1-1-2018

Regression Models For Analyzing Clustered Multinomial and Continuous Outcomes Under The Assumption of Exchangeability

Peter Ogutu Ngutu

Regression models for analyzing clustered Multinomial and Continuous Outcomes under the assumption of exchangeability

by

Peter O. Ngutu

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

Major: Mathematical Sciences

The University of Memphis

August 2018

DEDICATION

To my grandparents Zablon Waga Ngutu and Dani Patricia who both passed away in the course of my PhD studies. Thank you for always praying for me and believing in me. May you rest in eternal peace. "Precious in the sight of the LORD is the death of His saints" Psalms 116:15

ACKNOWLEDGEMENTS

I am grateful to my Lord and savior Jesus Christ for His great grace in making this work possible and without whom I would not have made it this far.

To my primary dissertation adviser, Dr. Dale Bowman. I thank God for you and I appreciate you for all the sacrifices you made in the course of working on this dissertation. You never hesitated to take time off to assist me with my paper even when it was inconvenient for you and for that please receive my heart felt appreciation. Your guidance and patience with me as I learned and implemented ideas and concepts have built me to be the student and researcher that I am today.

To my secondary adviser, Dr. E. O. George. Thank you for seeing the potential in me to achieve greater things than I even thought possible. You never allowed me to settle for mediocre and always pushed me to challenge myself beyond my personal limitations. I am especially grateful for your unassuming nature that taught me to be a more diligent researcher. I attribute a great deal of my success to your guidance throughout my graduate studies.

To Dr. Hongmei Zhang and Dr. Su Chen, I thank you for all your help, advise and incisive suggestions which have greatly helped in broadening the scope of this dissertation.

I would also like to acknowledge my family, especially my parents Mr. Charles and Grace Ngutu. Thank you for always being my support system. All my brothers and sisters: Patricia, Naomi, Josephine, Ian, Jackline, Jael and nephew Kirk, I thank you for being patient with me. I know it has not been easy not having your brother at home for the past 10 years but thank you for understanding and cheering me on all the while. To my uncle and aunt, Pst. Silvanus and Anne Oluoch, as well as my cousins; Ednauh, Fame and Givean, thank you for always being there for me when I needed you.

Lastly, to all friends and family who have assisted me in the course of my PhD study, I thank you all and God bless you abundantly.

ABSTRACT

Peter Ngutu. The University of Memphis. August 2018. Regression Models for Analyzing Clustered Multinomial and Continuous Outcomes Under the Assumption of Exchangeability. Major Professors: E. O. George, Ph.D. and D. Bowman, Ph.D.

We derive an expression for the joint distribution of exchangeable multinomial random variables and continous random variables for the purpose of analyzing clustered multinomial and continous data. In the past such clustered discrete and continous data could be analyzed with the use of quasi-likelihood procedures and generalized estimating equations to estimate marginal mean response parameters. Recently, the idea of exchangeability has been introduced to handle such data but research has focused primarily on analysis of clustered binary and continous outcomes. In applications to areas such as developmental toxicity studies, where discrete and contionus measures are recorded on each fetus, the discrete data may not necessarily be binary. For example, we may want to look at fetal death, malformation and normal fetuses as three possible outcomes separately. An impediment to a full likelihood-based analysis of such clustered multinomial data is the lack of a mathematically tractable representation of the joint distribution. The assumption of exchangeability is often reasonable in these fields of study where outcomes are measured within clusters and cluster responses can be assumed to be exchangeable in the sense that their joint distribution is invariant to permutation. We use this assumption to formulate fully parametric regression models for clusters of bivariate data with multinomial and contionus components. Tractable expressions for likelihood equations are derived and iterative schemes are given for computing efficient maximum likelihood estimates of the marginal mean, correlations, variances and higher moments. Regression models are then proposed having marginal interpretations and reproducible model structures. We demonstrate the use of the exchangeable procedure with an application to a developmental toxicity study involving fetal weight, malformation and death outcomes.

CONTENTS

Chapter Page

List of Tables vii

1 Introduction 1

2 Distribution Theory and likelihood function 5
- The Marginal Distribution For Exchangeable Multinomial Data 5
- Conditional Distribution of the Vector **W** Given **X** 6
- The Conditional Likelihood Function for the Joint Data 8

3 Non-Parametric Estimation of Parameters of Exchangeable Multinomial Distribution 9
- Combining Clusters of Different Sizes Assuming Interpretability 9
- E-Step 10
- M-Step 13

4 Parametric Estimation of Parameters of the Exchangeable Multinomial Distribution 15
- Joint Distribution of Trinary Responses 15
- Estimation of Parameters of the Discrete Model 16
- Weighted Least Squares Algorithm 18
- Modeling Using the Bivariate Exponential Function of Marshall and Olkin . 19

5 Estimation of Parameters of the Weight Model 23
- Parameter Estimation With Common Exchangeable Covariance Matrix . . . 23
- Estimations With Dose-Dependent Exchangeable Correlation Matrix 25

6 Application 26
- Data Summary 26
- Computed Parameter Estimates 28
- Joint Quantitative Risk Assessment 35
- Simulation Study 39
- Conclusion 44

References 46

A R-Codes 49
- EM Algorithm Assuming Interpretability 49
- Weighted Least Squares Algorithm 53

LIST OF TABLES

Table		Page
1	Discrete data for response of CD-1 Mice following exposure to ethylene glycol	27
2	Summary of response following exposure to ethylene glycol	28
3	Summary of observed Mean weights with Standard Deviations for Ethylene-Glycol Data	28
4	Estimates and standard errors of the Parameters for Malformation and Death Rate	29
5	Estimates of the Death and Malformation Rate under the Exchangeable Multinomial Model	30
6	Estimates of the weight parameters and standard errors with dose response functions	32
7	Estimates of the weight parameters with dose group fitted Jointly with the Exchangeable Regression Model	33
8	BMD and LED corresponding to q=10% excess risk for EG Data	37
9	Probabilities and CDF for each r,s pair for data with cluster size of n=2	40
10	Simulation results for the discrete parameters with cluster sizes of $n = 5$ and number of clusters, N	41
11	Simulation results for the discrete parameters with cluster sizes of $n = 10$ and number of clusters, N	42
12	Simulation results for weight parameters with cluster sizes of $n = 5$ and number of clusters, N	43
13	Simulation Results For weight parameters with cluster sizes of $n = 10$ and number of clusters, N	44

CHAPTER 1

INTRODUCTION

A major concern in developmental toxicity studies is the assessment of risk to unborn fetuses as a result of maternal exposure to hazardous components. Pre-clinical teratology experiments, usually with rodents, are thus performed to assess such risks in humans. These experiments usually have multiple endpoints and sometimes involve correlated discrete and continuous outcomes measured on the same subject. For example, pregnant mice are treated with the compound under investigation, and several measurements are recorded on the fetuses or pups in each litter. The end points recorded sometimes include multinomial indicators of fetal death and malformation, together with continuous measurements, such as fetal weight or length. The common feature in such studies is that measurements and responses within each cluster are usually highly correlated. For example, there is correlation between endpoints on a single fetus and there is also correlation between one endpoint and the same (or different) endpoints on a different fetus in the same litter.

Previous methods for analyzing mixed discrete and continuous outcomes include the general location model (GLOM) of Olkin and Tate [30] in which the discrete outcome is assumed to have a multinomial distribution and the continuous outcome is modeled by a multivariate Gaussian distribution; and a model by Cox [7] in which the marginal distribution of the continuous outcome is Gaussian, while the conditional distribution of the binary response given the continous outcomes is modeled by a logistic regression. Extensions of the location model have been proposed by Little and Schluchter [28], Little and Rubin [27] and Laird [23].

In the context of analyzing developmental toxicity data, Ryan et al. [32], Catalano and Ryan [6] and Fitzmaurice and Laird [13] proposed different models for the joint distribution of fetal malformation and weight. Catalano and Ryan [6] assumed that corresponding to the binary malformation variable, there exists an unobservable

continuous (Gaussian) latent variable, and that within each litter the malformation latent variables and fetal weights have a multivariate normal distribution. By conditioning on the fetal weight, they obtain a loglikelihood function as the sum of the marginal loglikelihood for fetal weight plus the conditional loglikelihood for malformation given fetal weight. The model of Fitzmaurice and Laird [13] is based on the general location model of Olkin and Tate [30] and the extension of this model by Laird [23] is implemented by the use of general estimating equations of Liang and Zeger [25] and Zeger and Liang [35] to model the marginal discrete outcomes and the conditional continuous outcomes given the discrete outcomes.

In this thesis, an extension of George et al. [17] is proposed where discrete data is multinomial as opposed to binary. Moreover, data from the same cluster is assumed to be exchangeable, as is the case in George et al. [18]. Specifically, for a single cluster, let $\{(X_1,W_1),\ldots,(X_n,W_n)\}$ be a set of observations from the same cluster, where X_i stands for the multinomial outcome, that is, $X_i = O_1$ if the ith fetus is dead, $X_i = O_2$ if the ith fetus is malformed and $X_i = O_3$ if the ith fetus is normal. Also let W_i denote the weight of the ith fetus. We assume that the vector of bivariate random variables $(X_1,W_1),\ldots,(X_n,W_n)$ is exchangeable in the sense of de Finetti [9]. That is, for any permutation $(\pi(1),\ldots,\pi(n))$ of $(1,\ldots,n)$,

$$\{(X_{\pi(1)},W_{\pi(1)}),\ldots,(X_{\pi(n)},W_{\pi(n)})\} \stackrel{\mathscr{L}}{=} \{(X_1,W_1),\ldots,(X_n,W_n)\}$$

Where "$\stackrel{\mathscr{L}}{=}$" denotes equality in distribution. Note that the exchangeability of $\{(X_1,W_1),\ldots,(X_n,W_n)\}$ implies that each of $(X_1,\ldots,X_n)$ and $(W_1,\ldots,W_n)$ is an exchangeable set of variables. However, $\{(X_1,W_1),\ldots,(X_n,W_n) \mid X_1 = x_1,\ldots,X_n = x_n\}$ is not necessarily exchangeable. Such a model, in which the marginal distributions of a set of bivariate random variables are exchangeable, but the conditional distribution is not, belongs to a class of distributions called quasi-symmetric (Darroch, [8]).

Similar to George et al. [17], we express the joint distribution of $\{(X_1,W_1),\ldots,(X_n,W_n)\}$ as a product of the marginal distribution of $(X_1,\ldots,X_n)$ and the

conditional distribution of $(W_1, \ldots, W_n)$ given $X_1, \ldots, X_n$. In application to developmental data, this has the interpretation of modeling fetal weights in a litter given the malformation, death or normal status of the fetuses in the same litter.

This thesis is organized as follows. In chapter 2, the joint distribution of the correlated discrete and continuous outcomes is discussed. The exchangeable multinomial model is then introduced and the conditional distribution of the vector of weights and the conditional likelihood function for the joint data are derived.

In chapter 3, an extension of Bowman and George's [4] approach is used to derive maximum likelihood estimates of the exchangeable multinomial data with an assumption of interpretability. The assumption of interpretability may be interpreted as assuming that the likelihood of responses within a cluster of any size is the same as a likelihood of those responses arising from a cluster with a maximum cluster size, with the observations not observed missing completely at random. EM algorithms are also derived when the cluster sizes are not equal and are missing.

In chapter 4, we review the likelihood estimation results of Bowman and George [3] for exchangeable multinomial data. An iterative weighted least squares approach is developed to maximize the likelihood and gain parameter estimates. Given the trinary outcomes, the mean weights are then estimated under various assumptions about the structure of the covariance matrix. Regression models are then constructed for efficient estimation of marginal mean parameters and correlations.

Chapter 5 looks at the estimations of the parameters of the weight model. Here, we model the weight parameters of the regression model using the specified covariates while assuming a multivariate normal distribution. For the estimation of the covariance matrix, we consider two cases, one where we have a common exchangeable covariance matrix and estimation with dose-dependent exchangeable correlation matrix.

In chapter 6, we look at a practical application using data from a developmental toxicity study conducted by the National Toxicology Program involving a study of

ethylene glycol (EG) using CD-1 mice. A risk assessment study is proposed where adverse effects of the dose-response model are examined to derive a safe dose of exposure for humans.

CHAPTER 2

DISTRIBUTION THEORY AND LIKELIHOOD FUNCTION

Consider a single cluster of outcomes $\{(X_1,W_1),\ldots,(X_n,W_n)\}$ where $(X_1,\ldots,X_n)$ is a sequence of exchangeable multinomial random variables and $(W_1,\ldots,W_n)$ is a set of exchangeable continous random variables. The joint density of $(X_1,W_1),\ldots,(X_n,W_n)$ can be factorized as

$$f_{\mathbf{X},\mathbf{W}}\big((x_1,w_1),\ldots,(x_n,w_n)\big) = f_{\mathbf{W}|\mathbf{X}}\big(w_1,\ldots,w_n \mid x_1,\ldots,x_n\big) f_{\mathbf{X}}\big(x_1,\ldots,x_n\big). \tag{1}$$

where $\mathbf{X} = (X_1,\ldots,X_n)'$and $\mathbf{W} = (W_1,\ldots,W_n)'$.

The joint log-likelihood function is written as a sum of the marginal log-likelihood for the multinomial endpoints and a conditional log-likelihood for the continuous endpoints given multinomial outcomes.

2.1 The Marginal Distribution For Exchangeable Multinomial Data

Consider an experiment consisting of multiple trials with each trial having $k+1$ possible categorical outcomes $O_1,O_2,\ldots,O_{k+1}$. Assume that each trial results in only one of these outcomes. Suppose we perform n trials under identical conditions, but in such a way that the outcomes from trial to trial are not independent but exchangeable and let $(X_1,X_2,\ldots,X_n)$ be random variables representing the outcomes of these trials.

The sequence $(X_1,X_2,\ldots,X_n)$ is exchangeable if

$$P(X_{\pi(1)} = x_1,\ldots,X_{\pi(n)} = x_n) = P(X_1 = x_1,\ldots,X_n = x_n) \tag{2}$$

for every permutation $\pi(1),\ldots,\pi(n)$ of $1,\ldots,n$.

Let $R_i =$ number of outcomes O_i, out of the n trials, $i = 1,\ldots,k+1$, so that $\sum_{i=1}^{k+1} R_i = n$. Then $(R_1,\ldots,R_k)$ is an exchangeable multinomial vector. The distribution of $(R_1,\ldots,R_k)$ is,

$$P(R_1 = r_1,\ldots,R_k = r_k) = \binom{n}{r_1,\ldots,r_k} P(X_1 = \cdots = X_{r_1} = O_1, X_{r_1+1} = \cdots$$

$$= X_{r_1+r_2} = O_2, \ldots, X_{\sum_{i=1}^{k-1} r_i+1} = \cdots = X_{\sum_{i=1}^{k} r_i} = O_k, X_{\sum_{i=1}^{k} r_i+1} = \cdots = X_{\sum_{i=1}^{k+1} r_i} = O_{k+1}) \quad (3)$$

where $\binom{n}{r_1,\ldots,r_k} = \frac{n!}{r_1!\cdots r_k!(n-\sum_{i=1}^{k} r_i)!}$.

If we let,

$$\tau_{r_1,\ldots,r_k;n} = P(X_1 = \cdots = X_{r_1} = O_1, X_{r_1+1} = \cdots$$
$$= X_{r_1+r_2} = O_2, \ldots, X_{\sum_{i=1}^{k-1} r_i+1} = \cdots = X_{\sum_{i=1}^{k} r_i} = O_k) \quad (4)$$

then, as in George et al. [15], it can be shown that

$$P(R_1 = r_1, \ldots, R_k = r_k) = p_{r_1,r_2,\ldots,r_k;n} =$$
$$\binom{n}{r_1,\ldots,r_k} \sum_{t=0}^{r_{k+1}} (-1)^t \binom{r_{k+1}}{t} \sum_{0 \le s_1+\cdots+s_k \le t} \binom{t}{s_1,\ldots,s_k} \tau_{r_1+s_1,\ldots,r_k+s_k} \quad (5)$$

2.2 Conditional Distribution of the Vector W Given X

Using the notation $(\mathbf{X}, \mathbf{W}) = \{(X_1, W_1), \ldots, (X_n, W_n)\}$, we assume that the conditional distribution of $\mathbf{W} \mid \mathbf{X}$,

$$\{\mathbf{W} \mid \mathbf{X} = \mathbf{x}\} = \{W_1, \ldots, W_n \mid X_1 = x_1, \ldots, X_n = x_n\},$$

has a multivariate normal distribution with mean vector $\boldsymbol{\mu}' = (a_1, \ldots, a_n)$ and covariance matrix $\boldsymbol{\Sigma}$, where

$$a_h = \begin{cases} \mu_1 & \text{if } x_h = O_1 \\ \vdots & \\ \mu_{k+1} & \text{if } x_h = O_{k+1} \end{cases} \quad (6)$$

for $h = 1, \ldots, n$. In the case of developmental response, $O_1, \cdots, O_{k+1}$ represent the discrete responses such as death, malformation etc while the W's are the continuous outcomes, for example, fetal weights or lengths.

Let

$$Cov[(W_h, W_l) \mid X_h = x_h, X_l = x_l]$$

$$= \begin{cases} \rho_u \sigma_u^2 \ , \text{if the } h\text{th and } l\text{th fetuses are of type } O_u \\ \rho_{uv} \sigma_u \sigma_v \ , \text{if } h\text{th fetus is type } O_u \text{ and } l\text{th fetus is type } O_v \end{cases} \tag{7}$$

for $1 \leq h, l \leq n$ and $1 \leq u, v \leq k+1$,

where

ρ_{uv} is the correlation between observations having different outcomes O_u and O_v,

$\sigma_u^2 = Var(W_i \mid X_i \text{ is type } O_u)$,

ρ_u is the correlation between observations having the same outcome O_u.

Let $\boldsymbol{\theta}' = (\mu_1, \ldots, \mu_{k+1}, \sigma_1^2, \ldots, \sigma_{k+1}^2, \rho_1, \ldots, \rho_{k+1}, \rho_{12}, \ldots, \rho_{k\ k+1}), (\mathbf{x}, \mathbf{w}) = \{(x_1, w_1), \ldots, (x_n, w_n)\}$, and $(\mathbf{x}^*, \mathbf{w}^*)$ be a permutation of the cluster components of $\{(x_1, w_1), \ldots, (x_n, w_n)\}$ such that $x_1^* = \cdots = x_{r_1}^* = O_1$,
$x_{(r_1+1)}^* = \cdots = x_{(r_1+r_2)}^* = O_2, \cdots, x_{(\sum_{u=1}^{k} r_u + 1)}^* = \cdots = x_{(\sum_{u=1}^{k+1} r_u)^*} = O_{k+1}$. That is, in $(\mathbf{x}^*, \mathbf{w}^*)$, the bivariate outcomes are ordered according to the multinomial responses. Then due to the exchangeability of $\{(X_1, W_1), \ldots, (X_n, W_n)\}$, the likelihood function of $(\boldsymbol{\tau}, \boldsymbol{\theta})$ can be expressed as

$$L(\boldsymbol{\tau}, \boldsymbol{\theta}; \mathbf{x}, \mathbf{w}) = L(\boldsymbol{\tau}, \boldsymbol{\theta}; \mathbf{x}^*, \mathbf{w}^*) = L_{\mathbf{x}^*}(\boldsymbol{\tau}) L_{\mathbf{w}^* | \mathbf{x}^*}(\boldsymbol{\tau}, \boldsymbol{\theta}) \tag{8}$$

where $\boldsymbol{\tau}$ is the vector of the parameters of the discrete model as in [2] and $L_{\mathbf{x}^*}(\boldsymbol{\tau})$ is the marginal likelihood based on the marginal density of $\mathbf{X}^*$, and $L_{\mathbf{w}^* | \mathbf{x}^*}(\boldsymbol{\tau}, \boldsymbol{\theta})$ is the conditional likelihood based on the conditional density of $\mathbf{W}^*$, given $\mathbf{X}^*$. We have that if $0 < r_1 + \cdots + r_k < n$, then

$$\mathbf{W}^* \mid \mathbf{X}^* = \mathbf{x}^* \sim N_n \left(\begin{bmatrix} \mu_1 \mathbf{1}_{r_1} \\ \mu_2 \mathbf{1}_{r_2} \\ \vdots \\ \mu_{k+1} \mathbf{1}_{r_{k+1}} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \cdots & \boldsymbol{\Sigma}_{1\ k+1} \\ \boldsymbol{\Sigma}_{12}' & \boldsymbol{\Sigma}_{22} & \cdots & \boldsymbol{\Sigma}_{2\ k+1} \\ \vdots & & \ddots & \\ \boldsymbol{\Sigma}_{1\ k+1}' & \boldsymbol{\Sigma}_{2\ k+1}' & \cdots & \boldsymbol{\Sigma}_{k+1\ k+1} \end{bmatrix} \right), \tag{9}$$

where $\boldsymbol{\Sigma}_{uu} = \sigma_u^2 \{(1 - \rho_u) \mathbf{I}_{r_u} + \rho_u \mathbf{J}_{r_u}\}$, for $u = 1, \ldots, k+1$ is the covariance structure for fetuses of response type O_u and $\boldsymbol{\Sigma}_{uv} = \rho_{uv} \sigma_u \sigma_v \mathbf{1}_{r_u} \mathbf{1}_{r_v}'$, for $1 \leq u < v \leq k+1$ is the covariance structure for fetuses of response types O_u and O_v. $\mathbf{I}_p$ is a $p \times p$ identity matrix, $\mathbf{J}_p$ is a $p \times p$ matrix of 1s, and $\mathbf{1}_p$ is a $p \times 1$ vector of 1s. Also, if any combination of the

sum of the number of responses $r_1, \ldots, r_k$, is 0 or n, then the $\boldsymbol{\mu}$ vector and the $\boldsymbol{\Sigma}$ matrix is reduced accordingly.

2.3 The Conditional Likelihood Function for the Joint Data

Consider a developmental toxicity study consisting of g dose groups. In the ith dose group, $i = 1, \ldots, g$, the number of clusters is m_i. In the jth cluster of the ith dose group there are n_{ij} fetuses. Let $(\mathbf{X}_{ij}, \mathbf{W}_{ij}) = \{(X_{ij1}, W_{ij1}), \ldots, (X_{ijn_{ij}}, W_{ijn_{ij}})\}$, where (X_{ijk}, W_{ijk}) is the multinomial and continous response of the kth fetus in the jth litter of the ith dose group, $k = 1, \ldots, n_{ij}$, $j = 1, \ldots, m_i$, $i = 1, \ldots, g$. Then

$$\{\mathbf{W}_{ij} \mid \mathbf{X}_{ij} = \mathbf{x}_{ij}\} \sim N_{n_{ij}}(\boldsymbol{\mu}_{ij}, \boldsymbol{\Sigma}_{ij}) \tag{10}$$

where

$$\boldsymbol{\mu}_{ij} = \begin{bmatrix} \mu_1 \mathbf{1}_{r_{1_{ij}}} \\ \mu_2 \mathbf{1}_{r_{2_{ij}}} \\ \vdots \\ \mu_{k+1} \mathbf{1}_{r_{k+1_{ij}}} \end{bmatrix},$$

and

$$\boldsymbol{\Sigma}_{ij} = \begin{bmatrix} \boldsymbol{\Sigma}_{11_{ij}} & \boldsymbol{\Sigma}_{12_{ij}} & \cdots & \boldsymbol{\Sigma}_{1\ k+1_{ij}} \\ \boldsymbol{\Sigma}'_{12_{ij}} & \boldsymbol{\Sigma}_{22_{ij}} & \cdots & \boldsymbol{\Sigma}_{2\ k+1_{ij}} \\ \vdots & & \ddots & \\ \boldsymbol{\Sigma}'_{1\ k+1_{ij}} & \boldsymbol{\Sigma}'_{2\ k+1_{ij}} & \cdots & \boldsymbol{\Sigma}_{k+1\ k+1_{ij}} \end{bmatrix}$$

for

$$\boldsymbol{\Sigma}_{uu_{ij}} = \sigma^2_{u_{ij}}\{(1-\rho_{u_{ij}})\mathbf{I}_{r_{u_{ij}}} + \rho_{u_i}\mathbf{J}_{r_{u_{ij}}}\}, \;\; u = 1, \ldots, k+1$$

$$\boldsymbol{\Sigma}_{uv_{ij}} = \rho_{uv_{ij}}\sigma_{u_{ij}}\sigma_{v_{ij}}\mathbf{1}_{r_{u_{ij}}}\mathbf{1}'_{r_{v_{ij}}}, \;\; 1 \le u < v \le k+1$$

where the $\rho_{u_{ij}}$ and $\rho_{uv_{ij}}$'s are the correlation parameters for the weights of litters in the jth litter of the ith treatment group and $(\sigma^2_{1_{ij}}, \ldots, \sigma^2_{k+1_{ij}})$ are the variances for the weights of litters in the jth cluster of the ith treatment group.

CHAPTER 3

NON-PARAMETRIC ESTIMATION OF PARAMETERS OF EXCHANGEABLE MULTINOMIAL DISTRIBUTION

An extension of George et al.[15] is adopted for a non parametric estimation of the τ's. For this estimation, George et al. [15] provide an expression of the τ's in terms of probabilities by inversion as

$$\tau_{r_1,\ldots,r_k} = \sum_{0\le s_1+\cdots+s_k\le r_{k+1}} \frac{\binom{r_{k+1}}{s_1,\ldots,s_k}}{\binom{n}{r_1+s_1,\ldots,r_k+s_k}} p_{r_1+s_1,\ldots,r_k+s_k;n}, \tag{11}$$

where $p_{r_1+s_1,\ldots,r_k+s_k;n}$ is as given in (5). Here, $r_1,\ldots,r_k$ represent the number of type $O_1,\ldots,O_k$ responses from a cluster of size n and $s_1,\ldots,s_k$ are all non-negative integers between 0 and r_{k+1} such that $s_1+\cdots+s_k$ is between 0 and r_{k+1}.

3.1 Combining Clusters of Different Sizes Assuming Interpretability

George et al. [15] obtain MLE's of marginal probabilities, τ, when cluster sizes are equal using the inversion expression in (11). When cluster sizes are unequal, MLE's of the τ's can be obtained using an EM - algorithm when an assumption of interpretability is reasonable.

Consider a single cluster of size n with $r_1,\ldots,r_{k+1}$ type $O_1,\ldots,O_{k+1}$ responses respectively. One way to interpret the assumption of marginal compatibility is to consider all observations as having originated from clusters with size M, where $M-n$ unobserved elements of the cluster are missing completely at random. For the exchangeable multinomial, the assumption of marginal compatibility or interpretability is that for all clusters sizes n, $\tau_{r_1,r_2,\ldots,r_k;n} = \tau_{r_1,r_2,\ldots,r_k;M} = \tau_{r_1,r_2,\ldots,r_k}$ for all $n = 1,\ldots,M$, where M is the maximum possible cluster size and $\tau_{r_1,\ldots,r_k;n}$ is as given in (4).

For each cluster where $n < M$, we will have n, observed members and $M-n$ unobserved members of the cluster. In a single dose group with N clusters, the observed

data is denoted as $\{R_{1_i} = r_{1_i}, R_{2_i} = r_{2_i}, \ldots, R_{k_i} = r_{k_i}; n_i\}_{i=1}^{N}$ where $r_{1_i}, r_{2_i}, \ldots, r_{k_i}$ represents the number of response types $O_1, O_2, \ldots, O_k$ respectively which have been observed in cluster i. Moreover, n_i is the observed size of the ith cluster and N is the total number of clusters. Similarly, the unobserved data take the form $\{R^*_{1_i} = r^*_{1_i}, R^*_{2_i} = r^*_{2_i}, \ldots, R^*_{k_i} = r^*_{k_i}; M - n_i\}_{i=1}^{N}$ where $r^*_{1_i}, r^*_{2_i}, \ldots, r^*_{k_i}$ represent the number of missing response types $O_1, O_2, \ldots, O_k$ respectively which are unobserved among the $(M - n_i)$ members of cluster i. The parameters of interest are

$$\Theta = \{p_{r_1,\ldots,r_k;M} \mid 0 \leq r_1, \ldots, r_k \leq M, \sum_{i=1}^{k} r_i \leq M\}$$

where $p_{r_1,\ldots,r_k;M}$ is as given in (5) for cluster size M. An E-M algorithm may be used to estimate the unknown parameters.

3.2 E-Step

For a single dose group, the complete data log-likelihood function is

$$\ell(\Theta) = \sum_{i=1}^{N} \log\left(p_{r_{1_i}+r^*_{1_i},\ldots,r_{k_i}+r^*_{k_i};n_i}\right), \tag{12}$$

where Θ is the set of parameters of interest.

The conditional distribution of the missing data given the observed and the parameter values at the tth iteration is found as

$$f(\{r^*_{1_i}, \ldots, r^*_{k_i}\}_{i=1}^{N} \mid \{r_{1_i}, \ldots, r_{k_i}\}_{i=1}^{N}, \Theta^{(t)}, M) = \prod_{i=1}^{N} \frac{P\left((r^*_{1_i},\ldots,r^*_{k_i} \text{ out of } M-n_i)\&(r_{1_i},\ldots,r_{k_i} \text{ out of } n_i)\right)}{P(r_{1_i},\ldots,r_{k_i} \text{ out of } n_i)} \tag{13}$$

For the numerator, there will be a total of $r_{j_i} + r^*_{j_i}$ observed and unobserved outcomes of type $O_j,\ j = 1, \ldots, k$, hence the probability of having these specific number of responses in the tth iteration for each outcome will be $p^{(t)}_{r_{1_i}+r^*_{1_i},\ldots,r_{k_i}+r^*_{k_i};M}$. Furthermore, the possible number of ways of choosing $r^*_{j_i}$ unobserved outcomes of type $O_j,\ j = 1, \ldots, k$ from a total of $M - n_i$ unobserved responses and r_{j_i} observed outcomes of type $O_j,\ j = 1, \ldots, k$ from a total of n_i observed responses is equivalent to a multinomial hypergeometric distribution

having $\binom{M-n_i}{r^*_{1_i},\ldots,r^*_{k_i}}$ different ways of choosing unobserved responses, $\binom{n_i}{r_{1_i},\ldots,r_{k_i}}$ different ways of choosing observed responses and a total of $\binom{M}{r_{1_i}+r^*_{1_i},\ldots,r_{k_i}+r^*_{k_i}}$ different ways of choosing both observed and unobserved responses. Multiplying the probability, $p^{(t)}_{r_{1_i}+r^*_{1_i},\ldots,r_{k_i}+r^*_{k_i};M}$ with the multinomial hypergeometric $\binom{n_i}{r_{1_i},\ldots,r_{k_i}}\binom{M-n_i}{r^*_{1_i},\ldots,r^*_{k_i}}/\binom{M}{r_{1_i}+r^*_{1_i},\ldots,r_{k_i}+r^*_{k_i}}$ will give the numerator as shown in (14).

$$= \prod_{i=1}^{N} \frac{p^{(t)}{}_{r_{1_i}+r^*_{1_i},\ldots,r_{k_i}+r^*_{k_i};M}\binom{n_i}{r_{1_i},\ldots,r_{k_i}}\binom{M-n_i}{r^*_{1_i},\ldots,r^*_{k_i}}/\binom{M}{r_{1_i}+r^*_{1_i},\ldots,r_{k_i}+r^*_{k_i}}}{p^{(t)}{}_{r_{1_i},\ldots,r_{k_i};n_i}} \tag{14}$$

$$= \prod_{i=1}^{N} \frac{\binom{n_i}{r_{1_i},\ldots,r_{k_i}}\binom{M-n_i}{r^*_{1_i},\ldots,r^*_{k_i}}p^{(t)}{}_{r_{1_i}+r^*_{1_i},\ldots,r_{k_i}+r^*_{k_i};M}}{\binom{M}{r_{1_i}+r^*_{1_i},\ldots,r_{k_i}+r^*_{k_i}}p^{(t)}{}_{r_{1_i},\ldots,r_{k_i};n_i}} \tag{15}$$

The conditional expectation of the complete data log-likelihood is then

$$Q(\Theta \mid \Theta^{(t)}) = E_{\{R^*_{1_i},\ldots,R^*_{k_i}\}_{i=1}^N \mid \{R_{1_i}=r_{1_i},\ldots,R_{k_i}=r_{k_i}\}_{i=1}^N,\Theta^{(t)}}\left(\sum_{i=1}^{N}\log\left(p_{r_{1_i}+r^*_{1_i},\ldots,r_{k_i}+r^*_{k_i};M}\right)\right)$$

$$= \sum_{r^*_{1_1}+\cdots+r^*_{k_1}=0}^{M-n_1}\cdots\sum_{r^*_{1_N}+\cdots+r^*_{k_N}=0}^{M-n_N}\sum_{i=1}^{N}\log p_{r_{1_i}+r^*_{1_i},\ldots,r_{k_i}+r^*_{k_i};M}\left(f(\{r^*_{1_i},\ldots,r^*_{k_i}\}_{i=1}^N \mid \{r_{1_i},\ldots,r_{k_i}\}_{i=1}^N,\Theta^{(t)},M)\right)$$

$$= \sum_{r^*_{1_1}+\cdots+r^*_{k_1}=0}^{M-n_1}\cdots\sum_{r^*_{1_N}+\cdots+r^*_{k_N}=0}^{M-n_N}\left[\sum_{i=1}^{N}\log p_{r_{1_i}+r^*_{1_i},\ldots,r_{k_i}+r^*_{k_i};M}\prod_{j=1}^{N}\frac{\binom{n_j}{r_{1_j},\ldots,r_{k_j}}\binom{M-n_j}{r^*_{1_j},\ldots,r^*_{k_j}}p^{(t)}{}_{r_{1_j}+r^*_{1_j},\ldots,r_{k_j}+r^*_{k_j};M}}{\binom{M}{r_{1_j}+r^*_{1_j},\ldots,r_{k_j}+r^*_{k_j}}p^{(t)}{}_{r_{1_j},\ldots,r_{k_j};n_j}}\right] \tag{16}$$

Let $Z_j = \frac{\binom{n_j}{r_{1_j},\ldots,r_{k_j}}\binom{M-n_j}{r^*_{1_j},\ldots,r^*_{k_j}}p^{(t)}{}_{r_{1_j}+r^*_{1_j},\ldots,r_{k_j}+r^*_{k_j};M}}{\binom{M}{r_{1_j}+r^*_{1_j},\ldots,r_{k_j}+r^*_{k_j}}p^{(t)}{}_{r_{1_j},\ldots,r_{k_j};n_j}}$, then our equation can be re-written as

$$= \sum_{r^*_{1_1}+\cdots+r^*_{k_1}=0}^{M-n_1}\cdots\sum_{r^*_{1_N}+\cdots+r^*_{k_N}=0}^{M-n_N}\left[\log p_{r_{1_1}+r^*_{1_1},\ldots,r_{k_1}+r^*_{k_1};M}\prod_{j=1}^{N}Z_j+\cdots+\log p_{r_{1_N}+r^*_{1_N},\ldots,r_{k_N}+r^*_{k_N};M}\prod_{j=1}^{N}Z_j\right] \tag{17}$$

$$= \sum_{r^*_{1_1}+\cdots+r^*_{k_1}=0}^{M-n_1}\cdots\sum_{r^*_{1_N}+\cdots+r^*_{k_N}=0}^{M-n_N}\log p_{r_{1_1}+r^*_{1_1},\ldots,r_{k_1}+r^*_{k_1};M}\prod_{j=1}^{N}Z_j+\cdots + \sum_{r^*_{1_1}+\cdots+r^*_{k_1}=0}^{M-n_1}\cdots\sum_{r^*_{1_N}+\cdots+r^*_{k_N}=0}^{M-n_N}\log p_{r_{1_N}+r^*_{1_N},\ldots,r_{k_N}+r^*_{k_N};M}\prod_{j=1}^{N}Z_j \tag{18}$$

$$= \sum_{r^*_{1_1}+\cdots+r^*_{k_1}=0}^{M-n_1}\log p_{r_{1_1}+r^*_{1_1},\ldots,r_{k_1}+r^*_{k_1};M}\cdot Z_1\sum_{r^*_{1_2}+\cdots+r^*_{k_2}=0}^{M-n_2}\cdots\sum_{r^*_{1_N}+\cdots+r^*_{k_N}=0}^{M-n_N}\prod_{j=2}^{N}Z_j+\cdots$$

$$+\sum_{r^*_{1_N}+\cdots+r^*_{k_N}=0}^{M-n_1}\log p_{r_{1_N}+r^*_{1_N},\ldots,r_{k_N}+r^*_{k_N};M}\cdot Z_N\sum_{r^*_{1_1}+\cdots+r^*_{k_1}=0}^{M-n_1}\cdots\sum_{r^*_{1_{N-1}}+\cdots+r^*_{k_{N-1}}=0}^{M-n_{N-1}}\prod_{j=1}^{N-1}Z_j \quad (19)$$

Due to the properties of distributions, the summations of the variables of the conditional distribution for each term will add up to 1. e.g. For the first term in equation (19) we will have

$$\sum_{r^*_{1_2}+\cdots+r^*_{k_2}=0}^{M-n_2}\cdots\sum_{r^*_{1_N}+\cdots+r^*_{k_N}=0}^{M-n_N}\prod_{j=2}^{N}\frac{\binom{n_j}{r_{1_j},\ldots,r_{k_j}}\binom{M-n_j}{r^*_{1_j},\ldots,r^*_{k_j}}p^{(t)}{}_{r_{1_j}+r^*_{1_j},\ldots,r_{k_j}+r^*_{k_j};M}}{\binom{M}{r_{1_j}+r^*_{1_j},\ldots,r_{k_j}+r^*_{k_j}}p^{(t)}{}_{r_{1_j},\ldots,r_{k_j};n_j}}=1 \quad (20)$$

Therefore, equation (19) reduces to

$$=\sum_{r^*_{1_1}+\cdots+r^*_{k_1}=0}^{M-n_1}\log p_{r_{1_1}+r^*_{1_1},\ldots,r_{k_1}+r^*_{k_1};M}\cdot Z_1+\cdots+\sum_{r^*_{1_N}+\cdots+r^*_{k_N}=0}^{M-n_N}\log p_{r_{1_N}+r^*_{1_N},\ldots,r_{k_N}+r^*_{k_N};M}\cdot Z_N \quad (21)$$

Replacing for Z_j we get

$$=\sum_{r^*_{1_1}+\cdots+r^*_{k_1}=0}^{M-n_1}\log p_{r_{1_1}+r^*_{1_1},\ldots,r_{k_1}+r^*_{k_1};M}\frac{\binom{n_1}{r_{1_1},\ldots,r_{k_1}}\binom{M-n_1}{r^*_{1_1},\ldots,r^*_{k_1}}p^{(t)}{}_{r_{1_1}+r^*_{1_1},\ldots,r_{k_1}+r^*_{k_1};M}}{\binom{M}{r_{1_1}+r^*_{1_1},\ldots,r_{k_1}+r^*_{k_1}}p^{(t)}{}_{r_{1_1},\ldots,r_{k_1};n_1}}+\cdots$$

$$+\sum_{r^*_{1_N}+\cdots+r^*_{k_N}=0}^{M-n_N}\log p_{r_{1_N}+r^*_{1_N},\ldots,r_{k_N}+r^*_{k_N};M}\frac{\binom{n_N}{r_{1_N},\ldots,r_{k_N}}\binom{M-n_N}{r^*_{1_N},\ldots,r^*_{k_N}}p^{(t)}{}_{r_{1_N}+r^*_{1_N},\ldots,r_{k_N}+r^*_{k_N};M}}{\binom{M}{r_{1_N}+r^*_{1_N},\ldots,r_{k_N}+r^*_{k_N}}p^{(t)}{}_{r_{1_N},\ldots,r_{k_N};n_N}} \quad (22)$$

This can be re-written as

$$\sum_{i=1}^{N}\sum_{r^*_{1_i}+\cdots+r^*_{k_i}=0}^{M-n_i}\log p_{r_{1_i}+r^*_{1_i},\ldots,r_{k_i}+r^*_{k_i};M}\frac{\binom{n_i}{r_{1_i},\ldots,r_{k_i}}\binom{M-n_i}{r^*_{1_i},\ldots,r^*_{k_i}}p^{(t)}{}_{r_{1_i}+r^*_{1_i},\ldots,r_{k_i}+r^*_{k_i};M}}{\binom{M}{r_{1_i}+r^*_{1_i},\ldots,r_{k_i}+r^*_{k_i}}p^{(t)}{}_{r_{1_i},\ldots,r_{k_i};n_i}}$$

$$=\sum_{i=1}^{N}\sum_{r^*_{1_i}=0}^{M-n_i}\sum_{r^*_{2_i}=0}^{M-n_i-r^*_{1_i}}\cdots\sum_{r^*_{k_i}=0}^{M-n_i-\sum_{j=1}^{k-1}r^*_{j_i}}\log p_{r_{1_i}+r^*_{1_i},\ldots,r_{k_i}+r^*_{k_i};M}\frac{\binom{n_i}{r_{1_i},\ldots,r_{k_i}}\binom{M-n_i}{r^*_{1_i},\ldots,r^*_{k_i}}p^{(t)}{}_{r_{1_i}+r^*_{1_i},\ldots,r_{k_i}+r^*_{k_i};M}}{\binom{M}{r_{1_i}+r^*_{1_i},\ldots,r_{k_i}+r^*_{k_i}}p^{(t)}{}_{r_{1_i},\ldots,r_{k_i};n_i}}$$

if we let $r_1=r_{1_i}+r^*_{1_i}, r_2=r_{2_i}+r^*_{2_i},\ldots,r_k=r_{k_i}+r^*_{k_i}$, we will now have

$$=\sum_{i=1}^{N}\sum_{r_1=r_{1_i}}^{M-(n_i-r_{1_i})}\sum_{r_2=r_{2_i}}^{M-(n_i-r_{1_i}-r_{2_i})-r_1}\cdots$$

$$\sum_{r_k=r_{k_i}}^{M-(n_i-\sum_{j=1}^{k}r_{j_i})-\sum_{j=1}^{k-1}r_j}\frac{\binom{n_i}{r_{1_i},\ldots,r_{k_i}}\binom{M-n_i}{r_1-r_{1_i},\ldots,r_k-r_{k_i}}p^{(t)}{}_{r_1,\ldots,r_k;M}}{\binom{M}{r_1,\ldots,r_k}p^{(t)}{}_{r_{1_i},\ldots,r_{k_i};n_i}}\log p_{r_1,\ldots,r_k;M}$$

Assuming $\binom{M}{r_1,\ldots,r_k}=0$, whenever $M<0, M<r_1+\cdots+r_k$ or $r_j<0$, for any j, such

that $j = 1, 2, \ldots, k$

$$Q(\Theta \mid \Theta^{(t)}) = \sum_{r_1=0}^{M} \sum_{r_2=0}^{M-r_1} \cdots \sum_{r_k=0}^{M-\sum_{j=1}^{k-1} r_j} \left[\sum_{i=1}^{N} \frac{\binom{n_i}{r_{1_i},\ldots,r_{k_i}} \binom{M-n_i}{r_1-r_{1_i},\ldots,r_k-r_{k_i}} p^{(t)}{}_{r_1,\ldots,r_k;M}}{\binom{M}{r_1,\ldots,r_k} p^{(t)}{}_{r_{1_i},\ldots,r_{k_i};n_i}} \right] \log p_{r_1,\ldots,r_k;M}$$

$$= \sum_{0 \le r_1+\cdots+r_k \le M} \left[\sum_{i=1}^{N} \frac{\binom{n_i}{r_{1_i},\ldots,r_{k_i}} \binom{M-n_i}{r_1-r_{1_i},\ldots,r_k-r_{k_i}} p^{(t)}{}_{r_1,\ldots,r_k;M}}{\binom{M}{r_1,\ldots,r_k} p^{(t)}{}_{r_{1_i},\ldots,r_{k_i};n_i}} \right] \log p_{r_1,\ldots,r_k;M} \tag{23}$$

3.3 M-Step

For the M-Step we need to maximize $Q(\Theta \mid \Theta^{(t)})$ subject to the constraint that

$$0 \le r_1 + \cdots + r_k \le M p_{r_1,\ldots,r_k;M} = 1 \tag{24}$$

Define

$$A_i = \frac{\binom{n_i}{r_{1_i},\ldots,r_{k_i}} \binom{M-n_i}{r_1-r_{1_i},\ldots,r_k-r_{k_i}} p^{(t)}{}_{r_1,\ldots,r_k;M}}{\binom{M}{r_1,\ldots,r_k} p^{(t)}{}_{r_{1_i},\ldots,r_{k_i};n_i}}$$

Using LaGrange multipliers define

$$L = \sum_{0 \le r_1+\cdots+r_k \le M} \left[\sum_{i=1}^{N} A_i \right] \log p_{r_1,\ldots,r_k;M} - \lambda \left(\sum_{0 \le r_1+\cdots+r_k \le M} p_{r_1,\ldots,r_k;M} - 1 \right).$$

For each $p_{r_1,\ldots,r_k;M}$, we have

$$\frac{\partial L}{\partial p_{r_1,\ldots,r_k;M}} = \frac{1}{p_{r_1,\ldots,r_k;M}} \sum_{i=1}^{N} A_i - \lambda \tag{25}$$

$$\frac{\partial L}{\partial \lambda} = - \sum_{0 \le r_1+\cdots+r_k \le M} p_{r_1,\ldots,r_k;M} + 1 \tag{26}$$

When we set the equations equal to zero and solve simultaneously we get

$$\frac{1}{\hat{p}_{r_1,\ldots,r_k;M}} \sum_{i=1}^{N} A_i - \lambda = 0$$

$$- \sum_{0 \le r_1+\cdots+r_k \le M} \hat{p}_{r_1,\ldots,r_k;M} + 1 = 0$$

The first equation (25) simplifies to

$$\hat{p}_{r_1,\ldots,r_k;M} = \frac{\sum_{i=1}^{N} A_i}{\lambda}.$$

When we substitute the results of the first equation (25) in the second equation (26) we get

$$- \sum_{0 \le r_1+\cdots+r_k \le M} \frac{\sum_{i=1}^{N} A_i}{\lambda} + 1 = 0.$$

Solving for λ yields the equation $\lambda = \sum_{0\le r_1+\cdots+r_k\le M}\sum_{i=1}^{N} A_i$, which on substituting back to the first equation (25) yields

$$\hat{p}_{r_1,\ldots,r_k;M} = \frac{\sum_{i=1}^{N} A_i}{\sum_{0\le r_1+\cdots+r_k\le M}\sum_{i=1}^{N} A_i}.$$

Due to the conditional probability interpretation of the addends

$$\sum_{0\le r_1+\cdots+r_k\le M} \frac{\binom{n_i}{r_{1_i},\ldots,r_{k_i}}\binom{M-n_i}{r_1-r_{1_i},\ldots,r_k-r_{k_i}} {p^{(t)}}_{r_1,\ldots,r_k;M}}{\binom{M}{r_1,\ldots,r_k}{p^{(t)}}_{r_{1_i},\ldots,r_{k_i};n_i}} = 1$$

$$\sum_{0\le r_1+\cdots+r_k\le M}\sum_{i=1}^{N} A_i = \sum_{i=1}^{N}\sum_{0\le r_1+\cdots+r_k\le M} A_i = \sum_{i=1}^{N} 1 = N$$

Therefore

$${p^{(t+1)}}_{r_1,\ldots,r_k;M} = \frac{\sum_{i=1}^{N} A_i^{(t)}}{N}$$

$${p^{(t+1)}}_{r_1,\ldots,r_k;M} = \frac{1}{N}\sum_{i=1}^{N} \frac{\binom{n_i}{r_{1_i},\ldots,r_{k_i}}\binom{M-n_i}{r_1-r_{1_i},\ldots,r_k-r_{k_i}} {p^{(t)}}_{r_1,\ldots,r_k;M}}{\binom{M}{r_1,\ldots,r_k}{p^{(t)}}_{r_{1_i},\ldots,r_{k_i};n_i}} \tag{27}$$

The iterative procedure of the EM-Algorithm proceeds as follows. Initial values of ${p^{(0)}}_{r_1,\ldots,r_k;M}$ and $p^{(0)}_{r_{1_i},\ldots,r_{k_i};n_i}$ are chosen. For instance, starting probabilities could be chosen such that all probabilities are equal. The value of ${p^{(1)}}_{r_1,\ldots,r_k;M}$ is obtained using the starting values from (27). This procedure is implemented iteratively until it converges. Upon convergence, estimation of the parameters $\tau_{r_1,\ldots,r_k;M}$ may be obtained from the inversion equation (11).

CHAPTER 4

PARAMETRIC ESTIMATION OF PARAMETERS OF THE EXCHANGEABLE MULTINOMIAL DISTRIBUTION

4.1 Joint Distribution of Trinary Responses

For parametric estimation, we will limit our model to the trinary case where for exchangeable multinomial response variables $X_1, \ldots, X_n$, there are three possible response categories, $O_1, O_2,$ and O_3. We define marginal probabilities as before

$$\tau_{r,s} = P(X_1 = \cdots = X_r = O_1, X_{r+1} = \cdots = X_{r+s} = O_2) \tag{28}$$

which, assuming exchangeability, is the same probability as for any permutation of $r+s$ of the random variables. As in George et al. [15], the probability of r type 1 responses and s type 2 responses in the n random variables, is given by

$$p_{r,s;n} = \binom{n}{r\ s} \sum_{k=0}^{n-r-s} (-1)^k \binom{n-r-s}{k} \sum_{\ell=0}^{k} \binom{k}{\ell} \tau_{r+\ell,s+k-\ell} \tag{29}$$

for $\binom{n}{r\ s} = \frac{n!}{r!s!(n-r-s)!}$. Letting R_1 be the random variable of the number of type O_1 outcomes and R_2, the random variable of the number of type O_2 outcomes, the joint moment generating function of R_1 and R_2 is

$$M_{R_1,R_2}(t_1,t_2) = \sum_{j=0}^{n} \sum_{\ell=0}^{j} (1-e^{t_1})^{\ell} (1-e^{t_2})^{j-\ell} \binom{n}{j} \binom{j}{\ell} \tau_{\ell,j-\ell} \tag{30}$$

From the moment generating function it can be shown that $E(R_1) = n\tau_{1,0},\ E(R_2) = n\tau_{0,1}$ and the covariance matrix of $(R_1,R_2)'$ is given by

$$Cov(R_1,R_2) = \begin{bmatrix} n\tau_{1,0}(1-n\tau_{1,0}) + n(n-1)\tau_{2,0} & n(n-1)\tau_{1,1} - n^2\tau_{1,0}\tau_{0,1} \\ n(n-1)\tau_{1,1} - n^2\tau_{1,0}\tau_{0,1} & n\tau_{0,1}(1-n\tau_{0,1}) + n(n-1)\tau_{0,2} \end{bmatrix}. \tag{31}$$

The correlation between type 1 responses is $\phi_{11} = \frac{\tau_{2,0}-\tau_{1,0}^2}{\tau_{1,0}(1-\tau_{1,0})}$, correlation between type 2 responses is $\phi_{22} = \frac{\tau_{0,2}-\tau_{0,1}^2}{\tau_{0,1}(1-\tau_{0,1})}$ and the correlation between between a type 1 and a type 2 response is given by $\phi_{12} = \frac{\tau_{1,1}-\tau_{1,0}\tau_{0,1}}{\sqrt{\tau_{1,0}(1-\tau_{1,0})\tau_{0,1}(1-\tau_{0,1})}}$.

4.2 Estimation of Parameters of the Discrete Model

Bergsma and Rudas [1] proved that any valid joint parametric modeling of marginal responses and pairwise correlations of multinomial responses cannot be achieved without some restrictions on the parameter space. For the correlated trinomial, when assuming a functional form for $\tau_{x,y} = q(x,y)$ a restriction on the parameter space is required. Bowman and George [3], show that if the $\tau's$ satisfy the condition of bivariate complete monotonicity they will result in a valid probability model. A bivariate function $q(x,y)$ is said to be bivariate completely monotonic on an interval I if $q(x,y)$ has derivatives of all orders on I and

$$(-1)^k q^{(k)}(x,y) \geq 0 \tag{32}$$

for positive integers k, where $q^{(k)}(x,y) = \frac{\partial^k q(x,y)}{\partial^j x \, \partial^{k-j} y}$ for $j = 0,\ldots,k$ and for all $k = 1,2,3...$ Bernstein's theorem, says that a function, f is completely monotone if and only if it is the Lapalace transform of a finite Borel measure, μ on $\mathbb{R}^+$

$$f(x) = \int_0^\infty e^{-xt} \partial\mu(t). \tag{33}$$

One such bivariate function that is a Laplace transform of the bivariate exponential distribution of Marshall and Olkin [29] is

$$\psi(r,s) = \frac{(\lambda + r + s)(\lambda_1 + \lambda_3)(\lambda_2 + \lambda_3) + \lambda_3 rs}{(\lambda + r + s)(\lambda_1 + \lambda_3 + r)(\lambda_2 + \lambda_3 + s)} \tag{34}$$

where $\lambda_1, \lambda_2, \lambda_3 \geq 0,\ \lambda_1 + \lambda_3 > 0,\ \lambda_2 + \lambda_3 > 0$, and $\lambda = \lambda_1 + \lambda_2 + \lambda_3$. Marshall and Olkin in [29], prove that $\psi(r,s)$ in equation (34) is indeed the Laplace transform of the bivariate exponential distribution, $f(r,s) = exp[-\lambda_1 r - \lambda_2 s - \lambda_3 \max(r,s)],\ \ r,s > 0$. Suppose we let $\tau_{r,s} = \psi(r,s)$, the probability of type 1 response, $\tau_{1,0} = \frac{\lambda_1 + \lambda_3}{\lambda_1 + \lambda_3 + 1}$ and the probability of type 2 response, $\tau_{0,1} = \frac{\lambda_2 + \lambda_3}{\lambda_2 + \lambda_3 + 1}$. To calculate the correlation between type 1 responses, we can use the correlation formula previously defined in Section 4.1, $\phi_{11} = \frac{\tau_{2,0} - \tau_{1,0}^2}{\tau_{1,0}(1 - \tau_{1,0})}$.Replacing

for $\tau_{2,0}$ and $\tau_{1,0}$ in the correlation formula yields

$$\phi_{11} = \left[\frac{\lambda_1+\lambda_3}{\lambda_1+\lambda_3+2} - \left(\frac{\lambda_1+\lambda_3}{\lambda_1+\lambda_3+1}\right)^2\right] \div \left[\frac{\lambda_1+\lambda_3}{\lambda_1+\lambda_3+1}\left(1-\frac{\lambda_1+\lambda_3}{\lambda_1+\lambda_3+1}\right)\right]$$

$$= \left[\frac{(\lambda_1+\lambda_3)(\lambda_1+\lambda_3+1)^2-(\lambda_1+\lambda_3+2)(\lambda_1+\lambda_3)^2}{(\lambda_1+\lambda_3+2)(\lambda_1+\lambda_3+1)^2}\right] \div \left[\frac{\lambda_1+\lambda_3}{\lambda_1+\lambda_3+1}\left(1-\frac{\lambda_1+\lambda_3}{\lambda_1+\lambda_3+1}\right)\right]$$

$$= \left[\frac{(\lambda_1+\lambda_3)(\lambda_1+\lambda_3+1)^2-(\lambda_1+\lambda_3+2)(\lambda_1+\lambda_3)^2}{(\lambda_1+\lambda_3+2)(\lambda_1+\lambda_3+1)^2}\right] \div \left[\frac{\lambda_1+\lambda_3}{(\lambda_1+\lambda_3+1)^2}\right]$$

$$= \left[\frac{(\lambda_1+\lambda_3+1)^2-(\lambda_1+\lambda_3+2)(\lambda_1+\lambda_3)}{(\lambda_1+\lambda_3+2)}\right]$$

$$11 = \frac{1}{\lambda_1+\lambda_3+2}. \tag{35}$$

Similarly, the correlation between type 2 responses can be calculated from the correlation formula $\phi_{22} = \frac{\tau_{0,2}-\tau_{0,1}^2}{\tau_{0,1}(1-\tau_{0,1})}$ also given in Section 4.1.

$$\phi_{22} = \left[\frac{\lambda_2+\lambda_3}{\lambda_2+\lambda_3+2} - \left(\frac{\lambda_2+\lambda_3}{\lambda_2+\lambda_3+1}\right)^2\right] \div \left[\frac{\lambda_2+\lambda_3}{\lambda_2+\lambda_3+1}\left(1-\frac{\lambda_2+\lambda_3}{\lambda_2+\lambda_3+1}\right)\right]$$

$$= \left[\frac{(\lambda_2+\lambda_3)(\lambda_2+\lambda_3+1)^2-(\lambda_2+\lambda_3+2)(\lambda_2+\lambda_3)^2}{(\lambda_2+\lambda_3+2)(\lambda_2+\lambda_3+1)^2}\right] \div \left[\frac{\lambda_2+\lambda_3}{\lambda_2+\lambda_3+1}\left(1-\frac{\lambda_2+\lambda_3}{\lambda_2+\lambda_3+1}\right)\right]$$

$$= \left[\frac{(\lambda_2+\lambda_3)(\lambda_2+\lambda_3+1)^2-(\lambda_2+\lambda_3+2)(\lambda_2+\lambda_3)^2}{(\lambda_2+\lambda_3+2)(\lambda_2+\lambda_3+1)^2}\right] \div \left[\frac{\lambda_2+\lambda_3}{(\lambda_2+\lambda_3+1)^2}\right]$$

$$= \left[\frac{(\lambda_2+\lambda_3+1)^2-(\lambda_2+\lambda_3+2)(\lambda_2+\lambda_3)}{(\lambda_2+\lambda_3+2)}\right]$$

$$\phi_{22} = \frac{1}{\lambda_2+\lambda_3+2}, \tag{36}$$

and the correlation between type 1 and type 2 responses from the correlation formula $\phi_{12} = \frac{\tau_{1,1}-\tau_{1,0}\tau_{0,1}}{\sqrt{\tau_{1,0}(1-\tau_{1,0})\tau_{0,1}(1-\tau_{0,1})}}$, is given by

$$= \left[\frac{(\lambda+2)(\lambda_1+\lambda_3)(\lambda_2+\lambda_3)+\lambda_3}{(\lambda+2)(\lambda_1+\lambda_3+1)(\lambda_2+\lambda_3+1)} - \frac{(\lambda_1+\lambda_3)(\lambda_2+\lambda_3)}{(\lambda_1+\lambda_3+1)(\lambda_2+\lambda_3+1)}\right] \div \sqrt{\frac{(\lambda_1+\lambda_3)}{(\lambda_1+\lambda_3+1)^2}\frac{(\lambda_2+\lambda_3)}{(\lambda_2+\lambda_3+1)^2}}$$

$$= \left[\frac{(\lambda+2)(\lambda_1+\lambda_3)(\lambda_2+\lambda_3)+\lambda_3-(\lambda+2)(\lambda_1+\lambda_3)(\lambda_2+\lambda_3)}{(\lambda+2)(\lambda_1+\lambda_3+1)(\lambda_2+\lambda_3+1)}\right] \cdot \frac{(\lambda_1+\lambda_3+1)(\lambda_2+\lambda_3+1)}{\sqrt{(\lambda_1+\lambda_3)(\lambda_2+\lambda_3)}}$$

$$\phi_{12} = \frac{\lambda_3}{(\lambda+2)\sqrt{(\lambda_1+\lambda_3)(\lambda_2+\lambda_3)}}. \tag{37}$$

4.3 Weighted Least Squares Algorithm

Bowman and George [3] show that maximum likelihood estimates of $\boldsymbol{\tau}$ can be found using an iterative weighted least squares procedure, where $\tau_{r,s} = f(r,s)$ is a generalization of the power family used in Kuk [22]. Let random variable $A_{r,s,n}$ be the number of clusters having r type 1 responses and s type 2 responses among all clusters of size n. Also let $p_{r,s;n}$ be the probability of r type 1 responses and s type 2 responses in a cluster of size n. Then, $A_{r,s,n}$ follows a multinomial distribution with parameters $(m_n, p_{r,s,n}; 0 \leq r+s \leq n)$ for $r = 0,\ldots,n$, $s = 0,\ldots,n$ where m_n refers to the number of clusters of size n. To establish a relationship between parameters and covariates that may vary for each cluster, a link function is chosen to relate the responses and correlations to covariates. We therefore model $p_{r,s,n}$ as a function of vector covariates $\boldsymbol{x}_{r,s,n}$ and parameters $\boldsymbol{\Omega}$ through the marginal probabilities of $\tau_{r,s;n}$ as $\tau_{r,s;n} = F_{r,s}(\boldsymbol{\Omega}'\boldsymbol{x}_{r,s,n})$ and obtain maximum likelihood estimates of $\boldsymbol{\Omega}$.

Bowman and George [5] show that the iterative score function for estimating $\boldsymbol{\Omega}$ can be written as a weighted least squares estimate as

$$\boldsymbol{\Omega}^{j+1} = \left(\mathbf{X}\mathbf{D}(\boldsymbol{\Omega}^j)\mathbf{X}'\right)^{-1}\mathbf{X}\mathbf{D}(\boldsymbol{\Omega}^j)\mathbf{Z}(\boldsymbol{\Omega}^j) \tag{38}$$

where

$$\mathbf{Z}(\boldsymbol{\Omega}^j) = \mathbf{X}'\boldsymbol{\Omega}^j + \mathbf{D}^{-1}(\boldsymbol{\Omega}^j)\mathbf{D}_1(\boldsymbol{\Omega}^j)\left(\mathbf{A} - \mathbf{M}\,\mathbf{P}(\boldsymbol{\Omega}^j)\right)$$

The terms of the iterative equation are defined as follows: $\mathbf{A}$ is the vector of observations $A_{r,s,n}$.$\mathbf{P}$ is the associated vector of probabilities $p_{r,s,n}$ and $\mathbf{X}$ is the matrix of covariates whose columns are the $\boldsymbol{x}_{r,s,n}$ covariate vectors. Further, $\mathbf{D}_1$ is a diagonal matrix

with diagonal elements given by $\frac{f_{r,s,n}}{p_{r,s,n}}$, where

$$f_{r,s,n} = \binom{n}{r,s} \sum_{k=0}^{n-r-s} (-1)^k \binom{n-r-s}{k} \sum_{l=0}^{k} \binom{k}{l} \frac{\partial F_{r+l,s+k-l}(\boldsymbol{\Omega}'\boldsymbol{x})}{\partial(\boldsymbol{\Omega}'\boldsymbol{x})}, \tag{39}$$

$F_{r+l,s+k-l}$ being a bivariate completely monotone function with the parameter $\boldsymbol{\Omega}$ and $\mathbf{D}_2$ is a diagonal matrix with diagonal elements $f_{r,s,n}$. $\mathbf{M}$ is a diagonal matrix with diagonal elements m_n, where m_n is the number of clusters of size n. The term $\mathbf{D}_1(\boldsymbol{\Omega})\mathbf{M}\mathbf{D}_2(\boldsymbol{\Omega})$ is represented by $\mathbf{D}(\boldsymbol{\Omega})$.

4.4 Modeling Using the Bivariate Exponential Function of Marshall and Olkin

In section 4.2, the parametric model for $\tau_{r,s}$ is given by $\tau(r,s) = \frac{(\lambda+r+s)(\lambda_1+\lambda_3)(\lambda_2+\lambda_3)+\lambda_3 rs}{(\lambda+r+s)(\lambda_1+\lambda_3+r)(\lambda_2+\lambda_3+s)}$ where $\lambda_1,\lambda_2,\lambda_3 \geq 0,\ \lambda_1+\lambda_3 > 0,\ \lambda_2+\lambda_3 > 0$, and $\lambda = \lambda_1+\lambda_2+\lambda_3$. A general link function is used for the parameters λ_1,λ_2 and λ_3.

$$\lambda_i = f(\boldsymbol{\Lambda}_i'\boldsymbol{x}_i) \tag{40}$$

where $\boldsymbol{\Lambda}_i$ is a vector of parameters and $\boldsymbol{x}_i$ is a vector of covariates, where $i = 1,2,3$ and $\boldsymbol{x} = (\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{x}_3)$. For $\boldsymbol{\Lambda} = (\boldsymbol{\Lambda}_1,\boldsymbol{\Lambda}_2,\boldsymbol{\Lambda}_3)$, we proceed with estimating equations for the weighted least squares algorithm as in Bowman and George [5].

Assuming that $p_{r,s,n}$ is a function of parameters $\boldsymbol{\Lambda}$, the log likelihood function is written as

$$\ell(\boldsymbol{\Lambda} \mid \mathbf{A}) = \sum_{n=1}^{M} \sum_{r=0}^{n} \sum_{s=0}^{n-r} A_{r,s,n} \log p_{r,s,n}(\boldsymbol{\Lambda}), \tag{41}$$

where M is the maximum cluster size. To maximize $\ell(\boldsymbol{\Lambda} \mid \mathbf{A})$ we will construct an iterative least squares algorithm by taking derivatives with respect to $\boldsymbol{\Lambda}$ subject to constraints

$$\sum_{r=0}^{n} \sum_{s=0}^{n-r} p_{r,s,n} = 1 \tag{42}$$

for $n = 1,\ldots,K$. Using LaGrange multipliers, Bowman and George [5] show that maximizing $\boldsymbol{\Lambda}$ for the complete function is equivalent to solving

$$S(\boldsymbol{\Lambda} \mid \mathbf{A}) = \sum_{n=1}^{M} \sum_{r=0}^{n} \sum_{s=0}^{n-r} \frac{A_{r,s,n} - m_n p_{r,s,n}}{p_{r,s,n}} \frac{\partial p_{r,s,n}(\boldsymbol{\Lambda})}{\partial \boldsymbol{\Lambda}} = 0 \tag{43}$$

Let

$$f_{1_{r,s,n}}(\boldsymbol{\Lambda}_1'\boldsymbol{x}_1) = \binom{n}{r,s} \sum_{k=0}^{n-r-s} (-1)^k \binom{n-r-s}{k} \sum_{l=0}^{k} \binom{k}{l} \frac{\partial}{\partial(\boldsymbol{\Lambda}_1'\boldsymbol{x}_1)} F_{r+l,s+k-l}(\boldsymbol{\Lambda}'\boldsymbol{x}), \tag{44}$$

$$f_{2_{r,s,n}}(\boldsymbol{\Lambda}_2'\boldsymbol{x}_2) = \binom{n}{r,s}\sum_{k=0}^{n-r-s}(-1)^k\binom{n-r-s}{k}\sum_{l=0}^{k}\binom{k}{l}\frac{\partial}{\partial(\boldsymbol{\Lambda}_2'\boldsymbol{x}_2)}F_{r+l,s+k-l}(\boldsymbol{\Lambda}'\boldsymbol{x}), \tag{45}$$

$$f_{3_{r,s,n}}(\boldsymbol{\Lambda}_3'\boldsymbol{x}_3) = \binom{n}{r,s}\sum_{k=0}^{n-r-s}(-1)^k\binom{n-r-s}{k}\sum_{l=0}^{k}\binom{k}{l}\frac{\partial}{\partial(\boldsymbol{\Lambda}_3'\boldsymbol{x}_3)}F_{r+l,s+k-l}(\boldsymbol{\Lambda}'\boldsymbol{x}). \tag{46}$$

where $F_{r+l,s+k-l}$ is the bivariate complete monotone function in (34), with parameters $\boldsymbol{\Lambda}_i$ from the link function $\lambda_i(\boldsymbol{\Lambda}_i) = f_i(\boldsymbol{\Lambda}_i'\boldsymbol{x}_i)$.

The score function can be written as

$$\mathbf{S}(\boldsymbol{\Lambda}_i \mid \mathbf{A}) = \sum_{n=1}^{M}\sum_{r=0}^{n}\sum_{s=0}^{n-r}\frac{A_{r,s,n} - m_n p_{r,s,n}}{p_{r,s,n}} f_{i_{r,s,n}}(\boldsymbol{\Lambda}_i'\boldsymbol{x}_i)\,\boldsymbol{x}_{r,s,n} \tag{47}$$

Define $\mathbf{D}_{1a}(\boldsymbol{\Lambda})$ as the diagonal matrix with elements $\frac{f_{1_{r,s,n}}(\boldsymbol{\Lambda}_1'\boldsymbol{x}_1)}{p_{r,s,n}(\boldsymbol{\Lambda}_1'\boldsymbol{x}_1)}$ and $\mathbf{D}_{2a}(\boldsymbol{\Lambda})$ as the diagonal matrix with elements $f_{1_{r,s,n}}(\boldsymbol{\Lambda}_1'\boldsymbol{x}_1)$. Define $\mathbf{D}_{1b}(\boldsymbol{\Lambda})$ as the diagonal matrix with elements $\frac{f_{2_{r,s,n}}(\boldsymbol{\Lambda}_2'\boldsymbol{x}_2)}{p_{r,s,n}(\boldsymbol{\Lambda}_2'\boldsymbol{x}_2)}$ and $\mathbf{D}_{2b}(\boldsymbol{\Lambda})$ as the diagonal matrix with elements $f_{2_{r,s,n}}(\boldsymbol{\Lambda}_2'\boldsymbol{x}_2)$. Define $\mathbf{D}_{1c}(\boldsymbol{\Lambda})$ as the diagonal matrix with elements $\frac{f_{3_{r,s,n}}(\boldsymbol{\Lambda}_3'\boldsymbol{x}_3)}{p_{r,s,n}(\boldsymbol{\Lambda}_3'\boldsymbol{x}_3)}$ and $\mathbf{D}_{2c}(\boldsymbol{\Lambda})$ as the diagonal matrix with elements $f_{3_{r,s,n}}(\boldsymbol{\Lambda}_3'\boldsymbol{x}_3)$ and $\mathbf{M}$ be a diagonal matrix with diagonal elements m_n, for m_n the number of clusters of size n, where

$$\mathbf{D}_a(\boldsymbol{\Lambda}) = \mathbf{D}_{1a}(\boldsymbol{\Lambda})\mathbf{M}\mathbf{D}_{2a}(\boldsymbol{\Lambda}) \tag{48}$$

$$\mathbf{D}_b(\boldsymbol{\Lambda}) = \mathbf{D}_{1b}(\boldsymbol{\Lambda})\mathbf{M}\mathbf{D}_{2b}(\boldsymbol{\Lambda}) \tag{49}$$

$$\mathbf{D}_c(\boldsymbol{\Lambda}) = \mathbf{D}_{1c}(\boldsymbol{\Lambda})\mathbf{M}\mathbf{D}_{2c}(\boldsymbol{\Lambda}) \tag{50}$$

The $(j+1)$th steps of scoring algorithms for $\boldsymbol{\Lambda}_1$,$\boldsymbol{\Lambda}_2$ and $\boldsymbol{\Lambda}_3$ using the scoring method are:

$$\boldsymbol{\Lambda}_1^{j+1} = \boldsymbol{\Lambda}_1^{j} + (\mathbf{X}_1\mathbf{D}_a(\boldsymbol{\Lambda}^j)\mathbf{X}_1')^{-1}\mathbf{X}_1\mathbf{D}_{1a}(\boldsymbol{\Lambda}^j)(\mathbf{A} - \mathbf{M}\,\mathbf{P}(\boldsymbol{\Lambda}^j)) \tag{51}$$

$$\boldsymbol{\Lambda}_2^{j+1} = \boldsymbol{\Lambda}_2^{j} + (\mathbf{X}_2\mathbf{D}_b(\boldsymbol{\Lambda}^j)\mathbf{X}_2')^{-1}\mathbf{X}_2\mathbf{D}_{1b}(\boldsymbol{\Lambda}^j)(\mathbf{A} - \mathbf{M}\,\mathbf{P}(\boldsymbol{\Lambda}^j)) \tag{52}$$

$$\boldsymbol{\Lambda}_3^{j+1} = \boldsymbol{\Lambda}_3^{j} + (\mathbf{X}_3\mathbf{D}_c(\boldsymbol{\Lambda}^j)\mathbf{X}_3')^{-1}\mathbf{X}_3\mathbf{D}_{1c}(\boldsymbol{\Lambda}^j)(\mathbf{A} - \mathbf{M}\,\mathbf{P}(\boldsymbol{\Lambda}^j)) \tag{53}$$

These are transformed into iterative weighted least squares steps as

$$\boldsymbol{\Lambda}_1^{j+1} = \left(\mathbf{X}_1\mathbf{D}_a(\boldsymbol{\Lambda}^j)\mathbf{X}_1'\right)^{-1}\mathbf{X}_1\mathbf{D}_a(\boldsymbol{\Lambda}^j)\mathbf{Z}_a(\boldsymbol{\Lambda}^j) \tag{54}$$

where

$$\mathbf{Z}_a(\boldsymbol{\Lambda}^j) = \mathbf{X}_1'\boldsymbol{\Lambda}_1^j + \mathbf{D}_a^{-1}(\boldsymbol{\Lambda}^j)\mathbf{D}_{1a}(\boldsymbol{\Lambda}^j)\left(\mathbf{A} - \mathbf{M}\,\mathbf{P}(\boldsymbol{\Lambda}^j)\right)$$

and

$$\boldsymbol{\Lambda}_2^{j+1} = \left(\mathbf{X}_2\mathbf{D}_b(\boldsymbol{\Lambda}^j)\mathbf{X}_2'\right)^{-1}\mathbf{X}_2\mathbf{D}_b(\boldsymbol{\Lambda}^j)\mathbf{Z}_b(\boldsymbol{\Lambda}^j) \tag{55}$$

where

$$\mathbf{Z}_b(\boldsymbol{\Lambda}^j) = \mathbf{X}_2'\boldsymbol{\Lambda}_2^j + \mathbf{D}_b^{-1}(\boldsymbol{\Lambda}^j)\mathbf{D}_{1b}(\boldsymbol{\Lambda}^j)\left(\mathbf{A} - \mathbf{M}\,\mathbf{P}(\boldsymbol{\Lambda}^j)\right)$$

and

$$\boldsymbol{\Lambda}_3^{j+1} = \left(\mathbf{X}_3\mathbf{D}_c(\boldsymbol{\Lambda}^j)\mathbf{X}_3'\right)^{-1}\mathbf{X}_3\mathbf{D}_c(\boldsymbol{\Lambda}^j)\mathbf{Z}_c(\boldsymbol{\Lambda}^j) \tag{56}$$

where

$$\mathbf{Z}_c(\boldsymbol{\Lambda}^j) = \mathbf{X}_3'\boldsymbol{\Lambda}_3^j + \mathbf{D}_c^{-1}(\boldsymbol{\Lambda}^j)\mathbf{D}_{1c}(\boldsymbol{\Lambda}^j)\left(\mathbf{A} - \mathbf{M}\,\mathbf{P}(\boldsymbol{\Lambda}^j)\right)$$

Parameter estimates for $\boldsymbol{\Lambda}_1$, $\boldsymbol{\Lambda}_2$ and $\boldsymbol{\Lambda}_3$ are computed simultaneously using the iterative weighted least squares estimation algorithm below.

For $j = 0$,

1. Obtain initial estimates of $\boldsymbol{\Lambda}_1^j$, $\boldsymbol{\Lambda}_2^j$ and $\boldsymbol{\Lambda}_3^j$.

2. Compute

$$\mathbf{D}_a(\boldsymbol{\Lambda}_1^j, \boldsymbol{\Lambda}_2^j, \boldsymbol{\Lambda}_3^j) = \mathbf{D}_{1a}(\boldsymbol{\Lambda}^j)\mathbf{M}\mathbf{D}_{2a}(\boldsymbol{\Lambda}^j)$$

   for $\mathbf{D}_{1a}(\boldsymbol{\Lambda}^j)$ and $\mathbf{D}_{2a}(\boldsymbol{\Lambda}^j)$ as in (48). Compute
   $\mathbf{Z}_a(\boldsymbol{\Lambda}_1^j, \boldsymbol{\Lambda}_2^j, \boldsymbol{\Lambda}_3^j) = \mathbf{X}_1'\boldsymbol{\Lambda}_1^j + \mathbf{D}_a^{-1}(\boldsymbol{\Lambda}^j)\mathbf{D}_{1a}(\boldsymbol{\Lambda}^j)\left(\mathbf{A} - \mathbf{M}\,\mathbf{P}(\boldsymbol{\Lambda}^j)\right)$

3. Using these estimates, compute $\boldsymbol{\Lambda}_1^{j+1}$ using (54).

4. Compute

$$\mathbf{D}_b(\boldsymbol{\Lambda}_1^{j+1}, \boldsymbol{\Lambda}_2^j, \boldsymbol{\Lambda}_3^j) = \mathbf{D}_{1b}(\boldsymbol{\Lambda}_1^{j+1}, \boldsymbol{\Lambda}_2^j, \boldsymbol{\Lambda}_3^j)\mathbf{M}\mathbf{D}_{2b}(\boldsymbol{\Lambda}_1^{j+1}, \boldsymbol{\Lambda}_2^j, \boldsymbol{\Lambda}_3^j)$$

for $\mathbf{D}_{1b}(\mathbf{\Lambda}_1^{j+1},\mathbf{\Lambda}_2^{j},\mathbf{\Lambda}_3^{j})$ and $\mathbf{D}_{2b}(\mathbf{\Lambda}_1^{j+1},\mathbf{\Lambda}_2^{j},\mathbf{\Lambda}_3^{j})$ using (49). Compute

$\mathbf{Z}_b(\mathbf{\Lambda}_1^{j+1},\mathbf{\Lambda}_2^{j},\mathbf{\Lambda}_3^{j}) =$

$\mathbf{X}_2'\mathbf{\Lambda}_2^{j} + \mathbf{D}_b^{-1}(\mathbf{\Lambda}_1^{j+1},\mathbf{\Lambda}_2^{j},\mathbf{\Lambda}_3^{j})\mathbf{D}_{1b}(\mathbf{\Lambda}_1^{j+1},\mathbf{\Lambda}_2^{j},\mathbf{\Lambda}_3^{j})\left(\mathbf{A} - \mathbf{M}\,\mathbf{P}(\mathbf{\Lambda}_1^{j+1},\mathbf{\Lambda}_2^{j},\mathbf{\Lambda}_3^{j})\right)$

5. Using these estimates, compute $\mathbf{\Lambda}_2^{j+1}$ using (55).

6. Compute

$$\mathbf{D}_c(\mathbf{\Lambda}_1^{j+1},\mathbf{\Lambda}_2^{j+1},\mathbf{\Lambda}_3^{j}) = \mathbf{D}_{1c}(\mathbf{\Lambda}_1^{j+1},\mathbf{\Lambda}_2^{j+1},\mathbf{\Lambda}_3^{j})\mathbf{M}\mathbf{D}_{2c}(\mathbf{\Lambda}_1^{j+1},\mathbf{\Lambda}_2^{j+1},\mathbf{\Lambda}_3^{j})$$

for $\mathbf{D}_{1c}(\mathbf{\Lambda}_1^{j+1},\mathbf{\Lambda}_2^{j+1},\mathbf{\Lambda}_3^{j})$ and $\mathbf{D}_{2c}(\mathbf{\Lambda}_1^{j+1},\mathbf{\Lambda}_2^{j+1},\mathbf{\Lambda}_3^{j})$ using (50). Compute

$\mathbf{Z}_c(\mathbf{\Lambda}_1^{j+1},\mathbf{\Lambda}_2^{j+1},\mathbf{\Lambda}_3^{j}) =$

$\mathbf{X}_3'\mathbf{\Lambda}_3^{j} + \mathbf{D}_c^{-1}(\mathbf{\Lambda}_1^{j+1},\mathbf{\Lambda}_2^{j+1},\mathbf{\Lambda}_3^{j})\mathbf{D}_{1c}(\mathbf{\Lambda}_1^{j+1},\mathbf{\Lambda}_2^{j+1},\mathbf{\Lambda}_3^{j})\left(\mathbf{A} - \mathbf{M}\,\mathbf{P}(\mathbf{\Lambda}_1^{j+1},\mathbf{\Lambda}_2^{j+1},\mathbf{\Lambda}_3^{j})\right)$

7. Using these estimates, compute $\mathbf{\Lambda}_3^{j+1}$ using (56).

8. Repeat steps $2-7$ until

$$\| \mathbf{\Lambda}_1^{j+1} - \mathbf{\Lambda}_1^{j} \| + \| \mathbf{\Lambda}_2^{j+1} - \mathbf{\Lambda}_2^{j} \| + \| \mathbf{\Lambda}_3^{j+1} - \mathbf{\Lambda}_3^{j} \| \leq tolerance$$

9. On convergence, the values of $\mathbf{\Lambda}_1^{j+1},\mathbf{\Lambda}_2^{j+1},\mathbf{\Lambda}_3^{j+1}$ are the iterated weighted least squares estimates of $\mathbf{\Lambda}_1,\mathbf{\Lambda}_2,\mathbf{\Lambda}_3$.

CHAPTER 5

ESTIMATION OF PARAMETERS OF THE WEIGHT MODEL

For the case of 3 multinomial outcomes, we assume the vector of weights $\mathbf{W}_{ij}$ has a multivariate normal distribution with mean vector $\boldsymbol{\mu}_{ij}$ and covariance matrix $\boldsymbol{\Sigma}_{ij}$. Let $\boldsymbol{\theta}_i' = \left(\mu_{1i}, \mu_{2i}, \mu_{3i}, \sigma_{1i}^2, \sigma_{2i}^2, \sigma_{3i}^2, \rho_{1i}, \rho_{2i}, \rho_{3i}, \rho_{12i}, \rho_{13i}, \rho_{23i}\right)$ be the parameters in the ith dose group as defined in section (2.2), for $i = 1, \ldots, g$ and $\boldsymbol{\theta}' = (\boldsymbol{\theta}_1', \ldots, \boldsymbol{\theta}_g')$. The ith dose group has a total of m_i litters. The kernel for the log likelihood function for $\boldsymbol{\theta}$ is given by

$$\ell(\boldsymbol{\theta}) \propto -\frac{1}{2}\sum_{i=1}^{g}\sum_{j=1}^{m_i}\log|\boldsymbol{\Sigma}_{ij}| - \frac{1}{2}\sum_{i=1}^{g}\sum_{j=1}^{m_i}(\mathbf{W}_{ij}-\boldsymbol{\mu}_{ij})'\,\boldsymbol{\Sigma}_{ij}^{-1}(\mathbf{W}_{ij}-\boldsymbol{\mu}_{ij}). \tag{57}$$

With the trinary outcomes given, we estimate the mean vector $\boldsymbol{\mu}_{ij}$ under various assumptions about the structure of the covariance matrix. First, a dose-dependent exchangeable covariance structure for all litters in each dose group is considered. Secondly, a common exchangeable correlation structure for all dose groups is assumed.

5.1 Parameter Estimation With Common Exchangeable Covariance Matrix

A dose-response model for weight is specified whereby, following George et al. [17], we simplify the covariance structure by assuming a common exchangeable correlation matrix across each dose group. The common correlation parameter for the weight within a litter is $\rho_{1i} = \rho_{2i} = \rho_{3i} = \rho_{12i} = \rho_{13i} = \rho_{23i} = \rho$, and the common variance for the weights is $\sigma_{1i}^2 = \sigma_{2i}^2 = \sigma_{3i}^2 = \sigma^2$. This simplifying assumption ascribes weight dose response only to the mean function.

$$\mathbf{W}_{ij} \mid \mathbf{x}_{ij} \sim N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_{ij}) \tag{58}$$

where $\boldsymbol{\Sigma}_{ij} = \sigma^2\{(1-\rho)\mathbf{I}_{n_{ij}} + \rho\mathbf{J}_{n_{ij}}\}$, for $j = 1, 2, \ldots, m_i,\ i = 1, 2, \ldots, g$, and $\boldsymbol{\mu}_i = (\mu_{1i}, \mu_{2i}, \mu_{3i})'$. We model mean weight parameters, as in George at al. [17], by

$$E(\mathbf{W}_{ij} \mid \mathbf{X}_{ij} = \mathbf{x}_{ij}) = \mathbf{Q}_{ij}\mathbf{D}_{ij}\boldsymbol{\eta} \tag{59}$$

where

$$\mathbf{Q}_{ij} = \begin{bmatrix} \mathbf{1}_{r_{ij}} & \mathbf{0}_{r_{ij}} & \mathbf{0}_{r_{ij}} \\ \mathbf{0}_{s_{ij}} & \mathbf{1}_{s_{ij}} & \mathbf{0}_{s_{ij}} \\ \mathbf{0}_{n_{ij}-r_{ij}-s_{ij}} & \mathbf{0}_{n_{ij}-r_{ij}-s_{ij}} & \mathbf{1}_{n_{ij}-r_{ij}-s_{ij}} \end{bmatrix}, \mathbf{D}_{ij} = \begin{bmatrix} 1 & d_i & \tau_{10} & \tau_{10}d_i \\ 1 & d_i & \tau_{01} & \tau_{01}d_i \\ 1 & d_i & 1-\tau_{10}-\tau_{01} & (1-\tau_{10}-\tau_{01})d_i \end{bmatrix}$$

and $\boldsymbol{\eta}' = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)$. This model is specifically chosen as it relates the malformation and death endpoints to dose using the $\tau_{r,s}$ parameter. Moreover, $\mathbf{1}_p$ is a $p \times 1$ vector of 1s and $\mathbf{0}_p$ is a $p \times 1$ vector of 0s. The conditional likelihood function for the weight data is obtained as

$$L(\boldsymbol{\eta}) = -\frac{1}{2}\sum_{i=1}^{g}\sum_{j=1}^{m_i}\left[n_{ij}\log\sigma_i^2 + (n_{ij}-1)\log(1-\rho) + \log\phi_{ij} + (\mathbf{W}_{ij} - \mathbf{Q}_{ij}\mathbf{D}_{ij}\boldsymbol{\eta})'\boldsymbol{\Sigma}_{ij}^{-1}(\mathbf{W}_{ij} - \mathbf{Q}_{ij}\mathbf{D}_{ij}\boldsymbol{\eta})\right] \tag{60}$$

where $\phi_{ij} = \{1 + (n_{ij}-1)\rho\}$. Taking derivatives of (60) with respect to $\boldsymbol{\eta}, \sigma^2$ and ρ and setting them to zero, we obtain the estimating equations for the MLEs of the parameters as follows:

$$\hat{\boldsymbol{\eta}} = \left[\sum_{i=1}^{g}\sum_{j=1}^{m_i}(\mathbf{D}_{ij}'\mathbf{Q}_{ij}'\hat{\boldsymbol{\Sigma}}_{ij}^{-1}\mathbf{Q}_{ij}\mathbf{D}_{ij})\right]^{-1}\sum_{i=1}^{g}\sum_{j=1}^{m_i}(\mathbf{D}_{ij}'\mathbf{Q}_{ij}'\hat{\boldsymbol{\Sigma}}_{ij}^{-1}\mathbf{W}_{ij}), \tag{61}$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{g}\sum_{j=1}^{m_i}(\mathbf{W}_{ij} - \mathbf{Q}_{ij}\mathbf{D}_{ij}\hat{\boldsymbol{\eta}})'\left[\mathbf{I}_{ij} - \hat{\rho}\frac{\mathbf{J}_{ij}}{\phi_{ij}}\right](\mathbf{W}_{ij} - \mathbf{Q}_{ij}\mathbf{D}_{ij}\hat{\boldsymbol{\eta}})}{N(1-\hat{\rho})} \tag{62}$$

$$N\sum_{i=1}^{g}\sum_{j=1}^{m_i}\{\hat{B}_{ij} - n_{ij}\hat{b}_{ij}[1+(n_{ij}-1)\hat{\rho}^2]\} = \hat{\rho}\sum_{i=1}^{g}\sum_{j=1}^{m_i}\left[n_{ij}(n_{ij}-1)/\hat{\phi}_{ij}\right]\sum_{i=1}^{g}\sum_{j=1}^{m_i}\left(\hat{B}_{ij} - \hat{\rho}n_{ij}\hat{b}_{ij}/\hat{\phi}_{ij}\right) \tag{63}$$

where $N = \sum_{i=1}^{g}\sum_j n_{ij}$ and

$$\hat{B}_{ij} = (\mathbf{W}_{ij} - \mathbf{Q}_{ij}\mathbf{D}_{ij}\hat{\boldsymbol{\eta}})'(\mathbf{W}_{ij} - \mathbf{Q}_{ij}\boldsymbol{D}_{ij}\hat{\boldsymbol{\eta}}),$$

$$\hat{b}_{ij} = (\mathbf{W}_{ij} - \mathbf{Q}_{ij}\mathbf{D}_{ij}\hat{\boldsymbol{\eta}})'\mathbf{J}_{n_{ij}}(\boldsymbol{W}_{ij} - \mathbf{Q}_{ij}\mathbf{D}_{ij}\hat{\boldsymbol{\eta}})/n_{ij},$$

To solve (61) and (62), an initial guess (for example 0) is used for ρ. With this, the first set of values of σ^2 and $\boldsymbol{\eta}$ are obtained and used in (63) to obtain new values of ρ's. This maximum likelihood estimate procedure is implemented iteratively until it converges.

5.2 Estimations With Dose-Dependent Exchangeable Correlation Matrix

In this model, we assume for dose group i, $(i = 1, \ldots, g)$,

$\rho_{1i} = \rho_{2i} = \rho_{3i} = \rho_{12i} = \rho_{13i} = \rho_{23i} = \rho_i$ and $\sigma_{1i}^2 = \sigma_{2i}^2 = \sigma_{3i}^2 = \sigma_i^2$. In this case, $\mathbf{\Sigma}_{ij} = \sigma_i^2[(1-\rho_i)\mathbf{I}_{n_{ij}} + \rho_i \mathbf{J}_{n_{ij}}]$, where $\mathbf{I}_p$ is a $p \times p$ identity matrix, $\mathbf{J}_p$ is a $p \times p$ matrix of 1s for $j = 1, 2, \ldots, m_i$. $i = 1, 2, \ldots, g$ and $\boldsymbol{\mu}_i = (\mu_{1i}, \mu_{2i}, \mu_{3i})'$. Using formulas from George et al. [17] and Seber [33], we have

$$|\mathbf{\Sigma}_{ij}| = \sigma_i^{2n_{ij}}(1-\rho_i)^{n_{ij}-1}\phi_{ij}, \tag{64}$$

$$\mathbf{\Sigma}_{ij}^{-1} = \frac{1}{\sigma_i^2(1-\rho_i)}\left(\mathbf{I}_{n_{ij}} - \frac{\rho_i}{\phi_{ij}}\mathbf{J}_{n_{ij}}\right). \tag{65}$$

We specify the dose-response model mean weight parameters as in section 5.1.

$$E(\mathbf{W}_{ij} \mid \mathbf{X}_{ij} = \mathbf{x}_{ij}) = \mathbf{Q}_{ij}\mathbf{D}_{ij}\boldsymbol{\eta} \tag{66}$$

where

$$\mathbf{Q}_{ij} = \begin{bmatrix} \mathbf{1}_{r_{ij}} & \mathbf{0}_{r_{ij}} & \mathbf{0}_{r_{ij}} \\ \mathbf{0}_{s_{ij}} & \mathbf{1}_{s_{ij}} & \mathbf{0}_{s_{ij}} \\ \mathbf{0}_{n_{ij}-r_{ij}-s_{ij}} & \mathbf{0}_{n_{ij}-r_{ij}-s_{ij}} & \mathbf{1}_{n_{ij}-r_{ij}-s_{ij}} \end{bmatrix}, \mathbf{D}_{ij} = \begin{bmatrix} 1 & d_i & \tau_{10} & \tau_{10}d_i \\ 1 & d_i & \tau_{01} & \tau_{01}d_i \\ 1 & d_i & 1-\tau_{10}-\tau_{01} & (1-\tau_{10}-\tau_{01})d_i \end{bmatrix}$$

and $\boldsymbol{\eta}' = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)$. The conditional likelihood function for the weight data is obtained as

$$L(\boldsymbol{\eta}) = -\frac{1}{2}\sum_{i=1}^{g}\sum_{j=1}^{m_i}\left[n_{ij}\log\sigma_i^2 + (n_{ij}-1)\log(1-\rho_i) + \log\phi_{ij} + (\mathbf{W}_{ij} - \mathbf{Q}_{ij}\mathbf{D}_{ij}\boldsymbol{\eta})'\mathbf{\Sigma}_{ij}^{-1}(\mathbf{W}_{ij} - \mathbf{Q}_{ij}\mathbf{D}_{ij}\boldsymbol{\eta})\right] \tag{67}$$

where $\phi_{ij} = \{1 + (n_{ij}-1)\rho_i\}$. Taking derivatives of (67) with respect to $\boldsymbol{\eta}, \sigma_i^2$ and ρ_i and setting them to zero, we obtain similar results to those of Section 5.1. A similar iterative scheme to the one described in Section 5.1 is implemented for obtaining the desired Maximum Likelihood Estimates.

CHAPTER 6

APPLICATION

6.1 Data Summary

We present an application of the multinomial exchangeable models by analyzing data from a developmental toxicity study conducted by the National Toxicology Program involving a study of ethylene glycol (EG) using CD-1 mice. The data was obtained from an experiment in which pregnant females were randomly assigned to one of four dose groups and subsequently exposed to the compound EG. The mice were administered dose levels of 0, 750, 1500 and 3000 mg/kg in the four dose groups. The number of dead and living fetuses was recorded within each litter. The data summarized in Table 1 is obtained from http://ntp.niehs.nih.gov. The live fetuses were also examined for several developmental endpoints including fetal weight and other visceral malformations. For our application we focus on analysis of malformation, which is the discrete endpoint, and weight which is the continuous endpoint collected on each fetus. Table 2 is essentially a summary of the general discrete data from table 1.

The summary data in table 3 shows the observed mean weights, standard deviations and Intra-Littter Correlations of fetuses with their respective standard errors in each dose level for the ethylene glycol data. This gives a preliminary indication of the dose effect on the response outcomes; increase in dose levels leads to corresponding increase in the average number of malformed and dead fetuses while the average weight of both malformed and dead fetuses decrease with increase in dose level. Though not expected, the average weight of the normal fetuses also appears to decrease with increase in dose level. The standard deviation and Intra-Litter Correlations however do not seem to be affected by the dose level for the weights across all response types.

Table 1: Discrete data for response of CD-1 Mice following exposure to ethylene glycol

Dose 1 (0 mg/kg)	Dose 2 (750 mg/kg)	Dose 3 (1500 mg/kg)	Dose 4 (3000 mg/kg)
$(9,1,0)_1$	$(7,0,2)_1$	$(3,3,0)_1$	$(9,2,2)_1$
$(9,2,0)_1$	$(9,2,1)_1$	$(5,4,0)_1$	$(9,2,4)_1$
$(11,0,0)_1$	$(10,0,1)_1$	$(7,4,1)_1$	$(9,5,3)_1$
$(11,1,0)_2$	$(11,0,2)_1$	$(8,0,2)_1$	$(10,1,0)_1$
$(11,5,0)_1$	$(11,1,3)_1$	$(9,2,0)_1$	$(10,6,1)_1$
$(12,0,0)_2$	$(11,2,1)_1$	$(9,2,6)_1$	$(11,2,5)_1$
$(12,1,0)_1$	$(12,0,1)_1$	$(10,0,8)_1$	$(12,0,1)_1$
$(13,0,0)_2$	$(12,1,0)_1$	$(12,0,0)_1$	$(12,0,4)_1$
$(13,1,0)_1$	$(13,0,3)_1$	$(12,1,3)_1$	$(12,1,5)_1$
$(13,3,0)_1$	$(13,1,0)_1$	$(12,1,6)_1$	$(12,2,7)_1$
$(14,0,0)_1$	$(13,1,1)_1$	$(12,2,4)_1$	$(12,2,9)_2$
$(14,2,0)_1$	$(13,2,1)_1$	$(12,3,6)_1$	$(12,5,2)_1$
$(15,0,0)_1$	$(13,5,1)_1$	$(13,0,1)_1$	$(13,1,3)_1$
$(15,1,0)_1$	$(14,0,0)_1$	$(13,0,7)_1$	$(13,1,4)_1$
$(15,2,0)_3$	$(14,1,3)_1$	$(13,1,1)_1$	$(13,1,5)_1$
$(15,5,0)_1$	$(14,2,1)_1$	$(13,2,8)_1$	$(14,2,12)_1$
$(16,0,1)_1$	$(14,4,0)_1$	$(14,1,3)_1$	$(14,4,9)_1$
$(16,2,0)_2$	$(15,0,0)_3$	$(14,2,5)_1$	$(15,3,3)_1$
$(17,3,0)_1$	$(15,1,3)_1$	$(14,3,5)_1$	$(15,3,12)_1$
	$(15,5,1)_1$	$(15,0,10)_1$	$(16,2,11)_1$
	$(18,5,1)_1$	$(15,2,0)_1$	$(16,9,7)_1$
		$(15,2,11)_1$	
		$(16,2,2)_1$	

The notation $(n,r,s)_t$ represents t litters of size n with r malformations and s deaths.

Table 2: Summary of response following exposure to ethylene glycol

Dose Group	1	2	3	4
Exposure Level	0	750	1500	3000
No. of Litters	24	23	23	22
No. of Fetuses	321	297	275	271
No. of Dead	36	33	37	56
No. of Malformed	1	26	89	118

Table 3: Summary of observed Mean weights with Standard Deviations for Ethylene-Glycol Data

Dose Level	Malformation Weight Mean (SE)	Death Weight Mean (SE)	Normal Weight Mean (SE)	Standard Deviation STDev (SE)	Intra-Litter Correlation Corr (SE)
0	0.9715 (0.1822)	0.5310 (0.1778)	1.0710 (0.0976)	0.1197 (0.0246)	0.4998 (0.0549)
750	0.8090 (0.1244)	0.4626 (0.3038)	0.8804 (0.1003)	0.1230 (0.0271)	0.3319 (0.0463)
1500	0.7067 (0.0977)	0.3340 (0.2361)	0.7950 (0.0983)	0.1153 (0.0257)	0.3990 (0.0388)
3000	0.6678 (0.1192)	0.3462 (0.2592)	0.7479 (0.1188)	0.1424 (0.0320)	0.3404 (0.0271)

6.2 Computed Parameter Estimates

For the malformation and death endpoints, we consider the regression model in equation (68) for $\tau_{r,s}^{(i)}$ where $\lambda_1 = f(\boldsymbol{\beta}_1' x) = \beta_{1_1} + \beta_{1_2} d_i$, $\lambda_2 = f(\boldsymbol{\beta}_2' x) = \beta_{2_1} + \beta_{2_2} d_i$ and $\lambda_3 = f(\boldsymbol{\beta}_3' x) = \beta_{3_1} + \beta_{3_2} d_i$. We choose all litters in the ith dose group:

$$F_{r,s}^{(i)}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3) = \tau_{r,s}^{(i)}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3) = \frac{(\lambda + r + s)(\lambda_1 + \lambda_3)(\lambda_2 + \lambda_3) + \lambda_3 rs}{(\lambda + r + s)(\lambda_1 + \lambda_3 + r)(\lambda_2 + \lambda_3 + s)} \tag{68}$$

where $\boldsymbol{\beta}_1' = (\beta_{1_1}, \beta_{1_2})$, $\boldsymbol{\beta}_2' = (\beta_{2_1}, \beta_{2_2})$, $\boldsymbol{\beta}_3' = (\beta_{3_1}, \beta_{3_2})$. $\lambda_1, \lambda_2, \lambda_3 \geq 0, \lambda_1 + \lambda_3 > 0$, $\lambda_2 + \lambda_3 > 0$ and $\lambda = \lambda_1 + \lambda_2 + \lambda_3$. As discussed in section 4.2, the chosen form for the parameters of the model are sufficient to guarantee that the $\tau_{r,s}$'s form a valid probability model. The parameter estimates for the EG data are as given in table 11.

Table 4: Estimates and standard errors of the Parameters for Malformation and Death Rate

Parameter	Estimate	Standard Error
$\hat{\beta}_{1_1}$	0.73534	0.08295
$\hat{\beta}_{1_2}$	0.00004	0.00001
$\hat{\beta}_{2_1}$	1.80526	0.91844
$\hat{\beta}_{2_2}$	-0.00083	0.00017
$\hat{\beta}_{3_1}$	1.95751	0.92332
$\hat{\beta}_{3_2}$	-0.00046	0.00009

Using the model in equation (68), the estimates of marginal response probabilities of malformation and death are obtained by evaluating equations (69) and (70) respectively at $\hat{\boldsymbol{\beta}}_1'$, $\hat{\boldsymbol{\beta}}_2'$ and $\hat{\boldsymbol{\beta}}_3'$.

$$\tau_{1,0}^{(i)}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_3) = \frac{\left[(\beta_{1_1} + \beta_{1_2} d_i) + (\beta_{3_1} + \beta_{3_2} d_i)\right]}{\left[(\beta_{1_1} + \beta_{1_2} d_i) + (\beta_{3_1} + \beta_{3_2} d_i) + 1)\right]} \tag{69}$$

$$\tau_{0,1}^{(i)}(\boldsymbol{\beta}_2, \boldsymbol{\beta}_3) = \frac{\left[(\beta_{2_1} + \beta_{2_2} d_i) + (\beta_{3_1} + \beta_{3_2} d_i)\right]}{\left[(\beta_{2_1} + \beta_{2_2} d_i) + (\beta_{3_1} + \beta_{3_2} d_i) + 1)\right]} \tag{70}$$

In table 5, we compare the observed rate of malformation and death with the values predicted by the model. Figure **??** shows a 3- Dimensional view of the marginal estimates of both the death and malformation rates.

Table 5: Estimates of the Death and Malformation Rate under the Exchangeable Multinomial Model

	Dose Group	Estimate (Standard Error)	Observed Rate
Malformation	1	.0044 (0.0082)	.0030
	2	.0592 (0.0329)	.0875
	3	.2207 (0.0327)	.3346
	4	.4673 (0.0170)	.4354
Death	1	.1125 (0.0438)	.1081
	2	.1317 (0.0432)	.1111
	3	.1563 (0.0489)	.1391
	4	.2081 (0.0495)	.2066

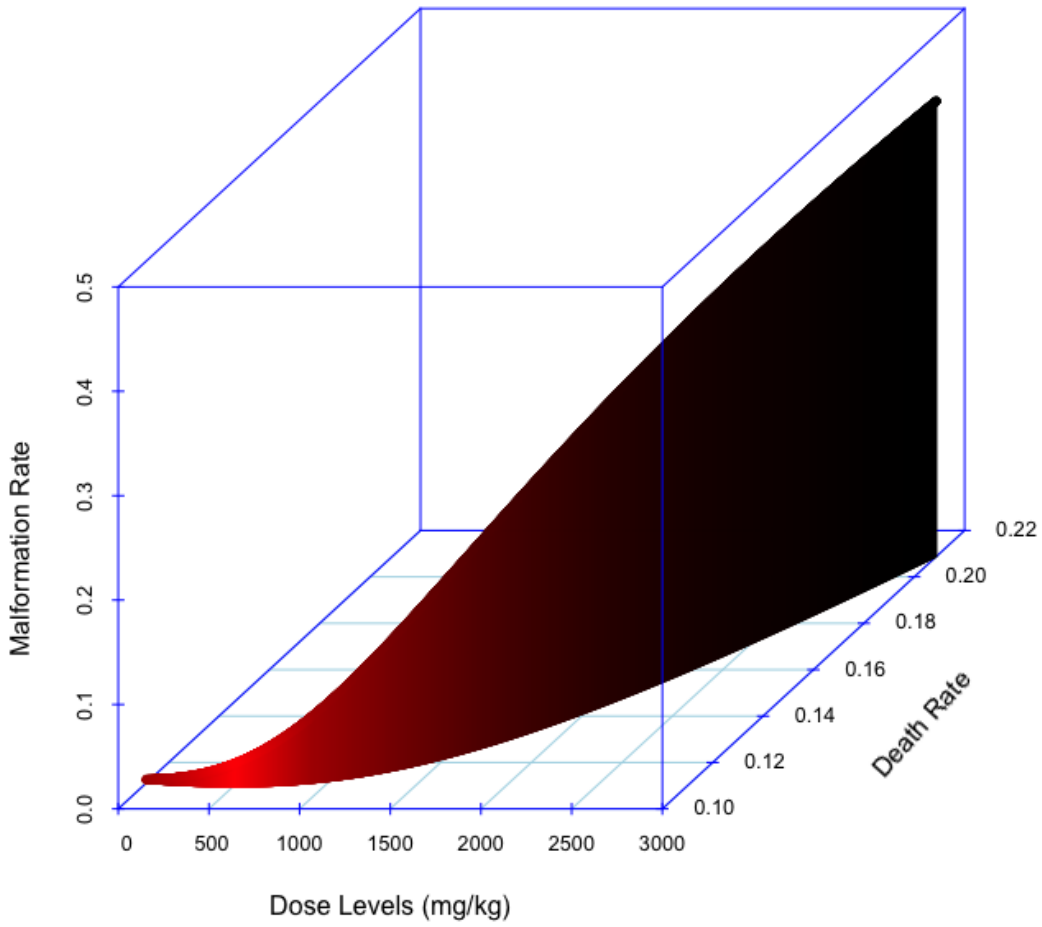

Figure 6.1: 3-D regression plot for malformation and Death Rates

To interpret this empirical result, it is important to recognize that for each dose group, the estimate of the malformation rate and death rate depends only on the marginal probabilities $\tau_{1,0}^{(i)}$ for the malformation endpoints in the *i*th dose group and $\tau_{0,1}^{(i)}$ for the death endpoints in the *i*th dose group. The malformation and death rates from the estimates are quite comparable to the observed rates confirming the results of the observations in summary data given in table 2.

Next we fit the weight data jointly using the regression model discussed in the Chapter 5 which includes the dose, malformation, death and normal parameters, as well as the interaction between the dose and discrete endpoints. The model paramteres are estimated using the dose-dependent exchangeable covariance structure for all litters in each dose group. The regression models with dose groups fitted jointly would appear as in equations (71), (72) and (73).

$$\hat{\mu}_{1_i} = \hat{\alpha}_1 + \hat{\alpha}_2 d_i + \hat{\alpha}_3 \hat{\tau}_{1,0} + \hat{\alpha}_4 \hat{\tau}_{1,0} d_i \tag{71}$$

$$\hat{\mu}_{2_i} = \hat{\alpha}_1 + \hat{\alpha}_2 d_i + \hat{\alpha}_3 \hat{\tau}_{0,1} + \hat{\alpha}_4 \hat{\tau}_{0,1} d_i \tag{72}$$

$$\hat{\mu}_{3_i} = \hat{\alpha}_1 + \hat{\alpha}_2 d_i + \hat{\alpha}_3 (1 - \hat{\tau}_{1,0} - \hat{\tau}_{0,1}) + \hat{\alpha}_4 (1 - \hat{\tau}_{1,0} - \hat{\tau}_{0,1}) d_i \tag{73}$$

Table 6: Estimates of the weight parameters and standard errors with dose response functions

Parameter	Estimate	Standard Error	P-Value
Intercept $(\hat{\alpha_1})$	0.8391	0.1252	$<.001$
Dose $(\hat{\alpha_2})$	-0.0002	0.0001	$<.001$
$\tau_{r,s}$ $(\hat{\alpha_3})$	-1.3488	0.2630	< 0.001
$\tau_{r,s} \times$ Dose $(\hat{\alpha_4})$	0.0007	0.0006	.013

In Table 10, a summary of parameter estimates, standard errors and P-values from the joint estimation of the model is provided. A negative estimate of α_2, $\hat{\alpha}_2 = (-0.0002)$, with significant P-value of less than 0.001 provides further evidence of association of lower fetal weight with malformed fetuses for Ethylene Glycol. Both the $\tau_{r,s}$ $(\hat{\alpha_3})$ and the interaction parameter $\tau_{r,s} \times$ Dose $(\hat{\alpha_4})$are found to be significant at level 0.05, with the $\tau_{r,s}(\hat{\alpha_3})$ having a P-Value of less than 0.001 and the $\tau_{r,s} \times$ Dose $(\hat{\alpha_4})$ parameter having a

P-Value of 0.013. This suggests that the regression model of mean weights proposed is highly dependent on the values of $\tau_{r,s}$, dose level and the iteration term between $\tau_{r,s}$ and dose level.

Table 7 shows the Joint estimates of the mean weight, standard deviation and intra-cluster correlation parameters for the EG data for each given dose level. The estimates of the mean weight for the normal fetuses, μ_3 given in table 7 are almost uniformly bigger than those for malformed fetuses μ_1 which are in turn also bigger than those of the dead fetuses μ_2. From table 7, it is evident that the estimates of the MLEs are quite comparable to the observed parameters in table 3.

Table 7: Estimates of the weight parameters with dose group fitted Jointly with the Exchangeable Regression Model

Dose Level	$\hat{\mu}_1$ (SE)	$\hat{\mu}_2$ (SE)	$\hat{\mu}_3$ (SE)	$\hat{\sigma}$ (SE)	$\hat{\rho}$ (SE)
0	0.8332 (0.0764)	0.6874 (0.0619)	0.9853 (0.0812)	0.1125 (0.0154)	0.4431 (0.04079)
750	0.6857 (0.0564)	0.5778 (0.0686)	0.8008 (0.0869)	0.1277 (0.0298)	0.3405 (0.0372)
1500	0.5894 (0.0473)	0.4860 (0.0312)	0.7043 (0.0721)	0.1118 (0.0292)	0.3532 (0.0322)
3000	0.5613 (0.0430)	0.3799 (0.0444)	0.6411 (0.0595)	0.1424 (0.0112)	0.3647 (0.0407)

Fig. 6.2 gives plots of fitted mean weight curves for the normal, malformed and dead fetuses in comparison to the observed average weights for the malformed fetuses. Examination of the graph shows considerable similarity between the estimated and observed mean weights. The plot supports the preliminary observations from the summary data concerning declining fetal weight with dose level among the normal, malformed and dead fetuses alike.

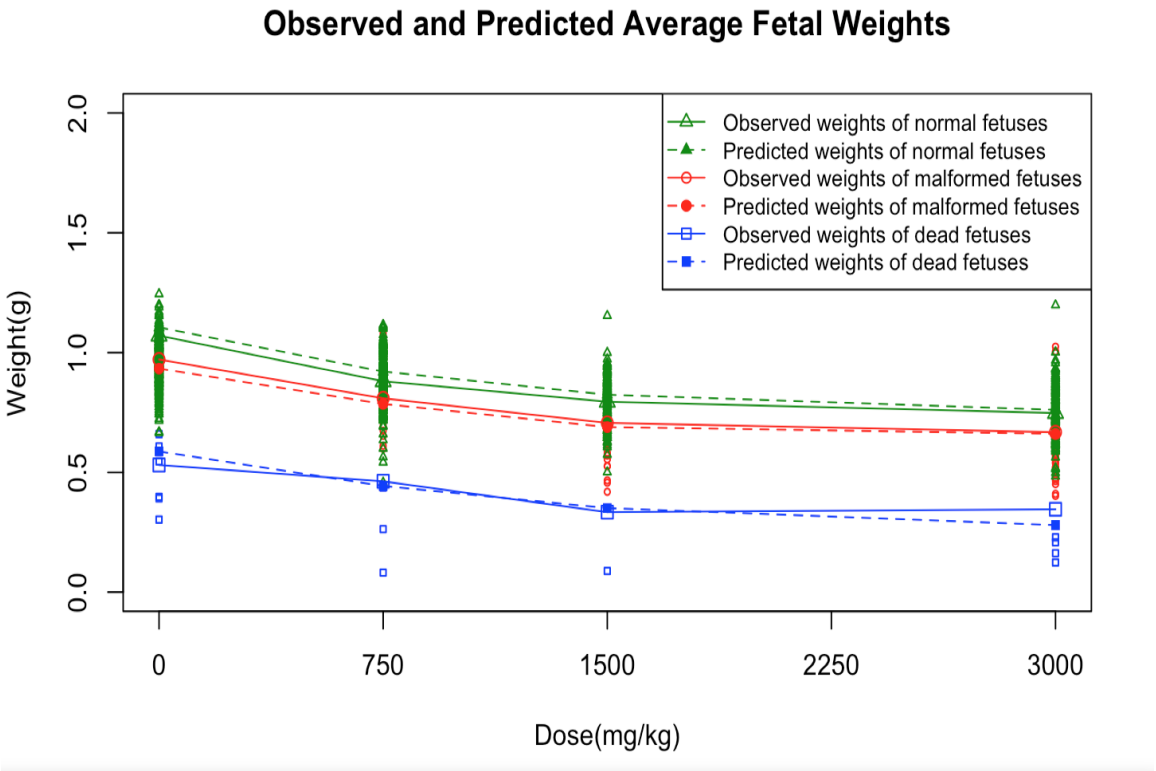

Figure 6.2: Graphs of observed and estimated fetal weights

6.3 Joint Quantitative Risk Assessment

A principal objective of the development of dose-response models for adverse effects in developmental toxicology is the determination of a safe level of exposure for humans. In the case where more than one end point is taken into consideration, we can choose either to examine the endpoints separately or jointly. A separate assessment of the endpoints assumes that a safe exposure level for the most sensitive endpoint will also result in a safe level for the other outcomes. This derivation is solely based on the endpoint that is most sensitive to the exposure. In contrast, the joint model approach bases its estimation on the correlation between the endpoints, such as weight and malformation or weight and death, to determine a safe dose level. Faes et al. [12], found that the correlation between endpoints can significantly affect the estimation of a safe dose level. For this reason, we will adopt the joint model approach to determine safe level of exposure.

The derivation of safe dose estimate requires the specification of an adverse event and a probability, $P(d)$, to represent the probability that a fetus exhibits this adverse event at dose level d. For example, in the joint model under consideration here, $P(d)$ may represent the probability that the kth fetus in litter j is either malformed, dead or has a low birth weight. Following Geys et al. [20] The cut off for low fetal birth weight, denoted W_c, can be specified as a certain number of standard deviations below the average weight in the control group that is low enough to be considered adverse. Define the probability of an adverse effect as

$$P(d) = P(W_{jk} < W_c \text{ or } X_{jk} \in {O_3}^c \mid d), \tag{74}$$

where W_{jk} is the weight of the kth fetus in the jth litter, W_c is the cut off for low birth weight, X_{jk} represents the discrete reponse of the kth fetus in the jth litter and ${O_3}^c$ is the complement of outcome O_3, that is, it is either outcome O_1, representing malformation or outcome O_2, representing death. One of the approaches proposed by Geys et al. [20] is the probit approach where the computational expression for $P(d)$, the probability of an

adverse effect at dose level d, is given as

$$
\begin{aligned}
P(d) &= P(W_{jk} < W_c \text{ or } X_{jk} = {O_3}^c \mid d) \\
&= P(W_{jk} < W_c) + P(X_{jk} = O_1) + P(X_{jk} = O_2) \\
&- P(W_{jk} < W_c \cap X_{jk} = O_1) - P(W_{jk} < W_c \cap X_{jk} = O_2) \\
&= P(W_{jk} < W_c \mid X_{jk} = O_1)P(X_{jk} = O_1) + P(W_{jk} < W_c \mid X_{jk} = O_2)P(X_{jk} = O_2) \\
&+ P(W_{jk} < W_c \mid X_{jk} = O_3)P(X_{jk} = O_3) + P(X_{jk} = O_1) + P(X_{jk} = O_2) \\
&- P(W_{jk} < W_c \mid X_{jk} = O_1)P(X_{jk} = O_1) - P(W_{jk} < W_c \mid X_{jk} = O_2)P(X_{jk} = O_2)) \\
&= P(W_{jk} < W_c \mid X_{jk} = O_3)P(X_{jk} = O_3) + P(X_{jk} = O_1) + P(X_{jk} = O_2) \\
&= \Phi\left(\frac{W_c - \mu_3(d)}{\sigma(d)}\right)\{1 - \tau_{1,0}(d) - \tau_{0,1}(d)\} + \tau_{1,0}(d) + \tau_{0,1}(d) \qquad (75)
\end{aligned}
$$

where $\Phi(\cdot)$ is the standard normal distribution function, $\mu_3(d)$ is the mean fetal weight for fetuses with a normal response outcome, that is neither malformed nor dead. The standard deviation of the fetal weights at dose level (d) is given as $\sigma(d)$, while $\tau_{1,0}(d)$ and $\tau_{0,1}(d)$ are the probabilities of malformation and death respectively at dose level d.

Using the probability $P(d)$, we can define the risk function, also called the excess risk, $r(d)$ as follows:

$$r(d) = \frac{P(d) - P(0)}{1 - P(0)} \qquad (76)$$

The purpose of the excess risk is to measure the increased response at dose level d over background risk in the control group with $d = 0$. To determine a safe exposure level, the excess risk function can be used to relate adverse outcomes to dose levels. One useful measure in determining safe level of exposure is the benchmark dose BMD_q, which is defined as the dose corresponding to a pre-specified small quantity q, where $q = 0.01\%,\ 0.1\%,\ 1\%$ or 5%, of increased response over background. We can represent the equation of excess risk as $r(d, \boldsymbol{\theta})$. The dose d that satisfies the equation of excess risk

$r(d,\boldsymbol{\theta})$ is the benchmark dose.

$$r(d,\boldsymbol{\theta}) = q. \tag{77}$$

The BMD_q can be estimated as a solution to $\hat{r}(d,\hat{\boldsymbol{\theta}}) = q$, where $\hat{\boldsymbol{\theta}}$ is the model estimate of the parameters. To acknowledge the fact that estimates rather than true known values are used to determine the BMD_q, it is customary to also compute its lower confidence limit which often replaces the estimated BMD_q. From the several approaches proposed on this, we implement the approach suggested by Kimmel and Gaylor [21]. This estimate is termed as the lower effective dose LED_q, and is the value of d that solves

$$q = \hat{r}(d,\hat{\boldsymbol{\theta}}) + 1.645\left\{ \hat{var}\{\hat{r}(d,\hat{\boldsymbol{\theta}})\} \right\}^{1/2} \tag{78}$$

where

$$\hat{var}\{\hat{r}(d,\hat{\boldsymbol{\theta}})\} = \left\{ \frac{\partial}{\partial \boldsymbol{\theta}} r(d) \right\}^{T} \hat{var}(\hat{\boldsymbol{\theta}}) \left\{ \frac{\partial}{\partial \boldsymbol{\theta}} r(d) \right\} \tag{79}$$

evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$.

Table 8: BMD and LED corresponding to q=10% excess risk for EG Data

Dose	Estimate
BMD_{10}	614
LED_{10}	586
Malformation BMD_{10}	732
Death BMD_{10}	808
Weight BMD_{10}	945

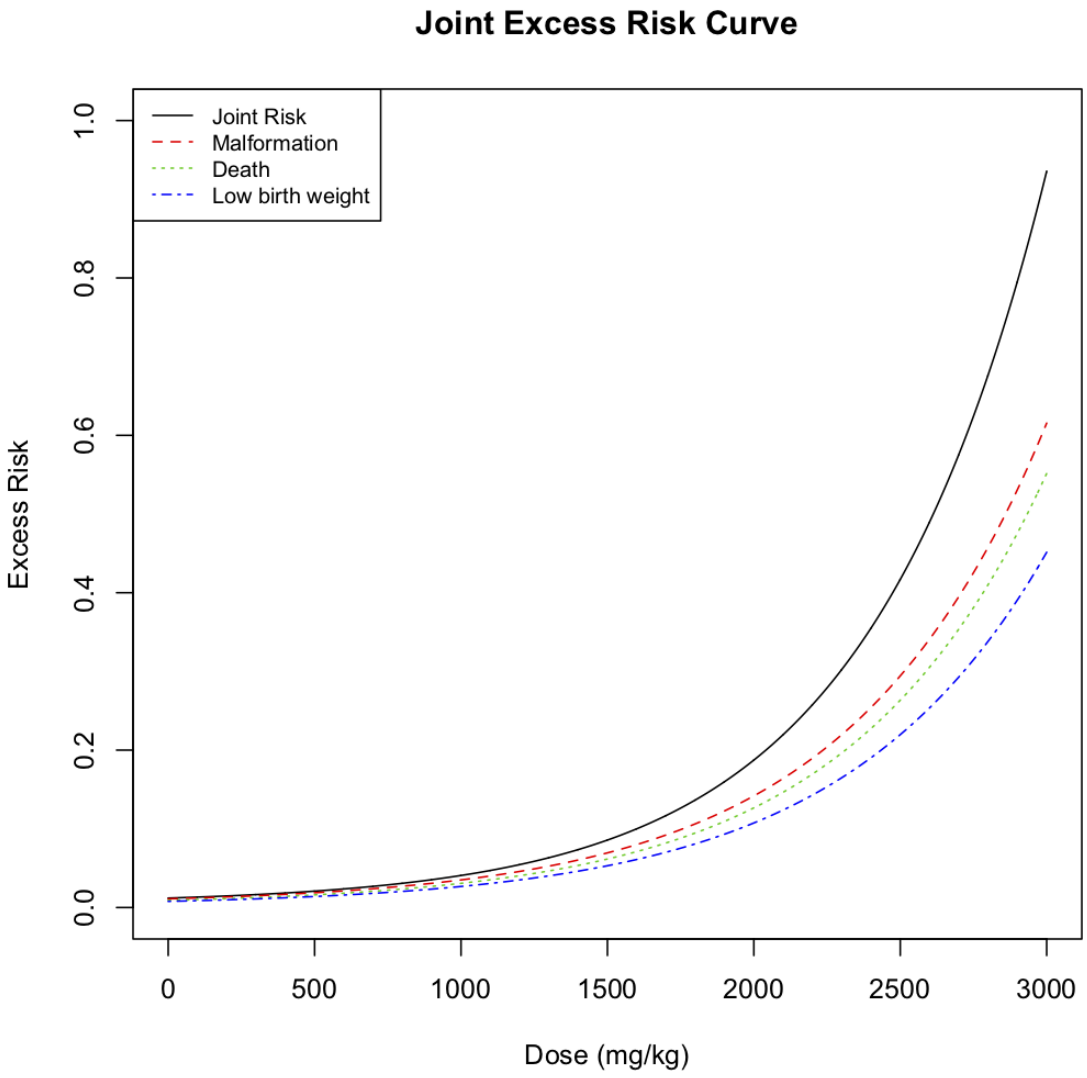

Figure 6.3: Graph of excess risk vs dose

Estimated BMDs and LEDs for the EG Data are shown in table 8 shows estimates for 10% excess risk over control. Following Geys et al. [20], the cut off value W_c used for low birth weight is the value two standard deviations (SDs) below the average birth weight in the control group, (that is where dose $= 0$). The cut off value W_c for the EG data is 0.5128 g. Included in table 8 for comparison are the estimated benchmark doses for each endpoint computed separately. A separate computation of benchmark dose involves basing the derivation of the safe exposure level on the endpoint that is most sensitive to the

exposure. This assumes that the safe dose for the most sensitive endpoint will be safe for other adverse outcomes. For the sake of these calculations, we assume the particular endpoint being considered, e.g. malformation as being most sensitive to the exposure.

The joint benchmark dose is more conservative than the malformation, death and low birth weight benchmark doses. In Figure 6.3, the additional risks of the joint, malformed, dead and low birth weight fetus as functions of dose are plotted. This example illustrates the importance of taking the relationship between the outcomes into account. The joint risk incorporates the relationship between the separate outcomes leading to a lower BMD_{10} than the individual outcomes separately.

6.4 Simulation Study

A simulation study was conducted to determine the effectiveness of the model in estimating parameters of the discrete model, $\boldsymbol{\tau}_{r,s}$ and the parameters of the continuous (weight) model $\boldsymbol{\theta}' = (\mu_1, \mu_2, \mu_3, \sigma, \rho)$ assuming a common exchangeable correlation matrix across each dose group. The MLEs of $\boldsymbol{\tau}_{r,s}$ were computed for $(r,s) = (1,0),(0,1)$. The following steps were used to generate the discrete data for the simulation.

1. The bivariate exponential distribution function of Marshall and Olkin [29] is used to generate the τ's by choosing appropriate values of $\lambda_i,\ i = 1,2,3$, that give the required values of $\tau_{r,s}$ to be estimated.

2. Using the exchangeable multinomial distribution (5), the probability of each (r,s) pair is calculated for a specific cluster of size n.

3. From the probability distribution we generate a cumulative distribution function (cdf).

4. To generate the (r,s) pair, a uniform $U(0,1)$ distribution is used to generate variates which are matched to the (r,s) pair by comparing the $U(0,1)$ variates to the cdf. Table 9 illustrates an example of how this pairing is done whereby the permutations

of clusters of size, $n = 2$ are matched with their corresponding probability and cdf. A random number is generated from a uniform $U(0,1)$ and if the value is between 0 and 0.1495 the (r,s) pair $(0,0)$ is chosen as a cluster. Again, for example, if the value of the uniform $U(0,1)$ is between 0.4694 and 0.6428 then the (r,s) pair $(1,1)$ will be chosen as the next cluster. This process is repeated until the required number of clusters is achieved.

Table 9: Probabilities and CDF for each r,s pair for data with cluster size of n=2

(r,s) Pair	Probability $p_{r,s}$	CDF
$(0,0)$	0.1495	0.1495
$(0,1)$	0.1602	0.3097
$(1,0)$	0.1597	0.4694
$(1,1)$	0.1734	0.6428
$(0,2)$	0.1550	0.7978
$(2,0)$	0.2022	1

We consider specific values of $\lambda_i,\ i = 1,2,3$, to estimate corresponding values of $\tau_{0,1}$ and $\tau_{1,0}$ such that we have the following three instances:

- When $\tau_{1,0} = 0.8$ and $\tau_{0,1} = 0.01$ to emulate instances of high malformation rate and low death rate.
- When $\tau_{1,0} = 0.01$ and $\tau_{0,1} = 0.8$ to emulate instances of low malformation rate and high death rate.
- When both $\tau_{1,0} = 0.5$ and $\tau_{0,1} = 0.5$ to emulate instances of equal malformation and death rate.

Moreover, we consider three scenarios with respect to the number of clusters $N = 10, 20, 30$. We also consider clusters of sizes $n = 5$ and 10, to see the effect that the

cluster size has on the estimates. With the generated data, we use the iterative weighted least squares procedure described in this thesis to come up with estimates of $\tau_{r,s}$. Table 10 and 11 show the $\tau_{r,s}$ estimates together with their respective standard errors. It was noted that the estimates slightly improve with increase in number of clusters N. The estimates also tend to improve with increase in cluster size n. The standard errors also get better with increase in cluster size, n and number of clusters, N.

Table 10: Simulation results for the discrete parameters with cluster sizes of $n = 5$ and number of clusters, N

Simulation Parameters	$\tau_{0,1}$	$\tau_{1,0}$
$\lambda_1 = 4,\ \lambda_2 = 0.005,\ \lambda_3 = 0.005$		
Actual	0.01	0.8
$N = 10$	0.0053 (0.0111)	0.6896 (0.0.1406)
$N = 20$	0.0072 (0.0082)	0.9493 (0.1012)
$N = 30$	0.0127 (0.0076)	0.8818 (0.0959)
$\lambda_1 = 0.005,\ \lambda_2 = 4,\ \lambda_3 = 0.005$		
Actual	0.8	0.01
$N = 10$	0.8201 (0.1218)	0.0066 (0.0129)
$N = 20$	0.7816 (0.1406)	0.0149 (0.0183)
$N = 30$	0.8144 (0.0977)	0.0160 (0.0098)
$\lambda_1 = 0.5,\ \lambda_2 = 0.5,\ \lambda_3 = 0.5$		
Actual	0.5	0.5
$N = 10$	0.3655 (0.3711)	0.5884 (0.2778)
$N = 20$	0.3872 (0.3257)	0.5707 (0.2611)
$N = 30$	0.4597 (0.3006)	0.5251 (0.2358)

Table 11: Simulation results for the discrete parameters with cluster sizes of $n = 10$ and number of clusters, N

Simulation Parameters	$\tau_{0,1}$	$\tau_{1,0}$
$\lambda_1 = 4,\ \lambda_2 = 0.005,\ \lambda_3 = 0.005$		
Actual	0.01	0.8
$N = 10$	0.0107 (0.0022)	0.8066 (0.0693)
$N = 20$	0.0089 (0.0014)	0.8082 (0.1003)
$N = 30$	0.0105 (0.0011)	0.7862 (0.0712)
$\lambda_1 = 0.005,\ \lambda_2 = 4,\ \lambda_3 = 0.005$		
Actual	0.8	0.01
$N = 10$	0.8035 (0.0243)	0.0098 (0.0255)
$N = 20$	0.7968 (0.0469)	0.0111 (0.0173)
$N = 30$	0.8015 (0.0392)	0.0101 (0.0083)
$\lambda_1 = 0.5,\ \lambda_2 = 0.5,\ \lambda_3 = 0.5$		
Actual	0.5	0.5
$N = 10$	0.4699 (0.3034)	0.5349 (0.1334)
$N = 20$	0.5320 (0.2628)	0.4610 (0.0893)
$N = 30$	0.5035 (0.0650)	0.4914 (0.1143)

Next we generate the continuous (weights) data by using the previous values of $\tau_{r,s}$ obtained from the simulated discrete data and choosing appropriate parameters α_1 and α_2, to where

$$\mu_1 = \alpha_{1_1} + \alpha_{2_1}\tau_{1,0} \tag{80}$$

$$\mu_2 = \alpha_{1_2} + \alpha_{2_2}\tau_{1,0} \tag{81}$$

$$\mu_3 = \alpha_{1_3} + \alpha_{2_3}(1 - \tau_{1,0} - \tau_{1,0}) \tag{82}$$

The values of μ_1, μ_2 and μ_3, representing the mean weights for malformed, dead and normal fetuses respectively, are hereafter used to generate random variables from the multivariate gaussian distribution in equation (58) with a common exchangeable covariance matrix with parameters σ and ρ. The same iterative procedure to estimate the continuous data in Chapter 5 is used to estimate the parameters $\boldsymbol{\theta}' = (\mu_1, \mu_2, \mu_3, \sigma, \rho)$. The estimates and corresponding standard errors for the weight data are recorded in tables 12 and 13. The results were estimated using fixed values of $\tau_{r,s}$ and number of clusters $N = 10, 20, 30$ of cluster sizes $n = 5$ and 10. It is noted that increase in number of clusters

and cluster size leads to estimates which are are much closer to the true estimates of the parameters and lower standard errors.

Table 12: Simulation results for weight parameters with cluster sizes of $n = 5$ and number of clusters, N

Simulation Parameters	μ_1	μ_2	μ_3	σ	ρ
$\tau_{1,0} = 0.8,\ \tau_{0,1} = 0.01$					
Actual	0.7	0.4	0.9	0.1	0.3
$N = 10$	0.7819 (0.2333)	0.4955 (0.2307)	0.8394 (0.3145)	0.1818 (0.0278)	0.2735 (0.1044)
$N = 20$	0.7765 (0.2005)	0.4646 (0.2315)	0.8801 (0.2829)	0.1375 (0.0212)	0.3443 (0.0997)
$N = 30$	0.7212 (0.1812)	0.4413 (0.2278)	0.8816 (0.2545)	0.1207 (0.0204)	0.2899 (0.0816)
$\tau_{1,0} = 0.01,\ \tau_{0,1} = 0.8$					
Actual	0.7	0.4	0.9	0.1	0.3
$N = 10$	0.7535 (0.2099)	0.4961 (0.2314)	0.8529 (0.2632)	0.1515 (0.0294)	0.2626 (0.0987)
$N = 20$	0.7417 (0.2096)	0.4708 (0.1797)	0.8637 (0.2524)	0.1275 (0.0257)	0.3291 (0.0976)
$N = 30$	0.7228 (0.2057)	0.4180 (0.1779)	0.8774 (0.2247)	0.1182 (0.0187)	0.3245 (0.0901)
$\tau_{1,0} = 0.5,\ \tau_{0,1} = 0.5$					
Actual	0.7	0.4	0.9	0.1	0.3
$N = 10$	0.7371 (0.2035)	0.3763 (0.2066)	0.9408 (0.2611)	0.1463 (0.0237)	0.3197 (0.1019)
$N = 20$	0.7301 (0.2017)	0.3899 (0.1963)	0.9364 (0.2449)	0.1365 (0.0219)	0.3305 (0.1013)
$N = 30$	0.7257 (0.1999)	0.4069 (0.1836)	0.9298 (0.2022)	0.0993 (0.0137)	0.3184 (0.0868)

Table 13: Simulation Results For weight parameters with cluster sizes of $n = 10$ and number of clusters, N

Simulation Parameters	μ_1	μ_2	μ_3	σ	ρ
$\tau_{1,0} = 0.8,\ \tau_{0,1} = 0.01$					
Actual	0.7	0.4	0.9	0.1	0.3
$N = 10$	0.7786 (0.1898)	0.3499 (0.1983)	0.9809 (0.2485)	0.1679 (0.0268)	0.3453 (0.0921)
$N = 20$	0.7398 (0.1377)	0.3592 (0.1782)	0.9548 (0.2176)	0.1132 (0.0112)	0.3281 (0.0835)
$N = 30$	0.6962 (0.1344)	0.3920 (0.1655)	0.9143 (0.1866)	0.1080 (0.0108)	0.3097 (0.0755)
$\tau_{1,0} = 0.01,\ \tau_{0,1} = 0.8$					
Actual	0.7	0.4	0.9	0.1	0.3
$N = 10$	0.7375 (0.1726)	0.4206 (0.1915)	0.9373 (0.2266)	0.1101 (0.0224)	0.3310 (0.0853)
$N = 20$	0.7175 (0.1210)	0.3592 (0.1651)	0.8958 (0.2139)	0.1065 (0.0152)	0.3234 (0.0827)
$N = 30$	0.6993 (0.1112)	0.3920 (0.1545)	0.8993 (0.2077)	0.1026 (0.0114)	0.3055 (0.0729)
$\tau_{1,0} = 0.5,\ \tau_{0,1} = 0.5$					
Actual	0.7	0.4	0.9	0.1	0.3
$N = 10$	0.7293 (0.1734)	0.4592 (0.2209)	0.9418 (0.2229)	0.1151 (0.0188)	0.3249 (0.0944)
$N = 20$	0.7192 (0.1399)	0.4359 (0.1910)	0.9294 (0.1930)	0.1123 (0.0135)	0.3101 (0.0881)
$N = 30$	0.6987 (0.1293)	0.4144 (0.1699)	0.8946 (0.1842)	0.1069 (0.0119)	0.3031 (0.0679)

6.5 Conclusion

In this thesis, we describe regression models for clustered multinomial and continuous responses under the assumption that data within clusters are exchangeable. We used the bivariate exponential function of Marshall and Olkin [29] to model the multinomial data and a multivariate Gaussian regression model for the continuous measurements, given the multinomial outcome. The assumption of exchangeability seems particularly suitable for data from developmental toxicity studies, where data on each fetus may include both discrete measurements, such as death or malformation and continuous measurements such as weight or length. The joint regression models are based on modeling the moments $\tau_{r,s}$ of the multinomial responses by completely monotonic functions, and a partially exchangeable Gaussian model for continuous responses, conditional on multinomial

responses. A joint quantitative model in risk assessment was compared to the typical approach of examining each end point separately. The joint risk takes into account the relationship between the outcomes which may be significant in estimating a safe dose for exposure. It was also noted that the estimates of τ and μ from the simulation study were quite comparable to the starting values. The assumption of exchangeability allows for accommodating the full cluster response, even when interest is in estimating the marginal response. The representation of the joint distributions of the discrete and continuous variables also allows for calculation of higher moments and yields tractable estimating equations for maximum likelihood estimation.

BIBLIOGRAPHY

[1] Bergsma, W. and Rudas, T. (2002). Marginal models for categorical data. The Annals of Statistics, 30, 149-159.

[2] Bowman,D and George,E O. A saturated model for analyzing exchangeable binary data: Applications to clinical and developmental toxicity studies. Journal of the American Statistical Association. 1995; 90; 871-879.

[3] Bowman, Dale, and E. Olusegun George. "Likelihood estimation for exchangeable multinomial data." Communications in Statistics-Theory and Methods 46.6 (2017): 2882-2892.

[4] Bowman, Dale, and E. Olusegun George. "Inference for Exchangeable Multinomial Data with Unequal and Missing Cluster Sizes", unpublished.

[5] Bowman, Dale, and E. Olusegun George. "Weighted least squares estimation for exchangeable binary data." Computational Statistics (2016): 1-19.

[6] Catalano, P.J., Ryan, L.M., 1992. Bivariate latent variable models for clustered discrete and continuous outcomes. J. Amer. Statist. Assoc. 87, 651658.

[7] Cox, D.R., 1972. The analysis of multivariate binary data. Appl. Statist. 21, 113120.

[8] Darroch, J.N., 1981. The MantelHaenszel test and tests of marginal symmetry, fixed-effects and mixed models for a categorical response. Internat. Statist. Rev. 49, 285307.

[9] de Finetti, B., 1974. Theory of Probability. Wiley, New York.

[10] Draper, D., Hodges, J.S., Mallows, C.L., Pregibon, D., 1993. Exchangeability and data analysis. J. Roy. Statist. Soc. A 156, 937.

[11] Efron, B. (1986). Double Exponential Families and Their Use in Generalized Linear Regression. Journal of the American Statistical Association, 81, 709-721.

[12] Faes, C., H. Geys, M. Aerts, G. Molenberghs, and P.J. Catalano. 2004. Modeling combined continuous and ordinal outcomes in a clustered setting. JABES 9(4):515-530.

[13] Fitzmaurice, G.M., Laird, N.M., 1995. Regression models for bivariate discrete and continuous outcomes with clustering. J. Amer. Statist. Assoc. 90, 845852.

[14] George, EO and Bowman, D. A full likelihood procedure for analyzing exchangeable binary data. Biometrics, 1995; 51; 512-523.

[15] E. Olusegun George, Kyeongmi Cheon, Yilian Yuan, Aniko Szabo; On exchangeable multinomial distributions. Biometrika 2016; 103 (2): 397-408. doi: 10.1093/biomet/asw009

[16] George, E.O., Bowman, D., and An, Q. (2013). Regression Models for Analyzing Clustered Binary and Continous outcomes under the assumption of Exchangeability. *Analysis of Mixed Data: Methods and Applications*, CRC Press.

[17] George EO, Armstrong A, Catalano PJ, Srivastava DK. Regression models for analyzing clustered binary and continuous outcomes under an assumption of exchangeability. Journal of Statistical Planning and Inference 2007; 137:3462-3474

[18] George, E.O., Srivastava, D.K., Seshadri, R., 2006. Parametric models for exchangeable binary data, Preprint.

[19] George, E.O., Ibrahim, J., Ryan, L.M., Seshadri, R., 2007. Bayesian analysis of clustered bivariate binary and continuous outcomes in teratology studies. (Preprint).

[20] Geys, H., Regan, M., Catalano, P. and Molenberghs, G. (2001). Two Latent Variable Risk Assessment Approaches for Mixed Continuous and Discrete Outcomes From Developmental Toxicity Data. Journal of Agricultural Biological and Environment Statistics, 6, 340355.

[21] Kimmel, C. A. and Gaylor, D. W. (1988), Issues in Qualitative and Quantitative Risk Analysis for Developmental Toxicology. Risk Analysis, 8: 15-20.

[22] Kuk, AYC. A litter-based approach to risk assessment in developmental toxicity studies via a power family of completely monotone functions. Applied Statistics. 2004; 53; 369-386.

[23] Laird, N.M., 1995. Longitudinal panel data: an overview of current methodology. In: Cox, D.R., Hinkley, D.V., Barndoff-Nielsen, O.E. (Eds.), Likelihood Time Series with Econometric and Other Applications, Chapman and Hall, London, 1996, pp. 143175.

[24] Self, S. G. and Liang, K. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. Journal of the American Statistical Association, 1987, 82: 605-610.

[25] Liang, K.-Y., Zeger, S.L., 1986. Longitudinal data analysis using generalized linear models. Biometrika 73, 1322.

[26] Liang, K.-Y., Zeger, S.L., Qaqish, B., 1992. Multivariate regression analyses for categorical data. J. Roy. Statist. Soc. B 54, 324.

[27] Little, R.J.A., Rubin, D.B., 1987. Statistical Analysis with Missing Data. Wiley, New York.

[28] Little, R.J., Schluchter, M.D., 1985. Maximum likelihood estimation for mixed continuous and categorical data with missing values. Biometrika 72, 497512.

[29] Marshall, Albert W., and Ingram Olkin. A Multivariate Exponential Distribution. Journal of the American Statistical Association, vol. 62, no. 317, 1967, pp. 3044.

[30] Olkin, I., Tate, R.F., 1961. Multivariate correlation models with mixed discrete and continuous variables. Ann. Math. Statist. 32, 448465 with correction in 36, 343344.

[31] Regan, M.M., Catalano, P.J., 1999. Likelihood models for clustered binary and continuous outcomes: application to developmental toxicology. Biometrics, 760768.

[32] Ryan, L.M., Catalano, P.J., Kimmel, C.A., Kimmel, G.L., 1991. Relationship between fetal weight and malformation in developmental toxicity studies. Teratology 44, 215223.

[33] Seber, G.A.F., 1984. Multivariate Observations. Wiley, New York.

[34] Stefanescu, C and Turnbull, BW. Likelihood inference for exchangeable binary data with varying cluster sizes. Biometrics.2003; 59: 18-24.

[35] Zeger, S.L., Liang, K.-Y., 1986. Longitudinal data analysis for discrete and continuous outcomes. Biometrics 42, 121130.

[36] Zhao, L.P., Prentice, R.L., Self, S.G., 1992. Multivariate mean parameter estimation by using a partial exponential model. J. Roy. Statist. Soc. Series B 54, 805811.

APPENDIX A

R-CODES

A.0.1 EM Algorithm Assuming Interpretability

```
rm(list=ls())
#EDGE Data Requires to be loaded
obs1<-obs[g.index==2,]
A<-obs1[,6]
rr<-obs1[,4:5]
n<-obs1[,3]
k<-2
B<-cbind(rr,n)
B1<-matrix(B,ncol=(k+1),byrow=F)
M<-max(n)

#define an array for parameter Theta depending
#on cluster size and number of outcomes
theta<-array(rep(0,(M+1)^k),dim=(rep(M+1,k)))
theta1<-theta
A1<-theta
N<-sum(A)
c<-1
for (i in 1:k){
cc<-(M+i)/i
c<-c*cc
}
t<-matrix(rep(0,k*(M+1)^k),ncol=k)

for(a1 in 1:k){
t[,a1]<-rep(rep(1:(M+1),each=(M+1)^(k-a1)),(M+1)^(a1-1)))
}
for(a2 in 1:nrow(t)){
a <- matrix(t[a2,], 1, byrow = TRUE)
theta[a]<-ifelse(sum(a)>(M+k),0,1/c)
}

#define a function to perform multinomial combinations
multchoose<-function(n,r){
m<-ifelse(n<0,0,ifelse(any(r<0),0,ifelse((n-sum(r))<0
,0,factorial(n)/
((prod(factorial(r)))*factorial(n-sum(r))))))
return(m)
}

#create a matrix with all possible arrangements of r_1,...,r_k
```

```
for (b1 in 1:k){
if(b1==1){
z1<-0:M
r1<-matrix(z1,ncol=1)
} else {
for (b2 in 1:nrow(r1)){
z1<-0:(M-sum(r1[b2,]))
z2<-r1[rep(b2,length(z1)),]
if(length(z1)==1){z3<-cbind(matrix(z2,nrow=1),z1)}
else{z3<-cbind(z2,z1)}
if(b2==1){z4<-z3}else{z4<-rbind(z4,z3)}
}
r1<-z4
}
}

#define the p_r1,..,rk,z1,...,zk function
pnrs<-function(n,r,s,theta,M){
num<-prod(choose(r,s))*choose((M-sum(r)),(n-sum(s)))
*theta[(matrix(r,nrow=1)+1)]

for (c1 in 1:k){
y1<-s[c1]:(M-n+sum(s[1:c1]))
if(c1==1){
y2<-matrix(y1,ncol=1)
} else {
y3<-y2[rep(1:nrow(y2), length(y1)),]
y4<-rep(y1,each=nrow(y2))
#if(nrow(y3)==1){y2<-cbind(matrix(y3,nrow=1),y4)}
else{y2<-cbind(y3,y4)}
y2<-cbind(y3,y4)
}
}
y5<-matrix(s,nrow=1)
y6<-y5[rep(1,nrow(y2)),]
if(nrow(y2)==1){y6<-matrix(y6,nrow=1)}else{y6<-y6}
y7<-apply(choose(y2,y6),1,prod)*choose((M-rowSums(y2)),
(n-rowSums(y6)))*theta[(y2+1)]
den<-sum(y7)

p<-num/den
return(p)
}

#perform the EM- Algorithm (10 iterations used)
```

```
tol<-1
for (u in 1:10){
for(d1 in 1:nrow(r1)){
arn<-0
for(d2 in seq(1,M,1)){
for (d3 in 1:k){
if(d3==1){
x1<-max(0,(r1[d1,d3]-M+d2)):min(r1[d1,d3],d2)
x2<-matrix(x1,ncol=1)
} else {
for (d4 in 1:nrow(x2)){
x1<-max(0,(r1[d1,d3]-M+d2+sum(r1[d1,(1:(d3-1))])
-sum(x2[d4,])))):min(r1[d1,d3],(d2-sum(x2[d4,])))
x3<-x2[rep(d4, length(x1)),]
if(length(x1)==1){x4<-cbind(matrix(x3,nrow=1),x1)}
else{x4<-cbind(x3,x1)}
if(d4==1){x5<-x4}else{x5<-rbind(x5,x4)}
}
x2<-x5
}
}
for(d5 in 1:nrow(x2)){
x.1<-as.vector(c(x2[d5,],d2))
x6<-apply(B1,1,identical,x.1)
x7<-sum(which(x6))
if(x7==0){x8<-0}else{x8<-A[which(x6)]}
arn<-arn+x8*pnrs(d2,r1[d1,],x2[d5,],theta,M)
}
}
A1[matrix(r1[d1,],nrow=1)+1]<-arn

}
A1[is.nan(A1)]<-0
theta1<-A1/N

tol<-sum(abs(theta-theta1))
theta<-theta1
}

# Apply the inversion function to find the values of tau
taurs<-function(r,M,theta){
for (g1 in 1:k){
if(g1==1){
y1<-0:(M-sum(r))
rk<-matrix(y1,ncol=1)
```

```
} else {
for (g2 in 1:nrow(rk)){
y1<-0:(M-sum(rk[g2,])-sum(r))
y2<-rk[rep(g2,length(y1)),]
if(length(y1)==1){y3<-cbind(matrix(y2,nrow=1),y1)}
else{y3<-cbind(y2,y1)}
if(g2==1){y4<-y3}else{y4<-rbind(y4,y3)}
}
rk<-y4
}
}
tau<-0
for (g3 in 1:nrow(rk)){
tau<-tau+multchoose((M-sum(r)),rk[g3,])
/multchoose(M,(rk[g3,]+r))*
theta[matrix((rk[g3,]+r),nrow=1)+1]
}
return(tau)
}

#calculate the estimates of each tau
tau11<-taurs(c(1,1),M,theta)
tau20<-taurs(c(2,0),M,theta)
tau02<-taurs(c(0,2),M,theta)
tau10<-taurs(c(1,0),M,theta)
tau01<-taurs(c(0,1),M,theta)
```

A.0.2 Weighted Least Squares Algorithm

```
obs<-data.matrix(obs)
library(MASS)

X<-as.matrix(rbind(rep(1,gk),
rep(c(dose1,dose2,dose3,dose4),each=gk1)))

beta1<-c(b11,b12)
beta2<-c(b21,b22)
beta3<-c(b31,b32)

n<-obs[,3]
r<-obs[,4]
s<-obs[,5]
A<-obs[,6]
m<-obs[,7]
g.index<-obs[,1]

p<-NA
f<-NA
h<-NA
mi<-ifelse(m==0,0,1/m)
tol<-1
dose<-rep(c(dose1,dose2,dose3,dose4),each=gk1)

## iterations
for (j in 1:2) {
  #while (tol>.1) {
  lambda1<-exp(beta1[1]+beta1[2]*dose)
  lambda2<-exp(beta2[1]+beta2[2]*dose)
  lambda3<-exp(beta3[1]+beta3[2]*dose)


  u1 <- abs(lambda1+lambda3)
  u2 <- abs(lambda2+lambda3)
  u3 <- abs(lambda3)
  u  <- abs(lambda1+lambda2+lambda3)

  tau <- (((u+r+s)*u1*u2)+(u3*r*s))
  /((u+r+s)*(u1+r)*(u2+s))

  ptaub1<-(((u+r+s)*u2*r)-(u3*r*s))
  /((u+r+s)*((u1+r)^2)*(u2+s))
```

```
ptaub2<-(((u+r+s)*u1*s)-(u3*r*s))
/((u+r+s)*(u1+r)*((u2+s)^2))

ptaub3<-(((((u+r+s)*(u2+u1))+(r*s))
         *((u+r+s)*(u1+r)*(u2+s)))
         -(((u+r+s)*((u2+s)+(u1+r)))
         *(((u+r+s)*u1*u2)+(u3*r*s))))
         /(((u+r+s)*(u1+r)*(u2+s))^2)

#p<-c(rep(0,9))
p<-NULL
f1<-NULL
f2<-NULL
f3<-NULL

#i=2
for (i in 1:nrow(obs)) {
  n0=n[i]
  r0=r[i]
  s0=s[i]
  np=n0-r0-s0
  g0=g.index[i]
  ts<-NULL
  ts1<-NULL
  ts2<-NULL
  ts3<-NULL

  #k=0
  for (k in 0:np) {
    t=tau[g.index==g0 & n==n0 & r>=r0
    & s>=s0  & r+s==k+r0+s0]
    tpb1<-ptaub1[g.index==g0 & n==n0
    & r>=r0 & s>=s0 & r+s==k+r0+s0]
    tpb2<-ptaub2[g.index==g0 & n==n0
    & r>=r0 & s>=s0 & r+s==k+r0+s0]
    tpb3<-ptaub3[g.index==g0 & n==n0
    & r>=r0 & s>=s0 & r+s==k+r0+s0]

    l=0:k
    ts=c(ts,sum(choose(k,l)*t))
    ts1<-c(ts1,sum(choose(k,l)*tpb1))
    ts2<-c(ts2,sum(choose(k,l)*tpb2))
    ts3<-c(ts3,sum(choose(k,l)*tpb3))

  }
```

```
    k=0:np
    fact<-factorial(n0)/(factorial(r0)
          *factorial(s0)*factorial(np))
    p[i]=fact*sum(((-1)^k)*choose(np,k)*ts)
    f1[i]<-fact*sum(((-1)^k)*choose(np,k)*ts1)
    f2[i]<-fact*sum(((-1)^k)*choose(np,k)*ts2)
    f3[i]<-fact*sum(((-1)^k)*choose(np,k)*ts3)

  }
  pi<-ifelse(p==0,0,1/p)
  f1[is.nan(f1)]<-0
  f2[is.nan(f2)]<-0
  f3[is.nan(f3)]<-0

  f1i<-ifelse(f1==0,0,1/f1)
  f2i<-ifelse(f2==0,0,1/f2)
  f3i<-ifelse(f3==0,0,1/f3)

  z1<-crossprod(X,beta1)+mi*f1i*(A-m*p)
  z2<-crossprod(X,beta2)+mi*f2i*(A-m*p)
  z3<-crossprod(X,beta3)+mi*f3i*(A-m*p)


  D1<-X%*%diag(f1^2*m*pi)%*%t(X)

  hess1<-ginv(X%*%diag(f1^2*m*pi)%*%t(X))
  test<-det(X%*%diag(f2^2*m*pi)%*%t(X))
  hess2<-ginv(X%*%diag(f2^2*m*pi)%*%t(X))
  hess3<-ginv(X%*%diag(f3^2*m*pi)%*%t(X))

  beta1p<-hess1%*%X%*%diag(f1^2*m*pi)%*%z1
  beta2p<-hess2%*%X%*%diag(f2^2*m*pi)%*%z2
  beta3p<-hess3%*%X%*%diag(f3^2*m*pi)%*%z3

  tol1<-sum((beta1p-beta1)^2)
  tol2<-sum((beta2p-beta2)^2)
  tol3<-sum((beta3p-beta3)^2)

  tol<-tol1+tol2+tol3

  beta1<-beta1p
  beta2<-beta2p
  beta3<-beta3p

}
```