Spring 2005

Statistical Analysis of Longitudinal and Multivariate Discrete Data

Deepak Mav
Old Dominion University

STATISTICAL ANALYSIS OF LONGITUDINAL AND MULTIVARIATE DISCRETE DATA

by

Deepak Mav
B.S. May 1997, University of Mumbai
M.S. August 1999, Indian Institute of Technology
M.S. May 2004, Old Dominion University

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirement for the Degree of

DOCTOR OF PHILOSOPHY

MATHEMATICS AND STATISTICS

OLD DOMINION UNIVERSITY
May 2005

Approved by:

N. Rao Chaganty (Director)

Davanand Naik

Michael Doviak

Larry Filer

ABSTRACT

STATISTICAL ANALYSIS OF LONGITUDINAL AND MULTIVARIATE DISCRETE DATA

Deepak Mav
Old Dominion University, 2005
Director: Dr. N. Rao Chaganty

Correlated multivariate Poisson and binary variables occur naturally in medical, biological and epidemiological longitudinal studies. Modeling and simulating such variables is difficult because the correlations are restricted by the marginal means via Fréchet bounds in a complicated way. In this dissertation we will first discuss partially specified models and methods for estimating the regression and correlation parameters. We derive the asymptotic distributions of these parameter estimates. Using simulations based on extensions of the algorithm due to Sim (1993, Journal of Statistical Computation and Simulation, 47, pp. 1-10), we study the performance of these estimates using infeasibility, coverage probabilities of the confidence ellipsoids, and asymptotic relative efficiencies as the criteria.

The second part of this dissertation is devoted to the study of fully specified models constructed using copulas, with special emphasis on the normal copula. Finding the maximum likelihood estimates and the Fisher information matrix for these models requires computation of multivariate normal probabilities. We also discuss several efficient algorithms for calculating multivariate normal integrals. For the multivariate probit and multivariate Poisson log-normal models, we implement maximum likelihood, derive the necessary equations, and illustrate it on two real life data sets. Next we study over and under dispersed models including quasi-multinomial and Lagrange families of distributions. We implement the maximum likelihood method for the quasi-multinomial model and illustrate the application of this model for market analysis of household preferences for saltine crackers.

ACKNOWLEDGMENTS

I wish to express deepest gratitude to my advisor Professor N. Rao Chaganty for numerous suggestions and discussions during the development of this dissertation. This laborious task would not have been completed without his continuous support and help, for which I am very grateful.

I would like to thank Professors Dayanand Naik, Michael Doviak and Larry Filer for serving on my dissertation committee. Thanks are also due to Dr. Naik for his immense encouragement, and Dr. Dahiya for arranging research support for me at the Eastern Virginia Medical School. Working alongside with Professors Michael Doviak and Trent Buskirk on several statistical consulting projects at the Eastern Virginia Medical School was a rewarding and educational experience for me. I thank them both for their support and mentorship. Special thanks to Ms. Jacquelyn LaCoste, Program Analyst at Constella Health Sciences for proof reading this dissertation.

Last but not the least, I would also like to thank my fellow students, friends, professors and staff members in the Department of Mathematics and Statistics at Old Dominion University, especially Professor Dorrepaal, for providing a pleasant and stimulating environment. Finally, I would like to thank the Department of Mathematics and Statistics for the financial support for my graduate studies.

I dedicate this thesis to my parents.

TABLE OF CONTENTS

CHAPTERS Page

CHAPTERS Page

APPENDIX

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Statistical methods for analyzing longitudinal and multivariate data are important tools in data analysis. They are used on data consisting of several measurements made on each individual or an experimental unit. Such data occur commonly in scientific disciplines including biology, medicine, psychology, business and many other fields. Historically, the majority of applications have been on multivariate measurements that are continuous and often normally distributed. However, in recent years, many researchers have taken up the challenging task of modifying and tailoring these methods for the analysis of repeated measurements or longitudinal data. The term "repeated or longitudinal data" refers to data consisting of responses taken on subjects or experimental units at different time points or on multiple occasions or under multiple conditions.

While the longitudinal data can be viewed as multivariate data, there are some key differences between the two. Multivariate data normally consists of a snapshot of different variables taken at a single time point, whereas longitudinal data consists of snapshots of the same variable taken at multiple time points. Therefore, the correlation structure in longitudinal data is different from the structure that arises in multivariate data. Thus, even though there are some similarities between the two types of data, each pose different challenges and require different approaches for statistical analysis. Much research has already been done on multivariate and longitudinal data when the response variables are continuous. Unlike the Gaussian distribution that is widely used to model continuous variables, there is not a single multivariate discrete distribution that is as prominent as the Gaussian distribution to model discrete random variables.

The goal of this thesis is to study partially specified and fully specified models for multivariate and longitudinal data when the response variables are discrete, such as Poisson counts, binary, and more generally multinomial or categorical variables. For the fully specified models, we could use the maximum likelihood method since it is the optimal method of estimation. For partially specified models, we develop

This dissertation follows the style of *Journal of the American Statistical Association.*

statistical methods for parameter estimation. Simulating partially specified discrete models is computationally a challenging problem. In this thesis we also study some simulation methods for partially specified models for multivariate Poisson counts and multivariate binary variables. We use those simulating techniques to compare the relative performance of the statistical methods for partially specified models.

I.1 Overview of Thesis

Apart from the introductory chapter, this dissertation consists of four chapters. In Chapter II, we first discuss the classical generalized linear models, which serves as the basis for modelling multivariate and longitudinal continuous and discrete responses. The multivariate Gaussian distribution is used extensively when analyzing continuous data. Unfortunately, to analyze discrete data, we have numerous multivariate distributions to model a wide variety of dependence structures and concepts. Most of these distributions are intractable in the sense that implementing maximum likelihood estimation could be difficult and computationally could be very intensive. As an alternative, in recent years several moment based methods have been put forward to analyze multivariate and longitudinal discrete data. The foundation for these methods is Godambe (1960)'s optimal theory of estimating equations. In particular the generalized estimating equations (GEE) have become a popular method for analyzing discrete longitudinal data. Despite its popularity the GEE method has significant problems particularly with the estimation of the correlation parameters. In Chapter II we give a description of GEE, its shortcomings, and other moment based methods including Gaussian estimation, modified Gaussian (MG) estimation, and quasi-least squares (QLS). A real life example is given to demonstrate the applicability of these methods.

In Chapter III, we derive the asymptotic distributions of MG and the QLS estimation methods. In order to compare the relative performance of these methods we will need to assume some multivariate distribution and a method of simulating from that distribution. In Chapter III, we discuss the algorithm due to Sim (1993) for simulating a multivariate Poisson distribution with given means and correlations. We study the asymptotic performance of MG and QLS methods via this simulation algorithm, using the joint coverage probabilities of simultaneous confidence regions,

asymptotic relative efficiencies and probability of infeasible or divergent solutions.

In Chapter IV we present fully specified discrete distributions constructed using copulas. We also give some insights into the infeasibility issues of moments based methods with the help of Frechét bounds. The bulk of this chapter is devoted to several fully specified discrete models including models based on normal copulas, probit models, mixture models, probit-normal models, Poisson log-normal models, and discrete choice models, in particular multinominal logit and probit models. Calculation of the maximum likelihood and the Fisher information for these models requires computation of the multivariate normal probabilities. In this chapter we also discuss several efficient algorithms for calculating multivariate normal probabilities. For the multivariate probit and multivariate probit-normal models, we implement maximum likelihood estimation deriving the necessary equations and illustrate it on two real life data sets.

In applications where the populations are subject to post-sampling effects there could be a non-ignorable over and under dispersion parameter. In Chapter V we introduce the quasi-multinomial and generalized Lagrange families of distributions, which account for the over and under dispersions. These distributions can be used to draw meaningful inferences on the strength of population dynamics. We implement the maximum likelihood for the quasi-multinomial distribution and illustrate it on the optical scanner panel data concerning the multiple purchase decisions and marketing predictors for the households in Rome, Georgia.

In Chapter VI we present a brief summary of results obtained in this dissertation. Finally, the Appendix contains important SAS programs that we developed for this dissertation.

CHAPTER II

LONGITUDINAL DATA ANALYSIS USING GLM

Generalized linear models (GLM) are one of the most popular and widely studied statistical models in statistical methodology. These statistical models that relate the random outcomes or responses to the covariates are extremely useful to understand variation in responses as a function of the covariates. They are also useful to make predictions based on the past behavior of the responses. In this chapter we present the background on these models and discuss various methods of estimating the model parameters. We also discuss criteria to check the adequacy and goodness of fit for the models.

The organization of this chapter is as follows. In Section II.1 we present generalized linear models in the univariate setup. Maximum likelihood which is the most efficient method of estimation of the parameters for generalized linear models will be discussed in Section II.1.1. Tests for model adequacy are presented in Section II.1.2. We present extensions of generalized linear models to longitudinal data analysis in Section II.2. The most popular method for estimating the parameters in longitudinal data analysis in the framework of the generalized linear models is the generalized estimating equations (GEE) proposed by Liang and Zeger (1986). In Section II.3 we present a summary of the method of generalized estimating equations. Despite its popularity, the GEE method has some pitfalls, and we discuss those in Section II.3.1. Several authors have proposed alternatives to the GEE method. Most important are the Gaussian, modified Gaussian and the Quasi-least squares estimating procedures. These methods are discussed in Sections II.4, II.4.1, and II.5 respectively. In Section II.6 we present simplifications of the estimating equations for the correlation parameter for commonly used correlation structures. Finally, a real life data example is presented to illustrate the various methods in Section II.7.

II.1 Generalized Linear Models for Univariate Responses

Suppose that we have a collection of n observations or responses of a dependent variable taken on n subjects, and suppose that we also have observations on some

covariates or explanatory variables. Let y_i denote the ith response or observation, and $\mathbf{x}_i$ be the corresponding vector of measurements taken on the explanatory variables. The simple linear model that relates the response variables with the explanatory variables is given by

$$y_i = \mathbf{x}_i'\beta + \epsilon_i \tag{2.1.1}$$

where β is the vector of unknown regression coefficients. The ϵ_i's are assumed to be independent, normal random variables with zero mean and constant variance. This traditional simple linear model (2.1.1) which has been extensively used in statistical data analysis has several limitations. First, the response variables y_i could be binary or counts and need not be continuous so that the assumption of normality for ϵ_i is invalid. Second, the response variable y_i could be continuous with restrictive range. In this case the traditional linear model may be inadequate since $\mathbf{x}_i'\beta$ is free to vary. For example, y_i could be a proportion and falls between 0 and 1, but the linear predictor $\mathbf{x}_i'\beta$ of y_i may not fall within this range. Third, it may not be realistic to assume that the variance of y_i is a constant. For example, if y_i represent Poisson counts then the variance depends on the mean of the y_i's.

To overcome the aforementioned limitations, Nelder and Wedderburn (1972) proposed a class of generalized linear models. These models are an extension of the traditional linear models in the sense that the mean of the responses depends on the regression parameter and the covariates through a nonlinear link function. Further, the variance is allowed to be a function of the mean. More importantly the probability distributions are within the exponential family of distributions and therefore possess some nice properties. These models have numerous applications. Below we briefly summarize the salient features of the generalized linear models.

Definition 2.1 *Exponential Dispersion Family. A probability mass function or a probability density function of a random variable y_i is said to be a member of the exponential dispersion family if it can be written in the form*

$$f(y_i;\theta_i,\phi) = \exp\left[\frac{y_i\theta_i - b(\theta_i)}{a_i(\phi)} + c(y_i,\phi)\right] \tag{2.1.2}$$

where θ_i is called the canonical form of the location parameter and ϕ is called the scale parameter.

Table 2.1: Canonical Link Functions

Distribution	Canonical link function		Variance function
Poisson	Log	$\eta_i = \log(\mu_i)$	μ_i
Binomial	Logit	$\eta_i = \log\left[\frac{\mu_i}{n_i - \mu i}\right]$	$\mu_i\left[1 - \frac{\mu_i}{n_i}\right]$
Normal	Identity	$\eta_i = \mu_i$	1
Gamma	Reciprocal	$n_i = \frac{1}{\mu_i}$	μ_i^2
Inverse Gaussian	Reciprocal^2	$\eta_i = \frac{1}{\mu_i^2}$	μ_i^3

For the probability density function (2.1.2) there exists a relationship between mean and variance. Indeed, we can verify that the mean $E(y_i) = \mu_i = b'(\theta_i)$ and $Var(y_i) = b''(\theta_i)\, a_i(\phi)$. The function $\tau(\mu_i) = \dfrac{d\mu_i}{d\theta_i} = b''(\theta_i)$ is known as the variance function. In many examples, $a_i(\phi) = \dfrac{\phi}{w_i}$, and the variance of y_i reduces to $\dfrac{\phi\tau(\mu_i)}{w_i}$, where w_i's are known constants, and they are called the prior weights.

In the generalized linear model we assume that the distribution of the ith sample outcome y_i is a member of the exponential dispersion family. Further, the relationship between the mean μ_i and the linear predictor $\eta_i = \mathbf{x}_i'\beta$ is given by $\mu_i = g^{-1}(\eta_i)$, where the function g is known as the link function. The link function is a monotonic and differentiable function. If $\theta_i = \eta_i$, then the generalized linear model is called the canonical model, and the corresponding link function is called the canonical link function. Table 2.1 has the canonical link and variance functions for some of the standard distributions.

II.1.1 Parameter Estimation in GLM

The popular method of estimating the unknown regression parameter β and the scale parameter ϕ in the generalized linear model is the method of maximum likelihood. The idea behind the maximum likelihood is simple. We would like to choose an estimate of the parameter value in the feasible region which makes the observed data most probable. This estimate need not be unique in general in the generalized linear model setting. However, for the models with the canonical link functions the maximum likelihood estimators for the location parameters are unique. To obtain

the maximum likelihood estimates, we need to maximize the log likelihood function

$$\log[L(\theta_i, \phi; y_i)] = \sum_{i=1}^{n} \left[\frac{[y_i\theta_i - b(\theta_i)]}{a_i(\phi)} + c(y_i, \phi) \right]. \tag{2.1.3}$$

Recall that $\mu_i = b'(\theta_i)$ and $g(\mu_i) = \eta_i = \mathbf{x}_i'\beta$. A common procedure to get the maximum likelihood estimate of β is to use the Newton-Raphson iterative algorithm. Starting with a trial value β_0 for β, at the rth step of the iterative algorithm we compute

$$\hat{\beta}_r = \hat{\beta}_{r-1} - H^{-1}S.$$

The vector S is known as the gradient vector. The j th component s_j of S is

$$\begin{aligned} s_j = \frac{\partial \log[L(\theta_i, \phi; y_i)]}{\partial \beta_j} &= \sum_{i=1}^{n} \left[\frac{y_i - \mu_i}{a_i(\phi)} \right] \frac{d\theta_i}{d\mu_i} \frac{d\mu_i}{d\eta_i} x_{ij}, \\ &= \sum_{i=1}^{n} \left[\frac{y_i - \mu_i}{a_i(\phi)} \right] \frac{1}{\tau(\mu_i)} \frac{d\mu_i}{d\eta_i} x_{ij}. \end{aligned}$$

where x_{ij} is the jth component of the vector $\mathbf{x}_i$. The Hessian matrix H consists of the second order partial derivatives of (2.1.3) with respect to β. The (j,k)th element of H is

$$h_{jk} = \frac{\partial^2 \log[L(\theta_i, \phi; y_i)]}{\partial \beta_j \partial \beta_k} = \sum_{i=1}^{n} \left\{ \left[\frac{y_i - \mu_i}{a_i(\phi)} \right] \frac{\partial^2 \theta_i}{\partial \beta_j \partial \beta_k} - \frac{b''(\theta_i)}{a_i(\phi)} \left(\frac{\partial \theta_i}{\partial \beta_j} \right) \left(\frac{\partial \theta_i}{\partial \beta_k} \right) \right\}. \tag{2.1.4}$$

Equation (2.1.4) can also be written as

$$h_{jk} = \sum_{i=1}^{n} \left\{ \left[\frac{y_i - \mu_i}{a_i(\phi)} \right] \frac{\partial^2 \theta_i}{\partial \eta_i^2} - \frac{1}{a_i(\phi)\tau(\mu_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 \right\} x_{ij} x_{ik}. \tag{2.1.5}$$

In matrix notation we have

$$H = \mathbf{X}'W\mathbf{X},$$

where $\mathbf{X} = (\mathbf{x}_1 \ldots, \mathbf{x}_n)'$ and W is the diagonal matrix with ith diagonal

$$w_i = \left[\frac{y_i - \mu_i}{a_i(\phi)} \right] \frac{\partial^2 \theta_i}{\partial \eta_i^2} - \frac{1}{a_i(\phi)\,\tau(\mu_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2.$$

The asymptotic covariance matrix of the maximum likelihood estimate of β is the inverse of the information matrix $\mathcal{I}$ where

$$\mathcal{I} = -E(H)$$

$$= \mathbf{X}'[-E(W)]\mathbf{X}. \tag{2.1.6}$$

It is easy to verify that the ith diagonal element of $-E(W)$, the expected value of $-W$, equals $\dfrac{1}{a_i(\phi)\,\tau(\mu_i)}\left(\dfrac{\partial \mu_i}{\partial \eta_i}\right)^2$.

II.1.2 Tests for Model Adequacy

It is well known that the maximum likelihood estimates are the most efficient estimates when the model is correctly specified. However, since the true model is unknown it is not possible to check whether we have correctly specified the model or not. We should at least make sure the model is adequately specified. A standard procedure to check the model adequacy is to compare the fitted model with the full model. The saturated or the full model is the model that contains the maximal number of parameters. Here the θ_i's are treated as the parameters, and therefore the total number of parameters is equal to the number of observations in the sample. Note that the saturated model fits the data exactly and hence it is of little use for any inference. The model adequacy is determined by comparing the estimated value L_c, of the likelihood for the model under consideration and the estimated value of the likelihood L_f, of the saturated model. Two statistics, the scaled deviance and scaled Pearson's chi-square are used to compare L_c with L_f. The scaled deviance or simply the deviance is defined as

$$D(c,\, f) \;=\; -2\log\left(\frac{L_f}{L_c}\right).$$

For exponential dispersion family,

$$D(c,\, f) \;=\; 2\sum_i [y_i(\hat{\theta}_i - \tilde{\theta}_i) + b(\hat{\theta}_i) - b(\tilde{\theta}_i)]/a_i(\phi),$$

where $\hat{\theta}_i$ and $\tilde{\theta}_i$ are the estimates of θ_i under current and saturated models respectively. Table 2.2 contains the deviance function $D(c,\, f)$ for standard distributions. The Pearson's chi-square statistic is defined as

$$\chi^2 \;= \sum_i \tau_i^2(\hat{\mu}_i)\;(y_i - \hat{\mu}_i)^2,$$

and the scaled Pearson's chi-square is χ^2/ϕ. Both the Pearson chi-square and $D(c,\, f)$ statistics are asymptotically distributed as chi-square and small values of these statistics are indicators of model adequacy.

Table 2.2: Deviance Functions

Distribution	$D(c,\ f)$
Normal	$\frac{1}{\sigma^2}\sum[(y_i-\hat{\mu}_i)^2]$
Poisson	$\sum[y_i \log\left(\frac{y_i}{\hat{\mu}_i}\right) - y_i + \hat{\mu}_i]$
Binomial	$\sum[y_i \log\left(\frac{y_i}{\hat{\mu}_i}\right) + (n_i - y_i)\log\left(\frac{n_i-y_i}{n_i-\hat{\mu}_i}\right)]$
Gamma	$2\nu\sum[\log\left(\frac{y_i}{\hat{\mu}_i}\right) + \frac{y_i-\hat{\mu}_i}{\hat{\mu}_i}]$

II.2 Overview of Longitudinal Data

Our discussion so far has been focused on data that consists of one response or one observation on each subject. In this section we turn our attention to the analysis of data that consists of several measurements taken on each subject. Such data occur in longitudinal studies where individuals are measured repeatedly through time or in a cross-sectional study where measurements are taken on each subject under several treatment plans. While it is often possible to address the same scientific questions in both longitudinal and cross-sectional studies, there are some key differences between the two. Many cross-sectional studies arise from cross over trials and the covariates tend to be fixed and not time varying. For example, Jones and Kenward (1987) report a cross trial. Three levels (control, low, and high) of an analgesic drug for relieving pain from primary menstrual cramps were given to 86 women in three time periods. The data consists of the number of women in each of the eight possible outcome categories of 0's and 1's (0 = no pain, 1 = pain), with the six orderings of the three levels of the analgesic drug. These data can be viewed as a cross-sectional study with the different treatment groups as the covariates.

Unlike cross-sectional studies, in longitudinal studies the covariates tend to change over time within individuals. The primary interest is to distinguish changes over time within individuals, and the differences among individuals in their cohort levels is of secondary importance. Longitudinal data can be collected either prospectively, following subjects forward in time, or retrospectively, by examining multiple measurements on each person from historical records. However, much longitudinal data are collected prospectively since the quality of repeated measurements collected from the past records may be inferior.

An example of a prospective longitudinal study is the data collected by Sommer

et al. (1984). The study by Alfred Sommer and colleagues was conducted in west Java, Indonesia to determine the causes of vitamin A deficiency in pre-school children. Over 3000 children were medically examined quarterly for up to six visits to assess whether they suffered from respiratory or diarrheal infection, an ocular manifestation of vitamin A deficiency. Several time varying covariates including age, weight of the children are also observed. The problem is to estimate the increase in risk of respiratory infection for children who are vitamin A deficient while controlling for other demographic factors, and to estimate the degree of heterogeneity in the risk of the disease among children.

In the examples described above, the outcome variable is binary, and therefore standard methods developed for the analysis of Gaussian data are not applicable. Statistical analysis of binary, and in general non-Gaussian longitudinal data is difficult. This is partly due to the fact multivariate extensions of the generalized linear models for modelling the joint distribution of the repeated observations, do not enjoy the nice properties that the multivariate normal distribution possesses.

During the last two decades several methods based on univariate generalized linear models were developed for the analysis of longitudinal measurements. These methods were useful to model both discrete and continuous response variables such as normal, binary, Poisson, and gamma responses. Furthermore, some of these methods were shown to accommodate time-independent as well as time dependent covariates. Some of these methods were also flexible enough to handle missing data provide the missing-ness mechanism is completely random.

The extensions of generalized linear models can be classified into three classes: marginal models, random effects models and transition models. In marginal models the marginal expectation is modelled as a function of explanatory variables; see for example Liang and Zeger (1986), Zeger and Liang (1992). The associations between within subject measurements are modelled separately from the marginal means and variance of the outcomes. Marginal models are effective when the scientific objectives are to characterize and contrast the populations of subjects.

In random effects models between subject variation due to unmeasured covariates is explained by inclusion of subject-specific random effects in the model. The random effects models also referred to as "subject-specific models", are often used when

the prime objective is to study individual's responses. Random effects models were studied by several authors including Laird and Ware (1982); Gilmour et al. (1985); Breslow and Clayton (1993) and Lin and Breslow (1996).

In transition models the conditional means of the outcomes are modelled as functions of the covariates and past responses. Hence the transition models have the interpretations similar to time series models. The application of transition models are limited, since it may not be reasonable to assume that responses follow a subject dependent stochastic process, that changes only with the covariates. Korn and Whittmore (1979), Zeger and Qaqish (1988) and Ware et al. (1988) discuss specific examples of transition models for binary and count data.

In this chapter we will study moment based models for analyzing longitudinal data. These models do not specify the full probability model, but assume only a functional form for the marginal distributions for the repeated measurements and specify a covariance structure for the within subject measurements. However, the covariance structure across time is treated as a nuisance parameter in many situations. The estimation methods exploit the independence across subjects to obtain consistent estimates of the parameters and their asymptotic standard errors. The earliest and the most popular method for analyzing longitudinal data is the generalized estimating equations method. In the next section we present details of this method.

II.3 Generalized Estimating Equations (GEE)

Suppose that we have a vector $Y_i = (y_{i1}, \ldots, y_{it_i})'$ consisting of repeated measurements taken on the ith subject and associated with each measurement y_{ij} we also have a vector of covariates $x_{ij} = (x_{ij1}, \ldots, x_{ijp})'$, $1 \leq j \leq t_i$, $1 \leq i \leq n$. We will assume that the Y_i's are independent. The distribution of Y_i is unknown in most situations. Because the repeated measurements on each subject are correlated, the correlation of the vector Y_i is not the identity. And any efficient statistical procedure should adjust for the within subject correlation.

In a seminal paper Liang and Zeger (1986) proposed a method based on minimal assumptions on the distribution of Y_i. Their method, now popularly known as the

GEE method, can be regarded as the extension of the quasi-likelihood method. As we have seen in Section 2.1, for many non-Gaussian distributions the variance is a function of the mean. This mean-variance relationship is the starting point of the GEE method.

The framework of the GEE method is the following. Here we assume that $E(y_{ij}) = \mu_{ij}$ and $\mathrm{Var}\,(y_{ij}) = \phi\,\tau(\mu_{ij})$, where $\phi > 0$ may be a known constant or an unknown scale parameter. The parameter ϕ is also known as the over-dispersion parameter. The variance function $\tau(\,\cdot\,)$ that relates the variance to the mean of y_{ij}, is assumed to be a known function. We also assume that the link function g is invertible and $\mu_{ij} = g^{-1}(x'_{ij}\,\beta)$, where $\beta = (\beta_1, \ldots, \beta_p)'$ is a p-dimensional vector of regression coefficients. Next we make some assumptions concerning the within subject correlation. To allow parsimonious modelling of the correlation, we assume that the correlation of Y_i is given by a structured correlation matrix $R_i(\alpha)$. In general the parameter $\alpha = (\alpha_1, \ldots, \alpha_q)'$ is a vector of dimension q. It is treated as a nuisance parameter, but is crucial to efficient estimation of the regression parameter.

We thus have $E(Y_i) = \mu_i(\beta)$ and $\mathrm{Cov}\,(Y_i) = \phi\,\Sigma_i(\beta, \alpha)$, where $\mu_i(\beta) = (\mu_{i1}, \ldots, \mu_{it_i})'$. To encompass several continuous and discrete marginal models, the covariance matrix $\Sigma_i(\beta, \alpha)$ is assumed to be a function of α as well as β and it is equal to $A_i^{\frac{1}{2}}(\beta)\,R_i(\alpha)A_i^{\frac{1}{2}}(\beta)$, where $A_i(\beta) = \mathrm{diag}\,(\tau(\mu_{i1}), \tau(\mu_{i2}), \ldots, \tau(\mu_{it_i}))$ is the diagonal matrix of variances of the y_{ij}'s. In the GEE method the regression and the correlation parameters β and α are obtained by an iterative procedure as described below.

Estimation procedure:

Step 1: Start with a trial value β_0.

Step 2: Compute the residuals $Z_i = A_i^{-1/2}(\beta_0)\,(Y_i - \mu_i(\beta_0))$.

Step 3: Estimate α by method of moments using Z_i, $1 \le i \le n$.

Step 4: Obtain β_1 solving the generalized estimating equations

$$\sum_{i=1}^{n} D_i'(\beta_j)\,\Sigma_i^{-1}(\beta,\,\alpha)\,(Y_i - \mu_i(\beta)) = 0, \quad 1 \le j \le p.$$

Step 5: Stop when $\beta_1 \approx \beta_0$, otherwise repeat steps 2, 3, 4.

In the above $D_i(\beta_j) = \partial \mu_i(\beta)/\partial \beta_j$, and equations in Step 4 are known as the GEEs. The estimates $\widehat{\beta}$ and $\widehat{\alpha}$ obtained by the above iterative procedure are known as the GEE estimates of β and α. This estimation procedure is now implemented in popular commercial statistical software packages like SAS, Splus and STATA.

Since the introduction of the GEE method by Liang and Zeger (1986), numerous authors have extended and suggested different versions of the method. All of these methods use the same estimation equation (2.3.1) but differ in how the correlation parameter is estimated in Step 3 of the above algorithm. Noteworthy to mention is the method by Prentice (1988) for analyzing correlated binary data. Prentice (1988) has suggested to replace the moment estimates of α used in Step 3, by another estimating equation, which is known as GEE1. Prentice and Zhao (1991) have given a single GEE type estimating equation treating both β and α as a single parameter. Their method is known as the GEE2 method, and it includes GEE and GEE1 as special cases. Finally, Hall and Severini (1998) suggested a unified approach for simultaneously estimating all the three parameters β, α and ϕ. Their approach, which uses ideas from extended quasi-likelihood, is known as the Extended Generalized Estimating Equation method (EGEE). However, Hall (2001) has shown that the EGEE approach is a special case of GEE1. In particular, EGEE amounts to estimating of the correlation parameter by maximizing the Gaussian likelihood function. We will study the Gaussian method of estimation in detail in Section II.4.

II.3.1 Shortcomings of the GEE Method

The GEE method has been very popular for analyzing longitudinal data because of its ability to estimate the regression parameter consistently requiring only correct specification of the marginal mean and variance. The GEE method is also computationally less demanding than the fully specified models. But despite its popularity the GEE method has some shortcomings. It falls short of the purpose it was introduced - that is, to handle correlated data efficiently. The GEE method has pitfalls in theoretical and in software implementations, particularly in the estimation of the correlation parameters.

Crowder (1995) was the first to point out these pitfalls in the GEE method. He argued with simple examples the working correlation, when it is misspecified, lacks

a proper definition and thus causes a break down of the asymptotic properties of the estimation procedure. Furthermore, even if the working correlation is correctly specified there is no guarantee that the moment estimate of α will fall within the set of feasible values, that is, α may not fall in the range where the correlation matrix is positive definite.

To overcome some of these pitfalls of the GEE method, Qu et al. (2000) presented a method that bypasses the estimation of the correlation parameter. Their method, based on quadratic inference functions, obtains an unbiased equation by combining basic score functions optimally. Vonesh et al. (2002) proposed a second order conditional generalized estimating equations (CGEE2) that also bypasses the estimation of the correlation parameter. The CGEE2 can also be used to estimate random-effect parameters in nonlinear mixed models setup. However, these extensions and several other extensions of the GEE method do not generalize easily for unbalanced and unequally timed longitudinal data that need to be analyzed using more complex correlation structures.

One of major problems with the GEE method is the absence of an objective function that is being minimized (maximized). Such an objective function, if exists will be useful to test adequacy or goodness of fit. A solution to this problem was given by Crowder (2001). He suggested the use of the Gaussian likelihood function as an objective function to estimate the correlation parameters. This method is known as the Gaussian estimation. We will discuss this method in the next section.

II.4 Gaussian Estimation

The Gaussian method of estimation was originally introduced by Whittle (1961) as a general method for estimating the parameters in time series data. It was brought into the limelight by Crowder (1995) for the analysis of correlated binomial data, and more recently by Crowder (2001) as a general and an alternative to the GEE method. Here the parameters are estimated by maximizing the Gaussian (normal) log likelihood. For the data setup described in Section 1, this amounts to minimizing with respect to β, α and ϕ, the objective function

$$\sum_{i=1}^{n} \left\{ \log |\phi \, \Sigma_i(\beta, \alpha)| + \frac{1}{\phi} \left(Y_i - \mu_i(\beta)\right)' \Sigma_i^{-1}(\beta, \alpha) \left(Y_i - \mu_i(\beta)\right) \right\}. \qquad (2.4.1)$$

We can rewrite equation (2.4.1) as

$$\sum_{i=1}^{n} \log|R_i(\alpha)| + \sum_{i=1}^{n} \log|A_i(\beta)| + \log(\phi) \sum_{i=1}^{n} t_i$$
$$+ \frac{1}{\phi} \sum_{i=1}^{n} (Y_i - \mu_i(\beta))' A_i^{-\frac{1}{2}}(\beta) R_i^{-1}(\alpha) A_i^{-\frac{1}{2}}(\beta) (Y_i - \mu_i(\beta)). \tag{2.4.2}$$

Taking partial derivative of (2.4.2) with respect to β_j and equating to zero we get

$$\sum_{i=1}^{n} D_i'(\beta_j) A_i^{-\frac{1}{2}}(\beta) R_i^{-1}(\alpha) Z_i$$
$$+ \sum_{i=1}^{n} \operatorname{tr}\left(\frac{\partial A_i^{-1}(\beta)}{\partial \beta_j} A_i(\beta) R_i^{-1}(\alpha) \left(\frac{1}{\phi} Z_i Z_i' - R(\alpha) \right) \right) = 0, \quad \text{for } j = 1, \ldots, p, \tag{2.4.3}$$

where $D_i(\beta_j) = \partial \mu_i(\beta) / \partial \beta_j$ and $Z_i = A_i^{-\frac{1}{2}}(\beta)(Y_i - \mu_i(\beta))$. Similarly, the estimating equations for the correlation parameter α can be derived as

$$\sum_{i=1}^{n} \operatorname{tr}\left[\frac{\partial R_i^{-1}(\alpha)}{\partial \alpha_j} \left(\frac{1}{\phi} Z_i Z_i' - R_i(\alpha) \right) \right] = 0 \quad \text{for } j = 1, \ldots, q, \tag{2.4.4}$$

Differentiating equation (2.4.2) with respect to ϕ and equating to zero yields following closed form estimate

$$\hat{\phi} = \frac{\sum_{i=1}^{n} (Y_i - \mu_i(\beta))' A_i^{-\frac{1}{2}}(\beta) R_i^{-1}(\alpha) A_i^{-\frac{1}{2}}(\beta) (Y_i - \mu_i(\beta))}{\sum_{i=1}^{n} t_i}. \tag{2.4.5}$$

Thus the Gaussian method of estimation involves solving equations (2.4.3), (2.4.4) and (2.4.5). Note that the equations (2.4.3) are the sums of two equations unbiased equations; a linear unbiased equation which is the GEE, and a quadratic unbiased equation. This is expected since not only the means but the variances of the y_{ij}'s are functions of the parameter β. Even though equations (2.4.3) are unbiased estimating equations, these are not the optimal linear unbiased estimating equations for estimating β in the sense of Godambe (1960). Furthermore, in some situations the presence of the quadratic part in the equations (2.4.3) could lead to an inconsistent estimate of β. This observation is the motivation for considering the following modification Gaussian estimation.

II.4.1 Modified Gaussian Estimation

Recently, Crowder (2001) suggested a modification of the Gaussian method, by decoupling the regression parameter β with the covariance parameters α and ϕ. The modified Gaussian method uses the same estimating functions as the Gaussian method for estimating α and ϕ, but differs in the estimating the regression parameters β. The estimating equation for the regression parameter is obtained by dropping the quadratic part in the equations (2.4.3). The resulting estimating equation is the GEE given by

$$\sum_{i=1}^{n} D_i'(\beta)\, A_i^{-\frac{1}{2}}(\beta)\, R_i^{-1}(\alpha)\, Z_i = 0, \tag{2.4.6}$$

where $D_i(\beta) = \partial \mu_i(\beta)/\partial \beta'$ and $Z_i = A_i^{-\frac{1}{2}}(\beta)\,(Y_i - \mu_i(\beta))$, $1 \leq i \leq m$. Thus the modified Gaussian method estimates the unknown parameters β, α and ϕ solving the equations (2.4.6), (2.4.4) and (2.4.5) iteratively.

II.5 Quasi-least Squares Estimation

The quasi-least squares method was introduced by Chaganty (1997) and further developed by Shults and Chaganty (1998) and Chaganty and Shults (1999). This provides an alternative method of estimating the correlation parameters in the GEE method and has been shown to overcome some of the pitfalls that occur with the moment estimates of the correlation parameters. Here we outline the quasi-least squares method of estimation of the unknown parameters β, α and ϕ. It is a two step procedure. The first step, motivated by the principle of (generalized) least squares, consists of partially minimizing with respect to β and α the generalized error sum of squares

$$\begin{aligned} Q(\beta, \alpha) &= \sum_{i=1}^{n} (Y_i - \mu_i(\beta))' \Sigma_i^{-1}(\beta, \alpha)\,(Y_i - \mu_i(\beta)) \\ &= \sum_{i=1}^{n} (Y_i - \mu_i(\beta))' A_i^{-\frac{1}{2}}(\beta)\, R_i^{-1}(\alpha)\, A_i^{-\frac{1}{2}}(\beta)\,(Y_i - \mu_i(\beta)). \end{aligned} \tag{2.5.1}$$

Straightforward differentiation of the above function with respect to β will result in an estimating equation that is the sum of GEE and an equation that is not unbiased. It is well known that unbiased estimating equations will yield consistent estimates. Since

consistency is a desirable property for an estimate, to obtain an unbiased estimating equation we first decouple the parameters β and α by introducing a variable β^* and modify (2.5.1) as

$$Q(\beta, \beta^*, \alpha, \phi) = \sum_{i=1}^{n} (Y_i - \mu_i(\beta))' A_i^{-\frac{1}{2}}(\beta^*) R_i^{-1}(\alpha) A_i^{-\frac{1}{2}}(\beta^*) (Y_i - \mu_i(\beta)). \tag{2.5.2}$$

Taking the partial derivatives of (2.5.2) with respect to β and α and replacing $\beta^* = \beta$ as in Chaganty (1997), we get the first stage estimating equations as

$$\sum_{i=1}^{n} D_i'(\beta) A_i^{-\frac{1}{2}}(\beta) R_i^{-1}(\alpha) Z_i = 0 \tag{2.5.3}$$

and

$$\sum_{i=1}^{n} Z_i' \frac{\partial R_i^{-1}(\alpha)}{\partial \alpha_j} Z_i = 0, \quad 1 \leq j \leq q, \tag{2.5.4}$$

where $D_i(\beta)$ and Z_i are defined before. The first step estimates $\widetilde{\beta}$ and $\widetilde{\alpha}$ are obtained starting with a trial value for β or α, and solving the estimating equations (2.5.3) and (2.5.4) recursively for β and α. We can show that the estimate $\widetilde{\beta}$ is a consistent estimate of β, whereas $\widetilde{\alpha}$ is not, since (2.5.3) is an unbiased estimating equation and (2.5.4) is not. The second step in the quasi-least squares procedure consists of solving the equation

$$\sum_{i=1}^{n} b_i(\widetilde{\alpha}, \alpha) = 0 \tag{2.5.5}$$

for α. Here $b_i(\widetilde{\alpha}, \alpha) = (b_{i1}(\widetilde{\alpha}, \alpha), \ldots, b_{iq}(\widetilde{\alpha}, \alpha))'$ and

$$b_{ij}(\alpha, \alpha) = \text{tr}\left(\frac{\partial R_i^{-1}(\alpha)}{\partial \alpha_j}\bigg|_{\alpha=\widetilde{\alpha}} R_i(\alpha) \right) = 0 \quad \text{for} \quad j = 1, \ldots, q. \tag{2.5.6}$$

We can show that the solution $\widehat{\alpha}$ of the equation (2.5.5) is a consistent estimate of α. A consistent and more efficient estimate of β is obtained solving equation (2.5.3) replacing α by $\widehat{\alpha}$. Finally, the quasi-least squares estimate of ϕ is given by

$$\widehat{\phi} = \frac{1}{n} \sum_{i=1}^{n} \frac{\widehat{Z}_i' R_i^{-1}(\widehat{\alpha}) \widehat{Z}_i}{t_i} = \frac{1}{n} \sum_{i=1}^{n} \frac{\text{tr}\left(R_i^{-1}(\widehat{\alpha}) \widehat{Z}_i \widehat{Z}_i' \right)}{t_i} \tag{2.5.7}$$

where $\widehat{Z}_i = A_i^{-\frac{1}{2}}(\widehat{\beta}) (Y_i - \mu_i(\widehat{\beta}))$.

II.6 Correlation Parameter Estimates for Common Structures

As we have seen in the previous sections, the three methods of estimation, GEE, MGE, and QLS use the same estimating equation for the regression parameter but differ in the choice of the estimating equation for the correlation parameter α. In this section we do further simplifications of the estimating equations for common correlation structures.

II.6.1 Exchangeable Correlation Structure

For repeated measurements that are not time dependent and are permutation invariant it is reasonable to assume that the correlation between any two measurements is constant. An example of this type of data is a health study in which the subjects are the clinics and the repeated measurements are patients within the clinics. The exchangeable correlation structure is also useful to model the correlation in data that is collected in clusters. Formally, the exchangeable correlation is a correlation matrix of the form

$$R_i(\alpha) = (1-\alpha)\mathbf{I}_{t_i} + \alpha\mathbf{J}_{t_i}$$

where $\mathbf{I}_t$ is an identity matrix of order t and $\mathbf{J}_t$ is $t \times t$ matrix with all elements equal 1. The determinant and the inverse of this matrix are $|R_i(\alpha)| = (1-\alpha)^{t_i-1}(1+(t_i-1)\alpha)$ and

$$R_i^{-1}(\alpha) = \frac{1}{(1-\alpha)}\left\{\mathbf{I}_{t_i} - \left[\frac{\alpha}{1+(t_i-1)\alpha}\right]\mathbf{J}_{t_i}\right\} \tag{2.6.1}$$

respectively. The feasible region, that is, the range of α where $R_i(\alpha)$ is positive definite, is a subset of the interval $[-1/(\max(t_i)-1), 1]$. Differentiating equation (2.6.1) with respect to α we get

$$\frac{\partial R_i^{-1}(\alpha)}{\partial\alpha} = \frac{1}{(1-\alpha)^2}\left[\mathbf{I}_{t_i} - C_{t_i}(\alpha)\mathbf{J}_{t_i}\right] \tag{2.6.2}$$

where

$$C_t(\alpha) = \left\{\frac{1+(t-1)\alpha^2}{[1+(t-1)\alpha]^2}\right\}.$$

Using (2.6.1) and (2.6.2), we can check that estimating equation (2.4.4) of the modified gaussian method simplifies to

$$\frac{1}{\phi(1-\alpha)}\sum_{i=1}^{n}\left\{\sum_{j=1}^{t_i} z_{ij}^2 - C_{t_i}(\alpha)\left(\sum_{j=1}^{t_i} z_{ij}\right)^2\right\} - \sum_{i=1}^{n}\left[\frac{t_i(t_i-1)\alpha}{1+(t_i-1)\alpha}\right] = 0. \quad (2.6.3)$$

Explicit solution for the above equation is not possible, and it has to be solved numerically.

The first step quasi-least square estimating equation (2.5.4) reduces to following

$$\sum_{i=1}^{n}\sum_{j=1}^{t_i} z_{ij}^2 - \sum_{i=1}^{n} C_{t_i}(\alpha)\left(\sum_{j=1}^{t_i} z_{ij}\right)^2 = 0. \quad (2.6.4)$$

Equation (2.6.4) can be rewritten as quadratic expression in α and the feasible solution is given by

$$\tilde{\alpha} = \frac{a}{b} - \sqrt{\left(\frac{a}{b}\right)^2 - \frac{c}{b}} \quad (2.6.5)$$

where

$$\begin{aligned} a &= \sum_{i=1}^{n}(t_i-1)\sum_{j=1}^{t_i} z_{ij}^2, \\ b &= \sum_{i=1}^{n}\left\{(t_i-1)^2\sum_{j=1}^{t_i} z_{ij}^2 - (t_i-1)\left(\sum_{j=1}^{t_i} z_{ij}\right)^2\right\}, \\ c &= \sum_{i=1}^{n}\left\{\sum_{j=1}^{t_i} z_{ij}^2 - (t_i-1)\left(\sum_{j=1}^{t_i} z_{ij}\right)^2\right\}. \end{aligned}$$

The quasi-least squares second step estimating equation (2.5.5) has a closed form solution given by

$$\widehat{\alpha} = \left\{\sum_{i=1}^{n} t_i\left[1 - C_{t_i}(\tilde{\alpha})\right]\right\}\left\{\sum_{i=1}^{n} t_i(t_i-1)C_{t_i}(\tilde{\alpha})\right\}^{-1} \quad (2.6.6)$$

II.6.2 Autoregressive Structure of Order One

The correlation in longitudinal data, typically decreases as the separation between the observations increases. The simplest correlation structure to model this correlation

pattern is the autoregressive structure of order one (AR(1)). In matrix form it is given by

$$R_i(\alpha) = \left[\alpha^{|j-k|}\right].$$

The feasible region for α is $(-1, 1)$. The determinant is $|R_i(\alpha)| = (1-\alpha^2)^{(t_i-1)}$, and the inverse of this matrix is

$$R_i^{-1}(\alpha) = \frac{1}{(1-\alpha^2)}\left[\mathbf{I}_{t_i} - 2\alpha\mathbf{C}_{1t_i} + \alpha^2\mathbf{C}_{2t_i}\right] \tag{2.6.7}$$

where

$$\begin{aligned}\mathbf{C}_{1t} &= \frac{1}{2}\left\{\sum_{j=1}^{t-1}[e_j e'_{(j+1)} + e_{(j+1)}e'_j]\right\},\\ \mathbf{C}_{2t} &= \sum_{j=2}^{t-1} e_j e'_j\end{aligned}$$

and e_j is jth unit vector in $\mathbb{R}^t$. Differentiating equation (2.6.7) with respect to α we get

$$\frac{\partial R_i^{-1}(\alpha)}{\partial\alpha} = \frac{2}{(1-\alpha^2)^2}\left[\alpha(\mathbf{I}_{t_i} + \mathbf{C}_{2t_i}) - (1+\alpha^2)\mathbf{C}_{1t_i}\right] \tag{2.6.8}$$

$$\begin{aligned}\operatorname{tr}\left(\frac{\partial R_i^{-1}(\alpha)}{\partial\alpha}R_i(\alpha)\right) &= -\frac{\partial \log|R_i(\alpha)|}{\partial\alpha}\\ &= \frac{2(t_i-1)\alpha}{(1-\alpha^2)}\end{aligned} \tag{2.6.9}$$

Using (2.6.8) and (2.6.9)) we can verify that the modified gaussian estimating equation (2.4.4) reduces to

$$\frac{1}{(1-\alpha^2)\phi}\sum_{i=1}^{n}\left\{\alpha\left(\sum_{j=1}^{t_i} z_{ij}^2 + \sum_{j=2}^{t_i-1} z_{ij}^2\right) - (1+\alpha^2)\sum_{j=1}^{t_i-1} z_{ij}z_{i(j+1)}\right\} - \sum_{i=1}^{n}(t_i-1)\alpha = 0. \tag{2.6.10}$$

This can be rewritten as the following cubic equation

$$\alpha^3 + a\alpha^2 + b\alpha + a = 0$$

where

$$\begin{aligned}a &= -\frac{1}{\phi}\left\{\sum_{i=1}^{n}\sum_{j=1}^{t_i-1} z_{ij}z_{i(j+1)}\right\}\left\{\sum_{i=1}^{n}(t_i-1)\right\}^{-1},\\ b &= \frac{1}{\phi}\left\{\sum_{i=1}^{n}\left(\sum_{j=1}^{t_i} z_{ij}^2 + \sum_{j=2}^{t_i-1} z_{ij}^2\right)\right\}\left\{\sum_{i=1}^{n}(t_i-1)\right\}^{-1} - 1.\end{aligned}$$

For the AR(1) structure, the first step quasi-least square estimating equation reduces to the quadratic equation

$$\alpha^2 - a^*\alpha + 1 = 0 \tag{2.6.11}$$

where

$$a^* = \left\{\sum_{i=1}^{n}\left(\sum_{j=1}^{t_i} z_{ij}^2 + \sum_{j=2}^{t_i-1} z_{ij}^2\right)\right\}\left\{\sum_{i=1}^{n}\sum_{j=1}^{t_i-1} z_{ij}z_{i(j+1)}\right\}^{-1}. \tag{2.6.12}$$

Feasible solution of (2.6.11) is given by

$$\tilde{\alpha} = \frac{1}{2}\left[a^* - \sqrt{(a^*+2)(a^*-2)}\right]. \tag{2.6.13}$$

The solution to the second step quasi-least squares estimating equation is

$$\widehat{\alpha} = \frac{2\tilde{\alpha}}{1+\tilde{\alpha}^2} = \frac{2}{a^*}. \tag{2.6.14}$$

A third commonly employed correlation structure for the analysis of longitudinal data is the moving average of order one correlation structure. Here the outcomes are assumed to be correlated with closest neighbors only. The (j, k)th element of $R_i(\alpha)$ is α if $|j-k| = 1$, 1 if $|j-k| = 0$, and 0 elsewhere. The determinant of $R_i(\alpha)$ is not in a closed form but can be computed easily using the following recurrence relation:

$$d_{t_i}(\alpha) = |R_i(\alpha)| = d_{t_i-1}(\alpha) - \alpha^2 d_{t_i-2}(\alpha) \tag{2.6.15}$$

where $d_0(\alpha) = 1$ and $d_1(\alpha) = 1$. The inverse of $R_i(\alpha)$ is not in simple form. Hence the estimating equations for the MGE and QLS have to be solved numerically.

II.7 Illustration of the Methods

In this section we present an example to illustrate data analysis using the methods discussed in this chapter. We consider the data studied in Preisser and Qaqish (1999) and made available publicly by the authors. The data is from a randomized clinical trial directed at assessing the Guidelines for Urinary Incontinence Discussion and Evaluation (GUIDE) as adopted by primary care providers. Data analysis goal is to identify factors among urinary incontinent men and women age 76 and older that are predictive of their response to the question: "Do you consider this accidental loss of urine a problem that interferes with your day to day activities or bothers you

in other ways?" A total of 137 patients from 38 practices were asked this question. In this example, the practices are independent and treated as subjects or clusters. The observations within practices are treated as equicorrelated measurements. The response y_{ij} of the jth patient from the ith practice, is a binary variable and equals 1 if the patient is bothered and 0 if not. There are five covariates: (i) age (ii) gender (male or female), (iii) the number of leaking accidents in a day (dayacc), (iv) severity of the leak, takes 1=if the leak creates a moisture, 2= wet their underwear, 3= trickle down the thigh, 4=wet the floor, (severe) and finally (v) number of times they go to the toilet to urinate (toilet). For the analysis, we introduced an indicator variable for female and the age variable is centered at 76 and scaled down by 0.1.

Table 2.3: GUIDE Data: Parameter Estimates

	GEE		MGE		QLS	
Covariate	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
INTERCEPT	-3.054	0.959	-2.929	0.959	-2.996	0.959
FEMALE	-0.745	0.600	-0.782	0.591	-0.762	0.596
AGE	-0.676	0.561	-0.694	0.556	-0.684	0.558
DAYACC	0.392	0.093	0.381	0.092	0.386	0.092
SEVERE	0.812	0.359	0.802	0.357	0.808	0.358
TOILET	0.108	0.099	0.106	0.098	0.107	0.098
α	0.093	-	0.153	0.074	0.120	0.076

Table 2.3 contains the parameter estimates for the three methods of estimation, GEE, Modified Gaussian (MGE) and quasi-least squares (QLS). We have used PROC GENMOD to obtain the estimates for the GEE method, which does not give the standard error for the estimate of α. Note that the estimates of the regression parameter obtained by the three methods are similar.

CHAPTER III

ASYMPTOTIC BEHAVIOR OF ESTIMATION METHODS

In this chapter we will first derive the asymptotic distributions of the moment based methods that we discussed in Chapter II. Next we will compare these methods via asymptotic relative efficiencies and coverage probabilities for Poisson counts and binary responses using simulations. However, simulating data from multivariate discrete distributions with given marginals and correlations is a challenging problem. Indeed a multivariate distribution with given marginals and correlations may not even exist and it need not be unique if it exists. There is a great deal of literature concerning generating bivariate discrete random vectors with given marginals beginning with McKendrick (1916) and Wicksell (1916). Unfortunately there is no straightforward generalization of these methods from bivariate case to higher dimensions. Due to complex nature of the generalized Poisson distributions, Tsionas (2001) and several authors proposed Bayesian multivariate Poisson regression.

In recent years a promising method for simulating a multivariate Poisson distribution with given correlations, due to Sim (1993), has gained popularity. In this chapter we will discuss a slight generalization of Sim (1993)'s algorithm. This algorithm, when successful, leads to a complicated multivariate Poisson distribution. Maximum likelihood estimation is intractable for this complex multivariate distribution, and there is a need to use moment based methods, like modified Gaussian and the quasi-least squares estimation. Therefore, it is of interest to know the relative performance of these two methods. In this chapter we will use simulated data generated using Sim's algorithm to compare performance of the two estimating methods; modified Gaussian and quasi-least squares.

The organizing of this chapter is as follows. In Section III.1 we derive the asymptotic distributions of the MGE and QLS methods. The simulation results comparing these methods for multivariate Poisson counts under exchangeable, autoregressive and moving average correlation assumptions are presented in Section III.2. In Section III.3 we use a real life data consisting of correlated binary responses to compare the two methods. A summary and conclusions of our results are presented at the end of chapter.

III.1 Asymptotic Distributions

In this section we will obtain the asymptotic distribution of the modified Gaussian and the quasi-least squares estimates, as $n \to \infty$. The following theorem due to Yuan and Jennrich (1998), plays a fundamental role in establishing the asymptotic distributions.

Theorem 3.1 *Let Z_i, $1 \leq i \leq n$, be independent random vectors of dimensions t_i generated from distribution $f_i(Z_i, \theta_0)$, $1 \leq i \leq n$. Assume that $t_i \leq t$ for all i and $\theta_0 \in \Theta$, which is a subset of $\mathbb{R}^k$. Let the multivariate functions $h_i(Z_i, \theta)$, $1 \leq i \leq n$, taking values in $\mathbb{R}^k$, be such that*

1. *For each i, $\mathrm{E}(h_i(Z_i, \theta_0)) = 0$ and $\mathrm{Var}(h_i(Z_i, \theta_0)) = M_i(\theta_0)$; and let $\frac{1}{n}\sum_{i=1}^{n} M_i(\theta_0) \to M(\theta_0)$, as $n \to \infty$. Assume that $M(\theta_0)$ is positive definite.*
2. *For all $\lambda \in \mathbb{R}^k$ of length one there exists a positive numbers B and δ such that for all i $\mathrm{E}\left[\frac{(\lambda' h_i(Z_i, \theta_0))^2}{1 + \lambda' M_i(\theta_0)\lambda}\right]^{1+\delta} \leq B.$*
3. *For each i, $h_i(Z_i, \theta)$ is twice differentiable almost surely on Θ.*
4. *For each $\theta \in \Theta$, $\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\left[\frac{\partial h_i(Z_i, \theta)}{\partial \theta'}\right] \to I(\theta)$ as $n \to \infty$. Assume that $I(\theta_0)$ is non-singular.*
5. *Suppose that $\left|\frac{\partial^2(\lambda' h_i(Z_i, \theta))}{\partial\theta\partial\theta'}\right| \leq T$ for all i and for all $\lambda \in \mathbb{R}^k$ of length one. Here $|\cdot|$ denotes the determinant.*

Under the above regularity conditions, if $\widehat{\theta}$ is the solution of the unbiased estimating equation

$$\frac{1}{n}\sum_{i=1}^{n} h_i(Z_i, \theta) = 0, \tag{3.1.1}$$

then we have

$$(\widehat{\theta} - \theta_0) \quad is \quad AMVN\left(0, \frac{[I(\theta_0)]^{-1} M(\theta_0) [I(\theta_0)]^{-1}}{n}\right). \tag{3.1.2}$$

Theorem 3.1 is proved by appealing to the Cramer-Wold device. The main steps of the proof are as follows. First, we reduce the multivariate estimating functions $h_i(Z_i, \theta)$ to univariate random variables using the linear transformation $\lambda' h_i(Z_i, \theta)$. Next under the five regularity conditions, the asymptotic normality of the univariate random variables is established as an application of the Lindeberg-Feller theorem, see Serfling (1980). In practice the matrices $I(\theta_0)$ and $M(\theta_0)$ in the asymptotic covariance matrix are unknown, and we can estimate them as

$$I(\widehat{\theta}) = \frac{1}{n} \sum_{i=1}^{n} \mathrm{E} \left[\frac{\partial h_i(Z_i, \theta)}{\partial \theta'} \right] \Bigg|_{\theta=\widehat{\theta}} \tag{3.1.3}$$

and

$$M(\widehat{\theta}) = \frac{1}{n} \sum_{i=1}^{n} M_i(\widehat{\theta}). \tag{3.1.4}$$

To establish the asymptotic normality of the modified Gaussian estimates, we first note that the three estimating equations (2.4.6), (2.4.4) and (2.4.5) used in that method can be rewritten as $\sum_{i=1}^{n} h_{mi}(Z_i, \theta) = 0$ where $h_{mi}(Z_i, \theta) = K'_{mi}(\theta)\, \xi_i(\theta)$ and

$$\begin{aligned}
K_{mi}(\theta) &= \begin{bmatrix} R_i^{-1}(\alpha) A_i^{-\frac{1}{2}}(\beta) D_i(\beta) & & \mathbf{0} \\ \mathbf{0} & b_{i1}(\alpha) \;\; b_{i2}(\alpha) \;\; \ldots \;\; b_{iq}(\alpha) & \mathrm{vec}(R_i^{-1}(\alpha)) \end{bmatrix}, \\
b_{ij}(\alpha) &= \mathrm{vec}\left(\frac{\partial R_i^{-1}(\alpha)}{\partial \alpha_j} \right), \quad 1 \le j \le q, \\
\xi_i(\theta) &= \begin{pmatrix} Z_i \\ \mathrm{vec}\left(Z_i Z_i' - \phi R_i(\alpha) \right) \end{pmatrix}.
\end{aligned}$$

The asymptotic distribution of the modified Gaussian estimates follows as a consequence of Theorem 3.1 and it is given below.

Theorem 3.2 *Let* $\theta = (\beta, \alpha, \phi)$ *and* $\widehat{\theta}_m = (\widehat{\beta}_m, \widehat{\alpha}_m, \widehat{\phi}_m)$ *be the MG estimates. If* $h_{mi}(Z_i, \theta)$*'s satisfy the regularity conditions 1–5 of Theorem 3.1 then we have*

$\sqrt{n}(\widehat{\theta}_m - \theta)$ is asymptotically multivariate normal with mean 0 and covariance matrix $\Sigma_m(\theta) = I_m^{-1}(\theta)M_m(\theta)I_m^{-1}(\theta)$ where

$$
\begin{aligned}
I_m(\theta) &= \frac{1}{n}\sum_{i=1}^{n} K'_{mi}(\theta)\nabla_i(\theta), \\
M_m(\theta) &= \frac{1}{n}\sum_{i=1}^{n} K'_{mi}(\theta)\Psi_i(\theta)K_{mi}(\theta), \\
\nabla_i(\theta) &= \mathrm{E}\left[-\frac{\partial \xi_i(\theta)}{\partial \theta}\right] \\
&= \begin{bmatrix} A_i^{-\frac{1}{2}}(\beta)D_i(\beta) & & \mathbf{0} & & \\ \mathbf{0} & \phi\, c_{i1}(\alpha) & \phi\, c_{i2}(\alpha) & \ldots & \phi\, c_{iq}(\alpha) \quad \mathrm{vec}(R_i(\alpha)) \end{bmatrix}, \\
c_{ij}(\alpha) &= \mathrm{vec}\left(\frac{\partial R_i(\alpha)}{\partial \alpha_j}\right), \quad 1 \le j \le q.
\end{aligned}
$$

and $\Psi_i(\theta) = \mathrm{Cov}(\xi_i(\theta))$.

Similarly, to derive the asymptotic distribution of the quasi-least squares estimates we first note that the estimating equations (2.4.6), (2.5.4), (2.5.5) and (2.5.7) can be rewritten as $\sum_{i=1}^{n} h_{qi}(Z_i, \theta) = 0$ where $h_{qi}(Z_i, \theta) = K'_{qi}(\theta)\xi_i(\theta)$ and

$$
K_{qi}(\theta) = \begin{bmatrix} R_i^{-1}(\alpha)A_i^{-\frac{1}{2}}(\beta)D_i(\beta) & & \mathbf{0} & & \\ \mathbf{0} & b_{1i}(\tilde{\alpha}) & b_{2i}(\tilde{\alpha}) & \ldots & b_{qi}(\tilde{\alpha}) \quad \mathrm{vec}(R_i^{-1}(\alpha)) \end{bmatrix}.
$$

Theorem 3.3 establishes the asymptotic distribution of the quasi-least squares estimates.

Theorem 3.3 *Let $\theta = (\beta, \alpha, \phi)$ and let $\widehat{\theta}_q = (\widehat{\beta}_q, \widehat{\alpha}_q, \widehat{\phi}_q)$ be the QLS estimates. Assume that for each α, there exists $\tilde{\alpha}$ such that*

$$
\sum_{i=1}^{n} \mathrm{tr}\left(\frac{\partial R_i^{-1}(\alpha)}{\partial \alpha}\bigg|_{\alpha=\tilde{\alpha}} R_i(\alpha)\right) = 0.
$$

If $h_{qi}(Z_i, \theta)$'s satisfy the regularity conditions 1–5 of Theorem 3.1 then we have $\sqrt{n}(\widehat{\theta}_q - \theta)$ is asymptotically multivariate normal with mean 0 and covariance matrix $\Sigma_q(\theta) = I_q^{-1}(\theta)M_q(\theta)I_q^{-1}(\theta)$ where

$$
\begin{aligned}
I_q(\theta) &= \frac{1}{n}\sum_{i=1}^{n} K'_{q_i}(\theta)\nabla_i(\theta) \\
M_q(\theta) &= \frac{1}{n}\sum_{i=1}^{n} K'_{q_i}(\theta)\Psi_i(\theta)K_{q_i}(\theta).
\end{aligned}
$$

Since both the quasi-least squares and modified Gaussian methods do not make any assumptions on the joint distribution of Y_i, the covariance matrix Ψ_i of ξ_i, which depends on moments of order up to four, is unknown. However, for continuous outcomes several authors have used a robust and consistent estimate of the covariance matrix Ψ_i given by $\widehat{\Psi}_i = \xi_i(\theta)\,\xi_i(\theta)'$. We could use $\widehat{\Psi}_i$ with θ replaced by $\widehat{\theta}_m$ and $\widehat{\theta}_q$ to estimate the asymptotic covariance matrices $\Sigma_m(\theta)$ and $\Sigma_q(\theta)$ respectively.

Our main goal in this chapter is to study the asymptotic performance of the QLS estimates with respect to the MG estimates for discrete outcomes, in particular for multivariate Poisson and multivariate binary outcomes. To achieve this goal, in the next section we will study some multivariate Poisson distributions that are characterized by given marginals and correlations. We will also study methods for simulating those distributions.

III.2 Multivariate Poisson Simulations

Unlike multivariate Gaussian distribution which is uniquely determined by the marginal means and covariance matrix, there could be none or several multivariate distributions with a specified covariance matrix and specified marginals such as Poisson with given means. In the bivariate case McKendrick (1916) constructed a bivariate distribution with Poisson marginals as follows. Suppose that U_1, U_2 and U_3 are independent Poisson with means κ_1, κ_2 and κ_3 respectively. Then $Y_1 = U_1 + U_3$ and $Y_2 = U_2 + U_3$ are distributed as bivariate Poisson with joint density function given by

$$g(y_1, y_2; \kappa_1, \kappa_2, \kappa_3) = \sum_{y_3=1}^{\min(y_1,y_2)} f(y_1 - y_3; \kappa_1 - \kappa_3) f(y_2 - y_3; \kappa_2 - \kappa_3) f(y_3; \kappa_3) \quad (3.2.1)$$

where $f(y;\kappa)$ is probability mass function of a Poisson random variable with mean κ. It is easy to see that the random variables Y_1 and Y_2 marginally are Poisson with means $\lambda_1 = \kappa_1 + \kappa_3$ and $\lambda_2 = \kappa_2 + \kappa_3$ and positive correlation $\alpha = \kappa_3/[\sqrt{(\kappa_1+\kappa_3)(\kappa_2+\kappa_3)}]$. This method works if for a given λ_1, λ_2, the correlation is in the range $0 < \alpha < \min\left[\sqrt{\lambda_1/\lambda_2}, \sqrt{\lambda_2/\lambda_1}\right]$. In general, given λ_1 and λ_2 the correlation α is restricted above by the Fréchet upper bound $\rho_U(\lambda_1, \lambda_2)$ as defined in (4.1.8). And there is no bivariate distribution with marginals as Poisson if $\alpha > \rho_U(\lambda_1, \lambda_2)$.

Using an alternative approach, Wicksell (1916) has independently derived an expression for the bivariate Poisson distribution (3.2.1) taking the limit of sums of bivariate Bernoulli distributions. For higher dimensions, Bernoulli random variables play an important role in the construction of multivariate Poisson distributions with given means and correlations. In particular, the binomial thinning operator, which has been extensively used in modelling reliability data involving imperfect repairs (see Barlow and Proschan (1975)), provides the foundation for simulating multivariate Poisson distributions.

Definition 3.1 *Let X be a non-negative integer valued random variable and $0 < \theta < 1$ be fixed. The binomial thinning of X with θ, is a random variable denoted by $\theta \star X$ that equals in distribution to $\sum_{i=0}^{X} I_i$ where $I_0 = 0$ and $I_i, i = 1, 2 \ldots, X$ are i.i.d. Bernoulli (θ) random variables independent of X.*

The binomial thinning is closed within the class of Poisson distributions as shown in the next lemma. Joe (1997) developed a unified approach to construct non-stationary Poisson time series models by exploiting this closure property of the binomial thinning operator.

Lemma 3.1 *Let X be a Poisson random variable with mean λ. Then*

1. *$\theta \star X$ is distributed as Poisson random variable with mean $\theta\,\lambda$.*
2. *$\mathrm{Cov}(\theta_1 \star X,\ \theta_2 \star X) = \theta_1\theta_2\lambda$.*

for all θ, θ_1 and $\theta_2 \in (0, 1)$.

Proof: Let $Y = \theta \star X$. Then Y given $X = x$ has Binomial(x, θ) distribution. The moment generating function of Y conditional on $X = x$ is

$$\mathrm{E}[e^{tY}|X = x] \;=\; \left[(1-\theta) + \theta e^{t}\right]^{x}.$$

Hence the moment generating function of the unconditional distribution of Y is

$$\begin{aligned}\mathrm{E}[e^{tY}] &= \mathrm{E}[\mathrm{E}[e^{tY}|X = x]] \\ &= \sum_{x=0}^{\infty}\left[(1-\theta) + \theta e^{t}\right]^{x}\frac{e^{-\lambda}\lambda^{x}}{x!} = e^{\lambda\theta(e^{t}-1)}\end{aligned}$$

which is the moment generating function of a Poisson random variable with mean $\theta\lambda$. To see the second part of the lemma, let $Y_1 = \theta_1 \star X$ and $Y_2 = \theta_2 \star X$. The random variables Y_1 and Y_2 are conditionally independent given $X = x$. Hence

$$\begin{aligned} \mathrm{E}[Y_1 Y_2|X = x] &= \mathrm{E}[Y_1|X = x]\,\mathrm{E}[Y_2|X = x] \\ &= \theta_1\,\theta_2\,x^2 \text{ and thus} \\ \mathrm{E}[Y_1 Y_2] &= \theta_1\theta_2\mathrm{E}(X^2) = \theta_1\theta_2(\lambda^2 + \lambda). \end{aligned}$$

Therefore

$$\begin{aligned} \mathrm{Cov}(Y_1, Y_2) &= \mathrm{E}(Y_1\,Y_2) - \mathrm{E}(Y_1)\mathrm{E}(Y_2) \\ &= \theta_1\theta_2(\lambda^2 + \lambda) - \theta_1\theta_2\lambda^2 = \theta_1\theta_2\lambda. \end{aligned}$$

This completes the proof.

Using the binomial thinning operator, we now introduce a constructive definition for the multivariate Poisson distribution.

Definition 3.2 *Let $Z_1, Z_2 \ldots, Z_q$ be independently distributed random variables such that Z_k is Poisson with mean λ_k, $1 \leq k \leq q$. Let $\Theta = [\theta_{jk}]$ be a $(p \times q)$ matrix of constants, $0 < \theta_{jk} < 1$. Define*

$$Y_j = \sum_{k=1}^{q} \theta_{jk} \star Z_k \quad j = 1, 2, \ldots, p, \tag{3.2.2}$$

then $\mathbf{Y} = (Y_1, Y_2 \ldots, Y_p)$ is distributed as multivariate poisson with mean vector $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_p)'$.

It is clear from the definition, the conditional probability mass function of Y_j given $(z_1, \ldots, z_q)$ is

$$P(y_j; \mathbf{z}) = \sum_{W \in A_{\mathbf{z}}(y_j)} \prod_{k=1}^{q} \binom{z_k}{w_k} \theta_{jk}^{w_k}(1 - \theta_{jk})^{z_k - w_k} \tag{3.2.3}$$

where $A_{\mathbf{z}}(y) = \left\{ W = (w_1, w_2, \ldots, w_q) : w_k \in \mathbb{N}, w_k \leq z_k \text{ and } \sum_{k=1}^{q} w_k = y \right\}$. Therefore the unconditional joint probability mass function of $\mathbf{Y}$ is

$$P(\mathbf{y}) = \sum_{\mathbf{z}} \left\{ \prod_{i=1}^{p} P(y_j; \mathbf{z}) \right\} \left\{ \prod_{k=1}^{q} \frac{e^{-\lambda_k} \lambda_k^{z_k}}{z_k!} \right\}. \tag{3.2.4}$$

Using Lemma 3.1, we can check that the moment generating function of $\mathbf{Y}$ is given by

$$\begin{aligned} M(\mathbf{t}) &= E\left[\prod_{j=1}^{p} e^{t_j Y_j}\right] \\ &= E_{\mathbf{z}}\left[\prod_{j=1}^{p}\prod_{k=1}^{q} \mathrm{e}^{t_j\,(\theta_{jk} \star Z_k)}\right] \\ &= = E_{\mathbf{z}}\left\{\prod_{k=1}^{q}\prod_{j=1}^{p}\left[\theta_{jk}e^{t_j} + (1-\theta_{jk})\right]^{Z_k}\right\}. \end{aligned} \tag{3.2.5}$$

Further, the moment generating function (3.2.5) can be written as

$$M(\mathbf{t}) = \prod_{k=1}^{q} \exp\{\lambda_k(e^{S_k(\mathbf{t})} - 1)\} \tag{3.2.6}$$

where $S_k(\mathbf{t}) = \sum_{j=1}^{p} \log[\theta_{jk}e^{t_j} + (1-\theta_{jk})]$. The representation (3.2.6) shows that the moment generating function of $\mathbf{Y}$ is equal to the joint moment generating function of q independently distributed Poisson random variables centered at $(S_1(\mathbf{t}), \ldots, S_q(\mathbf{t}))$. Also $S_k(\mathbf{t})$ can be viewed as the joint cumulant generating function of p independently distributed Bernoulli random variables.

The central moment generating function of $\mathbf{Y}$ is given by

$$\begin{aligned} K(\mathbf{t}) &= E\left[\exp(\mathbf{t}'(\mathbf{Y}-\boldsymbol{\mu}))\right] \\ &= \prod_{k=1}^{q} \exp\{\lambda_k(U_k(\mathbf{t}) - 1)\} \end{aligned} \tag{3.2.7}$$

where $U_k(\mathbf{t}) = e^{S_k(\mathbf{t})} - \sum_{j=1}^{p} \theta_{jk}t_j$.

The moments up to order four of the multivariate Poisson mass function (3.2.4) needed for calculating the asymptotic relative efficiencies of the estimating methods MGE and QLS, can be obtained by differentiating the central moment generating function (3.2.7). We will need the following derivatives to derive simplified expressions for the higher order moments.

Note that $U_k(\mathbf{t})$ can also be written as following power-series

$$U_k(\mathbf{t}) = \sum_{d_1=0}^{\infty} \cdots \sum_{d_p=0}^{\infty} \prod_{j=1}^{p} \left[\frac{t_j^{d_j}}{d_j!}\pi_{jk}^{(d_j)}\right] - \sum_{j=1}^{p} \theta_{jk}t_j \tag{3.2.8}$$

where

$$\pi_{jk}^{(d)} = \begin{cases} 1 & \text{if } d = 0 \\ \theta_{jk} & \text{otherwise.} \end{cases}$$

Therefore, for all $d_j = 0, 1, \ldots; j = 1, 2, \ldots, p$ such that $\sum_{j=1}^{p} d_j \geq 1$ we have

$$\begin{aligned} U_k^{(\mathbf{d})}(\mathbf{t}) &= \frac{\partial^{\sum_{j=1}^{p} d_j} U_k(\mathbf{t})}{\partial t_1^{d_1} \ldots \partial t_p^{d_p}} \\ &= \left\{ \sum_{r_1=0}^{\infty} \cdots \sum_{r_p=0}^{\infty} \prod_{j=1}^{p} \left[\frac{t_j^{r_j}}{r_j!} \pi_{jk}^{(r_j+d_j)} \right] \right\} - \delta_k^{(\mathbf{d})}(\mathbf{t}) \end{aligned} \tag{3.2.9}$$

where

$$\delta_k^{(\mathbf{d})}(\mathbf{t}) = \begin{cases} \theta_{jk} & \text{if } d_j = 1 \text{ and } d_{j'} = 0 \text{ for all } j' \neq j \\ 0 & \text{otherwise.} \end{cases}$$

Differentiating (3.2.7) with respect to t_{j_1} we get

$$\frac{\partial K(\mathbf{t})}{\partial t_{j_1}} = \omega_{j_1}(\mathbf{t}) K(\mathbf{t}) \tag{3.2.10}$$

where

$$\omega_{j_1}(\mathbf{t}) = \sum_{k=1}^{q} U_k^{(\mathbf{d}_1)}(\mathbf{t}) \lambda_k.$$

and $\mathbf{d}_1 = e_{j_1}$ (j_1th unit vector in $\mathbb{R}^p$).

The covariance matrix of $\mathbf{Y}$ can be obtained by differentiating (3.2.10) with respect to t_{j_2} and substituting $\mathbf{t}$ with $\mathbf{0}$ as

$$\frac{\partial^2 K(\mathbf{t})}{\partial t_{j_1} \partial t_{j_2}} = \left\{ \omega_{j_1}(\mathbf{t}) \omega_{j_2}(\mathbf{t}) + \omega_{j_1 j_2}^{(2)}(\mathbf{t}) \right\} K(\mathbf{t}) \tag{3.2.11}$$

where

$$\omega_{j_1 j_2}^{(2)}(\mathbf{t}) = \sum_{k=1}^{q} U_k^{(\mathbf{d}_2)}(\mathbf{t}) \lambda_k$$

and $\mathbf{d}_2 = (e_{j_1} + e_{j_2})$. Therefore the elements of $\mathrm{Cov}(\mathbf{Y}) = \Sigma$ are given by

$$\begin{aligned} \sigma_{j_1 j_2} &= \mathrm{Cov}(Y_{j_1}, Y_{j_2}) \\ &= \omega_{j_1 j_2}^{(2)}(\mathbf{0}) \\ &= \sum_{k=1}^{q} \prod_{l=1}^{p} \pi_{lk}^{(d_{2l})} \lambda_k \quad \text{for all } 1 \leq j_1, j_2 \leq p. \end{aligned} \tag{3.2.12}$$

Differentiating equation (3.2.11) with respect to t_{j_3} we get

$$\frac{\partial^3 K(\mathbf{t})}{\partial t_{j_1}\partial t_{j_2}\partial t_{j_3}} = \Big\{\omega_{j_1}(\mathbf{t})\omega_{j_2}(\mathbf{t})\omega_{j_3}(\mathbf{t}) + \omega_{j_1}(\mathbf{t})\omega^{(2)}_{j_2j_3}(\mathbf{t}) + \omega_{j_2}(\mathbf{t})\omega^{(2)}_{j_1j_3}(\mathbf{t}) + \omega_{j_3}(\mathbf{t})\omega^{(2)}_{j_1j_2}(\mathbf{t}) + \omega^{(3)}_{j_1j_2j_3}(\mathbf{t})\Big\} K(\mathbf{t}) \quad (3.2.13)$$

where

$$\omega^{(3)}_{j_1j_2j_3}(\mathbf{t}) = \sum_{k=1}^{q} U_k^{(\mathbf{d}_3)}(\mathbf{t})\lambda_k$$

is defined using (3.2.9) and $\mathbf{d}_3 = (e_{j_1} + e_{j_2} + e_{j_3})$. Let $\zeta = \mathrm{vec}((\mathbf{Y}-\boldsymbol{\mu})(\mathbf{Y}-\boldsymbol{\mu})')$. A typical element of $\mathrm{Cov}(\zeta,\ (\mathbf{Y}-\boldsymbol{\mu})) = \boldsymbol{\tau}$ is given by:

$$\begin{aligned}\tau_{j_1j_2j_3} &= \mathrm{Cov}((Y_{j_1}-\mu_{j_1})(Y_{j_2}-\mu_{j_2}), (Y_{j_3}-\mu_{j_3})) \\ &= \omega^{(3)}_{j_1j_2j_3}(\mathbf{0}) \\ &= \sum_{k=1}^{q}\prod_{l=1}^{p}\pi_{lk}^{(d_{3l})}\lambda_k. \end{aligned} \quad (3.2.14)$$

Similarly, differentiating equation (3.2.13) with respect to t_{j_4} we get

$$\begin{aligned}\frac{\partial^4 K(\mathbf{t})}{\partial t_{j_1}\ldots\partial t_{j_4}} = \Big\{&\omega_{j_1}(\mathbf{t})\omega_{j_2}(\mathbf{t})\omega_{j_3}(\mathbf{t})\omega_{j_4}(\mathbf{t}) + \omega^{(2)}_{j_1j_2}(\mathbf{t})\left[\omega^{(2)}_{j_3j_4}(\mathbf{t}) + \omega_{j_3}(\mathbf{t})\omega_{j_4}(\mathbf{t})\right] \\ &+\omega^{(2)}_{j_1j_3}(\mathbf{t})\left[\omega^{(2)}_{j_2j_4}(\mathbf{t}) + \omega_{j_2}(\mathbf{t})\omega_{j_4}(\mathbf{t})\right] + \omega^{(2)}_{j_2j_3}(\mathbf{t})\left[\omega^{(2)}_{j_1j_4}(\mathbf{t}) + \omega_{j_1}(\mathbf{t})\omega_{j_4}(\mathbf{t})\right] \\ &+\omega_{j_1}(\mathbf{t})\omega^{(3)}_{j_2j_3j_4}(\mathbf{t}) + \omega_{j_2}(\mathbf{t})\omega^{(3)}_{j_1j_3j_4}(\mathbf{t}) + \omega_{j_3}(\mathbf{t})\omega^{(3)}_{j_1j_2j_4}(\mathbf{t}) + \omega_{j_4}(\mathbf{t})\omega^{(3)}_{j_1j_2j_3}(\mathbf{t}) \\ &+\omega^{(4)}_{j_1j_2j_3j_4}(\mathbf{t})\Big\} K(\mathbf{t})\end{aligned}$$

where

$$\omega^{(4)}_{j_1j_2j_3j_4}(\mathbf{t}) = \sum_{k=1}^{q} U_k^{(\mathbf{d}_4)}(\mathbf{t})\lambda_k$$

is defined using (3.2.9) and $\mathbf{d}_4 = (e_{j_1} + e_{j_2} + e_{j_3} + e_{j_4})$. A typical element of the matrix $\mathrm{Cov}(\zeta) = \boldsymbol{\kappa}$ is given by the following equations:

$$\begin{aligned}\kappa_{j_1j_2j_3j_4} &= \mathrm{Cov}((Y_{j_1}-\mu_{j_1})(Y_{j_2}-\mu_{j_2}), (Y_{j_3}-\mu_{j_3})(Y_{j_4}-\mu_{j_4})) \\ &= \omega^{(2)}_{j_1j_3}(\mathbf{0})\omega^{(2)}_{j_2j_4}(\mathbf{0}) + \omega^{(2)}_{j_2j_3}(\mathbf{0})\omega^{(2)}_{j_1j_4}(\mathbf{0}) + \omega^{(4)}_{j_1j_2j_3j_4}(\mathbf{0}) \\ &= \sigma_{j_1j_3}\sigma_{j_2j_4} + \sigma_{j_1j_4}\sigma_{j_2j_3} + \sum_{k=1}^{q}\prod_{l=1}^{p}\pi_{lk}^{(d_{l4})}\lambda_k. \end{aligned} \quad (3.2.15)$$

Finally the covariance matrix of first and second order standardized deviations is

$$\Psi = \mathrm{Cov}(\xi) = D^{-\frac{1}{2}}(\boldsymbol{\sigma})\,\mathbf{V}\,D^{-\frac{1}{2}}(\boldsymbol{\sigma}) \quad (3.2.16)$$

where

$$\begin{aligned}\xi &= D^{-\frac{1}{2}}(\boldsymbol{\sigma})\,\xi^*,\\ \xi^* &= \begin{pmatrix} (\mathbf{Y}-\boldsymbol{\mu}) \\ \zeta - \mathrm{vec}(\Sigma) \end{pmatrix}\\ \mathbf{V} &= \begin{pmatrix} \Sigma & \boldsymbol{\tau}' \\ \boldsymbol{\tau} & \boldsymbol{\kappa} \end{pmatrix},\end{aligned}$$

$D(\boldsymbol{\sigma}) = \mathrm{diag}(\sigma_{11}, \ldots \sigma_{pp}, \sigma_{11}^2, \sigma_{11}\sigma_{22}, \ldots, \sigma_{pp}\sigma_{(p-1)(p-1)}, \sigma_{pp}^2)$. Note that $\mathbf{V}$ is the covariance matrix of ξ^*.

III.2.1 Simulation Procedure

We could use the constructive definition 3.2 to simulate multivariate Poisson vectors with fixed variance covariance matrix Σ. The first step in this approach is to find a vector q dimensional vector $\boldsymbol{\lambda}$ and $(p \times q)$ matrix Θ as a function of Σ satisfying equation (3.2.12).

Sim (1993) proposed an algorithm that uses one-to-one transformation by restricting $p = q$ and Θ to lower triangular matrix with unit diagonals. Given $\boldsymbol{\lambda}$ and Θ, we can find Σ using (3.2.12). Conversely, given Σ, we can calculate $\boldsymbol{\lambda}$ and Θ uniquely, using the recursive formulas

$$\lambda_j = \sigma_{jj} - \sum_{k=1}^{j-1} \theta_{jk}\lambda_k \tag{3.2.17}$$

$$\theta_{kj} = \begin{cases} \dfrac{1}{\lambda_j}\left(\sigma_{jk} - \displaystyle\sum_{l=1}^{j-1} \theta_{jl}\theta_{kl}\lambda_l\right) & \text{if } j < k \\ 1 & \text{if } j = k \\ 0 & \text{otherwise.} \end{cases} \tag{3.2.18}$$

The algorithm for generating multivariate Poisson distribution using equations (3.2.17) and (3.2.18) fails if λ_j's are not positive or θ_{jk} lie outside $[0, 1]$. In this case the Σ may not consistent, in the sense that there may not be a multivariate Poisson distribution with variance covariance matrix Σ. In general there could be more than one pair solution to equation (3.2.12). This is often the case when the covariance matrix Σ is structured or the targeted value of q is greater than p. For example, when $\Sigma = \sigma^2 R(\alpha)$ with $\sigma^2 > 0$ is and $R(\alpha) = (1-\alpha)\mathbf{I}_p + \alpha\mathbf{J}_p$ is exchangeable correlation matrix, we can check that both formulations

$$\{\boldsymbol{\lambda}_1 = \sigma^2[(1-\alpha)\mathbf{1}_p' \ \alpha]' \ , \ \Theta_1 = [\mathbf{I}_p \ \mathbf{1}_p] \}\,,$$

and

$$\{\boldsymbol{\lambda}_2 = \sigma^2\mathbf{1}_{(p+1)} \ , \ \Theta_2 = [(1-\sqrt{\alpha})\mathbf{I}_p \ \sqrt{\alpha}\mathbf{1}_p] \}$$

satisfy equations (3.2.12). In the above $\mathbf{I}_p$ is the identity matrix of order p, and $\mathbf{1}_p$ is a column vector of ones of order p. The selection of pair $(\boldsymbol{\lambda}, \boldsymbol{\Theta})$ introduces additional constraints on the feasibility of the Σ matrix. The study of the mappings of Σ to $(\boldsymbol{\lambda}, \boldsymbol{\Theta})$ is an important problem, whose solution is beyond the scope of this dissertation. However it is intuitive when Σ is proper covariance matrix of positively correlated random variables there exist at least one pair of $(\boldsymbol{\lambda}, \boldsymbol{\Theta})$ which satisfies equations (3.2.12).

III.2.2 Epileptic Seizure Data

In this section we will study the performance of the MG and QLS estimation methods for the multivariate Poisson model given in Definition 3.2, using simulations based on the algorithm described in Section III.2.1. For the simulations we will use a model fitted for a real life data using some well established methods. Leppik et. al. (1985) conducted a 2×2 randomized double-blinded crossover clinical trial to study the effectiveness of anti-epileptic progabide drug on 59 patients suffering from simple or complex seizures. At each of the four successive post-randomization visits the number of seizures occurring during past two weeks were reported. The four pre-crossover responses were made available in a seminal paper by Thall and Vail (1990) and are reproduced in Table 3.1.

An eyeball of the data clearly indicates that the patient with id #207 is an outlier, since the baseline and the first visit seizure count is more than 100. As a preliminary analysis we have fitted the log-linear model that was given by Thall and Vail (1990) for the mean responses:

$$\log(\mu_{ij}) \ = \ \beta_0 + \text{BASE}_i\,\beta_1 + \text{TRT}_i\,\beta_2 + (\text{BASE}\times\ \text{TRT})_i\,\beta_3 + \text{AGE}_i\,\beta_5 + \text{VISIT4}_j\,\beta_6. \tag{3.2.19}$$

where BASE = log(0.25× baseline seizure count), TRT is a binary indicator for inclusion in treatment group, AGE = log(Age of patient) and VISIT4 is a binary indicator for the fourth visit. The modified Gaussian and the quasi-least squares

Table 3.1: Successive two-week seizure counts for 59 epileptics

Placebo group							Progabide group						
Id	Visit				Base	Age	Id	Visit				Base	Age
	1	2	3	4				1	2	3	4		
104	5	3	3	3	11	31	101	11	14	9	8	76	18
106	3	5	3	3	11	30	102	8	7	9	4	38	32
107	2	4	0	5	6	25	103	0	4	3	0	19	20
114	4	4	1	4	8	36	108	3	6	1	3	10	30
116	7	18	9	21	66	22	110	2	6	7	4	19	18
118	5	2	8	7	27	29	111	4	3	1	3	24	24
123	6	4	0	2	12	31	112	22	17	19	16	31	30
126	40	20	23	12	52	42	113	5	4	7	4	14	35
130	5	6	6	5	23	37	117	2	4	0	4	11	27
135	14	13	6	0	10	28	121	3	7	7	7	67	20
141	26	12	6	22	52	36	122	4	18	2	5	41	22
145	12	6	8	4	33	24	124	2	1	1	0	7	28
201	4	4	6	2	18	23	128	0	2	4	0	22	23
202	7	9	12	14	42	36	129	5	4	0	3	13	40
205	16	24	10	9	87	26	137	11	14	25	15	46	33
206	11	0	0	5	50	26	139	10	5	3	8	36	21
210	0	0	3	3	18	28	143	19	7	6	7	38	35
213	37	29	28	29	111	31	147	1	1	2	3	7	25
215	3	5	2	5	18	32	203	6	10	8	8	36	26
217	3	0	6	7	20	21	204	2	1	0	0	11	25
219	3	4	3	4	12	29	207	102	65	72	63	151	22
220	3	4	3	4	9	21	208	4	3	2	4	22	32
222	2	3	3	5	17	32	209	8	6	5	7	41	25
226	8	12	2	8	28	25	211	1	3	1	5	32	35
227	18	24	76	25	55	30	214	18	11	28	13	56	21
230	2	1	2	1	9	40	218	6	3	4	0	24	41
234	3	1	4	2	10	19	221	3	5	4	3	16	32
238	13	15	13	12	47	22	225	1	23	19	8	22	26
							228	2	3	0	1	25	21
							232	0	0	0	0	13	36
							236	1	4	3	2	12	37

Table 3.2: Seizure Data: MG Estimates

Covariate	With Patient # 207			Without Patient # 207		
	Estimate	S.E.	p-value	Estimate	S.E.	p-value
Intercept	-2.7729	0.9489	0.0035	-2.3407	0.8766	0.0076
BASE	0.9499	0.0974	0.0000	0.9505	0.0973	0.0000
TRT	-1.3401	0.4272	0.0017	-0.5206	0.4164	0.2112
BASE× TRT	0.5627	0.1742	0.0012	0.1383	0.1941	0.4763
AGE	0.9011	0.2756	0.0011	0.7722	0.2550	0.0025
VISIT4	-0.1611	0.0656	0.0140	-0.1479	0.0763	0.0527
α	0.1906	0.2731	0.4853	0.1819	0.2765	0.5106

Table 3.3: Seizure Data: QLS Estimates

Covariate	With Patient # 207			Without Patient # 207		
	Estimate	S.E.	p-value	Estimate	S.E.	p-value
Intercept	-2.7939	0.9561	0.0035	-2.3579	0.8838	0.0076
BASE	0.9504	0.0987	0.0000	0.9509	0.0983	0.0000
TRT	-1.3386	0.4296	0.0018	-0.5196	0.4185	0.2145
BASE× TRT	0.5633	0.1749	0.0013	0.1388	0.1947	0.4758
AGE	0.9066	0.2772	0.0011	0.7768	0.2567	0.0025
VISIT4	-0.1611	0.0656	0.0140	-0.1479	0.0763	0.0527
α	0.3582	0.2547	0.1596	0.3393	0.2621	0.1955

estimates with and and without patient #207 are given in Tables 3.2 and 3.3. The results with patient #207 included are different from the results without that patient. For example, we can see from the table, the treatment and the baseline-treatment interaction become insignificant if we exclude patient #207.

For our simulations we have used the above model (3.2.19). for the mean of the responses. To avoid additional variability of estimates, patient #207 was excluded from our simulations. Details of the simulations are described in the following steps:

Description of the simulation steps:

1. We fixed the true parameter $\boldsymbol{\beta} = \boldsymbol{\beta}_0$, where

$$\boldsymbol{\beta}_0 = (-2.3574, 0.9509, -0.5196, 0.1388, 0.7767, -0.1479)'$$

is the GEE estimate for the epileptic seizure data given in Table 3.1 obtained using an exchangeable correlation structure and excluding patient with id #207.

2. For the true correlation matrix we used the exchangeable structure and later repeated the whole simulation procedure for the autoregressive and then for the moving average of order one correlation structures.

3. We varied the true correlation parameter α from 0 to $\alpha_{\max}$ in increments of 0.0125, where $\alpha_{\max}$ is the upper extreme of feasible region.

4. For each value of α, $N = 5000$ datasets consisting of 59 multivariate Poisson random vectors were simulated using Sim (1993)'s lower triangular mapping.

5. For the ith simulated data set we calculated $\widehat{\theta}_{m_i}$ and $\widehat{\theta}_{q_i}$, the MG and QLS estimates of θ, respectively. We also computed the covariance matrices $\widehat{\mathrm{Cov}}(\widehat{\theta}_{m_i}) = \Sigma_m(\widehat{\theta}_{m_i})$ and $\widehat{\mathrm{Cov}}(\widehat{\theta}_{q_i}) = \Sigma_q(\widehat{\theta}_{q_i})$. In calculating these matrices we have used formulas given by (3.2.16) for the Ψ matrix.

6. The infeasibility or divergent solution probability is estimated for each of the estimates $\widehat{\theta}_m$ and $\widehat{\theta}_q$ as the proportion of times the estimate did not converge or the estimate is deemed to be inconsistent according to Sim's mapping (3.2.18). The infeasible/divergent cases were discarded from further analysis.

7. The joint asymptotic relative efficiency of $\widehat{\theta}_q$ with respect to $\widehat{\theta}_m$ is computed using trace as well as the determinant of the matrix

$$\Gamma^* = \left\{ \frac{1}{N_c} \sum_{i=1}^{N_c} \Sigma_q(\widehat{\theta}_{qi}) \right\}^{-1} \left\{ \frac{1}{N_c} \sum_{i=1}^{N_c} \Sigma_m(\widehat{\theta}_{mi}) \right\} \tag{3.2.20}$$

where N_c is the number of simulated data sets that yielded convergent estimates for both MG and the QLS methods.

8. We calculated the coverage probability for simultaneous confidence region of the MG method as

$$\frac{1}{N_c} \sum_{i=1}^{N_c} I\left((\widehat{\theta}_{mi} - \theta_0)' \widehat{\Sigma}_{mi}^{-1} (\widehat{\theta}_{mi} - \theta_0) \leq \chi^2_{K+q,0.05} \right), \tag{3.2.21}$$

where I is the indicator function. The coverage probability for QLS method is calculated similarly.

III.2.3 Discussion of Simulation Results

In this section we report the findings of our simulations. Consider first the case where the true correlation structure is exchangeable. Figure 3.1 contains the plots of estimated probabilities (proportions of simulated data sets) as a function of α, where the MG and QLS estimates of $\theta = (\beta, \alpha, \phi)$ did not converge or the final values of the estimates are deemed to be invalid according the mappings (3.2.17) and (3.2.18). It is clear from the plots, both the MG and QLS methods have similar infeasibility problems. The estimated probabilities are low when α is in the interior of the feasible range. On the other hand, the proportions of invalid estimates are high when α is close to zero; the reason being the estimate of α could be negative in this case and Sim's algorithm is valid only for positively correlated Poisson variables and negative estimates of α are automatically discarded.

Figure 3.2 contains plot of simultaneous coverage probability of the 95% confidence ellipsoids as defined in (3.2.21) for both the MG and QLS estimates. The coverage probabilities for the QLS method are closer to the nominal level compared to the coverage probabilities of the MG method, over the entire range of α. Thus when the true structure is exchangeable, confidence ellipsoids constructed using QLS are preferable than those constructed using the MG method. Table 3.4 contains the asymptotic relative efficiencies of QLS estimates with respect to the MG estimates. We have presented efficiencies for two regression coefficients; $\widehat{\beta}_0$ is the coefficient of a time independent covariate, and $\widehat{\beta}_5$ is the coefficient of a time dependent covariate. Table 3.4 shows that QLS estimate of the regression coefficient for time independent covariate is more efficient, whereas the opposite is true for the regression coefficient of time dependent covariate.

Our simulations also show that the QLS estimate of α is less efficient than the MG estimate, and furthermore, the efficiency is decreasing as α increases. However, for the regression parameter as a whole, the overall efficiency of the QLS method as measured by the determinant (Λ_1) or the trace (ν_1) criteria of the submatrix of Γ^* defined in (3.2.20), is more than the MG estimate, over the entire feasible range of the correlation parameter α. Plots of these efficiencies are in Figure 3.3. Note that the efficiencies are increasing functions of α. Hence, when the regression parameter is of primary interest, and if the covariates are time independent or mixed, and if the

correlation is exchangeable but the parameter α is treated as a nuisance parameter, then QLS is preferable over the MG method.

Table 3.4 also shows that the overall efficiency of the QLS estimates of the regression and correlation parameter with respect to the MG estimates, as measured by the determinant (Λ_2) or the trace criteria (ν_2) of the matrix Γ^* defined in (3.2.20). For small values of α, the overall performance of the QLS is better than the MG estimates. But when there is a strong correlation, MG estimates tend to be more efficient than the QLS estimates.

Simulation results concerning infeasibility, coverage probabilities of the confidence ellipsoids, and the asymptotic relative efficiencies of the regression parameter, when the true correlation structure is autoregressive of order 1 (AR(1)) are presented in Figures 3.4, 3.5 and 3.6, respectively. Table 3.5 contains asymptotic relative efficiencies for the AR(1) structure, similar to the ones that we presented for the exchangeable case in Table 3.4. An examination of the figures and the table of efficiencies shows that the behavior of the QLS method with respect to MG method is very similar to that when the true correlation is exchangeable. However, when the true correlation is moving average of order one (MA(1)), the behavior of the efficiency of the QLS estimate of the regression parameter is different from the other structures as shown in Figure 3.9. The efficiency as a function of α is approximately concave with a maximum in a neighborhood of $\alpha = 0.3$. But note that the efficiency is more than over the entire range of α. The infeasible probabilities and coverage probabilities for the MA(1) structure are plotted in Figures 3.7 and 3.8 respectively. Table 3.6 contains various efficiencies for the MA(1) structure. When the true structure is MA(1), the coverage probabilities of the MG estimates are closer to the nominal level than the coverage probabilities of the QLS estimates.

In summary, the simulation results show that for all the three commonly occurring correlation structures, when the covariates are time independent or mixed the QLS estimate of the regression parameter as a whole is more efficient than the MG estimate. But if all the covariates are time varying or if the correlation parameter is not a nuisance parameter, MG is preferable compared to the QLS estimation method.

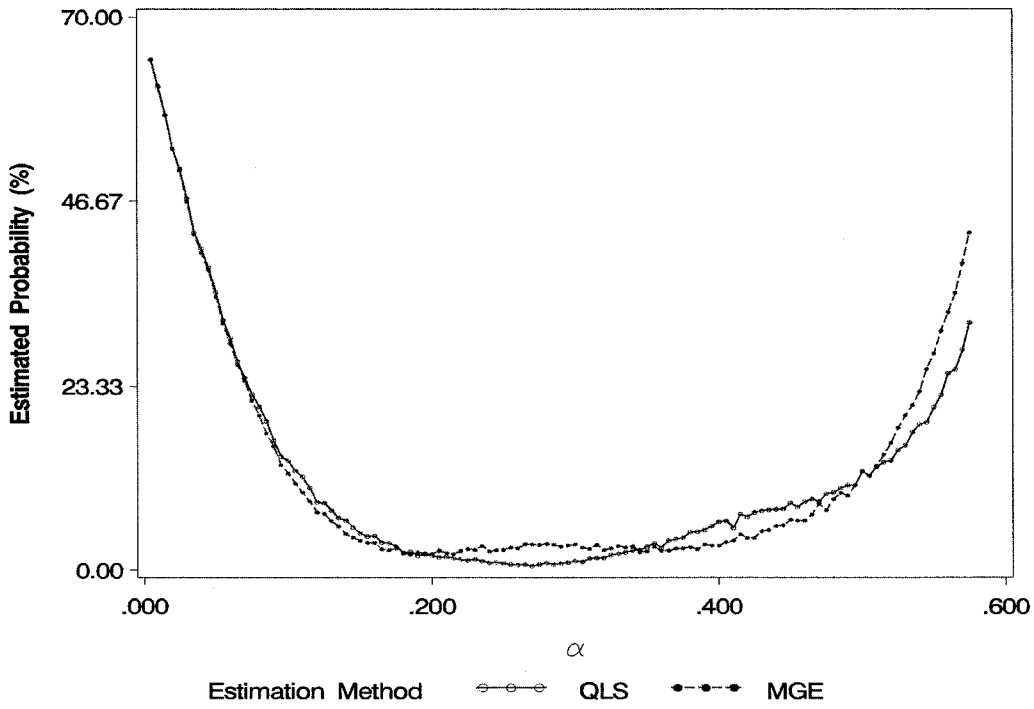

Figure 3.1: EXCH: Infeasibility/divergent solutions probabilities.

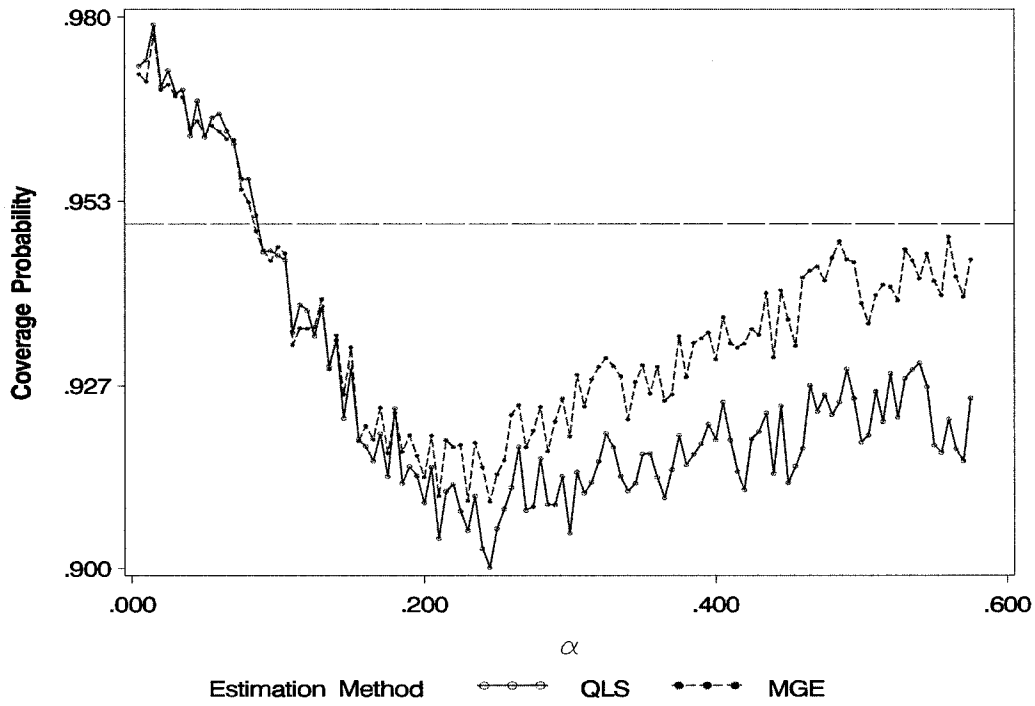

Figure 3.2: EXCH: Coverage probability of simultaneous confidence region.

Table 3.4: EXCH: ARE of QLS vs. MGE

α_0	N_c	$\widehat{\beta}_0$	$\widehat{\beta}_5$	$\widehat{\alpha}$	Λ_1	ν_1	Λ_2	ν_2
0.02	2294	1.0059	0.99765	0.99474	1.0277	6.0274	1.0226	7.0234
0.04	2929	1.0069	0.99716	0.99041	1.0323	6.0319	1.0228	7.0242
0.06	3470	1.0079	0.99668	0.98661	1.0365	6.0360	1.0231	7.0251
0.08	3909	1.0087	0.99616	0.97974	1.0402	6.0396	1.0195	7.0227
0.10	4256	1.0108	0.99499	0.97301	1.0501	6.0492	1.0221	7.0266
0.12	4516	1.0119	0.99421	0.96431	1.0548	6.0538	1.0173	7.0237
0.14	4644	1.0131	0.99324	0.95438	1.0603	6.0590	1.0118	7.0205
0.16	4722	1.0159	0.99137	0.94271	1.0729	6.0711	1.0107	7.0226
0.18	4826	1.0174	0.98997	0.93000	1.0794	6.0772	1.0023	7.0177
0.20	4836	1.0190	0.98835	0.91434	1.0862	6.0837	0.9905	7.0103
0.22	4826	1.0205	0.98673	0.89914	1.0922	6.0894	0.9782	7.0026
0.24	4844	1.0226	0.98457	0.88340	1.1009	6.0975	0.9672	6.9968
0.26	4834	1.0244	0.98231	0.86586	1.1084	6.1045	0.9525	6.9881
0.28	4789	1.0265	0.97970	0.84824	1.1167	6.1123	0.9381	6.9800
0.30	4781	1.0284	0.97690	0.82919	1.1240	6.1191	0.9208	6.9695
0.32	4801	1.0300	0.97402	0.80735	1.1295	6.1243	0.8981	6.9547
0.34	4729	1.0322	0.97057	0.78784	1.1376	6.1319	0.8800	6.9443
0.36	4743	1.0323	0.96867	0.76709	1.1358	6.1305	0.8527	6.9235
0.38	4631	1.0343	0.96482	0.74564	1.1420	6.1363	0.8303	6.9094
0.40	4556	1.0357	0.96127	0.72526	1.1456	6.1398	0.8073	6.8937
0.42	4488	1.0344	0.96027	0.70581	1.1375	6.1326	0.7771	6.8680
0.44	4388	1.0361	0.95564	0.68210	1.1414	6.1365	0.7503	6.8492
0.46	4318	1.0351	0.95428	0.66302	1.1340	6.1299	0.7217	6.8244
0.48	4175	1.0351	0.95124	0.64035	1.1302	6.1266	0.6915	6.7994
0.50	3960	1.0350	0.94873	0.62295	1.1267	6.1236	0.6681	6.7795
0.52	3842	1.0356	0.94502	0.60452	1.1258	6.1232	0.6453	6.7612
0.54	3510	1.0344	0.94392	0.5878	1.1181	6.1162	0.6209	6.7379
0.56	2991	1.0374	0.93672	0.57174	1.1258	6.1241	0.6063	6.7300

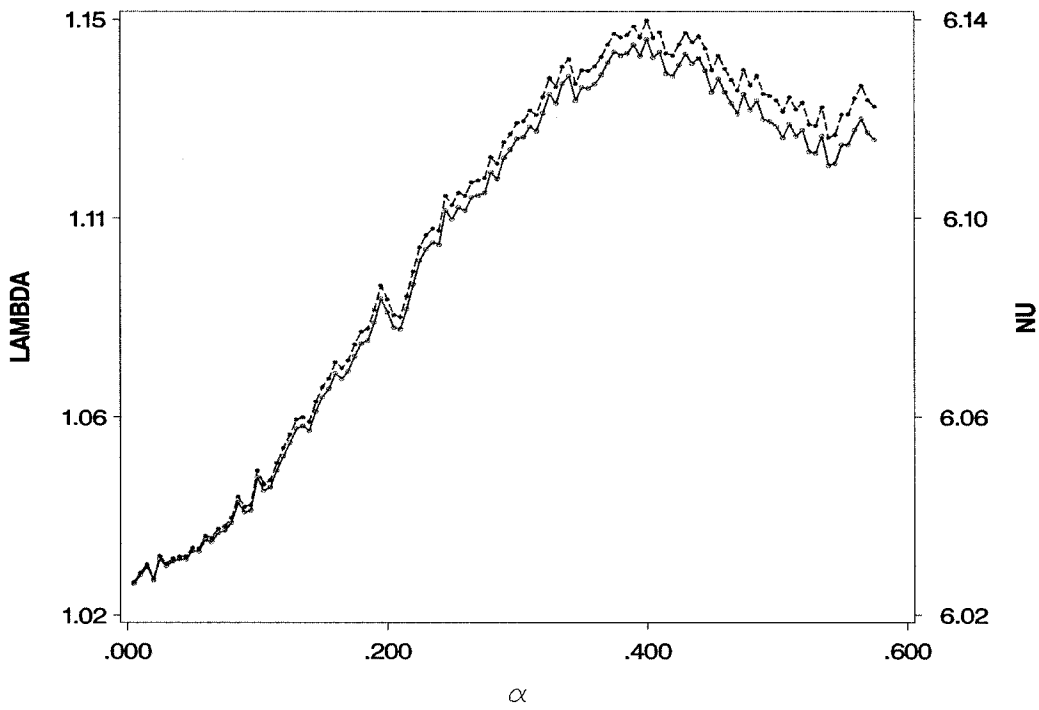

Figure 3.3: EXCH: ARE of $\hat{\beta}_q$ vs. $\hat{\beta}_m$.

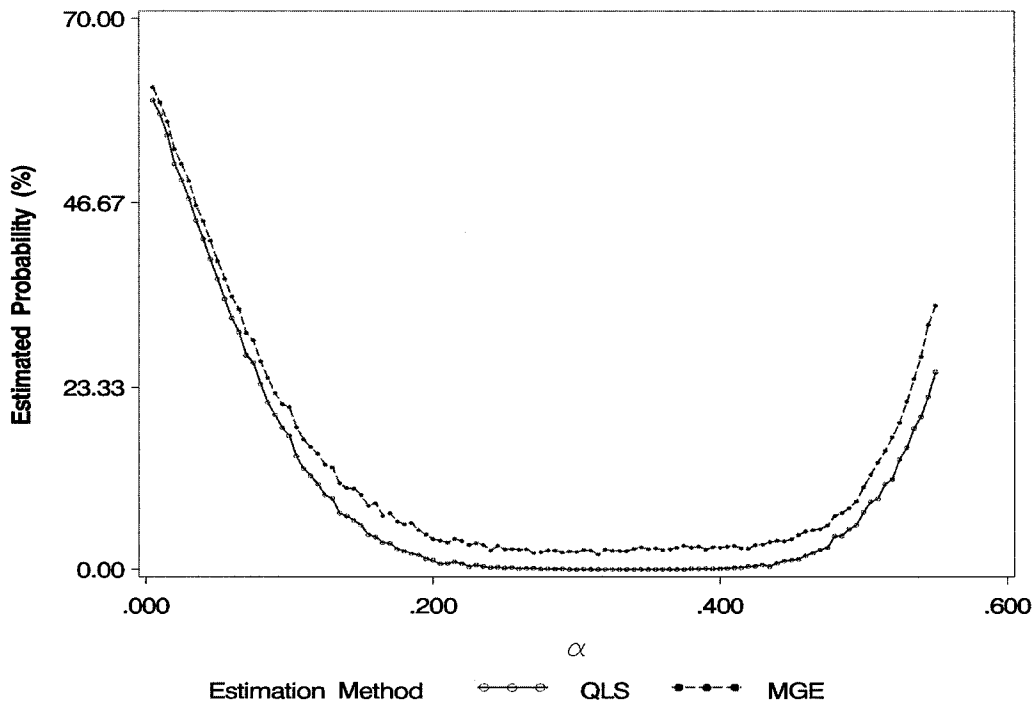

Figure 3.4: AR(1): Infeasibility/divergent solutions probabilities.

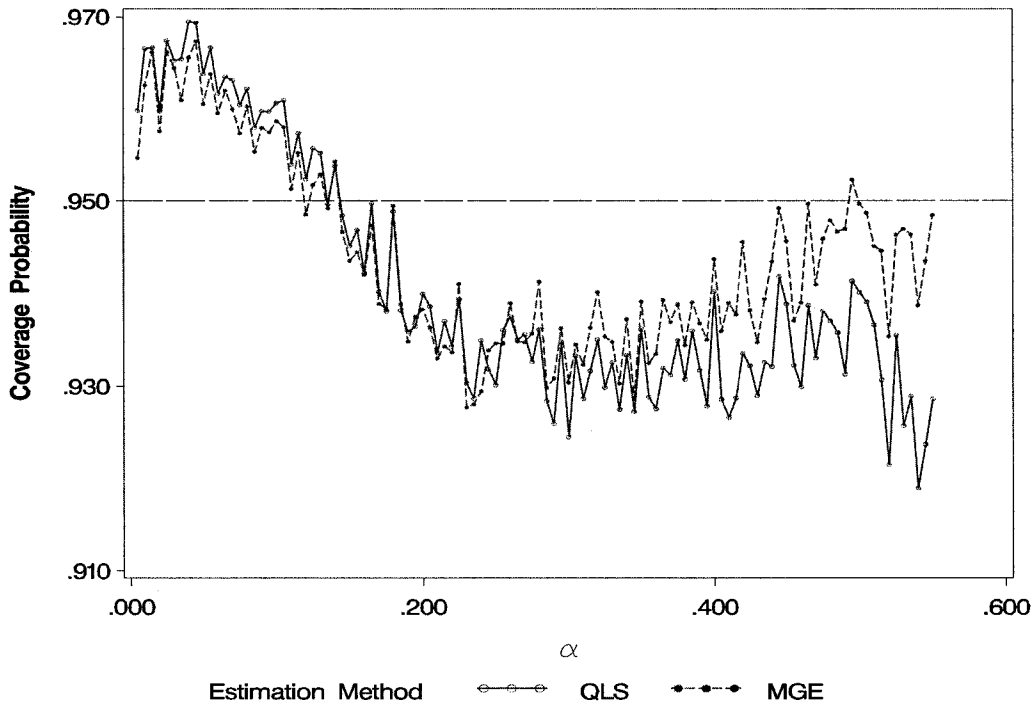

Figure 3.5: AR(1): Coverage probability of simultaneous confidence region.

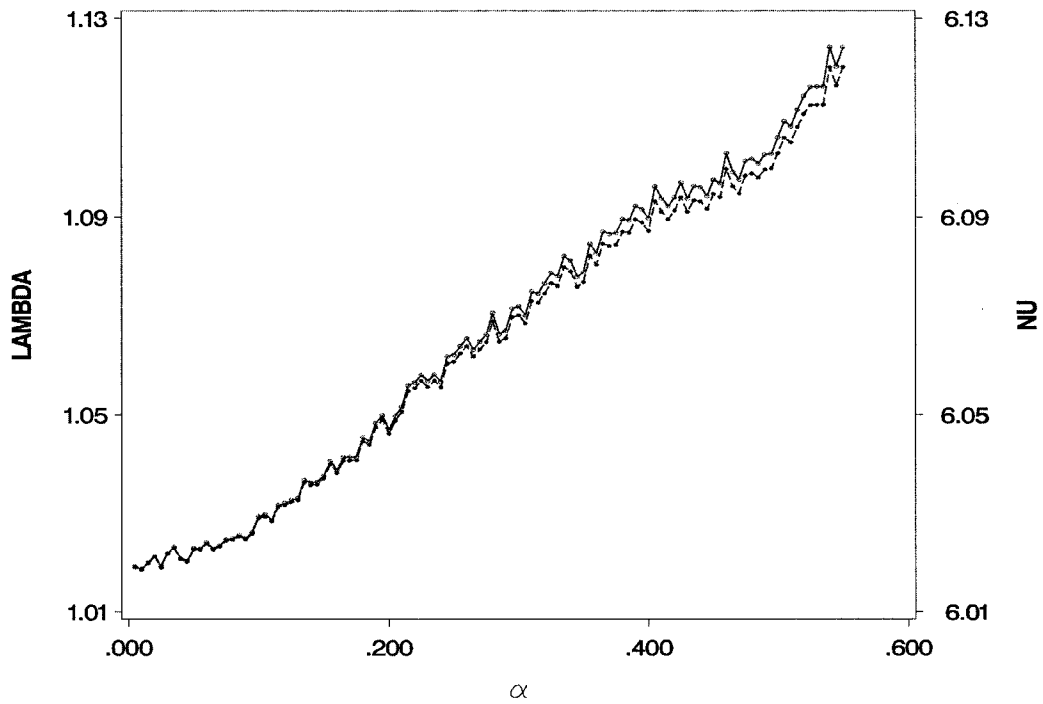

Figure 3.6: AR(1): ARE of $\hat{\beta}_q$ vs. $\hat{\beta}_m$.

Table 3.5: AR(1): ARE of QLS vs. MGE

α_0	N_c	$\widehat{\beta_0}$	$\widehat{\beta_5}$	$\widehat{\alpha}$	Λ_1	ν_1	Λ_2	ν_2
0.02	2328	1.0045	0.99881	0.98929	1.0213	6.0212	1.0107	7.0115
0.04	2784	1.0044	0.9988	0.9868	1.0209	6.0207	1.0078	7.0088
0.06	3256	1.0051	0.99855	0.98446	1.0243	6.0241	1.0088	7.0101
0.08	3669	1.0053	0.99843	0.98114	1.0251	6.0249	1.0063	7.008
0.10	3958	1.0062	0.99801	0.9757	1.0296	6.0292	1.0051	7.0075
0.12	4254	1.0068	0.9977	0.97061	1.0323	6.0319	1.0025	7.0057
0.14	4474	1.0077	0.99729	0.96438	1.0364	6.0359	1.0001	7.0042
0.16	4572	1.0083	0.9969	0.95625	1.0389	6.0383	0.994	6.9995
0.18	4707	1.0097	0.99611	0.94722	1.0454	6.0447	0.99067	6.9978
0.20	4801	1.0101	0.99579	0.93832	1.047	6.0462	0.9826	6.9915
0.22	4812	1.0122	0.99443	0.92653	1.0565	6.0554	0.97865	6.99
0.24	4875	1.0123	0.99396	0.91488	1.0567	6.0555	0.9659	6.9797
0.26	4869	1.0142	0.99255	0.90177	1.0653	6.0638	0.95906	6.976
0.28	4874	1.0155	0.99134	0.88755	1.0706	6.0688	0.9476	6.968
0.30	4885	1.0159	0.99067	0.87507	1.0719	6.0701	0.93437	6.9577
0.32	4867	1.0171	0.98917	0.85847	1.0765	6.0745	0.91908	6.9467
0.34	4864	1.0182	0.98764	0.84185	1.0812	6.0789	0.90341	6.9355
0.36	4873	1.0188	0.98653	0.8269	1.0826	6.0804	0.88682	6.9228
0.38	4852	1.0205	0.98427	0.80864	1.0896	6.087	0.87058	6.912
0.40	4852	1.0208	0.983	0.79114	1.0898	6.0872	0.84933	6.8954
0.42	4847	1.0221	0.98087	0.77332	1.0941	6.0913	0.83091	6.8822
0.44	4791	1.0228	0.97898	0.7559	1.0961	6.0933	0.81089	6.8672
0.46	4710	1.0248	0.97593	0.73788	1.103	6.0999	0.79363	6.8561
0.48	4546	1.0249	0.97456	0.72189	1.1019	6.0989	0.77272	6.8393
0.50	4312	1.0262	0.97185	0.7051	1.1062	6.1031	0.75469	6.8268
0.52	3948	1.0286	0.96797	0.68886	1.1145	6.111	0.74001	6.8185
0.54	3381	1.0313	0.96383	0.67498	1.1243	6.1202	0.72904	6.8138

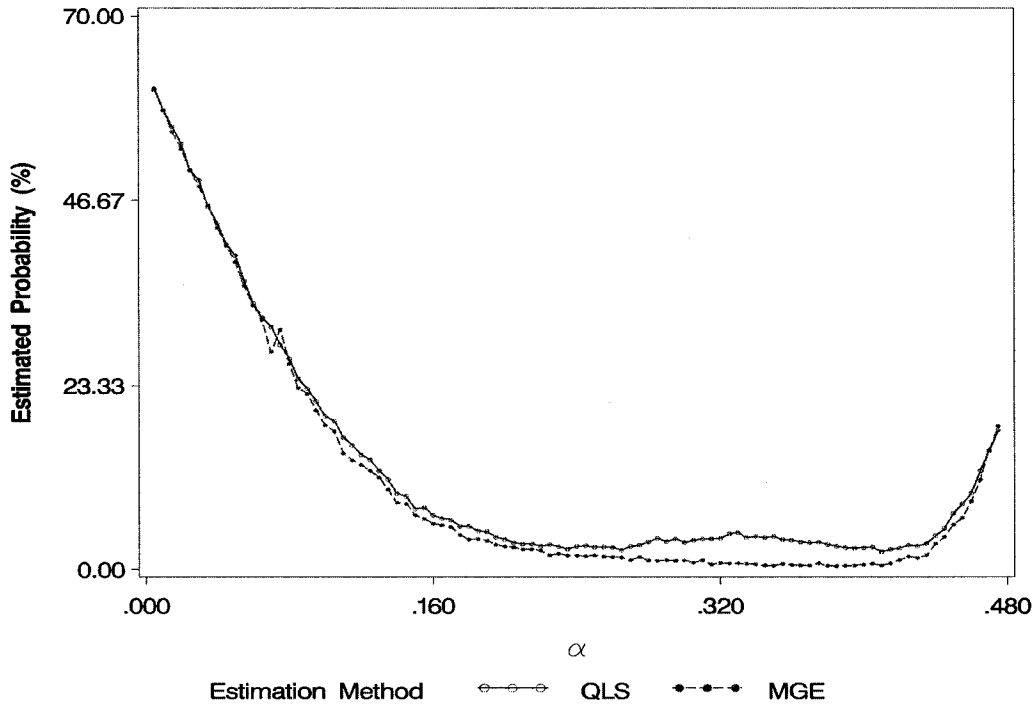

Figure 3.7: MA(1): Infeasibility/divergent solutions probabilities.

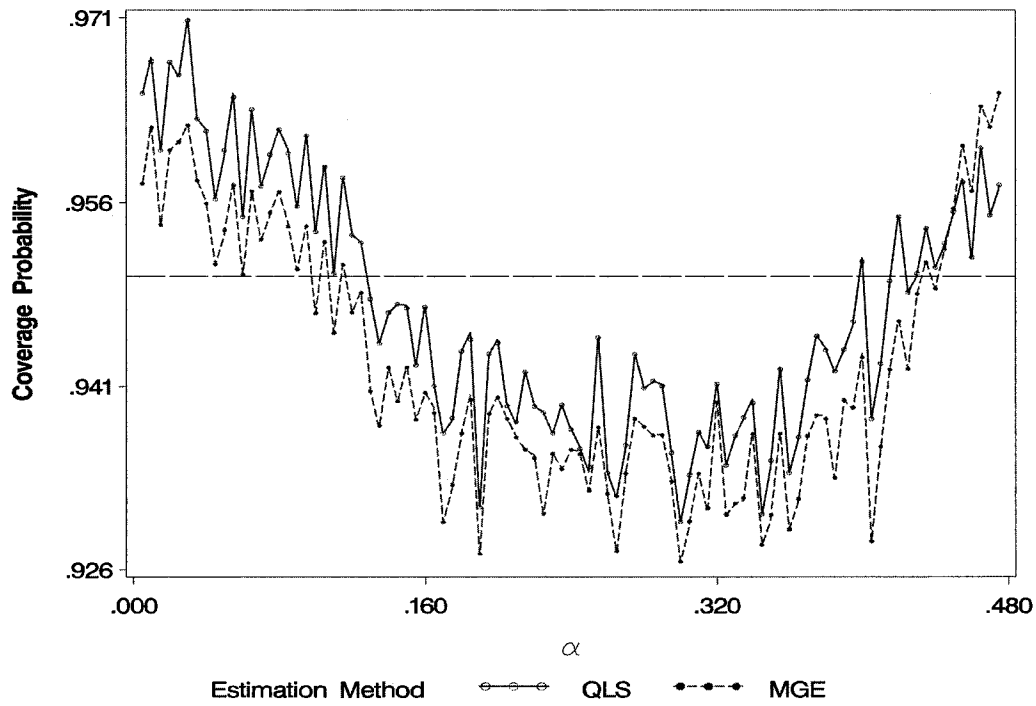

Figure 3.8: MA(1): Coverage probability of simultaneous confidence region.

Table 3.6: MA(1): ARE of QLS vs. MGE

α_0	N_c	$\widehat{\beta_0}$	$\widehat{\beta_5}$	$\widehat{\alpha}$	Λ_1	ν_1	Λ_2	ν_2
0.02	2240	1.0055	0.99851	0.98277	1.0263	6.026	1.0089	7.0098
0.04	2722	1.0055	0.99837	0.97905	1.0263	6.0261	1.0053	7.0064
0.06	3211	1.0055	0.99833	0.97535	1.0261	6.0258	1.0013	7.0027
0.08	3544	1.0061	0.99801	0.96947	1.0289	6.0286	0.99808	7.0001
0.10	3924	1.0069	0.99756	0.96254	1.0325	6.0321	0.99451	6.9972
0.12	4157	1.0077	0.99709	0.95388	1.0360	6.0355	0.98899	6.9926
0.14	4420	1.0086	0.99656	0.9429	1.0400	6.0394	0.98144	6.9863
0.16	4553	1.0091	0.99596	0.93282	1.0421	6.0415	0.97306	6.9791
0.18	4632	1.0104	0.99509	0.91897	1.0478	6.0470	0.96387	6.9717
0.20	4706	1.0109	0.99429	0.90369	1.0499	6.0490	0.94976	6.9595
0.22	4766	1.0113	0.99351	0.88659	1.0510	6.0501	0.93266	6.9445
0.24	4784	1.0119	0.99242	0.86683	1.0527	6.0517	0.91317	6.9275
0.26	4786	1.0126	0.99122	0.84730	1.0553	6.0542	0.89456	6.9115
0.28	4769	1.0131	0.98963	0.82459	1.0562	6.0552	0.87098	6.8907
0.30	4773	1.0134	0.98801	0.79927	1.0559	6.0549	0.8435	6.8661
0.32	4761	1.0129	0.98728	0.77505	1.0525	6.0516	0.81476	6.8396
0.34	4756	1.0122	0.98580	0.74544	1.0473	6.0467	0.77898	6.8059
0.36	4787	1.0122	0.98385	0.71751	1.0455	6.0450	0.74775	6.7769
0.38	4819	1.0114	0.98221	0.68986	1.0396	6.0393	0.71395	6.744
0.40	4836	1.0113	0.97905	0.65648	1.0357	6.0356	0.67576	6.7071
0.42	4826	1.0109	0.97622	0.62663	1.0307	6.0309	0.64077	6.6725
0.44	4687	1.0105	0.97231	0.59418	1.0246	6.0250	0.60266	6.6340
0.46	4293	1.0112	0.96488	0.55572	1.0202	6.0210	0.55974	6.5910

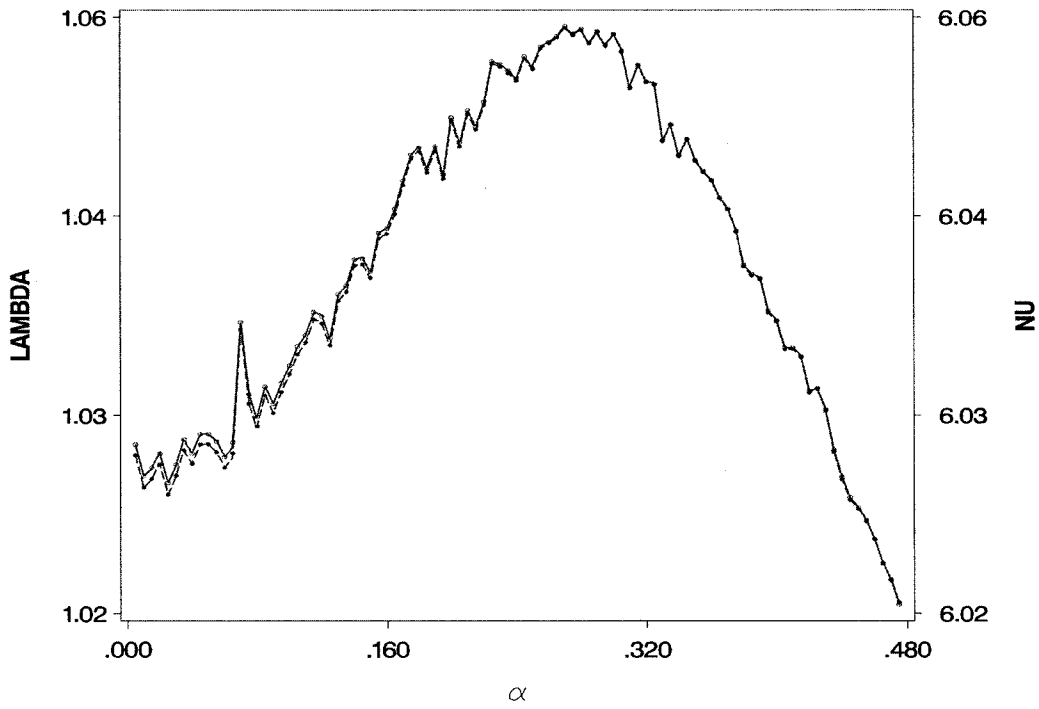

Figure 3.9: MA(1): ARE of $\hat{\beta}_q$ vs. $\hat{\beta}_m$.

III.3 Multivariate binary simulations

In this section we will compare performance of MG and the QLS methods for binary outcomes. As with the case of Poisson counts, we will first discuss a method of simulating correlated binary data in an efficient manner. Numerous simulation methods for generating binary variables with given means and correlations have been proposed in the literature. Emrich and Piedmonte (1991) described a method based on dichotomization of multivariate standard normal variables. Their method allows unequal means and negative correlations but requires solving a system of complex non-linear equations, that are computationally intensive. Park et al. (1996) developed a method which generates first correlated Poisson variables and those variables are truncated to yield correlated binary variables. Another method based on finite mixture of independent Bernoulli variables was discussed by Lunn and Davies (1998). Their method is easy to implement, but it is restricted to the case of equal means. Lee (1993) developed a method that computes the full joint distribution of correlated binary random variables using linear programming approach. Recently, Gange (1995) described an iterative log-linear model formulation that enables computing

the full joint distribution of the binary variables with given marginals and correlations. More recently, Qaqish (2003) presented a method of generating conditional linear family of multivariate Bernoulli distributions.

We introduce below a new family of multivariate binary distributions using the multivariate Poisson distributions discussed in the previous section. This family, which is easy to simulate, contains positively correlated binary random variables with fixed marginal probabilities and a given correlation matrix.

Definition 3.1 *Let* $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_p)$ *be a random vector distributed as multivariate Poisson, as defined in 3.2. Suppose that* Σ*, the covariance matrix of* $\mathbf{Y}$*, can be expressed as a function of a* $(p \times q)$ *weight matrix* Θ *and* $\Lambda = (\lambda_1, \lambda_2, \ldots, \lambda_q)$ *as given by the equations (3.2.12). If*

$$W_j = I(Y_j = 0); \quad j = 1, 2 \ldots, p,$$

then $\mathbf{W} = (W_1, W_2, \ldots, W_p)$ *is a multivariate correlated binary random vector.*

Note that neither the joint distribution nor the moment generating function of $\mathbf{W}$ is in a closed form. But the joint moments can be obtained replacing the quantities $e^{t_j}, j = 1, 2 \ldots, p$ with zero in the joint moment generating function of the random vector $\mathbf{Y}$ given in (3.2.6)). Hence we have

$$E\left[\prod_{k=1}^{c} W_{j_k}^{r_{j_k}}\right] = \exp\left[\sum_{j=1}^{q}\left(\prod_{k=1}^{c}(1-\theta_{j_k j}) - 1\right)\lambda_j\right] \tag{3.3.1}$$

where $j_k = 1, 2, \ldots, p$; $r_{j_k} = 1, 2, \ldots;$; $k = 1, 2, \ldots, c$ and $c = 1, 2, \ldots, p$.

After a long tedious derivations we obtain the following lemma consisting of higher order central moments needed to perform asymptotic analysis of MG and the QLS estimation methods.

Lemma 3.2 *Let* $\mathbf{W} = (W_1, W_2, \ldots, W_p)$ *be a random vector as defined in 3.1. Then*

$$
\begin{aligned}
\pi_{j_1} &= \mathrm{E}(W_{j_1}) = e^{-\sigma_{j_1 j_1}} \\
\sigma^*_{j_1 j_2} &= \mathrm{Cov}((W_{j_1} - \pi_{j_1}), (W_{j_2} - \pi_{j_2})) = e^{-\sigma_{j_1 j_1} - \sigma_{j_2 j_2}}(e^{\sigma_{j_1 j_2}} - 1) \\
\tau^*_{j_1 j_2 j_3} &= \mathrm{Cov}((W_{j_1} - \pi_{j_1})(W_{j_2} - \pi_{j_2}) - \sigma^*_{j_1 j_2}, (W_{j_3} - \pi_{j_3})) \\
&= e^{-\sigma_{j_1 j_1} - \sigma_{j_2 j_2} - \sigma_{j_3 j_3}} \left[e^{\sigma_{j_1 j_2} + \sigma_{j_1 j_3} + \sigma_{j_2 j_3} - \tau_{j_1 j_2 j_3}} - e^{\sigma_{j_1 j_2}} - e^{\sigma_{j_1 j_3}} - e^{\sigma_{j_2 j_3}} + 2 \right] \\
\kappa^*_{j_1 j_2 j_3 j_4} &= \mathrm{Cov}((W_{j_1} - \pi_{j_1})(W_{j_2} - \pi_{j_2}), (W_{j_3} - \pi_{j_3})(W_{j_4} - \pi_{j_4})) \\
&= e^{-\sigma_{j_1 j_1} - \sigma_{j_2 j_2} - \sigma_{j_3 j_3} - \sigma_{j_4 j_4}} \{ \exp(\sigma_{j_1 j_2} + \sigma_{j_1 j_3} + \sigma_{j_1 j_4} + \sigma_{j_2 j_3} + \sigma_{j_2 j_4} + \sigma_{j_3 j_4} \\
&\quad - \sigma_{j_1 j_3}\sigma_{j_2 j_4} - \sigma_{j_1 j_4}\sigma_{j_2 j_3} - \tau_{j_1 j_2 j_3} - \tau_{j_1 j_2 j_4} - \tau_{j_1 j_3 j_4} - \tau_{j_2 j_3 j_4} + \kappa_{j_1 j_2 j_3 j_4}) \\
&\quad - e^{\sigma_{j_1 j_2} + \sigma_{j_1 j_3} + \sigma_{j_2 j_3} - \tau_{j_1 j_2 j_3}} - e^{\sigma_{j_1 j_2} + \sigma_{j_1 j_4} + \sigma_{j_2 j_4} - \tau_{j_1 j_2 j_4}} \\
&\quad - e^{\sigma_{j_1 j_3} + \sigma_{j_1 j_4} + \sigma_{j_3 j_4} - \tau_{j_1 j_3 j_4}} - e^{\sigma_{j_2 j_3} + \sigma_{j_2 j_4} + \sigma_{j_3 j_4} - \tau_{j_2 j_3 j_4}} - (e^{\sigma_{j_1 j_2}} - 1)(e^{\sigma_{j_3 j_4}} - 1) \\
&\quad + e^{\sigma_{j_1 j_2}} + e^{\sigma_{j_1 j_3}} + e^{\sigma_{j_1 j_4}} + e^{\sigma_{j_2 j_3}} + e^{\sigma_{j_2 j_4}} + e^{\sigma_{j_3 j_4}} - 3 \}
\end{aligned}
$$

where $\sigma_{j_1 j_2}, \tau_{j_1 j_2 j_3}$ and $\kappa_{j_1 j_2 j_3 j_4}$ are defined in (3.2.12), (3.2.14) and (3.2.15).

Given a marginal mean vector $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_p)$ and a correlation matrix $R = [r_{jk}]$, we can simulate binary random variables defined in 3.1 by first computing the inverse relation

$$
\sigma_{j_1 j_2} = \begin{cases} -\log(\pi_j), & \text{if } j = j_1 = j_2; \\ \log\left(1 + r_{j_1 j_2}\sqrt{\left(\frac{\pi_{j_1}}{1-\pi_{j_1}}\right)\left(\frac{\pi_{j_2}}{1-\pi_{j_2}}\right)}\right), & \text{otherwise} \end{cases} \tag{3.3.2}
$$

and using the Sim's algorithm. We omit the details.

III.3.1 Coronary Artery Risk Development in Young Adults (CARDIA)

Here we present real life example of a correlated binary data. The example is a study of coronary artery risk development in young adults (CARDIA). This is a collaborative, longitudinal study designed to increase understanding of the contributors to changes in cardiovascular disease risk factors during the critical years of transition from adolescence through young adulthood to middle age. The study is sponsored by the National Heart, Lung and Blood Institute (NHLBI). It involved data collection at approximately two to three year intervals for 5115 participants aged 18 to 30 years recruited at four centers, Chicago,IL; Minneapolis, MN; Birmingham, AL;

and Oakland, CA; during 1985-1986. The demographic data such as race, gender, education level and age were also collected in addition to standard measurements of blood pressure, anthropy, blood lipids, self reported smoking behavior, physical diet, pulpmonary function and many psychological factors. The participants were followed for 15 years and repeated measurements were collected at years 0, 2, 5, 7, 10, and 15.

The data consisting of the first four smoking status measures and demographic variables for 5078 participants, was studied by Preisser et al. (2000) and was made available online at *http://www.bios.unc.edu/~jpreisse*. The objective of their analysis was to identify patterns of smoking behavior with respect to demographic variables. As is always the case with long-duration longitudinal studies CARDIA study also has some drop-outs and missing responses.

Table 3.7 contains a detailed description of missing data categorized by the demographic variables. One can see that the study cohort is evenly balanced with respect to demographical measures such as gender, race, age and education. Table 3.7 suggests that participants with high school or less education are more difficult to follow up than participants with some college experience, an alarming level of 33.6 % percentage of participants were lost to follow up. The likelihood of missing-ness for younger participants is greater than that of adult participants. The white participants are easier to follow up than black participants. The missing-ness patterns of female participants is similar to that of male participants.

Preisser et al. (2000) conducted a test of covariate dependent missing-ness assumption using logistic regression analysis to predict the probability of missing response treating race and sex as predictors. They concluded that the missing-ness patterns are covariate dependent also known as (MAR-X). As a result the standard GEE approach is not applicable. The four different models for each race and sex combinations is proposed using artificially created monotonically missing data which treats all the responses observed after first missing occurrence as missing even if they are not. The regression parameters are computed using the weighted generalized estimating equations (WGEE) approach of Robins et al. (1995) in conjunction with odd-ratio based modelling approach of Fitzmaurice and Laird (1993) by assigning reciprocals of missing-ness probabilities as weights.

Table 3.7: CARDIA Data: Missing-ness Patterns

		% of missing at year		
Variable	Total	2	5	7
Age(yrs)				
18-22	1421	13.30	19.42	25.62
23-26	1604	9.41	14.71	19.70
27-30	2052	8.48	12.43	17.35
Sex				
Female	2722	9.81	14.47	19.81
Male	2405	10.50	15.58	21.13
Race				
Black	2618	13.98	19.75	25.63
White	2459	6.02	10.17	14.85
Education				
High School or less	1286	17.13	25.04	33.36
Some College	1934	10.55	14.89	19.75
College Degree	1857	4.42	8.45	12.12
Center				
1	1171	8.63	20.41	24.25
2	1104	9.96	16.67	22.10
3	1381	8.40	12.89	18.83
4	1421	13.16	11.68	17.45
Baseline Smoker				
No	3531	8.58	13.17	18.10
Yes	1546	13.65	19.53	25.68

As mentioned earlier the CARDIA not only exhibits covariate dependent missing-ness (MAR-X) but also response dependent missing-ness (MAR-Y) characteristics. Therefore the WGEE approach is not suitable. In this section we propose following marginal logistic regression model for analysis of subset data arising from 3692 participants with no missing data.

$$\begin{aligned}\log\left(\frac{\mu_{ij}}{1-\mu_{ij}}\right) &= \beta_0 + (\text{Age})_i\,\beta_1 + I(\text{Race=White})_i\,\beta_2 + I(\text{Sex=Male})_i\,\beta_3 \\ &\quad + I(\text{Ed. = Some college})_i\,\beta_4 + I(\text{Ed. = College degree})_i\,\beta_5 \\ &\quad + (\text{Followup year})_j\,\beta_6 + I(\text{Center=2})_i\,\beta_7 + I(\text{Center=3})_i\,\beta_8 \\ &\quad + I(\text{Center=4})_i\,\beta_9\end{aligned}$$

where I is the indicator function. The modified Gaussian and the quasi-least squares, as well as the GEE estimates are displayed in Table 3.8 under the assumption of exchangeable correlation structure. Since the data set is large both MG and QLS

estimates are in agreement. The statistically significant positive estimate β_2 (0.0238) implies that likelihood of smoking increases with age. Table 3.8 also indicates that

Table 3.8: CARDIA Data: Parameter Estimates

	GEE			MG			QLS		
	Est.	S.E.	p-value	Est.	S.E.	p-value	Est.	S.E.	p-value
β	-0.8121	0.2638	0.0021	-0.8099	0.2638	0.0021	-0.8098	0.2638	0.0021
	0.0239	0.0100	0.0170	0.0238	0.0100	0.0174	0.0238	0.0100	0.0174
	-0.1344	0.0772	0.0817	-0.1348	0.0772	0.0807	-0.1348	0.0772	0.0807
	0.0709	0.0713	0.3201	0.0703	0.0713	0.3243	0.0703	0.0713	0.3242
	-0.6249	0.0855	0.0000	-0.6241	0.0855	0.0000	-0.6241	0.0855	0.0000
	-1.8747	0.1012	0.0000	-1.8745	0.1012	0.0000	-1.8746	0.1012	0.0000
	-0.0126	0.0044	0.0040	-0.0126	0.0044	0.0040	-0.0126	0.0044	0.0040
	0.1777	0.1106	0.1083	0.1777	0.1106	0.1083	0.1776	0.1106	0.1084
	0.5618	0.1024	0.0000	0.5619	0.1024	0.0000	0.5619	0.1024	0.0000
	-0.1408	0.1076	0.1907	-0.1393	0.1076	0.1956	-0.1393	0.1076	0.1956
α		-	-	0.7238	0.0113	0.0000	0.7230	0.0106	0.0000

the white participants are less likely to smoke than black participants. The p-value corresponding to the regression parameter β_3 is 0.324, and therefore it is inferred that the smoking is equally prevalent in female and male participants. The likelihood of participants with high-school or less education are more likely to smoke than participants with some college education or college graduates. Overall the probability of smoking decreased in the follow-up period. Also the participants recruited at study center 3 have statistically significant higher probability of smoking than participants recruited at study centers 1, 2 or 4. This could be explained by the demographic heterogeneity of the center locations.

Next we drew a stratified sample of 105 participants from the complete CARDIA study data using race, sex, education level and study centers as stratification variables. Table 3.9 gives the MG and QLS parameter estimates for the sample data for the following reduced model

$$\log\left(\frac{\mu_{ij}}{1-\mu_{ij}}\right) = \beta_0 + (\text{Age})_i\,\beta_1 + I(\text{Race=White})_i\,\beta_2 + I(\text{Ed. = Some college})_i\,\beta_3 + I(\text{Ed. = College degree})_i\,\beta_4 + (\text{Followup year})_j\,\beta_5. \quad (3.3.3)$$

The parameter estimates of reduced model computed using stratified sample have larger standard errors and are not comparable with the estimates of Table 3.8. In next section we present simulation results that compare the asymptotic performance of MG estimates and QLS estimates for the reduced model.

Table 3.9: CARDIA Sample: Parameter Estimates

	GEE			MGE			QLS		
	Est.	S.E.	p-value	Est.	S.E.	p-value	Est.	S.E.	p-value
β	-3.0195	1.5410	0.0501	-3.0259	1.5421	0.0497	-3.0271	1.5423	0.0497
	0.1183	0.0638	0.0635	0.1186	0.0638	0.0630	0.1187	0.0638	0.0629
	1.1869	0.5220	0.0230	1.1881	0.5220	0.0228	1.1883	0.5220	0.0228
	-1.3272	0.5597	0.0177	-1.3292	0.5602	0.0177	-1.3295	0.5603	0.0176
	-3.0755	0.7208	0.0001	-3.0780	0.7211	0.0000	-3.0785	0.7212	0.0000
	-0.0368	0.0230	0.1088	-0.0369	0.0230	0.1089	-0.0369	0.0230	0.1090
α	0.8346	-	-	0.8387	0.0425	0.0000	0.8394	0.0514	0.0000

III.3.2 Simulation Results

To check the performance of the MG and QLS methods for binary data, we used the GEE estimate $\beta_0 = (-3.0195, 0.1183, 1.1869, -1.3272, -3.0755, -0.0368)$ given in Table 3.9 as the true value for β, and repeated the eight simulation steps in Section III.2.2 using the model (3.3.3), and the covariates in the sample CARDIA data. The Poisson variables generated in step 4 were dichotomized using the indicator of zero, to obtain correlated binary variables.

When the true correlation is exchangeable, plots of estimated infeasibility/divergence probabilities for the MG and QLS methods are in Figure 3.10. The plots clearly show that both methods have the same pattern, the estimated probability is generally below 20% over a wide range in the interior of the feasible range, but shoots up at the boundary. Clearly, infeasibility and divergence will be problematic if the data has a small correlation or if the correlation is very high. Therefore, for weakly correlated binary data, we should use GEE with identity as the correlation matrix, and for strongly dependent data, we should use fully specified models that are described in the next chapter instead of MG or the QLS. Plots of the coverage probabilities of the 95% confidence ellipsoids are given in Figure 3.11. Clearly, the confidence ellipsoids constructed by the MG method are superior than those obtained using the QLS method. But note that the asymptotic relative efficiency of the QLS estimate of the regression parameter with respect to the MG estimate is more than one on a wide interval as shown in Figure 3.12. As in the Poisson case, we can see from the fifth column of Table 3.10 the MG estimate of α is more efficient than the QLS estimate. In general QLS estimates regressions coefficients for the time independent covariates more efficiently than the MG method. The opposite is true for

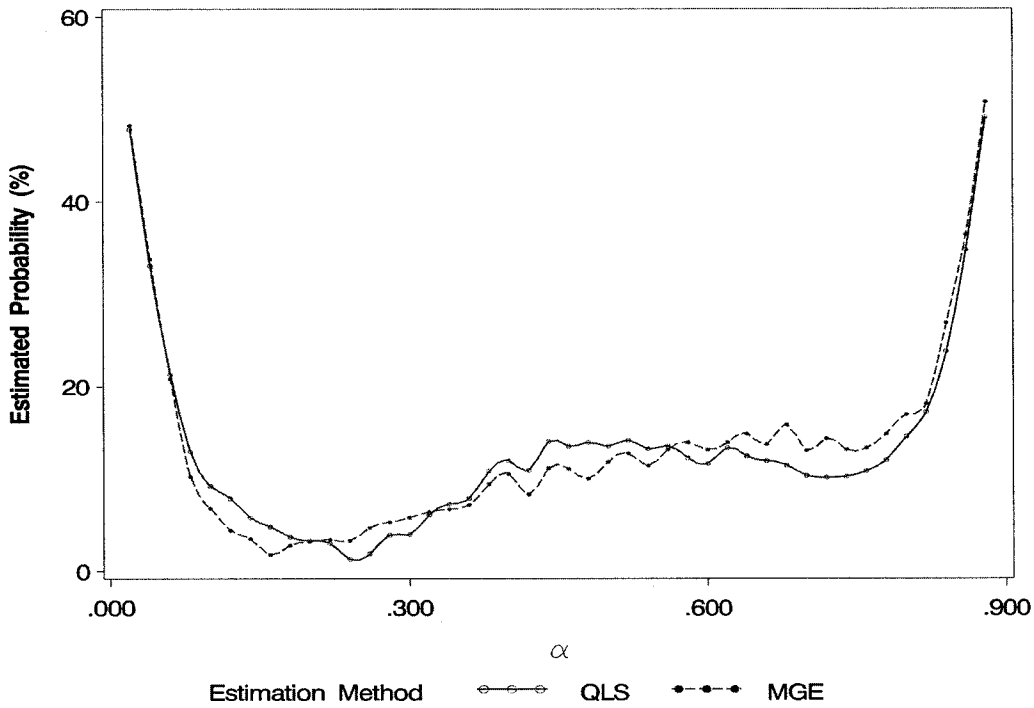

Figure 3.10: EXCH: Infeasibility/divergent solutions probabilities.

the time dependent covariates.

The analysis of the simulation results for the AR(1) structure regarding infeasibility, coverage probabilities, and asymptotic relative efficiencies are given in Figures 3.13, 3.14 and 3.15 respectively. Since QLS estimate of α is in a closed form for the AR(1) structure, the infeasibility probability is uniformly lower than for the MG estimate over the entire feasible region for α. The confidence ellipsoids of both the MG and QLS methods have a higher coverage probability for weakly correlated binary data, that is, for small values of α, and lower coverage probability for strongly correlated binary data, that is, for large values of α. Figure 3.15 shows that the asymptotic relative efficiency of QLS with respect to the MG method has a zig-zag behavior but it is more than one almost everywhere. Unlike the previous cases, Table 3.11 shows that QLS estimates of the regression coefficients for time independent and time dependent covariates more efficiently than the MG method. In summary, when the true correlation has AR(1) structure, QLS method is highly recommended over the MG method.

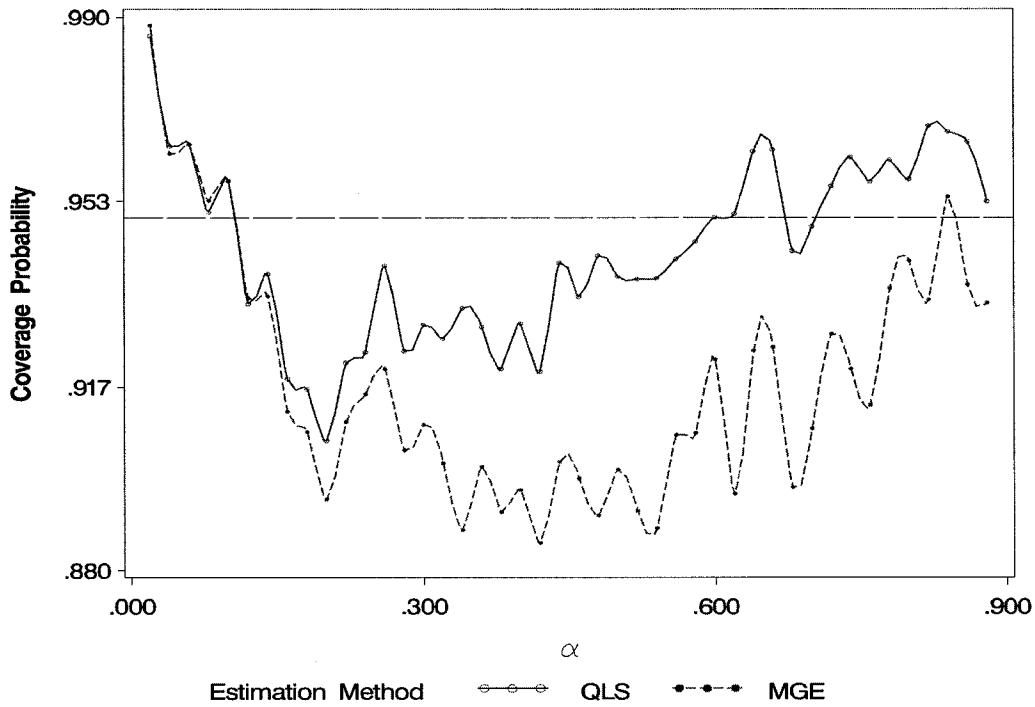

Figure 3.11: EXCH: Coverage probability of simultaneous confidence region.

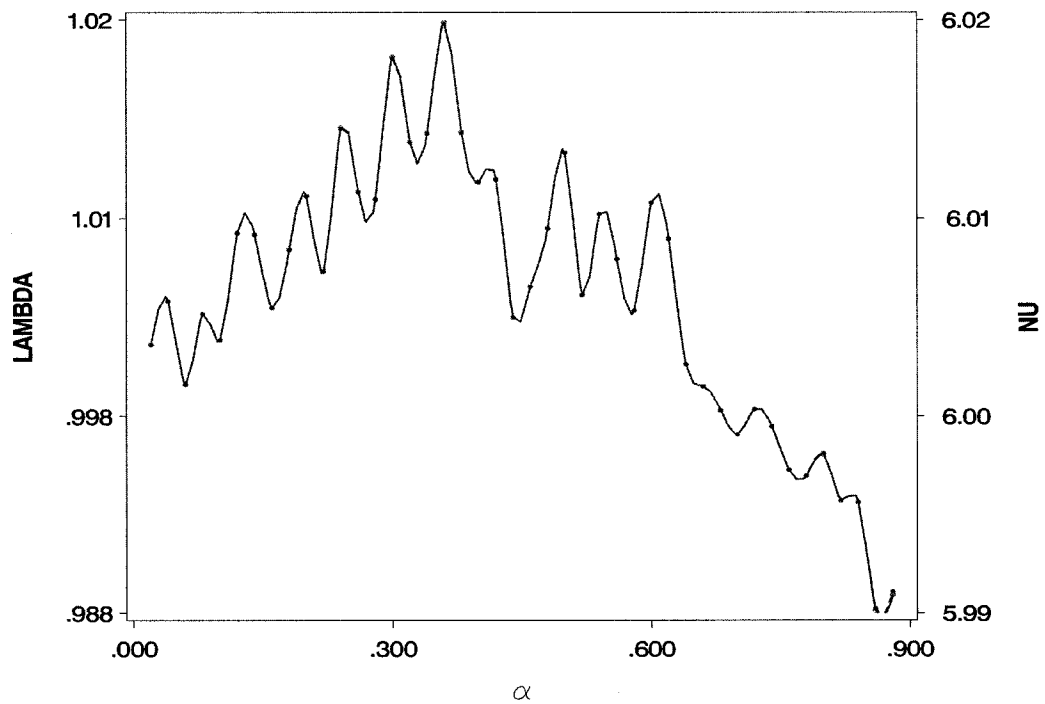

Figure 3.12: EXCH: ARE of $\hat{\beta}_q$ vs. $\hat{\beta}_m$.

Table 3.10: EXCH: ARE of QLS vs. MGE

α_0	N_c	$\widehat{\beta}_0$	$\widehat{\beta}_5$	$\widehat{\alpha}$	Λ_1	ν_1	Λ_2	ν_2
0.04	642	1.0009	.99962	.98250	1.0044	6.0044	.99887	7.0147
0.08	859	1.0008	.99962	.96466	1.0037	6.0037	.99107	7.0240
0.12	911	1.0017	.99913	.94719	1.0079	6.0079	.99342	7.0555
0.16	941	1.0009	.99947	.92594	1.0040	6.0040	.98309	7.0777
0.20	936	1.0021	.99863	.91165	1.0098	6.0098	.98928	7.1221
0.24	955	1.0031	.99777	.89625	1.0134	6.0133	.98806	7.1539
0.28	918	1.0023	.99808	.87871	1.0097	6.0096	.97900	7.1909
0.32	894	1.0030	.99724	.86779	1.0126	6.0126	.97903	7.2255
0.36	867	1.0045	.99543	.85766	1.0189	6.0188	.98122	7.2625
0.40	818	1.0029	.99671	.84612	1.0106	6.0105	.96616	7.2741
0.44	815	1.0010	.99852	.83770	1.0035	6.0035	.95318	7.2880
0.48	817	1.0025	.99637	.82850	1.0081	6.0081	.94838	7.3022
0.52	805	1.0015	.99747	.82600	1.0047	6.0047	.94272	7.3197
0.56	808	1.0025	.99518	.82376	1.0066	6.0066	.93792	7.3288
0.60	821	1.0034	.99304	.81984	1.0095	6.0095	.93219	7.3295
0.64	814	1.0013	.99598	.82400	1.0011	6.0011	.92389	7.3228
0.68	813	1.0007	.99688	.82714	.99864	5.9986	.91772	7.3068
0.72	823	1.0018	.99280	.82652	.99869	5.9987	.90356	7.2743
0.76	840	.99903	1.0013	.83194	.99553	5.9955	.89287	7.2283
0.80	803	1.0025	.98715	.82033	.99632	5.9964	.86910	7.1977
0.84	698	1.0011	.99013	.84963	.99386	5.9939	.86896	7.1414
0.88	449	1.0022	.97891	.84659	.98895	5.9891	.84177	7.0912

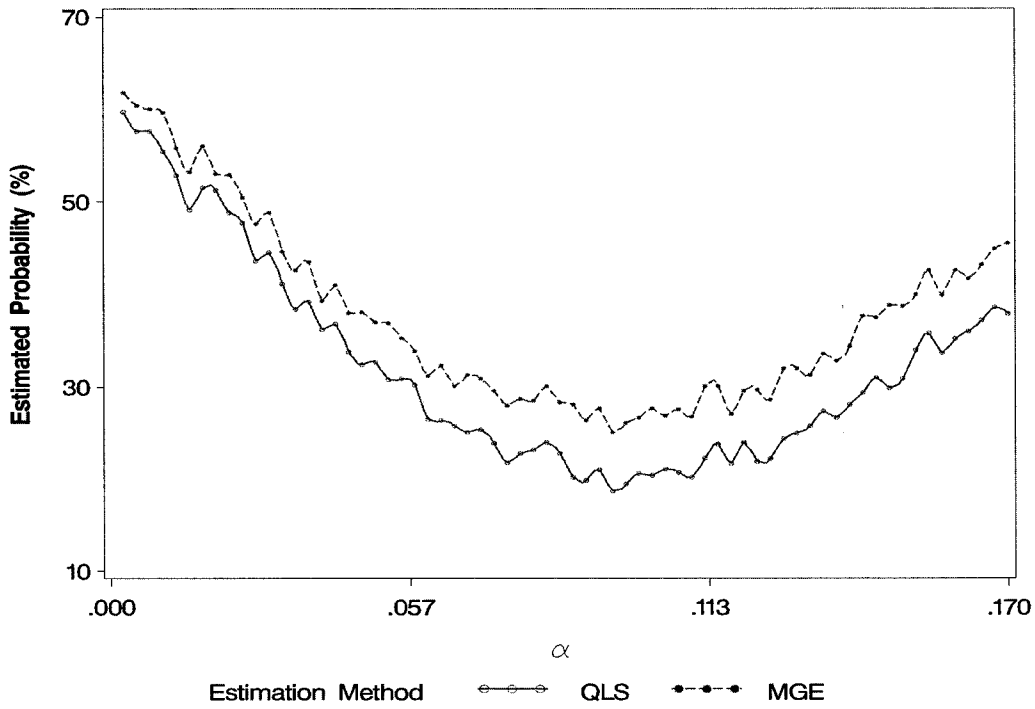

Figure 3.13: AR(1): Infeasibility/divergent solutions probabilities.

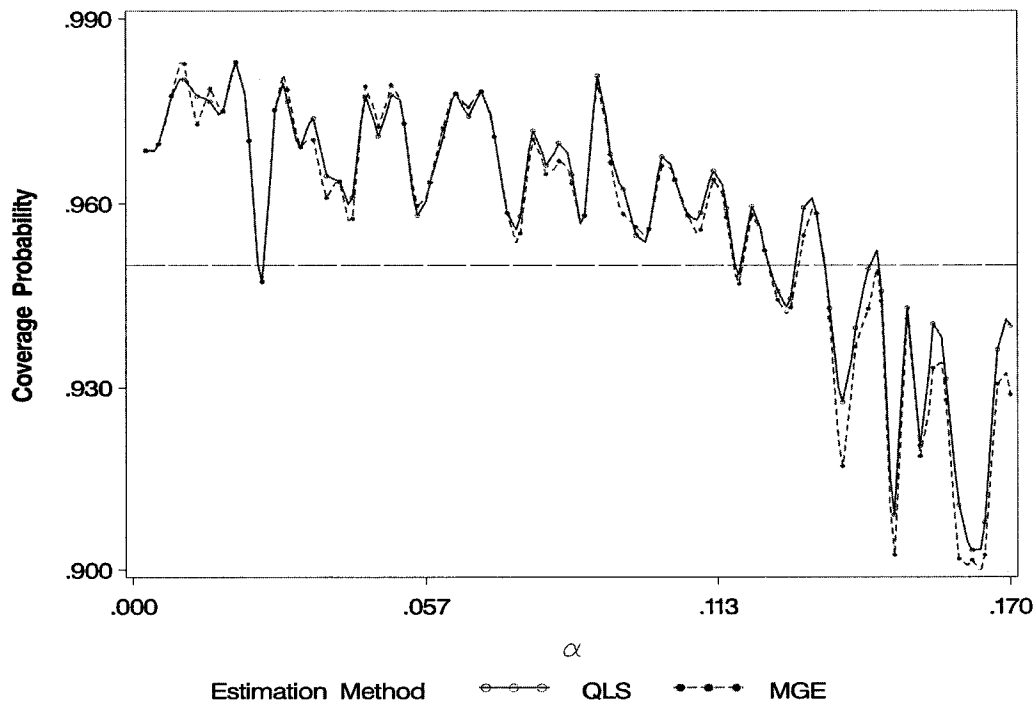

Figure 3.14: AR(1): Coverage probability of simultaneous confidence region.

Table 3.11: AR(1): ARE of QLS vs. MGE

α_0	N_c	$\widehat{\beta}_0$	$\widehat{\beta}_5$	$\widehat{\alpha}$	Λ_1	ν_1	Λ_2	ν_2
0.01	403	1.0006	1.0001	.99073	1.0029	6.0029	.99940	7.0067
0.02	469	1.0005	1.0001	.99098	1.0027	6.0027	.99918	7.0062
0.03	511	.99973	.99990	.98790	.99841	5.9984	.99292	7.0014
0.04	604	1.0004	1.0001	.98918	1.0021	6.0021	.99804	7.0068
0.05	625	1.0002	1.0000	.98746	1.0010	6.0010	.99601	7.0058
0.06	684	1.0005	1.0001	.98627	1.0028	6.0028	.99801	7.0099
0.07	685	1.0004	1.0000	.98504	1.0019	6.0019	.99665	7.0095
0.08	710	1.0004	1.0000	.98454	1.0019	6.0019	.99639	7.0097
0.09	726	1.0004	1.0001	.98279	1.0019	6.0019	.99603	7.0113
0.10	726	1.0002	1.0000	.98098	1.0010	6.0010	.99437	7.0113
0.11	723	1.0008	1.0001	.98116	1.0040	6.0040	.99819	7.0162
0.12	692	1.0008	1.0001	.97908	1.0041	6.0041	.99814	7.0190
0.13	664	1.0006	1.0001	.97762	1.0031	6.0031	.99652	7.0187
0.14	647	1.0006	1.0000	.97509	1.0033	6.0033	.99557	7.0201
0.15	597	1.0008	1.0001	.97434	1.0040	6.0040	.99658	7.0228
0.16	560	1.0007	1.0001	.97371	1.0034	6.0034	.99568	7.0224
0.17	534	1.0009	1.0001	.97263	1.0045	6.0045	.99678	7.0253

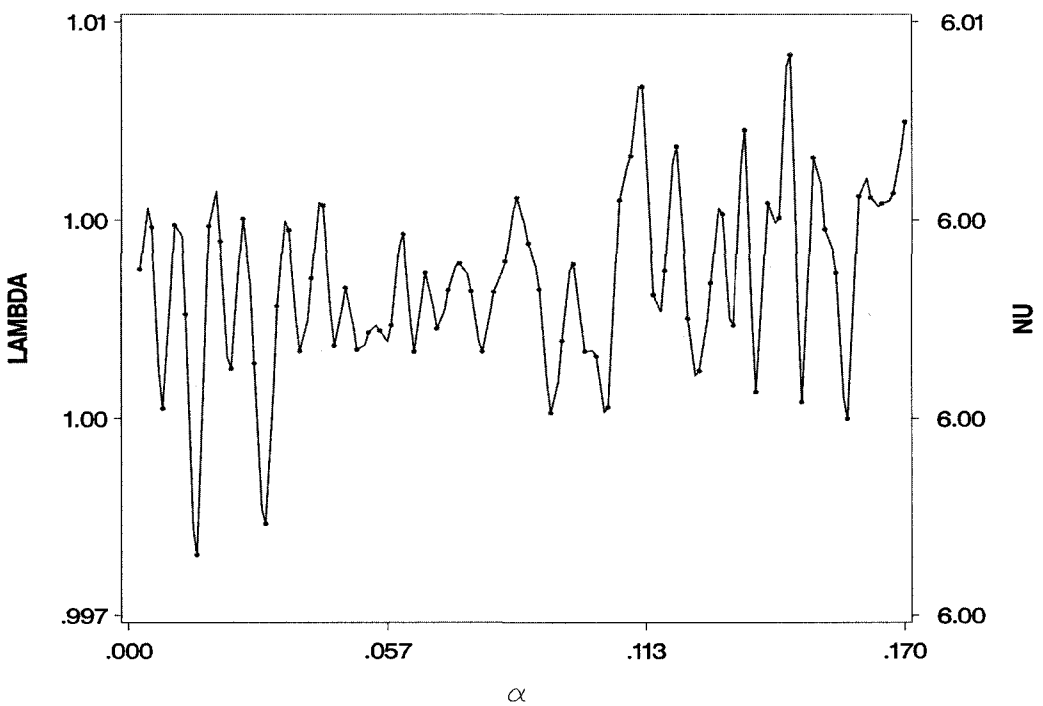

Figure 3.15: AR(1): ARE of $\hat{\beta}_q$ vs. $\hat{\beta}_m$.

CHAPTER IV

FULLY SPECIFIED MODELS FOR DISCRETE DATA

In previous chapters we have studied multivariate models which are partially specified in the sense that the joint distributions are unknown but the functional forms of the marginals were given. Furthermore, we have assumed in the multivariate models the covariance matrix $\Sigma_i(\beta, \alpha, \phi)$ can be decomposed as $\phi A_i^{\frac{1}{2}}(\beta)\, R_i(\alpha) A_i^{\frac{1}{2}}(\beta)$ (see Section 2.3). However, this assumption of decomposability of the covariance matrix has been the subject of debate especially when the response variables are non-Gaussian. For example, when the responses are binary or Poisson the correlations are functions of the marginal means and hence cannot be independent of the covariates. Ignoring this dependence of the correlations on the covariates may lead to erroneous estimates and misleading conclusions. In this chapter we will study fully specified multivariate models which can be used for the analysis of discrete multivariate and discrete longitudinal data.

The organization of this chapter is as follows. In Section IV.1 we provide background on the copulas and present several examples of copulas. Section IV.2 we describe probability models based on normal copula, including multivariate probit models. We also discuss several methods for computing multivariate normal probabilities. Maximum likelihood and theory of estimating equations methods were discussed in detail for the multivariate probit model. A real life data example is presented to demonstrate the practical use of these methods.

In Section IV.3 we discuss multivariate mixture models. Important special cases are the multivariate probit-normal model and the multivariate Poisson log-normal model. These two models are discussed in depth, in particular we present algebraic details for implementation of the maximum likelihood for the multivariate probit-normal model. We demonstrate the use of these models on the epileptic seizure data given in Thall and Vail (1990). Finally, we present multivariate discrete choice models in Section IV.4. These models include the multinomial logit and probit models have been widely used in the econometric literature.

IV.1 Background on Copulas and Fréchet Bounds

The multivariate normal distribution is widely used to model continuous data because of its simplicity and the fact that it is completely specified by the first two moments. Even though multivariate discrete distributions are not as simply characterized as the multivariate Gaussian distribution, they do admit a simple decomposition however. Basically, a multivariate discrete joint distribution F can be specified by the marginals and a function C, known as a copula that characterizes the structure of the dependence. The copula C can be viewed as the joint distribution function of p random variables $U_1, \ldots, U_p$, where marginally U_i is distributed as uniform on the interval (0, 1) for $1 \leq i \leq p$.

Definition 4.1 *A function of p variables* $C : [0,1]^p \to [0,1]$ *is known as a copula if it satisfies the following properties:*

1. $C(1, 1, \ldots, 1, u_j, 1, \ldots, 1) = u_j$ *for all* $j = 1, 2, \ldots, p$
2. $C(u_1, \ldots, u_p) = 0$ *if* $\min(u_1, \ldots, u_p) = 0$
3. *For all,* $a_{i_1} < a_{i_2}$, $i = 1, 2 \ldots, p$,

$$\sum_{j_1=1}^{2}\sum_{j_2=1}^{2}\cdots\sum_{j_p=1}^{2}(-1)^{j_1+j_2+\ldots+j_p}C(a_{1_{j_1}}, a_{2_{j_2}}, \ldots, a_{p_{j_p}}) \geq 0.$$

The following fundamental theorem due to Sklar (1959) gives the relation between the multivariate distribution function F and the copula C.

Theorem 4.1 *Let* $Y_1, Y_2, \ldots, Y_p$ *be random variables with marginal distribution functions* $F_1, F_2, \ldots, F_p$ *and joint distribution function* F. *Then the following hold.*

1. *There exist a p-dimensional copula* C *such that for all* $y_1, y_2, \ldots y_p \in \Re$,

$$F(y_1, y_2, \ldots, y_p) = C(F_1(y_1), F_2(y_2), \ldots, F_p(y_p))$$

2. *If* $Y_1, Y_2, \ldots, Y_p$ *are continuous random variables defined on real line, then* C *is unique. Otherwise,* C *is uniquely determined on the p dimensional rectangle* $Range(F_1) \times Range(F_2) \times \ldots \times Range(F_p)$.

IV.1.1 Examples of Copulas

We now present some popular and well known copulas.

Example 4.1 *(Independence Copula) The p-dimensional function given by*

$$C(u_1, u_2, \ldots, u_p) = \prod_{j=1}^{p} u_j \tag{4.1.1}$$

is known as the Independence copula.

Example 4.2 *(Morgenstern Copula) A simple version of the Morgenstern copula is given by*

$$C(u_1, u_2, \ldots, u_p) = \left[1 + \sum_{j=1}^{p} \sum_{k=j+1}^{p} \theta_{jk}(1-u_j)(1-u_k)\right] \prod_{j=1}^{p} u_j \tag{4.1.2}$$

where the parameters θ_{jk}'s are chosen such that (4.1.2) is indeed a copula. Note that θ_{jk} determines the dependence structure between U_j and U_k. The density function, which is the derivative of (4.1.2), is given by

$$c(u_1, u_2, \ldots, u_p) = 1 + \sum_{j=1}^{p} \sum_{k=j+1}^{p} \theta_{jk}(1-2u_j)(1-2u_k).$$

A general version of the p-dimensional Morgenstern copula which includes some higher order terms is defined as

$$C(u_1, u_2, \ldots, u_p) = \left[1 + \sum_{j_1=1}^{p} \sum_{j_2=j_1+1}^{p} \theta_{j_1 j_2} \prod_{k=1}^{2}(1-u_{j_k}) + \sum_{j_1=1}^{p} \sum_{j_2=j_1+1}^{p} \sum_{j_3=j_2+1}^{p} \theta_{j_1 j_2 j_3} \times \prod_{k=1}^{3}(1-u_{j_k}) + \ldots + \theta_{12\ldots p} \prod_{j=1}^{p}(1-u_j)\right] \prod_{j=1}^{p} u_j.$$

The above copula C has following density function

$$c(u_1, u_2, \ldots, u_p) = \left[1 + \sum_{j_1=1}^{p} \sum_{j_2=j_1+1}^{p} \theta_{j_1 j_2} \prod_{k=1}^{2}(1-2u_{j_k}) + \sum_{j_1=1}^{p} \sum_{j_2=j_1+1}^{p} \sum_{j_3=j_2+1}^{p} \theta_{j_1 j_2 j_3} \times \prod_{k=1}^{3}(1-2u_{j_k}) + \ldots + \theta_{12\ldots p} \prod_{j=1}^{p}(1-2u_j)\right].$$

The Morgenstern copula is also known as the Farlie-Gumbel-Morgenstern copula.

Example 4.3 *(Archimedean Copula) Let M be a univariate distribution function of a positive random variable. Note that $M(0) = 0$. Let*

$$\phi(u) = \int_0^\infty e^{-ut} dM(t), \qquad u \geq 0$$

be the Laplace transform of M. The p-dimensional Archimedean copula is defined as

$$C(u_1, u_2, \ldots, u_p) = \phi\left(\sum_{j=1}^{p} \phi^{-1}(u_j)\right). \tag{4.1.3}$$

This multivariate copula is permutation-symmetric in the p arguments so that it can be viewed as the distribution function of p exchangeable random variables, each distributed as uniform on the interval $(0,\ 1)$.

Example 4.4 *(Multivariate Normal Copula) Let $\Phi_p(.; \boldsymbol{\mu}, \Sigma)$ be a probability distribution function of p dimensional normal vector $\mathbf{Z}$ with mean $\boldsymbol{\mu}$ and variance covariance matrix Σ defined as*

$$\Phi_p(z_1, \ldots, z_p; \boldsymbol{\mu}, \Sigma) = \int_{-\infty}^{z_p} \cdots \int_{-\infty}^{z_1} \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{t}-\boldsymbol{\mu})' \Sigma^{-1} (\mathbf{t}-\boldsymbol{\mu})} dt_1 \ldots dt_p \tag{4.1.4}$$

The p-dimensional normal copula with latent correlation matrix $\mathbf{R}$ is given by

$$C(u_1, u_2, \ldots, u_p; \mathbf{R}) = \Phi_p(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_p); \mathbf{0}, \mathbf{R}), \tag{4.1.5}$$

where $\Phi^{-1}(.)$ is inverse function of univariate standard normal distribution $\Phi_1(.; 0, 1)$

(o $C(u_1, u_2, \ldots, u_p; \mathbf{R}) = \Phi_p(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_p); \mathbf{0}, \mathbf{R}),$ (4.1.5)

wl where $\Phi^{-1}(.)$ is inverse function of univariate standard normal distribution $\Phi_1(.; 0, 1)$
ate normal copula for modelling discrete outcomes.

It is easy to establish bounds for a copula. These bounds, known as Fréchet bounds, are given below.

Lemma 4.1 *(Fréchet bounds) Let C be a p-dimensional copula. Then for all $u_i \in [0, 1]$ $i = 1, 2, \ldots, p$,*

$$C_L(u_1, u_2, \ldots, u_p) \leq \ C(u_1, u_2, \ldots, u_p) \ \leq C_U(u_1, u_2, \ldots, u_p)$$

where Fréchet lower bound C_L and upper bound C_U are defined as

$$\begin{aligned} C_L(u_1, u_2, \ldots, u_p) &= \max(0, u_1 + u_2 + \ldots + u_p - (p-1)), \\ C_U(u_1, u_2, \ldots, u_p) &= \min(u_1, u_2, \ldots, u_p). \end{aligned}$$

The Fréchet upper bound C_U is a copula, whereas the Fréchet lower bound C_L is a copula only when $p = 2$.

As a special case, let us consider bivariate Poisson random variables. We are interested in studying the range of the correlation as a function of the marginal means, as determined by C_U and C_L, the Fréchet upper and lower bound copulas. Let Y_1 and Y_2 be two Poisson random variables with means λ_1 and λ_2. Suppose that $F_i(y_i)$ denotes the marginal cumulative distribution function of Y_i and the joint cumulative distribution function of (Y_1, Y_2) is denoted by $F(y_1, y_2)$. Then

$$F_L(y_1, y_2) \leq F(y_1, y_2) \leq F_U(y_1, y_2) \tag{4.1.6}$$

where $F_L(y_1, y_2) = C_L(F_1(y_1), F_2(y_2)) = \max(0, F_1(y_1)+F_2(y_2)-1)$ and $F_U(y_1, y_2) = C_U(F_1(y_1), F_2(y_2)) = \min(F_1(y_1), F_2(y_2))$.

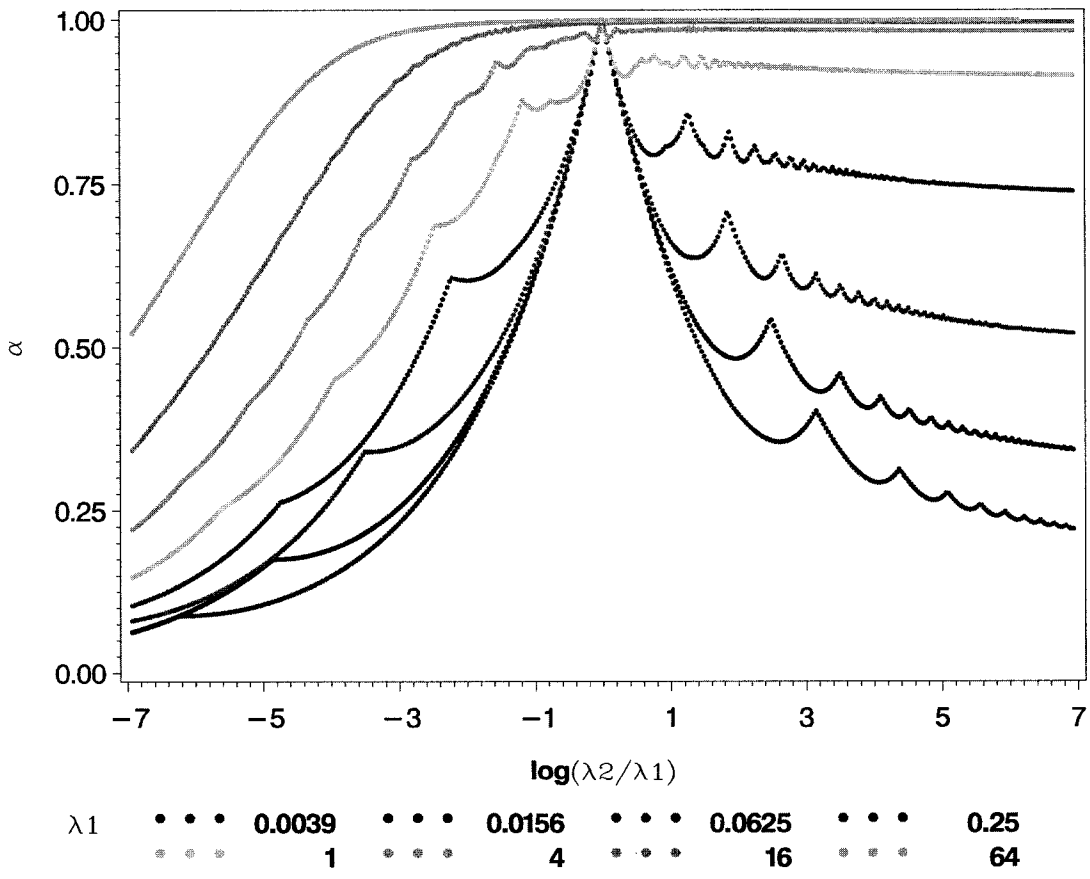

Figure 4.1: Bivariate Poisson: Fréchet upper bounds.

The correlation coefficient corresponding to the joint distribution $F_U(y_1, y_2)$ is given by

$$\alpha_U(\lambda_1, \lambda_2) = \frac{1}{\sqrt{\lambda_1 \lambda_2}} \sum_{y_1=1}^{\infty} \sum_{y_2=1}^{\infty} \min\left[F_1(y_1)(1 - F_2(y_2)), F_2(y_2)(1 - F_1(y_1))\right], \quad (4.1.7)$$

and the correlation coefficient corresponding to the joint distribution $F_L(y_1, y_2)$ is

$$\alpha_L(\lambda_1, \lambda_2) = \frac{1}{\sqrt{\lambda_1 \lambda_2}} \sum_{y_1=1}^{\infty} \sum_{y_2=1}^{\infty} - \min\left[(1 - F_1(y_1))(1 - F_2(y_2)),\ F_1(y_1)F_2(y_2)\right]. \quad (4.1.8)$$

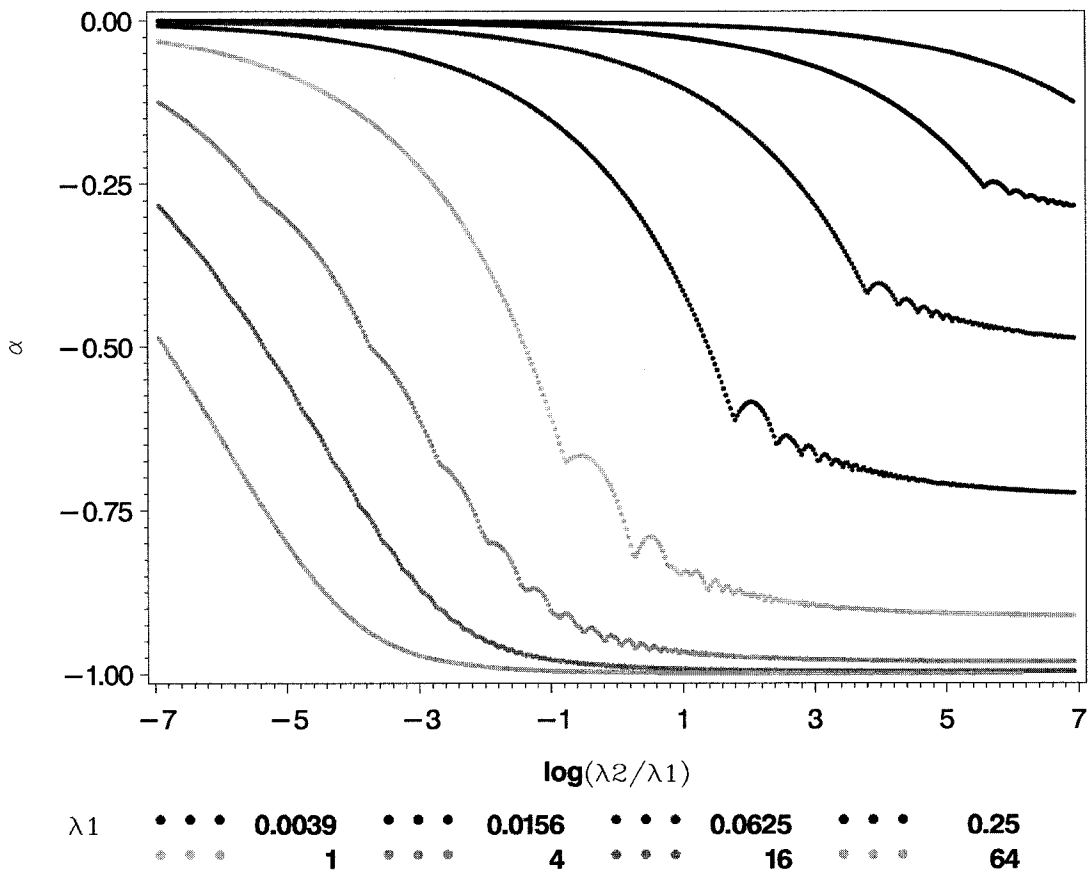

Figure 4.2: Bivariate Poisson: Fréchet lower bounds.

Thus the range of the correlation between Y_1 and Y_2 is given by the interval $[\alpha_L(\lambda_1, \lambda_2),\ \alpha_U(\lambda_1, \lambda_2)]$. Figures (4.1) and (4.2) demonstrate the complex behavior of $\alpha_U(\lambda_1, \lambda_2)$ and $\alpha_L(\lambda_1, \lambda_2)$ as a function of $(\lambda_1,\ \lambda_2)$. This complex behavior poses difficulty in incorporating the bounds on the correlation in moments based estimation methods that we studied in Chapters II and III. The next lemma gives some

properties of the functions $\alpha_U(\lambda_1, \lambda_2)$ and $\alpha_L(\lambda_1, \lambda_2)$.

Lemma 4.2 *Let Y_1 and Y_2 be two Poisson random variables with means λ_1 and λ_2 respectively. Assume without loss of generality $\lambda_1 \leq \lambda_2$. Let $\alpha_L(\lambda_1, \lambda_2)$ and $\alpha_U(\lambda_1, \lambda_2)$ be the correlations corresponding to the distribution functions $F_L(y_1, y_2)$ and $F_U(y_1, y_2)$ respectively. Then we have*

(a) $\alpha_U(\lambda_1, \lambda_2) \leq \dfrac{1}{\sqrt{\lambda_1 \lambda_2}} \left(\dfrac{(\lambda_2 - \lambda_1)^2}{2} + \lambda_1 \right)$

(b) $\alpha_L(\lambda_1, \lambda_2) = -\sqrt{\lambda_1 \lambda_2}$, *if* $\lambda_2 \leq \log(2)$.

Proof: If (Y_1, Y_2) is distributed as $F_U(y_1, y_2)$ then

$$E(Y_1 Y_2) = \sum_{y_1=1}^{\infty} \sum_{y_2=1}^{\infty} \min[\, \overline{F_1}(y_1), \overline{F_2}(y_2) \,]$$

where $\overline{F_i}(y_i) = P(Y_i \geq y_i)$ for $i = 1, 2$. Since $\lambda_2 \geq \lambda_1$, and the Poisson family possesses TP_2 property we have $\overline{F_2}(y_2) \geq \overline{F_1}(y_2) \geq \overline{F_1}(y_1)$ for $y_1 \geq y_2$. Now

$$\begin{aligned} E(Y_1 Y_2) &= \sum_{y_1=1}^{\infty} \sum_{y_2=1}^{y_1} \min[\, \overline{F_1}(y_1), \overline{F_2}(y_2) \,] + \sum_{y_1=1}^{\infty} \sum_{y_2=y_1+1}^{\infty} \min[\, \overline{F_1}(y_1), \overline{F_2}(y_2) \,] \\ &= \sum_{y_1=1}^{\infty} y_1 \overline{F_1}(y_1) + \sum_{y_2=2}^{\infty} \sum_{y_1=1}^{y_2-1} \min[\, \overline{F_1}(y_1), \overline{F_2}(y_2) \,] \\ &\leq \sum_{y_1=1}^{\infty} y_1 \overline{F_1}(y_1) + \sum_{y_2=2}^{\infty} (y_2 - 1)\, \overline{F_2}(y_2). \end{aligned} \tag{4.1.9}$$

The first term on the right hand side of (4.1.9) is

$$\begin{aligned} \sum_{y_1=1}^{\infty} y_1\, \overline{F_1}(y_1) &= \sum_{y_1=1}^{\infty} y_1\, P(Y_1 \geq y_1) \\ &= \sum_{y_1=1}^{\infty} \sum_{y=y_1}^{\infty} y_1\, P(Y_1 = y) \\ &= \sum_{y=1}^{\infty} \sum_{y_1=1}^{y} y_1\, P(Y_1 = y) \\ &= \sum_{y=1}^{\infty} \frac{y(y+1)}{2}\, P(Y_1 = y) \\ &= \frac{(E(Y_1^2) + E(Y_1))}{2} \end{aligned}$$

$$= \frac{\lambda_1^2}{2} + \lambda_1. \tag{4.1.10}$$

Now the second term on the right hand side of (4.1.9) is

$$\begin{aligned} \sum_{y_2=2}^{\infty}(y_2-1)\,\overline{F_2}(y_2) &= \sum_{y_2=1}^{\infty} y_2\,\overline{F_2}(y_2) - \sum_{y_2=1}^{\infty} y_2\,P(Y_2=y_2) \\ &= \frac{\lambda_2^2}{2} + \lambda_2 - \lambda_2 = \frac{\lambda_2^2}{2}. \end{aligned} \tag{4.1.11}$$

Substituting (4.1.10) and (4.1.11) in (4.1.9) we get

$$E(Y_1\,Y_2) \leq \frac{\lambda_1^2+\lambda_2^2}{2} + \lambda_1. \tag{4.1.12}$$

It is easy to verify inequality (a) using (4.1.12). The proof of (b) is simple. When $0 \leq \lambda_1 \leq \lambda_2 \leq \log(2)$, we have $P(Y_i = 0) = e^{-\lambda_i} \geq 0.5$ for $i = 1, 2$. Hence $2\,F_i(y) \geq 1$ or equivalently $F_i(y) \geq (1 - F_i(y))$ for all $x \geq 1$ and $i = 1, 2$. Therefore

$$F_1(y_1)F_2(y_2) \geq (1 - F_1(y_1))(1 - F_2(y_2))$$

and (4.1.8) simplifies to

$$\alpha_L(\lambda_1,\lambda_2) = \frac{-1}{\sqrt{\lambda_1\lambda_2}} \sum_{y_1=1}^{\infty}\sum_{y_2=1}^{\infty}(1 - F_1(y_1))(1 - F_2(y_2)) = \frac{-\lambda_1\,\lambda_2}{\sqrt{\lambda_1,\lambda_2}} = -\sqrt{\lambda_1,\lambda_2}.$$

This completes the proof of the lemma.

There are situations where the range of the correlation between Y_1 and Y_2 could be very narrow. For example, when $\lambda_1 = \lambda_2 = \lambda$ converges to zero, the random variables Y_1 and Y_2 converge to zero in probability and thus the range of the correlation between Y_1 and Y_2 converges to the $\{0, 1\}$. Also for fixed λ_1 when λ_2 is too large or too small the range of the correlation becomes a singleton set containing zero. In next section we present the models which are primarily characterized by the copulas.

IV.2 Probability Models Based on Normal Copula

The multivariate copula models are preferable alternatives to moment based methods of Chapter II. These models provide flexibility in the choice of marginal distributions such as Poisson, generalized Poisson, negative binomial and many more. Further, the marginal distributions need not be members of an exponential dispersion family.

Also, the rich collection of copulas allows us to model a wide variety of dependence structures.

In a multivariate copula model the cumulative distribution function of a random vector $\mathbf{Y} = (Y_1, \ldots, Y_p)$ is defined as

$$F(\mathbf{y}) = C(F_1(y_1), \ldots, F_p(y_p)), \tag{4.2.1}$$

where C is a p-dimensional copula and F_i is cumulative distribution function of discrete random variable Y_i. Hence if $\mathbf{Y}$ is a discrete random vector then its probability mass function can be written as

$$P(\mathbf{y}) = \sum_{j_1=0}^{1} \sum_{j_2=1}^{1} \cdots \sum_{j_p=0}^{1} (-1)^{\left(\sum_{k=1}^{p} j_k\right)} C(a_1^{j_1}(y_1), a_2^{j_2}(y_2), \ldots, a_p^{j_p}(y_p)) \tag{4.2.2}$$

where $a_j^1(y_j) = F_j(y_j)$ and $a_j^0(y_j) = F_j(y_{j-})$. Whereas when $\mathbf{Y}$ is continuous its probability density function is

$$f(\mathbf{y}) = \prod_{i=1}^{p} f_i(y_i)\, c(F(y_1), F(y_2), \ldots, F(y_p)) \tag{4.2.3}$$

where $f_i(.)$ is marginal probability density function of random variable Y_i and

$$c(u_1, u_2, \ldots, u_p) = \frac{\partial^p C(u_1, u_2, \ldots, u_p)}{\partial u_1 \partial u_2 \ldots \partial u_p} \tag{4.2.4}$$

is density of the copula C.

In practice several copulas could be used to model the joint distribution of discrete responses, but the best choice is the copula that yields a wide range for the dependence measures. The multivariate normal copula given in (4.1.5) provides great flexibility in modelling various types of dependence. In this section we study the multivariate probability models where the joint distribution depends on the marginals through the multivariate normal copula. When C is the multivariate normal copula, equations (4.2.2) and (4.2.3) simplify to

$$P(\mathbf{y}) = \sum_{j_1=0}^{1} \sum_{j_2=0}^{1} \cdots \sum_{j_p=0}^{1} (-1)^{\left(\sum_{k=1}^{p} j_k\right)} \Phi_p \left[d_1^{j_1}(y_1), d_2^{j_2}(y_2), \ldots, d_p^{j_p}(y_p); \mathbf{0}, \mathbf{R}\right] \tag{4.2.5}$$

and

$$f(\mathbf{y}) = \left[\prod_{i=1}^{p} f_i(y_i)\right] \frac{1}{|\Sigma|^{-\frac{1}{2}}} \exp\left\{\frac{1}{2}\mathbf{d}(\mathbf{y})'(\mathbf{I}_p - \mathbf{R}^{-1})\mathbf{d}(\mathbf{y})\right\} \tag{4.2.6}$$

respectively. Here $d_i^0(y_i) = \Phi^{-1}(F_i(y_i-)), d_i(y_i) = d_i^1(y_i) = \Phi^{-1}(F_i(y_i)); i = 1, 2, \ldots, p$ and $\mathbf{d}(\mathbf{y}) = (d_1(y_1), \ldots, d_p(y_p))'$.

For arbitrary marginal distributions $F_j(\cdot)$ and $F_k(\cdot)$, Joe (1997) has shown that the pairwise correlation coefficients α_{jk} corresponding to the mass function (4.2.5) approaches the Fréchet upper bound as the latent correlation $\rho_{jk} \to 1$. And α_{jk} equals the Fréchet lower bound when $\rho_{jk} = -1$. This result motivated Emrich and Piedmonte (1991) to advocate using multivariate normal copula as a tool to generate correlated binary random variables with widest possible range of associations. However, their approach requires simultaneous solutions of complex non-linear equations and a detailed understanding of the relationship between the correlation α and the latent correlation ρ is necessary. When $p = 2$ and the marginal distributions of Y_1 and Y_2 are Poisson with means λ_1 and λ_2. The probability mass function (4.2.5) reduces to

$$\begin{aligned} P(\mathbf{y}) &= \{\Phi_2(d(y_1), d_2(y_2), \rho) + \Phi_2(d_1(y_1 - 1), d_2(y_2 - 1), \rho) \\ &\quad -\Phi_2(d(y_1 - 1), d_2(y_2), \rho) - \Phi_2(d_1(y_1), d_2(y_2 - 1), \rho)\} \quad \text{if } y_1, y_2 = 0, 1, \ldots \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

Figure 4.3 shows the relation between α and ρ as a function of λ_2 when λ_1 is fixed at 5. It is clear from the figure, when λ_2 is large the correlation α is approximately equal to the latent correlation ρ. On the other hand, there is a weak dependence between α and ρ when λ_2 is close to 0. As expected when the latent correlation ρ is zero then $\alpha = 0$.

For discrete outcomes, flexibility in dependence modelling comes at cost of estimating computationally challenging p dimensional integrations. It is well known that when the mean parameters are large the Poisson random variables exhibit the behavior common to continuous variables. Therefore we propose replacing the probability mass function by the bivariate normal copula density:

$$f^*(\mathbf{y}) = \left\{\prod_{j=1}^{2} \frac{e^{-\lambda_j}\lambda_j^{y_j}}{y_j!}\right\}\left\{\frac{1}{\sqrt{1-\rho^2}} \exp\left[\frac{\rho d_1(y_1) d(y_2)}{1+\rho} - \frac{\rho^2(d_1(y_1) - d_2(y_2))^2}{2(1-\rho^2)}\right]\right\} \tag{4.2.7}$$

Note that (4.2.7) is not a proper density function, since the summation over all possible values of $\mathbf{y}$ does not equal 1. Figure 4.4 contains the plot of the bias

$$b(\lambda_1, \lambda_2, \rho) = \mathrm{E}[f^*(\mathbf{Y}) - P(\mathbf{Y})], \tag{4.2.8}$$

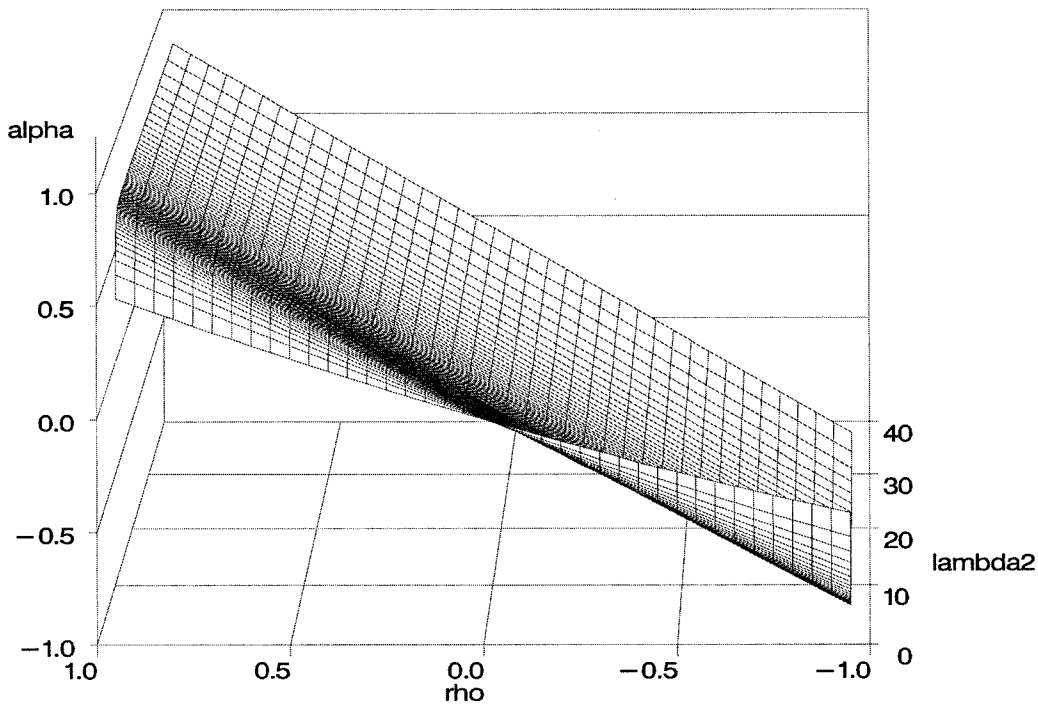

Figure 4.3: Bivariate Poisson: Latent correlation ρ vs. α for normal copulas.

of this continuity approximation.

Both Figures 4.3 and 4.4 are generated fixing the value of $\lambda_1 = 5$ and varying $\lambda_2 = \left(\frac{p}{1-p}\right)\lambda_1$; $p = 0.05, 0.1, \ldots, 0.95$; and $\rho = -0.95, 0.90, \ldots, 0.95$.

Clearly the bias is zero for large value of λ_2 (greater than 10). But for small values of λ_2 the approximation results in a large negative bias. Therefore the continuity approximation is not recommended for small marginal means. We describe in the next section methods for accurately estimating the joint distribution function by computing numerical integration via simulations.

IV.2.1 Multivariate Normal Interval Probabilities

The problem of evaluating multivariate normal probabilities and their derivatives is important because such expressions appear in leading econometric and clinical models. Fitting these models for real data requires evaluation of such probabilities and their derivatives for each trial parameter vector and each observation in a sample.

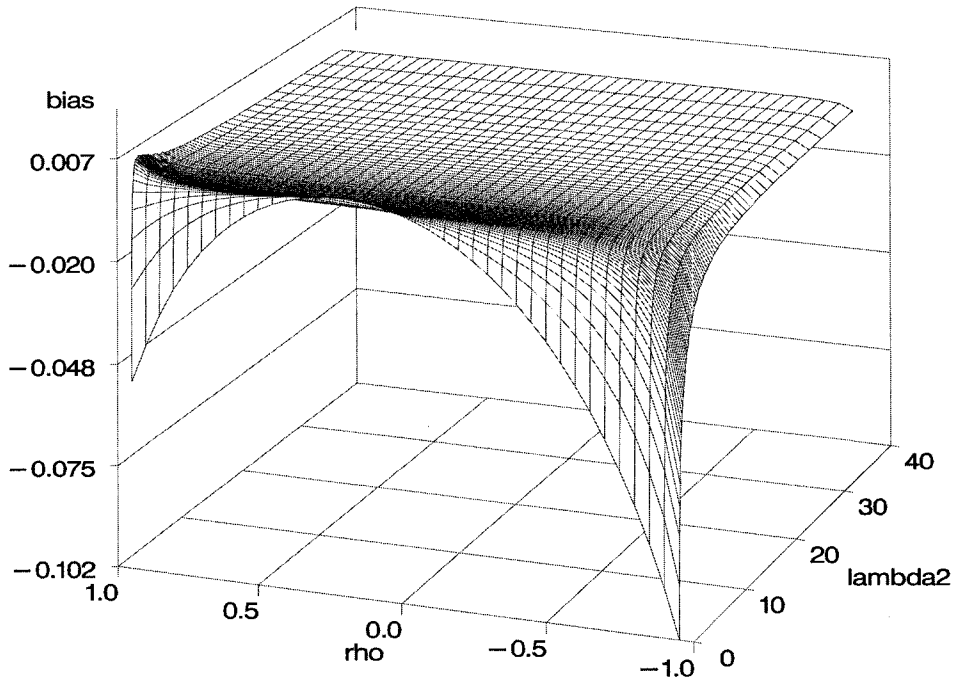

Figure 4.4: Bivariate Poisson: Bias of continuity approximation for normal copula.

There is a plethora of computational techniques in numerical analysis and in the econometrics literature for evaluating the multiple integral

$$\begin{aligned} P(\mathbf{a} \leq \mathbf{Y} \leq \mathbf{b}) &= \int_{\mathbf{a}}^{\mathbf{b}} \phi^p(\mathbf{y} - \mu, \Sigma) d\mathbf{y} \\ &= E\left[\prod_{i=1}^{p} I(a_i \leq Y_i \leq b_i)\right], \end{aligned} \tag{4.2.9}$$

where $\mathbf{Y}$ is a p-dimensional normal random vector with mean μ and covariance matrix Σ. This problem is computationally difficult unless $p < 4$ or the covariance matrix Σ has a special structure. In this section we describe the most frequently used methods for computing (4.2.9), see Hajivassiliou et al. (1996) for extensive details.

Crude frequency (Monte Carlo) simulator(CFS) : This is perhaps the most intuitive simulation method. We use the stochastic representation

$$\mathbf{Y} = \mu + C\mathbf{Z}$$

where $\mathbf{Z}$ is an independent standard normal vector and C is a lower triangular matrix derived from the Cholesky decomposition of $\Sigma = CC^T$. The method simulates n

independent and identically distributed $\mathbf{Y}_i, 1 \leq i \leq n$, vectors and approximates the multivariate interval probability (4.2.9) by

$$\hat{P} = \frac{1}{n}\sum_{i=1}^{n} I(\mathbf{a} \leq \mathbf{Y}_i \leq \mathbf{b}). \tag{4.2.10}$$

Kernel-smoothed frequency simulator(KFS): This method was developed by McFadden (1989) and it is a modification of the crude frequency simulator. Here the indicator function in (4.2.10) is replaced with

$$f(\mathbf{Y}_i) = g\left(\frac{(\mathbf{Y}_i - \mathbf{b})}{w}\right) - g\left(\frac{(\mathbf{Y}_i - \mathbf{a})}{w}\right) \tag{4.2.11}$$

where $g(.)$ is smooth kernel function such that

$$\begin{aligned} \lim_{x \to +\infty} g(\mathbf{x}) &= 0 \\ \lim_{x \to -\infty} g(\mathbf{x}) &= 1 \end{aligned}$$

and the constant w is known as the window width parameter. This method produces biased estimate of (4.2.9) for positive w, but in statistical applications the bias can be reduced by shrinking w as the sample size increases. A variety of kernel functions g were suggested in literature such as polynomial kernel: $g(\mathbf{x}) = \prod_{i=1}^{p} h(x_i)$ where

$$h(x) = \begin{cases} 1, & \text{for } x \leq -1; \\ (1 - x(2+x))/2, & \text{for } -1 < x \leq 0; \\ (1 - x(2-x))/2, & \text{for } 0 < x < 1; \\ 0, & \text{for } x \geq 1. \end{cases}$$

and normal kernel $g(\mathbf{x}) = \prod_{i=1}^{p} \Phi(-x_i)$.

Geweke-Hajivassiliou-Keane (GHK) simulator: This method uses the fact that the multivariate normal distribution can be expressed as the product of sequentially conditioned univariate normal distributions, that is, equation (4.2.9) can be rewritten as

$$P(\mathbf{a} \leq \mathbf{Y} \leq \mathbf{b}) = \prod_{i=1}^{p}\left[\Phi^{(i)}\left(\frac{(b_i - \mu_i) - \sum_{j=1}^{i-1} c_{ij}e_j}{c_{ii}}\right) - \Phi^{(i)}\left(\frac{(a_i - \mu_i) - \sum_{j=1}^{i-1} c_{ij}e_j}{c_{ii}}\right)\right] \tag{4.2.12}$$

where e_1 is standard normal variate with distribution function $\Phi^{(1)}$ and e_{i+1} is the truncated univariate standard normal variable with lower truncation point $\left(\frac{(a_i-\mu_i)-\sum_{j=1}^{i-1}c_{ij}e_j}{c_{ii}}\right)$ and upper truncation point $\left(\frac{(b_i-\mu_i)-\sum_{j=1}^{i-1}c_{ij}e_j}{c_{ii}}\right)$. The distribution function of e_{i+1} is denoted by $\Phi^{(i+1)}$.

Basically, the GHK simulator amounts to estimating each term in the product on the right hand side of (4.2.12) by the crude frequency simulator. Hajivassiliou et al. (1996) compared CFS, KFS, and GHK methods with several other methods, using several generator seed values. They found that GHK method has less bias and best convergence rate among all the methods. In the literature the GHK method is treated as the benchmark.

Genz's simulator: Genz (1992) proposed a two-stage transformation of (4.2.9). The first stage transformation is based on the conditional probabilities as in the GHK simulator. In the second stage the uniform random variables $u_i = \Phi(e_i), i = 2\ldots, p$ are used to reduce equation (4.2.12) to a $p-1$ dimensional integration as

$$P[\mathbf{a} \leq \mathbf{Y} \leq \mathbf{b}] = \int_{\mathbf{0}}^{\mathbf{1}} \prod_{i=1}^{p} [h_i(\mathbf{u}) - l_i(\mathbf{u})] d\mathbf{u} \tag{4.2.13}$$

where

$$\begin{aligned} h_i(\mathbf{u}) &= \Phi\left(\frac{(b_i-\mu_i)-\sum_{j=1}^{i-1}c_{ij}\Phi^{-1}(u_j h_j(\mathbf{u})+(1-u_j)l_j(\mathbf{u}))}{c_{ii}}\right) \\ l_i(\mathbf{u}) &= \Phi\left(\frac{(a_i-\mu_i)-\sum_{j=1}^{i-1}c_{ij}\Phi^{-1}(u_j h_j(\mathbf{u})+(1-u_j)l_j(\mathbf{u}))}{c_{ii}}\right). \end{aligned}$$

Finally the above integration is performed numerically using the randomized lattice rule techniques (see Cranley and Patterson (1976) and Keast (1973)).

In the next section we present widely used multivariate copula models.

IV.2.2 Multivariate Probit Models

The classical multivariate model for a binary response vector $\mathbf{Y} = (Y_1, \ldots, Y_p)$ is the multivariate probit model. The mass function is given by equation (4.2.2) where C is the multivariate normal copula with mean 0 and latent correlation matrix $\mathbf{R} = [r_{ij}]$

and $F_j(y_j)$ is the distribution function of a Bernoulli random variable with mean $p_j = \Phi(\mu_j)$. Thus the probability mass function of $\mathbf{Y}$ can be expressed as

$$P(\mathbf{y}) = \int_{\mathcal{C}_p} \cdots \int_{\mathcal{C}_1} \frac{1}{(2\pi)^{\frac{p}{2}}|\mathbf{R}|^{\frac{1}{2}}} \exp\left\{-\frac{\mathbf{t}'\mathbf{R}^{-1}\mathbf{t}}{2}\right\} d\mathbf{t} \tag{4.2.14}$$

where

$$\mathcal{C}_j = \begin{cases} (-\infty, \mu_j) & \text{if } y_j = 1; \\ (\mu_j, \infty) & \text{if } y_j = 0. \end{cases}$$

for $1 \leq j \leq p$. Note that equation (4.2.14) can also be written as

$$P(\mathbf{y}; \boldsymbol{\mu}, \mathbf{R}) = \Phi_p\left[(-1)^{1-y_1}\mu_1, \ldots, (-1)^{1-y_p}\mu_p; \mathbf{0}, \mathbf{R}\right] \tag{4.2.15}$$

Although equation (4.2.15) defines a proper probability distribution function for any positive definite matrix $\mathbf{R}$, in order to retain likelihood identifiability $\mathbf{R}$ is restricted to be a correlation matrix. Lesaffre and Kaufmann (1992) have studied (4.2.15) and gave some necessary and sufficient conditions for the existence of unique maximum likelihood estimates of $\boldsymbol{\mu}$ and $\mathbf{R}$ based on a random sample of n observations. We establish some notation to present some details of the maximum likelihood estimation.

We will denote by $\Sigma_{\mathbf{j}_1\mathbf{j}_2}$, the $(p_1 \times p_2)$ dimensional sub-matrix of Σ obtained by selecting elements corresponding to rows $\mathbf{j}_1 = (j_{11}, j_{12}, \ldots, j_{1p_1})'$ and columns $\mathbf{j}_2 = (j_{21}, j_{22}, \ldots, j_{2p_2})'$. Similarly, $\boldsymbol{\mu}_{\mathbf{j}_1}$ denotes the sub-vector of $\boldsymbol{\mu}$ obtained by selecting elements corresponding to $\mathbf{j}_1$.

Let $\mathbf{Z} \sim \mathbf{N}_p(\boldsymbol{\mu}, \Sigma)$. It is well known that the conditional distribution of $\mathbf{Z}_{\mathbf{j}_2}$ given $\mathbf{Z}_{\mathbf{j}_1} = \mathbf{0}$ is also multivariate normal with mean $\boldsymbol{\mu}_{\mathbf{j}_2|\mathbf{j}_1} = \boldsymbol{\mu}_{\mathbf{j}_2} - \Sigma_{\mathbf{j}_2\mathbf{j}_1}\Sigma^{-1}_{\mathbf{j}_1\mathbf{j}_1}\boldsymbol{\mu}_{\mathbf{j}_1}$, and covariance matrix $\Sigma_{\mathbf{j}_2|\mathbf{j}_1} = \Sigma_{\mathbf{j}_2\mathbf{j}_2} - \Sigma_{\mathbf{j}_2\mathbf{j}_1}\Sigma^{-1}_{\mathbf{j}_1\mathbf{j}_1}\Sigma_{\mathbf{j}_1\mathbf{j}_2}$. Further, if $\phi_p(\mathbf{z}; \boldsymbol{\mu}, \Sigma)$ is the p-dimensional multivariate normal density function with mean $\boldsymbol{\mu}$ and covariance matrix $\Sigma = [\sigma_{ij}]$, then

$$\frac{\partial \phi_p(\mathbf{z}; \boldsymbol{\mu}, \Sigma)}{\partial \sigma_{ij}} = \frac{\partial^2 \phi_p(\mathbf{z}, \boldsymbol{\mu}, \Sigma)}{\partial z_i \partial z_j} \tag{4.2.16}$$

$$\text{and } \frac{\partial \phi_p(\mathbf{z}; \boldsymbol{\mu}, \Sigma)}{\partial \sigma_{ii}} = \frac{1}{2}\frac{\partial^2 \phi_p(\mathbf{z}; \boldsymbol{\mu}, \Sigma)}{\partial^2 z_i}. \tag{4.2.17}$$

Equations (4.2.16) and (4.2.17), known as the Plackett's identities, are useful in the next lemma. The derivatives given in Lemma 4.3 are needed for computing the maximum likelihood estimates and the Fisher information.

Lemma 4.3 *Let* $\mathbf{Y}$ *be binary random vector with joint mass function* $P(\mathbf{y};\boldsymbol{\mu},\mathbf{R})$ *given by (4.2.14). Then*

$$
\begin{aligned}
\nabla_{j_1}(\boldsymbol{\mu},\mathbf{R}) &= \frac{\partial P(\mathbf{y};\boldsymbol{\mu},\mathbf{R})}{\partial \mu_{j_1}} \\
&= (-1)^{1+y_{j_1}}\phi(0;\boldsymbol{\mu}_{\mathbf{j}_1},\mathbf{R}_{\mathbf{j}_1\mathbf{j}_1})P(\mathbf{y}_{\mathbf{j}_1^*};\boldsymbol{\mu}_{\mathbf{j}_1^*|\mathbf{j}_1},\mathbf{R}_{\mathbf{j}_1^*|\mathbf{j}_1}) \\
\nabla^2_{j_1j_1}(\boldsymbol{\mu},\mathbf{R}) &= \frac{\partial^2 P(\mathbf{y};\boldsymbol{\mu},\mathbf{R})}{\partial \mu_{j_1}^2} \\
&= (-1)^{y_{j_1}}\phi(0;\boldsymbol{\mu}_{\mathbf{j}_1},\mathbf{R}_{\mathbf{j}_1\mathbf{j}_1})\mathbf{R}^{-1}_{\mathbf{j}_1\mathbf{j}_1}\Big[\boldsymbol{\mu}_{\mathbf{j}_1}P(\mathbf{y}_{\mathbf{j}_1^*};\boldsymbol{\mu}_{\mathbf{j}_1^*},\mathbf{R}_{\mathbf{j}_1^*|\mathbf{j}_1}) \\
&\quad +\mathbf{R}'_{\mathbf{j}_1^*\mathbf{j}_1}\nabla(\boldsymbol{\mu}_{\mathbf{j}_1^*|\mathbf{j}_1},\mathbf{R}_{\mathbf{j}_1^*|\mathbf{j}_1})\Big] \\
\nabla^2_{j_1j_2}(\boldsymbol{\mu},\mathbf{R}) &= \frac{\partial P(\mathbf{y};\boldsymbol{\mu},\mathbf{R})}{\partial r_{j_1j_2}} = \frac{\partial^2 P(\mathbf{y};\boldsymbol{\mu},\mathbf{R})}{\partial \mu_{j_1}\partial \mu_{j_2}} \\
&= (-1)^{y_{j_1}+y_{j_2}}\phi_2(\mathbf{0};\boldsymbol{\mu}_{\mathbf{j}_2},\mathbf{R}_{\mathbf{j}_2\mathbf{j}_2})P(\mathbf{y}_{\mathbf{j}_2^*},\boldsymbol{\mu}_{\mathbf{j}_2^*|\mathbf{j}_2},\mathbf{R}_{\mathbf{j}_2^*|\mathbf{j}_2}) \\
\nabla^3_{j_1j_2j_3}(\boldsymbol{\mu},\mathbf{R}) &= \frac{\partial^2 P(\mathbf{y};\boldsymbol{\mu},\mathbf{R})}{\partial r_{j_1j_2}\partial \mu_{j_3}} \\
&= (-1)^{1+y_{j_1}+y_{j_2}+y_{j_3}}\phi_3(\mathbf{0};\boldsymbol{\mu}_{\mathbf{j}_3},\mathbf{R}_{\mathbf{j}_3\mathbf{j}_3})P(\mathbf{y}_{\mathbf{j}_3^*};\boldsymbol{\mu}_{\mathbf{j}_3^*|\mathbf{j}_3},\mathbf{R}_{\mathbf{j}_2^*|\mathbf{j}_3}) \\
\nabla^3_{j_1j_2}(\boldsymbol{\mu},\mathbf{R}) &= \frac{\partial^2 P(\mathbf{y};\boldsymbol{\mu},\mathbf{R})}{\partial r_{j_1j_2}\partial \mu_{j_1}} \\
&= (-1)^{1+y_{j_1}+y_{j_2}}\phi_2(\mathbf{0};\boldsymbol{\mu}_{\mathbf{j}_2},\mathbf{R}_{\mathbf{j}_2\mathbf{j}_2})e'_{1_2}\mathbf{R}^{-1}_{\mathbf{j}_2\mathbf{j}_2}\Big[\boldsymbol{\mu}_{\mathbf{j}_2}P(\mathbf{y}_{\mathbf{j}_2^*};\boldsymbol{\mu}_{\mathbf{j}_2^*|\mathbf{j}_2},\mathbf{R}_{\mathbf{j}_2^*|\mathbf{j}_2}) \\
&\quad +\mathbf{R}'_{\mathbf{j}_2^*\mathbf{j}_2}\nabla(\boldsymbol{\mu}_{\mathbf{j}_2^*|\mathbf{j}_2},\mathbf{R}_{\mathbf{j}_2^*|\mathbf{j}_2})\Big] \\
\nabla^4_{j_1j_2j_1j_2}(\boldsymbol{\mu},\mathbf{R}) &= \frac{\partial^2 P(\mathbf{y};\boldsymbol{\mu},\mathbf{R})}{\partial r^2_{j_1j_2}} \\
&= (-1)^{y_{j_1}+y_{j_2}}\phi_2(\mathbf{0};\boldsymbol{\mu}_{\mathbf{j}_2},\mathbf{R}_{\mathbf{j}_2\mathbf{j}_2})e'_{1_2}\mathbf{R}^{-1}_{\mathbf{j}_2\mathbf{j}_2}\Big[(\boldsymbol{\mu}_{\mathbf{j}_2}\boldsymbol{\mu}'_{\mathbf{j}_2}+\mathbf{R}_{\mathbf{j}_2\mathbf{j}_2}) \\
&\quad \times P(\mathbf{y}_{\mathbf{j}_2^*};\boldsymbol{\mu}_{\mathbf{j}_2^*|\mathbf{j}_2},\mathbf{R}_{\mathbf{j}_2^*|\mathbf{j}_2})+\mathbf{R}'_{\mathbf{j}_2^*\mathbf{j}_2}\nabla^2(\boldsymbol{\mu}_{\mathbf{j}_2^*|\mathbf{j}_2},\mathbf{R}_{\mathbf{j}_2^*|\mathbf{j}_2})\mathbf{R}_{\mathbf{j}_2^*\mathbf{j}_2}\Big]\mathbf{R}^{-1}_{\mathbf{j}_2\mathbf{j}_2}e_{2_2} \\
\nabla^4_{j_1j_2j_1j_3}(\boldsymbol{\mu},\mathbf{R}) &= \frac{\partial^2 P(\mathbf{y};\boldsymbol{\mu},\mathbf{R})}{\partial r_{j_1j_2}\partial r_{j_1j_3}} \\
&= (-1)^{y_{j_1}+y_{j_2}+y_{j_3}}\phi_3(\mathbf{0};\boldsymbol{\mu}_{\mathbf{j}_3},\mathbf{R}_{\mathbf{j}_3\mathbf{j}_3})e'_{1_3}\mathbf{R}^{-1}_{\mathbf{j}_3\mathbf{j}_3}\Big[\boldsymbol{\mu}_{\mathbf{j}_3}P(\mathbf{y}_{\mathbf{j}_3^*};\boldsymbol{\mu}_{\mathbf{j}_3^*|\mathbf{j}_3},\mathbf{R}_{\mathbf{j}_3^*|\mathbf{j}_3}) \\
&\quad +\mathbf{R}'_{\mathbf{j}_3^*\mathbf{j}_3}\nabla(\boldsymbol{\mu}_{\mathbf{j}_3^*|\mathbf{j}_3},\mathbf{R}_{\mathbf{j}_3^*|\mathbf{j}_3})\Big] \\
\nabla^4_{j_1j_2j_3j_4}(\boldsymbol{\mu},\mathbf{R}) &= \frac{\partial^2 P(\mathbf{y};\boldsymbol{\mu},\mathbf{R})}{\partial r_{j_1j_2}\partial r_{j_3j_4}} \\
&= (-1)^{y_{j_1}+y_{j_2}+y_{j_3}+y_{j_4}}\phi_4(\mathbf{0},\boldsymbol{\mu}_{\mathbf{j}_4},\mathbf{R}_{\mathbf{j}_4\mathbf{j}_4})P(\mathbf{y}_{\mathbf{j}_4^*};\boldsymbol{\mu}_{\mathbf{j}_4^*|\mathbf{j}_4},\mathbf{R}_{\mathbf{j}_4^*|\mathbf{j}_4})
\end{aligned}
$$

where $\mathbf{j}_k=(j_1,\ldots,j_k)'$ *and* $\mathbf{j}_k^*$ *is vector of indexes that are complimentary to* $\mathbf{j}_k$, e_{j_k} *is* j*th unit vector in* $\mathbb{R}^k$. *In the above identities* $\nabla(\boldsymbol{\mu},\mathbf{R})=[\nabla_1(\boldsymbol{\mu},\mathbf{R}),\ldots,\nabla(\boldsymbol{\mu},\mathbf{R})]'$ *and* $\nabla^2(\boldsymbol{\mu},\mathbf{R})=[\nabla_{jk}(\boldsymbol{\mu},\mathbf{R})]$.

Lemma 4.3 is proved with routine differentiation and tedious algebra. We will need the following matrices in the next section:

$$\nabla^3(\boldsymbol{\mu}, \mathbf{R}) = \begin{bmatrix} \nabla^3_{121}(\boldsymbol{\mu}, \mathbf{R}) & \dots & \nabla^3_{12p}(\boldsymbol{\mu}, \mathbf{R}) \\ \vdots & \dots & \vdots \\ \nabla^3_{(p-1)p1}(\boldsymbol{\mu}, \mathbf{R}) & \dots & \nabla^3_{(p-1)pp}(\boldsymbol{\mu}, \mathbf{R}) \end{bmatrix}$$

and

$$\nabla^4(\boldsymbol{\mu}, \mathbf{R}) = \begin{bmatrix} \nabla^4_{1212}(\boldsymbol{\mu}, \mathbf{R}) & \dots & \nabla^4_{121p}(\boldsymbol{\mu}, \mathbf{R}) \\ \vdots & \dots & \vdots \\ \nabla^4_{(p-1)p12}(\boldsymbol{\mu}, \mathbf{R}) & \dots & \nabla^4_{(p-1)p(p-1)p}(\boldsymbol{\mu}, \mathbf{R}) \end{bmatrix}.$$

Estimation methods for multivariate probit model: Suppose that we have a random sample $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ of observations from the distribution (4.2.15) and corresponding explanatory variables $\mathbf{X}_1, \dots, \mathbf{X}_n$. Assume that the mean $\boldsymbol{\mu}$ is linked to the covariates via a link function and a regression parameter $\boldsymbol{\beta}$, as in Chapter II. In many application it is reasonable to assume that the latent correlation matrix $\mathbf{R}$ is fully characterized by q dimensional parameter vector $\boldsymbol{\alpha}$. Then the unknown parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\alpha})$ can be estimated using the maximum likelihood estimation procedure or by solving Godambe's optimal estimating equations. For a square matrix $A = [a_{ij}]$ of order p, define $\overline{A}$ as the vector $(a_{12}, a_{13}, \dots, a_{(p-1),p})$ obtained taking the elements of A above the diagonal. For maximum likelihood estimation, the log-likelihood function is given by

$$L(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log P(\mathbf{y}_i; \boldsymbol{\theta}) \tag{4.2.18}$$

and the likelihood estimating equation is

$$\begin{aligned} \nabla(\boldsymbol{\theta}) &= \frac{\partial L}{\partial \boldsymbol{\theta}} \\ &= \sum_{i=1}^{n} \frac{1}{P(\mathbf{y}_i; \boldsymbol{\theta})} J_i(\boldsymbol{\theta})' \nabla^*_i(\boldsymbol{\theta}) = 0, \end{aligned}$$

where $\nabla^*_i(\boldsymbol{\theta}) = \left[\nabla(\boldsymbol{\mu}_i, \mathbf{R}), \overline{\nabla^2(\boldsymbol{\mu}_i, \mathbf{R})}\right]'$ and the jacobian matrix $J_i(\boldsymbol{\theta})$ is given by

$$J_i(\boldsymbol{\theta}) = \begin{bmatrix} \left(\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\eta}_i}\right) \mathbf{X}_i & \mathbf{0} \\ \mathbf{0} & \left(\frac{\partial \overline{\mathbf{R}(\boldsymbol{\alpha})}}{\partial \boldsymbol{\alpha}}\right) \end{bmatrix}.$$

The Hessian matrix of $L(\boldsymbol{\theta})$ is given by

$$\begin{aligned} H(\boldsymbol{\theta}) &= \frac{\partial^2 L}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}} \\ &= \sum_{i=1}^{n} \frac{1}{P(\mathbf{y}_i,\boldsymbol{\theta})} J_i(\boldsymbol{\theta})' \left[\nabla_i^{*2}(\boldsymbol{\theta}) - \frac{1}{P(\mathbf{y}_i;\boldsymbol{\theta})}\nabla_i^*(\boldsymbol{\theta})\nabla_i^{*\prime}(\boldsymbol{\theta})\right] J_i(\boldsymbol{\theta}) \end{aligned}$$

where

$$\nabla_i^{*2}(\boldsymbol{\theta}) = \begin{bmatrix} \nabla^2(\boldsymbol{\mu}_i,\mathbf{R}) & \nabla^3(\boldsymbol{\mu}_i,\mathbf{R})' \\ \nabla^3(\boldsymbol{\mu}_i,\mathbf{R}) & \nabla^4(\boldsymbol{\mu}_i,\mathbf{R}) \end{bmatrix}.$$

The maximum likelihood estimate $\widehat{\boldsymbol{\theta}}$ can be obtained by the usual Newton-Raphson iterations:

$$\widehat{\boldsymbol{\theta}}_{k+1} = \widehat{\boldsymbol{\theta}}_k - H^{-1}(\widehat{\boldsymbol{\theta}}_k)\nabla(\widehat{\boldsymbol{\theta}}_k).$$

It is well known that $\widehat{\boldsymbol{\theta}}$ is AMVN$(\boldsymbol{\theta}, \mathcal{I}^{-1}(\boldsymbol{\theta}))$ where

$$\mathcal{I}(\boldsymbol{\theta}) = \mathrm{E}[-H(\boldsymbol{\theta})] = \sum_{i=1}^{n} J_i'(\boldsymbol{\theta})\mathrm{E}\left[\frac{1}{P(\mathbf{y}_i;\boldsymbol{\theta})^2}\nabla_i^*(\boldsymbol{\theta})\nabla_i^*(\boldsymbol{\theta})'\right] J_i(\boldsymbol{\theta}) \tag{4.2.19}$$

is the Fisher information.

The maximum likelihood estimation is very time consuming and computationally challenging, since it requires higher dimensional numerical integrations at each iteration. An alternative and simpler method is to estimate $\boldsymbol{\theta}$ solving Godambe's optimal estimating equation given by

$$\sum_{i=1}^{n} J_i'(\boldsymbol{\theta})D_i'V_i^{*-1}(\xi_i - \boldsymbol{\eta}_i^*) = 0, \tag{4.2.20}$$

where $\xi_i = \left[\mathbf{Y}_i, \overline{\mathbf{Z}_i\mathbf{Z}_i'}\right]$, $\mathbf{Z}_i = (\mathbf{Y}_i - \mathrm{E}(\mathbf{Y}_i))$, $\boldsymbol{\eta}_i^*$, V_i^* are the mean and covariance of ξ_i. The matrix $D_i = \left(\frac{\partial\boldsymbol{\eta}_i^*}{\partial\boldsymbol{\mu}_i} \;\; \frac{\partial\boldsymbol{\eta}_i^*}{\partial\mathbf{R}}\right)'$ can be computed using the derivatives in Lemma 4.3.

Suppose that $\widehat{\boldsymbol{\theta}}_g$ is the solution of the equation (4.2.20), then it is well known that (Godambe (1991)), $\widehat{\boldsymbol{\theta}}_g$ is AMVN$(\boldsymbol{\theta}, \mathcal{G}^{-1}(\boldsymbol{\theta}))$ where $\mathcal{G}(\boldsymbol{\theta})$ is the Godambe information is given by

$$\mathcal{G}(\boldsymbol{\theta}) = \sum_{i=1}^{n} J_i'(\boldsymbol{\theta})D_i'V_i^{*-1}\, D_i\, J_i(\boldsymbol{\theta}). \tag{4.2.21}$$

An example: To illustrate the two methods of estimation for the multivariate probit model, we consider a subset of data from the "Six Cities Study," that was analyzed

by Ware et al. (1984), Fitzmaurice and Laird (1993) and Chib and Greenberg (1996). The Six Cities study is a longitudinal investigation of indoor and outdoor air pollution on respiratory health of child. A subset of the original data given in Table 4.1, contains complete record of wheezing status at ages 7, 8, 9 and 10, as well as information about maternal smoking of 537 children from Stuebenville, Ohio. One of the objectives of this study was to determine the effect of maternal smoking on respiratory health of a child over time. Fitzmaurice and Laird (1993) used the marginal logit model to analyze the probability of wheezing using alternating logistic regression method of Carey et al. (1993). This method involves modelling dependence via bivariate odds ratio instead of the bivariate correlations to accommodate pair-wise restrictions on correlations. The method then proposes GEE-like estimating equation for odds-ratio estimation.

Table 4.1: Six Cities Data: Wheezing Status of 537 Children

No maternal smoking					No maternal smoking				
Age				Count	Age				Count
7	8	9	10		7	8	9	10	
0	0	0	0	237	0	0	0	0	118
0	0	0	1	10	0	0	0	1	6
0	0	1	0	15	0	0	1	0	8
0	0	1	1	4	0	0	1	1	2
0	1	0	0	16	0	1	0	0	11
0	1	0	1	2	0	1	0	1	1
0	1	1	0	7	0	1	1	0	6
0	1	1	1	3	0	1	1	1	4
1	0	0	0	24	1	0	0	0	7
1	0	0	1	3	1	0	0	1	3
1	0	1	0	3	1	0	1	0	3
1	0	1	1	2	1	0	1	1	1
1	1	0	0	6	1	1	0	0	4
1	1	0	1	2	1	1	0	1	2
1	1	1	0	5	1	1	1	0	4
1	1	1	1	11	1	1	1	1	7

We use following multivariate probit model to incorporate dependence between repeated measures.

$$\mu_{ij} = \beta_0 + (\text{Age}_j)\,\beta_1 + (\text{MS}_i)\,\beta_2 + (\text{Age}_i \times \text{MS}_j)\,\beta_3$$

where Y_{ij} is the indicator of wheezing for child i at jth year of study, 'AGE' is age in years since the child's 9th birthday and 'MS' is indicator of maternal smoking.

We model the latent correlations between the four repeated responses using four correlation structures: Independence, Unstructured, Exchangeable and Autoregressive. Maximum likelihood estimates of β and α and standard errors obtained from the diagonal elements of the Fisher information (σ_1) and Godambe information (σ_2) methods are summarized in Table 4.2.

Table 4.2: Six Cities Data: Maximum Likelihood Estimates

	Independence					Unstructured				
	Est.	S.E.	σ_1	σ_2	ARE	Est.	S.E.	σ_1	σ_2	ARE
β	-1.1259	.0472	.0471	.0471	1.0000	-1.1226	.0636	.0623	.0625	1.0040
	-0.0768	.0376	.0375	.0375	1.0000	-0.0784	.0315	.0313	.0313	1.0007
	0.1709	.0762	.0761	.0761	1.0000	0.1596	.1014	.1009	.1012	1.0071
	0.0367	.0615	.0611	.0611	1.0000	0.0374	.0511	.0506	.0506	1.0009
α	-	-	-	-	-	0.5835	.0723	.0659	.0662	1.0090
	-	-	-	-	-	0.5232	.0744	.0720	.0728	1.0207
	-	-	-	-	-	0.6870	.0601	.0578	.0580	1.0070
	-	-	-	-	-	0.5789	.0767	.0691	.0695	1.0116
	-	-	-	-	-	0.5577	.0816	.0719	.0726	1.0221
	-	-	-	-	-	0.6305	.0706	.0663	.0667	1.0128
L	-909.7206					-794.7184				

	Exchangeable					Autoregressive				
	Est.	S.E.	σ_1	σ_2	ARE	Est.	S.E.	σ_1	σ_2	ARE
β	-1.1195	.0629	.0619	.0621	1.0046	-1.1368	.0601	.0605	.0606	1.0021
	-0.0777	.0304	.0303	.0303	1.0010	-0.0816	.0362	.0363	.0364	1.0009
	0.1611	.1007	.1003	.1006	1.0071	0.1598	.0972	.0982	.0983	1.0030
	0.0384	.0492	.0491	.0491	1.0006	0.0438	.0589	.0592	.0592	1.0008
α	0.5984	.0415	.0405	.0411	1.0296	0.6447	.0275	.0318	.0320	1.0106
L	-797.6538					-804.1492				

The estimate of β_1 in all the four models is negative indicating the decline in the probability of wheezing with age. All the four models indicate that children of mothers who smoke are more like to be diagnosed for respiratory disease than the children of mothers who do not. Note that all the four models present very similar estimates of β. The log-likelihood statistic for the model which assumes independence among the repeated measure is smallest indicating inadequate dependence modelling. Using log-likelihood ratio test we can conclude that the model with exchangeable

latent correlation structure is an adequate model.

The observed asymptotic relative efficiency (ARE) of Godambe's optimal estimating functions versus maximum likelihood estimation are also displayed in Table 4.2. Since the data is large the ARE's are close to 1 for all the four models indicating both maximum likelihood estimation and Godambe optimal estimating function present equally efficient estimates β. This suggests that Godambe's optimal estimation function could be used as an alternative to the maximum likelihood estimation. We have repeated the analysis using logit link function and found that estimates of latent correlations and their standard deviations are invariant to the choice of the link function. Also the log-likelihood statistic is also unaffected by the link function.

In next section we present probability models derived by compounding several known marginal distributions.

IV.3 Multivariate Mixture Models

In this section we will study multivariate mixture models. The basic idea of multivariate mixture models is similar to modelling a cluster of outcomes with random effects for the subjects in a cluster and common fixed effects through distribution of random effects.

Let $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_p)$ be a multivariate discrete response vector. We assume that for given $\mathbf{\Lambda} = (\Lambda_1, \ldots, \Lambda_q)$, the probability mass function of $\mathbf{Y}$ is $f(y_1, y_2, \ldots, y_p; \mathbf{\Lambda})$. Suppose further that $\mathbf{\Lambda}$ is random vector, and the distribution of λ is a multivariate distribution $G(\lambda_1, \lambda_2, \ldots, \lambda_q)$ which can be expressed in terms of a copula function $C(u_1, u_2, \ldots, u_q)$ with univariate margins $G_j, j = 1, 2, \ldots, q$. The marginal probability mass function of $\mathbf{Y}$ is

$$P(\mathbf{y}) = \int f(\mathbf{y}; \boldsymbol{\lambda})\, c(G_1(\lambda_1), \ldots, G_q(\lambda_q)) \prod_{j=1}^{q} g_j(\lambda_j) d\boldsymbol{\lambda}, \tag{4.3.1}$$

where $c(u_1, \ldots, u_q)$ is the density of the copula $C(.)$. A special case of (4.3.1) is obtained when given $\mathbf{\Lambda}$, the random variables Y_i are independent and $P_i(y_i)$, the marginal probability mass function of Y_i depends on a parameter $\gamma_i = \gamma_i(\mathbf{\Lambda})$. In this

case (4.3.1) reduces to

$$P(\mathbf{y}) = \int \cdots \int \prod_{i=1}^{n} f(y_i;\gamma_i) c(G_1(\lambda_1),\ldots,G_q(\lambda_q)) \prod_{j=1}^{q} g_j(\lambda_j) d\boldsymbol{\lambda} \qquad (4.3.2)$$

In the next section we will study the popular multivariate mixture models for analyzing binary and count responses.

IV.3.1 Multivariate Probit-normal Model

The multivariate probit model is obtained by introducing latent variables $\mathbf{A} = (A_1,\ldots,A_p)$ such that the variable Y_i assumes value 1 if $A_j < \lambda_j$, and 0 otherwise, for $j = 1,2\ldots,p$. Let us suppose that the latent variable $\mathbf{A}$ is distributed as multivariate normal with mean 0 and variance covariance matrix Σ. Furthermore, the vector of threshold values $\lambda = (\lambda_1,\ldots,\lambda_p)$ is assumed to be multivariate normal vectors with mean μ_0 and variance covariance matrix Σ_0. For the model to be identifiable, the matrix Σ is restricted to a non-singular correlation matrix. Covariates can be incorporated in model easily by assuming $\lambda_j = \mathbf{x}_j'\boldsymbol{\beta}$, where $\boldsymbol{\beta} \sim N_k(0,\Sigma^*)$. The distribution of $\mathbf{Y} = (Y_1,\ldots,Y_p)$ is the multivariate probit-normal and it is given by

$$P(\mathbf{y}) \;=\; \int_{\mathbb{R}^p} P(\mathbf{y};\beta,\Sigma)\, \phi_k(\boldsymbol{\beta};\Sigma^*)\, d\beta, \qquad (4.3.3)$$

where $P(.\,;.)$ is the multivariate probit mass function defined in (4.2.14). The multivariate probit-normal reduces to the multivariate probit model when $\Sigma = \mathbf{O}$.

We can introduce another layer of randomness by assuming a prior distribution of Σ. Chib and Greenberg (1996) have analyzed the Six Cities data described in section IV.2.2 using multivariate probit-normal distribution. They assumed multivariate normal prior for $p(p-1)/2$ off-diagonals of Σ. Chib and Greenberg (1996) have used Markov Chain Monte Carlo method along with Gibb's simulator to estimate the posterior distribution of the regression parameters β and latent correlation matrix Σ. The marginal means of this posterior distribution were close to the maximum likelihood estimates reported in Table 4.2 but the variance of the posterior distribution was greater due to the additional layer of randomness.

IV.3.2 Multivariate Poisson Log-normal Models

The multivariate Poisson log-normal models are sub-class of general mixture models defined by (4.3.1). Suppose that $\mathbf{\Lambda}$ is distributed as multivariate log-normal distribution with parameters $(\boldsymbol{\mu}, \boldsymbol{\sigma} R \boldsymbol{\sigma})$ where $\boldsymbol{\sigma} = \text{diag}(\sigma_1, \ldots, \sigma_p)$ and R is a correlation matrix characterized by a q dimensional vector $\boldsymbol{\alpha}$. Given $\mathbf{\Lambda} = (\Lambda_1, \ldots, \Lambda_q)$, let Y_j, $1 \leq j \leq p$ be independent Poisson random variables with means λ_i, $1 \leq i \leq p$. The joint probability mass function of $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_p)$ is given by

$$P(\mathbf{y}) = \int_{\mathbb{R}_+^p} \left\{ \prod_{j=1}^{p} \frac{e^{-\lambda_j} \lambda_j^{y_j}}{y_j!} \right\} \left\{ \frac{\exp\left[-\frac{1}{2}(\log \boldsymbol{\lambda} - \boldsymbol{\mu})'(\boldsymbol{\sigma} R \boldsymbol{\sigma})^{-1}(\log \boldsymbol{\lambda} - \boldsymbol{\mu})\right]}{(2\pi)^{\frac{p}{2}} |\boldsymbol{\sigma} R \boldsymbol{\sigma}|^{\frac{1}{2}} \prod_{j=1}^{p} \lambda_j} \right\} d\boldsymbol{\lambda} \tag{4.3.4}$$

Let $C = [c_{ij}]$ be the lower triangular matrix in the cholesky factorization of R. If $\mathbf{Z} = C^{-1}\boldsymbol{\sigma}^{-1}(\log(\mathbf{\Lambda}) - \boldsymbol{\mu})$, then equation (4.3.4) can be rewritten as

$$P(\mathbf{y}) = \int_{\mathbb{R}^p} \left\{ \prod_{j=1}^{p} \frac{\exp[-e^{\mu_j + \sigma_j e_j' C\mathbf{z}} + y_j(\mu_j + \sigma_j e_j' C\mathbf{z})]}{y_j!} \right\} \left\{ \frac{e^{-\frac{1}{2}\mathbf{z}'\mathbf{z}} d\mathbf{z}}{(2\pi)^{\frac{p}{2}}} \right\} d\mathbf{z} \tag{4.3.5}$$

where e_j is jth unit vector in $\mathbb{R}^p$. Also for $d_j = 0, 1, 2, \ldots$; $c = 1, 2, \ldots, p$, the joint factorial moments can be obtained as

$$\begin{aligned}
\boldsymbol{\mu}'_{(\mathbf{d})}(\mathbf{Y}) &= \mathrm{E}\left\{ \prod_{j=1}^{p} \left[\frac{Y_j!}{(Y_j - d_j)!} \right] \right\} \\
&= \mathrm{E}\left\{ \exp\left[\sum_{j=1}^{p} d_j(\mu_j + \sigma_j e_j' C Z) \right] \right\} \\
&= \exp\left[\sum_{j=1}^{p} d_j \mu_j + \frac{1}{2} \left(\sum_{j=1}^{p} d_j \sigma_j e_j \right)' R \left(\sum_{j=1}^{p} d_j \sigma_j e_j \right) \right] \\
&= \exp\left[\sum_{j=1}^{p} d_j \mu_j + \frac{1}{2} \sum_{j=1}^{p} \sum_{k=1}^{p} d_j d_k \sigma_j \sigma_k r_{jk} \right]
\end{aligned}$$

According to (Johnson et al., 1997, p.5) the joint moments of $\mathbf{Y}$ can be expressed as

$$\begin{aligned}
\boldsymbol{\mu}'_{\mathbf{d}}(\mathbf{Y}) &= \mathrm{E}\left\{ \prod_{j=1}^{p} Y_j^{d_j} \right\} \\
&= \sum_{j_1=0}^{d_1} \cdots \sum_{j_p=0}^{d_p} \left[\prod_{k=1}^{p} S(d_k, j_k) \right] \boldsymbol{\mu}'_{(\mathbf{j})}(\mathbf{Y})
\end{aligned}$$

where $S(.,.)$ is Sterling number of second kind defined by recurrence relation

$$S(n,1)=1,\ \ S(n,i)=S(n-i,i-1)+iS(n-1,i)$$
$$\text{for } i=1,\ldots,n-1,\ \text{and } S(n,n)=1.$$

Also the joint central moments can be derived using the relation

$$\begin{aligned}\boldsymbol{\mu}_{\mathbf{d}}(\mathbf{Y}) &= \mathrm{E}\left\{\prod_{j=1}^{p}[Y_j-\mathrm{E}(Y_j)]^{d_j}\right\}\\ &= \sum_{j_1=0}^{d_1}\cdots\sum_{j_p=0}^{d_p}(-1)^{\mathbf{1'j}}\left[\prod_{k=1}^{p}\binom{d_k}{j_k}\mathrm{E}(Y_k)^{j_k}\right]\boldsymbol{\mu}'_{\mathbf{d-j}}(\mathbf{Y})\end{aligned}$$

Therefore

$$\begin{aligned}\mu_j^* &= \mathrm{E}(Y_j)\\ &= \exp\left\{\mu_j+\frac{1}{2}\sigma_j^2\right\}\\ \sigma_{jk}^* &= \mathrm{cov}(Y_j,Y_k)\\ &= \begin{cases}\mu_j^*+\mu_j^{*2}\left[\exp(\sigma_j^2)-1\right] & \text{if } j=k\\ \mu_j^*\mu_k^*\left[\exp(r_{jk}\sigma_j\sigma_k)-1\right] & \text{otherwise}\end{cases}\end{aligned}$$

Note that $\mathrm{Var}(Y_j)>\mathrm{E}(Y_j)$ and $\mathrm{Corr}(Y_j,Y_k)\leq\mathrm{Corr}(\Lambda_j,\Lambda_k)$. Hence the multivariate Poisson log normal model is suitable for over-dispersed count data.

Aitchison and Ho (1989) have used special of case of multivariate Poisson log-normal model with $\sigma_1=\sigma_2\ldots\sigma_p=\sigma$, to model the growth of competing spices of bacteria which depend on certain type of plants for livelihood. Next we discuss maximum likelihood estimates of the unknown parameters based on a random sample from the multivariate Poisson log-normal model.

Maximum likelihood estimation: Let $\mathbf{Y}_1,\mathbf{Y}_2,\ldots,\mathbf{Y}_n$ be a random sample of n independent observation from distribution (4.3.4). Then the log-likelihood of the observed data as a function of $\boldsymbol{\Theta}=(\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\alpha})$ is given by

$$\log(L)=\sum_{i=1}^{n}\log P(\mathbf{y}_i).$$

The next lemma gives expressions for the derivatives needed for the computation of the maximum likelihood estimates and the Fisher information matrix.

Lemma 4.4 *Let* $\mathbf{Y}$ *be a* p *dimensional random vector distributed as* $P(\mathbf{y})$ *given by (4.3.4). Then the following hold:*

$$
\begin{aligned}
\nabla_{\mu} &= \frac{\partial \log P(\mathbf{y})}{\partial \boldsymbol{\mu}} \\
&= \mathrm{E}\left[(\mathbf{y}-\boldsymbol{\Lambda})|\mathbf{y}\right] \\
\nabla_{\sigma} &= \frac{\partial \log P(\mathbf{y})}{\partial \boldsymbol{\sigma}} \\
&= \mathrm{E}\left[\boldsymbol{\sigma}^{-1}D(\log\boldsymbol{\Lambda}-\boldsymbol{\mu})(\mathbf{y}-\boldsymbol{\Lambda})|\mathbf{y}\right] \\
\nabla_{\alpha_j} &= \frac{\partial \log P(\mathbf{y})}{\partial \boldsymbol{\alpha}} \\
&= \mathrm{E}\left[(\mathbf{y}-\boldsymbol{\Lambda})'\left(\frac{\partial C}{\partial \alpha_j}\right)C^{*}(\log\boldsymbol{\Lambda}-\boldsymbol{\mu})|\mathbf{y}\right] \\
\nabla_{\mu\mu'} &= \frac{1}{P(\mathbf{y})}\frac{\partial^2 P(\mathbf{y})}{\partial\boldsymbol{\mu}\partial\boldsymbol{\mu}'} \\
&= \mathrm{E}\left[(\mathbf{y}-\boldsymbol{\Lambda})(\mathbf{y}-\boldsymbol{\Lambda})'-D(\boldsymbol{\Lambda})|\mathbf{y}\right] \\
\nabla_{\mu\sigma'} &= \frac{1}{P(\mathbf{y})}\frac{\partial^2 P(\mathbf{y})}{\partial\boldsymbol{\mu}\partial\boldsymbol{\sigma}'} \\
&= \mathrm{E}\left\{\boldsymbol{\sigma}^{-1}D(\log\boldsymbol{\Lambda}-\mu)\left[(\mathbf{y}-\boldsymbol{\Lambda})(\mathbf{y}-\boldsymbol{\Lambda})'-D(\boldsymbol{\Lambda})\right]|\mathbf{y}\right\} \\
\nabla_{\sigma\sigma'} &= \frac{1}{P(\mathbf{y})}\frac{\partial^2 P(\mathbf{y})}{\partial\boldsymbol{\sigma}\partial\boldsymbol{\sigma}'} \\
&= \mathrm{E}\left\{\boldsymbol{\sigma}^{-1}D(\log\boldsymbol{\Lambda}-\boldsymbol{\mu})\left[(\mathbf{y}-\boldsymbol{\Lambda})(\mathbf{y}-\boldsymbol{\Lambda})'-D(\Lambda)\right]\boldsymbol{\sigma}^{-1}D(\log\boldsymbol{\Lambda}-\boldsymbol{\mu})|\mathbf{y}\right\} \\
\nabla_{\mu\alpha_j} &= \frac{1}{P(\mathbf{y})}\frac{\partial^2 P(\mathbf{y})}{\partial\boldsymbol{\mu}\partial\alpha_j} \\
&= \mathrm{E}\left\{\left[(\mathbf{y}-\boldsymbol{\Lambda})(\mathbf{y}-\boldsymbol{\Lambda})'-D(\boldsymbol{\Lambda})\right]\left(\frac{\partial C}{\partial\alpha_j}\right)C^{*}(\log\boldsymbol{\Lambda}-\boldsymbol{\mu})|\mathbf{y}\right\} \\
\nabla_{\sigma\alpha_j} &= \frac{1}{P(\mathbf{y})}\frac{\partial^2 P(\mathbf{y})}{\partial\boldsymbol{\sigma}\partial\alpha_j'} \\
&= \mathrm{E}\left\{\boldsymbol{\sigma}^{-1}D(\log\boldsymbol{\Lambda}-\mu)\left[(\mathbf{y}-\boldsymbol{\Lambda})(\mathbf{y}-\boldsymbol{\Lambda})'-D(\boldsymbol{\Lambda})\right]\left(\frac{\partial C}{\partial\alpha_j}\right)C^{*}(\log\boldsymbol{\Lambda}-\boldsymbol{\mu})\right. \\
&\quad \left.+\left(\frac{\partial C}{\partial\alpha_j}\right)C^{*}\boldsymbol{\sigma}^{-1}D(\log\boldsymbol{\Lambda}-\boldsymbol{\mu})(\mathbf{y}-\boldsymbol{\Lambda})|\mathbf{y}\right\} \\
\nabla_{\alpha_j\alpha_k} &= \frac{1}{P(\mathbf{y})}\frac{\partial^2 P(\mathbf{y})}{\partial\alpha_j\partial\alpha_k} \\
&= \mathrm{E}\left\{(\log\boldsymbol{\Lambda}-\boldsymbol{\mu})'C^{*'}\left(\frac{\partial C}{\partial\alpha_j}\right)'\left[(\mathbf{y}-\boldsymbol{\Lambda})(\mathbf{y}-\boldsymbol{\Lambda})'-D(\Lambda)\right]\right. \\
&\quad \left.\times\left(\frac{\partial C}{\partial\alpha_j}\right)C^{*}(\log\boldsymbol{\Lambda}-\boldsymbol{\mu})+(\mathbf{y}-\boldsymbol{\Lambda})'\left(\frac{\partial^2 C}{\partial\alpha_j\partial\alpha_k}\right)C^{*}(\log\boldsymbol{\Lambda}-\boldsymbol{\mu})|\mathbf{y}\right\}
\end{aligned}
$$

where $D(\mathbf{x}) = diag(x_1 \ldots, x_p)$ *and* C^{*} *is Choleskey decomposition matrix of* R^{-1}.

By Lemma 4.4 the likelihood estimating equations for estimating Θ are

$$\frac{\partial \log L}{\partial \Theta} = \sum_{i=1}^{n} \nabla_{i\Theta} = 0$$

where $\nabla_{i\Theta} = \left(\nabla_{i\mu}\ \nabla_{i\sigma}\ \nabla_{i\alpha_1} \ldots, \nabla_{i\alpha_q}\right)'$. Next the variance covariance matrix of maximum likelihood estimators is estimated as $[-F(\Theta)]^{-1}$ where $F(\Theta)$ is observed Fisher information matrix given as

$$\begin{aligned} F(\Theta) &= -\frac{\partial^2 \log L}{\partial \Theta \partial \Theta'} \\ &= \sum_{i=1}^{n} [\nabla_{i\Theta}\ \nabla'_{i\Theta} - \mathbf{I}_{1_i}(\Theta)] \end{aligned}$$

and

$$\mathbf{I}_1(\Theta) = \begin{bmatrix} \nabla_{\mu\mu'} & \nabla_{\mu\sigma'} & \nabla_{\mu\alpha_1} & \cdots & \nabla_{\mu\alpha_q} \\ \nabla'_{\mu\sigma'} & \nabla_{\sigma\sigma'} & \nabla_{\sigma\alpha_1} & \cdots & \nabla_{\sigma\alpha_q} \\ \nabla'_{\mu\alpha_1} & \nabla'_{\sigma\alpha_1} & \nabla_{\alpha_1\alpha_1} & \cdots & \nabla_{\alpha_1\alpha_q} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \nabla'_{\mu\alpha_q} & \nabla'_{\sigma\alpha_q} & \nabla_{\alpha_1\alpha_q} & \cdots & \nabla_{\alpha_q\alpha_q} \end{bmatrix}$$

If the objective is to model the mean of the observed counts on covariates then we can assume that $E(Y_i) = \boldsymbol{\mu}_i = \mathbf{x}_i'\beta$, where β is the regression parameter. Also it is possible to study whether different set of covariates $\mathbf{z}_{ij}$ contribute to additional variability of outcomes by using the relation $\sigma_{ij} = \exp(\mathbf{z}'_{ij}\nu)$. The maximum likelihood estimates of $\boldsymbol{\theta}^* = (\beta, \nu, \alpha)'$ can be computed solving the likelihood equation

$$\begin{aligned} \nabla_{\boldsymbol{\theta}^*} &= \frac{\partial \log L}{\partial \boldsymbol{\theta}^*} \\ &= \sum_{i=1}^{n} J_i' \nabla_{i\Theta} = 0, \end{aligned}$$

where the jacobian matrix J_i is given by

$$J_i = \begin{bmatrix} \mathbf{X}_i & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & D(\boldsymbol{\sigma}_i)\mathbf{Z}_i & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix}$$

and the estimated covariance of the maximum likelihood estimates is by

$$\mathrm{Cov}(\widehat{\boldsymbol{\theta}^*}) = \left\{ \sum_{i=1}^{n} J_i' \left[\nabla_{i\Theta}\ \nabla'_{i\Theta} - \mathbf{I}_{1_i}(\Theta)\right] J_i \right\}^{-1}.$$

Table 4.3: Seizure Data: MLE for Multivariate Poisson Log-normal Model

	Dispersion Type					
	Time Independent		Time Dependent		Covariate Dependent	
Parameter	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
β	-1.3174	1.1339	-1.6818	1.2316	-1.0175	1.1611
	0.8297	0.0993	0.9354	0.1173	0.7893	0.1202
	-0.8275	0.3104	-0.8326	0.3432	-1.2369	0.5121
	0.3278	0.1396	0.2499	0.1524	0.4979	0.2481
	0.4725	0.3379	0.5624	0.3666	0.4070	0.3366
	-0.0913	0.0824	-0.0473	0.0894	0.0024	0.0903
ν	-0.5317	0.0667	-0.6141	0.1367	-0.6406	1.7367
	-	-	-0.4810	0.1213	0.1375	0.1675
	-	-	-0.2592	0.1219	0.8047	0.6322
	-	-	-0.9061	0.2100	-0.3302	0.3030
	-	-	-	-	-0.0572	0.5043
	-	-	-	-	-0.4015	0.2002
α	0.5710	-	0.6504	-	0.6233	-
$\log(L)$	-624.476		-619.5648		-620.3948	

We now illustrate the application of the multivariate Poisson log-normal models using epileptic seizure data described in Section III.2. Thall and Vail (1990) and various other authors have argued for a need to incorporate heteroscedasticity of the observed bi-weekly seizure counts. This is valid concern as the first hand descriptive analysis show that variance of seizure counts reported on week 2 is significantly greater than that of seizure counts reported on week 4. One may be interested in testing whether this heteroscedasticity is due to any of covariates such as Progabide drug.

To address the concerns we present three different models. All three model possess the same log-linear relationship between marginal means and covariates (see (3.2.19)), along with the exchangeable correlation assumption for R. In the first model the dispersion parameter is assumed to be time independent or constant, that is, $\sigma_{ij} = \exp(\nu)$. In the second model we assume the dispersion parameter is time varying, that is, $\sigma_{ij} = \exp(\nu_j)$. In the third model, we allow the dispersion parameter to be covariate dependent via log-linear relation, that is, $\sigma_{ij} = \exp(\mathbf{x}'_{ij}\nu)$. We use Monte Carlo Markov Chain method (using 5000 simulation points) for numerical computation of the joint probability mass function.

The maximum likelihood estimates for these models are reported in Table 4.3.

Standard errors for the parameter α were found to be unreliable due to the lack of convergence and hence not reported. Note that the estimates of parameters β, α in Table 4.3 are not comparable to the quasi-least squares and modified Gaussian estimates as over-dispersion is not considered in moment based methods.

Since the second and third models are generalizations of the first model, we have conducted goodness of fit for the second model relative to the first model by testing the null hypothesis $H_0 : \nu_1 = \nu_2 = \ldots = \nu_4$ using the likelihood ratio test. The observed value of χ^2 test statistic is equal to 10.21 with p-value 0.016. Therefore we conclude that the second model is superior to the first model. Similarly the relative performance of the third model with respect to the first model is checked testing the hypothesis $H_0 : \nu_2 = \nu_3 = \ldots = \nu_6 = 0$. The observed value of chi-squared likelihood ratio test statistic is 8.12 resulting in a p-value of 0.1495. Therefore the third model is not a significant improvement over the first. We conclude that the second model which incorporates time dependent over-dispersion is the best model for this data. Also the extra-variation in reported seizure counts is progressively decreasing which could mean that improved consistency of reported seizure counts.

In next section we present discrete choice models which are widely used in various applications especially in econometrics. These models are variants of the multivariate mixture and the multivariate copula models.

IV.4 Multivariate Discrete Choice Models

The main objective in an economic analysis is to find cause-and-effect relationship between marketing-mix variables such as product features, advertisements, and marketing performance variables such market share, total sales etc.

If the outcomes are unordered representing a choice or a decision, we can use the following multinomial discrete choice model to analyze the data. The discrete choice approach attempts to model the choice probability via a utility function which is intended to capture the perceived attractiveness to a decision maker. However, many users with identical attributes (covariates) vectors do not always make the decision. Hence, for each decision maker the random perceived utility as a function

of attributes is defined as

$$U_{ij}(\mathbf{x}_{ij}, \beta) = V_{ij}(\mathbf{x}_{ij}, \beta) + e_{ij}(\mathbf{x}_{ij}, \beta) \tag{4.4.1}$$

where the subscript i is an index for the individual, the subscript j is an index for the alternative, $V_{ij}(\mathbf{x}_{ij}, \beta)$ is a deterministic or measured utility, and e_{ij} is a random error or un-deterministic utility, which captures unexplained characteristics of alternatives and/or individuals. More general scenario can be envisioned when the effect of attributes varies across the alternatives. For example, the effect of price increase on utility of new brand will be severe than that of the well established brand. In these circumstances one needs to incorporate different sets β for each alternatives. The ideal models will have the perfect definition of the functions $V_{ij}(\mathbf{x}_{ij}, \beta)$ and magnitudes of the error terms will be very small.

In practice the most often used multinomial discrete choice models is the deterministic utility, modelled as a linear function of choice characteristics or covariates, that is,

$$V_{ij}(\mathbf{x}_{ij}, \beta) = \mathbf{x}'_{ij}\beta. \tag{4.4.2}$$

The individual i chooses alternative j if and only if its perceived utility is greater than or equal to that of any other alternative in his choice set. Hence the event of an individual i opting for alternative j, $(Y_i = j)$, can be expressed using a random utility function as follows:$U_{ij}(\mathbf{x}_{ij}, \beta) = \max_{k \in C_i} U_{ik}(\mathbf{x}_{ik}, \beta)$

$$U_{ij}(\mathbf{x}_{ij}, \beta) = \max_{k \in C_i} U_{ik}(\mathbf{x}_{ik}, \beta)$$

where C_i is the choice set of subject i. Hence the probability that individual i chooses alternative j from his choice set C_i is

$$P(Y_i = j; \mathbf{x}_i, \beta) = P\left[U_{ij}(\mathbf{x}_{ij}, \beta) = \max_{k \in C_i} U_{ik}(\mathbf{x}_{ik}, \beta)\right]. \tag{4.4.3}$$

For simplicity of notations we assume that the choice set C is same for all the individual and consists of K alternatives. When the variability of unobserved utility $e_{ij}(\mathbf{x}_{ij}, \beta)$ is assumed to be very small compared to that of deterministic utilities across the alternatives the *rational models* of Manheim (1979) can be used. These models have very limited scope of applications as they assume that the error terms are all equal to zero.

IV.4.1 Multinomial Logit Models

McFadden (1973) proposed a special case of multivariate discrete choice models. His model is known as multinomial logit (MNL) model. Here the unmeasured utilities $e_{ij}(\mathbf{x}_{ij}, \beta)$ are assumed to be independently and identically distributed as Gumbel (also known as type I extreme value distribution) with zero mean and independent of β and $\mathbf{x}_{ij}$, that is, the cumulative distribution function of $e_{ij}(\mathbf{x}_{ij}, \beta)$ is

$$F(e_{ij}; \mathbf{x}_{ij}, \beta) = \exp\left[-e^{-e_{ij}/\theta_j}\right] \tag{4.4.4}$$

where $(\theta_1, \ldots \theta_K)$ are heteroscedastic parameters. It can easily be shown using (4.4.3) the above distribution reduces to

$$P(Y_i = j; \mathbf{x}_i, \beta) = \frac{\exp[V_{ij}(\mathbf{x}_{ij}, \beta)/\theta_j]}{\sum_{j=1}^{K} \exp[V_{ij}(\mathbf{x}_{ij}, \beta)/\theta_j]}. \tag{4.4.5}$$

This formulation of random utilities has its own advantages as well as disadvantages. The advantages being closed form formulation of choice probabilities leads to ease of interpretation and readily available maximum likelihood estimators. But the problematic aspect of the multinomial logit models lies in the so called *independence of the irrelevant alternatives* (IIA) property. That is the relative probability of choice between two alternatives is only affected by the utilities of corresponding choices because

$$\frac{P(Y_i = j; \mathbf{x}_i, \beta)}{P(Y_i = l; \mathbf{x}_i, \beta)} = \exp\left[\frac{V_{ij}(\mathbf{x}_{ij}, \beta)}{\theta_j} - \frac{V_{il}(\mathbf{x}_{il}, \beta)}{\theta_l}\right]. \tag{4.4.6}$$

In applications where the un-deterministic utilities associated with alternatives are correlated as they share some neglected attributes the use of multinomial logit model is appropriate as the IIA assumption is violated. Next we present multinomial probit model which alleviates such restrictions by incorporating the associations between error terms.

IV.4.2 Multinomial Probit Models

The multinomial probit models is a random utility model in which the error terms $e'_{ij}s$ are assumed to have multivariate normal distribution with zero mean and arbitrary

variance-covariance matrix $\Sigma = [\sigma_{ij}]$. Thus variances of un-observable or error terms can be different and error terms may be correlated. The attribute effect are assumed to be independent of alternatives. Using the notations of McFadden (1989) define

$$\begin{aligned}
\mathbf{Z}_{i_{-j}} &= (\mathbf{x}_{i1} - \mathbf{x}_{ij}, \ldots, \mathbf{x}_{ij-1} - \mathbf{x}_{ij}, \mathbf{x}_{ij+1} - \mathbf{x}_{ij}, \ldots, \mathbf{x}_{iK} - \mathbf{x}_{ij}), \\
\mathbf{U}^*_{i_{-j}} &= (U_{i1} - U_{ij}, \ldots, U_{ij-1} - U_{ij}, U_{ij+1} - U_{ij}, \ldots, U_{iK} - U_{ij}), \\
\mathbf{e}^*_{i_{-j}} &= (e_{i1} - e_{ij}, \ldots, e_{ij-1} - e_{ij}, e_{ij+1} - e_{ij}, \ldots, e_{iK} - e_{ij}).
\end{aligned}$$

Then $\mathbf{e}^*_{i_{-j}} \sim \mathbf{N}(\mathbf{0}, \Sigma^*)$. For model identifiability and to characterize Σ^*, one of the diagonal element σ_{ll} of Σ is assumed to be 1 along with assumption $\sigma_{jl} = \sigma_{lj} = 0; j = 1, \ldots, K$.

Using equation (4.4.1), we can check that $\mathbf{U}^*_{i_{-j}}$ has a multivariate normal with mean $\mathbf{Z}'_{i_{-j}}\beta$ and variance covariance matrix $\mathbf{Z}'_{i_{-j}}\Sigma\mathbf{Z}_{i_{-j}}$. The event that individual alternative j is utilized by individual i can be written as

$$\begin{aligned}
P(Y_i = j|\mathbf{x}_i, \beta, \Sigma) &= P[\mathbf{U}^*_{i_{-j}} \leq \mathbf{0}|\mathbf{Z}_{i_{-j}}, \beta, \Sigma] \\
&= \Phi^p[\mathbf{Z}_{i_{-j}}\beta; \Sigma^*]. \qquad (4.4.7)
\end{aligned}$$

The multivariate interval probabilities and maximum likelihood estimates of parameters can be obtained by using the techniques described before. A comprehensive review of the multinomial probit model can be found in Daganzo (1979) and McFadden (1989).

CHAPTER V

MARGINAL DISPERSION MODELS

In the multivariate mixture models studied in Chapter IV, the random variables could be subject to an extra-variability that is due to variation in the unobservable latent variables. Hence, no meaningful inference can be made on the causes and nature of the dispersion parameters. Also most multivariate mixture models are not applicable to under-dispersed measurements. In this chapter we will introduce generalized versions of ordinary binomial and Poisson distributions which account for over and under dispersions.

This chapter is organized as follows. In Section V.1, we present a brief introduction showing the need to accommodate over and under dispersion parameters in a meaningful way with the help of some examples. In Section V.2, we present generalizations of multinomial distributions which are variations of the urn models of Mishra et al. (1992) and Consul (1974). These generalizations are known as quasi-multinomial distributions. In Section V.3 we outline the properties of the quasi-multinomial distributions and discuss maximum likelihood estimation of the parameters. We illustrate the estimation procedure using optical scanner panel data of Paap and Franses (2000) in Section V.4. In Section V.5 we present generalized Lagrange family of distributions, which lead to a generalization of standard Poisson and negative binomial distributions. Generalized Lagrange distributions could be used to generate multivariate over dispersed Poisson vectors using the thinning operator defined on quasi-binomial distributions.

V.1 Examples

The observed binary variables in many practical experiments do not fit a binomial model and exhibit either much greater or smaller variation than the regular binomial models. This is more evident when the experiments involve human subjects. When the human subjects take some decisions at different times in their lives, the consequences have a high impact on the person in charge or on the group as a whole.

Feller (1968) asserted that "..the apparent aftereffect of sampling was at first misinterpreted as an effect of true contagion, and so statisticians now speak of contagion in a vague and misleading manner." In the following scenarios the over dispersion parameter cannot be treated as nuisance since it is an essential component of the population characteristics.

1. Systems monitoring: Consider studies aimed at establishing efficient monitoring methods of a complex system comprising of n components, such as power transmissions lines or computer networks consisting of multiple web servers. These components dynamically transit between several stages with certain probabilities. Each of these transition stages corresponds to a certain level of throughput to total system performance. Further, due to interconnectedness of system components the event of any component transiting from one stage to another affects the transition probabilities of others. Therefore the multivariate counts of observed frequencies of components in several stages cannot be viewed as sum of independent multinomial vectors.

2. The survey data analysis: In several clinical studies the survey questionnaire contains multiple questions that address identical attributes in the study. The recorded responses to these questions are summarized as univariate outcomes, for example, Pittsburgh sleep quality index (PSQI), SF-25 quality of life index, and many more. In many scenarios these indices can be viewed as sum of identical but not independent ordinal variables. As a result the ordinary Poisson distributions cannot be used for analysis without compromising the strength of inference. Note that attentiveness or quality of questionnaire design can be assessed via observed dispersion when scale variables are modelled accurately.

3. Reliability analysis: In most manufacturing applications the final product is assembly of various components and each component can be manufactured using one of many multiple designs. The product reliability scores are computed after the finished product undergoes several reliability tests, and therefore are subject to extra variability. Any parametric model must incorporate the over-dispersions.

4. Credit risk modelling: In the analysis of payment default statistics over a period of time, it was observed that the variance is a complex function of the mean. The liability of an individual defaulting on a payment is affected by various events.

As each individual tends to exhibit different spending habits, their financial strategies tend to change with previous experiences.

The above examples and many more motivate the need to search for models with over and under dispersions. In next section we present classes of multinomial distributions that accommodate post-sampling effects.

V.2 Quasi-Multinomial Distribution

Urn Model ($(p+1)$-Stage Game):

1. Suppose that we have $(p+1)$ urns, such that urn i contains θ_i chips labelled i for $1 \leq i \leq p$. The $(p+1)$th urn contains θ_i chips numbered i, and a total of $\boldsymbol{\theta}_{(\cdot)} = (\sum_{i=1}^{p} \theta_i)$ chips.

2. In stage 0 of the game, a player selects randomly a partition $\mathbf{Y} = \mathbf{y} = (y_1, y_2, \ldots, y_p)$ of m and adds for each i, $1 \leq i \leq p$, $\gamma\, y_i$ chips labelled number 0 to urn i, and adds the same number of chips labelled number i to urn $(p+1)$. Here γ is a positive integer.

3. In stage j of the game the player draws s_j chips with replacement from urn j. If any of the chips drawn is labelled 0 the player looses the game, otherwise he/she moves on to stage $(j+1)$.

4. In stage $(p+1)$, the player draws m chips from urn $(p+1)$.

5. The player is declared winner if the m chips drawn at stage $(p+1)$ matches exactly the pre-determined partition $\mathbf{y}$.

The probability of a player winning a game with a pre-selected partition $\mathbf{y}$ is given by

$$P[\,\text{Winning} \mid \mathbf{Y} = \mathbf{y}\,] \;=\; \left\{\prod_{j=1}^{p}\left(\frac{\theta_j}{\theta_j + \gamma\, y_j}\right)^{s_j}\right\}\left\{\frac{m!}{y_1!\ldots y_p!}\prod_{j=1}^{p}\left(\frac{\theta_j + \gamma\, y_j}{\boldsymbol{\theta}_{(\cdot)} + \gamma\, m}\right)^{y_j}\right\}$$

where $\boldsymbol{\theta}_{(\cdot)} = \sum_{j=1}^{p} \theta_j$. Since the distribution of $\mathbf{Y}$ is uniform, the conditional distribution of $\mathbf{Y}$ given that the player has won is

$$P[\,\mathbf{Y} = \mathbf{y} \mid Won\,] = \frac{1}{\mathbf{B}(m, \boldsymbol{\theta}, \mathbf{s}, \gamma)}\left\{\frac{m!}{y_1!\ldots y_p!}\prod_{j=1}^{p}\left(\frac{\theta_j + y_j\gamma}{\boldsymbol{\theta}_{(\cdot)} + m\gamma}\right)^{y_j - s_j}\right\} \qquad (5.2.1)$$

where the normalizing constant $\mathbf{B}(\mathbf{n}, \theta, \mathbf{s}, \gamma)$ is defined as

$$\mathbf{B}(m, \boldsymbol{\theta}, \mathbf{s}, \gamma) = \sum_{\mathbf{y}} \left\{ \frac{m!}{y_1! \ldots y_p!} \prod_{j=1}^{p} \left(\frac{\theta_j + y_j \gamma}{\boldsymbol{\theta}_{(\cdot)} + m\gamma} \right)^{y_j - s_j} \right\},$$

$\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)$ and $\mathbf{s} = (s_1, s_2, \ldots, s_p)$. The distribution (5.2.1) is known as the quasi-multinomial distribution. Note that equation (5.2.1) is a proper probability mass function even if $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$ or γ are not integers but they have to satisfy the constraints

$$\left\{ (\boldsymbol{\theta}, \gamma) : \boldsymbol{\theta} \in \mathbb{R}_+^p \text{ and } \gamma \geq \max_j(-\theta_j/m) \right\}.$$

The following re-parameterized version of the probability mass function (5.2.1) is convenient because it has fewer parameters.

$$P[\mathbf{Y} = \mathbf{y}] = \frac{1}{\mathbf{B}(m, \boldsymbol{\pi}, \mathbf{s}, \gamma^*)} \left\{ \frac{m!}{y_1! \ldots y_p!} \prod_{j=1}^{p} \left(\frac{\pi_j + y_j \gamma^*}{1 + m\gamma^*} \right)^{y_j - s_j} \right\} \tag{5.2.2}$$

where $\pi_j = \theta_j / \boldsymbol{\theta}_{(\cdot)}$ and $\gamma^* = \gamma / \boldsymbol{\theta}_{(\cdot)}$. Note that $\sum_{j=1}^{p} \pi_j = 1$.

Remarks:

1. The function $\mathbf{B}(\cdot)$ reduces to Abel's general class of sums when $p = 2$, see Riordan (1968).

2. The above distribution (5.2.2) is not closed under convolutions unless $\gamma^* = 0$.

3. The marginal variances of the quasi-multinomial distribution are greater (less) than that of the traditional multinomial distribution when $\gamma^* > 0$ (< 0). The quasi-multinomial distribution reduces to the traditional multinomial distribution with parameters m and $\boldsymbol{\theta}$ when $\mathbf{s} = 0$ and $\gamma^* = 0$.

Special cases:

1. When $p = 2$, $s_1 = 1$, $s_2 = 0$, then $B(m, \boldsymbol{\theta}, \mathbf{s}, \gamma) = \theta_1^{-1}(\boldsymbol{\theta}_{(\cdot)} + m\gamma)$ hence the probability mass function (5.2.2) reduces to

$$Pr(\mathbf{Y} = y) = \left(\frac{m!}{y!(m-y)!} \right) \frac{\theta_1(\theta_1 + y_1\gamma)^{y-1}(\theta_2 + (m-y)\gamma)^{m-y}}{(\boldsymbol{\theta}_{(\cdot)} + m\gamma)^m}. \tag{5.2.3}$$

Applying transformation $\pi = \theta_1/(\boldsymbol{\theta}_{(\cdot)} + m\gamma)$ and $\gamma^* = \gamma/(\boldsymbol{\theta}_{(\cdot)} + m\gamma)$, (5.2.5) simplifies to

$$Pr(\mathbf{Y} = y) = \binom{m}{y} \pi(\pi + y\gamma^*)^{y-1}(1 - \pi - y\gamma^*)^{m-y} \tag{5.2.4}$$

The above mass function was proposed by Consul (1974) and it is known as Type I quasi-binomial distribution.

2. When $p = 2, s_1 = 1, s_2 = 1$, the probability mass function (5.2.2) can be written as

$$Pr(\mathbf{Y} = y) = \binom{m}{y} \frac{\theta_1(\theta_1 + \gamma^* y)^{y-1}\theta_2(\theta_2 + (m-y)\gamma)^{m-y-1}}{\boldsymbol{\theta}_{(\cdot)}(\boldsymbol{\theta}_{(\cdot)} + m\gamma)^{m-1}}. \tag{5.2.5}$$

This is known as Type II quasi-binomial distribution, see Consul (1990).

3. Mishra et al. (1992) defined a class of quasi-binomial distributions of dimension $p = 2$ using an urn model scheme. The probability mass function is given by

$$Pr(\mathbf{Y} = y) = \binom{m}{y} \frac{(\theta_1 + \gamma^* y)^{y-s}(\theta_2 + (m-y)\gamma^*)^{m-y-t}}{B(m, \boldsymbol{\theta}, \mathbf{s}, \gamma^*)}. \tag{5.2.6}$$

Also for $d_j = 0, 1, 2, \ldots, m; \quad j = 1, 2, \ldots, (p-1)$ and $d_p = 0$ such that $\mathbf{d}_{(\cdot)} = \sum_{j=1}^{p} d_j \leq m$, the joint factorial moments of $\mathbf{Y}$ can be obtained using the recurrence relations

$$\begin{aligned} \boldsymbol{\mu}'_{(\mathbf{d})}(\mathbf{Y}) &= \mathrm{E}\left\{\prod_{j=1}^{p-1}\left[\frac{Y_j!}{(Y_j - d_j)!}\right]\right\} \\ &= \left\{\frac{m!}{(m - \mathbf{d}_{(\cdot)})!}\right\} \frac{B\left[m - \mathbf{d}_{(\cdot)}, (\boldsymbol{\theta} + \gamma\mathbf{d}), \gamma, (\mathbf{s} - \mathbf{d})\right]}{B(m, \boldsymbol{\theta}, \gamma, \mathbf{s})}. \end{aligned} \tag{5.2.7}$$

Unfortunately, equations (5.2.7) do not always lead to closed form solutions for mean and covariance of $\mathbf{Y}$ for arbitrary integer vector $\mathbf{s}$. The moment based methods fail to provide quick estimates of $\boldsymbol{\theta}$ and γ. But closed form representations do exists for special cases. For example when $\mathbf{s} = \mathbf{1}$, we have

$$B(m, \boldsymbol{\theta}, \gamma, \mathbf{1}) = \boldsymbol{\theta}_{(\cdot)}(\boldsymbol{\theta}_{(\cdot)} + m\gamma)^{m-1}\left\{\prod_{j=1}^{p}\theta_j\right\}^{-1}$$

and the corresponding probability mass function is known as the type II quasi-multinomial distribution or QMD-II. This is given by

$$P(\mathbf{y}) = \left\{ \frac{m_i!}{y_{i1}! \ldots y_p! \boldsymbol{\theta}_{(\cdot)}} \prod_{j=1}^{p} \theta_{ij}(\theta_{ij} + y_{ij}\gamma)^{y_{ij}-1} \right\}. \tag{5.2.8}$$

In next lemma we present the mean and variance covariance matrix for the type II quasi-multinomial distribution.

Lemma 5.1 *Let* $\mathbf{Y} \sim$ *QMD-II*$(m, \boldsymbol{\pi}, \gamma)$ *and* $\boldsymbol{\pi}$ *be a probability mass function. Then*

(1) $\boldsymbol{\mu} = \mathrm{E}(\mathbf{Y}) = m\boldsymbol{\pi}$

(2) $\Sigma = \mathrm{Cov}(\mathbf{Y}) = m\phi(m, \gamma)(D(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}')$

where $D(\boldsymbol{\pi}) = diag(\pi_1, \ldots, \pi_{p-1})$ *and*

$$\phi(m, \gamma) = m - (m-1)(1+m\gamma)^{-1} \left[\sum_{k=2}^{m} \frac{(m-2)!}{(m-k)!} \left(\frac{\gamma}{1+m\gamma} \right)^{k-2} \right].$$

Proof: The proof is by induction. Part (1) of the lemma holds trivially if $m = 1$, since Y_j's reduce to Bernoulli (π_j) random variables and $\phi = 1$. Next assume that the (1) holds for $m - 1$. That is for all $\boldsymbol{\pi} \in [0,1]^p$ and $j = 1, \ldots, p$, we have

$$\mu_j(m-1, \boldsymbol{\pi}, \gamma) = \mathrm{E}(Y_j; m-1, \boldsymbol{\theta}, \gamma) = m\pi_j. \tag{5.2.9}$$

As special case of equation (5.2.7) we have the recurrence relation

$$\begin{aligned} \mu_j(m, \boldsymbol{\theta}, \gamma) &= \mathrm{E}(Y_j; m, \boldsymbol{\pi}, \gamma) \\ &= \frac{m\pi_j(1+\gamma)}{(1+m\gamma)} \left[1 + \frac{\gamma}{\pi_j + \lambda} \mu_j(m-1, \boldsymbol{\pi} + \gamma\, e_j, \gamma) \right] \end{aligned} \tag{5.2.10}$$

It can be easily verified that proposition 1. hold for m by substituting identity (5.2.9). This completes proof of 1. Similarly the recurrence relation for second order factorial moment is of the form

$$\begin{aligned} \tau_j(m, \boldsymbol{\pi}, \lambda) &= \mathrm{E}(Y_j(Y_j - 1); m, \boldsymbol{\pi}, \gamma) \\ &= \frac{m(m-1)\pi_1}{(1+m\gamma)} \left[(\pi_j + 2\gamma) + \frac{\gamma(m-2)(\pi_j + 3\gamma)}{(1+m\gamma)} \right. \\ &\quad \left. + \frac{\gamma^2(1+2\gamma)}{(\pi_j + 2\gamma)(1+n\gamma)} \tau(m-2, \boldsymbol{\pi} + 2\gamma\, e_j, \lambda) \right] \end{aligned} \tag{5.2.11}$$

The solution of above recurrence relation is

$$\tau_j(m, \boldsymbol{\pi}, \gamma) = \pi_1 \left[\sum_{k=2}^{n} \frac{m!}{(m-k)!} \frac{\gamma^{k-2}(\pi_k + k\gamma)}{(1+n\gamma)^{k-1}} \right] \tag{5.2.12}$$

Therefore the diagonal elements of variance of covariance matrix Σ is given by

$$\begin{aligned}\sigma_{jj} &= \mathrm{Var}(Y_j) \\ &= E(Y_j(Y_j-1)) + E(Y_j)(1-E(Y_j)) \\ &= m\pi_j(1-\pi_j)\phi(m,\gamma) \end{aligned} \tag{5.2.13}$$

Finally the off-diagonals of Σ is obtained using conditional distributions.

$$\begin{aligned}\sigma_{jk} &= \mathrm{Cov}(Y_j, Y_k) \\ &= \mathrm{E}\left[Y_k(\mathrm{E}(Y_j|Y_k) - \mu_j)\right] \\ &= -m\pi_j\pi_j\phi(m,\gamma) \end{aligned} \tag{5.2.14}$$

This completes the proof of the lemma.

Lemma 5.1 shows that the moments of type-II quasi-multinomial distribution have resemble the moments of the traditional multinomial distribution. and thus they are subject to similar interpretations.

V.3 Maximum Likelihood Estimation

Suppose that we have a random sample $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ of observations from type-II quasi-multinomial distribution. The explanatory variables $\mathbf{X}_1, \ldots, \mathbf{X}_n$ can be incorporated in the model by the relation $\theta_{ij} = \exp(\mathbf{x}_{ij}\boldsymbol{\beta}_j)$. Note that this setting reduces to that of multinomial logit discrete choice model of Section IV.4.1 when $\gamma = 0$. The unknown parameters $\boldsymbol{\Theta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_p, \gamma)$ can be estimated by maximizing the log-likelihood function

$$\begin{aligned} L = \sum_{i=1}^{n} \Big\{ &\left[\log(m_i!) - (m_i - 1)\log(\boldsymbol{\theta}_{i(\cdot)} + m_i\gamma) - \log(\boldsymbol{\theta}_{i(\cdot)})\right] \\ &+ \sum_{j=1}^{p} \left[(y_{ij}-1)\log(\theta_{ij} + y_{ij}\gamma) + \log(\theta_{ij}) - \log(y_{ij}!)\right] \Big\} \end{aligned} \tag{5.3.1}$$

which amounts to solving the following estimating equations

$$\frac{\partial L}{\partial \boldsymbol{\beta}_j} = \sum_{i=1}^{n} \mathbf{x}'_{ij}\theta_{ij}\Big\{-(m_i-1)[\boldsymbol{\theta}_{i(\cdot)} + m_i\gamma]^{-1} - \boldsymbol{\theta}_{i(\cdot)}^{-1} + (y_{ij}-1)[\theta_{ij} + y_{ij}\gamma]^{-1} + \theta_{ij}^{-1}\Big\} = 0$$
$$j = 1, 2 \ldots, p; \tag{5.3.2}$$

$$\frac{\partial L}{\partial \gamma} = \sum_{i=1}^{n} \left\{ -m_i(m_i-1)[\boldsymbol{\theta}_{i(\cdot)} + m_i\gamma]^{-1} + \sum_{j=1}^{p} y_{ij}(y_{ij}-1)[\theta_{ij} + y_{ij}\gamma]^{-1} \right\} = 0 \tag{5.3.3}$$

The covariance matrix of maximum likelihood estimates $\widehat{\Theta}$ is $-H^{-1}(\widehat{\Theta})$ where $H(\Theta)$ is the Hessian matrix. The robust estimate of covariance matrix is obtained using Fisher information (or $\mathrm{E}[-H(\widehat{\Theta})]$). The following lemma provides the negative incomplete moments needed for computation of the Fisher information matrix.

Lemma 5.2 *Let* $\mathbf{Y} \sim QMII(m, \boldsymbol{\pi}, \gamma)$ *and* $\boldsymbol{\pi}$ *be a probability mass function then*

$$
\begin{aligned}
\mathrm{E}\left[\frac{Y_j!}{(Y_j-k)!}(\pi_j+Y_j\gamma)^{-k}\right] &= \left[\frac{m!}{(m-k)!}(\pi_j+m\gamma)^{-k}\right]\pi_j\left(\frac{\pi_j+k\gamma}{1+k\gamma}\right)^{-1} \\
\mathrm{E}\left[\frac{Y_j!}{(Y_j-(k+1))!}(\pi_j+Y_j\gamma)^{-k}\right] &= \left[\frac{m!}{(m-(k+1))!}(\pi_j+Y_j\gamma)^{-k}\right]\pi_j \\
\mathrm{E}[(Y_j-1)(\pi_j+Y_j\gamma)^{-1}] &= m(1+\gamma)(1+m\gamma)^{-1}-\pi_j^{-1} \\
\mathrm{E}[(\pi_j+Y_j\gamma)^{-1}] &= -m\gamma(1+\gamma)(\pi_j+\gamma)^{-1}(1+m\gamma)^{-1}+\pi_j^{-1} \\
\mathrm{E}[Y_j(Y_j-1))(\pi_j+Y_j\gamma)^{-1}] &= m(m-1)\pi_j(1+m\gamma)^{-1} \\
\mathrm{E}[(Y_j-1)(\pi_j+Y_j\gamma)^{-2}] &= m\pi_j^{-1}(1+\gamma)(1+m\gamma)^{-1}-\pi_j^{-2} \\
&\quad -m(m-1)\gamma(\pi_j+2\gamma)^{-1}(1+m\gamma)-2(1+2\gamma) \\
\mathrm{E}[Y_j(Y_j-1)(\pi_j+Y_j\gamma)^{-2}] &= m(m-1)\pi_j(\pi_j+2\gamma)^{-1}(1+2\gamma)(1+m\gamma)^{-2} \\
\mathrm{E}[Y_j^2(Y_j-1)(\pi_j+Y_j\gamma)^{-2}] &= m(m-1)\pi_j(1+m\gamma)^{-2} \\
&\quad \times\left[m-2+2(\pi_j+2\gamma)^{-1}(1+2\gamma)\right].
\end{aligned}
$$

The proof of the lemma is straightforward and it is omitted.

V.4 Panel Data Analysis

We illustrate application of the quasi-multinomial distribution using the optical scanner panel data on the purchases of saltine crackers by Rome (Georgia) households. The data set consist of 3292 brand choice decisions made by 136 households over a two year period. The study included three major national brands namely, Sunshine, Keebler and Nabisco. The local brands are collapsed into one category 'Private '. The information on actual price of brand selected and shelf price of competing brands and whether there was display and/or newspaper feature of brands under study at the time of purchase is also recorded. Jain et al. (1994) have analyzed the subset of this dataset using random effect multinomial logit model. Recently, Paap and Franses (2000) presented an alternative dynamic multinomial probit model which

incorporates the long-term and short-term effects of marketing variables differently, and assumes the multivariate autoregressive time series structure for error terms.

Since the household choices tend to exhibit certain level of long term brand loyalties, they undermine the short term effect of marketing variables. Also the non-measurable or error utilities arising from brand choices of single household cannot be assumed to be independent of each other. Figure 5.1 is a visual display of the number purchases made by households. It shows that most households have made 15-35 purchases, and one household has made 77 purchases in the two years of study. Maximum likelihood estimation for the dynamic probit model of Paap and Franses (2000) requires numerical computation of joint error utilities and it is computationally challenging and may not be reliable.

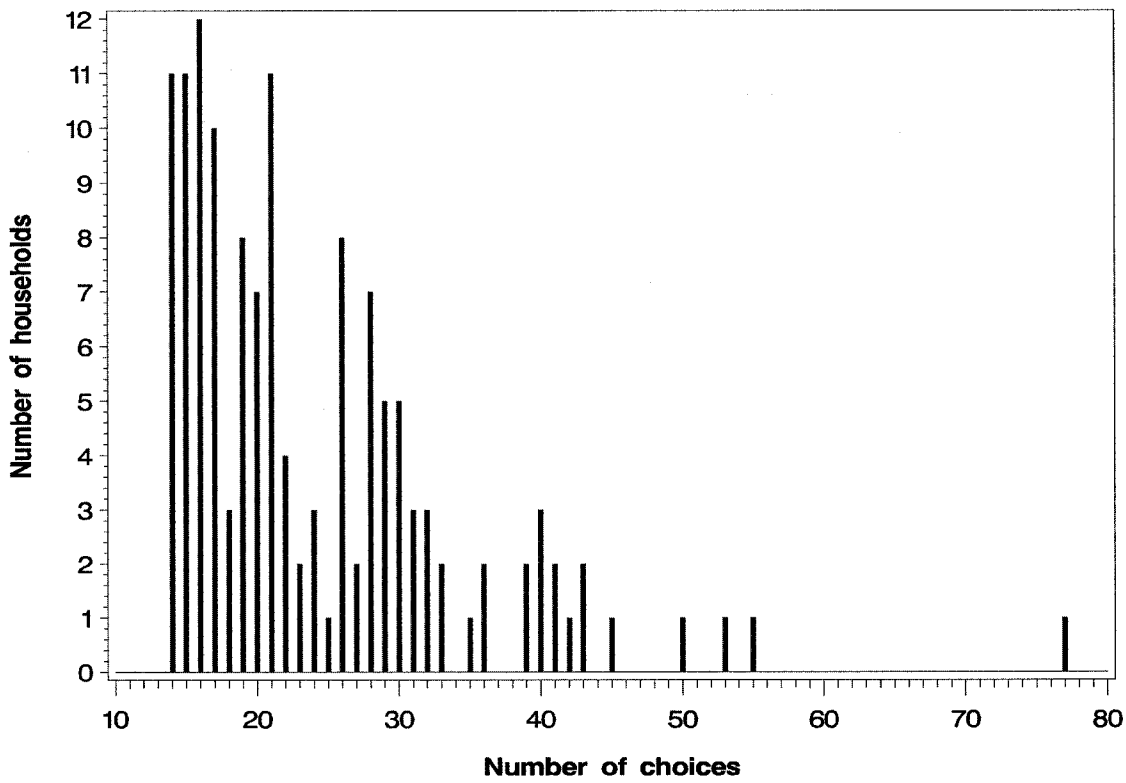

Figure 5.1: Cracker data: Household frequency plots.

We first artificially create a compact data set. For household i, this compressed dataset records purchase frequency of brand j (y_{ij}), along with the frequency of display (x_{ij_1}) and the frequency of feature (x_{ij_2}). The average shelf/purchase price (x_{ij_3}) is also recorded. Table 5.1 has the records of the first 10 households. Initial

analysis of Table 5.1 suggests that Nabisco brand ($j = 3$) is market leader in terms of total number of purchases and is more often on display than its competitors. The Keebler brand ($j = 2$) seems to be the costliest. Private brands ($j = 4$) are the cheap ones and hence account for the largest number of purchases. We can also notice that Nabisco brand is more consistently featured than other brands. As some of the brands are dominant and more powerful, it would be interesting to test whether the marketing strategies have distinct impact on utility of a brand.

Table 5.1: Cracker Data: Compressed Data of First 10 Household

i	j	y_{ij}	x_{ij_1}	x_{ij_2}	x_{ij_3}	i	j	y_{ij}	x_{ij_1}	x_{ij_2}	x_{ij_3}
1	1	2	3	0	1.0044	6	1	0	2	0	0.9782
	2	0	1	0	1.1125		2	0	1	0	1.1261
	3	14	7	0	1.1406		3	5	7	0	1.0875
	4	0	0	0	0.8106		4	23	0	0	0.5868
2	1	10	4	2	0.9313	7	1	1	1	0	1.0141
	2	2	3	1	1.2150		2	4	5	3	1.1706
	3	4	6	1	1.0719		3	12	6	1	0.9918
	4	0	1	1	0.6675		4	0	2	1	0.7000
3	1	0	1	0	0.9614	8	1	0	1	0	0.9715
	2	0	1	1	1.1243		2	0	1	1	1.1292
	3	7	2	0	1.0586		3	26	8	0	1.1173
	4	7	2	2	0.6600		4	0	3	0	0.7888
4	1	0	0	0	0.9841	9	1	13	5	1	0.9258
	2	0	1	0	1.1479		2	1	1	0	1.1695
	3	0	9	0	1.1155		3	4	5	2	1.1563
	4	29	0	0	0.5790		4	1	0	0	0.8058
5	1	0	1	0	0.9736	10	1	0	12	1	0.9298
	2	10	1	2	1.1079		2	0	0	0	1.1541
	3	4	3	2	1.1086		3	39	12	3	1.0771
	4	0	1	0	0.6936		4	2	3	1	0.7449

No short-term cause-effect analysis can be performed using this artificial data due to loss of dynamic information of actual explanatory variables. However, elementary study of long-term impact of marketing variables and brand utilities and consumer loyalty can be quickly performed using the quasi-multinomial distributions.

We present four different versions of the multinomial-logit brand choice models to analyze compressed data by expressing the distributional parameter θ as function of the observed utilites, that is, $\theta_{ij} = \exp(\mathbf{x}'_{ij}\boldsymbol{\beta}_j)$.

Table 5.2: Cracker Data: Maximum Likelihood Estimates

Brand dependent effects						
	QMD-II			Multinomial		
	Est.	SE.	σ	Est.	SE	σ
β_1	0.0196	0.0450	0.0429	-0.1142	0.0255	0.0120
	-0.0077	0.1270	0.1367	0.3519	0.0588	0.0631
	-3.1810	0.2112	0.2113	-5.9188	0.1234	0.1032
β_2	0.0406	0.0747	0.0676	-0.1086	0.0455	0.0381
	0.1148	0.0869	0.0970	0.0865	0.0641	0.0656
	-2.9866	0.2150	0.2015	-4.9092	0.1011	0.0938
β_3	-0.0568	0.0301	0.0287	-0.0984	0.0111	0.0091
	0.3074	0.0635	0.0628	0.3017	0.0243	0.0229
	-1.6253	0.2219	0.2134	-3.1419	0.0873	0.0737
β_4	-0.0015	0.0745	0.0729	0.0425	0.0192	0.0225
	0.0556	0.0933	0.0873	-0.0713	0.0205	0.0258
	-3.8285	0.2825	0.2879	-6.2001	0.0895	0.0974
γ	0.0890	0.0107	0.0106	-	-	-
L	-744.0209			-2127.889		

Brand independent effects						
	QMD-II			Multinomial		
	Est.	SE.	σ	Est.	SE	σ
β	0.0680	0.0187	0.0168	0.0927	0.0079	0.0054
	0.1541	0.0400	0.0387	0.2446	0.0125	0.0131
	-0.4205	0.1200	0.1154	-1.6526	0.0466	0.0359
γ	1.4852	0.1713	0.1720	-	-	-
L	-785.7753			-2893.453		

The first model allows brand dependent covariate effects and we assume the purchase counts ($\mathbf{Y}_i$) of household i, are distributed as $QMII(m_i, \boldsymbol{\theta}_i, \gamma)$. The second model also allows brand dependent effect but assumes traditional multinomial distribution for the counts.

The third and fourth models assume the covariate effects are brand independent ($\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta$). The third model assumes type II quasi-multinomial distribution for the purchase counts, whereas model four uses traditional multinomial distribution. Estimates of the unknown parameters for each model are computed using maximum-likelihood. Table 5.2 contains parameter estimates and their standard errors computed using observed and expected Fisher information matrices.

Maximum likelihood estimates for the first and second model suggest that the

utilities of 'Private' brands are most sensitive to price variation, whereas the adverse effects of price increase on utilities of Nabisco cracker are minimal. The log-likelihood statistics indicate that both models 2 and 4 are adequate. Also the first model indicates that effect of display is statistically insignificant for Private brands. The positive estimate of γ can be attributed to brand loyalty. The question whether the effect of marketing variables is brand dependent can be addressed by testing the hypothesis $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta$ using log-likelihood estimates of model 1 and 3. The chi-squared value of likelihood ratio test is 83.5088 with p-value less than 0.00001. Hence, we conclude that the first model is the most adequate model for analysis of cracker data.

In the next section we introduce families of generalized Lagrange distributions (GLPD). These families include distributions that can be considered as extensions of standard Poisson and negative binomial distributions.

V.5 Lagrange Distributions

Definition 5.1 *Let* $\mathbf{D}$ *be an open set in* $\mathbb{R}$ *and* $\mathbf{C}$ *be a subset of* $\mathbb{R}$. *A complex function* $G : \mathbf{D} \to \mathbf{C}$ *is said to be analytic in* $\mathbf{D}$ *if for each point* $t_0 \in \mathbf{D}$ *there exist an open subset* $\mathbf{D}^*$ *of* $\mathbf{D}$ *such that for each point* $t \in \mathbf{D}^*$, $G(t)$ *can be expressed as a power series in* $(t - t_0)$.

The following theorem, known as the Lagrange inversion formula, is the basis for the generalized Lagrange probability distributions. See Dieudonné (1971).

Theorem 5.1 *Let* G *be any analytic function in* $\mathbf{D} = [-r, r]$; $(r \geq 0)$ *such that* $M = \sup_{|t| \leq r} |G(t)|$ *and* $\mathbf{C} = (-r/M, r/M)$. *Then for each point* $u \in \mathbf{C}$ *there exist a unique analytic function* $G^* : \mathbf{C} \to (-r, r)$ *such that* $t = G^*(u)$ *is a solution of the equation*

$$t - u\, G(t) = 0. \tag{5.5.1}$$

Furthermore, for each analytic function H *in* $\mathbf{D}$ *and for all* $u \in \mathbf{C}$ *we have*

$$H(G^*(u)) = H(0) + \sum_{k=1}^{\infty} \frac{u^k}{k!} \left[\frac{\partial^{k-1}}{\partial t^{k-1}} \left(\frac{\partial H(t)}{\partial t} (G(t))^k \right) \right]_{t=0}. \tag{5.5.2}$$

Generalized Lagrange families of distributions are defined using the well known fact that the probability generating functions of non-negative integer valued discrete random variables are analytic by definition and their convolutions are yield probability generating functions.

Therefore, for given probability generating functions G and H, $(H \odot G^*)(u) = H(G^*(u))$ is also a probability generating function of random variable Y with probability mass function

$$P(Y=k) = \frac{1}{k!}\left[\frac{\partial^{k-1}}{\partial t^{k-1}}\left(\frac{\partial H(t)}{\partial t}(G(t))^k\right)\right]_{t=0} \qquad k = 0, 1, \ldots \tag{5.5.3}$$

where $G^*(u) = u\, G(G^*(u))$.

Jain and Consul (1971) defined generalized negative binomial (GNB) distribution as a special case of a Lagrange distribution by compounding binomial random variables. They have shown that the negative binomial and binomial distributions are special cases of GNB. Consul and Jain (1973) presented the following generalization of Poisson distributions.

Definition 5.2 *The count random variable Y is said to be distributed as generalized poisson distribution denoted as $GP(\lambda, \gamma)$, if the probability mass function is given by*

$$P(Y=j;\theta,\lambda) = \frac{\lambda(\lambda+j\gamma)^{j-1}\exp(-(\lambda+j\gamma))}{j!} \qquad j = 0, 1, 2, \ldots \tag{5.5.4}$$

The generalized Poisson distributions possess some of the most desired properties such as multiplicative additivity, unimodality and log-concavity and closed form representations of moments. A generalized Poisson variate ($\mathrm{GP}(\lambda, \gamma)$) can also be viewed as sum of N independent and identically distributed variables, where N is a Poisson random variable. This characterization makes generalized Poisson distribution an ideal candidate for modelling various queueing theory and actuarial science processes.

Theorem 5.2 *Let X be a generalized Poisson random variable with parameters (θ, λ) , $0 \leq \alpha \leq 1$. Let Y be such that the conditional distribution of Y given $X = n$ is type II Quasi-binomial with parameters $(n, \alpha\theta, (1-\alpha)\theta, \lambda)$. Then the unconditional distribution of Y is GP $(\alpha\theta, \lambda)$.*

Devroye (1989) has presented a modification of the acceptance-rejection algorithm for generating random variables from generalized Poisson distributions. This break-through was followed by Devroye (1992), wherein the branching method and uniform bound methods were tailored to generate random variables from several classes of Lagrange distributions. Using the derivations of higher moments of generalized distributions described in Consul and Shenton (1972) and generalized thinning operators based on Theorem 5.2, it is possible to replicate the asymptotic analysis of Chapter III for over-dispersed count data. We could also construct multivariate generalized Poisson distributions using the copulas described in Chapter IV. These will be pursued elsewhere.

CHAPTER VI

SUMMARY

In this thesis we presented an in-depth study of the statistical analysis of longitudinal and clustered data. Traditionally, the multivariate Gaussian distribution has been adequate when the data is continuous. But in recent years numerous applications have arisen where the longitudinal or clustered data are discrete. Unfortunately, multivariate discrete distributions are much more complex than the multivariate Gaussian distribution. Furthermore, many of these discrete distributions are intractable in the sense that implementing maximum likelihood estimation could be difficult and computationally very intensive. As an ad hoc solution several moment based methods, based on the optimal theory of estimating equations, were suggested. In this thesis we have studied these methods including the Gaussian, modified Gaussian (MG) estimation, and quasi-least squares (QLS) estimation procedures. We derived the asymptotic distributions of the estimates, and compared relative performance of these methods using simulations, primarily for correlated Poisson count data.

Although moment based methods are useful for a quick analysis, they are not optimal and as efficient as the likelihood based methods. In this thesis we have studied fully specified discrete distributions constructed using copulas. These distributions include models based on normal copulas, probit models, mixture models, probit-normal models, Poisson log-normal models, and discrete choice models, in particular multinominal logit and probit models. Calculation of maximum likelihood estimates and the Fisher information for these models requires computation of the multivariate normal probabilities, and we discussed several efficient algorithms for those computations. For the multivariate probit and probit-normal models, we have derived expressions for software implementation of the maximum likelihood estimates.

For modelling data that is subject to over or under dispersions, we have introduced the quasi-multinomial and generalized Lagrange families of distributions. These distributions can be used to draw meaningful inferences on the strength of population dynamics. We have implemented maximum likelihood for the quasi-multinomial distribution and tested it on a real life data dealing with multiple purchase decisions and marketing predictors of some products.

REFERENCES

Aitchison, J., and Ho, C. H. (1989), "The Multivariate Poisson-log Normal Distribution," *Biometrika*, 76, 643–653.

Bahadur, R. R. (1961), *A Representation of the Joint Distribution of Responses to n Dichotomous Items*, Stanford, CA: Stanford University Press, in Studies in Item Analysis and Prediction (Stanford Mathematical Studies in the Social Sciences VI).

Barlow, R. E., and Proschan, F. (1975), *Statistical Theory of Reliability and Life Testing: Probability Models*, New York: Holt, Rinehart and Winston, Inc.

Breslow, N. E., and Clayton, D. G. (1993), "Approximate Inference in Generalized Linear Mixed Models," *Journal of the American Statistical Association*, 88, 125–134.

Campbell, J. T. (1934), "The Poisson Correlation Function," *Proceedings of the Edinburg Mathematical Society*, 18–26.

Carey, V., Zeger, S. L., and Diggle, P. (1993), "Modelling Multivariate Binary Data With Alternating Logistic Regressions," *Biometrika*, 80, 517–26.

Chaganty, N. R. (1997), "An Alternative Approach to the Analysis of Longitudianl Data Via Generalized Estimating Equations," *Journal of Statistical Planning and Inference*, 63, 39–54.

Chaganty, N. R., and Naik, D. N. (2002), "Analysis of Multivariate Longitudinal Data Using Quasi-least Squares," *Journal of Statistical Planning and Inference*, 103, 421–436.

Chaganty, N. R., and Shults, J. (1999), "On Eliminating the Asymptotic Bias in the Quasi-least Squares Estimate of the Correlation Parameter," *Journal of Statistical Planning and Inference*, 76, 145–161.

Chib, S., and Greenberg, E. (1996), "Bayesian Analysis of Multivariate Probit Models," Econometrics 9608002, Economics Working Paper Archive at WUSTL, available at http://ideas.repec.org/p/wpa/wuwpem/9608002.html.

Chib, S., Greenberg, E., and Winkelmann, R. (1998), "Posterior Simulation and Bayes Factors in Panel Count Data Models," *Journal of Econometrics*, 86, 33–54.

Consul, P. C. (1974), "A Simple Urn Model Dependent Upon Predetermined Strategy," *Sankhyá: The Indian Journal of Statistics*, 36, Series B., 391–399.

— (1990), "On Some Properties and Applications of Quasi-binomial Distributions," *Communications in Statistics-Theory Methods*, 19, 474–504.

Consul, P. C., and Jain, G. C. (1973), "A Generalization of the Poisson Distributions," *Technometrics*, 15, 791–799.

Consul, P. C., and Shenton, L. R. (1972), "Use of Lagrange Expansion For Generating Discrete Generalized Probability Distributions," *SIAM: Journal of Applied Mathematics*, 23, 239–248.

Cranley, R., and Patterson, T. (1976), "Randomization of Number Theoretic Methods for Multiple Integration," *SIAM Journal on Numerical Analysis*, 13, 909–914.

Crowder, M. (1995), "On the Use of Working Correlation Matrix in Using Generalised Linear Models for Repeated Measures," *Biometrika*, 82, 407–410.

— (2001), "On Repeated Measures Analysis With Misspecified Covariance Structure," *Journal of Royal Statistical Society*, 63, 55–62.

Daganzo, C. (1979), *Multinomial Probit: The Theory and Its Application to Demand Forecasting*, New York, NY: Academic Press, Inc.

Devroye, L. (1989), "Random Variate Generators for Poisson-Poisson and Related Distributions," *Computational Statistics & Data Analysis*, 8, 247–278.

— (1992), "The Branching Process Method in Lagrange Random Variate Generation," *Communications in Statistics-Simulation*, 21, 1–14.

Dieudonné, J. (1971), *Infintesimal Calculus*, Boston: Houghton Mills.

Diggle, P. J., Heagerty, P., Liang, K.-Y., and Zeger, S. L. (2002), *Analysis of Longitudianal Data*, Oxford: Oxford University Press Inc., 2nd ed.

Emrich, L. J., and Piedmonte, M. R. (1991), "A Method for Generating High-Dimensional Multivariate Binary Variables," *The American Statistician*, 45, 302–204.

Feller, W. (1968), *An Introduction to Probability Theory and its Applications*, New York: John Wiley & Inc.

Fitzmaurice, G. M., and Laird, N. M. (1993), "A Likelihood-based Method for Analysing Longitudinal Binary Responses," *Biometrika*, 80, 141–51.

Gange, S. J. (1995), "Generating Multivariate Categorical Variables Using the Iterative Proportional Fitting Algorithm," *The American Statistician*, 49, 134–138.

Genz, A. (1992), "Numerical Computation of the Multivariate Normal Probabilities," *Journal of Compuational and Graphical Statistics*, 1, 141–150.

Gilmour, A. R., Anderson, R. D., and Rae, A. L. (1985), "The Analysis of Binomial Data by a Generalized Linear Mixed Model," *Biometrika*, 72, 593–599.

Godambe, V. P. (1960), "An Optimum Property of Regular Maximum Likelihood Estimation," *Annals of Mathematical Statistics*, 1208–1212.

Godambe, V. P. (ed.) (1991), *Estimating Functions*, Oxford: Oxford University Press.

Hajivassiliou, V., McFadden, D., and Rudd, P. (1996), "Simulation of Multivariate Normal Rectange Probabilities and Their Derivatives Theoritical and Computational Results," *Journal of Econometrics*, 72, 85–134.

Hall, D. B. (2001), "On the Application of Extended Quasi-likelihood to the Clustered Data Case," *Candadian Journal of Statistics*, 29, 77–97.

Hall, D. B., and Severini, T. A. (1998), "Extended Generalized Estimating Equations for Clustered Data," *Journal of the American Statistical Association*, 93, 1365–1375.

Jain, D. C., Vilcassim, N. J., and Chintagunta, P. K. (1994), "A Random-Effects Logit Brand-choice Model Applied to Panel Data," *Journal of Business & Economic Statistics*, 12, 317–328.

Jain, G. C., and Consul, P. C. (1971), "A Generalized Negative Binomial Distributions," *SIAM: Journal of Applied Mathematics*, 21, 501–513.

Joe, H. (1997), *Multivariate Models and Dependence Concepts*, New York: Chapman and Hall.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1997), *Discrete Multivariate Distributions*, New York: John Wiley & Sons, Inc.

Keast, P. (1973), "Optimal Parameters for Multidimensional Intergration," *SIAM Journal on Numerical Analysis*, 10, 831–838.

Kocherlakota, S., and Kocherlakota, K. (2001), "Regression in the Bivariate Poisson Distribution," *Communications in Statistics: Theory and Methods*, 30, 815–825.

Korn, E. L., and Whittmore, A. S. (1979), "Methods for Analyzing Panel Studies of Acute Health Effects of Air Pollution," *Biometrics*, 35, 795–802.

Laird, N. M., and Ware, J. H. (1982), "Random-effects Models for Longitudinal Data," *Biometrics*, 38, 963–974.

Lee, A. J. (1993), "Generating Random Binary Deviates Having Fixed Marginal Distribution and Specified Degrees of Association," *The American Statistician*, 47, 209–215.

Lee, L.-F. (2001), "On the Range of Correlation Coefficients of Bivariate Ordered Discrete Random Variables," *Econometric Theory*, 17, 247–256.

Lesaffre, E., and Kaufmann, H. (1992), "Existence and Uniqueness of the Maximum Likelihood Estimator for Multivariate Probit Model," *Journal of the American Statistical Association*, 87, 805–811.

Liang, K.-Y., and Zeger, S. L. (1986), "Longitudinal Data Analysis Using Generalized Linear Models," *Biometrika*, 73, 13–22.

Lin, X., and Breslow, N. E. (1996), "Bias Correction in Generalized Linear Mixed Models With Multiple Components of Dispersion," *Journal of the American Statistical Association*, 96, 1007–1016.

Lunn, A. D., and Davies, S. J. (1998), "A Note On Generating Correlated Binary Variables," *Biometrika*, 85, 487–490.

Manheim, M. (1979), *Fundamentals of Transportation System Analysis*, Cambrige, MA: MIT Press.

McCullagh, P., and Nelder, J. A. (1989), *Generalized Linear Models*, New York: Chapman and Hall, 2nd ed.

McFadden, D. (1973), *Conditional Logit Analysis of Qualitative Choice Behavior*, New York, NY: Academic Press, (P. Zarembka, ed.).

— (1989), "A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Intergration," *Econometrica*, 57, 995–1026.

McKendrick, A. G. (1916), "Applications of the Kinetic Theory of Gases to Vital Phenomena," *Indian Journal of Medical Research*, 3, 667–687.

Mishra, A., Tiwary, D., and Singh, S. K. (1992), "A Class of Quasi-binomial Distributions," *Sankhyá: The Indian Journal of Statistics*, 54, 67–76.

Nelder, J. A., and Wedderburn, R. (1972), "Generalized Linear Models," *Journal of Royal Statistical Society A*, 135, 370–384.

Paap, R. and Franses, P. H. (2000), "A Dynamic Multinomial Probit Model for Brand Choice With Different Long-run and Short-run Effects of Marketing-mix Variables," *Journal of Applied Econometrics*, 15, 717–744.

Park, C. G., Park, T., and Shin, D. W. (1996), "A Simple Method for Generating Correlated Binary Variables," *The American Statistician*, 50, 306–310.

Plackett, R. L. (1954), "A Reduction Formula for Normal Multivariate Integrals," *Biometrika*, 41, 351–360.

Preisser, J. S., Galecki, A. T., Lohman, K. K., and Wagenknecht, L. E. (2000), "Analysis of Smoking Trends with Incomplete Longitudinal Binary Responses," *Journal of the American Statistical Association*, 95, 1021–1031.

Preisser, J. S., and Qaqish, B. F. (1999), "Robust Regression for Clustered Data with Applcation to Binary Responses," *Biometrics*, 55, 574–579.

Prentice, R. L. (1988), "Correlated Binary Regression With Covariates Specific to Each Binary Observation," *Biometrics*, 44, 1033–1048.

Prentice, R. L., and Zhao, L. P. (1991), "Estimating Equations for Parameters in the Means and Covariances of Multivariate Discrete and Continuous Responses," *Biometrika*, 47, 825–839.

Qaqish, B. F. (2003), "A Family of Multivariate Binary Distributions for Simulating Correlated Binary Variables With Specified Marginal Means and Correlations," *Biometrika*, 90, 455–463.

Qu, A., Lindsay, B. G., and Li, B. (2000), "Improvised Generalised Estimating Equations Using Quadratic Inference Functions," *Biometrika*, 87, 823–836.

Riordan, J. (1968), *Combinatorial Identities*, New York: John Wiley & Sons, Inc.

Robins, A., Rotnitzky, J. M., and Zhao, L. P. (1995), "Analysis of Semi-parametric Regression Models for Repeated Outcomes in the Presence of Missing Data," *Journal of the American Statistical Association*, 90, 1321–1339.

Serfling, R. J. (1980), *Approximation Theorems of Mathematical Statistics*, New York: John Wiley.

Shults, J., and Chaganty, N. R. (1998), "Analysis of Serially Correlated Data Using Quasi-least Squares," *Biometrics*, 54, 1622–1630.

Sim, C. H. (1993), "Generation of Poisson and Gamma Random Vectors With Given Marginals and Covariance Matrix," *Journal of Statistical Computation and Simulation*, 47, 1–10.

Sklar, A. (1959), "Fonctions de Répartitions á n Dimensions et Leurs Marges," *Public Institute of the University of Paris*, 8, 229–231.

Sommer, A., Katz, J., and Tarwotjo, I. (1984), "Increased Risk of Respiratory Infection and Diarrhea in Chidren With Pre-existing Mild Vitamin A Deficiency," *American Journal of Clinical Nutrition*, 40, 1090–95.

Thall, P. F., and Vail, S. C. (1990), "Some Covariance Models for Longitudinal Count Data with Overdispersion," *Biometrics*, 46, 657–671.

Tsionas, E. G. (2001), "Baysian Multivariate Poisson Regression," *Communications in Statistics-Theory Methods*, 30, 243–255.

Vonesh, E. F., Wang, H., Nie, L., and Majumdar, D. (2002), "Conditional Second-Order Generalized Estimating Equations for Generalized Linear and Nonlinear Mixed-Effexts Models," *Journal of the American Statistical Association*, 97, 271–283.

Wang, Y.-G., and Carey, V. (2003), "Working Correlation Structure Misspecification, Estimation and Covariate Design: Implications for Generalised Estimating Equatins Performance," *Biometrika*, 90, 29–41.

Ware, J. H., Dockery, D. W., Spiro, A. I., Speizer, F. E., and Ferris, B. G. J. (1984), "Passive Smoking, Gas Cooking, and Respiratory Health of Children Living in Six Cities," *American Review of Respiratory Diseases*, 129, 366–374.

Ware, J. H., Lipsitz, S., and Speizer, F. E. (1988), "Issues in the Analysis of Repeated Categorical Outcomes," *Statistics in Medicine*, 7, 95–107.

Wicksell, S. D. (1916), "Some Theorems in the Theory of Probablility, With Special Reference to Their Importance in the Theory of Homograde Correlation," *Svenska Aktaurie-föreningens Tidskrift*, 165–213.

Yuan, K.-H., and Jennrich, R. I. (1998), "Asymptotics of Estimating Equations under Natural Conditions," *Journal of Multivariate Analysis*, 65, 245–260.

Zeger, S. L., and Liang, K. (1992), "An Overview of Methods for the Analysis of Longitudinal Data," *Statistics in Medicine*, 11, 1825–1839.

Zeger, S. L., and Qaqish, B. (1988), "Markov Regression Models for Time Series: A Quasi-likelihood Approach," *Biometrics*, 44, 1019–1031.

Zhao, L. P., and Prentice, R. L. (1990), "Correlated Binary Regression Using Quadratic Exponential Model," *Biometrika*, 77, 642–648.

APPENDIX

THE COLLECTION OF SAS PROGRAMS

A.1 Subroutines for Multivariate Poisson Simulations

```
/*-------------------------------------------------------------*/
/* The following subroutine simulates nsims multivariate Poisson */
/* obs from given covariance matrix Sigma using Sim's algorithm */
/*-------------------------------------------------------------*/

START SIMPOI(seed, Sigma, nsims);
   RUN Decompose(Sigma, alpha, lambda, Error, m);
   if (Error < 0) then do;
      print "Simulations Failure";
      return(Error);
   end;

   Z = J(m, nsims, 0);
   do k = 1 to nsims;
      X = J(m, 1, 0);
      do j = 1 to m;
         do i = 1 to j-1;
            if(X[i] & (alpha[j,i] > 0)) then
               Z[j,k] = Z[j,k] +
                  RANBIN(seed, X[i], alpha[j,i]);
         end;
         X[j] = RANPOI(seed, lambda[j]);
         Z[j,k] = Z[j,k] + X[j];
      end;
   end;
   return(Z);
Finish SIMPOI;
```

```
Start Decompose(Sigma, alpha, lambda, Error, m);
   m = nrow(Sigma); alpha = I(m); lambda = J(m, 1, 0); Error=1;
   lambda[1] = Sigma[1,1];
   do j = 2 to m;
      alpha[j,1] = Sigma[1,j]/lambda[1];
      if((0 > alpha[j,1]) | (alpha[j,1] > 1)) then Error = -1;
      do i = 2 to (j-1);
         do k = 1 to (i-1);
            alpha[j,i] = alpha[j,i] +
               alpha[i,k]*alpha[j,k]*lambda[k];
         end;
         alpha[j,i] = (Sigma[i,j] - alpha[j,i])/lambda[i];
         if((0 > alpha[j,i]) | (alpha[j,i] > 1)) then Error = -1;
      end;
      do k = 1 to (j-1);
         lambda[j] = lambda[j] + alpha[j,k]*lambda[k];
      end;
      lambda[j] = Sigma[j,j] - lambda[j];
      if(lambda[j] <= 0) then Error = -2;
   end;
Finish Decompose;


/*-----------------------------------------------------------------*/
/* This subroutine computes first four central moments of Poisson */
/* random variables.The functional arguments are alpha and lambda */
/*-----------------------------------------------------------------*/
Start Moments(alpha, lambda);
   m = nrow(lambda); dim = m+m*(m+1)/2; V = J(dim, dim, .);
   /* Second order moments */
   do i = 1 to m;
      do j = 1 to i;
         value = alpha[Unique(i||j),]; value =value[#,]*lambda;
         V[i,j] = value; V[j,i] = value;
      end;
```

```
   end;
   /* Third order moments */
   do i = 1 to m;
      do j = 1 to i;
         index1 = m + i*(i-1)/2+j;
         do k = 1 to m;
            value = alpha[Unique(i||j||k),];
            value = value[#,]*lambda;
            V[index1, k] = value; V[k, index1] = value;
         end;
      end;
   end;
   /* Fourth order moments */
   do i = 1 to m;
      do j = 1 to i;
         index1 = m + i*(i-1)/2+j;
         do k = 1 to m;
            do l = 1 to k;
               index2 = m + k*(k-1)/2+l;
               if (index1 >= index2) then do;
                  value = alpha[Unique(i||j||k||l),];
                  value = value[#,]*lambda;
                  V[index1,index2] = value + V[i,k]*V[j,l]
                                   + V[i,l]*V[j,k];
                  V[index2,index1] = V[index1,index2];
               end;
            end;
         end;
      end;
   end;
   return(V);
Finish Moments;
```

A.2 To Generate Fréchet Bounds Plots for Poisson Variables

```
/*-------------------------------------------------------------
   Let Y1 and Y2 be a random variables with marginal Poisson
   distributions with means lambda1 and lambda2. This program
   computes the correlation coefficient when joint distribution
   of (Y1,Y2) is Frechet upper and lower bounds copula.
   Finally, this program produces Figures 4.1 and 4.2 of thesis.
-------------------------------------------------------------*/
%let Outdir=%str(C:\Deepak\Paper 4);
%let xmin=8.0;
%let xmax=8.0;
%let xstepsize = 2;
%let dmin =-10.0;
%let dmax = 10.0;
%let dstepsize = 0.05;
%let beta = %sysfunc(log(2));
%let epsilon=0.1e-9;
%let xpoints = %sysevalf(%sysfunc
            (abs(%sysevalf((&xmin - &xmax) / &xstepsize) -1)));
libname OutLib  "&Outdir";

data OutLib.FrechetPoisson;
    do x1 = &xmin to &xmax by &xstepsize;
        lambda1 = exp(x1*&beta);
        do d = &dmin to &dmax by &dstepsize;
            x2 = d+x1;
            logratio = (x2-x1)*&beta;
            lambda2 = exp(x2*&beta);
            CovU = 0; CovL = 0;
            y1 = 0; S1 = 1; logf1 = -lambda1;
            do while(S1 > &epsilon);
                S1 = S1 - exp(logf1);
                y2 = 0; S2 = 1; logf2 = -lambda2;
```

```
                do while(S2 > &epsilon);
                    S2 = S2 - exp(logf2);
                    CovU = CovU + min((1-S1)*S2,(1-S2)*S1);
                    CovL = CovL - min((1-S2)*(1-S1),S1*S2);
                    y2 = y2+1;
                    logf2 = logf2 + log(lambda2/y2);
                end;
                y1 = y1+1;
                logf1 = logf1 + log(lambda1/y1);
            end;
            CorrU = CovU/(sqrt(lambda1*lambda2));
            CorrL = CovL/(sqrt(lambda1*lambda2));
            output;
        end;
    end;
    drop  CovU  CovL logf1 logf2;
run;

data frechet;
    set OutLib.FrechetPoisson;
    format lambda1 best6.5 lambda2 best6.5;
run;

goptions  reset=all   gaccess=gsasfile   gunit=pct  htitle=6
htext=3 vorigin=0in   horigin=0in  ftext=swiss   ftitle=swissb
colors=(blue black red green yellow cyan violet pink brown orange
cyan) cback=white hsize=7.5in  vsize=6in device=pslepsfc autofeed;

%macro defineSymbols(points);
%do j = 1 %to &points;
    symbol&j interpol=none  value=dot;
%end;
%mend defineSymbols;
```

```
%defineSymbols(&xpoints);

axis1 color = black  order=(-7 to 7 by 2)
  label = (color=black h=3 font=swissb "log" font=cgreek "(l2/l1)");
axis2 color = black order=(0 to 1 by 0.25)
      label = (a=90 h=3 font=cgreek "a");
legend1  label=(h=3 color=black font=cgreek "l1")
         value=(h=3 color=black font=swissb);
filename gsasfile "&Outdir\upper.ps";
proc gplot data=frechet;
    plot CorrU*logratio=lambda1 / haxis=axis1 vaxis=axis2
                                  legend=legend1;
run;
quit;

filename gsasfile clear; filename gsasfile "&Outdir\lower.ps";
axis2 color=black order=(-1 to 0 by 0.25)
     label = (a=90 h=3 font=greek "a");
proc gplot data=frechet;
    plot CorrL*logratio=lambda1 / haxis=axis1 vaxis=axis2
                                  legend=legend1;
run;
quit;
```

A.3 To Study Latent Correlations Using Bivariate Normal Copula

```
/*-------------------------------------------------------------
   Let Y1 and Y2 be a random variables with marginal Poisson
   distributions with means(lambda1) and (lambda2). Also let
   the joint pmf of (Y1,Y2) is defined using bivariate normal
   copula C(.;rho). For fixed value of lambda1=5 and variable
   values of lambda2 and rho, the following program computes
   the correlation coefficient and bias of pmf estimate
   under continuity approximation.
```

```
   Finally, this program produces Figures 4.3 and 4.4 of thesis.
-------------------------------------------------------------*/

%let outdir=%str(C:\Deepak\Paper 3\Count);

data Plotdata;
 do rho = -0.95 to 0.95 by 0.05;
   const = 1/(1-rho**2);
   lambda1 = 5.0;
   do p = 0.05 to 0.95 by 0.05;
     lambda2 = p/(1-p)*lambda1;
     bias = 0; mse = 0; cov = 0;
     y1 = 0; G1 = 0; f1 = exp(-lambda1); b1 = .M;
     do while(1-G1 > 0.1e-7);
       a1 = b1; G1 = G1 + f1; b1 = PROBIT(G1);
       y2 = 0; G2 = 0; f2 = exp(-lambda2); b2 = .M;
       do while(1-G2 > 0.1e-7);
         a2 = b2; G2 = G2 + f2; b2 = PROBIT(G2);
         f12Star = f1*f2*sqrt(const)*
             exp(-0.5*const*((rho**2)*(b1**2 + b2**2)-2*rho*b1*b2));
         if min(y1, y2) then
             f12 = (PROBBNRM(b1, b2, rho) + PROBBNRM(a1, a2, rho))
                  -(PROBBNRM(a1, b2, rho) + PROBBNRM(b1, a2, rho));
         else if (max(y1, y2) = 0) then f12 = PROBBNRM(b1, b2, rho);
         else if y1 then
             f12 = (PROBBNRM(b1, b2, rho) - PROBBNRM(a1, b2, rho));
         else f12 = (PROBBNRM(b1, b2, rho) - PROBBNRM(b1, a2, rho));
         bias = bias + f12*(f12Star-f12);
         cov = cov + f12*(y1*y2);
         y2 = y2+1; f2 = f2*lambda2/y2;
       end;
       y1 = y1+1; f1 = f1*lambda1/y1;
     end;
     alpha = (Cov - lambda1*lambda2)/sqrt(lambda1*lambda2);
```

```
     output;
   end;
 end;
 drop const a1 a2 b1 b2 G1 G2 f1 f2 y1 y2;
run;

data Plotdata;
    set Plotdata end=last;
    if (lambda2 < 40) then output;
    if last then do;
        alpha=-1; rho=-1; lambda2=0; bias=-0.09; output;
        alpha=1; rho=1; lambda2=40; bias=0.01; output;
    end;
run;

goptions  reset=all gaccess=gsasfile gunit=pct htitle=6 htext=3
vorigin=0in horigin=0in  ftext=swiss ftitle=swissb cback=white
hsize=7.5in vsize=6in  device=pslepsfc  autofeed;

filename gsasfile "&outdir\relation.ps";
proc g3d data=Plotdata;
    plot rho*lambda2=alpha / grid rotate=90 xticknum=5 yticknum=5
                             zticknum=5 zmin=-1 zmax=1
                             cbottom=blue ctop=green;
run;
quit;

filename gsasfile "&outdir\bias.ps";
proc g3d data=Plotdata;
    plot rho*lambda2=bias/ grid xticknum=5 yticknum=5 zticknum=5
                               cbottom=blue ctop=green;
run;
quit;
```

VITA

Deepak Mav
Department of Mathematics and Statistics
Old Dominion University
Norfolk, VA 23529

Education

Ph.D.	Old Dominion University, Norfolk, VA. (May 2005)
	Major: Computational and Applied Mathematics (Statistics)
MS	Old Dominion University, Norfolk, VA. (May 2004)
	Major: Computational and Applied Mathematics (Statistics),
MS	Indian Institute of Technology, Mumbai, India. (August 1999)
	Major: Applied Statistics and Informatics,
BS	University of Mumbai, Mumbai, India. (June 1997)
	Major: Statistics.

Experience

Biostatistics Graduate Assistant (05/2001 - 03/2003 & 12/2003 - 01/2005)
Eastern Virginia Medical School, Norfolk, VA

Teaching Assistant (08/1999 - 05/2001)
Old Dominion University, Norfolk, VA

Publications

Mav, D. and Chaganty, N. R., (2004) *"Bivariate Models for Identifying Differentially Expressed Genes in Microarray Experiments"*, Journal of Statistical Theory and Applications, 3(2), 111-124.

Chaganty, N. R. and **Mav, D.**, *"Asymptotic behavior of statistical methods for the analysis of correlated Poisson outcomes"*, under preparation.

Typeset using LaTeX.