To derive conditionally unbiased estimating equations we start with (2.1) and observe that

$$[Y_{is}^\star | W_{is}^\star] \sim \text{Bernoulli}(P_{is}^\star), \tag{2.4}$$

where $P_{is}^\star = W_{is}^\star \mathcal{H}(\alpha_s + x_i^T \beta)$. If $W_{is}^\star = 0$, then, by definition $Y_{is}^\star = 0$ as well, so that (2.4) holds trivially. Defining $0 * \log(0) = 0$ through the usual limiting argument, we have the marginal pseudo-log-likelihood:

$$\sum_{i=1}^{n} \sum_{s=1}^{S} Y_{is}^\star \log(P_{is}^\star) + (1 - Y_{is}^\star) \log(1 - P_{is}^\star). \tag{2.5}$$

Now $W_{is}^{\star}Y_{is}^{\star} = Y_{is}^{\star}$. Moreover, if $W_{is}^{\star} = 0$, then $P_{is}^{\star} = 0$ and $\log(1 - P_{is}^{\star}) = 0$. Therefore, the criterion in (2.5) is equal to the weighted logistic regression criterion,

$$\sum_{i=1}^{n}\sum_{s=1}^{S} W_{is}^{\star}\left\{Y_{is}^{\star}\log(\mathcal{H}_{is}) + (1 - Y_{is}^{\star})\log(1 - \mathcal{H}_{is})\right\}, \tag{2.6}$$

where $\mathcal{H}_{is} = \mathcal{H}(\alpha_s + x_i^T\beta)$. The criterion in (2.6) corresponds to the estimating equation

$$\sum_{i=1}^{n}\mathbf{X}_i^T\mathbf{W}_i\{Y_i - \mathcal{H}(\mathbf{X}_i\theta)\} = 0, \tag{2.7}$$

where $Y_i^{\star} = (Y_{i1}^{\star}, \ldots, Y_{iS}^{\star})^T$, and where $\mathbf{X}_i$, $\mathbf{W}_i$ and θ have the same form as in equation (2.2). Equation (2.1) implies that (2.7) is a conditionally unbiased, and hence marginally unbiased, estimating equation. For further discussion of conditional versus marginal unbiased estimating equations see Kunsch, Stefanski and Carroll (1989).

Because the weights are 0-1 valued, the maximizer of (2.6) can be computed by ordinary logistic regression of the binary responses in the active set. We refer to the resulting estimates as the simultaneous logistic regression (SLR) estimates. At each level s, model (2.7) can be viewed as a logistic regression on the "active" set at level s. Simultaneously fitting logistic regression using SLR allows for pooling information. Simpson et al. (1996) introduced the proportional odds version of SLR, referring to the method as "pseudo-strata." They used the pseudo-strata estimates as starting values for maximum likelihood, but did not discuss the possibility of basing inferences on them. The conditional odds version of SLR is in fact equivalent to maximum likelihood estimation. A number of authors have used this fact about continuation ratios to simplify the computations for that special case; see, e.g., Agresti (1990, page 319).

The SLR estimation produces consistent, asymptotically normal estimates, as can be shown by adapting results of Liang and Zeger (1986). Except for the special case of the conditional odds model, the inverse of the pseudo-information matrix will give inconsistent estimates of the asymptotic variance of the parameter estimates. The theory of estimating equations provides a consistent estimator for the asymptotic covariance of

$\hat{\alpha}$ and $\hat{\beta}$, namely, the sandwich estimator,

$$\hat{\text{var}}_{slr}(\hat{\alpha},\hat{\beta}) = H_1(\hat{\alpha},\hat{\beta})^{-1} H_2(\hat{\alpha},\hat{\beta}) H_1(\hat{\alpha},\hat{\beta})^{-1} \tag{2.8}$$

where

$$H_1(\hat{\alpha},\hat{\beta}) = \sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{W}_i^* \hat{\Delta}_i \mathbf{W}_i^* \mathbf{X}_i,$$

$$H_2(\hat{\alpha},\hat{\beta}) = \sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{W}_i^* (Y_i^* - \hat{\mathcal{H}}_i)(Y_i^* - \hat{\mathcal{H}}_i)^T \mathbf{W}_i^* \mathbf{X}_i,$$

$\mathcal{H}_i = (\mathcal{H}(\alpha_1 + \beta_1^T x_i), \ldots, \mathcal{H}(\alpha_S + \beta_S^T x_i))^T$, and $\Delta_i = \text{diag}\{\mathcal{H}'(\alpha_1 + x_i^T\beta_1), \ldots, \mathcal{H}'(\alpha_S + x_i^T\beta_S)\}$. For background and further references on estimating equations see Diggle, Liang and Zeger (1994).

Computationally, one can simply exclude the inactive indicator variables and use a standard logistic regression procedure on the active indicator variables. The SLR method is easy to implement and, as will be seen, it can be highly efficient. Inferences based on SLR provide a fast method for model selection. Furthermore, SLR can provide good starting values for the more efficient ELR approach, described below, and for maximum likelihood estimation.

More efficient estimates can be obtained by including the covariance in the ELR approach. Liang and Zeger (1986), Zeger and Liang (1986) have developed moment-based GEE methods for regression models for longitudinal categorical responses, where the repeated measurements on the same individual are correlated. Similarly, we can treat the indicator variables obtained from one observation as from one cluster, and consider the correlations among them. In the present case we can work out the exact covariance functions for the coded response variables. Under certain conditions, the ELR approach produces consistent estimators of the regression parameters, under only the correct specification of the form of the marginal mean function. The ELR for α and β solve the estimating equation

$$\sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{W}_i^* \Delta_i V_i^{-1} \mathbf{W}_i^* (Y_i^* - \mathcal{H}_i) = 0, \tag{2.9}$$

where $V_i = E[\mathbf{W}_i^*(Y_i^* - \mathcal{H}_i)(Y_i^* - \mathcal{H}_i)^T\mathbf{W}_i^*]$. Because the assumed covariance is correct under the model, the variance of $\hat{\alpha}$ and $\hat{\beta}$ is consistently estimated by

$$\hat{\text{var}}_{elr}(\hat{\alpha}, \hat{\beta}) = (\sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{W}_i^* \hat{\Delta} \hat{V}_i^{-1} \hat{\Delta}_i \mathbf{W}_i^* \mathbf{X}_i)^{-1}. \tag{2.10}$$

For the three model types mentioned previously V_i has elements v_{ist} $(s, t \in \{1, \ldots, S\})$, where

$$\text{Proportional odds}: \quad v_{ist} = \mathcal{H}_{is}(1 - \mathcal{H}_{it}), \quad \text{if } s \geq t;$$

$$\text{Continuation ratio}: \quad v_{ist} = \begin{cases} \mathcal{H}_{is} \prod_{j=s}^{S}(1 - \mathcal{H}_{ij}), & \text{if } s = t; \\ 0, & \text{if } s \neq t; \end{cases}$$

$$\text{Adjacent odds}: \quad v_{ist} = \begin{cases} \pi_{is}\{\mathcal{H}_{is} + (1 - \mathcal{H}_{is})^2\}, & \text{if } s = t; \\ -\pi_{is}(1 - \mathcal{H}_{is})\mathcal{H}_{i,s+1}, & \text{if } t = s + 1; \\ 0, & \text{if } t > s + 1; \end{cases}$$

where $\pi_{is} = \prod_{j=1}^{s} \mathcal{H}_{ij}\{1 + \sum_{j=1}^{S} \prod_{k=1}^{j} \mathcal{H}_{ik}\}^{-1}$ and $v_{its} = v_{ist}$.

Consistency of the ELR method is an open question in general because of the double occurrence of the stochastic weight matrix W_i^* in (2.9). This could possibly bias the estimating equation if the matrix $\Delta_i V_i^{-1}$ has nonzero off-diagonal terms. If the latter matrix is diagonal, then unbiasedness of the estimating equation follows from relation (2.1) because $(W_{is}^*)^2 = W_{is}^*$. In particular, the ELR estimating equation is unbiased for the proportional odds model and for the continuation ratio model.

2.4 Asymptotic Relative Efficiencies of SLR and ELR

Here we address the issue of asymptotic efficiencies of the SLR and ELR estimates. The efficiencies are generally high for the SLR method, and both methods yield fully efficient estimates for special cases.

Table 2.2: ARE of simultaneous logistic method for proportional odds model

Design	β_1	β_2	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	α_9
(i)	0.97	0.99	0.98	0.98							
(ii)	0.93	0.96	0.96	0.95	0.95	0.96	0.96				
(iii)	0.92	0.95	0.95	0.94	0.94	0.94	0.94	0.94	0.95	0.95	0.95

2.4.1 Proportional Odds Models

In general, it is difficult to compare the efficiency of SLR estimates to MLE analytically, so we consider 2×2 cross-sectional design configurations with different number of response levels. Here x_{i1} and x_{i2} are the dichotomous covariates indicating group membership for the i^{th} individual. The parameters are $\beta_1 = 1$, $\beta_2 = 2$; and three designs with 3, 6, and 10 response levels are selected:

(i) $\alpha_1 = 0; \alpha_2 = -2;$

(ii) $\alpha_1 = 0; \alpha_2 = -0.5; \alpha_3 = -1; \alpha_4 = -1.5; \alpha_5 = -2;$

(iii) $\alpha_1 = 0; \alpha_2 = -0.25; \alpha_3 = -.5; \alpha_4 = -.75; \alpha_5 = -1; \alpha_6 = -1.25; \alpha_7 = -1.5;$ $\alpha_8 = -1.75; \alpha_9 = -2;$

The efficiencies were calculated in Mathematica. We assume balance in all designs, that is each of the four combinations occurs with probability 0.25. Table 2.2 lists the asymptotic relative efficiencies for the three designs. It shows that the asymptotic efficiencies are high, the minimum efficiency is 92%.

The log-likelihood of the indicator variables, $Y_i^\star = \{Y_{i1}^\star, \ldots, Y_{iS}^\star\}$, can be written as

$$l(Y_i^\star) = (1 - Y_{i1}^\star)\log(1 - P_{i1}^\star) + (Y_{i1}^\star - Y_{i2}^\star)\log(P_{i1}^\star - P_{i2}^\star)$$
$$+ \cdots + (Y_{iS-1}^\star - Y_{iS}^\star)\log(P_{iS-1}^\star - P_{iS}^\star) + Y_{iS}^\star \log(P_{iS}^\star).$$

which follows an exponential family distribution. By the theory of generalized linear models, the ELR estimation is equivalent to MLE for the proportional odds models.

Table 2.3: ARE of simultaneous logistic method for adjacent category odds model

Design	β_1	β_2	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	α_9
(i)	0.99	0.98	0.99	0.99							
(ii)	0.96	0.88	0.98	0.95	0.95	0.98	0.96				
(iii)	0.92	0.79	0.99	0.98	0.96	0.96	0.96	0.94	0.94	0.97	0.92

2.4.2 Continuation Ratio Models

For the continuation ratio model. We first consider the multinomial representation. Let $n_s, s = 0, 1, \ldots, S$ denote the response count in each cell and let $n = \sum_{s=0}^{S} n_s$. Define by $q_s = Pr(Y = s \mid Y \leq s)$. The multinomial mass function has factorization

$$b(n, n_S, q_S)\ b(n - n_S, n_{S-1}, q_{S-1}) \cdots b(n - n_S - \cdots - n_2, n_1, q_1).$$

where $b(n, y, q)$ denote the binomial probability of y "successes" in n trials, when the success probability is q on each trial (see Agresti 1990, page 319). So the log-likelihood function is

$$l = \sum_{i=1}^{n} \sum_{s=1}^{S} W_{is}^{\star} \{Y_{is}^{\star} \log(P_{is}^{\star}) + (1 - Y_{is}^{\star}) \log(1 - P_{is}^{\star})\}.$$

which is the same as (2.6). Therefore the SLR is equivalent to MLE (and ELR) for the continuation ratio models, and it is asymptotically efficient.

2.4.3 Adjacent Category Odds Models

Using the same design configurations above, we compare the asymptotic relative efficiency of SLR to MLE. The results are listed in Table 2.3 In our examples, we have attempted to assess the influence of the number of categories. In general, the asymptotic relative efficiencies are high throughout, although occasionally SLR is inefficient for individual parameters. Because SLR treats each binary response separately in the estimation phase rather than borrowing strength from the correlations between binary responses, certain configurations of response probabilities my lower its efficiency. In particular, the low efficiency for estimating β_2 in Design (iii) appears to be due to extreme probabilities for

inclusion in group 2. Table 2.3 also suggests a decrease in efficiency as the number of response levels increase.

2.5 Goodness of Fit and Score Tests

Rao (1948) introduced score statistics having the form

$$S(\hat{\theta})^T I^{-1} S(\hat{\theta}) \tag{2.11}$$

where $S(\theta)$ is the vector of partial derivatives of the log likelihood function, $\hat{\theta}$ is the vector of restricted maximum likelihood estimates under H_0, and I is the Fisher information of the sample evaluated at $\hat{\theta}$. These test statistics are attractive because they only require computation of the null estimates $\hat{\theta}$ and are asymptotically equivalent to Wald and likelihood ratio statistics under both null and Pitman alternative hypotheses (Serfling 1980, pape156). The generalizations of Rao's score tests has been discussed by Boos (1992) in the general estimating equations situation, and by White (1982) in the context of model misspecification. These generalizations are able to account for lack of knowledge about the correlation structure by using semiparametric variance estimates. The same ideas may be adapted to score testing, as described by Boos (1992). Advantages of the Rao-type score test in comparison with the Wald-type parameter estimates test, include invariance to nonlinear transformations of parameters and estimation for the reduced model only. Generalized score tests can be developed for the SLR approach and the ELR approach to assess goodness of fit of the different ordinal regression models.

In the test for parallelism we consider model (1) as the full model, and we test the hypothesis $H_0 : g(\theta) = 0$, where $g : R^p \rightarrow R^r$ is a continuous vector function of θ such that its Jacobian at θ, $G(\theta) = \partial g(\theta)/\partial\theta$, is finite with full row rank r, against the alternative $H_1 : g(\theta) \neq 0$.

Let $\tilde{\theta}$ solve the constrained maximization problem

$$\max_{\theta\in\Theta} Q(\theta) \quad \text{subject to} \quad g(\theta) = 0,$$

and let

$$S(\theta) = \frac{\partial Q(\theta)}{\partial \theta}.$$

With the SLR approach, $Q(\theta)$ is given by (2.6), and $S(\theta)$ is given by (2.7) with unrestricted parameters. We need to solve (2.7) subject to $g(\theta) = 0$ to get $\tilde{\theta}$, the restricted parameter estimates. The estimates $\tilde{\theta}$ which maximizes $Q(\theta)$ subject to H_0 satisfies

$$S(\tilde{\theta}) - G(\tilde{\theta})^T \lambda = 0, \quad g(\tilde{\theta}) = 0,$$

where λ is an $r \times 1$ vector of Lagrange multipliers. This form is general but may not be easy to implement using existing software. For some special cases this can be accomplished by changing the design matrix to simplify the computation of restricted parameter estimates. For instance, for testing parallelism discussed before, the restricted parameterization is that $\beta_1 = \beta_2 = \ldots = \beta_S$. The parameter estimation for the restricted model is simplified by using the design matrix assuming common slope parameter. The score test statistic is as follows

$$S_{slr} = S(\tilde{\theta})^T H_1(\tilde{\theta})^{-1} G(\tilde{\theta})^T [G(\tilde{\theta}) H_1(\tilde{\theta})^{-1} H_2(\tilde{\theta}) H_1(\tilde{\theta})^{-1} G(\tilde{\theta})^T]^{-1} G(\tilde{\theta}) H_1(\tilde{\theta})^{-1} S(\tilde{\theta}). \tag{2.12}$$

Under H_0 and suitable regularity conditions, $S_{slr} \rightarrow \chi^2_r$. As for the ELR approach for the proportional odds model, the score test statistic has the form

$$S_{elr} = S(\tilde{\theta})^T I^{-1}(\tilde{\theta})^{-1} S(\tilde{\theta}). \tag{2.13}$$

Under H_0 and suitable regularity conditions, $S_{elr} \rightarrow \chi^2_r$. The SLR approach and its generalized score test provide a very flexible and simple way of modeling ordinal response data and checking for lack of fit.

As an application, we can use generalized score tests to assess proportionality in proportional odds models and parallelism of slope parameters in other models. Consider the general model without making the parallel line assumption $P_i^* = \mathbf{W}_i \mathcal{H}(X_i\theta)$ where $\theta = (\alpha_1, \ldots, \alpha_s, \beta_{11}, \ldots, \beta_{1S}, \ldots\ldots, \beta_{p1}, \ldots, \beta_{pS})^T$. Under the null hypothesis $H_0 : \beta_{k1} = \beta_{k2} = \cdots = \beta_{kS}$, $k = 1, \ldots, p$, there is a single common slope parameter for each of the p explanatory variables. Let $\beta_1, \ldots, \beta_p$ be the common slope parameters under H_0. Let $\hat{\alpha}_1, \ldots, \hat{\alpha}_S$, and $\hat{\beta}_1, \ldots, \hat{\beta}_p$ be the estimated parameters under H_0. So the score statistics S_{slr} and S_{elr} each have an asymptotic chi-square distribution with $p(S-1)$ degree of freedom. The score tests can also be used as model selection statistics for testing individual variables not in the model. We have implemented the procedures of estimation and testing in S-plus (MathSoft, Inc.). Two examples are discussed in the following section to illustrate the methodology.

2.6 Examples

Example 1. Analysis of Mental Health Data

To illustrate the methodology, we first consider data from Srole et al. (1962) on the relationship between an individual's mental health status and the socioeconomic status of his or her parents. The data are displayed in Table 2.4. The probability bar plot in Figure 2.1 shows a decreasing trend of mental health as the socioeconomic status goes down. It is not immediately clear which model is likely to provide a better description of these data. We therefore examine the empirical logit transformations using different logits. We plot the empirical logits versus the response levels. This technique is simple and often useful for choosing the parsimonious model form. Figure 2.2 shows the empirical logit transformation plots for the three types of logits. Each curve connects the transformed empirical logits of contiguous levels for each socioeconomic group. None of the logit transformation plots deviate greatly from parallelism. So we fit all the three models with common slopes parameters cross levels and compute the score statistics for testing the parallelism assumption.

The 10 degree of freedom score tests of parallel slopes for the proportional odds model using both the ELR and SLR are $S_{elr} = 7.78$, $p_{elr} = 0.65$, and $S_{slr} = 7.89$, $p_{slr} = 0.64$.

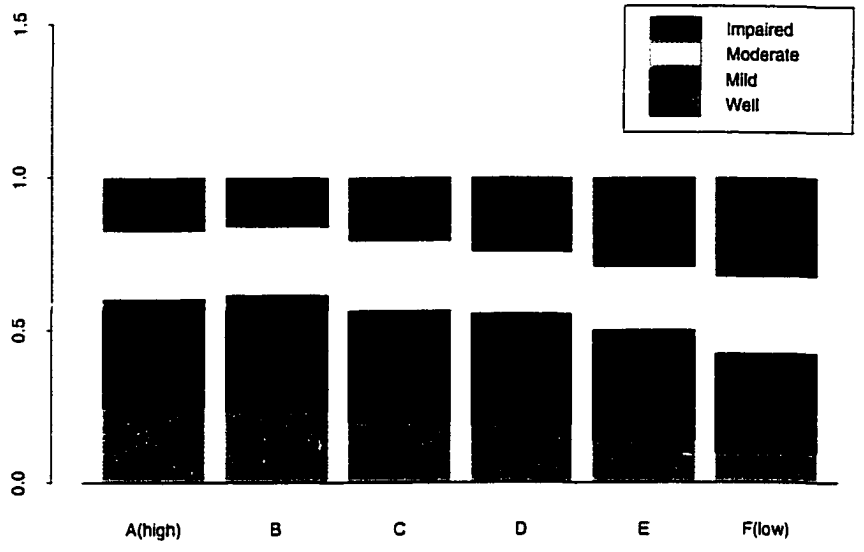


Figure 2.1: Bar plot of mental health data

The score test of parallel slopes for the continuation ratio model is $S = 6.89$, $p = 0.74$. And the score test of parallel slopes for the adjacent category odds model is $S = 5.41$, $p = 0.86$. Table 2.4 gives the fitted counts using the three models. We also use the bar plot of observed cell counts versus fitted cell counts as a diagnostic tool for checking individual fit. Parameter estimates and their standard errors are displayed in Table 2.5.

For further simplification, we assign scores to the six socioeconomic status levels, which $A = 5$, $B = 4$, $C = 3$, $D = 2$, $E = 1$, and $F = 0$, and fit the three models with common slopes at different levels. The 14 degree of freedom score tests for the proportional odds model, the continuation ratio model, and the adjacent category odds model are 7.91, 6.88, and 5.41 respectively. The fitted values are given in Table 2.6. All the three simple models are appropriate for mental health data.

Table 2.4: Data on mental health status

Parent's Socioeconomic Status	Mental Health Status			
	Well	Mild Symptom Formation	Moderate Symptom Formation	Impaired
$A(high)$	64 (60.5)[a] (60.5)[b] (62.7)[c] (59.2)[d]	94 (102.5) (102.5) (99.0) (103.2)	58 (51.6) (51.5) (50.5) (51.6)	46 (47.4) (47.6) (49.9) (45.2)
B	57 (57.6) (57.3) (58.5) (57.6)	94 (96.2) (96.0) (92.6) (96.4)	54 (47.7) (47.8) (47.2) (48.4)	40 (43.5) (43.9) (46.7) (42.6)
C	57 (56.2) (56.3) (56.8) (56.9)	105 (107.9) (107.8) (106.7) (107.8)	65 (61.4) (61.4) (60.7) (61.3)	60 (61.4) (61.6) (62.8) (61.0)
D	72 (69.4) (69.9) (70.8) (69.9)	141 (140.9) (140.9) (141.4) (140.3)	77 (85.0) (84.7) (83.8) (84.5)	94 (88.7) (88.5) (88.0) (89.3)
E	36 (38.3) (38.6) (36.8) (37.4)	97 (89.1) (89.3) (92.3) (88.1)	54 (62.9) (62.6) (63.9) (62.3)	78 (74.7) (74.5) (72.0) (77.2)
$F(low)$	21 (25.0) (25.1) (21.7) (24.4)	71 (65.4) (65.5) (68.8) (65.5)	54 (53.4) (53.2) (56.8) (52.7)	71 (73.3) (73.2) (69.7) (74.5)

[a]Fitted values with the proportional odds model using SLR method

[b]Fitted values with the proportional odds model using ELR method

[c]Fitted values with the continuation ratio model using SLR method

[d]Fitted values with the adjacent category odds model using SLR method

Table 2.5: Parameter Estimates and Standard Errors for Mental Health Data

		α_1	α_2	α_3	β_1	β_2	β_3	β_4	β_5
Proportional odds model	ELR	2.033	0.333	-0.676	-0.830	-0.847	-0.622	-0.530	-0.263
		(0.133)	(0.124)	(0.125)	(0.166)	(0.168)	(0.162)	(0.153)	(0.165)
	SLR	2.040	0.338	-0.673	-0.836	-0.861	-0.628	-0.529	-0.262
		(0.132)	(0.124)	(0.125)	(0.168)	(0.169)	(0.163)	(0.154)	(0.166)
Continuation ratio model	SLR	1.156	-0.465	-0.748	-0.699	-0.698	-0.525	-0.465	-0.237
		(0.122)	(0.113)	(0.108)	(0.137)	(0.138)	(0.134)	(0.136)	(0.136)
Adjacent category odds model	SLR	0.988	-0.216	0.345	-0.477	-0.473	-0.349	-0.291	-0.130
		(0.130)	(0.123)	(0.125)	(0.152)	(0.154)	(0.148)	(0.140)	(0.151)

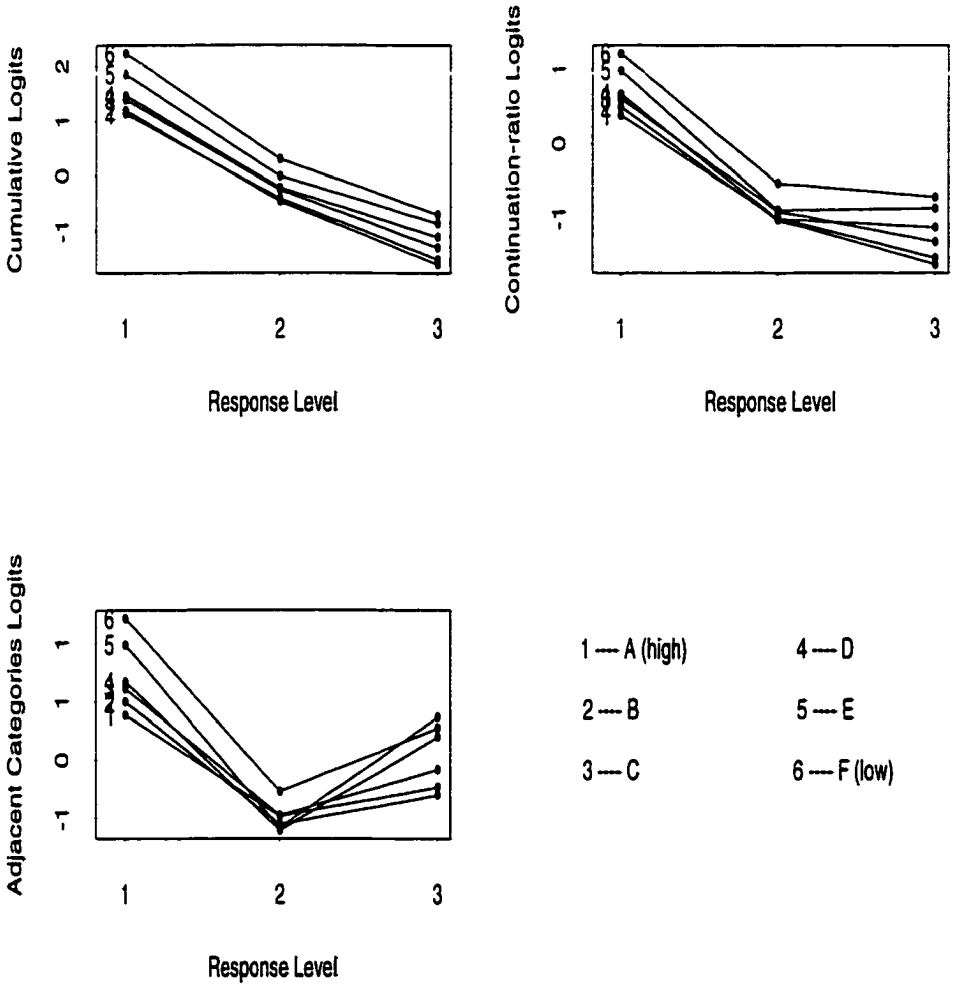
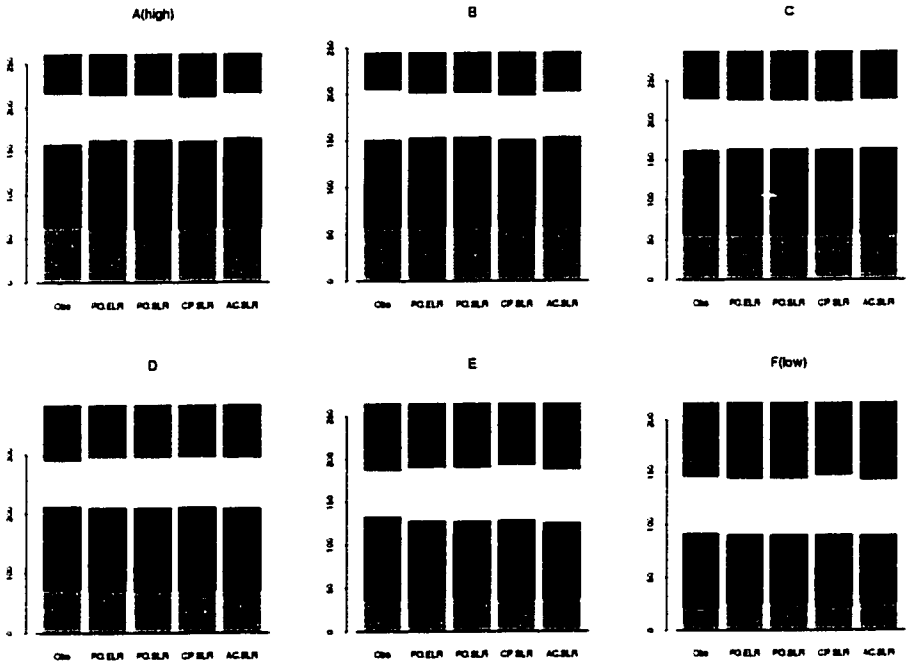


Figure 2.2: Empirical logit transformation plot of mental health data



Obs: Observed counts
PO.ELR: Proportional odds model with ELR PO.SLR: Proportional odds model with SLR
CP.SLR: Continuation ratio model with SLR AD.SLR: Adjacent category odds model

Figure 2.3: Observed values vs. Fitted values

Table 2.6: Fitted values with continuous scores on socioeconomic status

Parent's Socioeconomic Status	Mental Health Status			
	Well	Mild Symptom Formation	Moderate Symptom Formation	Impaired
5	64 (66.1)[a] (65.9)[b] (68.9)[c] (66.7)[d]	94 (104.0) (103.7) (99.0) (104.8)	58 (48.8) (48.9) (47.8) (49.5)	46 (43.1) (43.5) (46.3) (41.0)
4	57 (54.3) (54.3) (55.9) (55.0)	94 (95.0) (94.8) (92.2) (95.2)	54 (49.3) (49.3) (48.4) (49.5)	40 (46.4) (46.6) (48.5) (45.3)
3	57 (55.6) (55.7) (56.3) (56.2)	105 (107.5) (107.3) (106.5) (107.3)	65 (61.6) (61.6) (61.0) (61.5)	60 (62.2) (62.4) (63.3) (61.9)
2	72 (64.8) (65.0) (63.9) (65.0)	141 (137.4) (137.2) (138.8) (136.7)	77 (87.0) (86.9) (87.2) (86.4)	94 (94.8) (94.9) (94.1) (95.8)
1	36 (38.7) (39.0) (37.1) (38.4)	97 (89.5) (89.5) (92.3) (89.0)	54 (62.6) (62.4) (63.8) (61.9)	78 (74.1) (74.1) (71.9) (75.7)
0	21 (27.4) (27.6) (25.2) (26.6)	71 (68.5) (68.6) (71.9) (68.0)	54 (52.7) (52.6) (55.0) (52.2)	71 (68.4) (68.2) (64.9) (70.3)

[a]Fitted values with the proportional odds model using SLR method

[b]Fitted values with the proportional odds model using ELR method

[c]Fitted values with the continuation ratio model using SLR method

[d]Fitted values with the adjacent category odds model using SLR method

Table 2.7: Experimental results for chicken embryos exposed to arboviruses

Virus	Inoculum titre (PFU/egg)	Number not deformed	Number deformed	Number of deaths	Total number of eggs
Control	0	17	0	1	18
Tinaroo	3	18	0	1	19
	20	17	0	2	19
	2400	2	9	4	15
	88000	0	10	9	19
Facey's	3	13	1	3	17
Paddock	18	14	1	4	19
	30	9	2	8	19
	90	2	1	17	20

Example 2. Analysis of Chicken Embryo Data

As a second example, Table 2.7 lists a subset of data from Jarrett, Morgan, and Liow (1981). The data were also given in Morgan (1992, page 10, Table 1.7). The objective was to investigate the effects of arboviruses injected into chicken embryos and to quantify the potency of arboviruses. In the experiments, eggs were inoculated with a range of viruses and several inoculation levels and candled daily for 14 days to check viability. The surviving embryos were then examined for gross abnormalities and the results were reported 4 days later; see Jarrett et al. (1981) and McPhee et al. (1984) for more detail. The resulting data for the control group and two arboviruses, the Facey's Paddock virus and the Tinaroo virus, are listed in Table 2.7. There are three levels of possible responses - death, alive but deformed, and alive but not deformed. The need to examine the dependence of the responses on the amount of injected viruses leads to an ordinal regression analysis.

The conventional approach is to model the exposure data together with the control data using a proportional odds model. Morgan (1992) suggested using $\log_{10}$-transformed doses to improve the fitting and eliminate the huge influences of the large doses levels. However, for the Tinaroo and control data, with the $\log_{10}$-scaled dose entering the conventional proportional odds model, the control response rates at both deformed and death severities should be zero. This contradict the fact that there is one observed death in the controls. In order to use the proportional odds model, McPhee et al. (1984) and Morgan (1992) argue that the observed control death rate (1/18) was small and was therefore ignored, i.e., they omitted the control observations. However, as indicated by

Table 2.8: Model fitting results for chicken embryo data

Model used	Data set	Score statistics	Degrees of freedom	p-value
Proportional odds	Tinaroo and control	13.3	7	0.065
	Paddock and control	1.3	7	0.98
	All data	79.9	10	5.3×10^{-13}
Continuation ratio	Tinaroo and control	1.6	7	0.97
	Paddock and control	6.9	7	0.44
	All data	3.8	10	0.95
Adjacent category odds	Tinaroo and control	30.6	7	7.4×10^{-5}
	Paddock and control	9.1	7	0.25
	All data	20.5	10	0.024

Morgan (1992, page 120), the proportional odds model does not fit the data. Xie and Simpson (1999) pointed out that ignoring low incidence rates at high severity levels may lead to the failure of the model. They used a ordinal regression model with nonzero control response probability on the chicken embryo data.

We first model the exposure data and the control data for each virus separately. The models we used are the proportional odds model, the continuation ratio model, and adjacent category odds model. We use the score statistics with the SLR approach described in section 5 for testing goodness of fit. The results are listed in Table 2.8. The proportional odds model and adjacent category odds model do not fit the Tinaroo and control data, whereas the continuation ratio model works well. For the Facey's Paddock and control data, all the three models work.

Since both experiments share the same control data, it is appropriate to analyze the entire data set (two viruses and control). We fit a model with three sets of parameters: a set of spontaneous baseline parameters, and two sets of virus-specific intercept and slope parameters. In Table 2.8, the last three rows summarize the fitted models. For each of the three model forms, the full model can be expressed as follows:

$$E(Y_{is}^{\star} \mid W_{is}^{\star} = 1) = \mathcal{H}(\alpha_{s0} + \alpha_{s1}x_{i1} + \alpha_{s2}x_{i2} + \beta_1 x_{i1}d_i + \beta_2 x_{i2}d_i), \quad s = 1, 2;$$

where

$$x_{i1} = \begin{cases} 1, & \text{Tinaroo;} \\ 0, & \text{Otherwise;} \end{cases} \qquad x_{i2} = \begin{cases} 1, & \text{Facey's Paddock;} \\ 0, & \text{Otherwise;} \end{cases}$$

Table 2.9: Fitted values for chicken embryos data

Virus	Inoculum titre (PFU/egg)	Number not deformed	Number deformed	Number of deaths	Total number of eggs
Control	0	17	0	1	18
		(16.15)	(0.46)	(1.38)	
Tinaroo	3	18	0	1	19
		(17.05)	(1.71)	(0.24)	
	20	17	0	2	19
		(15.48)	(3.05)	(0.47)	
	2400	2	9	4	15
		(5.43)	(7.36)	(2.06)	
	88000	0	10	9	19
		(1.58)	(9.29)	(8.12)	
Facey's	3	13	1	3	17
Paddock		(14.34)	(0.60)	(2.06)	
	18	14	1	4	19
		(10.45)	(1.75)	(6.80)	
	30	9	2	8	19
		(8.09)	(2.11)	(8.81)	
	90	2	1	17	20
		(3.64)	(2.48)	(13.88)	

and d_i is log(dose + 1) for the ith observation. We use the log(dose + 1) transformation to allow non-zero deformed and death rates for the control group. The score test results show that, assuming parallel slopes across severity categories, only the continuation ratio model works for the entire data set. The fitted cell counts are listed in Table 2.9.

2.7 Discussion

In this chapter we have described a unified form for different logit models of ordinal responses. This general formulation provides a very flexible way of modeling ordinal response data. The binary coding method for ordinal responses is not only useful in the context of fitting the unified ordinal models, but also in fitting other ordinal regression models such as the latent structure ordinal models (Chapter 3) and nonparametric ordinal regression models (Chapter 4), and in dealing with censored ordinal data (Chapter 5). The methods we propose avoid the need to develop different model-fitting procedures

for different logit models. The simulation study shows that the efficiencies of the SLR approach are in general high. In an effort to obtain more efficient estimates, the ELR approach to consider the correlations among the indicator variables has been developed. We have also developed a generalized score tests for inferences based on the unified method, they provide fast tests for parallelism and other model selection hypotheses.

The SLR method can be implemented using conventional software. The SAS procedure GENMOD can produce the SLR estimates and the covariance matrix of the parameter estimates as well with the option TYPE=IND in the REPEATED statement. The implementation of the binary coding and the generalized score tests require additional work, though they are straightforward. Other software such as SUDAAN can also provide sandwich type covariance estimates.

Finally, we note that a potential use of our work is to provide a framework for modeling correlated ordinal data. Various correlation patterns can be incorporated by modeling the covariance between the indicator variables from the correlated observations. There are two main methods in the literature on measures of association between correlated ordinal data. Using the correlation coefficient as a measure of association, Miller, Davis and Landis(1993) have used the GEE method for proportional odds model. Williamson, Kim and Lipsitz (1995) and Heagerty and Zeger (1996) used the global odds ratios to measure the association. A promising direction for further research is to build a longitudinal GEE approach on the binary coding approaches of SLR and ELR.

Chapter 3

Latent Structure Models for Correlated Ordinal Data

3.1 Introduction

Statistical methods appropriate for the analysis of longitudinal categorical data are not as well developed as their continuous counterparts. The models for longitudinal ordinal response data have studied by various researchers. There are three commonly used approaches for modeling longitudinal response variables: (1) marginal models address the average responses in sub-populations whose members share the common values of explanatory variables, and utilize various methodological strategies to account for the correlation between repeated measurements; (2) random effects models consider the natural heterogeneity among subjects in a subset of the regression coefficients; (3) transitional models describe the probability distribution of a subject's future events given the subject's prior history. Diggle, Liang, and Zeger (1996) discussed existing methodologies for the analysis of discrete and continuous longitudinal data. In this chapter, we further develop the analysis of longitudinal ordinal outcomes through the latter two types of models, assuming a latent continuous structure.

In many situations it is reasonable to assume that the observed ordinal scale is a manifestation of a latent continuous variable partitioned by a set of unknown threshold

values. We develop a general latent structure framework for the analysis of ordinal longitudinal data, in which the inferences we make about the regression parameters of primary interest recognize the likely correlation structure in the data. This approach is general and can often be given intuitive meaning. Closely related work includes Keenan (1982), who studied longitudinal binary data generated by an underlying continuous-valued time series, as well as Xie, Simpson and Carroll (2000), who discussed random effects model for clustered ordinal data and a Gibbs sampling approach to fit the regression model.

Another approach to analyzing clustered ordinal data is to use the generalized estimating equation (GEE) approach of Liang and Zeger (1986). Several researchers have proposed GEE approach for the analysis of clustered ordinal response data. (Clayton 1992; Gange, Linton, Scott, and Klein 1993; Kim, Willamson, and Lipsitz 1993; Miller, Davis and Landis 1993, Heagerty and Zeger 1996; Simpson et al. 1996). For analyzing categorical time series. Liang and Zeger (1989) proposed a class of logistic regression models for multivariate binary time series. Hidden Markov models have been used to analyze discrete-valued time series (see for example MacDonald and Zucchini (1997)). Kitagawa (1987) studied non-Gaussian state space modeling of nonstationary time series.

A common strategy, the EM algorithm (Dempster, Laird, and Rubin 1977) is often used for maximum likelihood estimation. One major difficulty lies in evaluation of the high dimensional integrals that may appear in the likelihood function. When the dimension is one or two, numerical integration techniques can be reasonably easy (e.g. Crouch and Spiegelman, 1990). For higher dimensional problems, Monte Carlo integration methods can be used. See, for example, the applications of Gibbs sampling in Zeger and Karim (1991), and Xie, Simpson and Carroll (2000). An alternative strategy is to use conditional modes rather than conditional means in the score function. This approach has been used by Stiratelli, Laird, and Ware (1984) for logistic models with Gaussian random effects; Karim (1991), Schall (1991), and Breslow and Clayton (1993) for random effects generalized linear models (GLMs); and Lindstrom and Bates (1990) for non-linear regression models with Gaussian random effects and errors. In this article, we apply the Monte Carlo EM ($MCEM$) described in Tanner (1996) to get maximum likelihood estimates. The importance sampling technique is used to approximate the expectations in the E step. Geweke (1989) has shown that the approximation converges to the true value almost surely. This method is easy to implement. The M-step is accomplished using the weighted least squares (WLS).

The rest of the chapter is organized as follows: Section 2 introduces the general latent structure framework for longitudinal ordinal data. Section 3 discusses the EM-algorithm for parameter estimation. Section 4 applies our methodology to two examples. Section 5 provides further comments and discussions.

3.2 Basic Framework

Let Z_{ij} be a sequence of continuous latent random variables on the i^{th} of n subjects, and $t_{ij}, j = 1, \ldots, m_i$ be the corresponding times at which the measurements are taken. Suppose that we cannot observe Z_{ij} directly, instead we are only able to measure Y_{ij}, a sequence of ordinal variables, which classify Z_{ij} into S categories by a set of ordered thresholds $-\infty = \alpha_0 < \alpha_1 < \cdots < \alpha_S = \infty$, defined by

$$Y_{ij} = s, \quad \text{iff} \quad Z_{ij} \in (\alpha_{s-1}, \alpha_s].$$

Associated with each Y_{ij} is a vector, x_{ij}, of p explanatory variables. Given the continuous latent structure of Z_{ij}, along with a set of thresholds points, an ordinal model for Y_{ij} is obtained. Before we discuss models for correlated ordinal data, we first consider the independent latent structure model. Suppose that Z_{ij} are independent of each other and follow a linear model

$$Z_{ij} = x_{ij}^T \beta + e_{ij}$$

where the e_{ij} are from a common distribution with CDF $F(\cdot)$. Then the corresponding ordinal model has the form

$$Pr(Y_{ij} \leq s \mid x_{ij}) = F(\alpha_s - x_{ij}^T \beta). \tag{3.1}$$

If e_{ij} follow a standard logistic distribution, then model (3.1) is a proportional odds model, and if e_{ij} are from a standard normal distribution, model (3.1) is a ordinal threshold probit regression model. In general, any monotone increasing function mapping $(-\infty, \infty)$ onto the unit interval $(0, 1)$ can be used as $F(\cdot)$. Some other common candidates are inverse log-log, inverse complementary log-log, and Cauchy functions.

We now discuss two types of continuous latent structure: random effects models and Markov (autoregressive) models, and their respective ordinal models.

(a) Random effects models

When subjects are sampled at random from a population, various aspects of their behavior may show stochastic variation between subjects. For instance, some subjects are intrinsically high responders, others are low responders. One way to incorporate this feature in specific models is to allow some regression coefficients to vary from one individual to the next. Suppose the Z_{ij} follow a random effects model

$$Z_{ij} = x_{ij}^T\beta + w_{ij}^T b_i + e_{ij}$$

where w_{ij} is $q \times 1$ vector of explanatory variables attached to individual measurements, and $b = (b_1, \ldots, b_q)^t$ is a vector of random coefficients. One simple case is the random intercept model, where $w_{ij} = 1$. Different intercepts for different individuals can be interpreted as the different location parameters for the underlying continuous responses or the shifts of the thresholds for individuals, depending on the practical situation. For example, if the measurements are taken from different individuals by one investigator or using one instrument, it is appropriate to think of different intercepts as having different location parameters for different individuals; if the measurements are taken from one individual by several investigators or using different instruments, the second interpretation is more appropriate. Another case is when $x_{ij} = w_{ij}$, so that each individual can be thought to have their own regression coefficients. With a large number of observations for each subject, we could estimate their individual coefficients. But in practice, we have limited data and must borrow strength across subjects to make inferences on β or the b. This is accomplished by assuming that the b_i are independent realizations from a distribution.

In the random effect models, we assume: (1) the distribution of each latent variable Z_{ij} conditional on b_i follows exponential family law with density $f(z_{ij} \mid b_i, \beta)$; (2) given b_i, $Z_{i1}, \ldots, Z_{im_i}$ are independent, so the repeated measurements $Y_{i1}, \ldots, Y_{im_i}$ are conditionally independent; (3) the b_i are independent and identically distributed with density function $f(b_i; G)$. Therefore the ordinal model for Y_{ij} given b has the form

$$Pr(Y_{ij} \leq s \mid b_i, x_{ij}, w_{ij}) = F(\alpha_s - x_{ij}^T\beta - w_{ij}^T b_i). \tag{3.2}$$

(b) Markov (autoregressive) models

When the time dependence is central, models for the conditional distribution of Y_{ij} given $Y_{i1}, \ldots, Y_{i,j-1}$ may be more appropriate. Correlation among $Y_{i1}, \ldots, Y_{im_i}$ exists because they inherit a certain structure from an underlying continuous process $\{Z_{ij}\}$. We will focus on the case where the observation times are equally spaced. The most useful transition models are Markov chains for which the conditional distribution of Z_{ij} given its history depends only on the q prior observations. The integer q is referred to as the model *order*.

A *qth-order autoregressive process* can be written as

$$Z_{ij} = x_{ij}^T \beta + e_{ij},$$

where

$$e_{ij} = \sum_{r=1}^{q} \gamma_r e_{i,j-r} + \epsilon_{ij},$$

and the ϵ_{ij} are mutually independent mean-zero random variables. This is a Markov model of the form

$$Z_{ij} = x_{ij}^T \beta + \sum_{r=1}^{q} \gamma_r (Z_{i,j-r} - x_{i,j-r}^T \beta) + \epsilon_{ij}.$$

So, given $Z_{i,j-1}, \ldots, Z_{i,j-q}$, the corresponding conditional ordinal model for Y_{ij} is given by

$$Pr(Y_{ij} \leq s \mid Z_{i,j-1}, \ldots, Z_{i,j-q}) = F(\alpha_s - x_{ij}^T \beta - \sum_{r=1}^{q} \gamma_r (Z_{i,j-r} - x_{i,j-r}^T \beta)). \tag{3.3}$$

A general latent structure model is given by

$$Z_{ij} = x_{ij}^T \beta + w_{ij}^T b_i + \eta_i(t_{ij}) + \epsilon_{ij}, \tag{3.4}$$

where b_i are a set of mutually independent Gaussian random vectors, and the $\eta_i(t_{ij})$ are realizations a stationary Gaussian process. The Z_{ij} are "missing" data. We apply the Monte Carlo EM approach for handling the missing data.

3.3 Estimation: *MCEM* Algorithm

The *EM* algorithm (Dempster et al 1977) is a very general algorithm for ML estimation in missing data problems. The idea can be stated as follows: Augment the observed data with latent data so that the augmented distribution is "simple" More specifically the *EM* algorithm is an iterative method for handling "missing" data: (1) *E* step finds the conditional expectation of the "missing" data given the observed data and current estimated parameters, then substitutes these expectations for the missing data. (2) *M* step performs maximum likelihood estimation as there were no "missing" data. The iterations continue until convergence. Since the computation of the expectations are difficult analytically, we apply the Monte Carlo *EM* approach (Wei and Tanner 1990). The importance sampling idea is used to facilitate the *E* step.

In the *M* step, we need to fit an independent latent structure model given the values for the latent variables. In order to fit a proportional odds model, we apply the binary coding of ordinal data discussed in Chapter 2. The ordinal response is represented as a vector of indicator variables

$$Y_{ij}^{*} = (Y_{ij1}^{*}, \ldots, Y_{ijS-1}^{*})^{T}$$

where $Y_{ijs}^{*} = 1$ if $Y_{ij} \leq s$, and $Y_{ijs}^{*} = 0$ if $Y_{ij} > s$. This describes an ordinal response by a set of binary responses at each level, each indicating whether the ordinal response is above or below the level. This coding system is convenient for getting the likelihood equation, and its advantage becomes more clear in the context of censored ordinal data (Chapter 5).

(a) Random effects models

The log-likelihood function for the unknown parameter θ, which is defined to include α, β, and the elements of G, given the *complete data* is

$$\begin{aligned} l(\theta \mid Y, b) &= \sum_{i=1}^{n} \sum_{j=1}^{m_i} \log[p(Y_{ij}, b_i \mid \theta)] \\ &= \sum_{i=1}^{n} \sum_{j=1}^{m_i} \log[p(Y_{ij} \mid b_i, \theta)] + \log[p(b_i \mid \theta)]. \end{aligned}$$

In the general EM setting, the E-step consists of computing

$$Q(\theta, \theta^l) = \int_b l(\theta \mid Y, b) p(b \mid Y, \theta^l) db.$$

where θ^l is the current parameter estimate in the l^{th} iteration. When applying the Monte Carlo method to approximate the integral, it is difficult to directly sample from $p(b_i \mid Y_i, \theta^l)$. To facilitate the E-step, we apply the importance sampling technique. Note that the importance sampling method only requires that $p(b_i \mid Y_i, \theta^l)$ be known up to a proportionality constant. This observation is key, since we have avoided the need to standardize $p(b_i \mid Y_i, \theta^l)$. We know that $p(b_i \mid Y_i, \theta^l) \propto p(Y_i \mid b_i, \theta^l) p(b_i \mid \theta^l)$, and $p(b_i \mid \theta^l)$ is easy to sample from. Therefore the method of importance sampling is as follows:

(i) Draw $b_i^{(1)}, \ldots, b_i^{(K)}$ from $p(b_i \mid \theta^l)$;

(ii) Let $v_i^{(k)} = Pr(Y_i = y_i \mid b_i^{(k)}, \theta^l) = \prod_{j=1}^{m_i} Pr(Y_{ij} = y_{ij} \mid b_i^{(k)}, \theta^l)$;

(iii) Let $\omega_i^{(k)} = v_i^{(k)} / \sum_{k=1}^{K} v_i^{(k)}$.

(iv) Approximate

$$Q(\theta, \theta^l) \approx \sum_{i=1}^{n} \sum_{j=1}^{m_i} \sum_{k=1}^{K} \omega_i^{(k)} \{\log[p(y_{ij} \mid b_i^{(k)}, \theta)] + \log[p(b_i^{(k)} \mid \theta)]\}.$$

Assume that b_i are independent Gaussian $N(0, G)$, in the M-step we solve two sets of score equations for (α, β) and G, which are given by

$$S(\alpha, \beta) = \sum_{i=1}^{n} \sum_{j=1}^{m_i} \sum_{k=1}^{K} \omega_i^{(k)} D_{ij}^T V_{ij}^{-1} (y_{ij}^* - F(\alpha - x_{ij}^T \beta - w_{ij}^T b_i^{(k)})) = 0 \qquad (3.5)$$

$$S(G) = \frac{1}{2} G^{-1} (\sum_{i=1}^{n} \sum_{k=1}^{K} \omega_i^{(k)} b_i^{(k)} b_i^{(k)T}) G^{-1} - \frac{n}{2} G^{-1} = 0 \qquad (3.6)$$

where $D_{ij} = \partial F / \partial(\alpha, \beta)$, and $V_{ij} = \mathrm{cov}(Y_{ij}^*)$. The first set of equations are the weighted likelihood equations of an independent latent structure model given the random coefficients, and the second set of equations are the weighted likelihood equations for a multivariate normal distribution. The algorithm proceeds by iteratively updating first the estimates of the regression coefficients and then the variance of the random effects, until the parameter estimates converge.

The parameters α and β are of main interest, and G is considered a nuisance parameter. Notice that $(\hat{\alpha}, \hat{\beta})$ is independent of $\hat{G}$. We apply Louis (1982) method to estimate the covariance matrix of $(\hat{\alpha}, \hat{\beta})$. Define

$$r_{ijk} = \omega_i^{(k)} D_{ij}^T V_{ij}^{-1}(y_{ij}^\star - F(\alpha - x_{ij}^T\beta - w_{ij}^T b_i^{(k)})),$$

and

$$s_{ri}^2 = \sum_{j=1}^{m_i}\sum_{k=1}^{K} \frac{(r_{ijk} - \bar{r}_i)(r_{ijk} - \bar{r}_i)'}{m_i K - 1} \tag{3.7}$$

The estimate for the information matrix of $\hat{\alpha}$ and $\hat{\beta}$ is given by

$$\text{var}(\alpha, \beta) = -\left.\frac{\partial^2 Q(\theta, \phi)}{\partial(\alpha, \beta)^2}\right|_{\theta,\phi=\hat{\theta}} - \sum_i^n s_{ri}^2 \mid_{\theta=\hat{\theta}} \tag{3.8}$$

$\partial^2 Q(\theta, \phi)/\partial(\alpha, \beta)^2$ is the information matrix for the complete data, which can be estimated using the standard algorithm for fitting proportional odds models.

(b) Markov (autoregressive) models

Let $e_{ij-} = (e_{i,j-1}, \ldots, e_{i,j-q})^T$, The log-likelihood function given the *complete data* is

$$\begin{aligned} l(\theta \mid Y, e) &= \sum_{i=1}^{n}\sum_{j=1}^{m_i} \log[p(Y_{ij}, e_i \mid \theta)] \\ &= \sum_{i=1}^{n}\sum_{j=1}^{m_i} \log[p(Y_{ij} \mid e_{ij-}, \theta)] + \log[p(e_{ij-} \mid \theta)], \end{aligned}$$

where $e = (e_i, \ldots, e_n)^T$, and $e_i = (e_{i1}, \ldots, e_{im_i})^T$. Similarly, we carry out the approximation in the E-step as follows:

(i) Draw $e_i^{(1)}, \ldots, e_i^{(K)}$ from $p(e_i \mid \theta^l)$,

(ii) Let $v_i^{(k)} = Pr(Y_i = y_i \mid e_i^{(k)}, \theta^l) = \prod_{j=1}^{m_i} Pr(Y_{ij} = y_{ij} \mid e_i^{(k)}, \theta^l)$;

(iii) Let $\omega_i^{(k)} = v_i^{(k)} / \sum_{k=1}^{K} v_i^{(k)}$.

(iv) Approximate

$$Q(\theta, \theta^l) = \int_b l(\theta \mid Y, e) p(e \mid Y, \theta) dZ$$

$$\approx \sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k=1}^{K}\omega_i^{(k)}\{\log[p(y_{ij} \mid e_i^{(k)},\theta)] + \log[p(e_i^{(k)} \mid \theta)]\}.$$

In practice, assigning each simulated series one weights can converge very slowly. To increase the effectiveness of the Monte Carlo sample, we divide the simulated series into several sub-series, and calculated the weights for each sub-series. These weights reflect the local likelihood of the sub-series given the observed ordinal outcome and current parameter estimates. We use these weights for approximating $Q(\theta,\theta^l)$ as if there were several independent short series.

In the M-step, the estimating equations are given by

$$S(\theta) = \sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k=1}^{K}\omega_i^{(k)}\begin{pmatrix} D_{ij}^T \\ e_{ij-}^{(k)T}\end{pmatrix}\begin{pmatrix} V_{ij}^{-1} & 0 \\ 0 & I\end{pmatrix}\begin{pmatrix} y_{ij}^* - F(\alpha - x_{ij}^T\beta - \sum_{r=1}^{q}\gamma_r e_{i.j-r}^{(k)}) \\ e_{ij}^{(k)} - \sum_{r=1}^{q}\gamma_r e_{i,j-r}^{(k)}\end{pmatrix} = 0 \tag{3.9}$$

The score functions also consist of two sets of equations, one from a independent latent structure model given the latent continuous times series, the other from a autoregressive model of the latent variable given the observed discretization. This methodology can be readily extended to the estimation of the general structure model (3.4).

3.4 Examples

Example 1. Analysis of Ulcer Data

Uesaka and Asano (1987) described a randomized controlled clinical trial of three treatments for ulcer. Three drugs under study are randomly allocated to a number of patients. At each of three visits during the follow-up period, the response status of each patient was measured. Table 3.1 shows the classification of the percentages of the diameter of ulcer at 2, 4 and 6 weeks of administration period to those of the first examination. The three classes are $0-20\%$, $21-50\%$, and $51\%-100\%$, which are scored 1, 2 and 3, respectively. In the study many patients missed X-ray inspection at least one observation period. When the ulcer had been sufficient small at one observation period, the

subsequent measurements were often skipped. In this case, score 1 is given. As a result 83 patients were available for the study.

To take into account the influence of heterogeneous individuals on the responses, we consider the cumulative logit random effects model

$$\text{logit}Pr(Y_{ijt} \leq s \mid b_i) = \alpha_s + b_i + \beta_j + \tau_t.$$

where τ_1 and τ_2 are the treatment effects of treatment A_1 and A_2 with the effect of A_3 as the baseline, β_1 and β_2 are the period effects of week 4 and week 6 respectively, and b_i are mutually independent $N(0, \sigma^2)$. This model states that each patient has their own cumulative probability at a response level, $Pr(Y_{ijt} \leq s)$ given by $\exp(\alpha_s + b_i + \beta_j + \tau_t)/\{1 + \exp(\alpha_s + b_i + \beta_j + \tau_t)\}$. It further states that a person's odds at a response level relative to treatment A_3 are multiplied by $\exp(\tau_t)$ when taking treatment A_t, $t = 1, 2$, regardless of the initial risk.

A random intercept for each individual was imputed from the appropriate normal distribution with standard deviation taken as the current estimate of σ^l. Then we calculated the likelihood ω_i of the observed outcome $(y_{i1}, y_{i2}, y_{i3})'$ given the current estimates of $\alpha_1^l, \alpha_2^l, \tau_1^l, \tau_2^l, \beta_1^l, \beta_2^l$. The process is repeated to obtain K augmented data sets. It follows that in the M-step, we updated the estimates $\alpha_1^{l+1}, \alpha_2^{l+1}, \tau_1^{l+1}, \tau_2^{l+1}, \beta_1^{l+1}, \beta_2^{l+1}$ through a weighted proportional odds model, and σ^{l+1} through weighted least squares estimation. It is usually inefficient to start with a large value of K when the current estimates are far from the mode. Rather, one may increase K as the current approximation moves closer to the true mode.

The algorithm was initiated with $\alpha_1 = 1.73$, $\alpha_2 = -0.175$, $\tau_1 = 0.829$, $\tau_2 = 0.432$, $\beta_1 = -2.81$, $\beta_2 = -4.02$, $\sigma = 1.00$. These values except σ were obtained by fitting an independent model. The value of K was equal to 50 (1000) for iterations 1-11 (12-15). Table 3.2 gives history of the Monte Carlo EM algorithm. The convergence criterion is $\max_\theta \mid \theta^{l+1} - \theta^l \mid < 0.01$.

The results are given in Table 3.3-3.4. This analysis indicates the treatment A_1 has the least effect on ulcer, A_3 is the most effective treatment. In fact, treatment A_3 is a mixture of A_1 and A_2 and has been expected to be most effective. And it is also known that ulcer tends to cure if life environment of the patient would be improved. Form the result we can see that the effect of time is highly significant.

Table 3.1: Data from the study of Anti-ulcer drugs

	$DrugA_1$			$DrugA_2$			$DrugA_3$		
No.	2w	4w	6w	2w	4w	6w	2w	4w	6w
1	2	1	1	3	3	2	3	2	2
2	2	2	1	2	1	1	1	1	1
3	2	1	1	2	1	1	3	2	1
4	3	1	1	1	1	1	3	1	1
5	2	1	1	2	1	1	3	1	1
6	2	1	1	2	1	1	2	1	1
7	3	2	1	1	1	1	3	1	1
8	2	1	1	3	1	1	1	1	1
9	3	3	3	2	2	1	3	1	1
10	3	2	1	2	1	1	3	1	1
11	3	2	1	3	1	1	2	1	1
12	3	3	2	3	1	1	2	2	1
13	2	1	1	3	3	2	3	1	1
14	2	1	1	2	1	1	3	2	1
15	3	3	2	3	1	1	3	1	1
16	2	1	1	3	2	1	1	1	1
17	3	1	3	2	1	1	1	1	1
18	3	1	1	2	1	1	3	1	1
19	3	1	1	3	2	1	1	1	1
20	3	2	2	3	2	1	2	1	1
21	3	1	1	2	1	1	2	1	1
22	3	2	1	2	1	1	3	1	1
23				3	2	1	2	1	1
24				3	1	1	3	3	1
25				2	1	1	3	1	1
26				3	1	1	2	2	1
27				3	1	1	2	2	1
28				2	1	1	2	1	1
29				3	3	2	3	2	1
30				3	2	2			
31				3	3	2			
32				3	1	1			

Table 3.2: History of MCEM for ulcer data

Iteration	α_1	α_2	τ_1	τ_2	β_1	β_2	σ
1	-1.705	0.214	-0.952	-0.501	2.846	4.094	1.037
2	-1.704	0.215	-0.950	-0.500	2.846	4.093	1.021
3	-1.692	0.223	-0.935	-0.494	2.848	4.090	1.030
4	-1.718	0.224	-0.890	-0.496	2.843	4.074	1.098
5	-1.685	0.225	-0.950	-0.496	2.851	4.096	1.106
6	-1.718	0.224	-0.890	-0.496	2.843	4.073	1.133
7	-1.718	0.224	-0.890	-0.496	2.843	4.074	1.187
8	-1.718	0.224	-0.890	-0.496	2.843	4.074	1.241
9	-1.685	0.225	-0.949	-0.496	2.851	4.096	1.245
10	-1.685	0.225	-0.950	-0.496	2.851	4.096	1.275
11	-1.685	0.229	-0.949	-0.495	2.851	4.098	1.277
12	-1.677	0.229	-0.949	-0.495	2.853	4.097	1.283
13	-1.684	0.226	-0.949	-0.496	2.851	4.097	1.287
14	-1.684	0.227	-0.949	-0.495	2.851	4.096	1.286
15	-1.683	0.226	-0.949	-0.496	2.854	4.096	1.287

Table 3.3: Estimates of treatment effects and their odds ratios

Effects	Estimate	sd	p-value	Odds ratio
A_3	0	-	-	1
A_2	-0.496	0.353	0.149	0.609
A_1	-0.949	0.372	0.011	0.387

Table 3.4: Estimates of period effects and their odds ratios

Effects	Estimate	sd	p-value	Odds ratio
2nd week	0	-	-	1
4th week	2.85	0.365	0.001	17.3
6th week	4.10	0.443	0.001	60.3

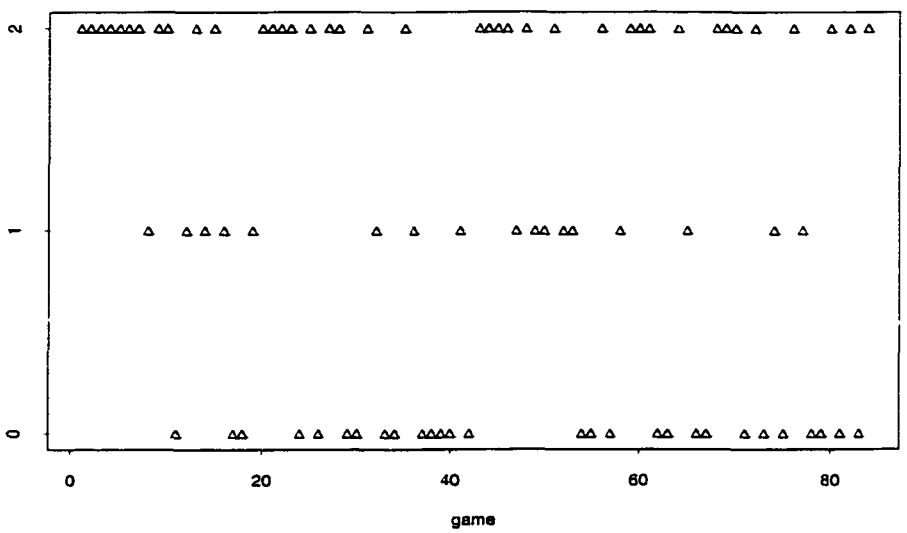


Figure 3.1: Maples Leafs' regular season results in 1993-1994

Example 2. Analysis of Toronto Maple Leafs games in 1993-1994 season

In the ordinal-valued time series case, the question of interest include: Is there serial dependence? Is a trend present? We consider the games of the Toronto Maple Leafs hockey team during the 1993-1994 season. The data are taken from Brillinger (1996), who recoded the data with two binary variables, "win" and "tie" and used a generalized linear model to analyze them separately. There were 84 regular season games, and the ordinal-valued time series, Y_t, take values 0,1 and 2 corresponding to the results loss, tie, win, respectively. The results correspond to the state of the game after regulation time. Figure 3.1 provides a graph of the results. The Toronto team began the season with a record setting winning streak of 10 games, and the team had 28 losses, 17 ties and 39 wins during the whole season.

Figure 3.2-3.3 provide smoothed estimates of the probability of a win and of a win or a tie respectively. The 95% confidence bands are also computed. The analysis is carried out as if the successive games are independent. These curves were produced employing a regression spline technique to be discussed in Chapter 4. Except for the early wins, the estimated win and loss curves rapidly moved to constants.

A basic question researchers usually pose when analyzing time series is the question of serial correlation: Does the result of the current game depend on its history? For time series of continuous variables, people usually summarize this dependence by the autocorrelation function (ACF) based on Pearson correlation, which describes the degree to which two dependent variables have a linear relationship. Ordinal variables do not have a defined metric, so the notion of linearity is not meaningful. However, the inherent ordering allows consideration of monotonicity, for instance, whether Y_{t+1} tends to increase as Y_t does. We use Kendall's tau analogously to the Pearson correlation in the autocorrelation function to describe the degree to which the relationship is monotone. Since the random variable Y is derived from Z, the correlation between Y_i and Y_{i-1} will inherit the relationship between Z_i and Z_{i-1}. Although less sensitive, the ACF function of the categorized variables can still reflect the autocorrelation of the latent continuous variables. For instance, if the continuous latent variables are independent, the derived ordinal random variables will also be independent.

The values of Kendall's tau from lag 0 to lag 10 are displayed in Figure 3.4. The autocorrelation at lag 0 is always 1 by definition. The horizontal dotted lines provide an approximate 95% confidence interval for the estimate of Kendall's tau at each lag. If no estimated value falls outside the confidence band, there is no significant evidence to suggest that there is serial correlation. Otherwise one should be concerned about the presence of serial correlation. In our example the plot does not indicate any strong serial correlation.

We also consider a parametric model assuming a latent continuous-valued time series and a set of cutoff points:

$$Pr(Y_i \leq s \mid Z_{i-1}) = F(\alpha_s - \gamma_r Z_{i-1}),$$

and

$$Z_i = \gamma Z_{i-1} + \epsilon_i.$$

The algorithm was initiated with $\alpha_1 = 0.693$, $\alpha_2 = -0.143$, $\gamma = 0.1$. These values except γ were obtained by fitting an independence model without serial correlation. The whole sequence is divided into 8 subsequences. The value of K was equal to 200 (3000)

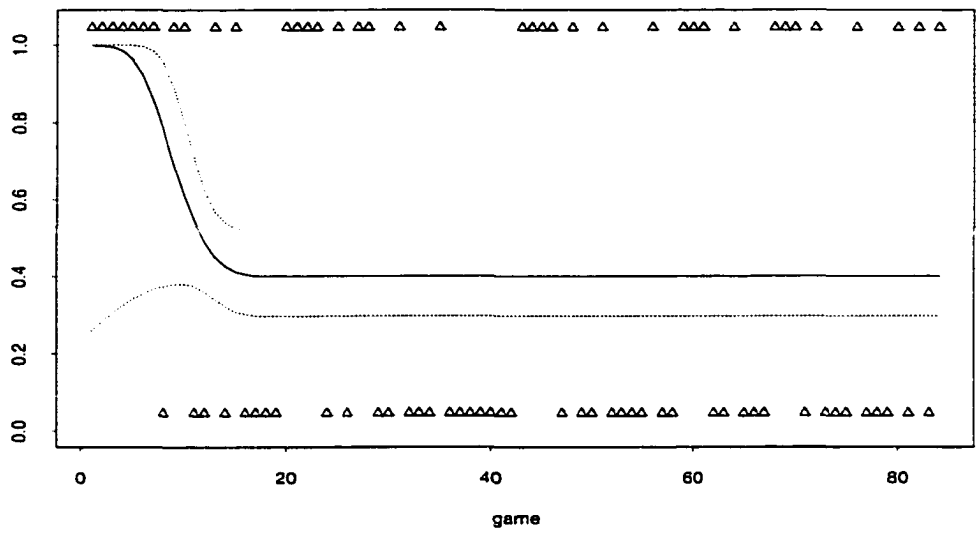


Figure 3.2: Smoothed "trend" of wins

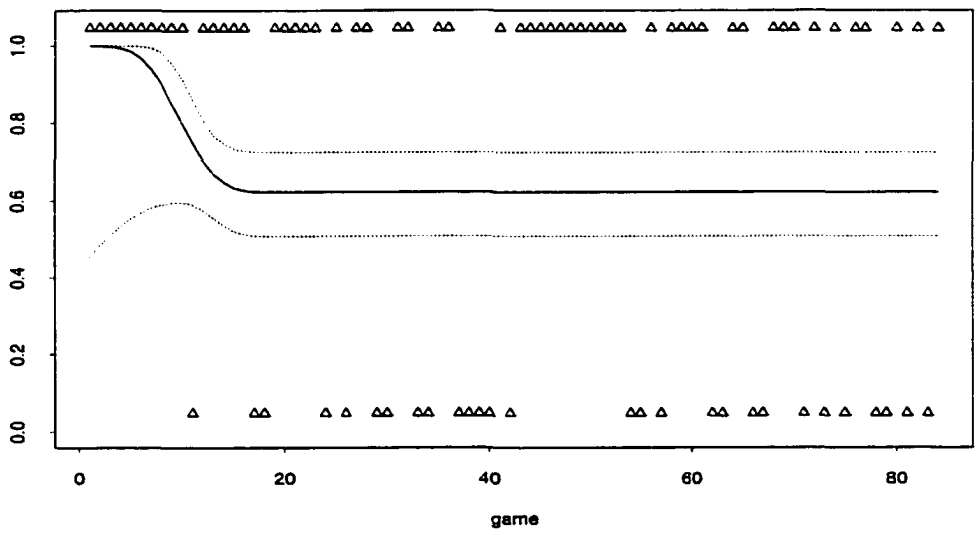


Figure 3.3: Smoothed "trend" of wins or ties vs. loss

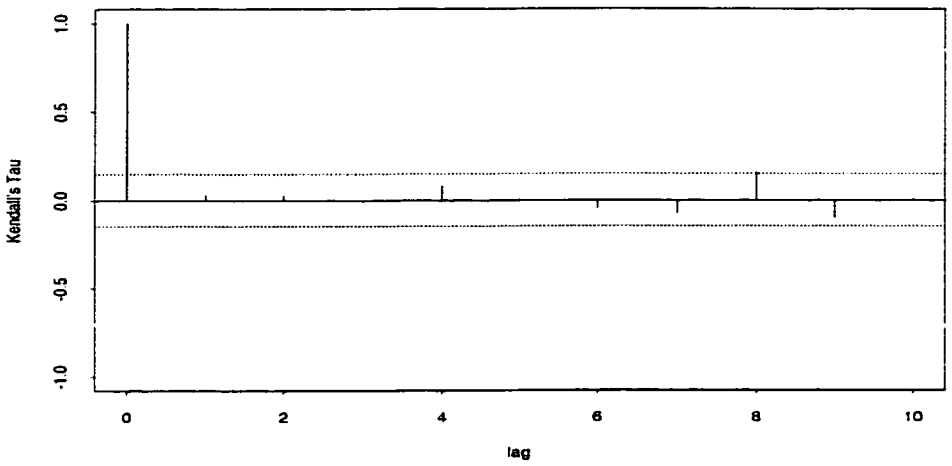


Figure 3.4: ACF using Kendall's tau

Table 3.5: History of MCEM for maples data

Iteration	α_1	α_2	γ
1	0.693	-0.143	0.117
2	0.693	-0.143	0.081
3	0.693	-0.143	0.028
4	0.693	-0.143	0.025
5	0.693	-0.143	-0.016
6	0.693	-0.143	-0.022
7	0.693	-0.143	-0.013
8	0.693	-0.143	0.000
9	0.693	-0.143	-0.012
10	0.693	-0.143	-0.013
11	0.693	-0.143	0.004
12	0.693	-0.143	-0.029
13	0.693	-0.143	-0.007
14	0.693	-0.143	0.001
15	0.693	-0.143	0.013
16	0.693	-0.143	0.015
17	0.693	-0.143	0.012
18	0.693	-0.143	0.013
19	0.693	-0.143	0.013

for iterations 1-16 (17-19). The convergence criterion is $\max_\theta \mid \theta^{l+1} - \theta^l \mid < 0.01$. Table 3.5 gives history of the Monte Carlo EM algorithm. We notice that the rate of convergence is much slower for the autoregression parameter than that for the intercept parameters.

The 1993-1994 Toronto team began the season with a winning streak, however, ultimately the results of the various games appear to be random. The analyses provide no real evidence of serial correlation.

3.5 Discussion

The analysis of ordinal longitudinal data is difficult partly because few models for the joint distribution of the repeated observations for a subject are available. The methods we propose provide a way of modeling the joint distribution. The within subject correlations between the ordinal responses is modeled through a latent continuous variable. A nice feature for the latent structure model is that the parameters do not change if two or more contiguous levels are combined, and thus are insensitive to the arbitrary or subjective nature of the category partition. This approach also offers ease of interpretation.

For ordinal data, we lost the information of the relative scale between the categories, so we should be cautious when we treat the ordered categories as scores and use methods for continuous data. The latent structure models the underlying scale is estimated in terms a set of intercepts α's, which represent the set of cutpoints, along with the parameters of interests.

For fitting random effect GLM, an alternative strategy to approximate conditional means is to use conditional modes. This approximation method gives reasonable estimates of α, and β in many problems. The approximation breaks down when there are few observations per subject and the GLM is far from the Gaussian. Karim (1991) and Breslow and Clayton (1993) have evaluated this approximate method for some special random effects GLMs. Using the importance sampling method, we can approximate the conditional means directly, the accuracy increases as K increases, and the approximation is independent of the form of the model. But this method requires more computational power, because each observation must be replaced by K observations.

There are other methods for modeling ordinal time series. Diggle et al. (1994) described a class of transitional models, in which a first order model can be expressed

as $\log\{Pr(Y_{ij} \leq s \mid Y_{i,j-1} = t)/Pr(Y_{ij} > s \mid Y_{i,j-1} = t)\} = \alpha_{st} + x'_{ij}\beta_t$. The ordered categorical time series is treated as a Markov Chain. The transition matrix become more and more complicated when the number of categories or the order of the chain increases. For example, a saturated second order Markov Chain without covariates for an ordinal time series with 5 levels has a transition matrix with $5^3 - 5^2$ parameters. The Latent structure model requires only 6 parameters, 4 cutpoints, 2 autoregression parameters. With latent structure models we can go beyond autoregressive models, Moving average (MA) models and other more complicated time series models can be formulated to analyze ordinal time series.

In some situations an ordinal response is not fully classified, the only information we know is that it lies in a set of contiguous categories. This is known as censoring. In chapter 5 we show how maximum likelihood estimates can be obtained using such partially classified observations.

Chapter 4

Multivariate Generalized Additive Models

4.1 Introduction

Regression is one of the most widely used of all statistical tools. Linear modeling in its widest sense is both well developed and and well understood, and there is in addition a variety of useful techniques for checking the assumptions involved. However, there are cases where such models cannot be applied because of intrinsic nonlinearity in the data. Non-parametric regression aims to provide a means of modeling such data. For the purpose of exploring data, smoothing techniques are useful by enhancing scatterplots to display the underlying structure of data, without reference to a parametric model. This can in turn lead to further useful suggestions of, or check on, appropriate parametric models. In this chapter, we introduce the use of non-parametric smoothing tools in the multinomial and ordinal data analysis problems, clustered data analysis problems, and longitudinal data analysis problems. The emphasis throughout is on using non-parametric techniques for exploration of data.

Generalized additive models were introduced by Hastie and Tibshirani (1984,1986). Their applications to logistic and proportional odds regression were discussed by Hastie and Tibshirani (1987, 1990). Yee and Wild (1996) proposed a class of vector generalized

additive models for marginal means, in which the correlation structure is assumed known and only depends on the marginal means. In this chapter, we extend the GAMs to multivariate data and propose a class of multivariate generalized additive models (MGAM), in which the components of a multivariate observation are correlated. Examples of models with this structure include multivariate linear regression, regression models for multivariate binary responses, and multinomial and ordinal regression models. MGAMs take into account the correlation between the components of a multivariate observation. In some applications, the correlation is only a function of the marginal means, while in other applications, the covariance matrix also depends on additional parameters. In MGAMs we model the correlation structure as well as the marginal means.

The penalized likelihood approach is used by Hastie and Tibshirani (1990) to fit generalized additive models. An alternative approach is regression splines. Regression splines are attractive because of their computational and statistical simplicity. For example, standard parametric inferential methods can be used to test the importance of any of the parameters, and standard procedures for fitting parametric models can be used. For fitting marginal means and the covariance structure, we use the maximum likelihood approach proposed by Zhao and Prentice (1990) for the "quadratic exponential family" If scientific interest is only in the marginal means as of a smooth function of the covariates, we apply the generalized estimating equation approach (GEE) proposed by Liang and Zeger (1986) to estimate the smooth function.

An important practical issue of using regression splines is that of "how many knots and where." Since the optimal number of knots depends on the assumption of the smoothness, and this is unknown, choosing the knots based only on data is an important issue. By choosing the number and location of knots, a parametric model as an approximation for the mean function is obtained. This is basically a model selection procedure. We use information based criterion, such as AIC or SIC, or a model selection based procedure for knots selection. For continuous response data, the issue of knot selection was discussed in He and Shi (1996).

Section 2 introduces a class of multivariate additive models. In section 3 we discuss the regression spline method, with a maximum likelihood estimation approach and a GEE approach for multivariate regression. Section 4 discusses the knot selection issue in using regression splines. An example is given in section 5 to illustrate the methodology. Further discussion is provided in section 6.

4.2 Multivariate Generalized Additive Models

A generalized additive model (GAM) differs from a generalized linear model (GLM) in that an additive predictor replaces the linear predictor. Specifically, we assume that the response y has a distribution in the exponential family. The mean $\mu = E(Y \mid x_1, \ldots, x_p)$ is linked to the predictor via

$$g(\mu) = \alpha + f_1(x_1) + \cdots + f_p(x_p).$$

where g is a known link function such as logit for binary response. With GAMs, instead of constraining the relationship between each x_i and $g(\mu)$ to be linear as in GLMs, the relationship is merely constrained to be smooth. This allows non-linear features of the data to be revealed.

Let us consider the situation for which the response is vector instead of a scalar. Suppose that for the ith individual under study an S-dimensional response vector $Y_i = (Y_{i1}, \ldots, Y_{iS})'$ and a p-dimensional covariate vector x are observed. Let $\mu_i = (\mu_{i1}, \ldots, \mu_{iS})'$ be the mean of Y given $x_1, \ldots, x_p$. The multivariate GAM model is defined as

$$g(\mu_i) = \alpha + \sum_{j=1}^{p} f_j(x_{ij}). \tag{4.1}$$

where $\alpha = (\alpha_1, \ldots, \alpha_S)'$ is a vector of intercepts, and $f_j = (f_{j1}, \ldots, f_{jS})'$ is a vector of smooth functions. The marginal variance depends on the marginal mean according to

$$\mathrm{Var}(Y_{is}) = v(\mu_{ij})\phi \tag{4.2}$$

where $v(\cdot)$ is a known variance function and ϕ is a scale parameter that may require to be estimated. The correlation between Y_{is} and Y_{it} is a function of the marginal means and perhaps of additional parameters γ, i.e.

$$\mathrm{Corr}(Y_{is}, Y_{it}) = \rho(\mu_{is}, \mu_{it}, \gamma), \tag{4.3}$$

where $\rho(\cdot)$ is a known function.

As discussed in Chapter 2, an ordinal response can be represented by a vector of binary variables. So we can treat an ordinal regression model as a multivariate binary regression model, usually with constraint on the smooth function f_j. The correlation between Y_{is} and Y_{it} is a function of only the marginal means. For example, an additive cumulative link model extended from a linear cumulative link model (1.4) is defined as

$$F^{-1}(\mu_{is}) = \alpha_s - \sum_{j=1}^{p} f_j(x_{ij}).$$

where $F^{-1}(\cdot)$ is the link function, α is a set of cutpoints, $\mu_{is} = Pr(O_i \leq s \mid x)$, and O_i is the ordinal outcome of the ith individual. The smooth function f_j satisfies $f_{js} = f_{jt} = f_j$. Letting $Y_{is} = I(O_i \leq s)$, the covariance between Y_{is} and Y_{it} is given by

$$\text{Corr}(Y_{is}, Y_{it}) = \frac{[\mu_{is}(1-\mu_{it})]^{\frac{1}{2}}}{[(1-\mu_{is})\mu_{it}]^{\frac{1}{2}}}, \qquad \text{if } s \leq t.$$

In longitudinal studies. The observations within each subject are correlated. In this case, the correlation between Y_{is} and Y_{it} usually does not depend only on the marginal means. Additional parameters are needed for describing the covariance structure. For example, a marginal repeated measurement model of this kind with uniform correlation structure is given by the following assumptions:

(i) $g(\mu_{ij}) = \alpha - \sum_{j=1}^{p} f_j(x_{ij})$,

(ii) $\text{Var}(Y_{is}) = v(\mu_{ij})\phi$,

(iii) $\text{Corr}(Y_{is}, Y_{it}) = \gamma$,

where $g(\cdot)$ is identity link for multivariate general regression models assuming Gaussian errors, and logit link for multivariate logistic models.

4.3 Regression Splines, Quadratic Exponential Family, and GEE

Regression splines represent the fit as a piecewise polynomial. The regions that define the pieces are separated by a sequence of knots. In addition, it is forced that the piecewise polynomial to join smoothly at these knots. Regression splines define a set of basis functions which are piecewise polynomials centered at the knots. B-spline basis functions are a common choice. Given knots and order of the splines, we can construct linearly independent B-spline basis function, denoted by $B_k(x)$. Every spline $f(x)$ in the space spanned by $\{B_k(x)\}$ then has a unique representation

$$f(x) = B(x)'\beta.$$

where $B(x)$ is the vector of B-spline basis functions representing the particular family of piecewise polynomials, evaluated at the observed values of the predictor x. The smooth function is estimated by a multiple regression on $B(x)$. So, given knots and order of the splines, the estimation of the spline function can be formulated as a parametric problem of estimating β. There are several efficient algorithms to generate the B-spline basis function for a given set of knots. We use function *spline.des* in S-plus for generating the B-spline design matrix $B(x)$.

The joint distribution of the Y_i can be expressed in the log-linear specification, which assumes the pdf of Y_i is of the form

$$f(y_i, \theta_i, \gamma_i) = \exp\{\theta_i' y_i + \gamma_i' w_i - A(\theta_i, \gamma_i)\} \tag{4.4}$$

where $W_i = (Y_{i1}Y_{i2}, \ldots, Y_{i,S-1}Y_{iS}, \ldots, Y_{i1}Y_{i2}\ldots Y_{iS})'$ is a $(2^S - S - 1) \times 1$ vector of two and hig-way cross products of Y_i, $\theta_i = (\theta_{i1}, \ldots, \theta_{iS})'$, and $\gamma_i = (\gamma_{i12}, \ldots, \gamma_{i,S-1,S}, \ldots, \gamma_{i12\ldots S})'$ are vectors of canonical parameters, and $A(\theta_i, \gamma_i)$ is a normalizing constant. Zhao and Prentice (1990) proposed a likelihood based approach that is based on the "quadratic exponential family" with the three- and high-way association parameters set to zero. When these parameters are set to zero, 4.4 holds with $W_i = (Y_{i1}Y_{i2}, \ldots, Y_{iS-1}Y_{iS})'$, $\theta_i = (\theta_{i1}, \ldots, \theta_{iS})'$, and $\gamma_i = (\gamma_{i12}, \ldots, \gamma_{iS-1,S})'$. Zhao and Prentice (1990) propose to model the mean $\boldsymbol{\mu_i}$, and the covariance of the response as a function of covariates by

some specified link function. Using this approach, given a set of knots and the order of the spline function, the set of likelihood equations for β and γ have the following form

$$\sum_{i=1}^{n} \begin{pmatrix} \frac{\partial \mu_i}{\beta} & 0 \\ \frac{\partial \sigma_i}{\beta} & \frac{\partial \sigma_i}{\gamma} \end{pmatrix}' \begin{pmatrix} V_i & K_i \\ K_i' & U_i \end{pmatrix}^{-1} \begin{pmatrix} y_i - \mu_i \\ s_i - \sigma_i \end{pmatrix} = 0 \tag{4.5}$$

where $S_{ist} = (Y_{is} - \mu_{is})(Y_{it} - \mu_{it})$, $\sigma_{ist} = E(S_{ist})$, $K_i = cov(Y_i, S_i)$, and $U_i = \text{cov}(S_i)$. The equations given in (4.5) yield maximum likelihood estimates when the true three- and high-way association parameters are zero. A serious drawback of this approach, however, is that consistency of $\hat{\beta}$ and $\hat{\gamma}$ requires the correct specification of the model for both the mean and the pairwise marginal correlations, namely, $\hat{\beta}$ may fail to be consistent when the model for the marginal association is misspecified even if the model for the mean is correct.

An alternative approach is to use the GEE approach proposed by Liang and Zeger (1986). They derived two set of estimating equations for β and γ separately.

$$S(\beta) = \sum_{i=1}^{n} D_i' V_i^{-1} (y_i - \mu_i) = 0. \tag{4.6}$$

and

$$S(\gamma) = \sum_{i=1}^{n} A_i' B_i^{-1} (s_i - \sigma_i) = 0. \tag{4.7}$$

where $D_i = \partial \mu_i / \partial \beta'$, $A_i = \partial \sigma_i / \partial \gamma'$, and V_i and B_i are the "working" covariance matrices of Y_i and S_i, respectively. One major attractive feature of the GEE approach is that it provides a consistent estimate, $\hat{\beta}$, that only requires that the model for the marginal means are correctly specified. Regardless of whether the "working" correlations are correctly specified, consistent estimates of the regression parameters are obtained.

4.4 Knot Selection

The degree of smoothness of the true regression function determines how well the function can be approximated. In practice the degree of smoothness of the true regression function is unknown and has to be pre-determined by specific consideration or by examination of

the data. Piecewise linear, quadratic, and cubic splines are common in practice. They avoid the oscillation problem often associated with higher order polynomials, and provide considerable flexibililty.

Given a set of potential knots, we use the AIC function proposed by Akaike (1973) to select knots for low values of

$$AIC(\hat{\beta}) = D(y;\hat{\beta}) + \frac{2p}{n}$$

Where $D(y;\hat{\beta})$ is the Deviance, p is the dimensionality of the approximating model. Another information criterion in favor of parsimonious model is the SIC function (Schwarz (1978)), in which one minimizes

$$SIC(\hat{\beta}) = D(y;\hat{\beta}) + \frac{p \log n}{2n}$$

Alternatively we can use a simple model selection procedure such as forward, backward, or stepwise procedure for knot selection. Cross validation is another approach adopted by some researchers (Burman 1990, and He and Shi 1996). Simulation study by He and Shi (1996) showed that the information-based criteria generally perform a little better than cross validation for knot selection.

When the true function does not change dramatically, uniform knots are usually sufficient. The uniform knots here refer to having about equal number of observations between two contiguous knots. In our experience, starting from a set of uniform knots and using a stepwise model selection procedure works well for ordinal regression models, such as proportional odds models.

4.5 An Example

Figure 4.1 shows the number of occurrences of rainfall, Y_i, over 1 mm in Tokyo for each day during the two years (1983-1984). The problem is to estimate the probability of occurrence of rainfall. For the purpose of exploring the data, we use a proportional odds model with regression splines to fit the data. The model for the smoothed probability of

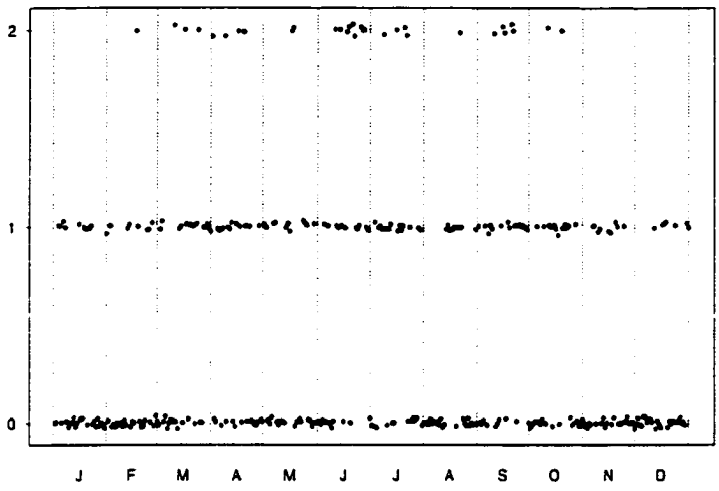


Figure 4.1: Number of rainfall occurrence in Tokyo, 1983-1984

number of occurrence is given by

$$\text{logit}\{Pr(Y_i \geq s)\} = \alpha_s + f(t_i), \qquad s = 1, 2.$$

We use a stepwise model selection procedure; the set of 12 potential knots is chosen at the beginning of each month. Figures 4.2 and 4.3 display the estimated probabilities of rainfall on either day ($Y_i \geq 1$) and of rainfalls on both days ($Y_i \geq 2$). The model is fitted as if the observations of successive days are independent. The dashed curves bound the 95% marginal confidence band. As mentioned before, one advantage of regression splines is that standard parametric inferential methods can be used to test statistical hypothesis. We used the score test defined in Chapter 2 to test the proportional odds assumption. The χ^2 test statistic is 7.79 with 4 degrees of freedom, and the p-value is 0.10.

Although the assumption of independence is not really statistical sound, the fitted curve can illustrate the gradual change of rainfall occurrence with season. Kitagawa (1987) used a state-space modeling approach to analyze this data set.

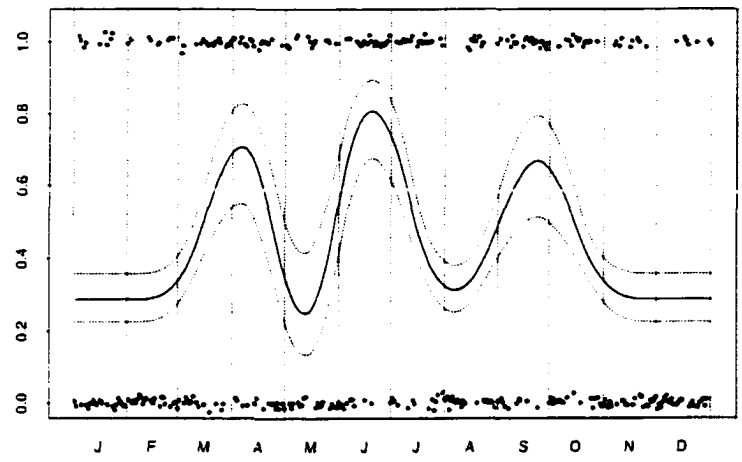


Figure 4.2: Estimated probability of rainfall

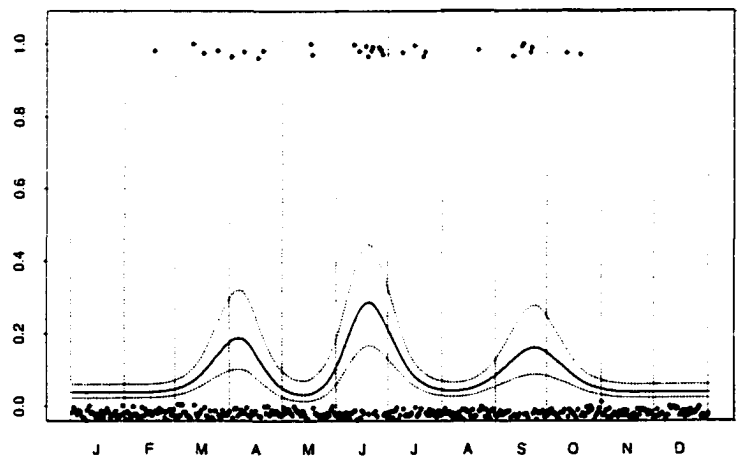


Figure 4.3: Estimated probability of rainfall occurred more than once

4.6 Discussion

In many situations, the additive extensions to GLMs provided by GAMs are invaluable for exploring data. This chapter proposes a further extension of this methodology to a multivariate setting in a natural way. The MGAMs have broad applicability for multinomial and ordinal data, where a response can be viewed as a vector of binary response, and for clustered and longitudinal data, where the observations for each subject are correlated. They can be used in a data analytic fashion to model and to test hypothesis about covariates. A more conservative approach is to use the non-parametric functions to suggest parametric transformations, and then proceed with the usual linear analysis on the transformed variables.

Regression splines provide a computationally convenient way of fitting additive models. When the knots are given, standard linear model estimation can be applied. However, the difficulty of choosing the number and location of the knots is a drawback of this approach. Information-based criteria or standard model selection procedures are applied in helping to select knots.

When dealing with more than one covariate, the backfitting algorithm discussed in Hastie and Tibshirani (1990) can be extended to MGAMs. It is an iterative algorithm for computing all f_j, based on the fact that $E(Y - \alpha - \sum_{j \neq k} f_j(x_j) \mid x_k) = f_k(x_k)$ if the MGAM (4.1) is correct. When the additivity assumption does not hold, a more general approach is to use multivariate splines, He and Shi (1996) discussed the use of bivariate tensor-product B-splines in a partly linear model.

Chapter 5

Censored data

5.1 Introduction

This chapter concerns the analysis of data when some of the values are not completely observed. Regression analysis is used to identify the relationship between the response variable and a set of covariates. A complicating phenomenon is censoring, where only partial information about the response is available. For example, in some toxicology studies the response is classified into several severity levels, such as no observable effect, mild effect, severe effect, and death, but in some circumstances part of the responses may be only recorded as either live or death, with the exact response level unknown. In some clinical settings, survival times are not observed directly. Instead, at intermittent examination times, it is determined whether or not the outcome has occurred or the observation has been censored. In such settings it is only known that the survival time is bracketed by the examination times immediately preceding and immediately following it. In the first example, the response variable is categorical, and in the second example the response is continuous. Data of this kind are termed interval censored data.

Relatively little has been published on the interval censored categorical data, particularly when a parametric regression model is assumed. Simpson et al (1996) discussed maximum likelihood estimation of ordinal regression in the presence of interval censoring. They derived the general form of the likelihood function subject to interval censoring. Related estimation problems for partially classified multinomial observations have been

studied by various authors, including Hartley (1958), Koch, Imrey, and Reinfurt (1972), Chen and Feinberg (1976), and Shipp, Howe, Watson, and Hogg (1991).

In survival analysis, when the data consist of the exact values and the right censored values, several parametric and non-parametric methods are available (Lawless 1982, Kalbfleisch and Prentice 1980). Interval censoring is another mechanism of censoring where the exact time of event is unknown; however the event is known to occur between two defined time points. Odell, Anderson, and D'Agostino (1992) applied a Weibull-based accelerated failure time model for interval censored data. They compared the maximum likelihood estimates for observed data with the method of substituting midpoints for interval censored data. The simulation studies indicate that for relatively large samples the maximum likelihood estimator is superior to the midpoint estimator, depending on the percentage of censored data. Kim (1997) used a loglinear model to incorporate interval censored failure times, assuming that the base line hazard is a step function on disjoint time intervals.

The EM algorithm, introduced in Dempster, Laird, and Rubin (1977) is a very general iterative algorithm for ML estimation in incomplete-data problems. For analyzing interval-censored data, the censored data are treated as "missing." We augment the fully observed data with "missing" censored data to simplify computations in the analysis. Tanner (1996) discussed the EM algorithm for simple linear regression with right censored data. In the case of censored regression data with normal errors, this approach reduces to an iterated series of least squares computations to estimate the regression coefficients. Kim (1997) discussed using the EM algorithm in fitting loglinear model with interval censored data. By augmenting the data, we replace the complicated analysis required in the observed data approach by a series of simple analyses. The main advantages of the EM algorithm are its generality and stability and, given a method to analyze complete data, its ease of implementation.

Although the EM algorithm is a general method for dealing with censored data, it has two major drawbacks. its typically slow rate of convergence, and its lack of the direct provision of a measure of precision for the estimator. There are more efficient algorithms for fitting some particular models. Agresti (1990) described a class of cumulative link models, which model the cumulative probabilities of the response via a strictly monotone function F^{-1} from $(0, 1)$ on the real line. For this type of model we derive a closed form

for the score function, and apply a weighted least squares algorithm to get parameter estimates.

In this chapter, we discuss the applications of EM algorithm for ML estimation in the analysis of interval-censored categorical and continuous response data. Section 3.1 introduces the EM algorithm for dealing with interval censored categorical data. Section 3.2 proposes a weighted least squares algorithm for fitting cumulative link models. Section 4 discusses the continuous response variable case. Section 5 gives further comments and discussion.

5.2 Censoring

Let Y_i be a discrete or continuous random variable with probability cumulative function F or density function f, Y_i is said to be censored in set C_i, if the only information we have about Y_i is that Y_i lies in C_i. For continuous variables, C_i usually is an interval $(C_{ia}, C_{ib}]$, which can take $\pm\infty$ as its boundary value. For example, two examinations at particular times to see whether a certain event has yet occurred will produce a censored observation of the time of occurrence of that event. Whether the observation is left-censored, right censored or interval censored depends on whether the event happened before the first examination, after the second examination or between two examinations. We make these distinction because the statistical methods are substantially different for the three types of censoring. For categorical responses, often referred as partially classified data, we do not make this distinction because of the limited response categories, or the lack of ordering for nominal data.

Assume *missing at random* defined in Little and Rubin (1987). likelihood inferences of censored data can ignore the stochastic nature of the censoring and treat observed data as if they were simple grouped data, for which the grouping is predetermined and non-stochastic. The likelihood for Y_i is

$$L(\theta \mid Y_i) = Pr(Y_i \in C_i) = \sum_{t \in C_i} Pr(Y_i = t)$$

if Y_i is categorical, and

$$L(\theta \mid Y_i) = Pr(Y_i \in C_i) = \int_{C_i} f(t)dt,$$

if Y_i is continuous.

5.3 Methods for Censored Categorical Responses

In the presence of censoring, usually the form of the likelihood functions is complicated, and computationally difficult to maximize directly. A general approach is to treat the interval censored observations as missing data, and use the EM algorithm to obtain the maximum likelihood estimates. Cumulative link models is a class of models for ordinal data, the likelihood function has a closed form. The weighted least square algorithm is derived for solving the score functions.

5.3.1 EM algorithm, a General Approach

Suppose we record the data so that the first m observations, denote by $y_1, \ldots, y_m$, are uncensored, and the remaining $n-m$ observations, denoted by $z_{m+1}, \ldots, z_n$, are censored, where $z_i \in C_i$. For ordinal response data, Z_i is often classified into fewer levels by collapsing contiguous categories. Let $p(\theta \mid Y, Z)$ denote the likelihood on completed data set, where Y represents the fully observed data, C represents the censored intervals, and Z represents the latent data; Let $l(y_i)$ denote the loglikelihood of y_i; and $p(Z \mid \theta^j, Y, C)$ denote the conditional predictive distribution of the latent data Z, conditional on the current estimates of the parameters. The E-step consists of computing

$$\begin{aligned} Q(\theta, \theta^j) &= \int_Z \log[p(\theta \mid Z, Y)]p(Z \mid \theta^j, Y, C)dZ \\ &= \sum_{i=1}^{m} l(y_i) + \sum_{i=m+1}^{n} \int_Z l(Z_i)p(Z_i \mid \theta^j, Y, C_i)dZ \end{aligned} \tag{5.1}$$

The conditional predictive distribution $p(Z_i \mid \theta^j, Y)$ is the conditional binomial or multinomial distribution, conditional on the fact that the unobserved response level Z_i is in C_i. Hence

$$\begin{aligned}\int_Z l(Z_i)p(Z_i \mid \theta^j, Y, C_i)dZ &= E(\ l(Z_i) \mid \theta^j, Z_i \in C_i) \\ &= \sum_{t_i \in C_i} Pr(Z_i = t_i \mid \theta^j, Z_i \in C_i)\ l(t_i) \\ &= \sum_{t_i \in C_i} w_{it_i}\ l(t_i)\end{aligned}$$

where $w_{it_i} = Pr(Z_i = t_i \mid \theta^j, Z_i \in C_i) = Pr(Z_i = t_i \mid \theta^j)/Pr(Z_i \in C_i \mid \theta^j)$ are the weights associated with the possible values of Z_i respectively. Therefore, $Q(\theta, \theta^j)$ is given by

$$Q(\theta, \theta^j) = \sum_{i=1}^{m} l(y_i) + \sum_{i=m+1}^{n} \sum_{t_i \in C_i} w_{it_i}\ l(t_i) \tag{5.2}$$

In the M-step, the Q function is maximized with respect to θ to obtain θ^{j+1}. Notice that the M-step is simply a weighted regression on the augmented data. The missing value Z_i is replaced by with all the possible values from c_{ia} to c_{ib} along with their weights. The algorithm is iterated until $\|\theta^{j+1} - \theta^j\|$ or $\mid Q(\theta^{j+1}, \theta^j) - Q(\theta^j, \theta^j) \mid$ is sufficiently small.

The EM algorithm is very attractive because of its simplicity. We can use standard software to facilitate the M-step. For example, the SAS procedures LOGISTIC and PROBIT provide ML fitting of proportional odds and threshold probit models, respectively. We have shown that a continuation ratio model can be fitted using only ordinary logistic regression procedure. As for the adjacent category model, it is known that we can express adjacent category logit models as baseline category logit models (see Agresti 1990, page 318), which can be fitted directly using SAS procedure CATMOD. In the E-step, only the weights need to be calculated. Usually, the E-step is easy to implement and the cost of computation is much lighter than that of the M-step.

The output of the EM algorithm, $\hat{\theta}$, is the maximum likelihood estimates. To get the covariance estimates of $\hat{\theta}$, one must compute the Hessian matrix of $\log p(\theta \mid Y)$.

For missing data problems, Louis (1982) stated the Missing Information Principle as following:

Observed Information = Complete Information - Missing Information.

A basic result due to Louis (1982) expresses the principle in the following form

$$-\frac{\partial^2 \log p(\theta \mid Y)}{\partial \theta^2} = -\left.\frac{\partial^2 Q(\theta,\phi)}{\partial \theta^2}\right|_{\phi=\theta} - \text{var}\left\{\frac{\partial \log p(\theta \mid Y,Z)}{\partial \theta}\right\},$$

where the variance is with respect to $p(Z \mid \theta, Y)$. The complete information $-\partial^2 Q(\theta,\phi)/\partial\theta^2$ is the complete information matrix obtained directly from the regression output at the final iteration for the augmented data. The missing information matrix is given by

$$\begin{aligned}\text{var}\left\{\frac{\partial \log p(\theta \mid Y,Z)}{\partial \theta}\right\} &= \text{var}\left\{\sum_{i=1}^{m}\frac{\partial l(y_i)}{\partial \theta} + \sum_{i=m+1}^{n}\frac{\partial l(z_i)}{\partial \theta}\right\} \\ &= \sum_{i=m+1}^{n}\text{var}\left\{\frac{\partial l(z_i)}{\partial \theta}\right\} \qquad (5.3)\end{aligned}$$

The information matrix of the observed data at the maximum likelihood estimate $\hat{\beta}$ can be estimated via

$$-\left.\frac{\partial^2 Q(\theta,\phi)}{\partial \theta^2}\right|_{\theta,\phi=\hat{\theta}} - \sum_{i=m+1}^{n}\left\{\sum_{t_i\in C_i} w_{it_i}\left(\left.\frac{\partial l(t_i)}{\partial \theta}\right|_{\theta=\hat{\theta}}\right)^2 - \left(\sum_{t_i\in C_i} w_{it_i}\left.\frac{\partial l(t_i)}{\partial \theta}\right|_{\theta=\hat{\theta}}\right)^2\right\} \qquad (5.4)$$

where w_{it_i} are the weights calculated from the last iteration.

Example 1. Continuation ratio models

As discussed in Chapter 1, we can obtain the maximum likelihood estimates for a continuation ratio model via fitting a simultaneous logistic regression on the recoded

binary data, defined as

$$Y_{is}^{*} = \begin{cases} 1 & \text{if } Y_i = s \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad W_{is}^{*} = \begin{cases} 1 & \text{if } Y_i \leq s \\ 0 & \text{otherwise} \end{cases} \quad s = 1, \ldots, S.$$

The likelihood, $l(y_i)$, is of form

$$W_{is}^{*}\{Y_{is}^{*}\log(P_{is}^{*}) + (1 - Y_{is}^{*})\log(1 - P_{is}^{*})\}$$

The M-step is accomplished by a standard logistic regression procedure.

5.3.2 Weighted Least Squares for Cumulative Link Models

Agresti (1990) presented a generalization of the proportional odds model that permits a variety of transformations for modeling ordinal response data. Let F denote the CDF of a continuous random variable having positive density over the entire real line. The F^{-1}, so called link function, is a strictly monotone function from $(0, 1)$ onto the real line. The cumulative link model has the form

$$F^{-1}\{Pr(Y \leq s \mid x)\} = \alpha_s - x'\beta,$$

or, equivalently

$$Pr(Y \leq s \mid x) = F(\alpha_s - x'\beta). \tag{5.5}$$

This model assumes that effects of x are the same for each cutpoint. This assumption holds if there is a linear regression for an underlying continuous response having standardized CDF F. McCullagh (1980) discussed several cumulative link models. The logit link, $F^{-1}(u) = \log[u/(1-u)]$, gives the proportional odds models. The standard normal CDF $F = \Phi$ gives the threshold probit models, a generalization of the binary probit model to ordinal data. The complementary log-log link, $F^{-1}(u) = \log[-\log(1-u)]$ is appropriate when the underlying distribution follows an exponential or extreme-value distribution. The ordinal model using this link is called a proportional hazard model,

because of the property

$$1 - Pr(Y \leq s \mid x_1) = [1 - Pr(Y \leq s \mid x_2)]^{\exp[(x_2 - x_1)'\beta]}$$

Proportional odds models and threshold probit models provide similar fits. The complementary log-log link, $\log[-\log(1-u)]$ is similar to the logit or probit for small u, but tends to ∞ much more slowly for large values.

Having y_i censored into the interval $\{c_{ia}, \ldots, c_{ib}\}$ can be viewed as collapsing the response into fewer levels. The response Z_i follows a conditional multinomial distribution, but the multinomial cells may differ for different observations, both respect to the number of available intervals as well as the interval boundaries. We use two sets of indicator variables to represent a censored response. Assume that $c_{ia} \leq y_i \leq c_{ib}$, we define

$$Y_{is}^{\star} = \begin{cases} 1 & \text{if } c_{ib} \leq s \\ 0 & \text{otherwise} \end{cases} \qquad \text{and} \qquad U_{is}^{\star} = \begin{cases} 1 & \text{if } c_{ib} \leq s \text{ or } c_{ia} > s \\ 0 & \text{otherwise} \end{cases}$$

The $U_{is}^{\star}$ can be viewed as an indicator variable, which specifies whether the information contained in $Y_{is}^{\star}$ is censored. The information contained in the censored ordinal response, Y_i, is only translated into $Y_{i1}^{\star}, \ldots, Y_{ic_{ia}-1}^{\star}, Y_{ic_{ib}}^{\star}, \ldots, Y_{iS}^{\star}$, and by collapsing contiguous levels, the parameters values do not change. Therefore intuitively, we only need to used those indicator variables without censoring.

Theorem 1 *the likelihood equation for the cumulative link model in the presence of censoring can be written as*

$$S(\alpha, \beta) = \sum_{i=1}^{n} D_i^T U_i^{\star} (U_i^{\star} V_i U_i^{\star})^{-} U_i^{\star} (y_i^{\star} - F_i) = 0 \tag{5.6}$$

where $D_{ij} = \partial F / \partial(\alpha, \beta)$, $V_{ij} = cov(Y_{ij}^{\star})$, $U_i^{\star} = diag(U_{i1}^{\star}, \ldots, U_{iS-1}^{\star})$, $F_i = (F_{i1}, \ldots, F_{iS-1})^T$, *and* $F_{is} = Pr(Y_{is}^{\star} = 1)$.

Proof: Let $Y_{i0}^{*}=0$, $Y_{iS}^{*}=1$, $F_{i0}=0$, and $F_{iS}=1$. without censoring, the log-likelihood of Y_i can be written as

$$l(Y_i)=\sum_{s=0}^{S}(Y_{i,s+1}^{*}-Y_{is}^{*})\log(F_{i,s+1}-F_{is}).$$

By the theory of generalized linear model, the likelihood equation is given by

$$\sum_{i=1}^{n}D_i^T V_i^{-1}(y_i^{*}-F_i)=0.$$

Now let $U_{i0}^{*}=1$, $U_{iS}^{*}=1$. In the presence of censoring, i.e. $c_{ia}\le y_i\le c_{ib}$, the loglikelihood of Y_i can be written as

$$\begin{aligned}
l(Y_i) &= \sum_{s=0}^{S}(Y_{i,s+1}^{*}-Y_{is}^{*})U_{i,s+1}^{*}U_{is}^{*}\log(F_{i,s+1}-F_{is}) \\
&\quad +\sum_{s=0}^{S}(Y_{i,s+1}^{*}-Y_{is}^{*})(1-U_{i,s+1}^{*}U_{is}^{*})\log[\sum_{s=0}^{S}(1-U_{i,s+1}^{*}U_{is}^{*})(F_{i,s+1}-F_{is})] \\
&= \sum_{s=0}^{c_{ia}-1}(Y_{i,s+1}^{*}-Y_{is}^{*})\log(F_{i,s+1}-F_{is})+\sum_{s=c_{ib}}^{S}(Y_{i,s+1}^{*}-Y_{is}^{*})\log(F_{i,s+1}-F_{is}) \\
&\quad +(Y_{ic_{ib}}^{*}-Y_{i,c_{ia}-1}^{*})\log(F_{ic_{ib}}-F_{i,c_{ia}-1}).
\end{aligned}$$

Therefore the loglikelihood equations have the form

$$\sum_{i=1}^{n}\tilde{D}_i^T\tilde{V}_i^{-}(\tilde{y}_i^{*}-\tilde{F}_i)=0, \tag{5.7}$$

where $\tilde{y}_i^{*}=(y_{i1}^{*},\ldots,y_{i,c_{ia}-1}^{*},y_{ic_{ib}}^{*},\ldots,y_{iS}^{*})^T$, $\tilde{F}_i=(F_{i1},\ldots,F_{i,c_{ia}-1},F_{ic_{ib}},\ldots,F_{iS})^T$, $\tilde{V}_i=\mathrm{cov}(\tilde{Y}_i)$, and $\tilde{D}_i=\partial\tilde{F}_i/\partial\theta$. It is easy to see that (5.7) is the same as (5.6). This completes the proof.

For the random effects and autoregressive latent structure model we discussed in Chapter 3, the score function for the M step can be obtained by replacing the first set of equations with (5.6). For continuation ratio models and adjacent category model, this approach is not applicable, because there is not a underlying continuous latent structure for these two types of models.

5.4 Methods for Censored Continuous Responses

Let us first consider regression model on completely observed continuous data. Suppose that continuous measurements from n subjects, denoted by $Y_1, Y_2, \ldots, Y_n$, are available, and X_i is a p-dimensional covariate associated with the ith object. Suppose that Y_i is related to the covariates X_i through a linear regression: for a p-dimensional regression coefficient β,

$$Y_i = X_i'\beta + \epsilon_i,$$

where ϵ_i are independent and identically distributed residuals. Denote the distribution function of the residuals by F, and its density by f. We assume that the form of the distribution is known. For example, with normal distributed residuals, we have the general linear models. If the Y_i are exponential or Weibull, we have accelerated failure time models.

In the censored case, among the n observations, $Y_1, Y_2, \ldots, Y_m$ are observed completely, $Z_{m+1}, Z_{m+2}, \ldots, Z_n$ are only partially observed. We only know that Z_i is in the interval $C_{m+1} = (c_{ia}, c_{ib}]$, where c_{ia} and c_{ib} can be $\pm\infty$. We assume that the censoring distribution does not involve the unknown parameter β. The likelihood of the entire sample is

$$\begin{aligned} L(\beta \mid Y, C) &= \prod_{i=1}^{m} f(y_i - X_i'\beta) \prod_{i=m+1}^{n} [F(c_{ib} - X_i'\beta) - F(c_{ia} - X_i'\beta)] \\ &= \prod_{i=1}^{m} f(y_i - X_i'\beta) \prod_{i=m+1}^{n} \int_{c_{ia}-X_i'\beta}^{c_{ib}-X_i'\beta} f(t)dt \end{aligned} \tag{5.8}$$

where $Y = (Y_1, \ldots, Y_m)'$, and $C = (C_{m+1}, \ldots, C_n)$. The likelihood on the complete data has a simpler form

$$L(\beta \mid Y, Z) = \prod_{i=1}^{m} f(y_i - X_i'\beta) \prod_{i=m+1}^{n} f(z_i - X_i'\beta)$$

In the E-step, we need to calculate

$$J_i(\beta, \beta^j) = \int \log f(z_i - X_i'\beta) f(z_i \mid \beta^j, Y, C_i) dz_i$$

where

$$f(z_i \mid \beta^j, Y, C_i) = \frac{I(z_i \in C_i) f(z_i - X_i'\beta^j)}{Pr(z_i \in C_i \mid \beta^j)},$$

β^j is the current estimate of β in the iteration, and I is the indicator function. When the predictive density $f(z_i \mid \beta^j, Y, C_i)$ is easy to sample, we can draw a random sample $z_{i1}, \ldots, z_{iK}$ directly, and approximate $J_i(\beta, \beta^j)$ by

$$\tilde{J}_i(\beta, \beta^j) = \frac{1}{K} \sum_{k=1}^{K} \log f(z_{ik} - X_i'\beta).$$

When we cannot directly sample from $f(z_i \mid \beta^j, Y, C_i)$, we apply the method of importance sampling to approximate $J_i(\beta, \beta^j)$ as

(i) draw $z_{i1}, \ldots, z_{iK}$ from $d_i(z)$,

(ii) $\tilde{J}_i(\beta, \beta^j) = [\sum_{i=1}^{K} w_{ik} \log f(z_{ik} - X_i'\beta)] / \sum_{k=1}^{K} w_{ik}$

where $d_i(z)$ is a density that is easy to sample from, and $w_{ik} = f(z_i \mid \beta^j, Y, C_i)/d_i(z_{ik})$. Geweke (1989) has shown that if the support of $d_i(z)$ includes the support of $f(z_i \mid \beta^j, Y, C_i)$, the z_{ik}'s are independent identically distributed from $d_i(z)$, and $J_i(\beta, \beta^j)$ exists and is finite, then

$$\tilde{J}_i(\beta, \beta^j) \to J_i(\beta, \beta^j) \qquad \text{a.s.}$$

The first condition is sensible, for if the support of $d_i(z)$ is strictly contained in the support of $f(z_i \mid \beta^j, Y, C_i)$, then there is no hope of generating deviates in the complement of the support of $d_i(z)$. The rate of convergence depends on how closely $d_i(z)$ mimics $f(z_i \mid \beta^j, Y, C_i)$. As noted by Geweke (1989), it is important that the tails of $d_i(z)$ do not decay faster that the tails of $f(z_i \mid \beta^j, Y, C_i)$. For interval censored data, where both c_{ia} and c_{ib} are finite, one choice for $d_i(z)$ is the uniform distribution between c_{ia} and c_{ib}. The support of $d_i(z)$ naturally contains the support of $f(z_i \mid \beta^j, Y, C_i)$, and $d_i(z)$ is flat over the interval (c_{ia}, c_{ib}), so almost sure convergence is guaranteed. For left or right censored data, we can use a truncated continuous distribution, such as a normal or t distribution, as our choice for $d_i(z)$ instead of a uniform distribution. We recomend to use a distribution with heavy tail to improve the rate of convergence.

If $d_i(z)$ poorly approximates $f(z_i \mid \beta^j, Y, C_i)$, the standard error of $\tilde{J}_i(\beta, \beta^j)$ is inflated, i.e., the effective Monte Carlo sample size is decreased. In the EM algorithm, the M-step is computationally expensive, and often there is limit on how much data software can

handle. So increasing the effectiveness of the Monte Carlo sample is very important in practice. An improved version of importance sampling can be implemented as following

(i) Draw $z_{i1}^*, \ldots, z_{iK^*}^*$ from $d_i(z)$,

(ii) Draw $z_{i1}, \ldots, z_{iK}$ from the sample $z_{i1}^*, \ldots, z_{iK^*}^*$ with weight w_{ik} on z_{ik}^*,

(iii) $\tilde{J}_i(\beta, \beta^j) = 1/K \sum_{k=1}^{K} \log f(z_{ik} - X_i'\beta)$.

The sample $z_{i1}, \ldots, z_{iK}$ is an independent identically distributed sample from $f(z_i \mid \beta^j, Y, C_i)$. Smith and Gelfand (1992) showed that the approximation improves as K^* increases. $Q(\beta, \beta^j)$ is approximated by

$$Q(\beta, \beta^j) \approx \sum_{i=1}^{m} \log f(y_i - X_i'\beta) + \sum_{i=m+1}^{n} \sum_{k=1}^{K} \frac{1}{K} \log f(z_{ik} - X_i'\beta) \tag{5.9}$$

The M-step is a weighted regression on the augmented data. The Monte Carlo EM algorithm is iterated until convergence. The observed information matrix can be estimated by

$$- \frac{\partial^2 Q(\beta, \phi)}{\partial \beta^2}\bigg|_{\beta, \phi = \hat{\beta}} - \sum_{i=m+1}^{n} \left\{ \frac{1}{K} \sum_{k=1}^{K} \left(\frac{\partial \log f(z_{ik} - X_i'\beta)}{\partial \beta}\bigg|_{\beta=\hat{\beta}} \right)^2 - \left(\frac{1}{K} \sum_{k=1}^{K} \frac{\partial \log f(z_{ik} - X_i'\beta)}{\partial \beta}\bigg|_{\beta=\hat{\beta}} \right)^2 \right\} \tag{5.10}$$

Example 2. Mean and Variance

In many situations with interval censored data, especially in rounding, the data are presented as the centers of the grouped sets. For univariate data, mean and variance can be estimated using those center values, the statistics are referred as simple mean $\bar{X}^*$ and simple variance s^{*2}. Heitjan (1989) summarized two main conclusions. First, the simple mean $\bar{X}^*$ is a good estimate of $\mu = EX$, and $s^*/\sqrt{n}$ is a good summary of the uncertainty about μ in $\bar{X}^*$. Second, s^{*2} is not a good estimate of $\sigma^2 = \mathrm{var}(X)$, but can be improved by a correction.

Assuming normality, we can draw from truncated normal distributions to fill in the censored data. The maximum likelihood estimates for μ and σ^2 are obtained by calcu-

lating the weighted sample mean and sample variance using the complete data when the EM algorithm reaches convergence.

Example 3. Life tables

The life table is a summary of the survival data grouped into convenient intervals. In some applications (e.g. actuarial), the data are often collected in such a a grouped form. For example, the data are grouped into intervals $C_1, \ldots, C_L$,such that $C_l = (c_{l-1}, c_l]$ with $c_0 = 0$ and $c_L = \infty$. The life table contains the number of failures and censored survival times falling in each interval.

Suppose that $\log Y$, the logarithm of survival time Y, is related to the covariate X via a linear model $Y = X'\beta + \epsilon$, where ϵ is an error variable with known density f. There are standard estimation procedures for accelerated failure time models with complete and right censored data (see for example Kalbfleisch and Prentice (1980), Chapter 3). The method discussed in this section provides a way of fitting an accelerated failure time model with data from a life table where interval censoring is inherently present.

5.5 Discussion

We propose a general method to obtain maximum likelihood estimates when interval censoring is present. Our main goal is to introduce a simple estimating method for the regression parameters. This method can be applied to both categorical and continuous responses.

Other numerical maximization techniques, such as Newton-Raphson algorithm used by Odell et al. 1992 can also be used to deal with maximizing the observed likelihood function. The likelihood contribution for a interval censored case involves integration of the density function over C_i, in which case the first and second derivatives of log-likelihood are quite complicated. If p is large, the Newton-Raphson algorithm requires calculation and inversion of a high dimensional matrix of second derivative. Divergence is quite frequent, especially when the matrix is close to singular. We use the EM and $MCEM$ algorithms in which interval censored responses are treated as missing data. The M-step is simply a regression on the augmented data which can be accomplished with standard software to facilitate the M-step. The implementation of E-step is straightforward.

In this chapter, our focus is on fitting parametric models with censored data. There are also non-parametric and semi-parametric approaches in analyzing censored data. Turnbull (1974) proposed a version of EM for non-parametric maximum likelihood estimation of a distribution function. Finkelstein (1986) proposed a proportional hazard model for interval censored failure time data.

References

Agresti, A. (1984). *Analysis of Ordinal Categorical Data.* New York: Wiley.

Agresti, A., (1990). *Categorical Data Analysis.* New York: Wiley.

Albert, J., and Chib, S., (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88 669-679.

Akaike, H., (1973). Information theory and an extension of the maximum likelihood principle. In *2nd International Symposium on Information Theory*, 267-281.

Aitchison, J., and Silvey, S.D., (1957). The generalization of probit analysis to the case of multiple responses. *Biometrika*, 41, 131-140.

Baker, S.G., and Laird, N.M., (1988). Regression analysis for categorical variables with outcome subject to nonignorable nonresponse (Corr: V83 p1232). *Journal of the American Statistical Association*, 83, 62-69.

Begg, C. B., and Gray R., (1984). Calculation of polychotomous logistic regression parameters using individualized regressions. *Biometrika*, 71, 11-18.

Blumenthal, S., (1968). Multinomial sampling with partially categorized data *Journal of the American Statistical Association*, 63, 542-551.

Bock, R.D., and Jones, L.V., (1968). *The Measurement and Prediction of Judgement and Choice.* San Francisco: Holden-Day.

Boos, D. D., (1992). On generalized score tests. *The American Statistician*, 46, 327-333.

Brant, R., (1990). Assessing proportionality in the proportional odds model for ordinal logistic regression. *Biometrics*, 46, 1171-1178.

Breslow, N.E., Clayton, D.G., (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88, 9-25.

Brillinger, D.R., (1996). An analysis of an ordinal-valued time series. *Athens Conference on Applied Probability and Time Series, Volume II: Time Series Analysis. Lecture Notes in Statistics, Vol. 115*, 73-87.

Chambers, R.L., and Welsh, A.H., (1993). Log-linear models for survey data with non-ignorable non-response. *Journal of the Royal Statistical Society, Series B*, 55, 157-170.

Clayton, D., (1992). Repeated ordinal measurements: a generalized estimating equation approach. *Medical Research Council Biostatistics Unit Technical Report*, Cambridge, England.

Clogg, C. C., and Shihadeh, E. S., (1994). Statistical models for ordinal variables (*Advanced Quantitative Techniques in the Social Sciences Series, Vol. 4)*. Thousand Oaks, CA: SAGE Publications.

Cox, D. R., (1970). *The Analysis of Binary Data.* London: Chapman and Hall.

Cox, D. R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society, Series B.*, 74 187-220.

Dempster, A.P., Laird, N.M., and Rubin, D.B., (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39 1-22.

Diggle, P.J., Liang, K., and Zeger, S.L., (1994). *Analysis of Longitudinal Data.* Oxford: Clarendon Press.

Feinberg, S. E., (1980). *The analysis of cross-classified categorical data.* Cambridge, MA: MIT Press.

Finkelstein, D.M. (1986). A proportional hazards model for interval-censored failure time data. *Biometrics*, 42, 845-854.

Gang, S.J., Linton, K.L.P., Scott, A.J., Demets, D.L., and Klein, R. (1993). Analysis of correlated ordinal measures with ophthalmic applications. Technical Report 71, University of Wisconsin Dept. of Biostatistics.

Genter, F.C., and Farewell, V.T., (1985). Goodness-of-link testing in ordinal regression models. *The Canadian Journal of Statistics*, 13, 37-44.

Geweke, J., (1989). Bayesian inference in econometric models using Monte Carlo integration. *Econometrica*, 57, 1317-1339.

Goodman, L.A., (1983). The analysis of dependence in cross-classifications having ordered categories, using log-linear models for frequencies and log-linear models for odd. *Biometrics*, 39, 149-160.

Heagerty, P.J. and Zeger, S.L., (1996). Marginal regression models for clustered ordinal measurements. *Journal of American Statistical Association*, 91, 1024-1036.

Jarrett, R.G., Morgan, G.J.T., and Liow, S., (1981). The effects of viruses on death and deformity rates in chicken eggs. Consulting Report VT 31/37, CSIRO Division of Mathematics and Statistics, Melbourne.

Hastie, T. and Tibshirani, R (1984). Generalized additive models. Tech, rep. 98, Dept. of Statistics, Stanford University.

Hastie, T. and Tibshirani, R (1986). Generalized additive models (with discussion) *Statist. Sci.*, 1, 297-318.

Hastie, T. and Tibshirani, R (1987). Non-parametric logistic and proportional-odds regression. *Appl. Statist.*, 36, 260-276.

Hastie, T. and Tibshirani, R (1990). *Generalized Additive Models.* London: Chapman and Hall.

He, X. and Shi, P. (1994). Convergence rate of B-spline estimators of nonparametric conditional quantile functions. *J. Nonparametric Statistics.* 3, 299-308.

He, X. and Shi, P. (1996). Bivariate tensor-product B-splines in a partly linear model. *J. Multivariate Anal.*, 58, No. 2, 162-181.

Heagerty, P.J. and Zeger, S.L. (1996). Marginal regression models for clustered ordinal measurements. *Journal of American Statistical Association*, 91, 1024-1036.

Imrey, P.B., Johnson, W.D. and Koch, G.G., (1976). An incomplete contingency table approach to paired-comparison experiment. *Journal of American Statistical Association*, 71, 614-623.

Kalbfleisch, J.D., and Prentice, R.L., (1980). *The statistical analysis of failure time data.* Wiley.

Keenan, D.M., (1981). Time series analysis of binary data. *ASA Proceedings of the Business and Economic Statistics Section*, 323-328.

Kim, D.K., (1997). Regression analysis of interval-censored survival data with covariates using log-linear models. *Biometrics*, 53, 1274-1283.

Kitagawa, G., (1987). Non-Gaussian state-space modeling of nonstationary time series. *Journal of the American Statistical Association*, 82, 1032-1041.

Koch, G.G., Imrey, P.B., and Reinfurt, D.W. (1972). Linear Model Analysis of Categorical Data with Incomplete Response Vectors. *Biometrics*, 28, 663-692.

Künsch, H. R. and Stefanski, L. A. and Carroll, R. J. (1989) Conditionally unbiased bounded-influence estimation in general regression models, with applications to generalized linear models. *JASA*, 84, 460-466.

Lawless, J.F., (1982). *Statistical models and methods for lifetime data.* Wiley.

Liang, K. Y., and Zeger, S. L., (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73, 13-22.

Lindstrom, M.J., and Bates, D.M., (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics*, 46, 673-687.

MacDonald, I.L., and Zucchini, W., (1980). *Hidden Markov and other models for discrete-valued time series.* Chapman and Hall.

McCullagh, P., (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society, Series B*, 42, 109-142.

McCullagh, P., and Nelder, J., (1989). *Generalized linear models.* 2nd ed, Chapman and Hall, New York.

McPhee, D.A., Parsonson, I.M., Della-Porta, A.J., and Jarrett, R.G., (1984). Teratogenicity of Australian simbuserogroup and some other bunyaviridae viruses: The embryonated chicken egg as a model. *Infection and Immunity*, 43, 413-420.

Miller, M.E., Davis,D.D., and Landis, J.R., (1993). The analysis of longitudinal polytomous responses: generalized estimation equations and connections with weighted least squares. *Biometrics*, 49, 1033-1044.

Molenberghs, G., and Goetghebeur, E., (1997). Simple fitting algorithms for incomplete categorical data. *Journal of the Royal Statistical Society, Series B.*, 59, 401-414.

Morgan, B.J.T., (1992). *Analysis of Quantal Response Data.* London: Chapman and Hall.

Odell, P.M., Anderson, K.M., and D'Augustino, R.B., (1992). Maximum likelihood estimation for interval-censored data using a Weibull-based accelerated failure time model. *Biometrics*, 48, 951-959.

Peterson, B. and Harrell, F.E., (1990). Partial proportional odds models for ordinal response variables. *Appl. Statist.*, 39, No.2, 205-217.

Prentice, R.L., (1988). Correlated binary regression with covariates specific to each binary observation, *Biometrics*, 44, 1033-1048.

Prentice, R.L., and Gloeckler, L.A.(1978). Regression analysis of grouped survival data with application to breast cancer data, *Biometrics*, 34, 57-67.

Rao, C.R., (1948). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Proceedings of the Cambridge Philosophical Society*, 44, 50-57.

Schall, R., (1991). Estimation in generalized linear models with random effects. *Biometrika*, 78, 719-727.

Simpson, D.G., Carroll, R. J., Xie, M. and Guth, D. L. (1996a). Interval censoring and marginal analysis in ordinal regression. *Journal of Agricultural, biological and Environmental Statistics*, 1, 354-376.

Simpson, D.G., Carroll, R.J., Xie, M. and Guth, D.L. (1996b). Weighted logistic regression and robust analysis of diverse toxicology data. *Communications in Statistics*. Theory and Method, 25, 2615-2632.

Srole, L., Langner, T.S., Michael, S.T., Opler, M.K., and Rennie, T.A.C., (1962). *Mental Health in the Metropolis: The Midtown Manhattan Study.* New York: McGraw-Hill.

Stiratelli, R., Laird, N., and Ware, J.H., (1984). Random-effects models for serial observations with binary response. *Biometrics*, 40, 961-971.

Tanner, M.A., (1996). *Tools for statistical inference. Methods for the exploration of posterior distributions and likelihood functions.* Springer-Verlag.

Thompson, W.A., (1977). On the treatment of grouped observations in life studies. *Biometrics*, 33, 463-470.

Turnbull, B., (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of the Royal Statistical Society, Series B*, 38, 290-295.

Uesaka, H. and Asano, C. (1987). Latent scale linear models for multivariate ordinal responses and their analysis by the method of weighted least squares, *Ann. Inst. Statist. Math.*, 39, 191-210.

White, H., (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50, 1-26.

Williamson, J.M., Kim, K.M., and Lipsitz, S.L. (1995), Analyzing bivariate ordinal data using a global odds ratio. *J. American Statistical Association*, 90, 1432-1437.

Xie, M., and Simpson D.G., (1999). Regression modeling of ordinal data with nonzero baselines. *Biometrics*, 55, 308-316.

Xie, M., Simpson, D.G. and Carroll, R.J., (1997). Scaled link functions for heterogeneous ordinal response data. In *Modeling Longitudinal and Spatially Correlated Data: Methods, Applications and Future Directions.* Springer-Verlag, Lecture Notes in Statistics, p. 23-26.

Xie, M., Simpson, D.G. and Carroll, R.J., (1997). Random effects in interval-censored ordinal regression: latent structure and Bayesian approach. *Biometrics*, In Press

Yee, T.W. and Wild, C.J. (1996). Vector generalized additive models. *J. R. Statist. Soc. B.*, 58, No.3, 481-493.

Zeger, S.L., and Liang, K. Y., (1986). Longitudinal data analysis for discrete and continuous outcomes, *Biometrics*, 42, 121-130.

Zeger, S.L., and Karim, M.R., (1991). Generalized linear models with random effects: A Gibbs sampling approach *Journal of the American Statistical Association*, 86, 79-86.

Zhao, L.P., and Prentice, R.L., (1990). Correlated binary regression using a quadratic exponential model. *Biometrika*, 77, 642-648.

Vita

LIMIN FU

Education

- Ph.D. in Statistics, University of Illinois at Urbana-Champaign, 01/00
- M.A. in Mathematics, Eastern Illinois University, 07/95
- B.S. in Applied Mathematics, Beijing Institute of Technology, 07/91

Work Experience

Intern, 06/99-08/99

Freddie Mac, Risk Assessment and Model Development Division.

- Evaluated and improved the repeat-sales model for forecasting home price index.

Graduate Consultant, 06/98-12/99

Illinois Statistics Office - Statistical Consulting

- Provided statistical consulting services, including designing experiments, constructing survey plans, analyzing data, using computers for statistical computations.

Research Assistant, 01/97-06/97, 07/96-12/99

University of Illinois, Children & Family Research Center

- Managed large data sets in Sybase.
- provided statistical analysis using SAS.

University of Illinois, Department of Statistics

- Developed statistical software in S-Plus and Fortran, including non parametric procedures and categorical regression procedures.

Teaching/Computer Assistant, 09/95-07/96, 08/94-08/95

University of Illinois, Department of Statistics

- Taught undergraduate algebra class.

Eastern Illinois University, Department of Mathematics

- Taught undergraduate statistics class.
- Worked as an computer lab instructor.

DBA / System Engineer / Program Analyst, 10/93-07/94, 08/91-10/93

Informix Software Inc., Beijing Office, China

- Taught customer training sessions.
- Provided consulting services on Informix RDBMS.

the Sixth National Research Institute, China

- Designed and developed management information systems.

Publications

- Travelling waves for reaction diffusion equations. *Journal of Partial Differential Equations*, 10 (1997), 149-157.
- Unified ordinal regression analysis via generalized estimating equations and generalized score tests. (With D. Simpson). *Journal of Statistical Planning and Inference*. In revision.
- Latent structure models for correlated ordinal data. (With D. Simpson). In preparation.

Professional Organization Membership

- American Statistical Association.