

A Model for Multivariate Binary Data with Covariates
Based on Compatible Conditionally Specified Logistic Regressions

by

Ying Liu

B.Sc., Hangzhou Normal University, 1985

M.Sc., Shanghai University of Finance and Economics, 1991

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

Department of Statistics

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

August 1994

©Ying Liu, 1994

In presenting this thesis in partial fulfillment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Statistics

The University of British Columbia
Vancouver, Canada

Date Aug. 30, 1994

Abstract

Rather than construction of a multivariate distribution from given univariate or bivariate margins, recently several papers seek to promote the development and usage of a simple but relatively unknown approach to the specification of models for dependent binary outcomes through conditional probabilities, each of which is assumed to be logistic. These recent proposals were all offered as heuristic approaches to specifying a multivariate distribution capable of representing the dependence of binary outcomes. However, they are limited in scope, for they all describe some special patterns of dependence. This thesis is concerned with a model for a multivariate binary response with covariates based on compatible conditionally specified logistic regressions. With this model, we allow for a general dependence structure for the binary outcomes.

Three likelihood-based computing methods are introduced to estimate the parameters in our model. An example on the coronary bypass surgery is presented for illustration.

Contents

Abstract	ii
Table of Contents	iii
List of Tables	v
Acknowledgement	vii
1 Introduction	1
2 Compatibility Conditions for Conditionally Specified Multivariate Distributions	5
2.1 Compatibility in 2 Dimensions	6
2.2 Compatibility in 3 Dimensions	7
3 Models for Multivariate Binary Responses with Covariates	9
3.1 Conditions for Compatibility in 2 or 3 Dimensions	10
3.1.1 2 Dimensional Case	10
3.1.2 3 Dimensional Case	14
3.2 General Case	20
3.3 Further Extension	23
4 Comparisons with Other Models	26
4.1 Regressive Logistic Models	26

4.2	Conditional Logistic Regression Models for Exchangeable Binary Data . .	28
5	Data Analysis and Computing	30
5.1	The Data Source and Description	30
5.2	Estimation Procedure	31
5.3	Comparison of Three Methods	35
5.4	Selection of Covariates	37
5.4.1	Odds Ratio Analysis	45
5.4.2	Univariate Analysis and Comparison of Models	46
5.5	Other Issues	50
5.5.1	General	50
5.6	Examples	52
5.6.1	Example 1	52
5.6.2	Example 2	57
	Appendix: C Program for Applying the Newton-Raphson Method	62

List of Tables

5.1	Description for the Subset of MCR Data	32
5.2	Description for the Subset of MCR Data (continued)	33
5.3	Results from Separate Logistic Regressions with Response Variables REM, NEM	34
5.4	Method Comparison under Case 1	37
5.5	Method Comparison under Case 2	38
5.6	Method Comparison under Case 3	38
5.7	Method Comparison under Case 4	39
5.8	Method Comparison under Case 5	40
5.9	Method Comparison under Case 5 (continued)	41
5.10	Method Comparison under Case 6	42
5.11	Method Comparison under Case 6 (continued)	43
5.12	Method Comparison under Case 6 (continued)	44
5.13	Odds Ratio for REM	47
5.14	Odds Ratio for NEM	48
5.15	Odds Ratio for PUM	49
5.16	The Association of REM (V21) and NEM (V22)	53
5.17	CI of Odds Ratio for REM (V21) and NEM (V22)	53
5.18	The Estimates of the Coefficients under All Five Cases	55
5.19	The Estimates of the Coefficients under All Five Cases (continued)	56

5.20 The Log-likelihood Functions with Different Number of Covariates . . . 57

5.21 The Log-likelihood Functions of Three Response Variables with Different
Number of Covariates 58

5.22 The Estimates of the Coefficients under the Third Case 59

5.23 The Marginal Odds Ratio for REM and NEM 60

Acknowledgements

I would like to thank my supervisor, Dr. Harry Joe for his guidance, help and support throughout the development of this thesis. I would also like to thank Dr. Martin Puterman for his careful reading of the manuscript and valuable comments. Finally, I must acknowledge the support of UBC Department of Statistics during my Masters program.

Chapter 1

Introduction

This thesis is concerned with a model for a multivariate binary response with covariates. The model is based on compatible logistic regressions. Association of outcomes may be a factor of primary or secondary interest in models for binary data. When measures of association are of interest, models for the marginal expectation may not be appropriate. Rather than construction of a multivariate distribution from given univariate or bivariate margins, recently several papers seek to promote the development and usage of a simple but relatively unknown approach to the specification of models for dependent binary outcomes through conditional probabilities. For binary response data, logistic regression models are useful for modelling the conditional probabilities.

Practically speaking, we frequently encounter the variety of circumstances in which several binary response variables are measured on associated explanatory variables. In such situations, interest may centre on modelling the multivariate response for analysis. Take two examples for illustration. The first example on coronary bypass surgery will be used in the thesis later. This example provides the pre-operation information such as patient's age, gender, etc. and the post-operation information such as patient's complications and survival status which are all given in the form of binary data. Since

we are interested in those outcome variables indicating the patient's complications and survival status after the surgery, we have a situation in which the analysis of a multivariate binary response with covariates is needed. A second example concerns the analysis of responses, college plans (Y_1) and parental encouragement (Y_2) and the covariates, IQ (X_1), sex (X_2), and socioeconomic status (X_3), where Y_i is 1 for yes, and 0 for no, $i = 1, 2$ (see Bonney, 1987). For this example, building a model for a multivariate binary response with covariates might be of interest.

Recently, regressive logistic models have been proposed when the data comprise sequences of binary outcomes possibly with measures on associated explanatory variables (Bonney, 1987). Regressive logistic models are based on the elementary decomposition of the joint probability so that the likelihood of a set of binary dependent outcomes, with or without explanatory variables, is expressed as a product of successive conditional probabilities each of which is assumed to be univariate logistic. In Bonney's models, each binary response variable is an explanatory variable for later response variables in the sequence. Another proposal (Rosner, 1984) is the polychotomous logistic regression model. In this model, Rosner converts the problem into a 2^n polychotomous logistic distribution for the number of successes when explanatory variables are zero. As a motivation for Rosner's model, Connolly and Liang (1988) describe a class of conditional logistic regression models for correlated binary data. This includes the polychotomous logistic model of Rosner as a special case.

These recent proposals were all offered as heuristic approaches for specifying a multivariate distribution capable of representing the dependence of binary outcomes. However, they are limited in scope, for they all describe some special patterns of dependence. Regressive logistic models proposed by Bonney deal with sequential binary outcomes, whereas Connolly and Liang's models might be useful when the dependent binary data

are almost exchangeable such as in the familial data situation.

In order to allow for a general dependence structure for the binary outcomes, we introduce a model for multivariate binary data with covariates based on compatible conditionally specified logistic regressions. Conditionally, each binary response variable is a covariate for the logistic regressions for the other binary response variables. Moreover, there are conditions on the regression parameters in order that these conditional distributions are compatible. Furthermore, we can initially use logistic regressions to fit the conditional probabilities for each binary outcome, given the remaining binary outcomes in addition to covariates, and use this to assess whether the multivariate model is reasonable.

Two different types of compatibility conditions are introduced in Chapter 2. One necessary and sufficient compatibility condition is proposed by Arnold and Strauss (1988); another compatibility condition comes from extending a result from Gelman and Speed (1993).

The focus of Chapter 3 is to discuss general conditions under which there exists a joint probability distribution with the given specified conditional probability distributions and to develop a model for multivariate binary data with covariates based on compatible conditionally specified logistic regression. Starting with 2 and 3 dimensions enlightens us on the conclusion of the general case (from Section 3.1 to Section 3.2). In Section 3.3, we briefly discuss what are conditions for compatibility when the conditional distributions are based on a distribution that is not logistic.

In Chapter 4, we compare our model with regressive logistic models proposed by Bonney and conditional logistic regression models for correlated binary data proposed by Connolly and Liang.

In Chapter 5, an example with the coronary bypass surgery data is used for illustration of our model. In Section 5.1, we describe the data source in the example. In Section 5.2, we introduce three computing methods which are all likelihood-based and are all used for estimating the parameters in our model. For our example, comparison of three methods is done in Section 5.3. While doing data analysis and computing among other things, answers to the following questions are sought (Section 5.4 and Section 5.5):

1. For our example, how do we choose the covariates to use in our model?
2. How much does the association of the response variables depend on the covariates?

Chapter 2

Compatibility Conditions for Conditionally Specified Multivariate Distributions

The goal of this chapter is to study conditions for compatibility of conditional probabilities; that is, conditions for which a given set of conditional probability functions leads to a proper joint multivariate distribution. This will be made clear through some examples, starting with the two dimensional case.

Let (X, Y) be a two dimensional random vector. Clearly, its probabilistic behavior for most purposes is adequately specified by knowledge of its joint cumulative distribution function

$$F_{X,Y}(x, y) = \Pr(X \leq x, Y \leq y); x, y \in \mathbf{R}$$

or joint density function $f_{X,Y}(x, y)$.

A variety of transforms can be used to characterize $F_{X,Y}$. It is clear that one marginal distribution and the family of corresponding conditional distributions, i.e., knowledge of $F_Y(y)$ and $F_{X|Y}(x|y) = \Pr(X \leq x|Y = y)$, for every y , will completely determine the

joint distribution of (X, Y) . On the other hand, knowledge of the marginal distributions $F_X(x)$ and $F_Y(y)$ has long been known to be inadequate to determine $F_{X,Y}(x, y)$. What if we are given both families of conditional distributions, $F_{X|Y}(x|y)$ for every possible value y of Y and $F_{Y|X}(y|x)$ for every possible value x of X ? Provided the families of conditional distributions are compatible, in a sense that there exists a valid joint distribution with the given conditional distributions, then indeed these families of conditional distributions will determine the joint distribution of (X, Y) .

2.1 Compatibility in 2 Dimensions

The joint, marginal and conditional densities of X and Y will be denoted by $f_{X,Y}(x, y)$, $f_X(x)$, $f_Y(y)$, $f_{X|Y}(x|y)$ and $f_{Y|X}(y|x)$. One compatibility condition proposed by Arnold and Strauss (1988) is that conditional densities $f_{X|Y}(x|y)$ and $f_{Y|X}(y|x)$ are compatible if and only if there exist nonnegative functions $a(x)$ and $b(y)$ such that

$$a(x)f_{Y|X}(y|x) = b(y)f_{X|Y}(x|y)$$

for all (x, y) .

Another compatibility condition is given in Gelman and Speed (1993). Suppose the conditional densities $f_{X|Y}(\cdot|y)$ are given for all y and $f_{Y|X}(\cdot|x_0)$ is given for a particular value x_0 . Assume that $f_{X|Y}(x_0|y) > 0$ for all y . Then for the joint density $f_{X,Y}(x, y)$,

$$f_{X,Y}(x, y) \propto \frac{f_{X|Y}(x|y)f_{Y|X}(y|x_0)}{f_{X|Y}(x_0|y)} \quad (2.1)$$

with the proportionality constant equal to

$$f_X(x_0) = \left\{ \int \int \frac{f_{X|Y}(x|y)f_{Y|X}(y|x_0)}{f_{X|Y}(x_0|y)} dx dy \right\}^{-1}.$$

That is, for two random variables, the set of conditional densities given one variable plus one conditional density given the other variable determine the joint bivariate distribution.

If the conditional densities $f_{X|Y}(\cdot|y)$ are given for all y and $f_{Y|X}(\cdot|x)$ are given for all x , then, Gelman and Speed stated that a condition for compatibility, derived from (2.1) is that:

$$\frac{f_{X|Y}(x|y)f_{Y|X}(y|x_1)}{f_{X|Y}(x_1|y)} / \frac{f_{X|Y}(x|y)f_{Y|X}(y|x_2)}{f_{X|Y}(x_2|y)} = \frac{f_{X|Y}(x_2|y)f_{Y|X}(y|x_1)}{f_{X|Y}(x_1|y)f_{Y|X}(y|x_2)} \quad (2.2)$$

does not depend on x, y for all choices of $x_1 \neq x_2$.

2.2 Compatibility in 3 Dimensions

Similarly, a three dimensional random vector (X, Y, Z) might be specified by giving the distributions of X given (Y, Z) , Y given (X, Z) and Z given (X, Y) . Such conditional specifications must be checked for compatibility in a manner analogous to that described in section 2.1.

On the basis of the idea provided by Arnold and Strauss, the compatibility condition for a three dimensional random vector (X, Y, Z) is that conditional densities $f_{X|Y,Z}(x|y, z)$, $f_{Y|X,Z}(y|x, z)$ and $f_{Z|X,Y}(z|x, y)$ are compatible if and only if there exist nonnegative functions $a(y, z)$, $b(x, z)$ and $c(x, y)$ such that

$$a(y, z)f_{X|Y,Z}(x|y, z) = b(x, z)f_{Y|X,Z}(y|x, z) = c(x, y)f_{Z|X,Y}(z|x, y)$$

for all (x, y, z) (see Arnold, Castillo and Sarabia (1992), Chapter 8).

The trivariate extension of Gelman and Speed's compatibility condition is as follows. Suppose the conditional densities $f_{X|Y,Z}(\cdot|y, z)$ are given for all y, z , $f_{Y|X,Z}(\cdot|x_0, z)$ are given for all z and a fixed x_0 , and $f_{Z|X,Y}(\cdot|x_0, y_0)$ is given for some fixed y_0 and the same x_0 as for $f_{Y|X,Z}(\cdot|x_0, z)$. Assume that $f_{X|Y,Z}(x_0|y, z) > 0$ for all y, z and $f_{Y|X,Z}(y_0|x_0, z) > 0$ for all z , then for the joint density $f_{X,Y,Z}(x, y, z)$,

$$f_{X,Y,Z}(x, y, z) \propto \frac{f_{X|Y,Z}(x|y, z)f_{Y|X,Z}(y|x_0, z)f_{Z|X,Y}(z|x_0, y_0)}{f_{X|Y,Z}(x_0|y, z)f_{Y|X,Z}(y_0|x_0, z)} \quad (2.3)$$

with the proportionality constant equal to $f_{X,Y}(x_0, y_0)$. If the conditional densities $f_{X|Y,Z}(\cdot|y, z)$ are given for all y, z , $f_{Y|X,Z}(\cdot|x, z)$ are given for all x, z , and $f_{Z|X,Y}(\cdot|x, y)$ are given for all x, y , then a condition for compatibility is that

$$\frac{f_{X|Y,Z}(x_2|y, z)f_{Y|X,Z}(y_2|x_2, z)f_{Y|X,Z}(y|x_1, z)f_{Z|X,Y}(z|x_1, y_1)}{f_{X|Y,Z}(x_1|y, z)f_{Y|X,Z}(y_1|x_1, z)f_{Y|X,Z}(y|x_2, z)f_{Z|X,Y}(z|x_2, y_2)} \quad (2.4)$$

does not depend on x, y, z for all choices of $(x_1, y_1) \neq (x_2, y_2)$.

Cases involving random variables of dimension greater than three can, of course, be considered. The general ideas, however, are clearly visible in the 2 and 3 dimensional settings and both concepts for the compatibility condition can be extended to higher dimensions.

Chapter 3

Models for Multivariate Binary Responses with Covariates

The ideas in the previous chapter will be applied to get a model of compatible logistic regressions for a multivariate binary response. Consider a set of n correlated binary variables $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)$, where Y_i is coded 1 and 0, and m covariate variables $\mathbf{X} = (X_1, X_2, \dots, X_m)$. Suppose that we are given the conditional probability distributions $\Pr(Y_i|Y_j, j \neq i, \mathbf{X})$ for $i = 1, 2, \dots, n$. From previous discussions, the joint probability distribution

$$\Pr(\mathbf{Y}|\mathbf{X}) = \Pr(Y_1, Y_2, \dots, Y_n|\mathbf{X})$$

will be determined if these conditional probability distributions are compatible.

With binary responses, the logit link function is a natural choice although, in principle, any link function could be chosen. In our study, we assume that all conditional distribution of Y_i given $Y_j, j \neq i$, and \mathbf{X} are logistic. That is, for a given binary response variable, the remaining binary response variables are used as covariates in addition to \mathbf{X} . Now our goal is to discuss general conditions under which there exists a joint probability distribution $\Pr(\mathbf{Y}|\mathbf{X})$ with the given conditional probability distributions. At the

same time, a model for multivariate binary data with covariates based on compatible conditionally specified logistic regressions can be constructed.

3.1 Conditions for Compatibility in 2 or 3 Dimensions

The two approaches of Arnold & Strauss and Gelman & Speed are illustrated in this section.

3.1.1 2 Dimensional Case

For ease of exposition, consider first the case of two binary response variables Y_1 and Y_2 , and a covariate vector \mathbf{x} . Suppose that Y_1 conditional on \mathbf{x} and $Y_2 = y_2$ is logit and that Y_2 conditional on \mathbf{x} and $Y_1 = y_1$ is logit, that is,

$$\begin{aligned} \text{logit}[\Pr(Y_1 = 1|y_2, \mathbf{x})] &= \log_e \frac{\Pr(Y_1 = 1|y_2, \mathbf{x})}{\Pr(Y_1 = 0|y_2, \mathbf{x})} \\ &= \alpha_1 + \beta_1 \mathbf{x} + \gamma_{12} y_2, \\ \text{logit}[\Pr(Y_2 = 1|y_1, \mathbf{x})] &= \log_e \frac{\Pr(Y_2 = 1|y_1, \mathbf{x})}{\Pr(Y_2 = 0|y_1, \mathbf{x})} \\ &= \alpha_2 + \beta_2 \mathbf{x} + \gamma_{21} y_1, \end{aligned}$$

with parameters $\alpha_1, \beta_1, \alpha_2, \beta_2, \gamma_{12}$, and γ_{21} .

After some rearrangement,

$$\Pr(Y_1 = y_1|y_2, \mathbf{x}) = \frac{\exp[(\alpha_1 + \beta_1 \mathbf{x} + \gamma_{12} y_2) y_1]}{1 + \exp(\alpha_1 + \beta_1 \mathbf{x} + \gamma_{12} y_2)}$$

and

$$\Pr(Y_2 = y_2|y_1, \mathbf{x}) = \frac{\exp[(\alpha_2 + \beta_2 \mathbf{x} + \gamma_{21} y_1) y_2]}{1 + \exp(\alpha_2 + \beta_2 \mathbf{x} + \gamma_{21} y_1)}. \quad (3.1)$$

(Note that for convenience, the following shorthand notation $\Pr(y_1|y_2, \mathbf{x})$ is used for $\Pr(Y_1 = y_1|Y_2 = y_2, \mathbf{X} = \mathbf{x})$, etc; when the context is clear.)

To ensure compatibility for the conditional probability distributions $\Pr(y_1|y_2, \mathbf{x})$ and $\Pr(y_2|y_1, \mathbf{x})$, we have the following theorem.

Theorem 3.1 *The conditional probability distributions $\Pr(y_1|y_2, \mathbf{x})$ and $\Pr(y_2|y_1, \mathbf{x})$ in (3.1) are compatible if and only if $\gamma_{12} = \gamma_{21}$, in which case, the joint probability distribution is:*

$$\Pr(y_1, y_2|\mathbf{x}) = \frac{1}{c} \exp[(\alpha_1 + \beta_1\mathbf{x})y_1 + (\alpha_2 + \beta_2\mathbf{x})y_2 + \gamma_{12}y_1y_2], \quad (3.2)$$

where $c = 1 + \exp(\alpha_1 + \beta_1\mathbf{x}) + \exp(\alpha_2 + \beta_2\mathbf{x}) + \exp[(\alpha_1 + \alpha_2) + (\beta_1 + \beta_2)\mathbf{x} + \gamma_{12}]$.

Proof 1: (using Arnold and Strauss's compatibility condition)

$\Pr(y_1|y_2, \mathbf{x})$ and $\Pr(y_2|y_1, \mathbf{x})$ are compatible if and only if there exist probabilities $\Pr(y_1, y_2|\mathbf{x})$, $\Pr(y_2|\mathbf{x})$ and $\Pr(y_1|\mathbf{x})$ such that:

$$\begin{aligned} \Pr(y_1, y_2|\mathbf{x}) &= \Pr(y_1|y_2, \mathbf{x}) \Pr(y_2|\mathbf{x}) \\ &= \Pr(y_2|y_1, \mathbf{x}) \Pr(y_1|\mathbf{x}). \end{aligned}$$

(see Arnold, Castillo and Sarabia (1992)).

Equivalently:

$$\frac{\Pr(Y_1 = 1|Y_2 = 1, \mathbf{x})}{\Pr(Y_2 = 1|Y_1 = 1, \mathbf{x})} = \frac{\Pr(Y_1 = 1|\mathbf{x})}{\Pr(Y_2 = 1|\mathbf{x})}, \quad (3.3)$$

$$\frac{\Pr(Y_1 = 1|Y_2 = 0, \mathbf{x})}{\Pr(Y_2 = 0|Y_1 = 1, \mathbf{x})} = \frac{\Pr(Y_1 = 1|\mathbf{x})}{\Pr(Y_2 = 0|\mathbf{x})}, \quad (3.4)$$

$$\frac{\Pr(Y_1 = 0|Y_2 = 1, \mathbf{x})}{\Pr(Y_2 = 1|Y_1 = 0, \mathbf{x})} = \frac{\Pr(Y_1 = 0|\mathbf{x})}{\Pr(Y_2 = 1|\mathbf{x})}, \quad (3.5)$$

$$\frac{\Pr(Y_1 = 0|Y_2 = 0, \mathbf{x})}{\Pr(Y_2 = 0|Y_1 = 0, \mathbf{x})} = \frac{\Pr(Y_1 = 0|\mathbf{x})}{\Pr(Y_2 = 0|\mathbf{x})}. \quad (3.6)$$

For example, dividing (3.3) by (3.4), and dividing (3.5) by (3.6), we have following equations respectively:

$$\frac{\Pr(Y_2 = 0|\mathbf{x})}{\Pr(Y_2 = 1|\mathbf{x})} = \frac{\Pr(Y_1 = 1|Y_2 = 1, \mathbf{x})}{\Pr(Y_2 = 1|Y_1 = 1, \mathbf{x})} \cdot \frac{\Pr(Y_2 = 0|Y_1 = 1, \mathbf{x})}{\Pr(Y_1 = 1|Y_2 = 0, \mathbf{x})}$$

$$\begin{aligned}
&= \frac{\exp(\alpha_1 + \beta_1 \mathbf{x} + \gamma_{12})}{1 + \exp(\alpha_1 + \beta_1 \mathbf{x} + \gamma_{12})} \cdot \frac{1 + \exp(\alpha_2 + \beta_2 \mathbf{x} + \gamma_{21})}{\exp(\alpha_2 + \beta_2 \mathbf{x} + \gamma_{21})} \\
&\quad \cdot \frac{1}{1 + \exp(\alpha_2 + \beta_2 \mathbf{x} + \gamma_{21})} \cdot \frac{1 + \exp(\alpha_1 + \beta_1 \mathbf{x})}{\exp(\alpha_1 + \beta_1 \mathbf{x})} \\
&= \frac{\exp(\gamma_{12}) \cdot [1 + \exp(\alpha_1 + \beta_1 \mathbf{x})]}{\exp(\gamma_{21}) \cdot \exp(\alpha_2 + \beta_2 \mathbf{x}) \cdot [1 + \exp(\alpha_1 + \beta_1 \mathbf{x} + \gamma_{12})]}, \quad (3.7)
\end{aligned}$$

$$\begin{aligned}
\frac{\Pr(Y_2 = 0|\mathbf{x})}{\Pr(Y_2 = 1|\mathbf{x})} &= \frac{\Pr(Y_1 = 0|Y_2 = 1, \mathbf{x})}{\Pr(Y_2 = 1|Y_1 = 0, \mathbf{x})} \cdot \frac{\Pr(Y_2 = 0|Y_1 = 0, \mathbf{x})}{\Pr(Y_1 = 0|Y_2 = 0, \mathbf{x})} \\
&= \frac{1}{1 + \exp(\alpha_1 + \beta_1 \mathbf{x} + \gamma_{12})} \cdot \frac{1 + \exp(\alpha_2 + \beta_2 \mathbf{x})}{\exp(\alpha_2 + \beta_2 \mathbf{x})} \\
&\quad \cdot \frac{1 + \exp(\alpha_1 + \beta_1 \mathbf{x})}{1 + \exp(\alpha_2 + \beta_2 \mathbf{x})} \\
&= \frac{1 + \exp(\alpha_1 + \beta_1 \mathbf{x})}{\exp(\alpha_2 + \beta_2 \mathbf{x}) \cdot [1 + \exp(\alpha_1 + \beta_1 \mathbf{x} + \gamma_{12})]}. \quad (3.8)
\end{aligned}$$

Comparing (3.7) with (3.8), only the condition $\gamma_{12}=\gamma_{21}$ makes them consistent. Other cases yield the same result. Therefore, we conclude that a necessary and sufficient condition for $\Pr(y_1|y_2, \mathbf{x})$ and $\Pr(y_2|y_1, \mathbf{x})$ to be compatible conditional distributions is $\gamma_{12}=\gamma_{21}$.

Since $\Pr(Y_2 = 0|\mathbf{x}) = 1 - \Pr(Y_2 = 1|\mathbf{x})$, solving (3.8) we have

$$\Pr(Y_2 = 1|\mathbf{x}) = \frac{\exp(\alpha_2 + \beta_2 \mathbf{x}) \cdot [1 + \exp(\alpha_1 + \beta_1 \mathbf{x})]}{1 + \exp(\alpha_1 + \beta_1 \mathbf{x}) + \exp(\alpha_2 + \beta_2 \mathbf{x}) + \exp[(\alpha_1 + \alpha_2) + (\beta_1 + \beta_2)\mathbf{x} + \gamma_{12}]}.$$

Similarly, we also can get

$$\Pr(Y_1 = 1|\mathbf{x}) = \frac{\exp(\alpha_1 + \beta_1 \mathbf{x}) \cdot [1 + \exp(\alpha_2 + \beta_2 \mathbf{x})]}{1 + \exp(\alpha_1 + \beta_1 \mathbf{x}) + \exp(\alpha_2 + \beta_2 \mathbf{x}) + \exp[(\alpha_1 + \alpha_2) + (\beta_1 + \beta_2)\mathbf{x} + \gamma_{12}]}.$$

Summarily,

$$\begin{aligned}
\Pr(Y_1 = y_1|\mathbf{x}) &= \frac{1}{c} \exp[(\alpha_1 + \beta_1 \mathbf{x})y_1][1 + \exp(\alpha_2 + \beta_2 \mathbf{x} + \gamma_{12}y_1)], \\
\Pr(Y_2 = y_2|\mathbf{x}) &= \frac{1}{c} \exp[(\alpha_2 + \beta_2 \mathbf{x})y_2][1 + \exp(\alpha_1 + \beta_1 \mathbf{x} + \gamma_{12}y_2)]. \quad (3.9)
\end{aligned}$$

where $c = 1 + \exp(\alpha_1 + \beta_1 \mathbf{x}) + \exp(\alpha_2 + \beta_2 \mathbf{x}) + \exp[(\alpha_1 + \alpha_2) + (\beta_1 + \beta_2) \mathbf{x} + \gamma_{12}]$.

Using (3.1) and (3.9), it is easy to prove that the joint probability distribution is (3.2). ■

Proof 2: (using Gelman and Speed's compatibility condition)

Assume that $\Pr(y_1' | y_2, \mathbf{x}), \Pr(y_1'' | y_2, \mathbf{x}) > 0$ for all y_2, \mathbf{x} and two fixed values y_1', y_1'' .

Implied by (2.2), a compatibility condition for $\Pr(y_1 | y_2, \mathbf{x})$ and $\Pr(y_2 | y_1, \mathbf{x})$ is that

$$\frac{\Pr(y_1 | y_2, \mathbf{x}) \Pr(y_2 | y_1', \mathbf{x})}{\Pr(y_1' | y_2, \mathbf{x})} / \frac{\Pr(y_1 | y_2, \mathbf{x}) \Pr(y_2 | y_1'', \mathbf{x})}{\Pr(y_1'' | y_2, \mathbf{x})} = \frac{\Pr(y_1'' | y_2, \mathbf{x}) \Pr(y_2 | y_1', \mathbf{x})}{\Pr(y_1' | y_2, \mathbf{x}) \Pr(y_2 | y_1'', \mathbf{x})} \quad (3.10)$$

does not depend on y_1, y_2 for all choices of $y_1' \neq y_1''$.

Using (3.1), (3.10) can be rewritten as follows:

$$\begin{aligned} \frac{\Pr(y_1'' | y_2, \mathbf{x}) \Pr(y_2 | y_1', \mathbf{x})}{\Pr(y_1' | y_2, \mathbf{x}) \Pr(y_2 | y_1'', \mathbf{x})} &= \left(\frac{\exp[(\alpha_1 + \beta_1 \mathbf{x} + \gamma_{12} y_2) y_1'']}{1 + \exp(\alpha_1 + \beta_1 \mathbf{x} + \gamma_{12} y_2)} \cdot \frac{\exp[(\alpha_2 + \beta_2 \mathbf{x} + \gamma_{21} y_1') y_2]}{1 + \exp(\alpha_2 + \beta_2 \mathbf{x} + \gamma_{21} y_1')} \right) / \\ &\quad \left(\frac{\exp[(\alpha_1 + \beta_1 \mathbf{x} + \gamma_{12} y_2) y_1']}{1 + \exp(\alpha_1 + \beta_1 \mathbf{x} + \gamma_{12} y_2)} \cdot \frac{\exp[(\alpha_2 + \beta_2 \mathbf{x} + \gamma_{21} y_1'') y_2]}{1 + \exp(\alpha_2 + \beta_2 \mathbf{x} + \gamma_{21} y_1'')} \right) \\ &= \frac{1 + \exp(\alpha_2 + \beta_2 \mathbf{x} + \gamma_{21} y_1')}{1 + \exp(\alpha_2 + \beta_2 \mathbf{x} + \gamma_{21} y_1'')} \cdot \exp[(\alpha_1 + \beta_1 \mathbf{x})(y_1'' - y_1')] \\ &\quad \cdot \exp[y_2(y_1'' - y_1')(\gamma_{12} - \gamma_{21})] \end{aligned} \quad (3.11)$$

With $y_1'' = 1, y_1' = 0$, (3.11) does not depend on y_1 and y_2 if and only if $\gamma_{12} = \gamma_{21}$. That is, a necessary condition for $\Pr(y_1 | y_2, \mathbf{x})$ and $\Pr(y_2 | y_1, \mathbf{x})$ to be compatible is $\gamma_{12} = \gamma_{21}$.

Letting $y_1' = 0$, from (2.1), (2.2) and (3.1), we get that the joint probability distribution $\Pr(y_1, y_2 | \mathbf{x})$ has the following property:

$$\begin{aligned} \Pr(y_1, y_2 | \mathbf{x}) &\propto \frac{\Pr(y_1 | y_2, \mathbf{x}) \Pr(y_2 | y_1 = 0, \mathbf{x})}{\Pr(y_1 = 0 | y_2, \mathbf{x})} \\ &= \frac{\exp[(\alpha_1 + \beta_1 \mathbf{x}) y_1 + (\alpha_2 + \beta_2 \mathbf{x}) y_2 + \gamma_{12} y_1 y_2]}{1 + \exp(\alpha_2 + \beta_2 \mathbf{x})} \end{aligned}$$

with the proportionality constant equal to

$$\frac{1 + \exp(\alpha_2 + \beta_2 \mathbf{x})}{1 + \exp(\alpha_1 + \beta_1 \mathbf{x}) + \exp(\alpha_2 + \beta_2 \mathbf{x}) + \exp[(\alpha_1 + \alpha_2) + (\beta_1 + \beta_2) \mathbf{x} + \gamma_{12}]}$$

In other words, the joint probability distribution is proved as (3.2). It is easy to prove that the condition $\gamma_{12} = \gamma_{21}$ is sufficient. ■

3.1.2 3 Dimensional Case

Next, in order to get clearer ideas about the general multivariate case, it is necessary to further discuss the 3 dimensional case with three binary response variables Y_1 , Y_2 and Y_3 , and a covariate vector \mathbf{x} . By assuming the linear logistic models for the conditional probability distributions $\Pr(Y_1 = y_1 | y_2, y_3, \mathbf{x})$, $\Pr(Y_2 = y_2 | y_1, y_3, \mathbf{x})$ and $\Pr(Y_3 = y_3 | y_1, y_2, \mathbf{x})$, we have

$$\Pr(y_1 | y_2, y_3, \mathbf{x}) = \frac{\exp[(\alpha_1 + \beta_1 \mathbf{x} + \gamma_{12} y_2 + \gamma_{13} y_3) y_1]}{1 + \exp(\alpha_1 + \beta_1 \mathbf{x} + \gamma_{12} y_2 + \gamma_{13} y_3)},$$

$$\Pr(y_2 | y_1, y_3, \mathbf{x}) = \frac{\exp[(\alpha_2 + \beta_2 \mathbf{x} + \gamma_{21} y_1 + \gamma_{23} y_3) y_2]}{1 + \exp(\alpha_2 + \beta_2 \mathbf{x} + \gamma_{21} y_1 + \gamma_{23} y_3)},$$

and

$$\Pr(y_3 | y_1, y_2, \mathbf{x}) = \frac{\exp[(\alpha_3 + \beta_3 \mathbf{x} + \gamma_{31} y_1 + \gamma_{32} y_2) y_3]}{1 + \exp(\alpha_3 + \beta_3 \mathbf{x} + \gamma_{31} y_1 + \gamma_{32} y_2)}, \quad (3.12)$$

where α 's, β 's and γ 's are parameters. (Note again that we are using shorthand notation, that is clear from its context.)

For a 3 dimensional case, an analog of Theorem 3.1 to guarantee compatibility for (3.12) is given in Theorem 3.2.

Theorem 3.2 *The conditional probability distributions $\Pr(y_1 | y_2, y_3, \mathbf{x})$, $\Pr(y_2 | y_1, y_3, \mathbf{x})$ and $\Pr(y_3 | y_1, y_2, \mathbf{x})$ in (3.12) are compatible if and only if $\gamma_{12} = \gamma_{21}$, $\gamma_{13} = \gamma_{31}$, and $\gamma_{23} =$*

γ_{32} . Furthermore, the joint probability distribution is of the form

$$\begin{aligned}\Pr(y_1, y_2, y_3|\mathbf{x}) &= \frac{1}{c} \exp[(\alpha_1 + \beta_1\mathbf{x})y_1 + (\alpha_2 + \beta_2\mathbf{x})y_2 + (\alpha_3 + \beta_3\mathbf{x})y_3 \\ &\quad + \gamma_{12}y_1y_2 + \gamma_{13}y_1y_3 + \gamma_{23}y_2y_3] \\ &= \frac{1}{c} \exp\left[\sum_{i=1}^3 (\alpha_i + \beta_i\mathbf{x})y_i + \sum_{1 \leq i < j \leq 3} \gamma_{ij}y_iy_j\right]\end{aligned}\quad (3.13)$$

where $c = 1 + \sum_{i=1}^3 \exp(\alpha_i + \beta_i\mathbf{x}) + \sum_{1 \leq i < j \leq 3} \exp[(\alpha_i + \alpha_j) + (\beta_i + \beta_j)\mathbf{x} + \gamma_{ij}] + \exp[(\alpha_1 + \alpha_2 + \alpha_3) + (\beta_1 + \beta_2 + \beta_3)\mathbf{x} + (\gamma_{12} + \gamma_{13} + \gamma_{23})]$.

Proof 1: (using the extension of Arnold and Strauss's compatibility condition)

$\Pr(y_1|y_2, y_3, \mathbf{x})$, $\Pr(y_2|y_1, y_3, \mathbf{x})$ and $\Pr(y_3|y_1, y_2, \mathbf{x})$ are compatible if and only if there are probabilities $\Pr(y_1, y_2, y_3|\mathbf{x})$, $\Pr(y_2, y_3|\mathbf{x})$, $\Pr(y_1, y_3|\mathbf{x})$ and $\Pr(y_1, y_2|\mathbf{x})$ such that

$$\begin{aligned}\Pr(y_1, y_2, y_3|\mathbf{x}) &= \Pr(y_1|y_2, y_3, \mathbf{x}) \cdot \Pr(y_2, y_3|\mathbf{x}) \\ &= \Pr(y_2|y_1, y_3, \mathbf{x}) \cdot \Pr(y_1, y_3|\mathbf{x}) \\ &= \Pr(y_3|y_1, y_2, \mathbf{x}) \cdot \Pr(y_1, y_2|\mathbf{x})\end{aligned}$$

for all $y_1, y_2, y_3, \mathbf{x}$.

Equivalently,

$$\begin{aligned}\frac{\Pr(y_1|y_2, y_3, \mathbf{x})}{\Pr(y_2|y_1, y_3, \mathbf{x})} &= \frac{\Pr(y_1, y_3|\mathbf{x})}{\Pr(y_2, y_3|\mathbf{x})}, \\ \frac{\Pr(y_1|y_2, y_3, \mathbf{x})}{\Pr(y_3|y_1, y_2, \mathbf{x})} &= \frac{\Pr(y_1, y_2|\mathbf{x})}{\Pr(y_2, y_3|\mathbf{x})}, \\ \frac{\Pr(y_2|y_1, y_3, \mathbf{x})}{\Pr(y_3|y_1, y_2, \mathbf{x})} &= \frac{\Pr(y_1, y_2|\mathbf{x})}{\Pr(y_1, y_3|\mathbf{x})}.\end{aligned}\quad (3.14)$$

Special cases of (3.14) are

$$\begin{aligned}\frac{\Pr(Y_1 = 0, Y_3 = 0|\mathbf{x})}{\Pr(Y_2 = 0, Y_3 = 0|\mathbf{x})} &= \frac{\Pr(Y_1 = 0|Y_2 = 0, Y_3 = 0, \mathbf{x})}{\Pr(Y_2 = 0|Y_1 = 0, Y_3 = 0, \mathbf{x})} \\ &= \frac{1 + \exp(\alpha_2 + \beta_2\mathbf{x})}{1 + \exp(\alpha_1 + \beta_1\mathbf{x})},\end{aligned}\quad (3.15)$$

$$\begin{aligned}
\frac{\Pr(Y_1 = 1, Y_3 = 0|\mathbf{x})}{\Pr(Y_2 = 0, Y_3 = 0|\mathbf{x})} &= \frac{\Pr(Y_1 = 1|Y_2 = 0, Y_3 = 0, \mathbf{x})}{\Pr(Y_2 = 0|Y_1 = 1, Y_3 = 0, \mathbf{x})} \\
&= \frac{1 + \exp(\alpha_2 + \beta_2\mathbf{x} + \gamma_{21})}{1 + \exp(\alpha_1 + \beta_1\mathbf{x})} \cdot \exp(\alpha_1 + \beta_1\mathbf{x}), \quad (3.16)
\end{aligned}$$

$$\begin{aligned}
\frac{\Pr(Y_1 = 0, Y_3 = 0|\mathbf{x})}{\Pr(Y_2 = 1, Y_3 = 0|\mathbf{x})} &= \frac{\Pr(Y_1 = 0|Y_2 = 1, Y_3 = 0, \mathbf{x})}{\Pr(Y_2 = 1|Y_1 = 0, Y_3 = 0, \mathbf{x})} \\
&= \frac{1 + \exp(\alpha_2 + \beta_2\mathbf{x})}{1 + \exp(\alpha_1 + \beta_1\mathbf{x} + \gamma_{12})} \cdot \frac{1}{\exp(\alpha_2 + \beta_2\mathbf{x})}, \quad (3.17)
\end{aligned}$$

$$\begin{aligned}
\frac{\Pr(Y_1 = 1, Y_3 = 0|\mathbf{x})}{\Pr(Y_2 = 1, Y_3 = 0|\mathbf{x})} &= \frac{\Pr(Y_1 = 1|Y_2 = 1, Y_3 = 0, \mathbf{x})}{\Pr(Y_2 = 1|Y_1 = 1, Y_3 = 0, \mathbf{x})} \\
&= \frac{1 + \exp(\alpha_2 + \beta_2\mathbf{x} + \gamma_{21})}{1 + \exp(\alpha_1 + \beta_1\mathbf{x} + \gamma_{12})} \\
&\quad \cdot \exp[(\alpha_1 - \alpha_2) + (\beta_1 - \beta_2)\mathbf{x} + (\gamma_{12} - \gamma_{21})]. \quad (3.18)
\end{aligned}$$

Dividing (3.15) by (3.16) and dividing (3.17) by (3.18), we have

$$\frac{\Pr(Y_1 = 0, Y_3 = 0|\mathbf{x})}{\Pr(Y_1 = 1, Y_3 = 0|\mathbf{x})} = \frac{1 + \exp(\alpha_2 + \beta_2\mathbf{x})}{1 + \exp(\alpha_2 + \beta_2\mathbf{x} + \gamma_{21})} \cdot \frac{1}{\exp(\alpha_1 + \beta_1\mathbf{x})}, \quad (3.19)$$

$$\frac{\Pr(Y_1 = 0, Y_3 = 0|\mathbf{x})}{\Pr(Y_1 = 1, Y_3 = 0|\mathbf{x})} = \frac{1 + \exp(\alpha_2 + \beta_2\mathbf{x})}{1 + \exp(\alpha_2 + \beta_2\mathbf{x} + \gamma_{21})} \cdot \frac{1}{\exp(\alpha_1 + \beta_1\mathbf{x} + \gamma_{12} - \gamma_{21})}. \quad (3.20)$$

Comparing (3.19) with (3.20), we conclude that one of the necessary and sufficient conditions for $\Pr(y_1|y_2, y_3, \mathbf{x})$, $\Pr(y_2|y_1, y_3, \mathbf{x})$ and $\Pr(y_3|y_1, y_2, \mathbf{x})$ to be compatible conditional distributions is that $\gamma_{12} = \gamma_{21}$.

By symmetry, we conclude that other necessary and sufficient conditions for compatibility are $\gamma_{13} = \gamma_{31}$ and $\gamma_{23} = \gamma_{32}$.

Using the same procedure which has resulted in (3.19) and (3.20), we have

$$\begin{aligned}
\frac{\Pr(Y_1 = 0, Y_3 = 1|\mathbf{x})}{\Pr(Y_1 = 1, Y_3 = 1|\mathbf{x})} &= \frac{\Pr((Y_1 = 0, Y_3 = 1|\mathbf{x}))}{\Pr(Y_2 = 0, Y_3 = 1|\mathbf{x})} \cdot \frac{\Pr(Y_2 = 0, Y_3 = 1|\mathbf{x})}{\Pr(Y_1 = 1, Y_3 = 1|\mathbf{x})} \\
&= \frac{\Pr(Y_1 = 0|Y_2 = 0, Y_3 = 1, \mathbf{x})}{\Pr(Y_2 = 0|Y_1 = 0, Y_3 = 1, \mathbf{x})} \cdot \frac{\Pr(Y_2 = 0|Y_1 = 1, Y_3 = 1, \mathbf{x})}{\Pr(Y_1 = 1|Y_2 = 0, Y_3 = 1, \mathbf{x})} \\
&= \frac{1 + \exp(\alpha_2 + \beta_2\mathbf{x} + \gamma_{23})}{1 + \exp(\alpha_2 + \beta_2\mathbf{x} + \gamma_{12} + \gamma_{23})} \cdot \frac{1}{\exp(\alpha_1 + \beta_1\mathbf{x} + \gamma_{13})} \quad (3.21)
\end{aligned}$$

and

$$\begin{aligned}
\frac{\Pr(Y_1 = 1, Y_3 = 0|\mathbf{x})}{\Pr(Y_1 = 1, Y_3 = 1|\mathbf{x})} &= \frac{\Pr(Y_1 = 1, Y_3 = 0|\mathbf{x})}{\Pr(Y_1 = 1, Y_2 = 0|\mathbf{x})} \cdot \frac{\Pr(Y_1 = 1, Y_2 = 0|\mathbf{x})}{\Pr(Y_1 = 1, Y_3 = 1|\mathbf{x})} \\
&= \frac{\Pr(Y_3 = 0|Y_1 = 1, Y_2 = 0, \mathbf{x})}{\Pr(Y_2 = 0|Y_1 = 1, Y_3 = 0, \mathbf{x})} \cdot \frac{\Pr(Y_2 = 0|Y_1 = 1, Y_3 = 1, \mathbf{x})}{\Pr(Y_3 = 1|Y_1 = 1, Y_2 = 0, \mathbf{x})} \\
&= \frac{1 + \exp(\alpha_2 + \beta_2\mathbf{x} + \gamma_{12})}{1 + \exp(\alpha_2 + \beta_2\mathbf{x} + \gamma_{12} + \gamma_{23})} \cdot \frac{1}{\exp(\alpha_3 + \beta_3\mathbf{x} + \gamma_{13})}. \quad (3.22)
\end{aligned}$$

From (3.21) and (3.22), $\Pr(Y_1 = 0, Y_3 = 1|\mathbf{x})$ and $\Pr(Y_1 = 1, Y_3 = 0|\mathbf{x})$ can be expressed in terms of $\Pr(Y_1 = 1, Y_3 = 1|\mathbf{x})$, respectively:

$$\Pr(Y_1 = 0, Y_3 = 1|\mathbf{x}) = \frac{1 + \exp(\alpha_2 + \beta_2\mathbf{x} + \gamma_{23})}{1 + \exp(\alpha_2 + \beta_2\mathbf{x} + \gamma_{12} + \gamma_{23})} \cdot \frac{\Pr(Y_1 = 1, Y_3 = 1|\mathbf{x})}{\exp(\alpha_1 + \beta_1\mathbf{x} + \gamma_{13})} \quad (3.23)$$

and

$$\Pr(Y_1 = 1, Y_3 = 0|\mathbf{x}) = \frac{1 + \exp(\alpha_2 + \beta_2\mathbf{x} + \gamma_{12})}{1 + \exp(\alpha_2 + \beta_2\mathbf{x} + \gamma_{12} + \gamma_{23})} \cdot \frac{\Pr(Y_1 = 1, Y_3 = 1|\mathbf{x})}{\exp(\alpha_3 + \beta_3\mathbf{x} + \gamma_{13})}. \quad (3.24)$$

Substituting (3.24) for (3.20), $\Pr(Y_1 = 0, Y_1 = 0|\mathbf{x})$ also can be expressed in terms of $\Pr(Y_1 = 1, Y_3 = 1|\mathbf{x})$:

$$\Pr(Y_1 = 1, Y_3 = 1|\mathbf{x}) = \frac{1 + \exp(\alpha_2 + \beta_2\mathbf{x})}{1 + \exp(\alpha_2 + \beta_2\mathbf{x} + \gamma_{12} + \gamma_{23})} \cdot \frac{\Pr(Y_1 = 1, Y_3 = 1|\mathbf{x})}{\exp[(\alpha_1 + \alpha_3) + (\beta_1 + \beta_3)\mathbf{x} + \gamma_{13}]} \quad (3.25)$$

Under the constraints of $\Pr(Y_1 = 0, Y_3 = 0|\mathbf{x}) + \Pr(Y_1 = 0, Y_3 = 1|\mathbf{x}) + \Pr(Y_1 = 1, Y_3 = 0|\mathbf{x}) + \Pr(Y_1 = 1, Y_3 = 1|\mathbf{x}) = 1$, a bit of calculation by substituting (3.23), (3.24) and (3.25) will result in following equation:

$$\frac{c \cdot \Pr(Y_1 = 1, Y_3 = 1|\mathbf{x})}{[1 + \exp(\alpha_2 + \beta_2\mathbf{x} + \gamma_{12} + \gamma_{23})] \exp[(\alpha_1 + \alpha_3) + (\beta_1 + \beta_3)\mathbf{x} + \gamma_{13}]} = 1$$

where $c = 1 + \sum_{i=1}^3 \exp(\alpha_i + \beta_i\mathbf{x}) + \sum_{1 \leq i < j \leq 3} \exp[(\alpha_i + \alpha_j) + (\beta_i + \beta_j)\mathbf{x} + \gamma_{ij}] + \exp[(\alpha_1 + \alpha_2 + \alpha_3) + (\beta_1 + \beta_2 + \beta_3)\mathbf{x} + (\gamma_{12} + \gamma_{13} + \gamma_{23})]$.

Thus,

$$\begin{aligned} \Pr(Y_1 = 1, Y_3 = 1 | \mathbf{x}) &= \frac{1}{c} \cdot \exp[(\alpha_1 + \alpha_3) + (\beta_1 + \beta_3)\mathbf{x} + \gamma_{13}] \\ &\quad \cdot [1 + \exp(\alpha_2 + \beta_2\mathbf{x} + \gamma_{12} + \gamma_{23})] \end{aligned}$$

Similarly, we can obtain other probabilities $\Pr(y_i, y_j | \mathbf{x})$. In brief, we can write $\Pr(y_i, y_j | \mathbf{x})$ as:

$$\begin{aligned} \Pr(y_i, y_j | \mathbf{x}) &= \frac{1}{c} \cdot \exp[(\alpha_i + \beta_i\mathbf{x})y_i + (\alpha_j + \beta_j\mathbf{x})y_j + \gamma_{ij}y_i y_j] \\ &\quad \cdot [1 + \exp(\alpha_k + \beta_k\mathbf{x} + \gamma_{ik}y_i + \gamma_{kj}y_j)] \end{aligned} \quad (3.26)$$

where $i < j$, $k \neq i$, $k \neq j$ and $i, j, k = 1, 2, 3$.

Finally, using (3.12) and (3.26), we can obtain the joint probability distribution as (3.13). ■

Proof 2: (using the extension of Gelman and Speed's compatibility condition)

From (2.4), a condition for $\Pr(y_1 | y_2, y_3, \mathbf{x})$, $\Pr(y_2 | y_1, y_3, \mathbf{x})$ and $\Pr(y_3 | y_1, y_2, \mathbf{x})$ to be compatible is that

$$\frac{\Pr(y_1'' | y_2, y_3, \mathbf{x}) \Pr(y_2'' | y_1'', y_3, \mathbf{x}) \Pr(y_2 | y_1', y_3, \mathbf{x}) \Pr(y_3 | y_1', y_2', \mathbf{x})}{\Pr(y_1' | y_2, y_3, \mathbf{x}) \Pr(y_2' | y_1', y_3, \mathbf{x}) \Pr(y_2 | y_1'', y_3, \mathbf{x}) \Pr(y_3 | y_1'', y_2'', \mathbf{x})} \quad (3.27)$$

does not depend on y_1 , y_2 and y_3 for all choices of $(y_1', y_2') \neq (y_1'', y_2'')$.

Applying (3.12), in this case we rewrite (3.27) as:

$$\begin{aligned} &\frac{\Pr(y_1'' | y_2, y_3, \mathbf{x}) \Pr(y_2'' | y_1'', y_3, \mathbf{x}) \Pr(y_2 | y_1', y_3, \mathbf{x}) \Pr(y_3 | y_1', y_2', \mathbf{x})}{\Pr(y_1' | y_2, y_3, \mathbf{x}) \Pr(y_2' | y_1', y_3, \mathbf{x}) \Pr(y_2 | y_1'', y_3, \mathbf{x}) \Pr(y_3 | y_1'', y_2'', \mathbf{x})} = \\ &\frac{1 + \exp(\alpha_3 + \beta_3\mathbf{x} + \gamma_{31}y_1'' + \gamma_{32}y_2'')}{1 + \exp(\alpha_3 + \beta_3\mathbf{x} + \gamma_{31}y_1' + \gamma_{32}y_2')} \cdot \exp[(\alpha_1 + \beta_1\mathbf{x})(y_1'' - y_1') + (\alpha_2 + \beta_2\mathbf{x})(y_2'' - y_2')] \\ &\cdot \exp[\gamma_{21}(y_1''y_2'' - y_1'y_2')] \cdot \exp[y_2(y_1'' - y_1')(\gamma_{12} - \gamma_{21})] \cdot \exp[y_3(y_1'' - y_1')(\gamma_{13} - \gamma_{31})] \\ &\cdot \exp[y_3(y_2'' - y_2')(\gamma_{23} - \gamma_{32})] \end{aligned} \quad (3.28)$$

Let $y_1'' = 1$, $y_1' = 0$, $y_2'' = 1$ and $y_2' = 0$, then (3.28) does not depend on y_1 , y_2 and y_3 if and only if $\gamma_{12} = \gamma_{21}$, $\gamma_{13} = \gamma_{31}$ and $\gamma_{23} = \gamma_{32}$. This indicates that necessary conditions for the trivariate case are $\gamma_{12} = \gamma_{21}$, $\gamma_{13} = \gamma_{31}$ and $\gamma_{23} = \gamma_{32}$.

Applying (2.3), let both fixed values of y_1 and y_2 be zeros, then we know for the joint probability distribution $\Pr(y_1, y_2, y_3, | \mathbf{x})$,

$$\begin{aligned} \Pr(y_1, y_2, y_3, | \mathbf{x}) &\propto \frac{\Pr(y_1|y_2, y_3, \mathbf{x}) \Pr(y_2|y_1 = 0, y_3, \mathbf{x}) \Pr(y_3|y_1 = 0, y_2 = 0, \mathbf{x})}{\Pr(y_1 = 0|y_2, y_3, \mathbf{x}) \Pr(y_2 = 0|y_1 = 0, y_3, \mathbf{x})} \\ &= \frac{\exp[(\alpha_1 + \beta_1 \mathbf{x})y_1 + (\alpha_2 + \beta_2 \mathbf{x})y_2 + (\alpha_3 + \beta_3 \mathbf{x})y_3 + \gamma_{12}y_1y_2 + \gamma_{13}y_1y_3 + \gamma_{23}y_2y_3]}{1 + \exp(\alpha_3 + \beta_3 \mathbf{x})} \end{aligned}$$

with the proportionality constant equal to $[1 + \exp(\alpha_3 + \beta_3 \mathbf{x})]/c$, where c is given in (3.13). Now we have proved that the joint probability distribution is the form of (3.13), and sufficiency of the conditions is easily proved. ■

In both Theorem 3.1 and Theorem 3.2, we use two kinds of techniques for the proof. The first technique uses Arnold and Strauss's compatibility conditions; the second uses Gelman and Speed's idea for compatibility conditions. It is obvious, from the lengths of the proofs in the 3 dimensional case, that the extension of Gelman and Speed's compatibility conditions is easier to use than the extension of Arnold and Strauss's. In fact, using the extension of Gelman and Speed's compatibility conditions, it is possible to prove an analogous theorem to Theorem 3.1 and Theorem 3.2 for the general multivariate case, but using the extension of Arnold and Strauss's compatibility conditions, the necessity of the condition might be too tedious to prove.

3.2 General Case

Our ultimate aim is to develop a conditionally specified logistic regression model for multivariate binary responses with covariates and to discuss necessary and sufficient conditions for compatibility under general dimensions. For notation, we change $\beta_j \mathbf{x}$ above to $\beta_{j1}x_1 + \dots + \beta_{jm}x_m$. Assume

$$\begin{aligned} \Pr(y_1|y_2, \dots, y_n, \mathbf{x}) &= \frac{\exp[(\alpha_1 + \sum_{j=1}^m \beta_{1j}x_j + \sum_{j \neq 1} \gamma_{1j}y_j)y_1]}{1 + \exp(\alpha_1 + \sum_{j=1}^m \beta_{1j}x_j + \sum_{j \neq 1} \gamma_{1j}y_j)}, \\ &\vdots \\ \Pr(y_n|y_1, \dots, y_{n-1}, \mathbf{x}) &= \frac{\exp[(\alpha_n + \sum_{j=1}^m \beta_{nj}x_j + \sum_{j \neq n} \gamma_{nj}y_j)y_n]}{1 + \exp(\alpha_n + \sum_{j=1}^m \beta_{nj}x_j + \sum_{j \neq n} \gamma_{nj}y_j)}. \end{aligned} \quad (3.29)$$

The above discussions in both 2 and 3 dimensional cases provide immediate generalization of the results for the n dimensional case. For the reason that we have mentioned at the end of last section, however, we prove the necessary conditions for conditional distributions (3.29) to be compatible only using the extension of Gelman and Speed's compatibility condition.

Theorem 3.3 *The conditional probability distributions $\Pr(y_i|y_j, j \neq i, \mathbf{x})$ for $i = 1, \dots, n$, in (3.29) are compatible if $\gamma_{ij} = \gamma_{ji}$ for all $i \neq j$. Under conditions for compatibility, a joint probability distribution is as follows:*

$$\Pr(y_1, \dots, y_n | \mathbf{x}) = \frac{1}{c} \exp\left\{ \sum_{i=1}^n [(\alpha_i + \sum_{j=1}^m \beta_{ij}x_j)y_i] + \sum_{1 \leq i < j \leq n} \gamma_{ij}y_i y_j \right\} \quad (3.30)$$

where c is the normalizing constant which involves a sum of 2^n exponential terms and is the form of

$$c = \sum_{y_1=0}^1 \dots \sum_{y_n=0}^1 \exp\left\{ \sum_{i=1}^n (\alpha_i + \sum_{j=1}^m \beta_{ij}x_j)y_i + \sum_{i < j} \gamma_{ij}y_i y_j \right\} \quad (3.31)$$

Proof: By extending (2.4) and (3.27) to the n dimensional case, we conclude that a condition for $\Pr(y_i|y_j, j \neq i, \mathbf{x})$ to be compatible is that

$$\frac{\prod_{i=1}^n \Pr(y_i|y'_1, \dots, y'_{i-1}, y_{i+1}, \dots, y_n, \mathbf{x}) \cdot \prod_{i=1}^n \Pr(y''_i|y''_1, \dots, y''_{i-1}, y_{i+1}, \dots, y_n, \mathbf{x})}{\prod_{i=1}^n \Pr(y_i|y''_1, \dots, y''_{i-1}, y_{i+1}, \dots, y_n, \mathbf{x}) \cdot \prod_{i=1}^n \Pr(y'_i|y'_1, \dots, y'_{i-1}, y_{i+1}, \dots, y_n, \mathbf{x})} \quad (3.32)$$

does not depend on y_i ($i = 1, \dots, n$) for all choices of $(y'_1, \dots, y'_n) \neq (y''_1, \dots, y''_n)$.

Given (3.29), we can write (3.32) in detail, that is:

$$\begin{aligned}
& \frac{\prod_{i=1}^n \Pr(y_i | y'_1, \dots, y'_{i-1}, y_{i+1}, \dots, y_n, \mathbf{x}) \cdot \prod_{i=1}^n \Pr(y''_i | y''_1, \dots, y''_{i-1}, y_{i+1}, \dots, y_n, \mathbf{x})}{\prod_{i=1}^n \Pr(y_i | y''_1, \dots, y''_{i-1}, y_{i+1}, \dots, y_n, \mathbf{x}) \cdot \prod_{i=1}^n \Pr(y'_i | y'_1, \dots, y'_{i-1}, y_{i+1}, \dots, y_n, \mathbf{x})} \\
&= \frac{\prod_{i=1}^n \exp[(\alpha_i + \sum_{j=1}^m \beta_{ij} x_j + \sum_{j<i} \gamma_{ij} y'_j + \sum_{j>i} \gamma_{ij} y_j) y_i]}{\prod_{i=1}^n \exp[(\alpha_i + \sum_{j=1}^m \beta_{ij} x_j + \sum_{j<i} \gamma_{ij} y''_j + \sum_{j>i} \gamma_{ij} y_j) y_i]} \\
&\quad \cdot \frac{\prod_{i=1}^n \exp[(\alpha_i + \sum_{j=1}^m \beta_{ij} x_j + \sum_{j<i} \gamma_{ij} y''_j + \sum_{j>i} \gamma_{ij} y_j) y'_i]}{\prod_{i=1}^n \exp[(\alpha_i + \sum_{j=1}^m \beta_{ij} x_j + \sum_{j<i} \gamma_{ij} y'_j + \sum_{j>i} \gamma_{ij} y_j) y'_i]} \\
&= \prod_{i=1}^n \exp[(\alpha_i + \sum_{j=1}^m \beta_{ij} x_j)(y''_i - y'_i) + \sum_{i>j} \gamma_{ij} (y''_j y''_i - y'_j y'_i)] \\
&\quad \cdot \prod_{i=1}^n \exp[\sum_{i<j} \gamma_{ji} y_j (y'_i - y''_i)] \cdot \prod_{i=1}^n \exp[\sum_{i<j} \gamma_{ij} y_j (y''_i - y'_i)] \tag{3.33}
\end{aligned}$$

If $(y'_1, \dots, y'_n) = \mathbf{1}$ and $(y''_1, \dots, y''_n) = \mathbf{0}$, (3.33) becomes

$$\begin{aligned}
& \frac{\prod_{i=1}^n \Pr(y_i | y'_1, \dots, y'_{i-1}, y_{i+1}, \dots, y_n, \mathbf{x}) \cdot \prod_{i=1}^n \Pr(y''_i | y''_1, \dots, y''_{i-1}, y_{i+1}, \dots, y_n, \mathbf{x})}{\prod_{i=1}^n \Pr(y_i | y''_1, \dots, y''_{i-1}, y_{i+1}, \dots, y_n, \mathbf{x}) \cdot \prod_{i=1}^n \Pr(y'_i | y'_1, \dots, y'_{i-1}, y_{i+1}, \dots, y_n, \mathbf{x})} \\
&= \prod_{i=1}^n \exp[-(\alpha_i + \sum_{j=1}^m \beta_{ij} x_j + \sum_{i>j} \gamma_{ij})] \cdot \prod_{i=1}^n \exp[\sum_{i<j} (\gamma_{ji} - \gamma_{ij}) y_j]. \tag{3.34}
\end{aligned}$$

Therefore, (3.32) does not depend on y_i ($i = 1, \dots, n$) if and only if $\gamma_{ij} = \gamma_{ji}$ for all $i \neq j$ in (3.34). Now we get the conclusion that the necessary conditions for $\Pr(y_i | y_j, j \neq i, \mathbf{x})$ to be compatible conditional distributions are $\gamma_{ij} = \gamma_{ji}$ for all $i \neq j$.

Again extending (2.1) to the n dimensional case, for the joint probability distribution $\Pr(y_1, \dots, y_n | \mathbf{x})$,

$$\Pr(y_1, \dots, y_n | \mathbf{x}) \propto \frac{\prod_{i=1}^n \Pr(y_i | y'_1, \dots, y'_{i-1}, y_{i+1}, \dots, y_n, \mathbf{x})}{\prod_{i=1}^n \Pr(y'_i | y'_1, \dots, y'_{i-1}, y_{i+1}, \dots, y_n, \mathbf{x})} \tag{3.35}$$

In this case, if $y'_i = 0$ for all i , we can write (3.35) as follows:

$$\Pr(y_1, \dots, y_n | \mathbf{X}) \propto \prod_{i=1}^n \exp[(\alpha_i + \sum_{j=1}^m \beta_{ij} x_j + \sum_{i<j} \gamma_{ij} y_j) y_i]$$

with the proportionality constant equal to $1/c$, where c is defined as (3.31). That is, the model for multivariate binary data with covariates based on compatible conditionally specified logistic regressions is:

$$\Pr(y_1, \dots, y_n | \mathbf{x}) = \frac{1}{c} \exp\left\{\sum_{i=1}^n [(\alpha_i + \sum_{j=1}^m \beta_{ij} x_j) y_i] + \sum_{1 \leq i < j \leq n} \gamma_{ij} y_i y_j\right\}$$

where $c = \sum_{y_1=0}^1 \cdots \sum_{y_n=0}^1 \exp\{\sum_{i=1}^n (\alpha_i + \sum_{j=1}^m \beta_{ij} x_j) y_i + \sum_{i < j} \gamma_{ij} y_i y_j\}$.

■

Finally, we prove the sufficient conditions for conditional distributions (3.29) to be compatible.

Theorem 3.4 *The conditional probability distributions of (3.30) are linear logistic distributions.*

Proof: Given the model (3.30) with $\gamma_{ij} = \gamma_{ji}$, $i \neq j$, we have

$$\begin{aligned} \Pr(y_i, i \neq k | X) &= \Pr(y_1, \dots, y_{k-1}, y_k = 1, y_{k+1}, \dots, y_n | X) \\ &\quad + \Pr(y_1, \dots, y_{k-1}, y_k = 0, y_{k+1}, \dots, y_n | X) \\ &= \frac{1}{c} \left\{ \exp\left[(\alpha_k + \sum_{j=1}^m \beta_{kj} x_j) + \sum_{i \neq k} (\alpha_i + \sum_{j=1}^m \beta_{ij} x_j) y_i \right. \right. \\ &\quad \left. \left. + \sum_{j \neq k} \gamma_{kj} y_j + \sum_{1 \leq i < j \leq n, i, j \neq k} \gamma_{ij} y_i y_j \right] \right. \\ &\quad \left. + \exp\left[\sum_{i \neq k} (\alpha_i + \sum_{j=1}^m \beta_{ij} x_j) y_i + \sum_{1 \leq i < j \leq n, i, j \neq k} \gamma_{ij} y_i y_j \right] \right\} \\ &= \frac{1}{c} \exp\left[\sum_{i \neq k} (\alpha_i + \sum_{j=1}^m \beta_{ij} x_j) y_i + \sum_{1 \leq i < j \leq n, i, j \neq k} \gamma_{ij} y_i y_j \right] \\ &\quad \cdot \left\{ 1 + \exp\left[\alpha_k + \sum_{j=1}^m \beta_{kj} x_j + \sum_{j \neq k} \gamma_{kj} y_j \right] \right\}. \end{aligned}$$

Further, we obtain

$$\Pr(y_k | y_j, j \neq k, X) = \frac{\Pr(y_1, y_2, \dots, y_n | X)}{\Pr(y_j, j \neq k | X)}$$

$$\begin{aligned}
&= \frac{\exp\{\sum_{i=1}^n [(\alpha_i + \sum_{j=1}^m \beta_{ij}x_j)y_i] + \sum_{1 \leq i < j \leq n} \gamma_{ij}y_i y_j\}}{\exp[\sum_{i \neq k} (\alpha_i + \sum_{j=1}^m \beta_{ij}x_j)y_i + \sum_{1 \leq i < j \leq n, i, j \neq k} \gamma_{ij}y_i y_j]} \\
&\quad \cdot \frac{1}{\{1 + \exp[\alpha_k + \sum_{j=1}^m \beta_{kj}x_j + \sum_{j \neq k} \gamma_{kj}y_j]\}} \\
&= \frac{\exp[(\alpha_k + \sum_{j=1}^m \beta_{kj}x_j)y_k + (\sum_{j \neq k} \gamma_{kj}y_j)y_k]}{1 + \exp[\alpha_k + \sum_{j=1}^m \beta_{kj}x_j + \sum_{j \neq k} \gamma_{kj}y_j]}.
\end{aligned}$$

This equation indicates that conditional probability distributions $\Pr(y_k|y_j, j \neq k, X)$ for $k = 1, \dots, n$ are linear logistic distributions. ■

3.3 Further Extension

The discussion so far assumes implicitly that the conditional distributions $\Pr(y_j|y_i, i \neq j, \mathbf{x})$ for $j = 1, \dots, n$, are logistic, so we say that the model (3.30) for multivariate binary data with covariates we have constructed is based on compatible conditionally specified logistic regressions. In fact, as mentioned before, with binary responses, any other link function could be chosen instead of the logit link function. From this point of view, a question to be addressed here is what are conditions for compatibility when the conditional distributions are based on a distribution F that is not logistic, i.e.,

$$\Pr(Y_j = 1|y_i, i \neq j, \mathbf{x}) = F(\alpha_j + \beta_j \mathbf{x} + \sum_{i \neq j} \gamma_{ji}y_i), j = 1, \dots, n.$$

In this section, we intend to motivate the development of conditions for compatibility when F is not logistic. For this reason, let us simply consider the case of two binary response variables Y_1 and Y_2 , and a covariate vector \mathbf{x} . Assume that the conditional distribution of Y_j is

$$\Pr(Y_j = 1|y_i, \mathbf{x}) = F(\alpha_j + \beta_j \mathbf{x} + \gamma_{ji}y_i)$$

for $j = 1, 2, i = 3 - j$.

Here we use Arnold and Strauss's compatibility condition instead of Gelman and Speed's. Consequently, the conditional distributions $\Pr(y_1|y_2, \mathbf{x})$ and $\Pr(y_2|y_1, \mathbf{x})$ are compatible if and only if there exist probabilities $\Pr(y_1, y_2|\mathbf{x})$, $\Pr(y_1|\mathbf{x})$ and $\Pr(y_2|\mathbf{x})$ such that:

$$\begin{aligned}\Pr(y_1, y_2|\mathbf{x}) &= \Pr(y_1|y_2, \mathbf{x}) \Pr(y_2|\mathbf{x}) \\ &= \Pr(y_2|y_1, \mathbf{x}) \Pr(y_1|\mathbf{x}).\end{aligned}$$

In detail, we write the above form as:

$$\frac{\Pr(Y_1 = 1|Y_2 = 1, \mathbf{x})}{\Pr(Y_2 = 1|Y_1 = 1, \mathbf{x})} = \frac{\Pr(Y_1 = 1|\mathbf{x})}{\Pr(Y_2 = 1|\mathbf{x})}, \quad (3.36)$$

$$\frac{\Pr(Y_1 = 1|Y_2 = 0, \mathbf{x})}{\Pr(Y_2 = 0|Y_1 = 1, \mathbf{x})} = \frac{\Pr(Y_1 = 1|\mathbf{x})}{\Pr(Y_2 = 0|\mathbf{x})}, \quad (3.37)$$

$$\frac{\Pr(Y_1 = 0|Y_2 = 1, \mathbf{x})}{\Pr(Y_2 = 1|Y_1 = 0, \mathbf{x})} = \frac{\Pr(Y_1 = 0|\mathbf{x})}{\Pr(Y_2 = 1|\mathbf{x})}, \quad (3.38)$$

$$\frac{\Pr(Y_1 = 0|Y_2 = 0, \mathbf{x})}{\Pr(Y_2 = 0|Y_1 = 0, \mathbf{x})} = \frac{\Pr(Y_1 = 0|\mathbf{x})}{\Pr(Y_2 = 0|\mathbf{x})}. \quad (3.39)$$

In particular, if we divide (3.36) by (3.37) and divide (3.38) by (3.39), we have

$$\begin{aligned}\frac{\Pr(Y_2 = 0|\mathbf{x})}{\Pr(Y_2 = 1|\mathbf{x})} &= \frac{\Pr(Y_1 = 1|Y_2 = 1, \mathbf{x})}{\Pr(Y_2 = 1|Y_1 = 1, \mathbf{x})} \cdot \frac{\Pr(Y_2 = 0|Y_1 = 1, \mathbf{x})}{\Pr(Y_1 = 1|Y_2 = 0, \mathbf{x})} \\ &= \frac{F(\alpha_1 + \beta_1\mathbf{x} + \gamma_{12})}{F(\alpha_1 + \beta_1\mathbf{x})} \cdot \frac{1 - F(\alpha_2 + \beta_2\mathbf{x} + \gamma_{21})}{F(\alpha_2 + \beta_2\mathbf{x} + \gamma_{21})}\end{aligned} \quad (3.40)$$

$$\begin{aligned}\frac{\Pr(Y_2 = 0|\mathbf{x})}{\Pr(Y_2 = 1|\mathbf{x})} &= \frac{\Pr(Y_1 = 0|Y_2 = 1, \mathbf{x})}{\Pr(Y_2 = 1|Y_1 = 0, \mathbf{x})} \cdot \frac{\Pr(Y_2 = 0|Y_1 = 0, \mathbf{x})}{\Pr(Y_1 = 0|Y_2 = 0, \mathbf{x})} \\ &= \frac{1 - F(\alpha_1 + \beta_1\mathbf{x} + \gamma_{12})}{1 - F(\alpha_1 + \beta_1\mathbf{x})} \cdot \frac{1 - F(\alpha_2 + \beta_2\mathbf{x})}{F(\alpha_2 + \beta_2\mathbf{x})}\end{aligned} \quad (3.41)$$

Obviously, (3.40) and (3.41) must be equal. In other words, a necessary and sufficient condition for $\Pr(y_1|y_2, \mathbf{x})$ and $\Pr(y_2|y_1, \mathbf{x})$ to be compatible conditional distributions is

$$\frac{F(\alpha_1 + \beta_1\mathbf{x} + \gamma_{12})}{1 - F(\alpha_1 + \beta_1\mathbf{x} + \gamma_{12})} \cdot \frac{F(\alpha_2 + \beta_2\mathbf{x})}{1 - F(\alpha_2 + \beta_2\mathbf{x})} = \frac{F(\alpha_2 + \beta_2\mathbf{x} + \gamma_{21})}{1 - F(\alpha_2 + \beta_2\mathbf{x} + \gamma_{21})} \cdot \frac{F(\alpha_1 + \beta_1\mathbf{x})}{1 - F(\alpha_1 + \beta_1\mathbf{x})} \quad (3.42)$$

After analysing (3.42), we find that both of the following conditions make (3.42) hold:

1. $\alpha_1 = \alpha_2$, $\beta_1 = \beta_2$, and $\gamma_{12} = \gamma_{21}$.

2. $\gamma_{12} = \gamma_{21} = 0$, i.e., the two binary response variables Y_1 and Y_2 are independent.

It is not obvious that general conditions can be obtained. There may be conditions that depend on F . As for higher dimensional cases, the general analysis is quite difficult.

Chapter 4

Comparisons with Other Models

Logistic regression is widely used to study the effects of explanatory variables on binary response variables. Several different specifications have been considered for dependent binary response variables. By means of logistic regression, several useful approaches to specifying a multivariate distribution capable of representing the dependence of binary response variables have recently been proposed. However, in contrast to model (3.30) which allows for a general dependence structure for binary response variables, these other models are limited in scope. Here we discuss two kinds of these models and compare them with model (3.30).

4.1 Regressive Logistic Models

To describe sequentially dependent binary response variables, Bonney (1987) proposed regressive logistic models in which the probability of an observation is conditioned on all preceding observations. The basic theory of Bonney's models is the follow. Since there exists a natural ordering to the indexing of all dependent binary response variables, it is reasonable to decompose the probability of \mathbf{Y} given \mathbf{X} into a product of n probabilities

such as:

$$\begin{aligned}\Pr(\mathbf{Y}|\mathbf{X}) &= \Pr(Y_1, Y_2, \dots, Y_n|\mathbf{X}) \\ &= \Pr(Y_1|\mathbf{X}) \Pr(Y_2|Y_1, \mathbf{X}) \cdots \Pr(Y_n|Y_1, Y_2, \dots, Y_{n-1}, \mathbf{X}).\end{aligned}\quad (4.1)$$

Correspondingly, he defines the i th logit as

$$\theta_i = \log\left[\frac{\Pr(Y_i = 1|Y_1, \dots, Y_{i-1}, \mathbf{X})}{\Pr(Y_i = 0|Y_1, \dots, Y_{i-1}, \mathbf{X})}\right],$$

and assumes that θ_i is a linear function of $Y_1, \dots, Y_{i-1}, \mathbf{X}$ and can be expressed by

$$\theta_i = \alpha + \gamma_1 Z_1 + \cdots + \gamma_{n-1} Z_{n-1} + \beta \mathbf{X},$$

where

$$Z_j = \begin{cases} 2Y_j - 1, & \text{if } j < i \\ 0, & \text{if } j \geq i. \end{cases}$$

Therefore, the joint probability is decomposed into a product of successive conditional probabilities each of which is assumed to be univariate logistic:

$$\Pr(\mathbf{Y}|\mathbf{X}) = \prod_{i=1}^n \frac{e^{\theta_i Y_i}}{1 + e^{\theta_i}}.\quad (4.2)$$

Model (4.2) provided by Bonney has the theoretical and practical advantage that it can be analyzed and fitted in the same way as the logistic regression model for independent binary response variables, and with the same computer programs. However, there exist differences between model (3.30) and model (4.2). Model (3.30) is based on compatible conditionally specified logistic regressions and deals with a general dependence structure for the Y_i 's, but model(4.2) is based on a product of a sequence of logistic regressions and only should be used if the binary variables are observed sequentially in time. The decomposition (4.1) is applicable only when a natural ordering of the dependent observations exists. Note that a different order generally implies a different model, and the joint probabilities (4.2) are not necessarily the same for different orders.

4.2 Conditional Logistic Regression Models for Exchangeable Binary Data

Rosner (1984) proposed a polychotomous logistic regression model which reduces to the beta-binomial distribution for the number of successes when explanatory variables are zero. Based on Rosner's idea, Connolly and Liang (1988) suggested a general model in which Rosner's model is considered as a special case:

$$\Pr(\mathbf{Y}|X_1, \dots, X_n) = c(\theta, \beta) \exp\left\{ \sum_{k=0}^{Y_- - 1} F_n(k, \theta) + \sum_{j=1}^n X_j' \beta Y_j \right\} \quad (4.3)$$

where c is a normalizing constant chosen to make the probabilities, $\Pr(\mathbf{Y}|\mathbf{X})$, sum to unity and Y_- is the sum of the Y_j 's. It follows from (4.3) that the logit conditional probability of $Y_j = 1$ given $Y_{-j} = (Y_1, \dots, Y_{j-1}, Y_{j+1}, \dots, Y_n)$ and X_j is

$$\text{logit } \Pr(Y_j = 1|Y_{-j}, X_j) = F_n(w_j; \theta) + \beta' X_j, \quad (4.4)$$

where $w_j = Y_- - Y_j$ is the sum of the Y 's excluding Y_j . Clearly, the conditional probability for $Y_j = 1$ in (4.4) depends on Y_{-j} only through the sum. In other words, this indicates a very important feature of the logistic representation in (4.4): when there are no covariate effects or when $X_1 = \dots = X_n$, then the models are such that the n Y_j 's are exchangeable.

Compared with model (3.30), model (4.3) introduced by Connolly and Liang has some limitations due to the restricted dependence structure. That is, model (4.3) might only be reasonable when the dependence among observations is approximately exchangeable such as in familial data.

On the other hand, in model (3.30), the logit conditional probability of $Y_i = 1$ given y_{-i} and X is

$$\text{logit } \Pr(Y_i = 1|y_{-i}, X) = \alpha_i + \sum_{j=1}^m \beta_{ij} x_j + \sum_{i \neq j} \gamma_{ij} y_j.$$

This logistic representation means that the dependence of each single y on the rest is only through the sum when all γ_{ij} 's are equal. Thus, we conclude that model (3.30) is a special case of Connolly and Liang's model (4.3) when $\gamma_{ij} \equiv \gamma$ for all $i \neq j$ and $\beta_{ij} = \beta_j$ for all i, j , since then $\sum_{i \neq j} \gamma_{ij} y_j = \gamma(y. - y_i)$.

Chapter 5

Data Analysis and Computing

An example with some coronary bypass surgery data will now be presented to illustrate the application of the model for multivariate binary responses with covariates that we have discussed in Chapter 3. From a clinical point of view, it is important to determine risk factors for different complications. Using multivariate binary model (3.30), we can determine risk factors for several response variables simultaneously and account for dependence between the binary response variables. Zhang (1993) did the risk factor analysis separately for each binary response variable (or complication variable).

In order to fit model (3.30), three likelihood-based computing methods will be introduced. After that, several topics will be discussed, including choice of covariates and inferences from the model.

5.1 The Data Source and Description

MCR (Merged, Multi-Center, Multi-Specialty Clinical Registries) is an international data-base developed by Health Data Research Institute in which information about patients who had heart-related surgery was recorded. When doing statistical analysis, instead of using the entire data set, we choose a random sample subset which contains

880 patients information (this is done partly in order to reduce computational time). The information available both pre and post operation is provided in Table 5.1 and Table 5.2. The pre-operation information includes patient's age, gender, prior myocardial infarction, existence or non-existence of other diseases, body surface area, etc. (the first 19 variables: V1 to V19). The post-operation information includes the patient's status during and after the bypass surgery; for example, complications, such as renal or neurological problems, and survival status to 30 days following surgery (the last 7 variables: V20 to V26). The variables of primary interest are those outcome variables indicating the patient's complications and survival status after the surgery. Here, the complication variables are related to the quality of life after surgery. For more documentation of the data set, see Zhang (1993).

5.2 Estimation Procedure

Our aim in this section is to discuss from a practical viewpoint, how to apply model (3.30) and how to estimate the parameters in model (3.37).

Firstly, a simple, indirect, yet important way is to use the existing computer programs such as the `glm()` function in Splus which fits the coefficients of the linear logistic model using maximum likelihood in the binomial family. Using the `glm()` function in Splus to fit the conditional probabilities $\Pr(y_i|y_j, j \neq i, \mathbf{x})$ separately, not only can we estimate the parameters, but also we can assess the compatibility conditions for the conditionally specified distributions.

For example, after using the `glm()` function twice to fit model (3.30) for two binary response variables REM (V21), NEM (V22) and one covariate AGE (V1) in the MCR data set, we have the results in Table 5.3. From Table 5.3, in addition to obtaining the

Table 5.1: Description for the Subset of MCR Data

Variable	Name	Codes/Values	Abbreviation
V1	Age	Years	AGE
V2	Sex	0=Male 1=Female	SEX
V3	Prior Myocardial Infarction	0=No 1=Yes	PMI
V4	Obesity	0=No 1=Yes	OBE
V5	Chronic Obstructive Pulmonary Disease	0=No 1=Yes	COP
V6	Diabetes	0=No 1=Yes	DIA
V7	Renal Disease	0=No 1=Yes	REN
V8	Hypertension	0=No 1=Yes	HTN
V9	Alcohol Abuse	0=No 1=Yes	ETO
V10	Cancer	0=No 1=Yes	CA
V11	Liver Disease	0=No 1=Yes	LIV
V12	Central Nervous System Disease	0=No 1=Yes	CNS
V13	Prior Cerebrovascular Accident	0=No 1=Yes	PCA
V14	Rheumatic Heart Disease	0=No 1=Yes	RHE

Table 5.2: Description for the Subset of MCR Data (continued)

Variable	Name	Codes/Values	Abbreviation
V15	Other Surgery	0=No 1=Yes	OTH
V16	No Surgery	0=No 1=Yes	NON
V17	Valve Replacement	0=No 1=Yes	VAL
V18	Left Ventricular Dysfunction	0=Normal 1= 40-49% 2= 30-39% 3= 20-29% 4= ≤ 20%	LVD
V19	Body Surface Area	Square meters	BSA
V20	Discharge/30 Day Status	0=Dead 1=Alive	STA
V21	Renal Complication (Mild or Severe)	0=No 1=Yes	REM
V22	Neurological Complication (Mild or Severe)	0=No 1=Yes	NEM
V23	Pulmonary (Mild or Severe)	0=No 1=Yes	PUM
V24	Myocardial Infarction	0=No 1=Yes	MI
V25	Low Out Syndrome (Mild or Severe)	0=No 1=Yes	LOM
V26	Sepsis	0=No 1=Yes	SEP

Table 5.3: Results from Separate Logistic Regressions with Response Variables REM, NEM

Parameter	Value	Std. Error	t value
α_1	-8.9190	1.5952	-5.5912
β_{11}	0.0818	0.0229	3.5700
γ_{12}	1.4760	0.3914	3.7712
α_2	-7.1316	1.0204	-6.9900
β_{21}	0.0705	0.0149	4.7296
γ_{21}	1.4521	0.3959	3.6676

estimate of the parameters, more important, we find that γ_{12} is quite close to γ_{21} and this gives an indication of the adequacy of fit of model (3.30) for these data.

Secondly, model (3.30) can be viewed as an exponential family model $c^{-1} \exp\{\lambda^{-1} \mathbf{s}\}$ with “sufficient” statistic vector

$$\mathbf{s}^T = (s_1, \dots, s_M) = (y_1, \dots, y_n, y_1 y_2, y_1 y_3, \dots, y_{n-1} y_n, x_i y_j, i = 1, \dots, m, j = 1, \dots, n).$$

Given data of the form $(y_{i1}, \dots, y_{in}, x_{i1}, \dots, x_{im}), i = 1, \dots, N$, the sufficient statistic vector is $\sum_{i=1}^N \mathbf{s}_i$. Using the Newton-Raphson procedure for an exponential family log-likelihood, we can obtain the maximum likelihood estimate of the parameters in model (3.30) by an iterative process as:

$$\hat{\lambda}_{r+1} = \hat{\lambda}_r - \mathbf{H}^{-1}(\hat{\lambda}_r) \mathbf{u}(\hat{\lambda}_r)$$

for $r = 0, 1, 2, \dots$, where $\mathbf{u}(\lambda)$ is a vector whose j th component is the first derivative of the log-likelihood with respect to the j th parameter and the matrix $\mathbf{H}(\lambda)$ (sometimes called the Hessian matrix) is the matrix of second partial derivatives of log-likelihood function. $\hat{\lambda}_0$, a vector of initial estimates of λ , can come from the results of using the `glm()` function, separately for each response variable.

In order to use the Newton-Raphson method to estimate parameters in model (3.30), A C program (see Appendix) was written for analyzing the MCR data set.

The Newton-Raphson procedure means that $M(M + 1)/2$ second derivative evaluations, M first derivative evaluations and a matrix inverse are needed even before the linear search is attempted. In order to simplify these problems, the quasi-Newton method is suggested as the third choice in obtaining the estimate of the parameters in model (3.30) since in the quasi-Newton method, a method is used to approximate \mathbf{H}^{-1} directly from the first derivative information available at each step of the iteration (sometimes this approximation can be inaccurate) (see Nash 1990). To use the quasi-Newton method for analyzing our data, another C program was written. As in the first C program, we get initial values from the separate logistic regressions.

5.3 Comparison of Three Methods

All three methods (the `glm()` function, the Newton-Raphson method and the quasi-Newton method) we have introduced can be used to fit model (3.30), but what are computational differences among them? By means of the MCR data set, we use the three methods to estimate the parameters in model (3.30) under different cases, for example:

- Case 1: one binary variable – REM (V21), one covariate – LIV (V11). (See Table 5.4).
- Case 2: one binary variable – REM (V21), two covariates – AGE (V1) and LIV (V11). (See Table 5.5).
- Case 3: two binary variables – REM (V21) and NEM (V22), one covariate – SEX (V2). (See Table 5.6).

- Case 4: two binary variables – REM (V21) and NEM (V22), two covariates – SEX (V2) and COP (V5). (See Table 5.7).
- Case 5: three binary variables – REM (V21), NEM (V22) and PUM (V23), two covariates – SEX (V2) and COP (V5). (See Table 5.8 and Table 5.9).
- Case 6: four binary variables – REM (V21), NEM (V22), PUM (V23) and MI (V24), three covariates – AGE (V1), SEX (V2) and PMI (V3). (see Table 5.10, Table 5.11 and Table 5.12).

From the computations for the MCR data set with three methods, we find that:

1. The `glm()` function in Splus is an approximate method, though from the standard errors in Table 5.4 to Table 5.12, it appears that both the `glm()` function and the Newton-Raphson method generate rather similar estimates of the parameters. It is less convenient than other two methods since the number of many times we need to run the `glm()` function depends on the number of binary response variables. An advantage of the `glm()` function is that one can assess the goodness of fit by a rough comparison of γ_{ij} and γ_{ji} relative to their standard errors from the `glm()` function before correctly applying model (3.30) to a data set.
2. Sometimes the quasi-Newton method generates inaccurate estimated standard errors. For example, in Table 5.10, the standard error of α_1 is 1.09 when using the quasi-Newton method, but when using the `glm()` function or the Newton-Raphson method, the standard errors of α_1 are 1.71 and 1.64, respectively. Also in Table 5.10, using the quasi-Newton method, the standard error of α_3 is 0.94, but using the `glm()` function or the Newton-Raphson method, the standard errors of α_3 are 0.52 and 0.52, respectively.

Table 5.4: Method Comparison under Case 1

		glm() function	Newton-Raphson	quasi-Newton
α_1	value	-3.2058	-3.2059	-3.2059
	Std. Error	0.1745	0.1749	0.1748
β_{11}	value	1.4141	1.4141	1.4140
	Std. Error	1.0941	1.0942	1.0108
loglikelihood			-146.5499	-146.5495

3. The computational time for using the Newton-Raphson method to fit model (3.30) is much less than that for using the quasi-Newton method, but it requires more programming effort to compute second derivatives, first derivatives and a matrix inverse, etc.
4. The model (3.30) appears to fit adequately in all cases considered here, based on a check of the requirement $\gamma_{ij} = \gamma_{ji}$ using glm() function. One could do other standard checks for the adequacy of the logistic regressions (eg. Hosmer and Lemeshow, 1989).

5.4 Selection of Covariates

Another point worth mentioning is that, as stated before, the clinical researchers or surgeons are only interested in those polychotomous responses indicating the patient's survival status and some complication variables which are related to the quality of life after surgery, and there is the question of how to choose the covariates to use. Generally speaking, only those covariates which show potential relation with the response variable are selected. In this section, two methods for selection of covariates are suggested.

Table 5.5: Method Comparison under Case 2

		glm() function	Newton-Raphson	quasi-Newton
α_1	value	-9.6722	-9.6723	-9.3933
	Std. Error	1.5862	1.5900	1.5782
β_{11}	value	0.0966	0.0966	0.0926
	Std. Error	0.0224	0.0225	0.0223
β_{12}	value	1.8047	1.8048	1.7787
	Std. Error	1.1188	1.1192	1.0257
loglikelihood			-135.1319	-135.1475

Table 5.6: Method Comparison under Case 3

		glm() function	Newton-Raphson	quasi-Newton
α_1	value	-3.7141	-3.7140	-3.7140
	Std. Error	0.2509	0.2536	0.2431
α_2	value	-2.5473	-2.5474	-2.5475
	Std. Error	0.1504	0.1515	0.1409
β_{11}	value	0.5764	0.5756	0.5756
	Std. Error	0.3586	0.3603	0.3335
β_{21}	value	0.1043	0.1043	0.1043
	Std. Error	0.2669	0.2683	0.2587
γ_{12}	value	1.8547	1.8547	1.8548
	Std. Error	0.3797	0.3810	0.3542
γ_{21}	value	1.8546		
	Std. Error	0.3806		
loglikelihood			-392.3369	-392.3369

Table 5.7: Method Comparison under Case 4

		glm() function	Newton-Raphson	quasi-Newton
α_1	value	-3.9967	-3.9947	-3.9947
	Std. Error	0.2835	0.2825	0.2721
α_2	value	-2.5258	-2.5251	-2.5252
	Std. Error	0.1530	0.1543	0.1273
β_{11}	value	0.5212	0.5175	0.5175
	Std. Error	0.3667	0.3651	0.3402
β_{12}	value	1.5225	1.5214	1.5214
	Std. Error	0.3997	0.3997	0.3580
β_{21}	value	0.1155	0.1116	0.1116
	Std. Error	0.2672	0.2686	0.1796
β_{22}	value	-0.2900	-0.2863	-0.2863
	Std. Error	0.4156	0.4186	0.3217
γ_{12}	value	1.9178	1.9166	1.9167
	Std. Error	0.3922	0.3921	0.3641
γ_{21}	value	1.9173		
	Std. Error	0.3914		
loglikelihood			-386.2089	-386.2089

Table 5.8: Method Comparison under Case 5

		glm() function	Newton-Raphson	quasi-Newton
α_1	value	-4.9749	-4.9244	-4.9245
	Std. Error	0.4487	0.4439	0.4485
α_2	value	-3.0887	-3.0889	-3.0890
	Std. Error	0.2194	0.2198	0.2222
α_3	value	-0.8458	-0.8435	-0.8436
	Std. Error	0.0921	0.0920	0.0924
β_{11}	value	0.5686	0.5204	0.5204
	Std. Error	0.3718	0.3693	0.3714
β_{12}	value	1.1336	1.0935	1.0934
	Std. Error	0.4100	0.4104	0.4129
β_{21}	value	0.1149	0.1141	0.1141
	Std. Error	0.2729	0.2723	0.2731
β_{22}	value	-0.5904	-0.5917	-0.5917
	Std. Error	0.4215	0.4208	0.4223
β_{31}	value	-0.0081	-0.0086	-0.0086
	Std. Error	0.1646	0.1646	0.1669
β_{32}	value	1.2974	1.2883	1.2883
	Std. Error	0.2408	0.2409	0.2420

Table 5.9: Method Comparison under Case 5 (continued)

		glm() function	Newton-Raphson	quasi-Newton
γ_{12}	value	1.5391	1.4974	1.4974
	Std. Error	0.4038	0.4040	0.4079
γ_{13}	value	1.8039	1.7854	1.7855
	Std. Error	0.4652	0.4712	0.4767
γ_{23}	value	1.2493	1.2503	1.2503
	Std. Error	0.2660	0.2662	0.2692
γ_{21}	value	1.4963		
	Std. Error	0.4039		
γ_{31}	value	1.8062		
	Std. Error	0.4655		
γ_{32}	value	1.2615		
	Std. Error	0.2661		
loglikelihood			-924.4926	-924.4912

Table 5.10: Method Comparison under Case 6

		glm() function	Newton-Raphson	quasi-Newton
α_1	value	-10.3369	-10.1998	-10.1091
	Std. Error	1.7149	1.6367	1.0864
α_2	value	-8.0159	-7.9365	-7.7579
	Std. Error	1.0736	1.0620	1.0129
α_3	value	-1.0627	-1.0352	-0.9465
	Std. Error	0.5241	0.5186	0.9363
α_4	value	-3.2307	-3.1701	-3.0885
	Std. Error	0.8946	0.8842	0.9967
β_{11}	value	0.0800	0.0752	0.0739
	Std. Error	0.0242	0.0230	0.0181
β_{12}	value	0.2847	0.2913	0.3136
	Std. Error	0.3843	0.3753	0.7905
β_{13}	value	0.6403	0.9493	0.9395
	Std. Error	0.2373	0.2790	0.3493
β_{21}	value	0.0770	0.0758	0.0732
	Std. Error	0.01554	0.0154	0.0157
β_{22}	value	-0.1451	-0.1271	-0.1166
	Std. Error	0.2807	0.2793	0.3682
β_{23}	value	-0.3320	-0.3429	-0.3420
	Std. Error	0.1887	0.2012	0.3460

Table 5.11: Method Comparison under Case 6 (continued)

		glm() function	Newton-Raphson	quasi-Newton
β_{31}	value	0.0021	0.0014	0.0001
	Std. Error	0.0084	0.0083	0.0148
β_{32}	value	0.1598	0.1603	0.1629
	Std. Error	0.1890	0.1888	0.2271
β_{33}	value	-1.6840	-1.6642	-1.6608
	Std. Error	0.1331	0.1313	0.1411
β_{41}	value	-0.0008	-0.0018	-0.0027
	Std. Error	0.0141	0.0139	0.0165
β_{42}	value	-0.2366	-0.2292	-0.2357
	Std. Error	0.3350	0.3342	0.9272
β_{43}	value	-0.4165	-0.4903	-0.5051
	Std. Error	0.2195	0.2291	0.3361
γ_{12}	value	1.2000	1.1181	1.1284
	Std. Error	0.4157	0.4216	0.9704
γ_{13}	value	2.3946	2.7162	2.7072
	Std. Error	0.4760	0.5087	0.5792
γ_{14}	value	0.1531	0.0198	0.0403
	Std. Error	0.6662	0.6542	0.9719

Table 5.12: Method Comparison under Case 6 (continued)

		glm() function	Newton-Raphson	quasi-Newton
γ_{23}	value	1.0296	0.9935	0.9925
	Std. Error	0.3040	0.3076	0.6812
γ_{24}	value	-0.2330	-0.1803	-0.1831
	Std. Error	0.4875	0.4765	0.9951
γ_{34}	value	1.1814	1.1522	1.114
	Std. Error	0.3862	0.3446	0.6171
γ_{21}	value	1.0650		
	Std. Error	0.4304		
γ_{31}	value	2.8826		
	Std. Error	0.5216		
γ_{32}	value	1.0161		
	Std. Error	0.3135		
γ_{41}	value	-0.0134		
	Std. Error	0.6564		
γ_{42}	value	-0.170		
	Std. Error	0.4762		
γ_{43}	value	1.1795		
	Std. Error	0.3533		
loglikelihood			-994.3497	-994.4018

5.4.1 Odds Ratio Analysis

When the variables are binary, a natural way to measure the association is the odds ratio (as well as its logarithm). Consider two binary variables Y_1 and Y_2 , where Y_i is coded 1 and 0. Let $p_{ij} = \Pr(Y_1 = i, Y_2 = j)$, $i, j = 0, 1$, so that the odds ratio, ψ , is defined as:

$$\psi = \frac{p_{00}p_{11}}{p_{01}p_{10}}$$

From a random sample of size N , we generate the following 2×2 contingency table with $N = a + b + c + d$:

	$Y_1 = 0$	$Y_1 = 1$
$Y_2 = 0$	a	b
$Y_2 = 1$	c	d

The estimated success probabilities in the two data sets are $\hat{p}_{00} = a/N$, $\hat{p}_{10} = b/N$, $\hat{p}_{01} = c/N$, and $\hat{p}_{11} = d/N$, and so the estimated odds ratio, $\hat{\psi}$, is given by

$$\hat{\psi} = \frac{\hat{p}_{00}\hat{p}_{11}}{\hat{p}_{10}\hat{p}_{01}} = \frac{ad}{bc}.$$

An interval estimate or confidence interval (CI) of the odds ratio is more useful than a point estimate for measuring the association. Also we note that the logarithm of the estimated odds ratio is better approximated by a normal distribution. The approximate standard error of the estimated log odds ratio, $\log \psi$, can be shown to be given by

$$s.e.(\log \hat{\psi}) \approx \sqrt{\left(\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}\right)}$$

An outline proof of this result is given in Schlesselman (1982). An approximate $100(1 - \alpha)\%$ confidence interval for ψ is

$$\exp[\log \hat{\psi} \pm z_{\alpha/2} s.e.(\log \hat{\psi})]$$

where Z_β is the upper β quantile of the standard normal distribution (see Hosmer and Lemeshow 1989).

Table 5.13, Table 5.14 and Table 5.15 provide the odds ratios for response variable REM (V21), NEM (V22) and PUM (V23) respectively. For more detailed information, see Zhang (1993).

Statistically, only those 95% CI not containing 1 are more strongly related with the variable. Using odds ratio analysis, we conclude that covariates REN (V7), LIV (V11), CNS (V12), COP (V5), DIA (V6), and HTN (V8) are strongly related with the variable REM (V21); covariates CNS (V12), CA (V10), OBE (V4), REN (V7), and HTN (V8) are strongly related with the variable NEM (V22); covariates COP (V5), OBE (V4), REN (V7), ETO (V9), CNS (V12), DIA (V6), and HTN (V8) are strongly related with the variable PUM (V23).

5.4.2 Univariate Analysis and Comparison of Models

An alternative for deciding the covariates to use is based on the differences of log-likelihoods. The difference in the log-likelihoods of two nested models measures the extent to which the additional terms improve the fit of the model to the observed response variables. Take model (3.30) which has n binary response variables, for example. Suppose that two linear logistic models, model (1) and model (2), say, are to be compared, where the two models are as follows:

- Model (1): $\text{logit Pr}(Y_i = 1 | y_j, j \neq i, x_{ik}, 1 \leq k \leq m-1) = \alpha_i + \sum_{j=1}^{m-1} \beta_{ij} x_{ik} + \sum_{i \neq j} \gamma_{ij} y_j, i = 1, \dots, n$
- Model (2): $\text{logit Pr}(Y_i = 1 | y_j, j \neq i, x_{ik}, 1 \leq k \leq m) = \alpha_i + \sum_{j=1}^m \beta_{ij} x_{ik} + \sum_{i \neq j} \gamma_{ij} y_j, i = 1, \dots, n$

Table 5.13: Odds Ratio for REM

Variable	Heading	$\hat{\psi}$	$s.e.(\log \hat{\psi})$	95 % CI of $\hat{\psi}$	Remark
V2	SEX	1.48	0.21	(0.97, 2.26)	
V4	OBE	1.13	0.37	(0.54, 2.37)	
V5	COP	1.86	0.26	(1.11, 3.11)	positively related
V6	DIA	1.80	0.24	(1.11, 2.91)	positively related
V7	REN	9.78	0.22	(6.30, 15.2)	positively related
V8	HTN	1.75	0.20	(1.18, 2.62)	positively related
V9	ETO	1.34	0.47	(0.53, 3.95)	
V11	LIV	3.58	0.62	(1.04, 12.3)	positively related
V12	CNS	3.40	0.33	(1.75, 6.61)	positively related
V13	PCA	1.33	0.60	(0.40, 4.34)	
V14	RHE	0.88	0.59	(0.27, 2.85)	
V15	OTH	0.90	0.52	(0.32, 2.50)	
V17	VAL	0.64	1.01	(0.08, 4.76)	

Table 5.14: Odds Ratio for NEM

Variable	Heading	$\hat{\psi}$	$s.e.(\log \hat{\psi})$	95 % CI of $\hat{\psi}$	Remark
V2	SEX	1.34	0.15	(0.99, 1.81)	
V4	OBE	2.11	0.21	(1.37, 3.25)	positively related
V5	COP	1.22	0.21	(0.80, 1.84)	
V6	DIA	1.38	0.18	(0.96, 1.98)	
V7	REN	1.62	0.23	(1.02, 2.57)	positively related
V8	HTN	1.49	0.14	(1.12, 1.97)	positively related
V9	ETO	1.40	0.32	(0.73, 2.68)	
V10	CA	2.96	0.43	(1.26, 6.93)	positively related
V11	LIV	2.20	0.55	(0.74, 6.53)	
V12	CNS	3.92	0.24	(2.41, 6.39)	positively related
V13	PCA	2.02	0.37	(0.97, 4.19)	
V14	RHE	0.81	0.43	(0.34, 1.88)	
V15	OTH	0.94	0.35	(0.47, 1.90)	
V17	VAL	0.58	0.73	(0.14, 2.46)	

Table 5.15: Odds Ratio for PUM

Variable	Heading	$\hat{\psi}$	$s.e.(\log \hat{\psi})$	95 % CI of $\hat{\psi}$	Remark
V2	SEX	1.08	0.09	(0.90, 1.30)	
V4	OBE	3.38	0.16	(2.45, 4.66)	positively related
V5	COP	3.45	0.13	(2.66, 4.48)	positively related
V6	DIA	1.91	0.11	(1.52, 2.40)	positively related
V7	REN	3.32	0.16	(2.41, 4.58)	positively related
V8	HTN	1.86	0.08	(1.57, 2.20)	positively related
V9	ETO	2.21	0.21	(1.46, 3.36)	positively related
V10	CA	0.56	0.41	(0.25, 1.25)	
V12	CNS	2.12	0.21	(1.39, 3.21)	positively related
V13	PCA	0.67	0.29	(0.37, 0.91)	
V14	RHE	0.97	0.23	(0.61, 1.54)	
V15	OTH	0.58	0.22	(0.37, 0.91)	
V17	VAL	1.03	0.34	(0.52, 2.01)	

Denote the maximized log-likelihoods under model (1) and model (2) by L_1 and L_2 , respectively, so that twice the difference in the log-likelihoods of two models is

$$2[L_2 - L_1]$$

Asymptotically (large sample size N), the distribution of this statistic has a χ^2 distribution with degrees of freedom $\nu_1 - \nu_2$ equal to the difference in the number of the parameters between two models (see Cox and Hinkley (1974)). In our example, since we have n binary response variables, that is, $\nu_1 - \nu_2 = n$ when adding an extra covariate x_m to model (1) which already has covariates x_1, x_2, \dots, x_{m-1} . Using the difference between the log-likelihood functions of two models, we can find out how much the covariate x_m improves the fit of the model to the observed response variables.

Other methods for selection of covariates will be introduced in the next section.

5.5 Other Issues

Before concluding this chapter, another question to discuss is how much the association of the response variables depends on the covariates. This is one inference of interest. Another reason that this issue arises is that some methods, for example, Generalized Estimating Equations, seem to assume the correlation of two response variables does not depend on covariates. What is the case for our data?

5.5.1 General

As we know, with binary variables, the odds ratio (as well as its logarithm) is a natural measure of association. Considering the 2 dimensional case for model (3.30) where there exists only two binary response variables, the odds ratio is

$$\frac{\Pr(1, 1|\mathbf{x}) \Pr(0, 0|\mathbf{x})}{\Pr(1, 0|\mathbf{x}) \Pr(0, 1|\mathbf{x})} = \frac{\exp[(\alpha_1 + \beta_1\mathbf{x}) + (\alpha_2 + \beta_2\mathbf{x}) + \gamma_{12}]}{\exp(\alpha_1 + \beta_1\mathbf{x}) \exp(\alpha_2 + \beta_2\mathbf{x})} = \exp(\gamma_{12}).$$

Hence, the logarithm of the odds ratio is γ_{12} and it does not depend on the covariates.

However, it is not true that the odds ratio does not depend on the covariates when the number of binary response variables is greater than 2. Take the 3 dimensional case in model (3.30), for a particular instance. Based on model (3.30), the joint distribution of Y_i and Y_j is

$$\begin{aligned} \Pr(y_i, y_j | \mathbf{x}) &= \frac{\Pr(y_i, y_j, y_k | \mathbf{x})}{\Pr(y_k | y_i, y_j, \mathbf{x})} \\ &= \frac{c^{-1} \exp\{\sum_{i=1}^3 (\alpha_i + \sum_{t=1}^m \beta_{it} x_t) y_i + \sum_{1 \leq i < j \leq 3} \gamma_{ij} y_i y_j\}}{\exp\{(\alpha_k + \sum_{t=1}^m \beta_{kt} x_t + \gamma_{ik} y_i + \gamma_{jk} y_j) y_k\} / 1 + \exp(\alpha_k + \sum_{t=1}^m \beta_{kt} x_t + \gamma_{ik} y_i + \gamma_{jk} y_j)} \\ &= c^{-1} [1 + \exp(\alpha_k + \sum_{t=1}^m \beta_{kt} x_t + \gamma_{ik} y_i + \gamma_{jk} y_j)] \\ &\quad \cdot \exp\{(\alpha_i + \sum_{t=1}^m \beta_{it} x_t) y_i + (\alpha_j + \sum_{t=1}^m \beta_{jt} x_t) y_j + \gamma_{ij} y_i y_j\} \end{aligned}$$

for (i, j, k) , where $c = 1 + \sum_{i=1}^3 \exp(\alpha_i + \sum_{t=1}^m \beta_{it} x_t) + \sum_{1 \leq i \neq j \leq 3} \exp[(\alpha_i + \alpha_j) + \sum_{t=1}^m (\beta_{it} + \beta_{jt}) x_t + \gamma_{ij}] + \exp[\sum_{i=1}^3 \alpha_i + \sum_{t=1}^m \sum_{i=1}^3 \beta_{it} x_t + \sum_{i=1}^3 \gamma_{ij}]$.

As a result, the odds ratio for Y_i and Y_j is

$$\begin{aligned} &\frac{\Pr(Y_i = 0, Y_j = 0 | \mathbf{x}) \Pr(Y_i = 1, Y_j = 1 | \mathbf{x})}{\Pr(Y_i = 0, Y_j = 1 | \mathbf{x}) \Pr(Y_i = 1, Y_j = 0 | \mathbf{x})} \\ &= \frac{[1 + \exp(\alpha_k + \sum_{t=1}^m \beta_{kt} x_t)][1 + \exp(\alpha_k + \sum_{t=1}^m \beta_{kt} x_t + \gamma_{ki} + \gamma_{kj})]}{[1 + \exp(\alpha_k + \sum_{t=1}^m \beta_{kt} x_t + \gamma_{ki})][1 + \exp(\alpha_k + \sum_{t=1}^m \beta_{kt} x_t + \gamma_{kj})]} \exp(\gamma_{ij}) \end{aligned} \quad (5.1)$$

This form implies that when there exist three binary response variables in model (3.30), the odds ratio for arbitrary two binary response variables indeed depends on the covariates \mathbf{x} . Using similar steps as above, we can get the same conclusions for the case of more than three binary response variables in model (3.30).

Next, we take two examples to illustrate how the association of binary response variables varies with covariates for our data. In each example, we will separately consider model (3.30) with

- one covariate: COP (V5);
- two covariates: COP (V5), and CNS (V12);
- three covariates: COP (V5), CNS (V12), and REN (V7);
- four covariates: COP (V5), CNS (V12), REN (V7), and HTN (V8);
- five covariates: COP (V5), CNS (V12), REN (V7), HTN (V8), and OBE (V4).

5.6 Examples

5.6.1 Example 1

In our first example, only two complication variables, REM (V21) and NEM (V22) are considered as the response variables in model (3.30). After using the Newton-Raphson method to fit model (3.30), the log-odds ratio for REM and NEM is simply equal to the estimate value of the parameter γ_{12} . Table 5.16 presents the log-odds ratio for REM, NEM and its standard error (showed in parenthesis) under different number of covariates. From Table 5.16, we find that the log-odds ratio for REM and NEM decreases when we add an extra covariate. As a matter of fact, we say the association of two binary response variables varies with covariates, but it is not significantly so when considering the standard errors.

Considering the confidence interval of the odds ratio for REM and NEM, we find that all 95% confidence intervals under the above five cases do not contain 1 (see Table 5.17). This implies that variable REM and variable NEM are strongly related each other whatever covariates we choose.

Now let us return to the topic of Section 5.4. We mentioned that the odds ratio analysis and another approach based on the differences of log-likelihoods, can suggest which

Table 5.16: The Association of REM (V21) and NEM (V22)

Covariates	Log-Odds Ratio	Std. Error
COP(V5)	1.9296	0.3909
COP CNS(V12)	1.7422	0.4070
COP CNS REN(V7)	1.7005	0.4180
COP CNS REN HTN(V8)	1.6938	0.4184
COP CNS REN HTN OBE(V4)	1.6842	0.4219

Table 5.17: CI of Odds Ratio for REM (V21) and NEM (V22)

Covariates	95% CI of odds ratio
COP(V5)	(3.2010, 14.8166)
COP CNS(V12)	(2.5715, 12.6785)
COP CNS REN(V7)	(2.4139, 12.4257)
COP CNS REN HTN(V8)	(2.3959, 12.3524)
COP CNS REN HTN OBE(V4)	(2.3568, 12.3186)

covariates to use. Here introduce two related methods: the first method is to use the ratios of the estimates of the coefficients for the covariates to their standard errors: $\hat{\beta}/s.e.(\hat{\beta})$; the second uses differences in the values of log-likelihood function to assess covariates.

For different number of covariates, Table 5.18 and Table 5.19 list the estimates of the coefficients for the covariates (β 's), corresponding standard errors and their ratios when applying the Newton-Raphson method to fit model (3.30). Statistically, if the absolute value of a ratio is large than 1.96, this might suggest that we need to consider the variable of this coefficient as our covariate in model (3.30).

After checking the ratios in Table 5.18 and Table 5.19, we conclude

1. Among the five covariates: COP (V5), CNS (V12), REN (V7), HTN (V8) and OBE (V4), covariate COP and covariate REN are most important for response variable REM.
2. For response variable NEM, covariate CNS and covariate OBE are more important than other three covariates: COP, REN and HTN.
3. Covariate HTN seems to be less important than other covariates for both response variable REM and response variable NEM. HTN appears not be significant when several covariates are included, possibly because it is strongly associated with the other covariates.

In last section, we know, from Table 5.13 and Table 5.14, that of the five covariates, four of them: COP, CNS, REN and HTN are individually strongly related with the response variable REN; four of them: CNS, REN, HTN and OBE are individually strongly related with the response variable NEM. Comparing with the conclusion from checking the ratios

Table 5.18: The Estimates of the Coefficients under All Five Cases

$Y_1 = \text{REM}$		Number of Covariates				
		1	2	3	4	5
COP (V5)	β_{11}	1.5516	1.5344	1.4579	1.4532	1.4644
	Std. Error	0.3982	0.4021	0.4128	0.4119	0.4110
	Ratio	3.8965	3.8160	3.5317	3.5280	3.5630
CNS (V12)	β_{12}		1.2536	0.9634	0.9432	0.9569
	Std. Error		0.5372	0.5666	0.5668	0.5637
	Ratio		2.3336	1.7003	1.6641	1.6975
REN (V7)	β_{13}			1.5753	1.5471	1.5515
	Std. Error			0.4242	0.4286	0.4272
	Ratio			3.7136	3.6097	3.6318
HTN (V8)	β_{14}				0.1658	0.1567
	Std. Error				0.3705	0.3722
	Ratio				0.4475	0.4210
OBE (V4)	β_{15}					0.1593
	Std. Error					0.5358
	Ratio					0.2973

Table 5.19: The Estimates of the Coefficients under All Five Cases (continued)

$Y_2=NEM$		Number of Covariates				
		1	2	3	4	5
COP (V5)	β_{21}	-0.2800	-0.3023	-0.3144	-0.3228	-0.3278
	Std. Error	0.4183	0.4251	0.4270	0.4258	0.4240
	Ratio	0.6694	-0.7111	-0.7363	-0.7581	-0.7731
CNS (V12)	β_{22}		1.5045	1.4802	1.4662	1.4477
	Std. Error		0.4082	0.4121	0.4130	0.4161
	Ratio		3.6857	3.5918	3.5501	3.4792
REN (V7)	β_{23}			0.1909	0.1489	0.1158
	Std. Error			0.4117	0.4154	0.4155
	Ratio			0.4637	0.3584	0.2787
HTN (V8)	β_{24}				0.2050	0.1244
	Std. Error				0.2545	0.2582
	Ratio				0.8055	0.4818
OBE (V4)	β_{25}					0.9050
	Std. Error					0.3432
	Ratio					2.6369

Table 5.20: The Log-likelihood Functions with Different Number of Covariates

Covariates	Log-likelihood
COP(V5)	-387.4225
COP CNS(V12)	-377.1679
COP CNS REN(V7)	-370.5885
COP CNS REN HTN(V8)	-370.0956
COP CNS REN HTN OBE(V4)	-366.7212

in Table 5.18 and Table 5.19, partly we can say that to choose covariates, checking the ratios of the estimates of the coefficients for the covariates to their standard errors is an additional method to the odds ratio analysis.

In the meanwhile, we also calculate the log-likelihood function when response variables are two complication variables, REM (V21) and NEM (V22), and covariates are one of the above five cases (see Table 5.20). Table 5.20 shows that the log-likelihood function increases when we add an extra covariate. Especially, we find that when we add covariate HTN to covariates COP, CNS and REN, the log-likelihood function only increases by 0.4929, less than other increase. Again, this may suggest that covariate HTN is less important than other covariates for both response variables.

To conclude, using multivariate binary model (3.30), we can do the risk factor analysis for several response variables (or complication variables) simultaneously.

5.6.2 Example 2

Now, we briefly discuss another example for checking the 3 dimensional case where there exist three binary response variables, REM (V21), NEM (V22), and PUM (V23) in model (3.30).

Table 5.21: The Log-likelihood Functions of Three Response Variables with Different Number of Covariates

Covariates	Log-likelihood	Difference
COP(V5)	-925.7063	
COP(V5) CNS(V12)	-910.6644	15.0419
COP(V5) CNS(V12) REN(V7)	-889.7837	20.8807
COP(V5) CNS(V12) REN(V7) HTN(V8)	-882.5742	7.2095
COP(V5) CNS(V12) REN(V7) HTN(V8) OBE(V4)	-873.7382	8.8360

In Table 5.21, we present the values of log-likelihoods under different number of covariates when using the Newton-Raphson method and the differences of log-likelihoods when adding another covariate to previous case. Clearly, as previous example the log-likelihood function increases when we add an extra covariate. Considering differences, it seems that the last two covariates HTN and OBE are not more important than other covariates, but comparing with χ_3^2 , they are still statistically significant.

We choose the third case: three covariates COP, CNS and REN for our final analysis. Similar to Table 5.18 and Table 5.19, Table 5.22 deals with three binary response variable REM, NEM, PUM and three covariates COP, CNS, REN in model (3.30). Checking the ratios in Table 5.22, we say that among these three covariates, CNS is not significant for binary response variable REM; COP and REN is not significant for binary response variable NEM.

After using the Newton-Raphson method to fit model (3.30) for the third case, in addition to the estimates of the coefficients listed in Table 5.22, we also obtained other estimates of the parameters in model (3.30). Therefore, through equation (5.1) the

Table 5.22: The Estimates of the Coefficients under the Third Case

		REM (V21)	NEM (V22)	PUM (V23)
COP (V5)	Coefficient	1.1318	-0.5786	1.3013
	Std. Error	0.4189	0.4254	0.2447
	Ratio	2.7018	-1.3601	5.3179
CNS (V12)	Coefficient	0.7580	1.2370	1.1185
	Std. Error	0.5554	0.4180	0.4131
	Ratio	1.3648	2.9593	2.7076
REN (V7)	Coefficient	1.2210	-0.1104	1.5234
	Std. Error	0.4280	0.4110	0.2988
	Ratio	2.8528	0.2686	5.0984

marginal odds ratio for REM and NEM is the form of

$$\begin{aligned}
 & \frac{[1 + \exp(\alpha_3 + \sum_{j=1}^3 \beta_{3j}x_j)][1 + \exp(\alpha_3 + \sum_{j=1}^3 \beta_{3j}x_j + \gamma_{13} + \gamma_{23})]}{[1 + \exp(\alpha_3 + \sum_{j=1}^3 \beta_{3j}x_j + \gamma_{13})][1 + \exp(\alpha_3 + \sum_{j=1}^3 \beta_{3j}x_j + \gamma_{23})]} \exp(\gamma_{12}) \\
 = & \exp(1.4036) \cdot \frac{1 + \exp(-0.9899 + 1.3013x_1 + 1.1185x_2 + 1.5234x_3)}{1 + \exp(-0.9899 + 1.3013x_1 + 1.1185x_2 + 1.5234x_3 + 1.5186)} \\
 & \cdot \frac{1 + \exp(-0.9899 + 1.3013x_1 + 1.1185x_2 + 1.5234x_3 + 1.5186 + 1.1775)}{1 + \exp(-0.9899 + 1.3013x_1 + 1.1185x_2 + 1.5234x_3 + 1.1775)} \\
 = & \frac{1 + \exp(-0.9899 + 1.3013x_1 + 1.1185x_2 + 1.5234x_3)}{1 + \exp(1.7062 + 1.3013x_1 + 1.1185x_2 + 1.5234x_3)} \\
 & \cdot \frac{1 + \exp(0.1687 + 1.3013x_1 + 1.1185x_2 + 1.5234x_3)}{1 + \exp(0.1876 + 1.3013x_1 + 1.1185x_2 + 1.5234x_3)} \cdot 4.0698 \tag{5.2}
 \end{aligned}$$

As a function of the covariates COP, CNS and REN, the marginal odds ratio for REM and NEM varies with the covariates for model (3.30). In fact, the association of the response variables depends strongly on the covariates in this 3 dimensional case.

From Table 5.1 and Table 5.2, we know all covariates COP, CNS and REN are binary variables. Substituting different values of the covariates for equation (5.2), we give Table 5.23 to show the changes of the marginal odds ratio for REM and NEM. From

Table 5.23: The Marginal Odds Ratio for REM and NEM

COP (V5)	CNS (V12)	REN (V7)	Marginal Odds Ratio
$x_1 = 0$	$x_2 = 0$	$x_3 = 0$	0.8490
$x_1 = 1$	$x_2 = 0$	$x_3 = 0$	0.4464
$x_1 = 0$	$x_2 = 1$	$x_3 = 0$	0.4800
$x_1 = 0$	$x_2 = 0$	$x_3 = 1$	0.4124
$x_1 = 1$	$x_2 = 1$	$x_3 = 0$	0.3291
$x_1 = 1$	$x_2 = 0$	$x_3 = 1$	0.3094
$x_1 = 0$	$x_2 = 1$	$x_3 = 1$	0.3173
$x_1 = 1$	$x_2 = 1$	$x_3 = 1$	0.2826

these marginal odds ratios in Table 5.23, we conclude that of three covariates, the most important covariate for the changes of the marginal odds ratio for REM and NEM is REN, next is COP.

Bibliography

- [1] Arnold, Barry C., Castillo, Enrique and Sarabia, José-Mariá (1992). Conditionally Specified Distributions. Springer-Verlag.
- [2] Arnold, B.C. and Strauss, D. (1988). Bivariate distributions with exponential conditionals. *JASA* **83**, 522-527.
- [3] Bonney, George Ebow (1987). Logistic regression for dependent binary observations. *Biometrics* **43**, 951-973.
- [4] Chambers, John M., Hastie, Trevor J. and AT&T Bell Laboratories (1992). Statistical Models in S. Wadsworth & Brooks / Cole Advanced Books & Software Pacific Grove, California.
- [5] Collett, D. (1994). Modelling Binary Data. Chapman & Hall.
- [6] Nash, J C (1990). Compact Numerical Methods For Computers *Linear Algebra and Function Minimisation*. Second Edition. Adam Hipger, Bristol and New York.
- [7] Connolly, Margaret A. and Liang, Kung-Yee (1988). Conditional logistic regression models for correlated binary data. *Biometrika* **75**, 3, 501-506.
- [8] Cox, D.R. and Hinkley, D.V. (1974). Theoretical Statistics. Chapman Hall, London.
- [9] Gelman, A. and Speed, T.P. (1993). Characterizing a joint probability distribution by conditionals. *J. Roy. Statist. Soc. B* **55** , 185-188.

- [10] Hosmer, D.W. and Lemeshow, S. (1989). *Applied Logistic Regression*. Wiley, New York.
- [11] McCullagh, P., and Nelder, J.A. (1989). *Generalized Linear Models*. Second Edition. Chapman Hall, London.
- [12] Press, W.H., Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T. (1988). *Numerical Recipes in C: the Art of Scientific Computing*. Cambridge University Press, Cambridge, Eng.
- [13] Schlesselman, J.J. (1982). *Case-Control Studies: Design, Conduct, Analysis*. Oxford University Press, New York.
- [14] Zhang, Hongbin (1993). *Risk prediction models for binary response variable for the coronary bypass operation (M.Sc. Thesis)*. Dept. of Statistics, University of British Columbia.

Appendix

C Program for Applying the Newton-Raphson Method

```
#include <stdio.h>
#include <math.h>
#define N 900
#define NP 150
#define NC 20
main()
{
    double t[NP][NP],h[NP][NP],s[N][NP],x[N][NC],y[N][NC];
    double c[N],lm[NP],suf[NP];
    double dif,mx,prod,sum,lik,logc;
    int iter,i,j,k,d,n,m,nn,kf,ii,M,M1,jj[6];
    double **a,**b,**dmatrix();
    FILE *fp;
    void d2b(),gaussj();

    if ((fp = fopen("name","r")) == NULL)
    { printf("file not found\n"); exit();
    }

    scanf("%d %d %d", &n, &m, &nn);
    /*
        n is dimension of binary response variables
        m is dimension of covariates
        nn is the number of observation
    */

    M=n+n*m+n*(n-1)/2; M1=M+1;
    /*
        M is the number of parameters
        Next step is to get initial values of the parameters
    */
}
```

```

        and read values for binary response variables
        and covariates from a file "name"
*/

for(i=1;i<=M;i++) fscanf(fp,"%F",&lm[i]);
for(i=1;i<=M;i++) printf("%6.3f", lm[i]); printf("\n");

for(k=1;k<=nn;k++)
{   for(i=1;i<=m;i++) fscanf(fp,"%F",&x[k][i]);
    for(j=1;j<=n;j++) fscanf(fp,"%F",&y[k][j]);
}
printf("\n");

for(i=1,kf=1;i<=n;i++) kf*=2;
printf("n=%d, m=%d, nn=%d, M=%d, kf=%d\n", n,m,nn,M,kf);

for(i=1;i<=M;i++) suf[i]=0.;
/*
    Next, calculate sufficient statistics -- suf
*/
for(k=1;k<=nn;k++)
{   for(i=1;i<=n;i++)
    {   s[k][((i-1)*m+i)]=y[k][i];
        for(j=1;j<=m;j++) s[k][((i-1)*m+i+j)]=y[k][i]*x[k][j];
    }
    for(i=1;i<=n-1;i++)
    {   for(j=i+1;j<=n;j++) s[k][n+n*m+n*(i-1)-i*(i-1)/2+j-i]
        =y[k][i]*y[k][j];
    }
}

for(i=1;i<=M;i++)
{   for(k=1;k<=nn;k++) suf[i]+=s[k][i];
    printf("%8.4f", suf[i]);
    if(i%10==0) printf("\n");
}

```

```

}
printf("\n");

dif=1.; iter=0;
a = dmatrix(1,M,1,M); b = dmatrix(1,M,1,1);
while(iter<30 && dif>1.e-4)
{ for(i=1;i<=M;i++)
  { for(j=1;j<=M1;j++) h[i][j]=0.0;
  }
  for(k=1;k<=nn;k++)
  { c[k]=0.;
    for(i=1;i<=M;i++)
    { for(j=1;j<=M1;j++) t[i][j]=0.0;
    }
    /*
    calls routine, d2b(n,ii,jj): returns possible binary vector
    jj of dimension n
*/

    for(ii=1;ii<=kf;ii++)
    { d2b(n,ii,jj);
      for(i=1;i<=n;i++)
      { s[k][i+(i-1)*m]=(double)jj[i];
        for(j=1;j<=m;j++) s[k][(i-1)*m+i+j]
          =(double)(jj[i])*x[k][j];
      }
      for(i=1;i<=n-1;i++)
      { for(j=i+1;j<=n;j++) s[k][n+n*m+n*(i-1)-i*(i-1)/2+j-i]
          =(double)jj[i]*jj[j];
      }
      for(i=1,sum=0.;i<=M;i++) sum+=lm[i]*s[k][i];
      prod=exp(sum); c[k]+=prod;
      for(i=1;i<=M;i++) t[i][M1]+=s[k][i]*prod;
      for(i=1;i<=M;i++)
      { for(j=1;j<=M;j++) t[i][j]+=prod*s[k][i]*s[k][j];
      }

```

```

    }
    for(i=1;i<=M;i++) t[i][M1]/=c[k];
    for(i=1;i<=M;i++)
    { for(j=1;j<=M;j++) t[i][j]=t[i][M1]*t[j][M1]-t[i][j]/c[k];
    }
    for(i=1;i<=M;i++)
    { for(j=1;j<=M+1;j++) h[i][j]+=t[i][j];
    }
}
for(i=1;i<=M;i++)
{ for(j=1;j<=M;j++) a[i][j]=h[i][j];
  b[i][1]=suf[i]-h[i][M1];
}
/*
        calls routine, gaussj(a,M,b,1)
        for solving a linear system
*/
gaussj(a,M,b,1);
for(i=1,dif=0.;i<=M;i++)
{ mx=fabs(b[i][1]); if(mx>dif) dif=mx;
}
for(i=1;i<=M;i++) lm[i]-=b[i][1];
iter++;
if(iter%1==0)
{ for(i=1;i<=M;i++)
  { printf("lm=%8.4f", lm[i]);
    if(i%10==0) printf("\n");
  }
  printf("iter=%4d\n",iter);
}
}
if(iter>=30) printf("did not converge\n");
for(i=1;i<=M;i++)
{ printf("%8.4f %8.4f %8.4f\n",lm[i],sqrt(-a[i][i]),
  lm[i]/sqrt(-a[i][i]));
}

```

```

}
printf("iter=%4d\n",iter);

for(lik=0.,i=1;i<=M;i++) lik+=suf[i]*lm[i];
for(i=1,logc=0.;i<=nn;i++) logc+=log(c[i]);
lik-=logc;
printf("loglikelihood= %9.4f\n", lik);
}

```

```

void d2b(n,ii,jj)
int n,ii,jj[];
{
    int i,d;
    d=ii-1;
    for(i=n;i>=1;i--)
    {
        jj[i]=d%2;
        d=d/2;
    }
}

```