MODEL ADEQUACY TESTS FOR EXPONENTIAL FAMILY REGRESSION MODELS

by

CHAMPA HEMANTHI MAGALLA

BSc, University of Colombo, Sri Lanka, 2002

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Statistics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas
2012

Abstract

The problem of testing for lack of fit in exponential family regression models is considered. Such nonlinear models are the natural extension of Normal nonlinear regression models and generalized linear models. As is usually the case, inadequately specified models have an adverse impact on statistical inference and scientific discovery. Models of interest are curved exponential families determined by a sequence of predictor settings and mean regression function, considered as a sub-manifold of the full exponential family. Constructed general alternative models are based on clusterings in the mean parameter components and allow likelihood ratio testing for lack of fit associated with the mean, equivalently natural parameter, for a proposed null model. A maximin clustering methodology is defined in this context to determine suitable clusterings for assessing lack of fit. In addition, a geometrically motivated goodness of fit test statistic for exponential family regression based on the information metric is introduced. This statistic is applied to the cases of logistic regression and Poisson regression, and in both cases it can be seen to be equal to a form of the Pearson χ^2 statistic. This same statement is true for multinomial regression. In addition, the problem of testing for equal means in a heteroscedastic Normal model is discussed. In particular, a saturated 3 parameter exponential family model is developed which allows for equal means testing with unequal variances. A simulation study was carried out for the logistic and Poisson regression models to investigate comparative performance of the likelihood ratio test, the deviance test and the goodness of fit test based on the information metric. For logistic regression, the Hosmer-Lemeshow test was also included in the simulations. Notably, the likelihood ratio test had comparable power with that of the Hosmer-Lemeshow test under both m- and n-asymptotics, with superior power for constructed alternatives. A distance function defined between densities and based on the information metric is also given. For

logistic models, as the natural parameters go to plus or minus infinity, the densities become more and more deterministic and limits of this distance function are shown to play an important role in the lack of fit analysis. A further simulation study investigated the power of a likelihood ratio test and a geometrically derived test based on the information metric for testing equal means in heteroscedastic Normal models.

MODEL ADEQUACY TESTS FOR EXPONENTIAL FAMILY REGRESSION MODELS

by

CHAMPA HEMANTHI MAGALLA

BSc, University of Colombo, Sri Lanka, 2002

A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Statistics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas
2012

Approved by:

Major Professor
James Neill

Abstract

The problem of testing for lack of fit in exponential family regression models is considered. Such nonlinear models are the natural extension of Normal nonlinear regression models and generalized linear models. As is usually the case, inadequately specified models have an adverse impact on statistical inference and scientific discovery. Models of interest are curved exponential families determined by a sequence of predictor settings and mean regression function, considered as a sub-manifold of the full exponential family. Constructed general alternative models are based on clusterings in the mean parameter components and allow likelihood ratio testing for lack of fit associated with the mean, equivalently natural parameter, for a proposed null model. A maximin clustering methodology is defined in this context to determine suitable clusterings for assessing lack of fit. In addition, a geometrically motivated goodness of fit test statistic for exponential family regression based on the information metric is introduced. This statistic is applied to the cases of logistic regression and Poisson regression, and in both cases it can be seen to be equal to a form of the Pearson χ^2 statistic. This same statement is true for multinomial regression. In addition, the problem of testing for equal means in a heteroscedastic Normal model is discussed. In particular, a saturated 3 parameter exponential family model is developed which allows for equal means testing with unequal variances. A simulation study was carried out for the logistic and Poisson regression models to investigate comparative performance of the likelihood ratio test, the deviance test and the goodness of fit test based on the information metric. For logistic regression, the Hosmer-Lemeshow test was also included in the simulations. Notably, the likelihood ratio test had comparable power with that of the Hosmer-Lemeshow test under both m- and n-asymptotics, with superior power for constructed alternatives. A distance function defined between densities and based on the information metric is also given. For

logistic models, as the natural parameters go to plus or minus infinity, the densities become more and more deterministic and limits of this distance function are shown to play an important role in the lack of fit analysis. A further simulation study investigated the power of a likelihood ratio test and a geometrically derived test based on the information metric for testing equal means in heteroscedastic Normal models.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to express my sincere gratitude to my major advisor Dr. Jim Neill for all the guidance given to me throughout this research. Without his thoughtfulness, patience and encouragement, this work would not have been a success. A very special thank to Dr. Forrest Miller for all the explanations in mathematical parts and his invaluable thoughts towards the analysis and the writing part. Special thanks to Dr. Paul Nelson and Dr. Haiyan Wang for being there in my committee, the knowledge imparted to me and for the suggestions made on this research work.

My sincere thanks to Dr. John Boyer, former department head who gave me the opportunity to study in this department. I would like to thank all the professors in the department of statistics for all the knowledge shared with me during the last few years. Special thanks to Pam and the secretaries in the department office for invaluable help during the time I spent here.

Special gratitude to the faculty and staff in the Department of Statistics, University of Colombo for giving me the basic knowledge and foundation in statistics which made this work possible.

Finally, my heartfelt gratitude for my parents, brothers and friends for their endless love and support throughout my education. Their love and encouragement made this achievement possible.

Dedication

To my loving parents and brothers...

Chapter 1

Introduction

The problem of testing for lack of fit in exponential family regression models is considered. Such nonlinear models are the natural extension of Normal nonlinear regression models and generalized linear models. As is usually the case, inadequately specified models have an adverse impact on statistical inference and scientific discovery. Models of interest are curved exponential families determined by a sequence of predictor settings and mean regression function, considered as a sub-manifold of the full exponential family. Constructed general alternative models are based on clusterings in the mean parameter components and allow likelihood ratio testing for lack of fit associated with the mean, equivalently natural parameter, for a proposed null model. A maximin clustering methodology is defined in this context to determine suitable clusterings for assessing lack of fit. In addition, a geometrically motivated goodness of fit test statistic for exponential family regression based on the information metric is introduced. This statistic is applied to the cases of logistic regression and Poisson regression, and in both cases it can be seen to be equal to a form of the Pearson χ^2 statistic. This same statement is true for multinomial regression. In addition, the problem of testing for equal means in a heteroscedastic Normal model is discussed. In particular, a saturated 3 parameter exponential family model is developed which allows for equal means testing with unequal variances. A simulation study was carried out for the logistic and Poisson

regression models to investigate comparative performance of the likelihood ratio test, the deviance test and the goodness of fit test based on the information metric. For logistic regression, the Hosmer-Lemeshow test was also included in the simulations. Notably, the likelihood ratio test had comparable power with that of the Hosmer-Lemeshow test under both m- and n-asymptotics, with superior power for constructed alternatives. A distance function defined between densities and based on the information metric is also given. For logistic models, as the natural parameters go to plus or minus infinity, the densities become more and more deterministic and limits of this distance function are shown to play an important role in the lack of fit analysis. A further simulation study investigated the power of a likelihood ratio test and a geometrically derived test based on the information metric for testing equal means in heteroscedastic Normal models.

Chapter 2 reviews the literature associated with exponential family nonlinear regression models. Several methods of checking lack of fit are reviewed, including tests for generalized linear models, a broad class of the exponential family regression models.

Chapter 3 reviews the necessary geometry for maximum likelihood estimation in exponential family nonlinear regression models, and presents general alternative models for assessing lack of fit associated with the mean function. The construction of such general alternatives is based on clusterings in the mean components. A maximin power clustering methodology is defined in the context of exponential family nonlinear regression models to determine suitable clusterings for assessing lack of fit. In addition, a geometrically motivated goodness of fit test statistic for exponential family regression based on the information metric is introduced. Curved exponential family models and generalizations are discussed to exemplify the general discussion, including tests for generalized linear models and equal means testing for heterogenous Normal models.

Chapter 4 presents the results of simulation studies for lack of fit tests discussed in Chap-

ter 3. A simulation study was carried out for the logistic regression model to investigate comparative performance of the likelihood ratio test with the deviance test and the Hosmer-Lemeshow test, along with a goodness of fit test based on the information metric. Subsequently, an analytical explanation for unusual patterns observed in the power results is presented. The analysis involves consideration of limits to infinity of an information based distance function, and provides guidance for the selection of parameters in the logistic regression model for appropriate study. A simulation study for the Poisson regression model is also considered. The comparative power performance of the likelihood ratio test, the deviance and a goodness of fit test based on the information metric is presented. An investigation of power for testing equal means with different data generators is considered for heteroscedastic Normal models.

Chapter 5 discusses the necessary asymptotic distributional results associated with the likelihood ratio statistics described in Chapter 3. The derivation of the asymptotic non-central chi-square distribution under local parameter alternatives is based in part on the corresponding array of distributions being locally asymptotic Normal (LAN). As described in Chapter 5, a key condition to ensure LAN for exponential family nonlinear regression models is that the array of exponential family distributions be uniformly differentiable in quadratic mean. Establishing the L_2 differentiability of the parametric array of densities associated with the array of exponential family distributions, and hence LAN, is carried out in Chapter 5. Based on LAN, the asymptotic non-central chi-square distribution for the sequence of log likelihood ratio statistics is provide by Theorem 1 in this chapter.

Chapter 6 summarizes conclusions and suggests topics for future research.

Chapter 2

Literature Review

The problem of assessing model adequacy is historically old and has generated much research related to certain statistical models. On the other hand, limited study of general exponential family regression models has been done, although much work regarding the statistical behavior for models in certain subclasses (e.g. generalized linear models) has been accomplished as noted by Wei (1998). In particular, checking model adequacy in the general case of exponential family regression models has not been well developed. A geometric development of exponential family regression models is given by Wei (1998). The following reviews general inference techniques for exponential family nonlinear regression models, along with model adequacy tests for particular models within the general class of exponential family regression models.

Cordeiro and Paula (1989) presented improved likelihood ratio tests for exponential family nonlinear models based on earlier work by Cordeiro (1983) on generalized linear models. Ferrari and Cordeiro (1996) developed a corrected score test which gives more accurate size performance for finite sample sizes in exponential family nonlinear models. The adequacy measures for varying dispersion in exponential family nonlinear models has been considered by Wei, Shi, Fung, and Hu (1998), developing a score statistic and an adjusted score statistic

for Normal, Inverse Gaussian and Gamma nonlinear models. For the large sample case, the accuracy of the diagnostic tools based on Normal approximations was considered by Cook and Tsai (1990). The authors mainly considered methods of assessing the magnitude of the difference between large sample and likelihood based confidence regions for exponential family parameters. Under a sequence of Pitman alternatives, Lemonte (2011) considered the local power of likelihood ratio, Wald, score and gradient test statistics for the nun-null distribution functions of exponential family nonlinear models, generalizing the results by Cordeiro and Ferrari. Considering the fact that the likelihood ratio test statistic does not follow the chi-square asymptotic limit, when the models are misspecified, Choi and Kiefer (2011) discussed the geometry of the log likelihood ratio statistic in misspecified models.

By converting the complicated nonparametric regression in exponential families to a normal theory regression problem, Brown, Cai, and Zhou (2010) developed a method of fitting regression models in exponential families. They have proposed an approach that uses a mean matching variance stabilizing transformation on the data. Using an R-square measure, a goodness of fit test was developed by Cameron and Windmeijer (1997). The authors have applied this method to a class of exponential family regression models, including logit, probit, Poisson, Geometric, Gamma and Exponential. The R-square measure considered here is based on the Kullback-Leibler divergence which measures the proportionate reduction in uncertainty due to the inclusion of predictor variables.

Stute and Zhu (2002) proposed a non-parametric test for checking the adequacy of the generalized linear model based on empirical processes given by the residuals of the observations. Extending the idea of four goodness of fit tests, Zheng (2000) developed a method for model selection in random effects and marginal models to the case of longitudinal data. Pan and Lin (2005) developed methods for checking the adequacy of generalized linear mixed models based on the cumulative sums of residuals over covariates or predicted values of the response variable. A lack of fit test for the mean function of generalized linear models was

developed by Su and Wei (1991), which can handle the case of non-replication. This test is designed to detect the inaccuracy of the mean function, even if the variance of the response variable is misspecified. Pregibon (1980) developed a test to check the accuracy of the link function in generalized linear models. Claeskens and Hjort (2008) discuss order selection tests and Neyman smooth-type tests to assess model adequacy in general, and in particular for the generalized linear model. Considering a generalized partially linear model, Hardle, Mammen, and Muller (1998) developed a test statistic to decide between a parametric and a semi-parametric model. The generalized linear model is perturbed here with a nonlinear function, and using two examples the authors tested the null hypothesis of a parametric model versus the semi-parametric alternative. Addressing the problem of over dispersion and under dispersion in count data, Sellers and Shmuelli (2010) introduced a method of fitting Poisson models. This method was based on the Conway-Maxwell-Poisson distribution. Using the exponential family properties, the leverages, Deviance and Pearson residuals are also computed and diagnostic methods presented.

Several approaches for assessing the goodness of fit of logistic regression models have been proposed. Tsiatis (1980) proposed a goodness of fit test based on partitioning the space of covariates into distinct regions, and using a score statistic for the coefficients for the grouping variable. A strategy for determining the groups was not indicated. Hosmer and Lemeshow (1980) considered a goodness of fit test for the multiple logistic regression model by using the chi-square test statistic for contingency tables. The method used different ways of calculating the expected frequencies with some predefined grouping strategies. Hosmer, Lemeshow, and Klar (1988) developed another goodness of fit statistic for logistic models when the estimated probabilities are small. Hosmer et al. (1997) made a comparison of the Hosmer and Lemeshow test with four other tests in an effort to assess the disadvantages of the previous test. They used simulation methods to compare their chi-square test with a test based on smoothed residuals developed by Cessie, Van Houwelingen and Royston, a score test for an extended logistic model by Stukel, unweighted residual sum of squares

and the Pearson's chi-square test. Archer et al. (2007) proposed a goodness of fit based on the residuals of the observations, when the data are from complex sampling designs. The importance of assessing the adequacy of a logistic model is discussed by Sarkar and Midi (2010), giving an overview of the likelihood ratio test, Hosmer Lemeshow test, Osius-Rojek large sample approximation test and the test developed by Stukel.

Foundational issues of statistical modeling are discussed by McCullagh (2002), and various approaches to assessing model adequacy are discussed by Lindsay and Liu (2009) and Shmueli (2010). The emphasis in this dissertation is on the development of statistical tests for checking the correctness of a proposed exponential family nonlinear regression model as compared to constructed general alternative models based on clusterings in the mean parameter components. Additionally, a goodness of fit test statistic for exponential family regression based on the information metric is defined and investigated for use in checking the correctness of a specified model.

Chapter 3

Lack of Fit Tests for Exponential Family Regression Models

3.1 Geometry of Exponential Families

The following reviews some geometrical aspects of exponential families and introduces notation that is used subsequently. Regression for the mean function is considered, and seen to determine a sub-manifold of the full exponential family. Maximum likelihood estimation equations are presented for the regression parameters. In addition, a geometrically motivated goodness of fit test statistic for exponential family regression based on the information metric is introduced.

Let $\Omega \subset \mathbb{R}^k$ for some k. For $g : \Omega \to \mathbb{R},\ g(x) > 0\ \forall\ x \in \Omega$, let S consist of the k-parameter manifold of densities, $p(x, \eta) = g(x)e^{\eta \cdot x - \varphi(\eta)}$ for $x \in \Omega,\ \eta \in \mathbb{R}^k$, with dominating measure $d\mu$ which is assumed to be σ-finite on the Borel sets of $\mathbb{R}^k$. Further, the parameter η must be restricted to the subset N of $\mathbb{R}^k$ such that $\int_\Omega g(x)e^{\eta \cdot x} d\mu(x) < \infty$. The corresponding exponential family of densities is assumed to be regular of order k (cf Brown (1986)).

The condition

$$\int_\Omega g(x)e^{\eta\cdot x-\varphi(\eta)}d\mu(x)=1$$

determines the function $\varphi(\eta)$ as

$$\varphi(\eta)=log\int_\Omega g(x)e^{\eta\cdot x}d\mu(x)$$

so that

$$\frac{\partial\varphi}{\partial\eta_i}(\eta)=E_\eta(x_i).$$

Then, writing $X=(x_1,x_2,...,x_k)$, it can be shown that

$$E_\eta(X)=\nabla\varphi=\left(\frac{\partial\varphi}{\partial\eta_1}(\eta),...,\frac{\partial\varphi}{\partial\eta_k}(\eta)\right)$$
$$Var_\eta(X)=\nabla'\nabla\varphi=\left(\frac{\partial^2\varphi}{\partial\eta_i\partial\eta_j}(\eta)\right).$$

Note that the log-likelihood for x is given by

$$l(x,\eta)=log\ p(x,\eta)=\eta\cdot x-\varphi(\eta)+log\ g(x)$$

so that $\frac{\partial l}{\partial\eta_i}=x_i-\frac{\partial\varphi}{\partial\eta_i}$. The *information metric* is then given by

$$\begin{aligned}g_{ij}(\eta)&=E_\eta\left(\frac{\partial l}{\partial\eta_i}\frac{\partial l}{\partial\eta_j}\right)\\&=\int_\Omega\left(x_i-\frac{\partial\varphi}{\partial\eta_i}\right)\left(x_j-\frac{\partial\varphi}{\partial\eta_j}\right)e^{\eta\cdot x-\varphi(\eta)}g(x)d\mu(x).\end{aligned}$$

Thus,

$$(g_{ij}(\eta))=\left(\frac{\partial^2\varphi}{\partial\eta_i\partial\eta_j}(\eta)\right)=Var_\eta(X).$$

The matrix $\left(\frac{\partial^2\varphi}{\partial\eta_i\partial\eta_j}(\eta)\right)$ is assumed to be positive definite, and $g^{ij}(\eta)$ will denote the inverse matrix of $g_{ij}(\eta)$.

Consider the mapping $\mu : N \to \mathbb{R}^k$ given by

$$\mu(\eta) = \nabla\varphi(\eta) = \left(\frac{\partial\varphi}{\partial\eta_1}, ..., \frac{\partial\varphi}{\partial\eta_k}\right).$$

The Jacobian matrix for the transformation is $\left(\frac{\partial^2\varphi}{\partial\eta_j\partial\eta_i}\right)$ which is nonsingular, so μ is a local diffeomorphism (e.g. Kass and Vos (1997)). The parameters $\eta = (\eta_1, \eta_2, ..., \eta_k)$ which are used to define S are called the *natural parameters* for S. The parameters $\mu = (\mu_1, \mu_2, ..., \mu_k)$ are called the *mean parameters* for S, and $\mu_i = \frac{\partial\varphi}{\partial\eta_i}(\eta)$. The tangent space to S at $p(x, \eta)$, denoted $T_\eta S$, is spanned by the random variables,

$$\frac{\partial l}{\partial\eta_i}(x, \eta) = x_i - \frac{\partial\varphi}{\partial\eta_i}(\eta).$$

Note

$$\begin{aligned}
\frac{\partial\mu_i}{\partial\eta_j} &= \frac{\partial^2\varphi}{\partial\eta_j\partial\eta_i} = g_{ij}(\eta) \\
\text{so that} \quad \frac{\partial\eta_j}{\partial\mu_i} &= g^{ij}(\eta) \\
\text{which gives} \quad \frac{\partial l}{\partial\eta_i} &= \sum_m \frac{\partial\mu_m}{\partial\eta_i}\frac{\partial l}{\partial\mu_m} = \sum_m g_{im}\frac{\partial l}{\partial\mu_m} \\
\text{and} \quad \frac{\partial l}{\partial\mu_j} &= \sum_n \frac{\partial\eta_n}{\partial\mu_j}\frac{\partial l}{\partial\eta_n} = \sum_n g^{jn}\frac{\partial l}{\partial\eta_n}.
\end{aligned}$$

Thus, writing $<,>_\eta$ for the information metric on $T_\eta S$ and using Einstein's summation convention, in which the summation is automatically taken without the symbol Σ for such indices as appear twice in one term, once as a subscript and once as a superscript,

$$\begin{aligned}
&\left\langle \frac{\partial l}{\partial\eta_i}, \frac{\partial l}{\partial\mu_j} \right\rangle_\eta = \left\langle \frac{\partial l}{\partial\eta_i}, g^{jn}\frac{\partial l}{\partial\eta_n} \right\rangle_\eta = g^{jn}\left\langle \frac{\partial l}{\partial\eta_i}, \frac{\partial l}{\partial\eta_n} \right\rangle_\eta \\
\text{so that} \quad &\left\langle \frac{\partial l}{\partial\eta_i}, \frac{\partial l}{\partial\mu_j} \right\rangle_\eta = g^{jn}g_{ni} = \delta_i^j.
\end{aligned}$$

Also,

$$\begin{aligned}\left\langle \frac{\partial l}{\partial \mu_i}, \frac{\partial l}{\partial \mu_j} \right\rangle_\eta &= \left\langle g^{im}\frac{\partial l}{\partial \eta_m}, g^{jn}\frac{\partial l}{\partial \eta_n} \right\rangle_\eta \\ &= g^{im}g^{jn}\left\langle \frac{\partial l}{\partial \eta_m}, \frac{\partial l}{\partial \eta_n} \right\rangle_\eta = g^{im}g^{jn}g_{mn} \\ &= g^{im}\delta^j_m = g^{ij}(\eta).\end{aligned}$$

Now consider regression for the mean as determined by a regression function $f(z, \beta)$ where $z \in \mathbb{R}^p$ is a vector of predictor variables and $\beta \in B \subset \mathbb{R}^q$ is a vector of regression parameters. Then a sequence $Z = (z_1, z_2, ..., z_k)$ of predictor settings and mean function f determines

$$M = \{p \in S : \mu_i(p) = f(z_i, \beta), \text{ for some } \beta \in B \text{ and each } i,\ 1 \leq i \leq k\ \},$$

a q-dimensional sub-manifold of S parameterized by $\beta = (\beta_1, \beta_2, ..., \beta_q)$ in terms of the mean parameters for S. Since the mapping $\eta \to \mu$ is one-to-one, the densities of M may be indexed by the corresponding η instead of μ. With additional structure, M becomes a curved exponential family (CEF).

An m-dimensional CEF can be defined as follows with two regularity conditions as given by Kass and Vos (1997). Starting with a full exponential family $S = \{p_\eta : \eta \in N\}$, consider a subfamily $S_0 = \{p_\eta \in S : \eta \in N_0\}$ where η is restricted to a subspace N_0 of N. A common way subfamilies are defined is through a mapping $\theta \to \eta(\theta)$, which generates a subset $N_0 = \eta(\Theta)$ and thus the corresponding S_0. Equivalently, S_0 may be defined by a mapping $\theta \to \mu(\theta)$, which would restrict μ to lie in a subset of the mean-value parameter space. The sub-manifold M defined above using a regression function is an instance of such S_0.

Given a mapping $\theta \to \eta(\theta)$ that defines S_0 [via $N_0 = \eta(\Theta)$], the subfamily S_0 is an m-dimensional CEF provided Θ is an open subset in $\mathbb{R}^m$ and

1. the mapping is one-to-one and smooth (infinitely differentiable), and of rank m, meaning that the $m \times k$ derivative matrix $D\eta(\theta)$ has rank m everywhere and

2. writing $\phi : N_0 \to \Theta$ for the inverse mapping, if a sequence $\{\eta_n \in N_0\}$ converges to a point $\eta_0 \in N_0$, then the corresponding sequence $\{\phi(\eta_n) \in \Theta\}$ must converge to $\phi(\eta_0) \in \Theta$.

Next note for $p \in M$ such that $\mu_i(p) = f(z_i, \beta)$, $\mu_i = \frac{\partial \varphi}{\partial \eta_i}(\eta)$. In addition, the tangent space to M at p, denoted T_pM, is spanned by the random variables

$$\frac{\partial}{\partial \beta_a}\left(l(x, \mu(\beta))\right) = \frac{\partial \mu_i}{\partial \beta_a}\frac{\partial l}{\partial \mu_i}\left(x, \mu(\beta)\right) = \frac{\partial f}{\partial \beta_a}(z_i, \beta)\frac{\partial l}{\partial \mu_i}(x, \mu).$$

Thus,

$$\left\langle \frac{\partial l}{\partial \beta_a}(x, \mu(\beta)), \frac{\partial l}{\partial \beta_b}(x, \mu(\beta)) \right\rangle_p = \frac{\partial f}{\partial \beta_a}(z_i, \beta)\frac{\partial f}{\partial \beta_b}(z_j, \beta)g^{ij}(\eta).$$

Given data $y \in \Omega$, maximum likelihood estimation (MLE) chooses β such that $l(y, \eta(\beta)) = l(y, \eta(\mu(\beta)))$ is maximized. Thus, for each $1 \leq a \leq q$, such β satisfies the equation

$$0 = \frac{\partial}{\partial \beta_a}(l(y, \eta(\mu(\beta)))) = \frac{\partial \mu_j}{\partial \beta_a}(\beta)\frac{\partial \eta_i}{\partial \mu_j}(\mu)\frac{\partial l}{\partial \eta_i}(y, \eta(\beta))$$

$$= \frac{\partial \mu_j}{\partial \beta_a}(\beta)g^{ij}(\eta)\left(y_i - \frac{\partial \varphi}{\partial \eta_i}(\eta)\right) = \frac{\partial \mu_j}{\partial \beta_a}(\beta)(y_i - \mu_i)\left\langle \frac{\partial l}{\partial \mu_i}(x, \mu), \frac{\partial l}{\partial \mu_j}(x, \mu) \right\rangle_p.$$

That is, for each $1 \leq a \leq q$,

$$0 = \left\langle (y_i - \mu_i) \frac{\partial l}{\partial \mu_i}(x, \mu), \frac{\partial \mu_j}{\partial \beta_a}(\beta) \frac{\partial l}{\partial \mu_j}(x, \mu) \right\rangle_p$$

$$= \left\langle v_y(\mu), \frac{\partial l}{\partial \beta_a}(x, \mu(\beta)) \right\rangle_p .$$

Thus, $v_y(\mu) \in T_{p(\mu)} M^{\perp}$ where

$$v_y(\mu)(x) = (y_i - \mu_i) \frac{\partial l}{\partial \mu_i}(x, \mu),$$

a random variable with respect to x. Here the superscript $\perp$ represents orthogonal complementation.

Summarizing, the geometry placed on exponential families above involves two tangent bundles: $TS = \{T_\eta S\}$ and $TM = \{T_p M\}$, such that $T_p M$ is a linear subspace of $T_p S$. In addition, the information metric (i.e. *Fisher information inner product*) $<, >_\eta$ is defined on $T_\eta S$ for each $p(\eta) \in S$, and hence on the tangent space for M. Each point $p \in M$ thus has two vector spaces associated with it: a k-dimensional vector space $T_p S$ and a q-dimensional subspace $T_p M$. If $\hat{p} = p(\mu(\hat{\beta}))$ maximizes the likelihood function then $v_y(\hat{\mu})$ is orthogonal to $T_{\hat{p}} M$.

Differential-geometrical methods for statistical inference are more completely discussed by Amari (1982, 1985), Amari et al. (1987) and Amari and Nagaoka (2000), for example. In particular, S as defined above is a Riemannian manifold with the information metric, and M is a smooth sub-manifold imbedded in S.

For an exponential family, a distance (squared) can now be defined between densities p_1 and p_2 as

$$S(p_1, p_2) = S(\mu_1, \mu_2) = \left\| \sum_{i=1}^{n} (\mu_{1i} - \mu_{2i}) \frac{\partial}{\partial \mu_i} \right\|^2_{Information\ metric\ at\ \mu_2}$$

where μ_1 and μ_2 are the mean vectors for p_1 and p_2, respectively. Note this is not symmetric in (p_1, p_2).

If M is a parametric subfamily of the exponential family S, then $S(p_1, p_2)$ can be used to define a goodness of fit statistic for M by, with y denoting the data vector,

$$S(y) = S(y, \hat{\mu}) = \left\| \sum_{i=1}^{n} (y_i - \hat{\mu}_i) \frac{\partial}{\partial \mu_i} \right\|^2_{Information\ metric\ at\ \hat{\mu}}$$

where $\hat{\mu}$ is the MLE for μ given y.

Since $\hat{\mu}$ is determined by the vector field $v_y = \sum_{i=1}^{n} (y_i - \mu_i) \frac{\partial}{\partial \mu_i}$ being perpendicular to $T_\mu M$, this represents a very reasonable definition for the distance (squared) from the data point y to M.

In the following, this is applied to the cases of logistic regression and Poisson regression. In both cases it can be seen to be equal to a form of the Pearson χ^2 statistic, by comparing the respective formulas to those given in Dobson and Barnett (2008). It can be shown that this same statement is true for multinomial regression.

In the case of logistic regression, $S(\mu_1, \mu_2)$ has the nice property that it has limits as the natural parameters go to $\pm\infty$. Further, for logistic models as the natural parameters go to $\pm\infty$, the densities become more and more deterministic and those limits can play an important role in the lack of fit analysis.

Other facts about $S(\mu_1, \mu_2)$ include:

1. For Normal families with fixed covariance matrix $\sum$, the information matrix in the

mean coordinates is constant equal to $\sum^{-1}$ and $S(\mu_1, \mu_2)$ is the Mahalanobis distance.

2. In general, for μ_1 close to μ_2, $S(\mu_1, \mu_2)$ gives the quadratic approximation of the Kullback-Leibler (KL) distance from p_1 to p_2.

3.2 Lack of Fit and Construction of General Alternative Models

The following presents the structure of the alternative full models to be considered and a basic orthogonality condition that will be imposed. In particular, constructed general alternative models are based on clusterings in the mean parameter components. A maximin clustering methodology is defined in the context of exponential family nonlinear regression models in order to determine suitable clusterings for assessing lack of fit.

An exponential family of probability densities for n independent but not identically exponentially distributed random variables can be written as

$$p(y_1, y_2, ..., y_n; \eta) = \left(\prod_{i=1}^{n} g(y_i) \right) e^{\sum_{i=1}^{n} \eta_i y_i - \sum_{i=1}^{n} \varphi(\eta_i)}$$

with the log-likelihood function given by

$$l(y_1, y_2, ..., y_n; \eta) = \sum_{i=1}^{n} \eta_i y_i - \sum_{i=1}^{n} \varphi(\eta_i) + \sum_{i=1}^{n} log\ g(y_i).$$

The natural parameters and mean parameters are, respectively,

$$\eta = (\eta_1, \eta_2, ..., \eta_n)$$
$$\mu = (\mu_1, \mu_2, ..., \mu_n)$$

where $\mu_i = E(y_i) = \varphi'(\eta_i)$ with corresponding variance $\sigma_i^2 = var(y_i) = \varphi''(\eta_i)$. Further, the information matrix with respect to the natural and mean parameters is

$$I(\eta) = diag(\sigma_1^2, \sigma_2^2, ..., \sigma_n^2) \text{ and}$$
$$I(\mu) = diag(\sigma_1^{-2}, \sigma_2^{-2}, ..., \sigma_n^{-2}) \text{ ,respectively.}$$

The basic model is $\mu_i = f(z_i, \beta)$ which is equivalently given by $\eta_i = u(z_i, \beta)$ where $u = (\varphi')^{-1} \circ f$. The adequacy of the mean regression function is to be tested.

A general full model for comparison is given by $\eta_i = u(z_i, \beta) + h_i(\beta, s)$ with $h_i(\beta, 0) = 0$.

The likelihood function can be written as

$$p(y_1, .., y_n; s, \beta) = \left[e^{\sum_{i=1}^n h_i(\beta,s)y_i - \sum_{i=1}^n \varphi(u(z_i,\beta)+h_i(\beta,s))}\right] \prod_{i=1}^n g(y_i) e^{\sum_{i=1}^n u(z_i,\beta)y_i}.$$

Note if $\beta = \beta_0$ is fixed then

$$p(y_1, .., y_n; s) = \left[e^{\sum_{i=1}^n h_i(\beta_0,s)y_i - \sum_{i=1}^n \varphi(u(z_i,\beta_0)+h_i(\beta_0,s))}\right] v(y_1, ..., y_n, \beta_0) \text{ where}$$
$$v(y_1, ..., y_n, \beta_0) = \prod_{i=1}^n g(y_i) e^{\sum_{i=1}^n u(z_i,\beta_0)y_i}.$$

Examples:

- Case 1: Let $h_i(\beta, s) = s\gamma_i(\beta), 1 \leq i \leq n$, so that the full model is $\eta_i(\beta, s) = u(z_i, \beta) + s\gamma_i(\beta)$. For $\beta = \beta_0$ fixed, this gives a curve with Efron's statistical curvature equal to 0 and

$$p(y_1,..,y_n;s)=\left[e^{s(\sum_{i=1}^n\gamma_i(\beta_0)y_i)-\sum_{i=1}^n\varphi(u(z_i,\beta_0)+s\gamma_i(\beta_0))}\right]v(y_1,...,y_n,\beta_0),$$

so that $\sum_{i=1}^n\gamma_i(\beta_0)y_i$ is a sufficient statistic for parameter s with $\beta=\beta_0$ held fixed.

- Case 2: Let $h_i(\beta,s_1,...,s_k)=\sum_{j=1}^k s_j\gamma_{ij}(\beta)$. For $\beta=\beta_0$ fixed,

  $$p(y_1,...,y_n,s_1,...,s_k)=\left[e^{\sum_{j=1}^k s_j\left(\sum_{i=1}^n\gamma_{ij}(\beta_0)y_i\right)-\sum_{i=1}^n\varphi_i\left(u(z_i,\beta_0)+\sum_{j=1}^k s_j\gamma_{ij}(\beta_0)\right)}\right]v(y_1,...,y_n,\beta_0)$$

  so that $(\sum_{i=1}^n\gamma_{i1}(\beta_0)y_i,\sum_{i=1}^n\gamma_{i2}(\beta_0)y_i,...,\sum_{i=1}^n\gamma_{ik}(\beta_0)y_i)$ is a sufficient statistic for $(s_1,s_2,...,s_k)$ with $\beta=\beta_0$ held fixed.

In constructing alternative models, it is natural to impose that $V_S(\beta)\in(T_{\eta(\beta,0)}M)^\perp$ where

$V_S(\beta)=\sum_i\frac{d\eta_i}{ds}\frac{\partial}{\partial\eta_i}$ is a generating vector field for the lack of fit part of the full model.

With $h_i(\beta,s)=s\gamma_i(\beta)$,

$V_S(\beta)=\sum_i\gamma_i(\beta)\frac{\partial}{\partial\eta_i}=\sum_i\sigma_i^2(\beta)\gamma_i(\beta)\frac{\partial}{\partial\mu_i}=\sum_i\alpha_i(\beta)\frac{\partial}{\partial\mu_i}$, say, since $\frac{\partial}{\partial\eta_i}=\sigma_i^2\frac{\partial}{\partial\mu_i}$.

Thus, letting $W_{\beta_k}=\sum_i\frac{\partial\eta_i}{\partial\beta_k}\frac{\partial}{\partial\eta_i}$ at $s=0$, the general orthogonality condition is

$0=\langle V_S,W_{\beta_k}\rangle_{\eta(\beta)}=\sum_i\sigma_i^2(\beta)\frac{\partial u(z_i,\beta)}{\partial\beta_k}\gamma_i(\beta)=\sum_i\frac{\partial u(z_i,\beta)}{\partial\beta_k}\alpha_i(\beta),1\le k\le q$, since $\gamma_i=\frac{\alpha_i}{\sigma_i^2}$.

In order to determine a full model in the mean components, it is proposed to use the maximin clustering methodology developed by Miller et al. (1998, 1999) and Neill et al. (2000, 2002) for univariate and multi-response linear models, and Neill and Miller (2003) and Munasinghe (2010) for nonlinear models with additive error, to construct general alternative models. The exponential family regression models are nonlinear models which allow for non-additive errors, and the generalization of the maximin method in this context is discussed next. Accordingly, the definition of a weighted inner product on $\mathbb{R}^n$ is reviewed, along with the corresponding orthogonal projection operator.

A weighted inner product on $\mathbb{R}^n$ can be defined as $< a, b >= a^T V^{-1} b$ where V is a symmetric positive definite matrix. For example, when working in the mean representation,

$$V = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{pmatrix}.$$

Next suppose A is an $n \times p$ matrix of rank p. Then $A^T V^{-1} A$ is an invertible $p \times p$ symmetric matrix.

Let

$$P_A = A(A^T V^{-1} A)^{-1} A^T V^{-1}.$$

Then

$$P_A^2 = A(A^T V^{-1} A)^{-1} A^T V^{-1} A(A^T V^{-1} A)^{-1} A^T V^{-1}$$
$$P_A^2 = A(A^T V^{-1} A)^{-1} A^T V^{-1} = P_A.$$

Thus, P_A is idempotent. Further,

$$V^{-1} P_A = V^{-1} A(A^T V^{-1} A)^{-1} A^T V^{-1}$$
$$P_A^T = V^{-1} A(A^T V^{-1} A)^{-1} A^T$$

so that $V^{-1} P_A = P_A^T V^{-1}$, the generalized symmetric property. Thus, $< a, P_A b >= a^T V^{-1} P_A b = a^T P_A^T V^{-1} b = (P_A a)^T V^{-1} b$ so that $< a, P_A b >=< P_A a, b >$, which is the inner product symmetric property. Hence, $< P_A a, P_A a >=< a, P_A^2 a >=< a, P_A a >$.

Note $P_A A = A$, so that each column of A is an eigenvector of P_A with eigenvalue 1. Thus, $P_A a = a$ for $a \in C(A)$ so that $C(A) \subset range\ P_A$. But by construction, for any vector v, $P_A v = Aw \in C(A)$ for some vector w so that $range\ P_A \subset C(A)$. Thus, $range\ P_A = C(A)$ and $C(P_A) = C(A)$.

Now suppose, $b \in C(A)^{\perp}$. For any vector a, $< a, P_A b >=< P_A a, b >= 0$ so that $P_A b = 0$. Thus, $b \in C(A)^{\perp}$ implies $P_A b = 0$. Conclude that P_A is the orthogonal projection operator onto $C(A)$ with the inner product defined using the matrix V^{-1}.

For comparison with the proposed null model, constructed alternative models can be determined by decomposing the lack of fit space according to clusterings (cf Christensen (1989, 1991)). A particular clustering can then be selected according to a maximin power strategy to specify a full model against which to test the null model for lack of fit. The maximin clustering methodology will now be defined for exponential family regression models. For specified $\eta_i = u_i(\beta_1, \beta_2, ..., \beta_q),\ q < n$, let

$$
\begin{aligned}
V_1 &= \sum_{i=1}^{n} \frac{\partial \eta_i}{\partial \beta_1} \frac{\partial}{\partial \eta_i} = \sum_{i=1}^{n} \sigma_i^2 \frac{\partial \eta_i}{\partial \beta_1} \frac{\partial}{\partial \mu_i} \\
V_2 &= \sum_{i=1}^{n} \sigma_i^2 \frac{\partial \eta_i}{\partial \beta_2} \frac{\partial}{\partial \mu_i} \\
&\vdots \\
V_q &= \sum_{i=1}^{n} \sigma_i^2 \frac{\partial \eta_i}{\partial \beta_q} \frac{\partial}{\partial \mu_i},
\end{aligned}
$$

a basis set of vectors for $T_\beta M$ under the null model. Note $\sigma_i^2 = \sigma_i^2(\beta_1, \beta_2, ..., \beta_q)$.

Then, for a particular value of β, say $\beta^* = (\beta_1^*, \beta_2^*, ..., \beta_q^*)$, σ_i^2 and $\frac{\partial \eta_i}{\partial \beta_j}$ are known, and the $n \times q$ matrix

$$
T_{\beta *} = \begin{bmatrix}
\sigma_1^2 \frac{\partial \eta_1}{\partial \beta_1} & \cdots & \sigma_1^2 \frac{\partial \eta_1}{\partial \beta_q} \\
\sigma_2^2 \frac{\partial \eta_2}{\partial \beta_1} & \cdots & \sigma_2^2 \frac{\partial \eta_2}{\partial \beta_q} \\
\vdots & \vdots & \vdots \\
\sigma_n^2 \frac{\partial \eta_n}{\partial \beta_1} & \cdots & \sigma_n^2 \frac{\partial \eta_n}{\partial \beta_q}
\end{bmatrix}
$$

is such that $C(T_{\beta^*})$ represents $T_{\beta^*} M$ by its mean parameter components.

Let Z be any clustering matrix. Then $C(Z)_{\beta^*}$ is considered to be that subspace of $T_{\beta^*} S$

obtained using the column elements of $C(Z)$ as the mean parameter components of a tangent vector. That is,

$$\begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_n \end{pmatrix} \in C(Z) \rightsquigarrow \sum_{i=1}^{n} \gamma_i \frac{\partial}{\partial \mu_i}\bigg|_{\beta^*} .$$

Now let Z be any clustering matrix. Then

$$B_{\beta^*}(Z) = C(Z)_{\beta^*} \cap C(T_{\beta^*})^{\perp}$$

represents the exponential family nonlinear analogue to so-called between-cluster lack of fit. Within-cluster lack of fit in this context is obtained by replacing $C(Z)_{\beta^*}$ with $(C(Z)_{\beta^*})^{\perp}$. All indicated subspaces are to be interpreted as subspaces of $T_{\eta(\beta^*)}S$. Regarding notation, it is emphasized that $T_pM \subset T_pS$ where $p = p(\mu) = p(\eta)$ with $\eta = \eta(\beta^*)$. Further, $T_pS = T_\mu S = T_\eta S$ and $(T_\eta M)^{\perp} \subset T_\eta S$.

The focus of the current work is to test for between-cluster lack of fit, and the maximin clustering methodology is employed in the simulation study of Chapter 4 for testing model adequacy for the cases of logistic regression and Poisson regression.

Given a collection of candidate clusterings, a maximin power clustering maximizes

$$l_{Z_{\beta^*}} = inf \left\{ \frac{\|\nu\|^2}{\tau(\beta^*)(\nu)} : \nu \in B_{\beta^*}(Z), \nu \neq 0 \right\}$$

where $\tau(\beta^*)$ is a positive definite quadratic form on the lack of fit subspace $(T_{\beta^*}M)^{\perp}$. In analogy with previous work (cf Munasinghe (2010)), let

$$\tau(\beta^*)(\nu) = \sum_{Z_{ij}} w_{ij} \left\| P_{B_{Z_{ij}}} \nu \right\|^2$$

where the sum is taken over all acceptable edge clustering matrices, and the weights w_{ij} are nonnegative and sum to one. Further, the norms are calculated with respect to the information inner product discussed above.

The following steps can be used to determine $B_{\beta^*}(Z)$ and $P_{B_{\beta^*}(Z)}$ for a given clustering matrix Z. For notational brevity, the dependence on β^* has been largely suppressed.

1. $P_Z = Z(Z^T V^{-1} Z)^{-1} Z^T V^{-1}$.
2. Let $X_0 = P_Z T_{\beta^*}$, an $n \times q$ matrix.
3. $P_{X_0} = X_0(X_0^T V^{-1} X_0)^{-1} X_0^T V^{-1}$.
4. $P_{B_Z} = P_Z - P_{X_0}$, an $n \times n$ matrix.
5. Determine the eigenvectors of P_{B_Z} having positive eigenvalues, $> 10^{-5}$ for computational purposes. These eigenvectors generate a basis for B_Z with respect to the mean parameter basis.

To calculate $\tau(\beta^*)$, for each acceptable edge clustering let Z_{ij} be the edge clustering matrix, and let $B_{Z_{ij}} = C(Z_{ij})_{\beta^*} \cap C(T_{\beta^*})^{\perp}$. Use steps one to four given above to obtain $P_{B_{Z_{ij}}}$. Then

$$\begin{aligned} \left\| P_{B_{Z_{ij}}} \nu \right\|^2 &= (P_{B_{Z_{ij}}} \nu)^T V^{-1} P_{B_{Z_{ij}}} \nu \\ &= \nu^T P_{B_{Z_{ij}}}^T V^{-1} P_{B_{Z_{ij}}} \nu \\ &= \nu^T V^{-1} P_{B_{Z_{ij}}}^2 \nu \\ &= \nu^T V^{-1} P_{B_{Z_{ij}}} \nu \\ \text{so that } \quad \tau(\beta^*)(\nu) &= \sum_{Z_{ij}} w_{Z_{ij}} \nu^T V^{-1} P_{B_{Z_{ij}}} \nu. \end{aligned}$$

Note in the case $dim\ B_{\beta^*}(Z) = 1$, step five above is used to obtain $0 \neq \nu \in B_{\beta^*}(Z)$ for a given clustering matrix Z. Then

$$l_{Z_{\beta^*}} = \frac{\|\nu\|^2}{\tau(\beta^*)(\nu)} = \frac{\nu^T V^{-1} \nu}{\tau(\beta^*)(\nu)}.$$

Regarding the constructed lack of fit part of the full model, when Z has been chosen and $\nu \in B_{\beta^*}(Z) = C(Z)_{\beta^*} \cap C(T_{\beta^*})^{\perp}$ is determined as

$$\begin{aligned} \nu &= \sum_{i=1}^{n} \alpha_i(\beta^*) \frac{\partial}{\partial \mu_i}, \\ \text{take} \quad V_S(\beta^*) &= \sum_{i=1}^{n} \gamma_i(\beta^*) \frac{\partial}{\partial \eta_i} = \nu \\ \text{with} \quad \gamma_i(\beta^*) &= \frac{\alpha_i(\beta^*)}{\sigma_i^2(\beta^*)}. \end{aligned}$$

3.3 Curved Exponential Family Models and Generalizations

Generalized linear models make up a broad class of exponential family regression models. Maximum likelihood estimation for general constructed alternatives discussed in the preceding section are considered for this class of models in the first subsection below. In the following subsection, the problem of testing for equal means in a heteroscedastic Normal model is discussed. In particular, a saturated 3 parameter exponential family model is developed which allows for equal means testing with unequal variances.

3.3.1 Generalized Linear Models

A generalized linear parametric model (GLM) in the exponential family with canonical link function is given by

$$\eta_i = \sum_{j=1}^{q} z_i^j \beta_j \text{ for } 1 \leq i \leq n.$$

A perturbation of this GLM within the exponential family is given by

$$\eta_i = \sum_{j=1}^{q} z_i^j \beta_j + s\gamma_i(\beta) \text{ for } 1 \leq i \leq n$$

with model parameters $(\beta, s) = (\beta_1, \beta_2, ..., \beta_q, s)$.

The information matrix for the (β, s) model is

$$I(\beta,s)_{jl} = \begin{cases} I(\eta)\left(\frac{\partial\eta}{\partial\beta_j}, \frac{\partial\eta}{\partial\beta_l}\right) = \sum_{i=1}^{n} \sigma_i^2(\beta,s)\left(z_i^j + s\frac{\partial\gamma_i}{\partial\beta_j}\right)\left(z_i^l + s\frac{\partial\gamma_i}{\partial\beta_l}\right) & \text{for } 1 \leq j \leq q,\ 1 \leq l \leq q \\ I(\eta)\left(\frac{\partial\eta}{\partial\beta_j}, \frac{\partial\eta}{\partial s}\right) = \sum_{i=1}^{n} \sigma_i^2(\beta,s)\left(z_i^j + s\frac{\partial\gamma_i}{\partial\beta_j}\right)\gamma_i(\beta) & \text{for } 1 \leq j \leq q, l = q+1 \\ I(\eta)\left(\frac{\partial\eta}{\partial s}, \frac{\partial\eta}{\partial s}\right) = \sum_{i=1}^{n} \sigma_i^2(\beta,s)\gamma_i^2(\beta) & \text{for } j = q+1, l = q+1. \end{cases}$$

Along the GLM given by $s = 0$, $I(\beta,0)\left(\frac{\partial\eta}{\partial\beta_j}, \frac{\partial\eta}{\partial s}\right) = \sum_{i=1}^{n} \sigma_i^2(\beta,0) z_i^j \gamma_i(\beta)$. In order to satisfy the general orthogonality condition this term is zero for $1 \leq j \leq q$. Thus, with $\alpha_i = \sigma_i^2\gamma_i$, $(\alpha_1, \alpha_2, ..., \alpha_n)$ are chosen so that, $\sum_{i=1}^{n} z_i^j \alpha_i = 0$ for $1 \leq j \leq q$ and $\gamma_i(\beta)$ are taken as $\gamma_i(\beta) = \frac{\alpha_i}{\sigma_i^2(\beta,0)}$ for $1 \leq i \leq n$.

Observed information for the (β, s) model is given by the negative value of the second partial derivative matrix of the log-likelihood evaluated at the MLE. Note that

$$l(y_1, y_2, ..., y_n, \beta, s) = \sum_{i=1}^{n} y_i \left(\sum_{j=1}^{q} z_i^j \beta_j + s\gamma_i(\beta) \right) - \sum_{i=1}^{n} \varphi \left(\sum_{j=1}^{q} z_i^j \beta_j + s\gamma_i(\beta) \right) + \sum_{i=1}^{n} log\ g(y_i)$$

$$\text{so that} \quad l(y_1, y_2, ..., y_n, \beta, s) = \sum_{j=1}^{q} \left(\sum_{i=1}^{n} z_i^j y_i \right) \beta_j + \left(\sum_{i=1}^{n} \gamma_i(\beta) y_i \right) s - \sum_{i=1}^{n} \varphi \left(\sum_{j=1}^{q} z_i^j \beta_j + s\gamma_i(\beta) \right) + \sum_{i=1}^{n} log\ g(y_i).$$

Thus,

$$\frac{\partial l}{\partial \beta_j} = \sum_{i=1}^{n} z_i^j y_i + \left(\sum_{i=1}^{n} \frac{\partial \gamma_i}{\partial \beta_j} y_i \right) s - \sum_{i=1}^{n} \varphi' \left(\sum_{l=1}^{q} z_i^l \beta_l + s\gamma_i(\beta) \right) \left(z_i^j + s \frac{\partial \gamma_i}{\partial \beta_j} \right)$$

$$\frac{\partial l}{\partial s} = \sum_{i=1}^{n} \gamma_i(\beta) y_i - \sum_{i=1}^{n} \varphi' \left(\sum_{l=1}^{q} z_i^l \beta_l + s\gamma_i(\beta) \right) \gamma_i(\beta)$$

$$\text{so that} \quad \frac{\partial^2 l}{\partial \beta_m \partial \beta_j} = \left(\sum_{i=1}^{n} \frac{\partial^2 \gamma_i}{\partial \beta_m \partial \beta_j} y_i \right) s - \sum_{i=1}^{n} \varphi' \left(\sum_{l=1}^{q} z_i^l \beta_l + s\gamma_i(\beta) \right) \frac{\partial^2 \gamma_i}{\partial \beta_m \partial \beta_j} s - \sum_{i=1}^{n} \varphi'' \left(\sum_{l=1}^{q} z_i^l \beta_l + s\gamma_i(\beta) \right) \left(z_i^m + s \frac{\partial \gamma_i}{\partial \beta_m} \right) \left(z_i^j + s \frac{\partial \gamma_i}{\partial \beta_j} \right)$$

$$\text{and} \quad \frac{\partial^2 l}{\partial s \partial \beta_j} = \sum_{i=1}^{n} \frac{\partial \gamma_i}{\partial \beta_j} y_i - \sum_{i=1}^{n} \varphi' \left(\sum_{l=1}^{q} z_i^l \beta_l + s\gamma_i(\beta) \right) \frac{\partial \gamma_i}{\partial \beta_j} - \sum_{i=1}^{n} \varphi'' \left(\sum_{l=1}^{q} z_i^l \beta_l + s\gamma_i(\beta) \right) \gamma_i(\beta) \left(z_i^j + s \frac{\partial \gamma_i}{\partial \beta_j} \right)$$

$$\text{and} \quad \frac{\partial^2 l}{\partial s^2} = -\sum_{i=1}^{n} \varphi'' \left(\sum_{l=1}^{q} z_i^l \beta_l + s\gamma_i(\beta) \right) \gamma_i(\beta)^2.$$

These equations can be written as

$$\frac{\partial^2 l}{\partial\beta_m\partial\beta_j} = \sum_{i=1}^{n}(y_i - \mu_i(\beta,s))\frac{\partial^2\gamma_i}{\partial\beta_m\partial\beta_j}s - \sum_{i=1}^{n}\sigma_i^2(\beta,s)\left(z_i^m + s\frac{\partial\gamma_i}{\partial\beta_m}\right)\left(z_i^j + s\frac{\partial\gamma_i}{\partial\beta_j}\right)$$
$$\frac{\partial^2 l}{\partial s\partial\beta_j} = \sum_{i=1}^{n}(y_i - \mu_i(\beta,s))\frac{\partial\gamma_i}{\partial\beta_j} - \sum_{i=1}^{n}\sigma_i^2(\beta,s)\left(z_i^j + s\frac{\partial\gamma_i}{\partial\beta_j}\right)\gamma_i(\beta)$$
$$\frac{\partial^2 l}{\partial s^2} = -\sum_{i=1}^{n}\sigma_i^2(\beta,s)\gamma_i(\beta)^2.$$

Thus, for fixed (β, s), taking the expectations of the $y_i's$ using $p(y_1, y_2, ..., y_n, \eta(\beta, s))$ the negative of the matrix $I(\beta, s)$ is obtained.

The MLE equations are

$$\frac{\partial l}{\partial\beta_j} = 0, \quad 1 \le j \le q, \quad \frac{\partial l}{\partial s} = 0.$$

These equations can be written as

$$\sum_{i=1}^{n}(y_i - \mu_i(\beta,s))\left(z_i^j + s\frac{\partial\gamma_i}{\partial\beta_j}\right) = 0, \quad 1 \le j \le q$$
$$\sum_{i=1}^{n}(y_i - \mu_i(\beta,s))\gamma_i(\beta) = 0,$$

which states that the vector $\sum_{i=1}^{n}(y_i - \mu_i(\beta,s))\frac{\partial}{\partial\mu_i}$ is perpendicular to the vectors

$$\sum_{i=1}^{n}\frac{\partial\eta_i}{\partial\beta_j}\frac{\partial}{\partial\eta_i}, \quad 1 \le j \le q$$
$$\text{and} \quad \sum_{i=1}^{n}\frac{\partial\eta_i}{\partial s}\frac{\partial}{\partial\eta_i}$$

with respect to the information inner product. As specific examples of the GLM, the logistic and Poisson regression models are considered in Chapter 4.

3.3.2 Testing for Equal Means in a Heteroscedastic Normal Model

Let $\Omega = \{(x_1, x_2) | x_2 = x_1^2\} \subset \mathbb{R}^2$ and let $\theta = (\theta_1, \theta_2)$ denote the natural parameters. Then, with $d\mu(x) = dx_1$ on Ω and $g(x_1, x_2) = 1$,

$$\begin{aligned}\int_\Omega g(x)e^{\theta \cdot x} d\mu(x) &= \int_\Omega g(x)e^{\theta_1 x_1 + \theta_2 x_2} d\mu(x) = \int_{-\infty}^{\infty} e^{\theta_1 x_1 + \theta_2 x_1^2} dx_1 \\ &= \int_{-\infty}^{\infty} e^{\theta_2\left(x_1^2 + \frac{\theta_1}{\theta_2} x_1\right)} dx_1 = \int_{-\infty}^{\infty} e^{\left[\theta_2\left(x_1 + \frac{\theta_1}{2\theta_2}\right)^2 - \frac{\theta_1^2}{4\theta_2}\right]} dx_1 \\ &= e^{\frac{-\theta_1^2}{4\theta_2}} \int_{-\infty}^{\infty} e^{\theta_2\left(x_1 + \frac{\theta_1}{2\theta_2}\right)^2} dx_1\end{aligned}$$

Write $\sigma^2 = -1/2\theta_2$ so that $\theta_2 = -1/2\sigma^2 < 0$. Then

$$\begin{aligned}\int_\Omega g(x)e^{\theta \cdot x} d\mu(x) &= e^{\frac{-\theta_1^2}{4\theta_2}} \int_{-\infty}^{\infty} e^{-\left(x_1 + \frac{\theta_1}{2\theta_2}\right)^2 / 2\sigma^2} dx_1 = e^{\frac{-\theta_1^2}{4\theta_2}} \sigma\sqrt{2\pi} \\ &= e^{\frac{-\theta_1^2}{4\theta_2}} \sqrt{\frac{\pi}{-\theta_2}}\end{aligned}$$

Thus,

$$\begin{aligned}\varphi(\theta) &= log \int_\Omega g(x)e^{\theta \cdot x} d\mu(x) \\ &= \frac{-\theta_1^2}{4\theta_2} + \frac{1}{2} log\left(\frac{\pi}{-\theta_2}\right) \\ &= -\frac{\theta_1^2}{4\theta_2} - \frac{1}{2} log(-\theta_2) + \frac{1}{2} log(\pi).\end{aligned}$$

Further, the mean parameters are given by

$$\mu_1 = E(x_1) = \frac{\partial \varphi}{\partial \theta_1} = -\frac{\theta_1}{2\theta_2} = \mu$$
$$\mu_2 = E(x_2) = \frac{\partial \varphi}{\partial \theta_2} = \frac{\theta_1^2}{4\theta_2^2} - \frac{1}{2\theta_2} = \mu^2 + \sigma^2.$$
$$\text{Also,} \quad \frac{\partial^2 \varphi}{\partial \theta_1^2} = -\frac{1}{2\theta_2} = \sigma^2$$
$$\text{and} \quad \frac{\partial^2 \varphi}{\partial \theta_2 \partial \theta_1} = \frac{\theta_1}{2\theta_2^2} = 2\mu\sigma^2.$$

Hence, on Ω, the exponential density is the Normal given by

$$p(x, \theta) = g(x) e^{\theta \cdot x - \varphi(\theta)} = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x_1-\mu)^2/2\sigma^2}.$$

With the parameters μ and σ^2 given above, tangent vectors are given by

$$\frac{\partial}{\partial \mu_1} = \frac{\partial \mu}{\partial \mu_1}\frac{\partial}{\partial \mu} + \frac{\partial \sigma^2}{\partial \mu_1}\frac{\partial}{\partial \sigma^2} = \frac{\partial}{\partial \mu} - 2\mu_1 \frac{\partial}{\partial \sigma^2}$$
$$\frac{\partial}{\partial \mu_2} = \frac{\partial \mu}{\partial \mu_2}\frac{\partial}{\partial \mu} + \frac{\partial \sigma^2}{\partial \mu_2}\frac{\partial}{\partial \sigma^2} = \frac{\partial}{\partial \sigma^2}.$$

Thus, given an observation $(x_1, x_2) = (x_1, x_1^2)$ on Ω, the vector field in terms of (μ_1, μ_2) coordinates is

$$\vec{v_x} = (x_1 - \mu_1)\frac{\partial}{\partial \mu_1} + (x_2 - \mu_2)\frac{\partial}{\partial \mu_2},$$

and in terms of (μ, σ^2) coordinates,

$$\vec{v_x} = (x_1 - \mu)\frac{\partial}{\partial \mu} + [(x_1 - \mu)^2 - \sigma^2]\frac{\partial}{\partial \sigma^2}.$$

The MLE $(\hat{\mu_1}, \hat{\mu_2})$ satisfies, $\vec{v_x}(\hat{\mu_1}, \hat{\mu_2}) \in T_{(\hat{\mu_1}, \hat{\mu_2})} M^{\perp}$.

Thus,in terms of (μ, σ^2) coordinates, the information metric is

$$g\left(\frac{\partial}{\partial\mu}, \frac{\partial}{\partial\mu}\right) = \frac{1}{\sigma^2}$$
$$g\left(\frac{\partial}{\partial\mu}, \frac{\partial}{\partial\sigma^2}\right) = 0$$
$$g\left(\frac{\partial}{\partial\sigma^2}, \frac{\partial}{\partial\sigma^2}\right) = \frac{1}{2\sigma^4}.$$

Now consider n independent samples on Ω, (x_{1i}, x_{2i}) for $1 \leq i \leq n$ where $x_{2i} = x_{1i}^2$. With $r > 0$, m real and k_i nonzero, taking

$$-\frac{\theta_{1i}}{2\theta_{2i}} = m \quad \text{so} \quad \theta_{1i} = -2m\theta_{2i}$$
$$-\frac{1}{2\theta_{2i}} = r^2k_i^2 \quad \text{so} \quad \theta_{2i} = -\frac{1}{2r^2k_i^2}$$

corresponds to the parametric model

$$x_i = m + rk_iZ$$

where $Z \sim N(0,1)$ and writing $x_{1i} = x_i$ for the data, $1 \leq i \leq n$. Thus, $E(x_i) = m$ and $Var\ (x_i) = r^2k_i^2, 1 \leq i \leq n$, where m and r are considered as the parameters and the k_i are specified nonzero covariate values.

Next consider a reparameterization by letting

$$t_1 = \frac{m}{r^2},\ t_2 = -\frac{1}{2r^2} \ \text{ so } \ m = -\frac{t_1}{2t_2},\ r^2 = -\frac{1}{2t_2}.$$

Also let $\xi_i = \frac{1}{k_i^2} > 0$. Then

$$\theta_{1i} = \frac{m}{r^2}\frac{1}{k_i^2} = \xi_i t_1$$
$$\theta_{2i} = \xi_i t_2, \quad t_2 < 0.$$

More generally, consider the model with unequal means given by

$$\begin{aligned}\theta_{1i} =& \xi_i t_1 + s\gamma_i \\ \theta_{2i} =& \xi_i t_2, \quad t_2 < 0, \quad \sum_{i=1}^{n} \gamma_i = 0\end{aligned}$$

where the condition $\sum_{i=1}^{n} \gamma_i = 0$ represents the general orthogonality condition on the lack of fit part of the full model, the null model corresponding to the preceding model with equal means.

Note that

$$E(x_i) = -\frac{\theta_{1i}}{2\theta_{2i}} = \frac{-t_1\xi_i - s\gamma_i}{2t_2\xi_i} = -\frac{t_1}{2t_2} - \frac{s}{2t_2}\frac{\gamma_i}{\xi_i}$$

so that $s = 0$ indeed corresponds to the equal means model. Maximum likelihood estimation of the parameters is considered in order to construct test statistics for testing equal means in a heteroscedastic Normal model.

The joint density function for the unequal means model is

$$\begin{aligned}p((x_{1i}, x_{2i}); s, t_1, t_2) &= e^{\left[\sum_{i=1}^{n} x_{1i}(\xi_i t_1 + s\gamma_i) + \sum_{i=1}^{n} x_{2i}(\xi_i t_2) - \sum_{i=1}^{n} \varphi(\xi_i t_1 + s\gamma_i, \xi_i t_2)\right]} \\ &= e^{\left[\overline{x_1} t_1 + \overline{x_2} t_2 + s\widetilde{x} - \sum_{i=1}^{n} \varphi(\xi_i t_1 + s\gamma_i, \xi_i t_2)\right]}.\end{aligned}$$

Note $(\overline{x_1} = \sum_{i=1}^{n} \xi_i x_{1i}, \overline{x_2} = \sum_{i=1}^{n} \xi_i x_{2i}, \widetilde{x} = \sum_{i=1}^{n} \gamma_i x_{1i})$ are the sufficient statistics for the parameters (t_1, t_2, s). The preceding provides a 3 parameter model with sufficient statistics $\overline{x_1} = \sum_{i=1}^{n} \xi_i x_{1i}, \overline{x_2} = \sum_{i=1}^{n} \xi_i x_{2i}$ and $\widetilde{x} = \sum_{i=1}^{n} \gamma_i x_{1i}$, functions of (x_{1i}, x_{2i}), $1 \le i \le n$, with $x_{2i} = x_{1i}^2$.

Now consider the sample space for $(\bar{x_1}, \bar{x_2}, \widetilde{x})$ for the case of general n. Write $x_{1i} = x_i,\ x_{2i} = x_i^2,\ 1 \le i \le n$, and $x = (x_i)$. The sufficient statistic mapping $\Phi : \mathbb{R}^n \to \mathbb{R}^3$ for the above

3-parameter model is given by

$$\Phi(x) = \left(\underbrace{\sum_{i=1}^{n} \xi_i x_{1i}}_{\overline{x_1}}, \underbrace{\sum_{i=1}^{n} \xi_i x_{2i}}_{\overline{x_2}}, \underbrace{\sum_{i=1}^{n} \gamma_i x_{1i}}_{\widetilde{x}} \right).$$

Recall that $\sum_{i=1}^{n} \gamma_i = 0$ and $\xi_i > 0,\ 1 \leq i \leq n$, so that the vectors (ξ_i) and (γ_i) are linearly independent. Since the derivative matrix of Φ is given by

$$D\Phi(x) = \begin{pmatrix} \xi_1 & \xi_2 & ... & \xi_n \\ 2\xi_1 x_1 & 2\xi_2 x_2 & ... & 2\xi_n x_n \\ \gamma_1 & \gamma_2 & ... & \gamma_n \end{pmatrix},$$

the (row) rank of $D\Phi(x)$ is at least 2 since (ξ_i) and (γ_i) are linearly independent. Further, the rank of $D\Phi(x)$ is equal to 2 if and only if there exist numbers α and β such that $(\xi_i x_i) = \alpha(\xi_i) + \beta(\gamma_i)$. That is, $\xi_i x_i = \alpha\xi_i + \beta\gamma_i$ so that $x_i = \alpha + \beta\left(\frac{\gamma_i}{\xi_i}\right),\ 1 \leq i \leq n$.

That is, if and only if (writing as column vectors)

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \alpha \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} + \beta \begin{pmatrix} \gamma_1/\xi_1 \\ \gamma_2/\xi_2 \\ \vdots \\ \gamma_n/\xi_n \end{pmatrix}$$

if and only if

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in Span \left[\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}, \begin{pmatrix} \gamma_1/\xi_1 \\ \gamma_2/\xi_2 \\ \vdots \\ \gamma_n/\xi_n \end{pmatrix} \right].$$

Otherwise, $D\Phi(x)$ has maximal rank. Note that if $D\Phi(x)$ has maximal rank then $\Phi(x)$ is an interior point of the image of Φ in $\mathbb{R}^3$.

Let $M = \{(x_i) \in \mathbb{R}^n \mid (x_i) \in span(\mathbf{1}, (\gamma_i/\xi_i))\}$. Suppose $(x_i) \in M$ so there exist α and β such that $x_i = \alpha + \beta\left(\frac{\gamma_i}{\xi_i}\right)$, $1 \le i \le n$. Using $\sum_{i=1}^n \gamma_i = 0$,

$$\begin{aligned}
\bar{x_1} &= \sum_i \xi_i x_i = \sum_i \xi_i \alpha + \sum_i \beta\gamma_i = \alpha\left(\sum_i \xi_i\right) \\
\widetilde{x} &= \sum_i \gamma_i x_i = \sum_i \gamma_i \alpha + \sum_i \beta\gamma_i^2/\xi_i = \beta\left(\sum_i \gamma_i^2/\xi_i\right) \\
\bar{x_2} &= \sum_i \xi_i x_i^2 = \sum_i \xi_i\left(\alpha^2 + 2\alpha\beta\gamma_i/\xi_i + \beta^2\gamma_i^2/\xi_i^2\right) \\
\bar{x_2} &= \left(\sum_i \xi_i\right)\alpha^2 + \left(\sum_i \gamma_i^2/\xi_i\right)\beta^2 \\
\bar{x_2} &= \frac{1}{(\sum_i \xi_i)}\bar{x_1}^2 + \frac{1}{(\sum_i \gamma_i^2/\xi_i)}\widetilde{x}^2.
\end{aligned}$$

Thus, $\Phi(M)$ is contained in a paraboloid in $\mathbb{R}^3$ given in Figure 3.1.

Given $(\bar{x_1}, \bar{x_2}, \widetilde{x})$ such that

$$\bar{x_2} = \frac{1}{(\sum_i \xi_i)}\bar{x_1}^2 + \frac{1}{(\sum_i \gamma_i^2/\xi_i)}\widetilde{x}^2,$$

take

$$\begin{aligned}
\alpha = \frac{\bar{x_1}}{\sum_i \xi_i}, \quad \beta &= \frac{\widetilde{x}}{\sum_i \gamma_i^2/\xi_i} \\
\text{and } (x_i) = \alpha 1 + \beta(\gamma_i/\xi_i) &\ \in M. \\
\text{Then } \sum_i \xi_i x_i &= \sum_i \xi_i \alpha + \sum_i \beta\gamma_i = (\sum_i \xi_i)\alpha = \bar{x_1} \\
\sum_i \gamma_i x_i &= \sum_i \gamma_i \alpha + \sum_i \beta\gamma_i^2/\xi_i = \widetilde{x} \\
\text{and } \sum_i \xi_i x_i^2 &= \sum_i \xi_i\left(\alpha^2 + 2\alpha\beta\frac{\gamma_i}{\xi_i} + \beta^2\frac{\gamma_i^2}{\xi_i^2}\right) \\
&= (\sum_i \xi_i)\alpha^2 + (\sum_i \gamma_i^2/\xi_i)\beta^2 = \bar{x_2}.
\end{aligned}$$

This shows that $\Phi(M)$ gives all of the paraboloid.

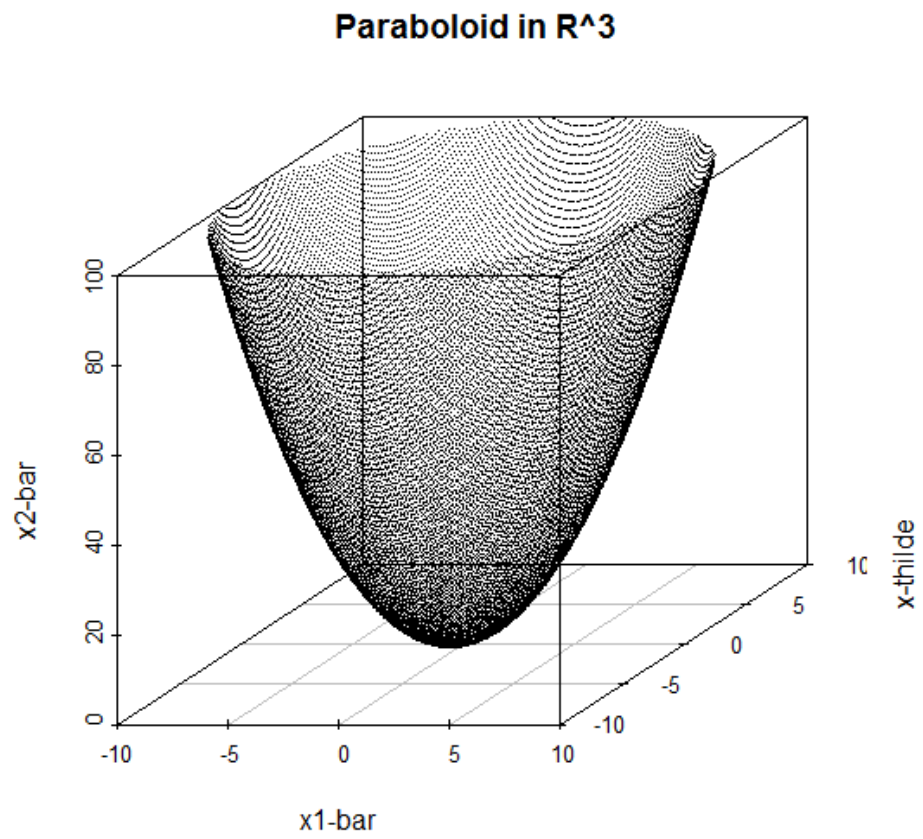

Figure 3.1: *Paraboloid in* $\mathbb{R}^3$ *containing* $\Phi(M)$

Let $\mu_i = E(x_i)$ and $\sigma_i^2 = Var(x_i),\ 1 \leq i \leq n$. Then

$$E(\bar{x_1}) = E(\sum_i \xi_i x_i) = \sum_i \xi_i \mu_i$$
$$E(\bar{x_2}) = E(\sum_i \xi_i x_i^2) = \sum_i \xi_i E(x_1^2) = \sum_i \xi_i \mu_i^2 + \sum_i \xi_i \sigma_i^2$$
$$E(\widetilde{x}) = E(\sum_i \gamma_i x_i) = \sum_i \gamma_i \mu_i.$$

It follows that

$$E(\Phi(x)) = (E(\bar{x_1}), E(\bar{x_2}), E(\widetilde{x}))$$
$$E(\Phi(x)) = \Phi((\mu_i)) + \left(0, \sum_i \xi_i \sigma_i^2, 0\right).$$

Note $\mu_i = -\frac{\theta_{1i}}{2\theta_{2i}} = -\frac{(\xi_i t_1 + s\gamma_i)}{2\xi_i t_2} = -\frac{t_1}{2t_2} - \frac{s}{2t_2}\left(\frac{\gamma_i}{\xi_i}\right)$ for $1 \leq i \leq n$.

Thus, $(\mu_i) \in span(\mathbf{1}, (\gamma_i/\xi_i))$. Hence, $(\mu_i) \in M$ so that $\Phi((\mu_i))$ is on the paraboloid and it follows from the above equation that $E(\Phi(x)) = (E(\bar{x_1}), E(\bar{x_2}), E(\widetilde{x}))$ is in the interior of S, where S denotes the set of points in $\mathbb{R}^3$ such that

$$\bar{x_2} \geq \frac{1}{(\sum_i \xi_i)} \bar{x_1}^2 + \frac{1}{(\sum_i \gamma_i^2/\xi_i)} \widetilde{x}^2.$$

If the parameters are changed from (t_1, t_2, s) to (τ_1, t_2, τ_2) where

$$\tau_1 = -\frac{t_1}{2t_2},\ t_2 = t_2,\ \tau_2 = -\frac{s}{2t_2},\quad t_2 < 0,$$

then

$$\mu = \mu(\tau_1, \tau_2) = \tau_1 \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + \tau_2 \begin{bmatrix} \gamma_1/\xi_1 \\ \gamma_2/\xi_2 \\ \vdots \\ \gamma_n/\xi_n \end{bmatrix}.$$

Also,

$$\sigma_i^2 = -\frac{1}{2\theta_{2i}} = -\frac{1}{2\xi_i t_2}$$
$$\text{so that } \xi_i \sigma_i^2 = -\frac{1}{2t_2} \text{ and thus,}$$
$$\sum_{i=1}^{n} \xi_i \sigma_i^2 = \left(\frac{n}{2}\right)\left(-\frac{1}{t_2}\right) \text{ and}$$
$$\sum_{i=1}^{n} (\xi_i \gamma_i) \sigma_i^2 = 0.$$

Hence,

$$E(\Phi(x)) = \Phi(\mu(\tau_1, \tau_2)) + \left(0, \left(\frac{n}{2}\right)\left(-\frac{1}{t_2}\right), 0\right)$$

where $\Phi(\mu(\tau_1, \tau_2))$ is on the paraboloid and $\left(\frac{n}{2}\right)\left(-\frac{1}{t_2}\right)$ ranges over $(0, +\infty)$.

This formula gives $E(\Phi(x))$ in terms of the parameters determining the probability distribution which determines the expectations.

For a given point $(\bar{x_1}, \bar{x_2}, \widetilde{x})$ in the interior of S, the following steps show that parameters (t_1, t_2, s), or equivalently (τ_1, t_2, τ_2), can be chosen so that $E(\Phi(x)) = (\bar{x_1}, \bar{x_2}, \widetilde{x})$.

1. First choose $t_2 < 0$ such that $\bar{x_2} - \frac{\bar{x_1}^2}{\sum_i \xi_i} - \frac{\widetilde{x}^2}{\sum_i \gamma_i^2/\xi_i} = \frac{n}{-2t_2}$.

2. Then take $\tau_1 = \frac{\bar{x_1}}{\sum_i \xi_i}$ and $\tau_2 = \frac{\widetilde{x}}{\sum_i \gamma_i^2/\xi_i}$.

3. Lastly take $t_1 = (-2t_2)\tau_1$ and $s = (-2t_2)\tau_2$.

Thus, given data (x_i), $(\bar{x_1}, \bar{x_2}, \widetilde{x})$ can be calculated and parameters (t_1, t_2, s) uniquely determined having $(\bar{x_1}, \bar{x_2}, \widetilde{x})$ as $E(\Phi(x))$.

These formulae for (t_1, t_2, s), equivalently (τ_1, t_2, τ_2), give the MLE for the full 3-parameter model.

Now consider the MLE for the model $s = 0$ within the (t_1, t_2, s) model. The log-likelihood function is,

$$l((x_{1i}, x_{2i}); s, t_1, t_2) = \bar{x_1}t_1 + \bar{x_2}t_2 + s\widetilde{x} - \sum_{i=1}^{n} \varphi\left(\xi_i t_1 + s\gamma_i, \xi_i t_2\right),$$

which, for $s = 0$, is

$$l((x_{1i}, x_{2i}); t_1, t_2) = \bar{x_1} t_1 + \bar{x_2} t_2 - \sum_{i=1}^{n} \varphi\left(\xi_i t_1, \xi_i t_2\right).$$

Given data (x_{1i}) with $x_{2i} = x_{1i}^2$, the MLE equations are

$$0 = \frac{\partial l}{\partial t_1} = \bar{x_1} - \sum_{i=1}^{n} \mu_{1i}(\xi_i t_1, \xi_i t_2)\xi_i \;\text{ since } \frac{\partial \varphi}{\partial \theta_1} = \mu_1$$
$$0 = \frac{\partial l}{\partial t_2} = \bar{x_2} - \sum_{i=1}^{n} \mu_{2i}(\xi_i t_1, \xi_i t_2)\xi_i \;\text{ since } \frac{\partial \varphi}{\partial \theta_2} = \mu_2$$

and $\mu_{2i} = E(x_{1i}^2) = \mu_{1i}^2 + \sigma_i^2$.

For $s = 0$, $\tau_2 = 0$ so that $\mu_{1i} = \tau_1 = -\frac{t_1}{2t_2}$, $1 \le i \le n$, and thus the first MLE equation gives

$$\bar{x_1} - \sum_{i=1}^{n} \tau_1 \xi_i = 0 \;\text{ so that }\quad \tau_1 = \frac{\bar{x_1}}{\sum_{i=1}^{n} \xi_i}.$$

The second MLE equation gives

$$\bar{x_2} - \sum_{i=1}^{n} (\tau_1^2 + \sigma_i^2)\xi_i = 0 \;\text{ so that }\quad \bar{x_2} - (\tau_1^2)\left(\sum_{i=1}^{n} \xi_i\right) - \sum_{i=1}^{n} \sigma_i^2 \xi_i = 0.$$

But $\sigma_i^2 \xi_i = -\frac{1}{2t_2}$ so that

$$\bar{x_2} - \frac{\bar{x_1}^2}{\sum_i \xi_i} + n\left(\frac{1}{2t_2}\right) = 0.$$

Thus, the MLE for the $s = 0$ model is

$$\begin{cases} \hat{\tau_1} = \frac{\bar{x_1}}{\sum_i \xi_i} \\ \frac{-n}{2\hat{t_2}} = \bar{x_2} - \frac{\bar{x_1}^2}{\sum_i \xi_i} = \bar{x_2} - (\sum_i \xi_i)\hat{\tau_1}^2. \end{cases}$$

Recall that (t_1, t_2, s) is equivalent (τ_1, t_2, τ_2) where $t_2 < 0$, and $s = 0$ if and only if $\tau_2 = 0$. Further, $\tau_2 = 0$ gives $E(x_i) = \mu_{1i} = \tau_1$ for $1 \le i \le n$ so that the model has all the $E(x_i)$ equal.

Comparing the parameter estimates gives

1. $(\tau_1)_{null} - (\tau_1)_{full} = 0.$

2. $(\tau_2)_{null} - (\tau_2)_{full} = -(\tau_2)_{full} = -\frac{\widetilde{x}}{\sum_i \gamma_i^2/\xi_i}$ since $(\tau_2)_{null} = 0.$

3. $\left(\frac{n}{-2t_2}\right)_{null} - \left(\frac{n}{-2t_2}\right)_{full} = \frac{\widetilde{x}^2}{\sum_i \gamma_i^2/\xi_i} = \left(-\frac{1}{2t_2}\right)_{true} S_1,$ thus defining S_1

as

$$S_1 = \frac{\widetilde{x}^2}{var(\widetilde{x})} = \frac{\left(\sum_i \gamma_i x_{1i}\right)^2}{\sum_i \gamma_i^2 \sigma_i^2}$$

where

$$Var(\widetilde{x}) = \sum_i \gamma_i^2 \sigma_i^2 = \sum \gamma_i^2 (-1/2\xi_i t_2)$$
$$Var(\widetilde{x}) = \left(-\frac{1}{2t_2}\right) \sum_{i=1}^{n} \frac{\gamma_i^2}{\xi_i}.$$

Further, $E(\widetilde{x}) = \sum_i \gamma_i \mu_i$, and under the null model the μ_i's are equal so $E(\widetilde{x}) = 0$ since $\sum_i \gamma_i = 0$. Now $\widetilde{x} = \sum_{i=1}^{n} \gamma_i x_{1i}$ is Normal, so that S_1 is distributed χ^2 with 1 degree of freedom under the null model.

To determine a statistic based on S_1 (and hereafter referred to as the S_1 statistic), $\left(-\frac{1}{2t_2}\right)_{full}$ is used in place of $\left(-\frac{1}{2t_2}\right)_{true}$. This comparison of parameter estimates can be graphically illustrated as in Figure 3.2.

A geometrical meaning of the S_1 statistic in terms of Fisher information is

$$S_1 = \frac{< \sum_i \gamma_i \frac{\partial}{\partial \theta_{1i}}, \sum_i x_{1i} \frac{\partial}{\partial \mu_{1i}} >^2}{\sum_i \gamma_i^2 \sigma_i^2}$$

$$\text{so that} \quad S_1 = \left[\frac{< \sum_i \gamma_i \frac{\partial}{\partial \theta_{1i}}, \sum_i x_{1i} \frac{\partial}{\partial \mu_{1i}} >}{\left| \sum_i \gamma_i \frac{\partial}{\partial \theta_{1i}} \right|} \right]^2$$

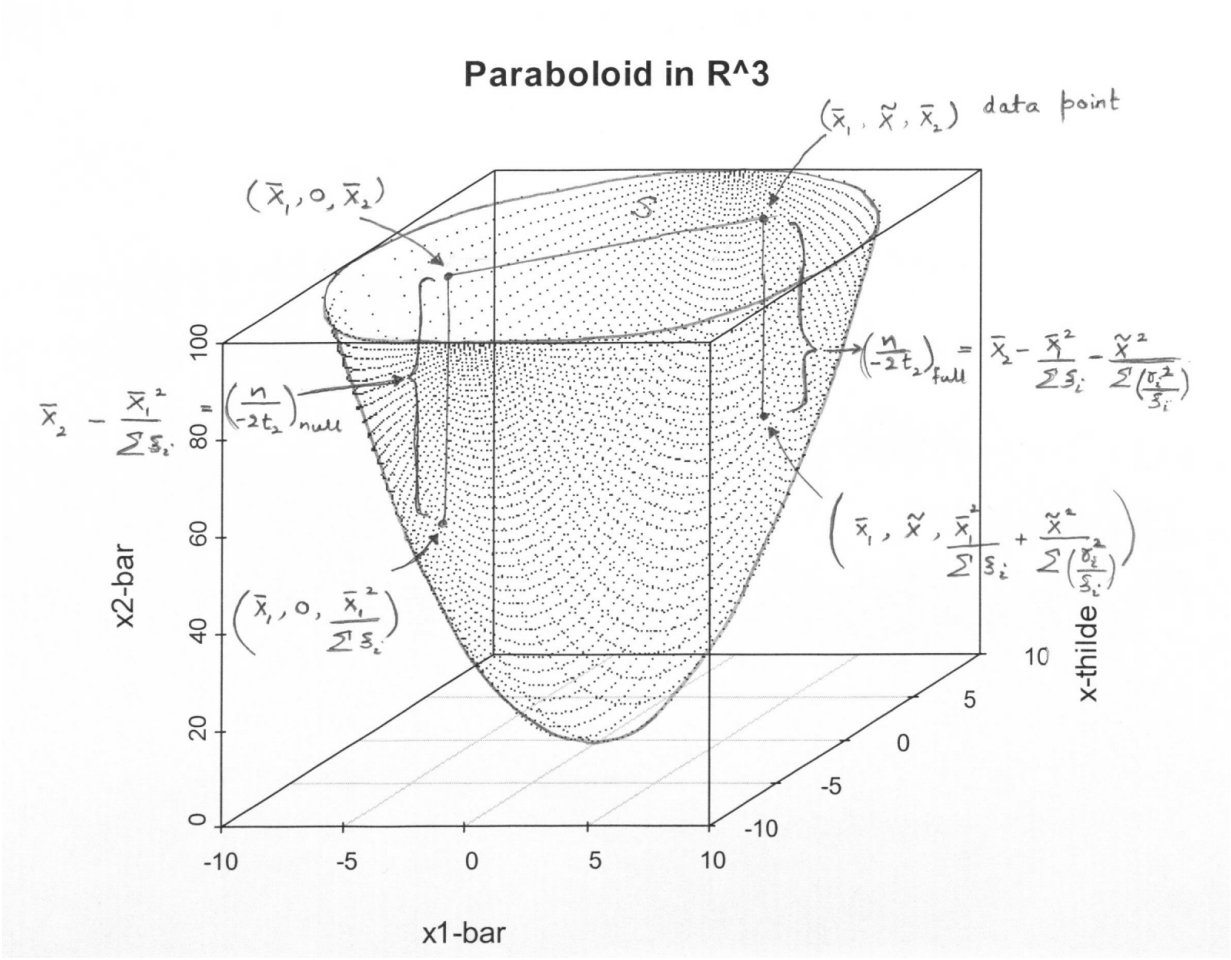

Figure 3.2: *Graphical Comparison of Parameter Estimates*

where the inner product and norm are calculated with respect to the information matrix which depends upon (estimated) parameters.

Note that the condition $\sum_{i=1}^{n} \gamma_i = 0$ gives the required orthogonalities. Recall

$$\sigma_i^2 = -\frac{1}{2\theta_{2i}} = -\frac{1}{2t_2\xi_i} \quad \text{so that} \quad \xi_i\sigma_i^2 = -\frac{1}{2t_2}.$$

Also,

$$\begin{bmatrix} I\left(\frac{\partial}{\partial\theta_1}, \frac{\partial}{\partial\theta_1}\right) & I\left(\frac{\partial}{\partial\theta_1}, \frac{\partial}{\partial\theta_2}\right) \\ I\left(\frac{\partial}{\partial\theta_2}, \frac{\partial}{\partial\theta_1}\right) & I\left(\frac{\partial}{\partial\theta_2}, \frac{\partial}{\partial\theta_2}\right) \end{bmatrix} = \begin{bmatrix} \sigma^2 & 2\mu\sigma^2 \\ 2\mu\sigma^2 & 2\sigma^4 + 4\mu^2\sigma^2 \end{bmatrix}$$

and

$$\frac{\partial}{\partial t_1} = \sum_i \xi_i \frac{\partial}{\partial \theta_{1i}}, \quad \frac{\partial}{\partial t_2} = \sum_i \xi_i \frac{\partial}{\partial \theta_{2i}}, \quad \frac{\partial}{\partial s} = \sum_i \gamma_i \frac{\partial}{\partial \theta_{1i}},$$

so that

$$\left\langle \frac{\partial}{\partial s}, \frac{\partial}{\partial t_1} \right\rangle = \sum_i \sigma_i^2 \xi_i \gamma_i = -\frac{1}{2t_2} \sum_i \gamma_i = 0$$
$$\left\langle \frac{\partial}{\partial s}, \frac{\partial}{\partial t_2} \right\rangle \Big|_{s=0} = \sum_i 2\mu_i \sigma_i^2 \xi_i \gamma_i = 2m \sum_i \sigma_i^2 \xi_i \gamma_i = 0,$$

since all the means are equal when $s = 0$.

Because there is only one orthogonality condition, clusters determined by the γ_i will use only two groups in order to accommodate a one dimensional lack of fit subspace. Further, in order to use the geometric analysis previously discussed for the unequal means model, the γ_i must not depend on the parameters (t_1, t_2) since $\sum_i \gamma_i x_i$ is part of a sufficient statistic. Thus, the $\frac{\partial}{\partial \theta_{1i}}$ components are to be clustered. That is, in the current context, for a clustering matrix Z the meaning of $C(Z)_{\beta^*}$ is as follows. For

$$\gamma = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_n \end{pmatrix} \in C(Z),$$

γ denotes the vector

$$\sum_i \gamma_i \frac{\partial}{\partial \theta_{1i}} \Big|_{\beta^*} = \sum_i \gamma_i \sigma_i^2 \frac{\partial}{\partial \mu_i} \Big|_{\beta^*}.$$

Since the γ_i cannot depend upon (t_1, t_2), this rules out the use of fuzzy clusters (cf Munasinghe (2010)).

Consider a clustering determined by the following structure

$$(\gamma_i) = \begin{pmatrix} a \\ \vdots \\ a \\ b \\ \vdots \\ b \end{pmatrix}$$

where there are l number of a's and $(n-l)$ number of b's. According to the condition $\sum_i \gamma_i = 0$,

$$la + (n-l)b = 0 \quad \text{so that} \quad a = -\left(\frac{n}{l} - 1\right) b.$$

Taking the length squared of the γ vector to be n,

$$la^2 + (n-l)b^2 = n$$

and thus,

$$l\left(\frac{n}{l} - 1\right)^2 b^2 + (n-l)b^2 = n$$

$$b^2\left[l\left(\frac{n}{l} - 1\right)^2 + (n-l)\right] = n$$

$$\text{so that} \quad b = \sqrt{\frac{l}{n-l}} \quad \text{and} \quad a = -\sqrt{\frac{n-l}{l}} = -\frac{1}{b}.$$

A method to determine the number of components in each of the two clusters (i.e. l)is given in Chapter 4 for use with the S_1 and likelihood ratio statistics to test for equal means in the heterogenous Normal model.

Chapter 4

Simulation Results

4.1 Logistic Regression Model

An exponential family model for Binomial proportions is the logistic regression model, which represents a generalized linear model with response variables measured on a binary scale. The canonical link function is the logit transformation, which with one predictor variable z is given by

$$\eta = log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 z.$$

The simulations carried out in this section involve n independent Binomial observations with parameters (m_i, π_i) for $1 \leq i \leq n$. The asymptotic distribution of goodness of fit or lack of fit test statistics is obtained by letting n become large. If the number of distinct predictor settings also increases with n, then each value of m_i tends to be small, leading to so-called n-asymptotics. If the number of distinct predictor settings is held fixed and n is allowed to become large, then each value of m_i also tends to become large, leading to so-called m-asymptotics. Note that when there is at least one continuous predictor in the model then the number of distinct predictor settings approximates n, and represents a frequently encountered case in applications. This terminology is utilized by Hosmer and Lemeshow

(2000), for example. The simulations for the logistic regression model will illustrate results corresponding to both m- and n-asymptotics.

Properties of the logistic model with canonical link function include

$$l(y,\eta) = \sum_{i=1}^{n}\left[y_i\eta_i - m_i log(1+e^{\eta_i}) + log\binom{m_i}{y_i}\right]$$
$$\varphi(\eta) = \sum_{i=1}^{n} m_i log(1+e^{\eta_i})$$

and

$$\eta_i = log\frac{\pi_i}{1-\pi_i}, \quad 0<\pi_i<1$$
$$\pi_i = \frac{e^{\eta_i}}{1+e^{\eta_i}}, \quad 1-\pi_i = \frac{1}{1+e^{\eta_i}}$$
$$\mu_i = \frac{\partial\varphi}{\partial\eta_i} = \frac{m_i}{1+e^{\eta_i}}e^{\eta_i} = m_i\pi_i$$
$$\sigma_i^2 = \frac{\partial^2\varphi}{\partial\eta_i^2} = m_i\left[\frac{(1+e^{\eta_i})e^{\eta_i} - e^{\eta_i}e^{\eta_i}}{(1+e^{\eta_i})^2}\right]$$
$$\sigma_i^2 = m_i\frac{e^{\eta_i}}{(1+e^{\eta_i})^2} = m_i\pi_i(1-\pi_i).$$

The power of the LRT statistic, the deviance (cf Dobson and Barnett (2008)), the Hosmer Lemeshow test statistic (cf Hosmer and Lemeshow (2000)), Wald test statistics as given in Boos (1992) and Shao (2003), and the test statistic S based on the information metric presented in Chapter 3 was investigated for detecting between-cluster lack of fit. In addition, the power for these tests was assessed when the data was generated by a functionally different model as compared to the constructed full model based on between-cluster lack of fit.

Specifically, Uniformly distributed predictor values for z on the interval $(-3,3)$ were used to generate data sets with sizes $n=50$ and $n=100$. One set of data was generated for each

specified size with $m_i = 50$, using R software (cf Appendix A). Such will illustrate the case of m-asymptotics. Later in this section, the case of n-asymptotics will be illustrated with $m_i = 1$. As discussed in Munasinghe (2010), a cover for the predictor space was determined by dividing the interval $(-3, 3)$ into the five cells $\{(-3, -1.8), [-1.8, -0.6), [-0.6, 0.6), [0.6, 1.8), [1.8, 3)\}$, with three associated overlapping subsets given by $\{(-3, -0.6), [-1.8, 1.8), [0.6, 3)\}$ specifying the cover elements. Since the cardinality of the set of crisp clustering consistent with this cover becomes very large with increasing data size, a subset comprised of *ordered* partitions from the full collection was utilized for computational feasibility in the simulation studies. This led to candidate clusterings consisting of 156 crisp clusterings for the data with $n = 50$ and 690 crisp clusterings for the data with $n = 100$. The maximin power clustering criterion discussed in Neill and Miller (2003) and adapted to the logistic regression model (cf section 3.2) was then used to determine the best crisp clustering $Z_n(\beta)$ at each of chosen β values.

A grid search was utilized to determine the MLE of parameters in the logistic and alternative models used for the simulation study. An appropriate grid over which to search the β values for a logistic regression model is motivated by noting that

$$\begin{aligned} \pi_i &= \frac{e^{\eta_i}}{1 + e^{\eta_i}} = \frac{1}{1 + e^{-\eta_i}} \\ \frac{d\pi_i}{d\eta_i} &= (-1)(1 + e^{-\eta_i})^{-2}(-e^{-\eta_i}) = \frac{e^{-\eta_i}}{(1 + e^{-\eta_i})^2} \\ \lim_{\eta_i \to -\infty} \pi_i &= 0, \quad \lim_{\eta_i \to +\infty} \pi_i = 1, \ \pi(0) = \frac{1}{2}. \end{aligned}$$

Since the densities become more and more deterministic as the natural parameters go to $\pm\infty$, the grid is obtained by constraining the η and thus the β. Later in this section, a simulation without such constraints shows that the test statistics can achieve zero power for specified parameter settings. In addition, an analytical explanation of such phenomena

is presented and is based on the information distance (squared) (cf section 3.1) involving limits to infinity.

Constraints of the form $\delta_1 \leq \eta_i \leq \delta_2$ with $\delta_1 < 0 < \delta_2$ will be imposed on the η. Such bounds are determined by excluding values of π nearly equal to 0 or 1. Specifically, the excluded values are $\pi_i < 10^{-5}$ and $\pi_i > 1 - 10^{-5}$. With $\pi_i = \frac{e^{\eta_i}}{1+e^{\eta_i}}$, the approximate values of δ_1 and δ_2 can be calculated as -12 and 12, respectively.

Now write $\eta_i = \beta_0 + \beta_1 z_i$, with z_i in some bounded interval for $1 \leq i \leq n$, and let $\bar{z}_i = a(z_i + c) = az_i + ac$ with $a > 0$. Then

$$\eta_i = \beta_0 + \beta_1 \left(\frac{1}{a}\bar{z}_i - c \right) = (\beta_0 - \beta_1 c) + \frac{\beta_1}{a}\bar{z}_i.$$

Thus,

$$\eta_i = \bar{\beta}_0 + \bar{\beta}_1 \bar{z}_i \text{ with } \bar{\beta}_0 = \beta_0 - \beta_1 c,\ \bar{\beta}_1 = \frac{\beta_1}{a}.$$

Without loss of generality, a and c can be chosen such that $0 \leq \bar{z}_i \leq 1$ for $1 \leq i \leq n$. Thus, it is desired to have $\delta_1 \leq \bar{\beta}_0 + \bar{\beta}_1 \bar{z}_i \leq \delta_2$ where $0 \leq \bar{z}_i \leq 1$ for $1 \leq i \leq n$. The region $\bar{R}$, say, in the $(\bar{\beta}_0, \bar{\beta}_1)$ plane for which these *all* hold is given by

$$\bar{R} \begin{cases} \delta_1 \leq \bar{\beta}_0 \leq \delta_2 \\ \\ \delta_1 - \bar{\beta}_0 \leq \bar{\beta}_1 \leq \delta_2 - \bar{\beta}_0. \end{cases}$$

Grid points can now be determined directly by transformation from the $(\bar{\beta}_0, \bar{\beta}_1)$ plane. Since the region $\bar{R}$ is horizontally simple, the grid on $\bar{R}$ can be obtained as follows:

1. Break up the $\bar{\beta}_0$ axis from δ_1 to δ_2.

2. Then, for each $\bar{\beta}_0$ value from step (1), break up the interval $\delta_1 - \bar{\beta}_0 \leq \bar{\beta}_1 \leq \delta_2 - \bar{\beta}_0$.

Then

$$\left.\begin{array}{c} \bar{\beta}_0 = \beta_0 - \beta_1 c \\ \bar{\beta}_1 = \frac{\beta_1}{a} \end{array}\right\} \Rightarrow \quad \begin{array}{c} \beta_0 = \bar{\beta}_0 + ca\bar{\beta}_1 \\ \beta_1 = a\bar{\beta}_1 \end{array}$$

This gives the grid points for (β_0, β_1) in the R region, say. Figure (4.1) depicts an example of this transformation.

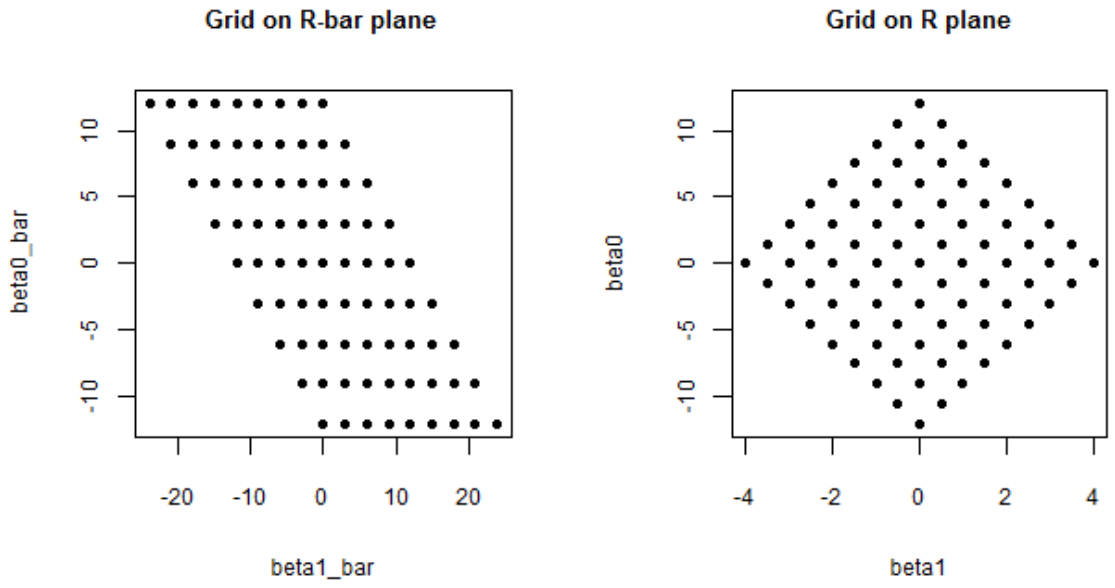

Figure 4.1: *Example of the Grid Transformation from the $\bar{R}$ Region to the R Region*

With the constraints specified by $\delta_1 = -12, \delta_2 = 12$, and with the predictor z values coming from the Uniform distribution on $(-3, 3)$, $a = 1/6$ and $c = 3$. Thus, $-12 \leq \bar{\beta}_0 \leq 12$ and $-24 \leq \bar{\beta}_1 \leq 24$ but restricted within the parallelogram shaped region $\bar{R}$, and this leads to $-12 \leq \beta_0 \leq 12$ and $-4 \leq \beta_1 \leq 4$ but restricted within the diamond shaped region R.

In this selected grid, twenty-one (β_1, β_2) values were chosen and the crisp maximin clusterings were calculated using the method described in section 3.2. The results are shown in

Figure 4.2. The plot is color coded in such a way that the points with like coloring correspond to the same maximin clustering. The numbered clusterings are listed in Appendix B.

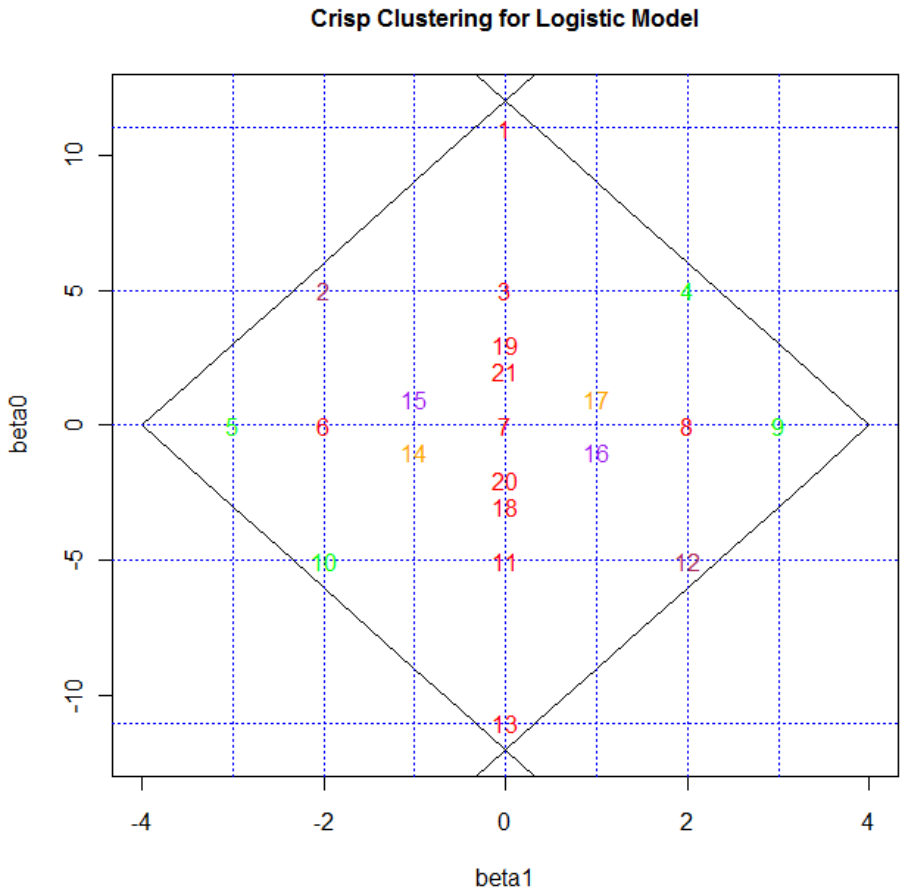

Figure 4.2: *Maximin Crisp Clusterings for the Logistic Model*

Using these crisp clusterings, a size analysis for the LRT was carried out to check the validity of the null χ^2 distribution using a grid search method. Before presenting the simulated size results corresponding to the twenty-one (β_1, β_2) values in Figure 4.2, the grid search method employed is discussed.

In particular, based on the grid developed in the preceding, estimation of the parameters

for alternative models of the sort given in Chapter 3 for the GLM is considered. That is,

$$\eta_i(s) = \sum_{j=1}^{q} z_i^j \beta_j + s\gamma_i(\beta) = \xi_i + s\gamma_i(\beta)$$

where

$$\xi_i = \eta_i(0) = \sum_{j=1}^{q} z_i^j \beta_j.$$

Let

$$\sigma_i^2(s) = \sigma_i^2(\eta_i(s)) = \sigma_i^2(\xi_i + s\gamma_i(\beta))$$
$$\mu_i(s) = \mu_i(\eta_i(s)) = \mu_i(\xi_i + s\gamma_i(\beta))$$
$$\sigma_i^2(0) = \sigma_i^2(\xi_i) \text{ and } \mu_i(0) = \mu_i(\xi_i).$$

Then $\mu_i(s) = m_i\pi_i(s)$ where $\pi_i(s) = \frac{e^{\eta_i(s)}}{1+e^{\eta_i(s)}}$

so that

$$\pi_i(s) = \frac{e^{\xi_i}e^{s\gamma_i(\beta)}}{1 + e^{\xi_i}e^{s\gamma_i(\beta)}} = \frac{e^{\xi_i}}{e^{\xi_i} + e^{-s\gamma_i(\beta)}}.$$

Next let $f(s)$ denote the derivative of the log-likelihood with respect to s for *fixed* β to obtain

$$f(s) = \sum_{i=1}^{n} (y_i - \mu_i(s))\gamma_i(\beta)$$
$$f(s) = \sum_{i=1}^{n} \left(y_i - \frac{m_i e^{\xi_i}}{e^{\xi_i} + e^{-s\gamma_i(\beta)}} \right) \gamma_i(\beta).$$

It is the root of $f(s) = 0$ that is of interest in order to obtain the MLE under the full model. Note that the derivative of $f(s)$ is less than zero. Also, since

$$\pi_i(s) = \frac{e^{\xi_i}}{e^{\xi_i} + e^{-s\gamma_i(\beta)}} = \frac{e^{\xi_i}}{e^{\xi_i} + e^{s(-\gamma_i(\beta))}}$$

it follows that

1. $\gamma_i(\beta) > 0$ gives
$$\lim_{s\to+\infty} \pi_i(s) = 1, \quad \lim_{s\to-\infty} \pi_i(s) = 0$$
2. $\gamma_i(\beta) < 0$ gives
$$\lim_{s\to+\infty} \pi_i(s) = 0, \quad \lim_{s\to-\infty} \pi_i(s) = 1$$
3. $\gamma_i(\beta) = 0$ gives
$$\pi_i(s) = \frac{e^{\xi_i}}{1 + e^{\xi_i}} = \pi_i(0).$$

Thus,

$$\lim_{s\to+\infty} f(s) = \sum_{i\in Positive} (y_i - m_i)\gamma_i(\beta) + \sum_{i\in Negative} y_i\gamma_i(\beta) + \sum_{i\in Zero} (y_i - m_i\pi_i(0))\gamma_i(\beta)$$
$$\lim_{s\to-\infty} f(s) = \sum_{i\in Positive} y_i\gamma_i(\beta) + \sum_{i\in Negative} (y_i - m_i)\gamma_i(\beta) + \sum_{i\in Zero} (y_i - m_i\pi_i(0))\gamma_i(\beta).$$

Note the two preceding limits are negative and positive valued, respectively. Further, $\sigma_i^2(s) = m_i\pi_i(s)(1 - \pi_i(s))$ so that unless $\gamma_i(\beta) = 0$, $\lim_{s\to\pm\infty} \sigma_i^2 = 0$. Based on the preceding, the following bisection method was used in estimating the value of s that maximizes the log-likelihood function under the constructed alternative:

1. Select negative s_1 and positive s_2, both large in magnitude.
2. Check the values of $f(s)$ at the chosen s_1 and s_2.

3. If $f(s_1)$ is not positive and/or $f(s_2)$ is not negative then adjust the s values.

4. Choose the midway value between s_1 and s_2, and check whether $f(s)$ is positive or negative.

5. If it is positive then assign s_1 for that chosen middle value. If it is negative then assign s_2 to that middle value.

6. Repeat steps 4 and 5 until the root is found.

This procedure gives the value of s that maximizes the log-likelihood function for fixed (β_0, β_1) on the selected grid. Then choosing the grid point corresponding to the maximal log-likelihood function provides the (approximated) MLE of the parameters (β_0, β_1, s) in the full model. Similarly, direct evaluation of the log-likelihood function for the null model across the grid provides the (approximated) MLE of the parameters (β_0, β_1) in the null model. Thus, the calculation of the log LRT statistic can be approximated for a simulated data vector.

Returning to the size analysis for the LRT corresponding to the twenty-one (β_1, β_2) values in Figure 4.2, the results are given in Figures 4.3, 4.4, 4.5, 4.6 for sample size $n = 50$, and in Figures 4.7, 4.8, 4.9, 4.10 for sample size $n = 100$. Different simulation parameters were considered in calculating the size values, including the number of simulations (B) and the density of the grid. Note the dimension of $T_\beta M$ is equal to two and the dimension of all between cluster-lack of fit subspaces is equal to one. Thus, χ^2_1 is the proposed null distribution and the specified nominal level was taken as .05.

According to the plots it can be concluded that the χ^2 distribution is particularly valid in the central part of the grid. With increasing values of the simulation parameters n, B and grid density, the desired nominal level is well approximated except at certain β settings near the grid boundary. In subsequent subsections, the four values of $\beta = (\beta_0, \beta_1)$ taken as

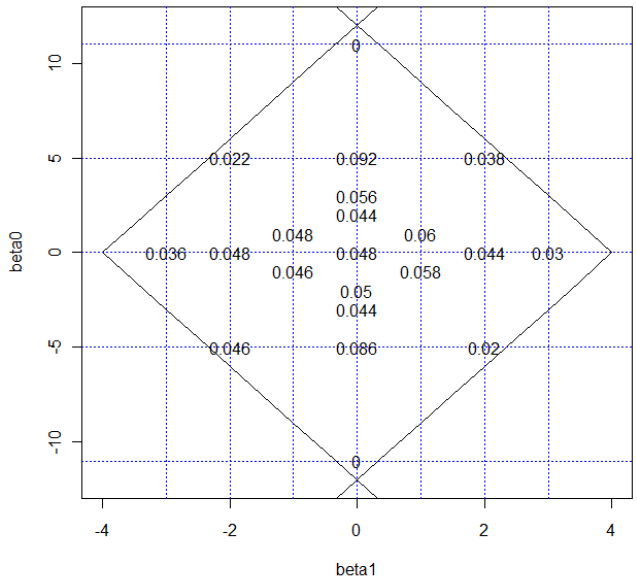

Figure 4.3: *LRT Size Values for Logistic Model with n=50, B=500 and 625 Grid Points*

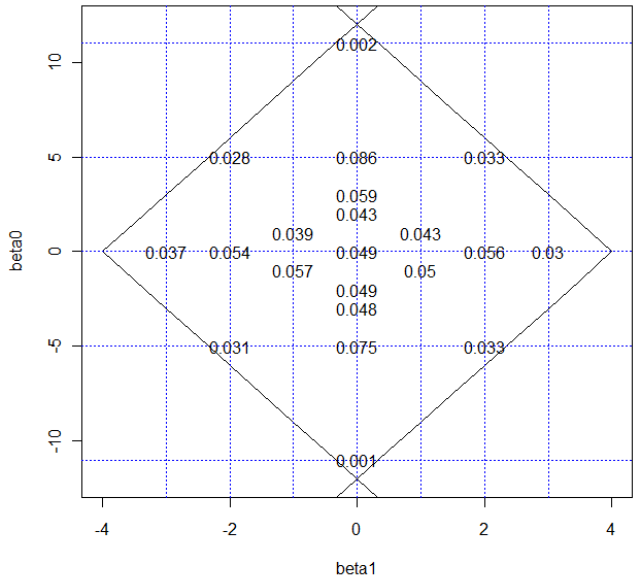

Figure 4.4: *LRT Size Values for Logistic Model with n=50, B=1000 and 625 Grid Points*

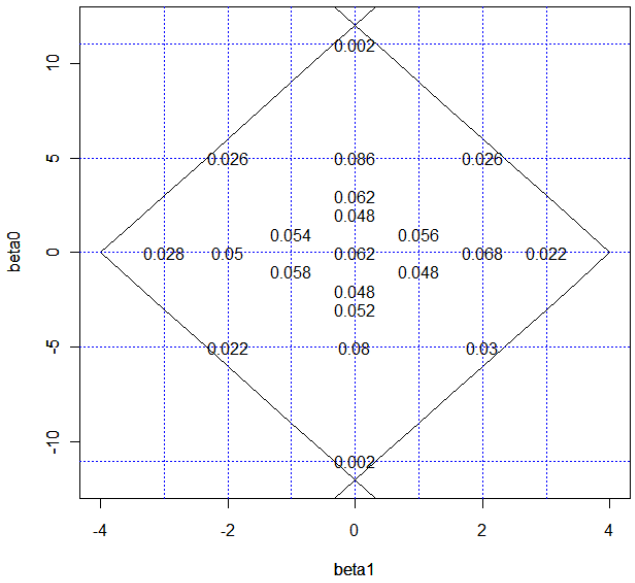

Figure 4.5: *LRT Size Values for Logistic Model with n=50, B=500 and 2401 Grid Points*

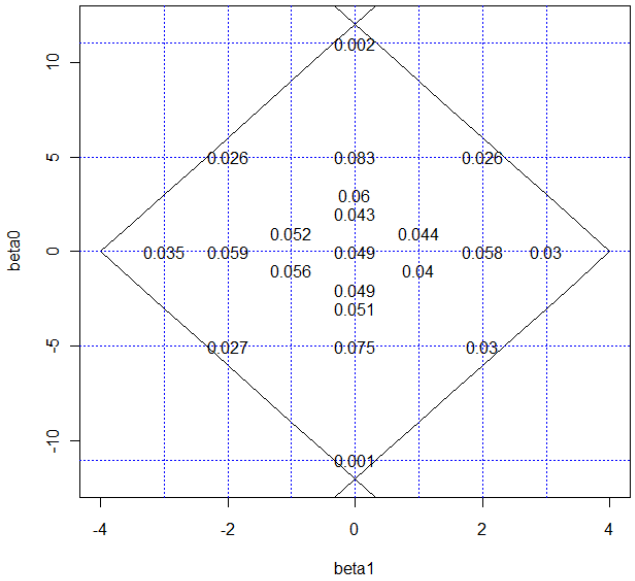

Figure 4.6: *LRT Size Values for Logistic Model with n=50, B=1000 and 2401 Grid Points*

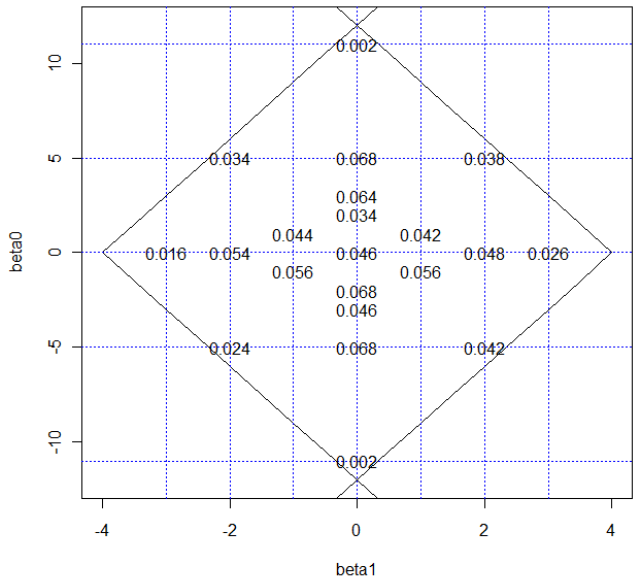

Figure 4.7: *LRT Size Values for Logistic Model with n=100, B=500 and 625 Grid Points*

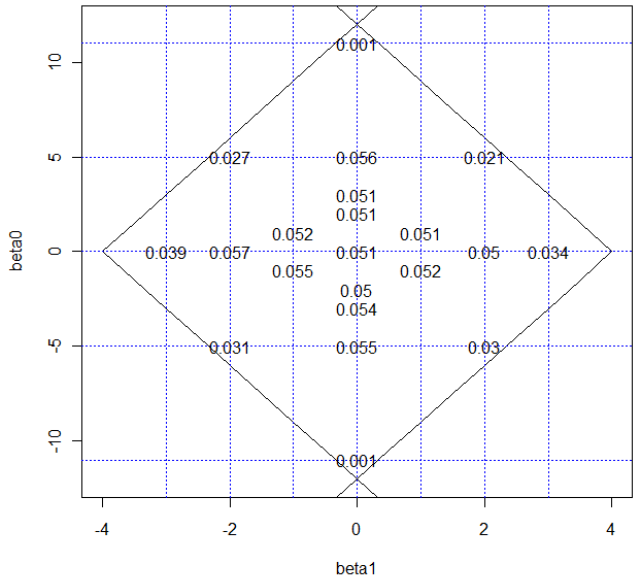

Figure 4.8: *LRT Size Values for Logistic Model with n=100, B=1000 and 625 Grid Points*

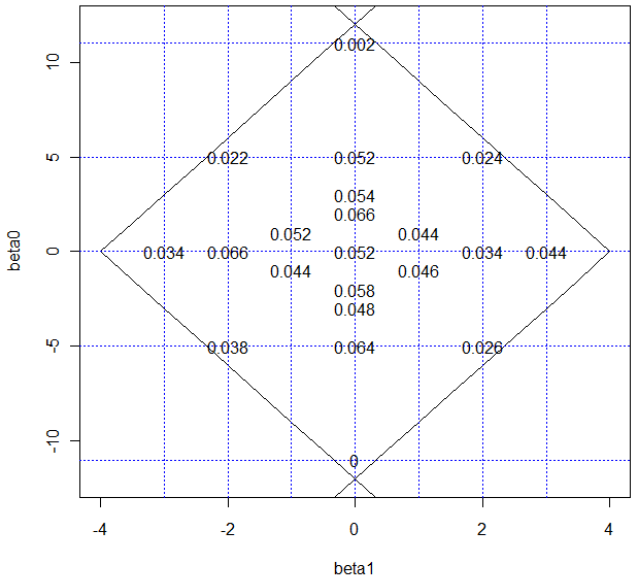

Figure 4.9: *LRT Size Values for Logistic Model with n=100, B=500 and 2401 Grid Points*

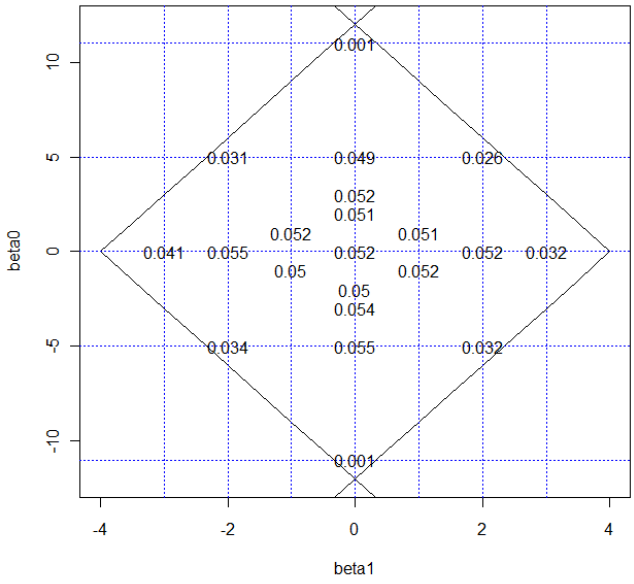

Figure 4.10: *LRT Size Values for Logistic Model with n=100, B=1000 and 2401 Grid Points*

$\{(-1,-1),(-1,1),(1,-1),(1,1)\}$ in the central part of the grid will be used for illustration purposes. Alternative full models based on the four multiple maximin clusterings will be considered subsequently, with fuzzy clusterings employed to facilitate the grid search for the MLE. Fuzzy clusterings, as discussed in Munasinghe (2010), are determined from the four maximin clusterings corresponding to the four selected β settings. Further, based on the selected cover of the predictor space, the dimension of the fuzzy clusterings is equal to three.

Table 4.1 gives the crisp maximin clusterings $Z_n(\beta)$ for the two sizes of data sets at each of the four selected β settings. The results indicate that the crisp clusterings are not constant across the chosen β settings, with considerable difference in selected clusters as n increases.

values of β	$Z_n(\beta)$	
	n=50	n=100
(-1,-1)	$\{x_1 : x_{17}\}$ $\{x_{18} : x_{33}\}$ $\{x_{34} : x_{50}\}$	$\{x_1 : x_{37}\}$ $\{x_{38} : x_{62}\}$ $\{x_{63} : x_{100}\}$
(-1,1)	$\{x_1 : x_{17}\}$ $\{x_{18} : x_{29}\}$ $\{x_{30} : x_{50}\}$	$\{x_1 : x_{37}\}$ $\{x_{38} : x_{80}\}$ $\{x_{81} : x_{100}\}$
(1,-1)	$\{x_1 : x_{17}\}$ $\{x_{18} : x_{29}\}$ $\{x_{30} : x_{50}\}$	$\{x_1 : x_{37}\}$ $\{x_{38} : x_{80}\}$ $\{x_{81} : x_{100}\}$
(1,1)	$\{x_1 : x_{17}\}$ $\{x_{18} : x_{33}\}$ $\{x_{34} : x_{50}\}$	$\{x_1 : x_{37}\}$ $\{x_{38} : x_{62}\}$ $\{x_{63} : x_{100}\}$

Table 4.1: *Maximin Clusterings for the Logistic Model with n=50 and n=100 for the Selected* (β_0, β_1) *Values*

4.1.1 Data Generation by Perturbing the Proposed Logistic Model (Using the Constructed Full Model)

The data were simulated by using the perturbed logistic model with $\eta_i = \beta_0 + \beta_1 z_i + s\gamma_i$ where (γ_i) is a basis vector for the between-cluster lack of fit subspace at selected values of $\beta = (\beta_0, \beta_1)$. The grid is chosen as discussed above with $\delta_1 = -12, \delta_2 = 12$ and with the z values drawn from the Uniform distribution on $(-3, 3)$. For each of twenty-five equally separated values of $\bar{\beta}_0$ between $\delta_1 = -12$ and $\delta_2 = 12$, twenty-five corresponding and equally separated values of $\bar{\beta}_1$ were chosen in the $\bar{R}$ region, resulting in 625 grid points for approximation of the MLE parameter estimates. The power of the LRT statistic with nominal size .05 was investigated. The results were obtained by using $B = 500$ simulations, and are given in Table 4.2 for different values of (β_0, β_1) with n=50. Table 4.2 gives the power of the LRT for both the cases of a single maximin clustering as well as for fuzzy clusterings. Also, the corresponding power values for the Hosmer-Lemeshow goodness of fit test for the logistic model was included for comparison. It can be observed from Table 4.2 that the LRT has superior power as compared to the Hosmer-Lemeshow test for the cases considered in this simulation study, especially for smaller values of the parameter s. Further, there is not an observed difference in the use of a single maximin clustering as compared to the use of fuzzy clusterings, except for the $(\beta_0, \beta_1) = (1, 1)$ case.

4.1.2 Data Generation by a Functionally Different Model than the Constructed Full Model

A functionally different model given by $\eta_i = \beta_0 + \beta_1 z_i + \alpha z_i^2$ is next considered as the data generator in checking the simulated power of several test statistics, including the LRT statistic, Wald statistics, the deviance, the Hosmer-Lemeshow statistic and the $S(y, \hat{\mu})$ statistic (cf section 3.1), which is based on the information metric.

(β_0,β_1)	s	with single maximin clusters	with multiple maximin clusters	H-L test	(β_0,β_1)	s	with single maximin clusters	with multiple maximin clusters	H-L test
(-1,-1)	0	0.036	0.036	0.036	(-1,1)	0	0.074	0.074	0.052
	0.4	0.236	0.236	0.072		0.4	0.164	0.164	0.074
	0.8	0.694	0.694	0.254		0.8	0.534	0.534	0.188
	1.2	0.95	0.95	0.58		1.2	0.84	0.84	0.436
	1.6	0.998	0.998	0.848		1.6	0.97	0.97	0.724
	2	1	1	0.972		2	1	1	0.9
	2.4	1	1	0.998		2.4	1	1	0.992
	2.8	1	1	1		2.8	1	1	0.998
	3.2	1	1	1		3.2	1	1	1
(1,-1)	0	0.056	0.056	0.052	(1,1)	0	0.052	0.052	0.034
	0.4	0.178	0.178	0.088		0.4	0.238	0.168	0.122
	0.8	0.532	0.532	0.242		0.8	0.736	0.58	0.344
	1.2	0.868	0.868	0.482		1.2	0.966	0.902	0.738
	1.6	0.984	0.984	0.778		1.6	1	0.992	0.96
	2	1	1	0.96		2	1	1	0.998
	2.4	1	1	0.994		2.4	1	1	1
	2.8	1	1	0.998		2.8	1	1	1
	3.2	1	1	1		3.2	1	1	1

Table 4.2: *Power of the LRT and Hosmer-Lemeshow Tests for Logistic Model with* $n = 50$ *at Selected* (β_0, β_1, s) *Values*

Recall from Chapter 3 that $S(y,\mu) = \|(y_i - \mu_i)\frac{\partial}{\partial \mu_i}\|^2_{Information\ at\ \mu}$, so that $S(y,\mu) = \sum_{i=1}^{n} \frac{(y_i-\mu_i)^2}{\sigma_i^2(\mu)}$ for the logistic model with $\mu_i = m_i\pi_i$ and $\sigma_i^2 = m_i\pi_i(1-\pi_i)$ where m_i is the number of Bernoulli trials and π_i is the probability of success. A test statistic for goodness of fit is thus $S(y,\hat{\mu})$ where $\hat{\mu}$ is the MLE of the proposed model. This can be seen to be equal to the Pearson χ^2 goodness of fit test statistic as given in Dobson and Barnett (2008), for example. Note that the Hosmer-Lemeshow, deviance and S statistics do not depend on a constructed alternative model, and hence do not depend on specified clusterings. For the LRT and Wald statistics, the single maximin power clustering for $\beta = (1,1)$ was used to investigate simulated power for these test statistics. The χ^2 distribution was used to

α	Power of the Test					
	LRT	$Wald_B$	$Wald_S$	S	deviance	H-L
0	0.052	0.048	0.046	0.058	0.054	0.04
2	0.41	0.324	0.308	0.072	0.148	0.16
4	0.548	0.352	0.326	0.074	0.018	0.274
8	0.352	0.432	0.422	0.074	0.02	0.228
12	0.242	0.188	0.214	0.282	0.002	0.116
16	0.122	0.0.038	0.034	0.420	0	0.046
20	0.012	0.004	0.002	0.0.124	0	0.016
24	0.002	0	0	0.022	0	0.006
28	0	0	0	0	0	0
32	0	0	0	0	0	0
36	0	0	0	0	0	0

Table 4.3: *Power of the LRT,* $Wald_B, Wald_S, S(y, \hat{\mu})$*, Deviance and Hosmer-Lemeshow Statistics for Logistic Model with* $n = 50$ *for* $(\beta_0, \beta_1, \alpha) = (1, 1, \alpha)$ *Values*

determine critical points, set according to the .05 nominal level.

To demonstrate the usefulness of the bounded grid discussed in the first part of this chapter, the simulated power results are given in Table 4.3, and graphically illustrated in Figure 4.11, where the data generators do not satisfy the constraints imposed by the grid. Note that the predictor space for these simulated powers was taken as Uniform on $(0, 1)$. Recall that the grid restricts the search for MLE estimates of the parameters in such a way as to avoid parameter values that correspond to (nearly) deterministic densities. Although the simulated size for all of the test statistics approximates the desired nominal size equal to .05, it can be observed that the simulated power ultimately decreases to zero for increasing values of the parameter α in the data generator. Upon further investigation, when the grid is allowed sufficiently large, the estimates for (β_0, β_1) were noted to become large and thus allow the null model to fit well, even though the data came from a rogue data generator. Accordingly, for the cases considered, the power decreases to zero and remains so.

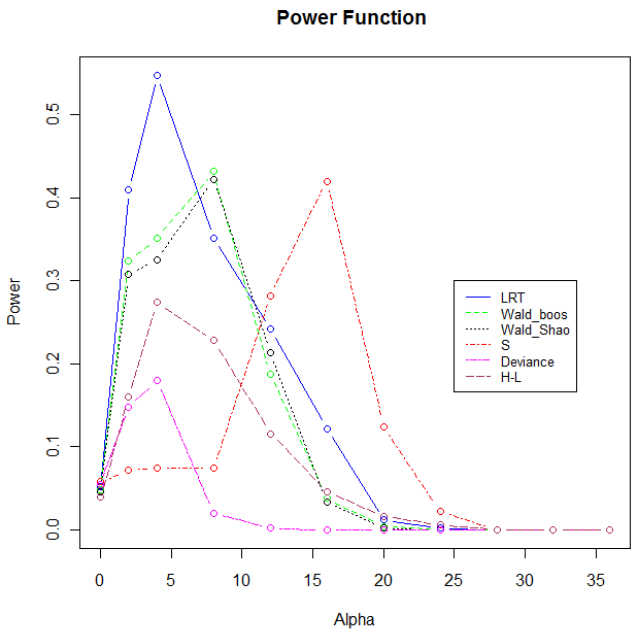

Figure 4.11: *Power of LRT,* $Wald_{Boos}, Wald_{Shao}, S(y, \hat{\mu})$*, Deviance and Hosmer-Lemeshow Tests for Logistic Model with* $n = 50$ *at Selected* $(\beta_0, \beta_1, \alpha)$ *Values*

For comparison, the results given in Table 4.4 reflect the case that the data generators do satisfy the bounded constraints of the grid. In particular, the simulated power for the test statistics is comparable, with the simulated power increasing to one for large α in all cases. In addition, it can be observed that the simulated power for the deviance and Hosmer-Lemeshow tests lags behind the other tests for small values of the parameter α. Note that the predictor space for these simulated powers was taken as Uniform on $(-3, 3)$.

An analytical explanation of the phenomena observed in Table 4.3, and graphically illustrated in Figure 4.11, is presented next. The analysis is based on the information distance (squared) (cf section 3.1) involving limits to infinity.

Consider the test statistic defined above as $S(y, \hat{\mu})$ where $\hat{\mu}$ is the MLE estimate of the mean, with $\mu_i = m_i\pi_i$ and variance $\sigma_i^2 = m_i\pi_i(1 - \pi_i)$. Then

α	Power of the Test					
	LRT	$Wald_B$	$Wald_S$	S	Deviance	H-L
0	0.044	0.05	0.044	0.056	0.074	0.04
0.05	0.222	0.22	0.2	0.156	0.118	0.112
0.1	0.638	0.584	0.628	0.694	0.346	0.548
0.15	0.884	0.782	0.892	0.976	0.724	0.932
0.2	0.976	0.96	0.978	0.982	0.956	0.998
0.25	0.998	0.994	0.998	1	0.998	1
0.3	1	1	1	1	1	1
0.35	1	1	1	1	1	1
0.4	1	1	1	1	1	1
0.45	1	1	1	1	1	1
0.5	1	1	1	1	1	1

Table 4.4: *Power of the LRT,* $Wald_B, Wald_S, S(y, \hat{\mu})$*, Deviance and Hosmer-Lemeshow Statistics for Logistic Model with* $n = 50$ *for* $(\beta_0, \beta_1, \alpha) = (1, 1, \alpha)$ *Values*

$$S(y, \pi) = \sum_i \frac{1}{m_i \pi_i (1 - \pi_i)} (y_i - m_i \pi_i)^2$$

$$S(y, \pi) = \sum_i \frac{m_i}{\pi_i (1 - \pi_i)} \left(\frac{y_i}{m_i} - \pi_i \right)^2 .$$

The following cases are considered, assuming that the null model is true. i.e. the limits for π are along the null model.

1. For $y_u = (m_1, m_2, ..., m_n)$,

   $S(y_u, \pi) = \sum_i \frac{m_i}{\pi_i (1-\pi_i)} (1 - \pi_i)^2 = \sum_i \frac{m_i}{\pi_i} (1 - \pi_i)$

   so that $\lim_{\pi \to 1} S(y_u, \pi) = 0$.

2. For $y_l = (0, 0, ..., 0)$,

$S(y_l, \pi) = \sum_i \frac{m_i}{\pi_i(1-\pi_i)}\pi_i^2 = \sum_i \left(\frac{m_i}{1-\pi_i}\right)\pi_i$

so that $\lim_{\pi \to 0} S(y_l, \pi) = 0.$

3. Suppose $y_\infty = (y_i)$ with $y_i = m_i$ or $y_i = 0$. Then

   $S(y_\infty, \pi) = \sum_{y_i = m_i} \frac{m_i}{\pi_i}(1-\pi_i) + \sum_{y_i = 0}\left(\frac{m_i}{1-\pi_i}\right)\pi_i$

   so that $\lim_{\begin{pmatrix} y_i = m_i \text{ gives } \pi_i \to 1 \\ y_i = 0 \text{ gives } \pi_i \to 0 \end{pmatrix}} S(y_\infty, \pi) = 0.$

4. Suppose $\pi_i \to 1$ or $\pi_i \to 0$ for each i

   but $\exists i$ with $m_i = y_i$ but $\pi_i \to 0$ or

   $\exists i$ with $y_i = 0$ but $\pi_i \to 1$.

   Then $\lim_{\pi} S(y_\infty, \pi) = +\infty.$

5. Suppose $y_i \neq m_i$ and $y_i \neq 0$ for some i and $\pi_i \to 0$ or $\pi_i \to 1$ for that i.

   Then $\lim_{\pi} S(y_\infty, \pi) = +\infty.$

From the above analysis, it is clear that the extreme situations (i.e. basically deterministic) present anomalies when assessing the adequacy of the logistic model. In order to simulate data corresponding to controlled π_i values in the full range of (0,1), without going to extremes, the method employed by Hosmer et al. (1997) was used to further investigate power for the test statistics.

In particular, the data generator for the logistic model with $\eta = \beta_0 + \beta_1 z + \beta_2 z^2$, for chosen values of the three parameters determined by $\pi(-1.5) = 0.05$, $\pi(3) = 0.95$ and $\pi(-3) = J$ where $J = 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08$, was used in a power simulation study

J	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08
β_0	-1.138	-1.489	-1.697	-1.846	-1.963	-2.059	-2.142	-2.214
β_1	1.257	1.139	1.070	1.020	0.981	0.949	0.922	0.898
β_2	0.035	0.113	0.159	0.192	0.218	0.239	0.258	0.274

Table 4.5: *Values of $\beta_0, \beta_1, \beta_2$ with Increasing J Values*

for the tests under consideration. Here the predictor z values are generated as Uniform on $(-3, 3)$, and π is evaluated corresponding to the three values $-1.5, -3, 3$. Eight models are thus considered to investigate the simulated power with omission of a squared predictor term. This method assures that the generated models with lack of linearity in the logit function become progressively more profound with increasing J values. Table 4.5 gives values of $\beta_0, \beta_1, \beta_2$ with increasing J values.

The grid is again selected such that $\delta_1 = -12, \delta_2 = 12$. With this setup, the power is investigated using a single maximin clustering for different values of J using the same test statistics as before. The results are given in Table 4.6. The clustering generated at $(\beta_0, \beta_1) = (1, 1)$ was again used in constructing the general alternative models. Note that the S, deviance and Hosmer-Lemeshow statistics do not depend on clusterings or the specification of an alternative model.

The simulated power for the test statistics is seen to be comparable, although the deviance and Hosmer-Lemeshow tests lag in power for lower J values. In addition, the simulated power corresponding to lower J values for the S statistic is best in Table 4.6.

Next the power is investigated with the same setup but using multiple maximin clusterings. The power values are given in Table (4.7). The same four (β_0, β_1) points that have been used previously were chosen in the grid for calculating the multiple clusterings. Note again that S, the deviance and Hosmer-Lemeshow statistics do not involve alternative models in checking for LOF, and hence also do not depend on the clusterings, thus giving the same

J	Power of the Test					
	LRT	$Wald_B$	$Wald_S$	S	deviance	H-L
0.01	0.142	0.154	0.158	0.312	0.094	0.1
0.02	0.866	0.844	0.874	0.878	0.406	0.74
0.03	0.976	0.98	0.978	0.968	0.782	0.976
0.04	0.998	0.998	1	0.998	0.934	1
0.05	1	1	1	1	0.986	1
0.06	1	1	1	1	0.998	1
0.07	1	1	1	1	0.998	1
0.08	1	1	1	1	1	1

Table 4.6: *Power of the LRT,* $Wald_B, Wald_S, S(y, \hat{\mu})$*, Deviance and Hosmer-Lemeshow Statistics for Logistic Model with* $n = 50$ *for Different* J *Values Using a Single Maximin Clustering*

power values as in the case for single maximin clusterings.

The conclusions using the multiple maximin clusterings are somewhat similar to the single maximin case. However, with clusterings selected more nearly equal to the chosen (β_0, β_1), the multiple clusterings setup may provide better results. In any case, a measure of robustness is illustrated with the current choice of clusterings.

4.1.3 n-Asymptotics for Logistic Model

Thus far, consideration has been given to the case of m-asymptotics in the logistic regression model. In this subsection, the case of n-asymptotics will be considered. The simulations are carried out with $n = 100$ and $m_i = 1$ to illustrate the n-asymptotics case. The same setup for the grid search to determine approximate MLE parameter estimates will be used as before. Also, the clusterings were selected and a size analysis is carried out as before. The size at selected grid points is shown in Figures 4.12, 4.13, 4.14, 4.15.

According to the plots it can be concluded that the χ^2 distribution provides a reasonable

J	Power of the Test					
	LRT	$Wald_B$	$Wald_S$	S	Deviance	H-L
0.01	0.128	0.158	0.132	0.312	0.094	0.1
0.02	0.766	0.742	0.792	0.878	0.406	0.74
0.03	0.926	0.93	0.934	0.968	0.782	0.976
0.04	0.98	0.982	0.98	0.998	0.934	1
0.05	0.996	0.994	0.996	1	0.986	1
0.06	1	1	1	1	0.998	1
0.07	1	1	1	1	0.998	1
0.08	1	1	1	1	1	1

Table 4.7: *Power of the LRT,* $Wald_B, Wald_S, S(y, \hat{\mu})$*, Deviance and Hosmer-Lemeshow Statistics for Logistic Model with* $n = 50$ *for Different* J *Values Using Multiple Maximin Clusterings*

approximation to the null distribution of the LRT statistic, for the values of the simulation parameters B and grid density considered. The desired nominal level of .05 is reasonably approximated except at certain β settings near the grid boundary.

Simulated power for the LRT is next investigated with constructed full models based on between-cluster lack of fit. Single maximin clusterings and multiple maximin clusterings corresponding to the same four values of $\beta = (\beta_0, \beta_1)$ (i.e. $\{(-1,-1),(-1,1),(1,-1),(1,1)\}$) will be used for illustration purposes. Note that the fuzzy clusterings were also calculated using the same four points considered previously.

The simulated power is given in Table 4.8. The power for the Hosmer-Lemeshow test is also included for comparison. It can be observed from Table 4.8 that the LRT generally has superior power as compared to the Hosmer-Lemeshow test for the cases considered in this simulation study, especially for smaller values of the parameter s.

Next the data are generated using a functionally different model as in m-asymptotics case. The same model, $\eta = \beta_0 + \beta_1 z + \beta_2 z^2$, is used to generate the data in calculating the power

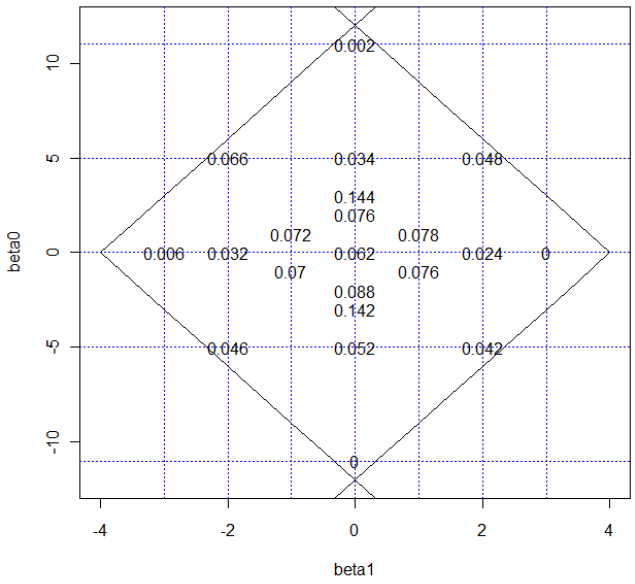

Figure 4.12: *Size Values for Logistic Model in the n-asymptotics Case with n=100, B=500 and 625 Grid Points*

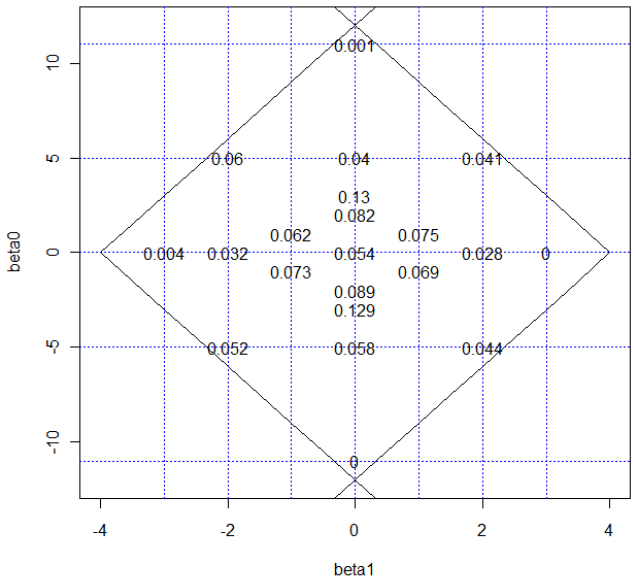

Figure 4.13: *Size Values for Logistic Model in the n-asymptotics Case with n=100, B=1000 and 625 Grid Points*

(β_0,β_1)	s	with single maximin clusters	with multiple maximin clusters	H-L test	(β_0,β_1)	s	with single maximin clusters	with multiple maximin clusters	H-L test
(-1,-1)	0	0.066	0.07	0.032	(-1,1)	0	0.076	0.082	0.052
	0.05	0.382	0.19	0.064		0.05	0.3	0.146	0.08
	0.1	0.91	0.422	0.272		0.1	0.762	0.26	0.242
	0.15	0.998	0.714	0.634		0.15	0.98	0.41	0.652
	0.2	1	0.912	0.92		0.2	1	0.576	0.934
	0.25	1	0.99	0.996		0.25	1	0.8	0.992
	0.3	1	0.998	1		0.3	1	0.94	1
	0.35	1	1	1		0.35	1	0.986	1
	0.4	1	1	1		0.4	1	1	1
(1,-1)	0	0.072	0.072	0.046	(1,1)	0	0.078	0.088	0.05
	0.05	0.286	0.264	0.138		0.05	0.396	0.618	0.236
	0.1	0.792	0.834	0.432		0.1	0.938	1	0.678
	0.15	0.99	0.99	0.816		0.15	1	1	0.984
	0.2	1	1	0.974		0.2	1	1	1
	0.25	1	1	1		0.25	1	1	1
	0.3	1	1	1		0.3	1	1	1

Table 4.8: *Power of the LRT and Hosmer-Lemeshow Tests for Logistic Model with* $n = 100, m_i = 1$ *at Selected* (β_0, β_1, s) *Values*

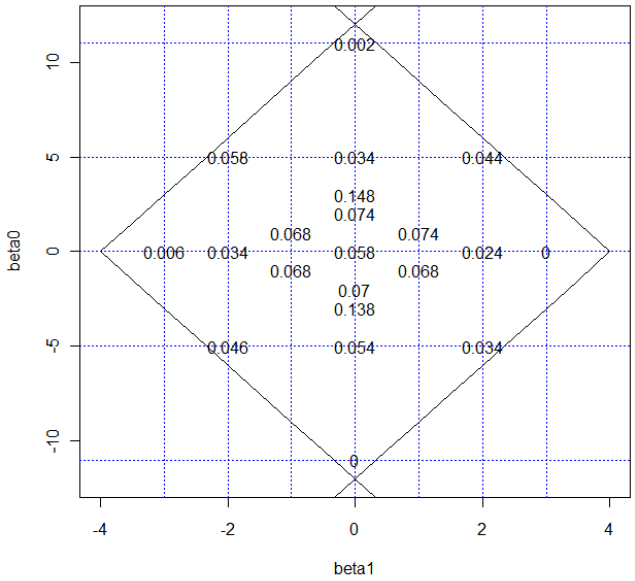

Figure 4.14: *Size Values for Logistic Model in the n-asymptotics Case with n=100, B=500 and 2401 Grid Points*

of the LRT, S, Wald and Hosmer-Lemeshow statistics. The data generator as explained in section 4.1.2 is used with different values of J controlling the π_i values as before. The corresponding $(\beta_0, \beta_1, \beta_2)$ values are given in Table 4.9, along with the simulated power values in Table 4.10 for the single maximin clusterings and Table 4.11 for multiple maximin clusterings. Note that the S and Hosmer-Lemeshow statistics do not depend on the clustering or an alternative model, thus giving the same values in both cases.

The simulated power for the test statistics is seen to be comparable, except for the fluctuations for the S statistic. Notably, the LRT has comparable power under both m- and n-asymptotics.

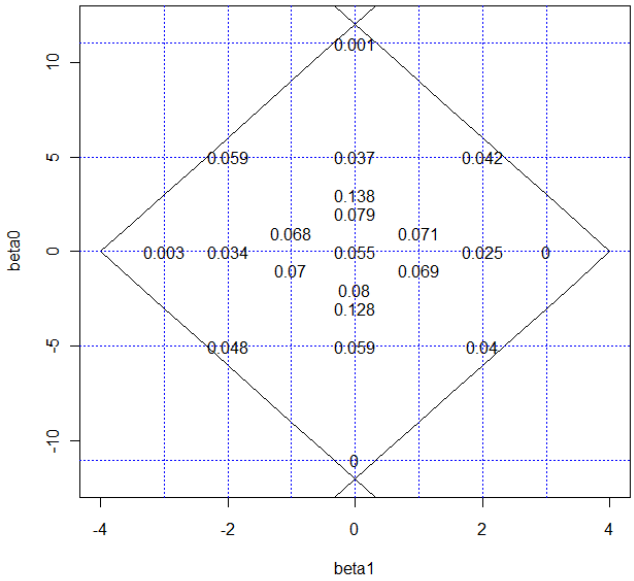

Figure 4.15: *Size Values for Logistic Model in the n-asymptotics Case with n=100, B=1000 and 2401 Grid Points*

J	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
β_0	-2.337	-2.742	-3.012	-3.232	-3.435	-3.638	-3.859	-4.128	-4.534
β_1	0.857	0.722	0.632	0.558	0.491	0.423	0.35	0.26	0.125
β_2	0.301	0.391	0.451	0.5	0.545	0.59	0.639	0.699	0.789

Table 4.9: *Values of $\beta_0, \beta_1, \beta_2$ with Increasing J Values for n-asymptotics*

4.2 Poisson Regression Model

Another exponential family model involves the Poisson regression model, which represents a generalized linear model with each response variable equal to the count or frequency of events associated with the corresponding predictor setting. The canonical link function is the natural logarithm transformation, which with one predictor variable z is given by

J	Power of the Test				
	LRT	$Wald_B$	$Wald_S$	S	H-L
0.1	0.402	0.552	0.428	0.632	0.476
0.2	0.636	0.718	0.66	0.772	0.728
0.3	0.798	0.812	0.82	0.85	0.834
0.4	0.894	0.874	0.902	0.834	0.914
0.5	0.96	0.9	0.956	0.802	0.956
0.6	0.976	0.906	0.978	0.78	0.972
0.7	0.996	0.95	0.994	0.744	0.996
0.8	0.998	0.97	0.998	0.69	0.998
0.9	1	1	1	0.454	0.998

Table 4.10: *Power of the LRT,* $Wald_B, Wald_S, S(y, \hat{\mu})$ *and Hosmer-Lemeshow Statistics for Logistic Model with* $n = 100$ *for Different* J *Values Using a Single Maximin Clustering*

J	Power of the Test				
	LRT	$Wald_B$	$Wald_S$	S	H-L
0.1	0.466	0.65	0.496	0.632	0.476
0.2	0.708	0.822	0.728	0.772	0.728
0.3	0.836	0.888	0.85	0.85	0.834
0.4	0.92	0.926	0.93	0.834	0.914
0.5	0.968	0.944	0.972	0.802	0.956
0.6	0.984	0.964	0.984	0.78	0.972
0.7	0.998	0.97	0.988	0.744	0.996
0.8	0.998	0.98	0.996	0.69	0.998
0.9	1	0.996	0.998	0.454	0.998

Table 4.11: *Power of the LRT,* $Wald_B, Wald_S, S(y, \hat{\mu})$ *and Hosmer-Lemeshow Statistics for Logistic Model with* $n = 100$ *for Different* J *Values Using a Multiple Maximin Clustering*

$$\eta = log(\lambda) = \beta_0 + \beta_1 z.$$

The simulations involve n independent Poisson observations with parameter λ_i where $1 \leq i \leq n$. Properties of the Poisson model with its canonical link function include

$$\begin{aligned} l(y, \eta) &= \sum_{i=1}^{n} [y_i \eta_i - e^{\eta_i} - log(y_i!)] \\ \varphi(\eta) &= \sum_{i=1}^{n} e^{\eta_i} \\ \eta_i &= log\ \lambda_i, \quad \lambda_i > 0 \\ \mu_i &= \frac{\partial \varphi}{\partial \eta_i} = e^{\eta_i} = \lambda_i \\ \sigma_i^2 &= \frac{\partial^2 \varphi}{\partial \eta_i^2} = e^{\eta_i} = \lambda_i \end{aligned}$$

The power of the LRT statistic, the deviance (cf Dobson and Barnett (2008)) and the test statistic S based on the information metric presented in Chapter 3 was investigated for detecting between-cluster lack of fit. In addition, the power for these tests was assessed when the data was generated by a functionally different model as compared to the constructed full model based on between-cluster lack of fit. The predictor variable was generated as Uniform on (-3,3) with size $n = 50$ and $n = 100$ (cf Appendix A).

Similar to logistic regression, an appropriate grid over which to search the β values for a Poisson regression model is considered. The method used in section 4.1 can also be used for the Poisson model by considering constraints on the mean of the distribution. In particular, with $\lambda > 0$ and excluding larger values of λ (i.e. excluding $\lambda > 10^5$, for example) leads to the approximate values of δ_1 and δ_2 as 1 and 12, respectively. With the z values drawn from the Uniform distribution on $(-3, 3)$, $a = 1/6$ and $c = 3$, $\bar{\beta}_0, \bar{\beta}_1$ range within the $\bar{R}$ region

according to $1 \leq \bar{\beta}_0 \leq 12$ and $-11 \leq \bar{\beta}_1 \leq 11$, and by transformation, β_0, β_1 range within the R region according to $1 \leq \beta_0 \leq 12$ and $-\frac{11}{6} \leq \beta_1 \leq \frac{11}{6}$.

In this selected grid, seventeen (β_0, β_1) values were chosen and the crisp maximin clusterings were calculated using the method described in section 3.2. The results are shown in Figure 4.16. The plot is color coded in such a way that the points with like coloring correspond to the same maximin clustering. The numbered clusterings are listed in Appendix B.

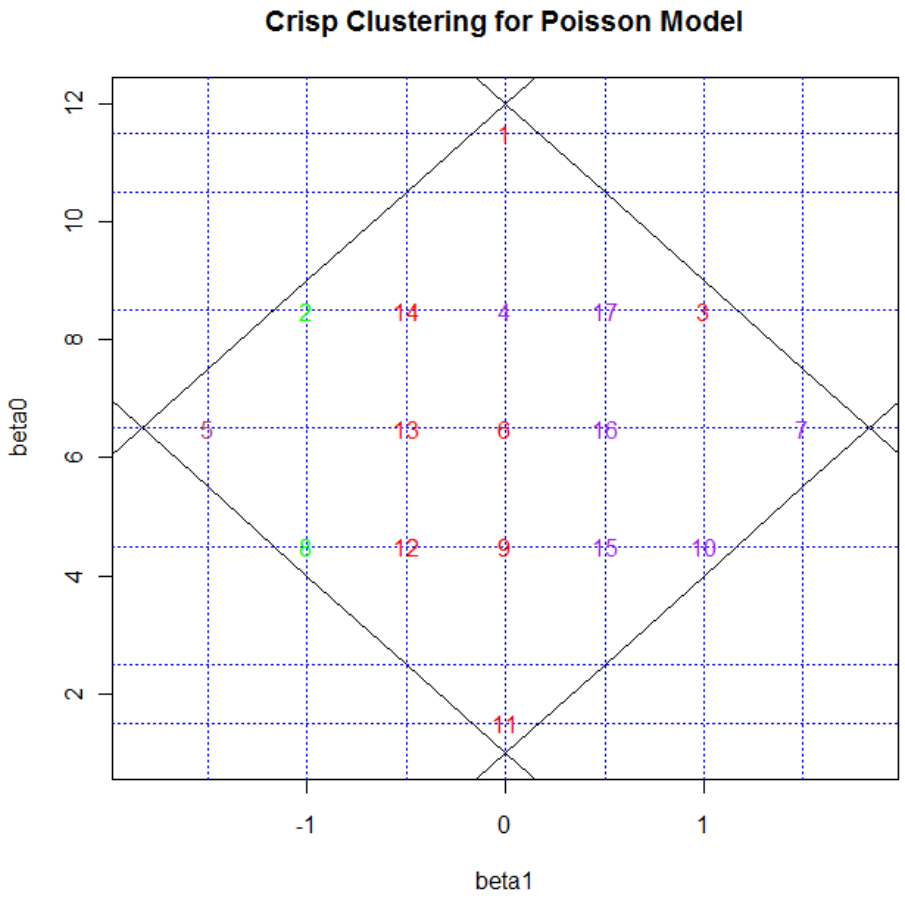

Figure 4.16: *Maximin Crisp Clusterings for the Poisson Model*

Using these crisp clusterings, a size analysis for the LRT was carried out to check the validity of the null χ^2 distribution. Approximation of the MLE of the parameters for alternative

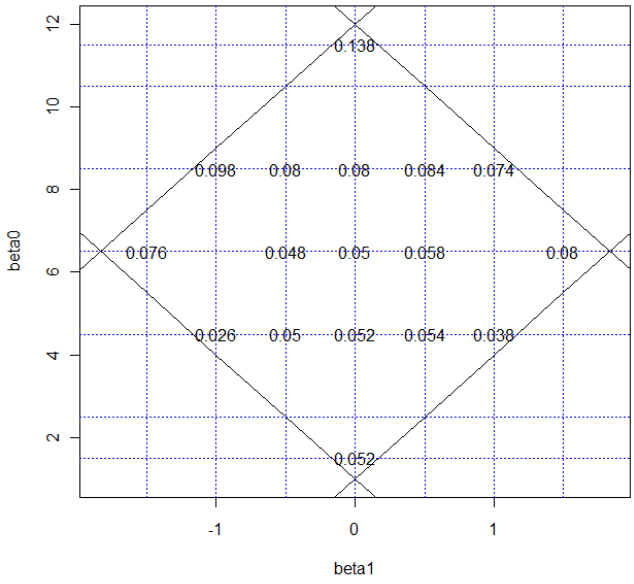

Figure 4.17: *LRT Size Values for Poisson Model with n=50, B=500 and 529 Grid Points*

models of the sort given in Chapter 3 for the GLM is considered, using a grid search based on the bisection method as in section 4.1. The simulated size for the LRT corresponding to the seventeen (β_0, β_1) values in Figure 4.16 is given in Figures 4.17, 4.18, 4.19, 4.20 for sample size $n = 50$, and Figures 4.21, 4.22, 4.23, 4.24 for sample size $n = 100$. Different simulation parameters were considered in calculating the size values, including the number of simulations (B) and the density of the grid. Note the dimension of $T_\beta M$ is equal to two and the dimension of all between cluster-lack of fit subspaces is equal to one. Thus, χ^2_1 is the proposed null distribution and the specified nominal level was taken as .05.

According to the plots it can be concluded that the χ^2 distribution is particularly valid in the lower central part of the grid. With increasing values of the simulation parameters n, B and grid density, the desired nominal level is well approximated except at certain β settings near the grid boundary. In subsequent subsections, the four values of $\beta = (\beta_0, \beta_1)$ taken as $\{(4.5, -0.5), (4.5, 0.5), (6.5, -0.5), (6.5, 0.5)\}$ in the lower central part of the grid will be

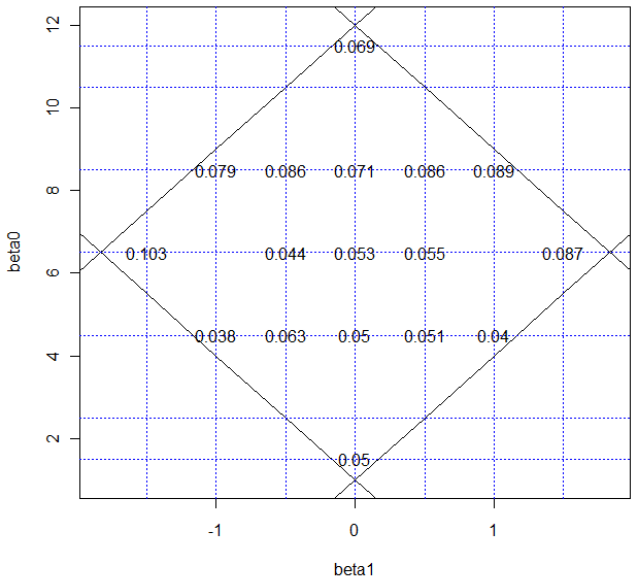

Figure 4.18: *LRT Size Values for Poisson Model with n=50, B=1000 and 529 Grid Points*

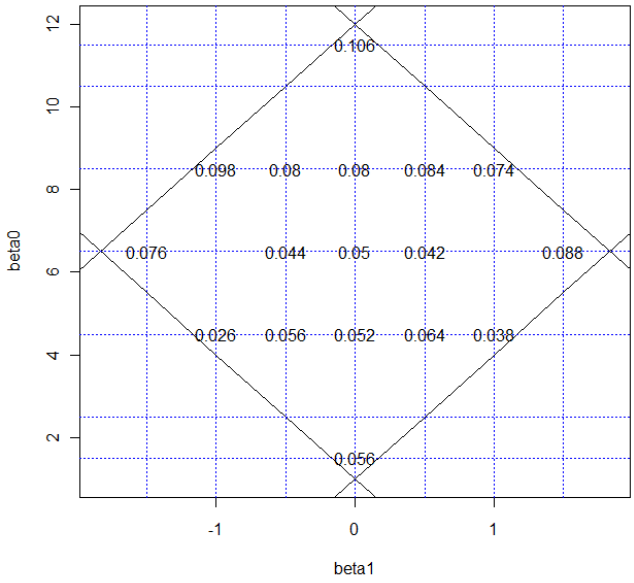

Figure 4.19: *LRT Size Values for Poisson Model with n=50, B=500 and 2025 Grid Points*

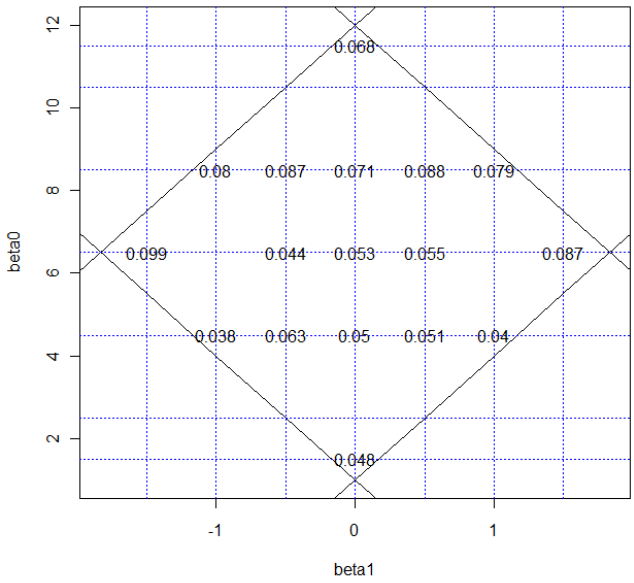

Figure 4.20: *LRT Size Values for Poisson Model with n=50, B=1000 and 2025 Grid Points*

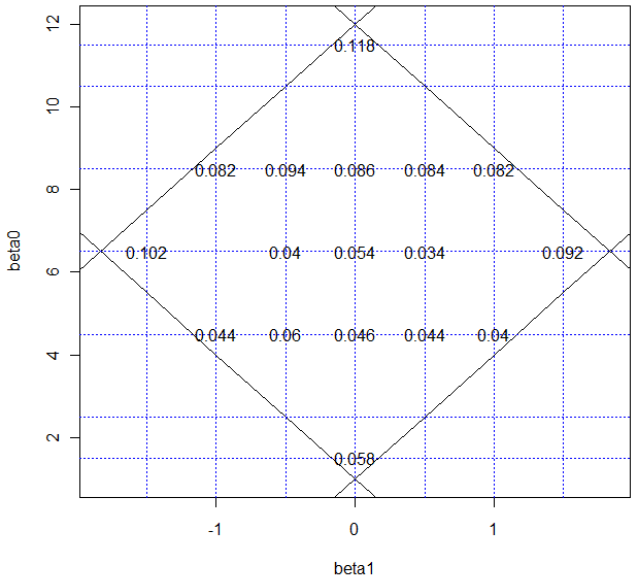

Figure 4.21: *LRT Size Values for Poisson Model with n=100, B=500 and 529 Grid Points*

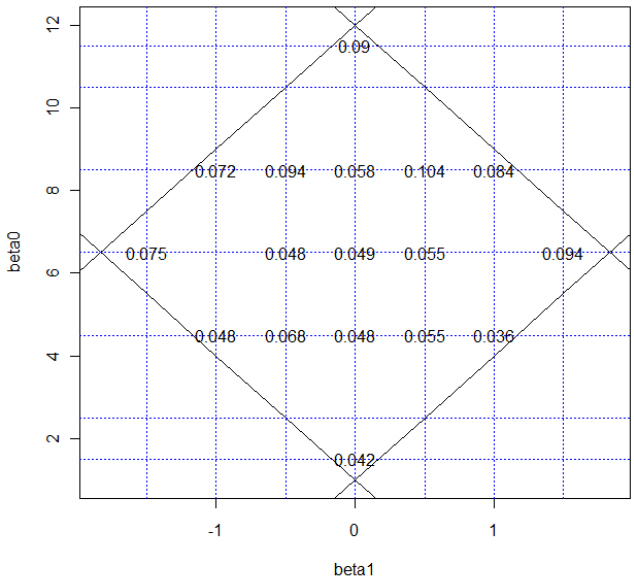

Figure 4.22: *LRT Size Values for Poisson Model with n=100, B=1000 and 529 Grid Points*

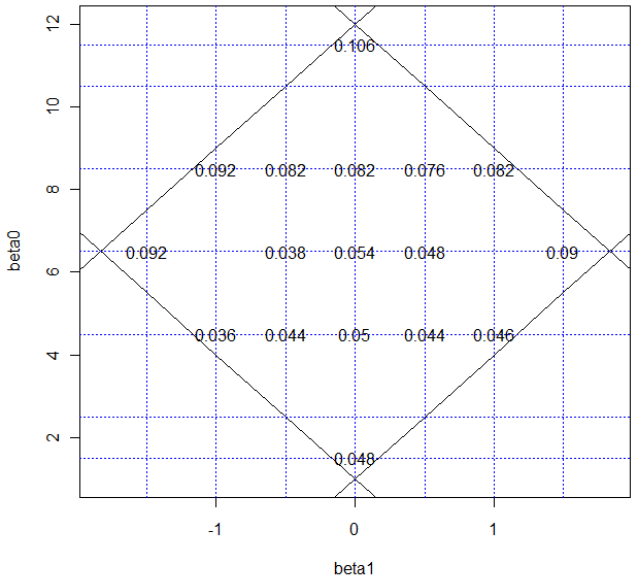

Figure 4.23: *LRT Size Values for Poisson Model with n=100, B=500 and 2025 Grid Points*

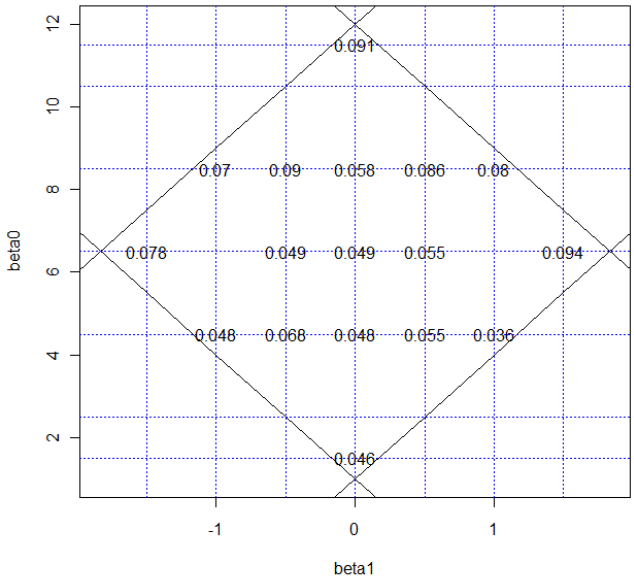

Figure 4.24: *LRT Size Values for Poisson Model with n=100, B=1000 and 2025 Grid Points*

used for illustration purposes. Alternative full models based on the four multiple maximin clusterings will be considered subsequently, with fuzzy clusterings employed to facilitate the grid search for the MLE. Fuzzy clusterings, as discussed in Munasinghe (2010), are determined from the four maximin clusterings corresponding to the four selected β settings. Further, based on the selected cover of the predictor space, the dimension of the fuzzy clusterings is equal to three.

Table 4.12 gives the crisp maximin clusterings $Z_n(\beta)$ for the two sizes of data sets at each of the four selected β settings. The results indicate that the crisp clusterings are not constant across the chosen β settings.

values of β	$Z_n(\beta)$	
	n=50	n=100
(4.5,-0.5)	$\{x_1 : x_{17}\}$ $\{x_{18} : x_{40}\}$ $\{x_{41} : x_{50}\}$	$\{x_1 : x_{37}\}$ $\{x_{38} : x_{75}\}$ $\{x_{76} : x_{100}\}$
(4.5,0.5)	$\{x_1 : x_{17}\}$ $\{x_{18} : x_{28}\}$ $\{x_{29} : x_{50}\}$	$\{x_1 : x_{37}\}$ $\{x_{38} : x_{80}\}$ $\{x_{81} : x_{100}\}$
(6.5,-0.5)	$\{x_1 : x_{17}\}$ $\{x_{18} : x_{40}\}$ $\{x_{41} : x_{50}\}$	$\{x_1 : x_{37}\}$ $\{x_{38} : x_{75}\}$ $\{x_{76} : x_{100}\}$
(6.5,0.5)	$\{x_1 : x_{17}\}$ $\{x_{18} : x_{28}\}$ $\{x_{29} : x_{50}\}$	$\{x_1 : x_{37}\}$ $\{x_{38} : x_{80}\}$ $\{x_{81} : x_{100}\}$

Table 4.12: *Maximin Clusterings for the Poisson Model with n=50 and n=100 for the Selected* (β_0, β_1) *Values*

4.2.1 Data Generation by Perturbing the Proposed Poisson Model

The data were simulated by using the perturbed Poisson model with $\eta_i = \beta_0 + \beta_1 z_i + s\gamma_i$ where (γ_i) is a basis vector for the between-cluster lack of fit subspace at selected values of $\beta = (\beta_0, \beta_1)$. Here the mean function of the distribution is $\mu_i = \exp(\eta_i)$. The grid is chosen as discussed above with $\delta_1 = 1, \delta_2 = 12$ and with the z values drawn from the Uniform distribution on $(-3, 3)$. For the LRT, 529 grid points were used for approximation of the MLE parameter estimates. For the $S(y, \hat{\mu})$ statistic, grids with 529 and 2025 points were used for parameter estimation under the null, as well as fitting the null model directly with the *glm* function in the *R* software package. The power of the LRT statistic, the deviance and the $S(y, \hat{\mu})$ statistic was investigated. Note $S(y, \hat{\mu})$ can be seen to be equal to the Pearson χ^2 goodness of fit test statistic as given in Dobson and Barnett (2008), for example. The χ^2 distribution was used to determine critical points, set according to the .05

nominal level.

The results were obtained by using $B = 500$ simulations, and are given in Table 4.13 for $n = 50$ and Table 4.14 for $n = 100$. Tables 4.13 and 4.14 give the power of the LRT for both the cases of a single maximin clustering as well as for fuzzy clusterings. It can be observed from the tables that the LRT tests have superior simulated power as compared to the deviance and S tests for the cases considered in this simulation study, especially for smaller values of the parameter s. The differences are more pronounced for larger n. Further, the power for the S test calculated with the R software lags the power obtain for S based on the grid search with either grid density.

4.2.2 Data Generation by a Functionally Different Model than the Constructed Full Model

A functionally different model given by $\eta_i = \beta_0 + \beta_1 z_i + \alpha z_i^2$ is next considered as the data generator in checking the power of the LRT statistic, the deviance and the $S(y, \hat{\mu})$ statistics. The crisp clustering at each point is used for the single clustering simulated power, while the four crisp clusterings are used in calculating the fuzzy clustering results for the LRT. The χ^2 distribution was used to determine critical points, set according to the .05 nominal level.

The results were obtained by using $B = 500$ simulations, and are given in Table 4.15 for $n = 50$ and Table 4.16 for $n = 100$. Tables 4.15 and 4.16 give the power of the LRT for both the cases of a single maximin clustering as well as for fuzzy clusterings. Notably, the S test outperformed the LRT and deviance tests. The LRT tests in turn have better simulated power than the deviance test, in particular for smaller values of the α parameter. The differences are more pronounced for larger n. Further, the power for the S test calculated with the R software lags the power obtain for S based on the grid search with either grid

(β_0,β_1)	s	with single maximin clusters	with multiple maximin clusters	deviance test	S statistic		
					529 g.p.	2025 g.p.	fit
(4.5,-0.5)	0	0.05	0.05	0.042	0.07	0.07	0.034
	1	0.17	0.17	0.068	0.094	0.094	0.054
	2	0.51	0.51	0.106	0.14	0.14	0.076
	3	0.884	0.884	0.278	0.304	0.304	0.212
	4	0.984	0.984	0.476	0.492	0.492	0.382
	5	1	1	0.738	0.754	0.754	0.654
	6	1	1	0.918	0.918	0.918	0.886
	7	1	1	0.986	0.988	0.988	0.974
	8	1	1	0.996	0.992	0.992	0.99
(4.5,0.5)	0	0.054	0.07	0.044	0.074	0.074	0.03
	1	0.126	0.22	0.05	0.082	0.082	0.044
	2	0.342	0.422	0.09	0.112	0.112	0.076
	3	0.658	0.628	0.158	0.174	0.174	0.132
	4	0.85	0.832	0.254	0.294	0.294	0.228
	5	0.98	0.926	0.416	0.476	0.476	0.388
	6	0.996	0.982	0.66	698	698	0.61
	7	1	0.99	0.822	0.862	0.862	0.808
	8	1	1	0.944	0.96	0.96	0.928
(6.5,-0.5)	0	0.048	0.048	0.048	0.074	0.074	0.044
	2	0.116	0.116	0.05	0.078	0.078	0.046
	4	0.312	0.312	0.082	0.116	0.116	0.066
	6	0.644	0.644	0.12	0.168	0.168	0.092
	8	0.874	0.874	0.234	0.29	0.29	0.198
	10	0.974	0.974	0.368	0.426	0.426	0.326
	12	0.998	0.998	0.584	0.638	0.638	0.528
	14	0.998	0.998	0.762	0.798	0.798	0.744
	16	1	1	0.91	0.92	0.92	0.886
(6.5,0.5)	0	0.058	0.066	0.038	0.054	0.054	0.032
	2	0.08	0.082	0.04	0.07	0.07	0.04
	4	0.196	0.144	0.06	0.086	0.086	0.044
	6	0.404	0.238	0.1	0.118	0.118	0.086
	8	0.614	0.394	0.14	0.19	0.19	0.118
	10	0.81	0.558	0.2	0.258	0.258	0.17
	12	0.924	0.696	0.3	0.354	0.354	0.258
	14	0.989	0.868	0.464	0.516	0.516	0.414
	16	1	0.914	0.588	0.654	0.654	0.55

Table 4.13: *Power of the LRT, Deviance and S Tests for Poisson Model with* $n = 50$ *at Selected* (β_0, β_1, s) *Values*

(β_0,β_1)	s	with single maximin clusters	with multiple maximin clusters	deviance test	S statistic		
					529 g.p.	2025 g.p.	fit
(4.5,-0.5)	0	0.048	0.048	0.058	0.07	0.07	0.054
	1	0.32	0.32	0.076	0.09	0.09	0.056
	2	0.812	0.812	0.166	0.174	0.174	0.136
	3	0.994	0.994	0.354	0.378	0.378	0.298
	4	1	1	0.696	0.696	0.696	0.626
	5	1	1	0.936	0.94	0.94	0.912
	6	1	1	0.996	0.996	0.996	0.994
	7	1	1	1	1	1	1
	8	1	1	1	1	1	1
(4.5,0.5)	0	0.058	0.046	0.046	0.068	0.068	0.032
	1	0.142	0.122	0.05	0.07	0.07	0.046
	2	0.428	0.374	0.084	0.098	0.098	0.066
	3	0.742	0.66	0.144	0.148	0.148	0.118
	4	0.928	0.906	0.242	0.252	0.252	0.198
	5	0.99	0.97	0.352	0.366	0.366	0.308
	6	1	1	0.58	0.582	0.582	0.496
	7	1	1	0.762	0.772	0.772	0.712
	8	1	1	0.884	0.894	0.894	0.856
(6.5,-0.5)	0	0.056	0.056	0.056	0.068	0.068	0.054
	2	0.186	0.186	0.08	0.098	0.098	0.072
	4	0.566	0.566	0.104	0.12	0.12	0.088
	6	0.904	0.904	0.198	0.23	0.23	0.174
	8	0.992	0.992	0.382	0.418	0.418	0.342
	10	0.998	0.998	0.592	0.61	0.61	0.556
	12	1	1	0.828	0.842	0.842	0.802
	14	1	1	0.94	0.948	0.948	0.934
	16	1	1	0.992	0.992	0.992	0.982
(6.5,0.5)	0	0.062	0.056	0.048	0.068	0.068	0.04
	2	0.116	0.106	0.52	0.068	0.07	0.044
	4	0.262	0.252	0.06	0.072	0.098	0.044
	6	0.51	0.46	0.086	0.112	0.148	0.068
	8	0.758	0.67	0.126	0.152	0.252	0.106
	10	0.908	0.86	0.186	0.206	0.366	0.156
	12	0.974	0.952	0.278	0.306	0.582	0.25
	14	0.988	0.992	0.42	0.45	0.772	0.376
	16	1	1	0.574	0.612	0.894	0.534

Table 4.14: *Power of the LRT, Deviance and S tests for Poisson Model with* $n = 100$ *at Selected* (β_0, β_1, s) *Values*

density, especially for smaller values of the α parameter.

4.3 Testing for Equal Means in a Heteroscedastic Normal Model

This subsection presents the results of a simulation study for the LRT and S_1 tests, as discussed in section 3.3.2. Specifically, power for these statistics is investigated for detecting unequal means under various data generators based on heteroscedastic Normal data.

First consider the data generator

$$x_i = m + \alpha k_i + r k_i Z$$

where $Z \sim N(0,1)$.

Then

$$x_i = m + (\alpha + rZ)k_i = \alpha \left[\frac{m}{\alpha} + \left(1 + r\frac{Z}{\alpha}\right) k_i \right]$$

so that for large α, $x_i \approx \alpha k_i$. Thus,

$$\begin{aligned} \bar{x_2} &\approx \sum_i \left(\frac{1}{k_i^2}\right) \alpha^2 k_i^2 = n\alpha^2 \\ \bar{x_1} &\approx \sum_i \left(\frac{1}{k_i^2}\right) \alpha k_i = \alpha \sum_i \left(\frac{1}{k_i}\right) \\ \tilde{x} &\approx \sum_i \gamma_i \alpha k_i = \alpha \sum_i \gamma_i k_i, \end{aligned}$$

and

(β_0,β_1)	α	with single maximin clusters	with multiple maximin clusters	Deviance test	S statistic		
					529 g.p.	2025 g.p.	fit
(4.5,-0.5)	0	0.05	0.05	0.042	0.07	0.07	0.034
	0.01	0.136	0.136	0.082	0.296	0.294	0.074
	0.02	0.384	0.384	0.23	0.878	0.874	0.21
	0.03	0.778	0.778	0.634	1	0.986	0.634
	0.04	0.944	0.944	0.944	1	0.992	0.94
	0.05	0.998	0.998	0.996	1	0.998	0.996
	0.06	1	1	1	1	1	1
(4.5,0.5)	0	0.054	0.036	0.044	0.074	0.074	0.03
	0.01	0.12	0.29	0.086	0.402	0.402	0.07
	0.02	0.408	0.806	0.258	0.98	0.976	0.266
	0.03	0.732	0.986	0.662	1	0.978	0.652
	0.04	0.95	1	0.934	1	0.986	0.942
	0.05	1	1	0.998	1	1	0.998
	0.06	1	1	1	1	1	1
(6.5,-0.5)	0	0.048	0.048	0.048	0.074	0.074	0.044
	0.005	0.226	0.226	0.112	0.432	0.432	0.098
	0.01	0.682	0.682	0.44	0.988	0.988	0.406
	0.015	0.964	0.964	0.896	1	1	0.894
	0.02	0.998	0.998	1	1	1	1
	0.025	1	1	1	1	1	1
	0.03	1	1	1	1	1	1
(6.5,0.5)	0	0.058	0.066	0.038	0.054	0.054	0.032
	0.005	0.216	0.408	0.098	0.668	0.668	0.076
	0.01	0.6	0.956	0.494	1	1	0.466
	0.015	0.916	1	0.908	1	1	0.9
	0.02	0.988	1	1	1	1	1
	0.025	1	1	1	1	1	1
	0.03	1	1	1	1	1	1

Table 4.15: *Power of the LRT, Deviance and S Tests for Poisson model with* $n = 50$ *at Selected* $(\beta_0, \beta_1, \alpha)$ *Values*

(β_0,β_1)	α	with single maximin clusters	with multiple maximin clusters	Deviance test	S statistic		
					529 g.p.	2025 g.p.	fit
(4.5,-0.5)	0	0.048	0.048	0.058	0.074	0.074	0.054
	0.01	0.238	0.238	0.126	0.494	0.494	0.126
	0.02	0.692	0.692	0.432	1	1	0.456
	0.03	0.97	0.97	0.936	1	1	0.94
	0.04	1	1	1	1	1	1
	0.05	1	1	1	1	1	1
	0.06	1	1	1	1	1	1
(4.5,0.5)	0	0.058	0.046	0.046	0.068	0.068	0.032
	0.01	0.584	0.496	0.11	0.734	0.734	0.102
	0.02	0.998	0.982	0.546	1	1	0.56
	0.03	1	1	0.944	1	1	0.956
	0.04	1	1	1	1	1	1
	0.05	1	1	1	1	1	1
	0.06	1	1	1	1	1	1
(6.5,-0.5)	0	0.056	0.056	0.056	0.068	0.068	0.054
	0.005	0.37	0.37	0.176	0.802	0.802	0.158
	0.01	0.898	0.898	0.834	1	1	0.826
	0.015	0.998	0.998	1	1	1	1
	0.02	1	1	1	1	1	1
	0.025	1	1	1	1	1	1
	0.03	1	1	1	1	1	1
(6.5,0.5)	0	0.062	0.056	0.048	0.068	0.068	0.04
	0.005	0.778	0.716	0.168	0.972	0.972	0.156
	0.01	1	1	0.846	1	1	0.842
	0.015	1	1	1	1	1	1
	0.02	1	1	1	1	1	1
	0.025	1	1	1	1	1	1
	0.03	1	1	1	1	1	1

Table 4.16: *Power of the LRT, Deviance and S Tests for Poisson model with* $n = 100$ *at Selected* $(\beta_0, \beta_1, \alpha)$ *Values*

$$(-2t_2)_{full} \approx n\left[n\alpha^2 - \alpha^2\frac{\left(\sum_i (1/k_i)\right)^2}{\sum_i (1/k_i)^2} - \alpha^2\frac{\left(\sum_i \gamma_i k_i\right)^2}{\sum_i \gamma_i^2 k_i^2}\right]^{-1}$$

$$S_1 \approx \frac{n}{\alpha^2}\left[\frac{1}{n - \frac{\left(\sum_i (1/k_i)\right)^2}{\sum_i (1/k_i)^2} - \frac{\left(\sum_i \gamma_i k_i\right)^2}{\sum_i \gamma_i^2 k_i^2}}\right]\left[\frac{\alpha^2\left(\sum_i \gamma_i k_i\right)^2}{\sum_i \gamma_i^2 k_i^2}\right].$$

Hence,

$$\lim_{\alpha\to\pm\infty} S_1 = \left[\frac{n}{n - \frac{\left(\sum_i (1/k_i)\right)^2}{\sum_i (1/k_i)^2} - \frac{\left(\sum_i \gamma_i k_i\right)^2}{\sum_i \gamma_i^2 k_i^2}}\right]\frac{\left(\sum_i \gamma_i k_i\right)^2}{\sum_i \gamma_i^2 k_i^2},$$

so that the clustering (i.e. γ) should be determined by maximizing $\frac{\left(\sum_i \gamma_i k_i\right)^2}{\sum_i \gamma_i^2 k_i^2}$.

Power for both LRT and S_1 statistics was investigated for different values of α by generating the k_i values from Uniform on $(1, 2)$ and Uniform on $(1, 5)$. Table 4.17 shows the simulated power for different values of l (cf section 3.3.2 which discusses the clustering structure) with $m = 1, r = 1$ for the case when the k_i values were drawn from the Uniform on $(1, 2)$ for the LRT statistic. Table 4.18 shows the simulated power for the S_1 statistic, along with the limit of S_1 in the last row.

Figure 4.25 illustrates the values of the S_1 limit with changing l.

Note that the limit of S_1 is maximized at $l = 17$. Thus, the power was investigated with changing parameter values of m and r for different values of α at $l = 17$. The power did not vary with changing m values so the power was calculated for different values of r and the results are given in Table 4.19 for the LRT test and Table 4.20 for the S_1 test.

The same data generator was used with the k_i drawn from Uniform on $(1, 5)$, and the simulated power values for changing l are given in Table 4.21 for the LRT and Table 4.22 for the S_1 statistic.

α	Power when m=1,r=1						
	l=10	l=15	l=20	l=25	l=30	l=35	l=40
0	0.064	0.06	0.059	0.058	0.048	0.053	0.05
1	0.184	0.207	0.2	0.175	0.158	0.149	0.118
2	0.523	0.602	0.582	0.523	0.437	0.362	0.27
3	0.843	0.901	0.889	0.825	0.766	0.636	0.483
4	0.979	0.99	0.989	0.973	0.937	0.839	0.669
5	0.999	1	1	0.998	0.985	0.95	0.829
6	1	1	1	1	0.999	0.985	0.931
7	1	1	1	1	1	0.998	0.979
8	1	1	1	1	1	1	0.991

Table 4.17: *Power of the LRT for Different Values of α with Changing l values in the Model $x_i = m + \alpha k_i + r k_i Z$ where $k_i \sim Uniform(1, 2)$*

α	Power when m=1,r=1						
	l=10	l=15	l=20	l=25	l=30	l=35	l=40
0	0.067	0.063	0.062	0.062	0.057	0.06	0.052
1	0.194	0.22	0.209	0.192	0.167	0.157	0.126
2	0.537	0.612	0.595	0.533	0.455	0.375	0.285
3	0.852	0.905	0.895	0.838	0.779	0.647	0.492
4	0.98	0.99	0.99	0.977	0.94	0.851	0.689
5	0.999	1	1	0.999	0.987	0.952	0.834
6	1	1	1	1	0.999	0.988	0.94
7	1	1	1	1	1	0.999	0.982
8	1	1	1	1	1	1	0.993
S_1 limit	101.0714	165.7049	146.6116	90.1098	58.4003	35.9093	22.102

Table 4.18: *Power of the S_1 Statistic for Different Values of α with Changing l Values in the model $x_i = m + \alpha k_i + r k_i Z$ where $k_i \sim Uniform(1, 2)$*

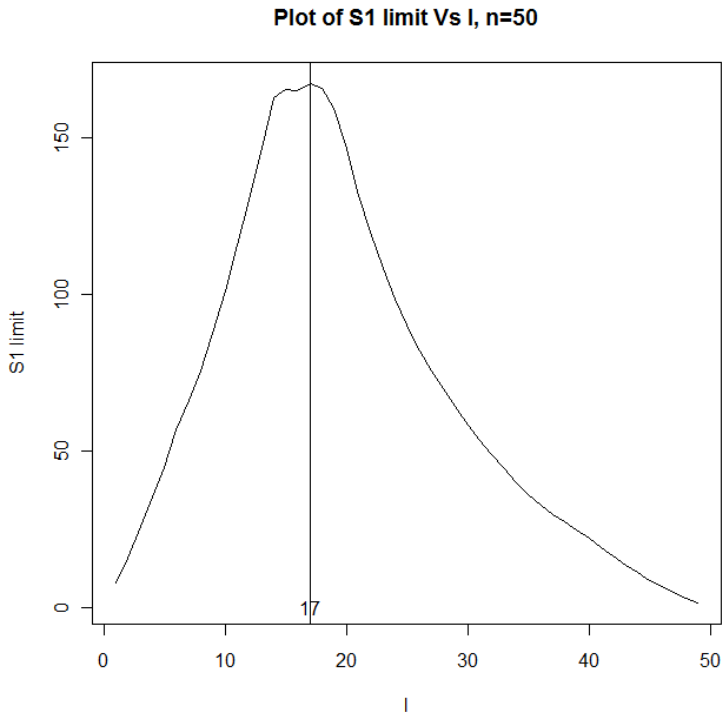

Figure 4.25: *Limit of S_1 for Changing l with $k_i \sim Uniform(1,2)$*

α	Power when l=17					
	r=1	r=2	r=3	r=-1	r=-2	r=-3
0	0.055	0.055	0.053	0.054	0.059	0.058
1	0.221	0.098	0.076	0.224	0.081	0.075
2	0.597	0.221	0.113	0.631	0.195	0.134
3	0.906	0.399	0.197	0.922	0.374	0.215
4	0.992	0.597	0.314	0.993	0.571	0.317
5	1	0.786	0.432	1	0.774	0.464
6	1	0.906	0.574	1	0.89	0.611
7	1	0.971	0.704	1	0.952	0.745
8	1	0.992	0.796	1	0.992	0.861

Table 4.19: *Power of the LRT for Different Values of α with Different (m,r) Combinations in the Model $x_i = m + \alpha k_i + r k_i Z$ where $k_i \sim Uniform(1,2)$*

α	Power when l=17,S_1 limit= 167.5127					
	r=1	r=2	r=3	r=-1	r=-2	r=-3
0	0.063	0.063	0.057	0.059	0.066	0.062
1	0.238	0.102	0.08	0.234	0.094	0.081
2	0.607	0.238	0.124	0.646	0.213	0.139
3	0.909	0.409	0.211	0.924	0.39	0.23
4	0.993	0.607	0.326	0.994	0.588	0.328
5	1	0.801	0.445	1	0.784	0.488
6	1	0.909	0.588	1	0.895	0.629
7	1	0.973	0.713	1	0.955	0.755
8	1	0.993	0.806	1	0.992	0.865

Table 4.20: *Power of the S_1 Test for Different Values of α with Different (m, r) Combinations in the Model $x_i = m + \alpha k_i + r k_i Z$ where $k_i \sim Uniform(1, 2)$*

α	Power when m=1,r=1						
	l=10	l=15	l=20	l=25	l=30	l=35	l=40
0	0.054	0.056	0.065	0.063	0.058	0.058	0.055
0.5	0.219	0.214	0.199	0.171	0.135	0.11	0.087
1	0.659	0.631	0.547	0.458	0.359	0.262	0.193
1.5	0.944	0.929	0.87	0.743	0.641	0.468	0.331
2	0.997	0.992	0.978	0.927	0.841	0.676	0.465
2.5	1	0.999	0.998	0.988	0.945	0.832	0.613
3	1	1	1	0.999	0.982	0.92	0.724
3.5	1	1	1	1	0.994	0.969	0.823
4	1	1	1	1	0.999	0.99	0.905

Table 4.21: *Power of the LRT for Different Values of α with Changing l Values in the Model $x_i = m + \alpha k_i + r k_i Z$ where $k_i \sim Uniform(1, 5)$*

α	Power when m=1,r=1						
	l=10	l=15	l=20	l=25	l=30	l=35	l=40
0	0.058	0.059	0.07	0.066	0.062	0.61	0.059
0.5	0.227	0.227	0.208	0.18	0.149	0.116	0.094
1	0.676	0.655	0.556	0.47	0.372	0.274	0.206
1.5	0.947	0.937	0.878	0.762	0.662	0.485	0.343
2	0.998	0.994	0.979	0.931	0.852	0.694	0.488
2.5	1	0.999	0.998	0.989	0.952	0.845	0.634
3	1	1	1	0.999	0.984	0.928	0.74
3.5	1	1	1	1	0.995	0.975	0.843
4	1	1	1	1	0.999	0.99	0.911
S1 limit	159.9188	137.667	82.3965	46.9392	29.8855	18.6205	11.4605

Table 4.22: *Power of the S_1 Test for Different Values of α with Changing l Values in the Model $x_i = m + \alpha k_i + rk_iZ$ where $k_i \sim Uniform(1, 5)$*

Figure 4.26 illustrates the values of the S_1 limit with changing l.

Note that the limit of S_1 is maximized at $l = 11$. Thus, the power was investigated with changing parameter values of m and r for different values of α at $l = 11$. The power did not vary with changing m values so the power was calculated for different values of r and the results are given in Table 4.23 for the LRT test and Table 4.24 for the S_1 test.

Next consider the data generator $x_i = m+\alpha k_i+\beta i+rk_iZ$ with $\beta = 0.001$. Again the power was investigated for different values of l and different parameter values of r for the LRT and S_1 statistics. The results are given in Tables 4.25 and 4.26 for changing l and Tables 4.27 and 4.28 for changing r with $l = 17$ when the k_i are drawn from Uniform on $(1, 2)$, and Tables 4.29, 4.30, 4.31, 4.32 for k_i Uniform on $(1, 5)$. Here Tables 4.29 and 4.30 show power values with changing l and Tables 4.31 and 4.32 show power values with changing r when l=11, where the limit of S_1 is maximized.

The data generator $x_i = m+\alpha k_i^2+rk_i^2Z$ was also used to investigate the power for different

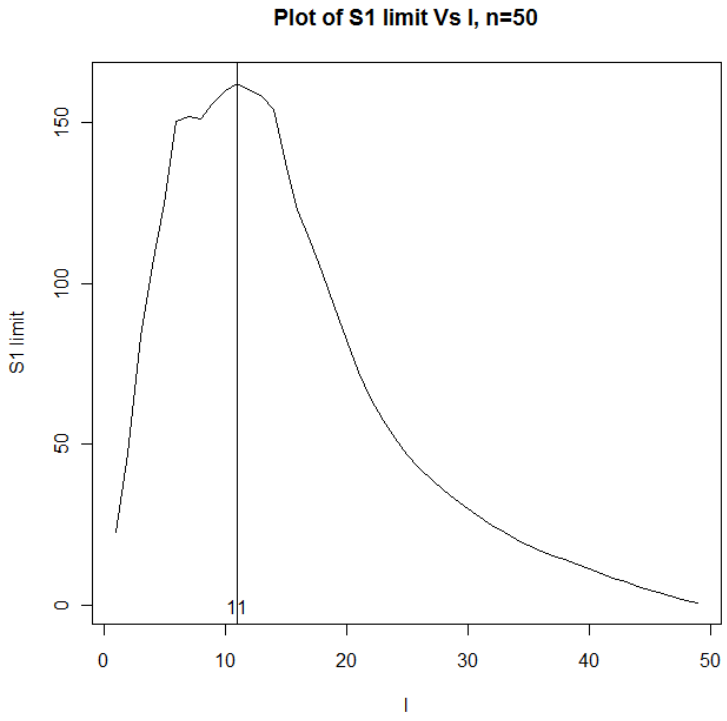

Figure 4.26: *Limit of S_1 for Changing l with $k_i \sim Uniform(1,5)$*

α	Power when l=11					
	r=1	r=2	r=3	r=-1	r=-2	r=-3
0	0.057	0.049	0.057	0.054	0.049	0.06
0.5	0.224	0.088	0.089	0.223	0.108	0.072
1	0.658	0.223	0.143	0.657	0.247	0.128
1.5	0.929	0.426	0.225	0.935	0.445	0.229
2	0.996	0.654	0.364	0.995	0.658	0.37
2.5	0.999	0.832	0.504	1	0.83	0.5
3	1	0.935	0.662	1	0.931	0.625
3.5	1	0.98	0.781	1	0.983	0.783
4	1	0.997	0.883	1	0.995	0.878

Table 4.23: *Power of the LRT for Different Values of α with Different (m,r) Combinations in the Model $x_i = m + \alpha k_i + r k_i Z$ where $k_i \sim Uniform(1,5)$*

α	Power when l=11,S_1 limit=162.155					
	r=1	r=2	r=3	r=-1	r=-2	r=-3
0	0.062	0.049	0.064	0.06	0.053	0.063
0.5	0.235	0.093	0.094	0.227	0.115	0.075
1	0.669	0.232	0.154	0.673	0.261	0.141
1.5	0.944	0.44	0.235	0.938	0.458	0.244
2	0.996	0.666	0.377	0.995	0.663	0.386
2.5	1	0.84	0.524	1	0.839	0.509
3	1	0.938	0.679	1	0.935	0.64
3.5	1	0.983	0.793	1	0.985	0.793
4	1	0.997	0.89	1	0.995	0.887

Table 4.24: *Power of the S_1 statistic for Different Values of α with Different (m, r) Combinations in the Model $x_i = m + \alpha k_i + rk_iZ$ where $k_i \sim Uniform(1, 5)$*

α	Power when m=1,r=1						
	l=10	l=15	l=20	l=25	l=30	l=35	l=40
0	0.067	0.056	0.062	0.063	0.055	0.062	0.062
0.75	0.157	0.165	0.148	0.148	0.129	0.115	0.1
1.5	0.388	0.396	0.377	0.343	0.301	0.242	0.194
2.25	0.66	0.708	0.693	0.617	0.536	0.467	0.332
3	0.86	0.903	0.893	0.848	0.742	0.64	0.507
3.75	0.965	0.981	0.976	0.957	0.899	0.806	0.657
4.5	0.994	0.997	0.995	0.991	0.971	0.907	0.775
5.25	0.999	1	1	0.999	0.996	0.962	0.872
6	1	1	1	1	0.999	0.992	0.93

Table 4.25: *Power of the LRT for Different Values of α with Changing l Values in the Model $x_i = m + \alpha k_i + \beta i + rk_iZ$ where $k_i \sim Uniform(1, 2)$*

α	Power when m=1,r=1						
	l=10	l=15	l=20	l=25	l=30	l=35	l=40
0	0.07	0.063	0.068	0.066	0.057	0.065	0.066
0.75	0.165	0.174	0.155	0.156	0.136	0.117	0.109
1.5	0.404	0.41	0.392	0.357	0.315	0.254	0.205
2.25	0.674	0.72	0.705	0.623	0.549	0.481	0.348
3	0.87	0.914	0.902	0.856	0.753	0.658	0.523
3.75	0.969	0.982	0.977	0.959	0.901	0.818	0.672
4.5	0.996	0.997	0.996	0.991	0.973	0.914	0.793
5.25	0.999	1	1	0.999	0.997	0.965	0.88
6	1	1	1	1	0.999	0.995	0.935
S1 limit	101.0174	165.7049	146.6116	90.1098	58.4003	35.9093	22.1027

Table 4.26: *Power of the S_1 Statistic for Different Values of α with Changing l Values in the Model $x_i = m + \alpha k_i + \beta i + r k_i Z$ where $k_i \sim Uniform(1, 2)$*

α	Power when l=17					
	r=1	r=2	r=3	r=-1	r=-2	r=-3
0	0.052	0.054	0.059	0.063	0.054	0.06
0.75	0.158	0.086	0.073	0.145	0.081	0.066
1.5	0.417	0.137	0.094	0.417	0.138	0.104
2.25	0.718	0.245	0.147	0.719	0.246	0.134
3	0.898	0.39	0.198	0.913	0.383	0.215
3.75	0.981	0.556	0.283	0.981	0.552	0.302
4.5	0.999	0.688	0.38	0.997	0.688	0.389
5.25	1	0.822	0.48	1	0.804	0.486
6	1	0.903	0.594	1	0.907	0.598

Table 4.27: *Power of the LRT for Different Values of α with Different (m, r) Combinations in the Model $x_i = m + \alpha k_i + \beta i + r k_i Z$ where $k_i \sim Uniform(1, 2)$*

α	Power when l=17,S_1 limit=167.5127					
	r=1	r=2	r=3	r=-1	r=-2	r=-3
0	0.057	0.063	0.066	0.067	0.061	0.063
0.75	0.163	0.088	0.078	0.154	0.09	0.074
1.5	0.425	0.149	0.1	0.438	0.146	0.111
2.25	0.733	0.266	0.153	0.733	0.257	0.143
3	0.904	0.404	0.205	0.915	0.402	0.221
3.75	0.982	0.573	0.294	0.983	0.57	0.315
4.5	0.999	0.704	0.394	0.99	0.705	0.399
5.25	1	0.831	0.5	1	0.817	0.504
6	1	0.91	0.617	1	0.907	0.616

Table 4.28: *Power of the S_1 Statistic for Different Values of α with Different (m, r) Combinations in the Model $x_i = m + \alpha k_i + \beta i + r k_i Z$ where $k_i \sim Uniform(1, 2)$*

α	Power when m=1,r=1						
	l=10	l=15	l=20	l=25	l=30	l=35	l=40
0	0.057	0.053	0.063	0.046	0.052	0.053	0.056
0.5	0.231	0.229	0.18	0.17	0.158	0.121	0.109
1	0.673	0.658	0.599	0.474	0.379	0.303	0.201
1.5	0.933	0.936	0.894	0.803	0.669	0.499	0.346
2	1	0.999	0.981	0.952	0.862	0.691	0.501
2.5	1	1	0.998	0.992	0.96	0.853	0.649
3	1	1	1	0.999	0.993	0.938	0.762
3.5	1	1	1	1	0.999	0.979	0.844
4	1	1	1	1	1	0.994	0.898

Table 4.29: *Power of the LRT for Different Values of α with Changing l Values in the Model $x_i = m + \alpha k_i + \beta i + r k_i Z$ where $k_i \sim Uniform(1, 5)$*

α	Power when m=1,r=1						
	l=10	l=15	l=20	l=25	l=30	l=35	l=40
0	0.065	0.058	0.065	0.051	0.058	0.061	0.06
0.5	0.246	0.235	0.192	0.181	0.164	0.127	0.115
1	0.686	0.668	0.614	0.495	0.391	0.316	0.216
1.5	0.939	0.939	0.899	0.82	0.684	0.517	0.365
2	1	0.999	0.984	0.957	0.868	0.714	0.517
2.5	1	1	0.998	0.994	0.966	0.863	0.666
3	1	1	1	0.999	0.994	0.945	0.777
3.5	1	1	1	1	0.999	0.984	0.857
4	1	1	1	1	1	0.996	0.905
S1 limit	159.9188	137.667	82.3965	46.9392	29.8855	18.6205	11.4605

Table 4.30: *Power of the S_1 Statistic for Different Values of α with Changing l Values in the Model $x_i = m + \alpha k_i + \beta i + r k_i Z$ where $k_i \sim Uniform(1, 5)$*

α	Power when l=11					
	r=1	r=2	r=3	r=-1	r=-2	r=-3
0	0.063	0.042	0.053	0.053	0.064	0.06
0.5	0.225	0.09	0.075	0.245	0.093	0.091
1	0.677	0.226	0.138	0.668	0.254	0.153
1.5	0.939	0.424	0.224	0.94	0.466	0.24
2	0.998	0.643	0.368	0.997	0.679	0.385
2.5	1	0.829	0.517	1	0.841	0.523
3	1	0.95	0.664	1	0.938	0.677
3.5	1	0.985	0.786	1	0.985	0.791
4	1	0.997	0.88	1	0.997	0.879

Table 4.31: *Power of the LRT for Different Values of α with Different (m, r) Combinations in the Model $x_i = m + \alpha k_i + \beta i + r k_i Z$ where $k_i \sim Uniform(1, 5)$*

α	Power when l=11,S_1 limit=162.155					
	r=1	r=2	r=3	r=-1	r=-2	r=-3
0	0.067	0.044	0.058	0.055	0.071	0.066
0.5	0.239	0.094	0.08	0.258	0.1	0.096
1	0.695	0.238	0.145	0.682	0.265	0.163
1.5	0.942	0.44	0.24	0.945	0.477	0.25
2	0.998	0.656	0.39	0.999	0.687	0.398
2.5	1	0.839	0.536	1	0.847	0.539
3	1	0.954	0.679	1	0.943	0.692
3.5	1	0.985	0.797	1	0.985	0.802
4	1	0.997	0.883	1	0.999	0.888

Table 4.32: *Power of the S_1 Statistic for Different Values of α with Different (m, r) Combinations in the Model $x_i = m + \alpha k_i + \beta i + r k_i Z$ where $k_i \sim Uniform(1, 5)$*

values of l and different parameter values of r for the LRT and S_1 statistics. The results are given in Tables 4.33, 4.34, 4.35, 4.36 with the k_i drawn from Uniform on $(1, 2)$, and Tables 4.37, 4.38, 4.39, 4.40 for k_i Uniform on $(1, 5)$. Tables 4.33 and 4.34 show power values with changing l, and Tables 4.35 and 4.36 are for changing values of r when the k_i are drawn from Uniform on $(1, 2)$. Tables 4.37 and 4.38 show power with different l values and Tables 4.39 and 4.40 are for different r values when the k_i are from Uniform on $(1, 5)$. The power for changing values of r was investigated with l that maximized the limit of S_1.

The conclusions are similar for all three data generators. When the power was investigated with changing l, the simulated power reached the value one more quickly with the l values that maximized the S_1 limit. This is true for both test statistics LRT and S_1. With changing r values, with l fixed at the point that the S_1 limit was maximized, the simulated power reached one more quickly with smaller values of r.

α	Power when m=1,r=1						
	l=10	l=15	l=20	l=25	l=30	l=35	l=40
0	0.058	0.052	0.06	0.058	0.061	0.055	0.058
0.5	0.308	0.368	0.361	0.328	0.303	0.249	0.202
1	0.809	0.893	0.886	0.839	0.778	0.673	0.539
1.5	0.985	0.997	0.995	0.989	0.975	0.94	0.845
2	1	1	1	1	0.998	0.994	0.968
2.5	1	1	1	1	0.999	1	0.995
3	1	1	1	1	1	1	0.999
3.5	1	1	1	1	1	1	1
4	1	1	1	1	1	1	1

Table 4.33: *Power of the LRT for Different Values of α with Changing l Values in the Model $x_i = m + \alpha k_i^2 + r k_i^2 Z$ where $k_i \sim Uniform(1, 2)$*

α	Power when m=1,r=1						
	l=10	l=15	l=20	l=25	l=30	l=35	l=40
0	0.06	0.053	0.068	0.069	0.066	0.063	0.062
0.5	0.325	0.381	0.373	0.348	0.32	0.256	0.21
1	0.821	0.897	0.89	0.849	0.789	0.689	0.556
1.5	0.987	0.997	0.996	0.99	0.977	0.941	0.857
2	1	1	1	1	0.999	0.994	0.971
2.5	1	1	1	1	0.999	1	0.996
3	1	1	1	1	1	1	0.999
3.5	1	1	1	1	1	1	1
4	1	1	1	1	1	1	1
S1 limit	101.0174	165.7049	146.6116	90.1098	58.4003	35.9093	22.102

Table 4.34: *Power of the S_1 Statistic for Different Values of α with Changing l Values in the Model $x_i = m + \alpha k_i^2 + r k_i^2 Z$ where $k_i \sim Uniform(1, 2)$*

α	Power when l=17					
	r=1	r=2	r=3	r=-1	r=-2	r=-3
0	0.058	0.061	0.061	0.065	0.059	0.048
0.5	0.357	0.123	0.084	0.383	0.135	0.084
1	0.88	0.346	0.221	0.887	0.359	0.17
1.5	0.995	0.657	0.398	0.996	0.672	0.344
2	1	0.899	0.594	1	0.871	0.552
2.5	1	0.978	0.782	1	0.974	0.734
3	1	0.998	0.906	1	0.996	0.864
3.5	1	1	0.959	1	1	0.959
4	1	1	0.993	1	1	0.988

Table 4.35: *Power of the LRT for Different Values of* α *with Different* (m, r) *Combinations in the Model* $x_i = m + \alpha k_i^2 + r k_i^2 Z$ *where* $k_i \sim Uniform(1, 2)$

α	Power when l=17,S_1 limit= 167.5127					
	r=1	r=2	r=3	r=-1	r=-2	r=-3
0	0.062	0.065	0.068	0.068	0.065	0.055
0.5	0.369	0.128	0.091	0.4	0.142	0.087
1	0.889	0.362	0.227	0.896	0.371	0.186
1.5	0.997	0.678	0.412	0.996	0.682	0.357
2	1	0.906	0.613	1	0.881	0.565
2.5	1	0.98	0.793	1	0.975	0.746
3	1	0.998	0.914	1	0.997	0.873
3.5	1	1	0.959	1	1	0.961
4	1	1	0.994	1	1	0.988

Table 4.36: *Power of the* S_1 *Statistic for Different Values of* α *with Different* (m, r) *Combinations in the Model* $x_i = m + \alpha k_i^2 + r k_i^2 Z$ *where* $k_i \sim Uniform(1, 2)$

α	Power when m=1,r=1						
	l=10	l=15	l=20	l=25	l=30	l=35	l=40
0	0.055	0.049	0.049	0.052	0.057	0.068	0.058
0.1	0.293	0.308	0.295	0.245	0.206	0.183	0.145
0.2	0.738	0.768	0.737	0.652	0.592	0.454	0.36
0.3	0.967	0.972	0.957	0.911	0.863	0.756	0.625
0.4	0.997	1	0.999	0.994	0.977	0.925	0.826
0.5	1	1	1	1	0.999	0.983	0.93
0.6	1	1	1	1	1	0.994	0.983
0.7	1	1	1	1	1	1	0.995
0.8	1	1	1	1	1	1	0.998

Table 4.37: *Power of the LRT for Different Values of α with Changing l Values in the Model $x_i = m + \alpha k_i^2 + r k_i^2 Z$ where $k_i \sim Uniform(1, 5)$*

α	Power when m=1,r=1						
	l=10	l=15	l=20	l=25	l=30	l=35	l=40
0	0.058	0.053	0.051	0.06	0.065	0.068	0.064
0.1	0.309	0.325	0.31	0.261	0.222	0.198	0.155
0.2	0.744	0.781	0.746	0.67	0.611	0.47	0.37
0.3	0.969	0.975	0.958	0.918	0.871	0.769	0.64
0.4	0.997	1	0.999	0.994	0.977	0.937	0.837
0.5	1	1	1	1	0.999	0.986	0.936
0.6	1	1	1	1	1	0.996	0.989
0.7	1	1	1	1	1	1	0.997
0.8	1	1	1	1	1	1	0.999
S1 limit	159.9188	137.667	82.3965	46.9392	29.8855	18.6205	11.4605

Table 4.38: *Power of the S_1 Statistic for Different Values of α with Changing l Values in the Model $x_i = m + \alpha k_i^2 + r k_i^2 Z$ where $k_i \sim Uniform(1, 5)$*

α	Power when l=11					
	r=1	r=2	r=3	r=-1	r=-2	r=-3
0	0.052	0.053	0.048	0.055	0.056	0.055
0.1	0.294	0.116	0.071	0.25	0.1	0.066
0.2	0.753	0.286	0.152	0.72	0.261	0.144
0.3	0.968	0.513	0.278	0.97	0.516	0.25
0.4	0.998	0.745	0.434	0.999	0.747	0.416
0.5	1	0.896	0.594	1	0.89	0.574
0.6	1	0.969	0.758	1	0.969	0.72
0.7	1	0.996	0.879	1	0.993	0.844
0.8	1	1	0.94	1	1	0.937

Table 4.39: *Power of the LRT for Different Values of* α *with Different* (m, r) *Combinations in the Model* $x_i = m + \alpha k_i^2 + r k_i^2 Z$ *where* $k_i \sim Uniform(1, 5)$

α	Power when l=11,S_1 limit=162.155					
	r=1	r=2	r=3	r=-1	r=-2	r=-3
0	0.056	0.058	0.053	0.056	0.06	0.056
0.1	0.313	0.126	0.075	0.263	0.109	0.073
0.2	0.762	0.298	0.159	0.738	0.268	0.153
0.3	0.97	0.528	0.289	0.971	0.534	0.263
0.4	0.998	0.749	0.456	0.999	0.757	0.429
0.5	1	0.9	0.609	1	0.896	0.591
0.6	1	0.973	0.765	1	0.971	0.738
0.7	1	0.996	0.89	1	0.993	0.854
0.8	1	1	0.946	1	1	0.939

Table 4.40: *Power of the* S_1 *Statistic for Different Values of* α *with Different* (m, r) *Combinations in the Model* $x_i = m + \alpha k_i^2 + r k_i^2 Z$ *where* $k_i \sim Uniform(1, 5)$

Chapter 5

Asymptotics

This chapter discusses the asymptotic non-central χ^2 distribution property of the log LRT statistic for testing $H_0 : s = 0$ versus $H_a : s \neq 0$ in an exponential family regression model with

$$\eta_i = u(z_i, \beta) + h_i(\beta, s)$$

where $h_i(\beta, 0) = 0$ as given in section 3.2. In particular, the case with $h_i(\beta, s) = s\gamma_i(\beta)$ is considered. With $\beta \in \mathbb{R}^q$ and $s \in \mathbb{R}$, let $\theta = (\beta, s) \in \mathbb{R}^c$ where $c = q + 1$ so that $\eta_i = \eta_i(\theta)$. For asymptotic purposes, a parametric array of densities $(p_{ni,\theta}, 1 \leq i \leq n, n \geq 1)$ is considered where

$$p_{ni}(y, \theta) = g(y)e^{\eta_{ni}(\theta)y - \varphi(\eta_{ni}(\theta))}$$

so that

$$l_{ni}(y,\theta) = \eta_{ni}(\theta)y - \varphi(\eta_{ni}(\theta)) + log\ g(y).$$

The proof of the asymptotic distribution of the log LRT statistic is based in part on the requirement that the array is locally asymptotic normal (LAN). A key condition to ensure LAN is that the array of exponential family densities satisfy the following definition.

Definition:

A parametric array $(p_{ni,\theta}, 1 \leq i \leq n, n \geq 1)$ of densities is *uniformly differentiable in quadratic mean* (i.e. uniform q.m.d.) at θ if for each $1 \leq i \leq n$ and $n \geq 1$ there exists a measurable function $\ell'_{ni,\theta}$ such that

$$\int \left(\sqrt{p_{ni,\theta+\tilde{h}}} - \sqrt{p_{ni,\theta}} - \frac{1}{2}\tilde{h}^T \ell'_{ni,\theta} \sqrt{p_{ni,\theta}} \right)^2 d\mu = o\left(\left\| \tilde{h} \right\|^2 \right)$$

independent of n and i as $\tilde{h} \to 0$ in $\mathbb{R}^c$ where $p_{ni,\theta}$ denotes the density of y_{ni}.

Lemma 1 below gives mild conditions under which uniform q.m.d. obtains for the array of exponential family densities.

Lemma 1: Fix $\theta \in \mathbb{R}^c$ and suppose that the predictor settings z_{ni} are contained in a compact set in $\mathbb{R}^p$ for all $1 \leq i \leq n, n \geq 1$. Suppose $\frac{\partial \eta_{ni}}{\partial \theta_l}$ and $\frac{\partial^2 \eta_i}{\partial \theta_k \partial \theta_l}$ are uniformly bounded in n and i on a neighborhood of θ, and that $\varphi(\eta_{ni})$, along with the first two derivatives, are uniformly bounded as well. Then the array of exponential family densities for nonlinear regression models is uniform q.m.d. at θ.

Proof:

Note

$$p_{ni}(y,\theta) = e^{l_{ni}(y,\theta)}$$

$$p_{ni}(y,\theta)^{1/2} = e^{\frac{1}{2}l_{ni}(y,\theta)}$$

and let $r(t) = \sqrt{p_{ni}(y,\theta+th)} = e^{\frac{1}{2}l_{ni}(y,\theta+th)}$ for $t \in \mathbb{R},\ h \in \mathbb{R}^c$. Then

$$r'(t) = \frac{1}{2}r(t)\sum_{k=1}^{c}\frac{\partial l_{ni}}{\partial\theta_k}(y,\theta+th)h_k$$

so that

$$r''(t) = \frac{1}{2}r'(t)\left[\sum_{k=1}^{c}\frac{\partial l_{ni}}{\partial\theta_k}(y,\theta+th)h_k\right] + \frac{1}{2}r(t)\left[\sum_{k=1}^{c}\sum_{l=1}^{c}\frac{\partial^2 l_{ni}}{\partial\theta_l\partial\theta_k}(y,\theta+th)h_k h_l\right].$$

That is,

$$r''(t) = \frac{1}{4}r(t)\left[\sum_{l=1}^{c}\frac{\partial l_{ni}}{\partial\theta_l}(y,\theta+th)h_l\right]\left[\sum_{k=1}^{c}\frac{\partial l_{ni}}{\partial\theta_k}(y,\theta+th)h_k\right] + \frac{1}{2}r(t)\left[\sum_{k=1}^{c}\sum_{l=1}^{c}\frac{\partial^2 l_{ni}}{\partial\theta_l\partial\theta_k}(y,\theta+th)h_k h_l\right]$$

where

$$\begin{aligned}
\frac{\partial l_{ni}}{\partial \theta_k}(y, \theta + th) &= \frac{\partial \eta_{ni}}{\partial \theta_k}(\theta + th)y - \varphi'(\eta_{ni}(\theta + th))\frac{\partial \eta_{ni}}{\partial \theta_k}(\theta + th) \\
&= \frac{\partial \eta_{ni}}{\partial \theta_k}(\theta + th)\left[y - \varphi'(\eta_{ni}(\theta + th))\right] \\
\frac{\partial^2 l_{ni}}{\partial \theta_k \partial \theta_l}(y, \theta + th) &= \frac{\partial^2 \eta_{ni}}{\partial \theta_k \partial \theta_l}(\theta + th)y - \varphi''(\eta_{ni}(\theta + th))\frac{\partial \eta_{ni}}{\partial \theta_l}(\theta + th)\frac{\partial \eta_{ni}}{\partial \theta_k}(\theta + th) \\
&\qquad - \varphi'(\eta_{ni}(\theta + th))\frac{\partial^2 \eta_{ni}}{\partial \theta_k \partial \theta_l}(\theta + th) \\
&= \left[y - \varphi'(\eta_{ni}(\theta + th))\right]\frac{\partial^2 \eta_{ni}}{\partial \theta_k \partial \theta_l}(\theta + th) - \varphi''(\eta_{ni}(\theta + th))\frac{\partial \eta_{ni}}{\partial \theta_l}(\theta + th)\frac{\partial \eta_{ni}}{\partial \theta_k}(\theta + th)
\end{aligned}$$

Thus,

$$\begin{aligned}
r(0) &= \sqrt{p_{ni}(y, \theta)} \\
r'(0) &= \frac{1}{2}\sqrt{p_{ni}(y, \theta)}\sum_{k=1}^{c}\frac{\partial l_{ni}}{\partial \theta_k}(y, \theta)h_k.
\end{aligned}$$

By Taylor's theorem, $r(t) = r(0) + r'(0)t + \frac{1}{2}r''(\tau)t^2$ for some $0 \leq \tau \leq t$. Thus, setting $t = 1$ and rearranging terms gives

$$r(1) - r(0) - r'(0) = \frac{1}{2}r''(\tau)$$

for some $0 \leq \tau \leq 1$. That is,

$$\sqrt{p_{ni}(y,\theta+h)}-\sqrt{p_{ni}(y,\theta)}-\frac{1}{2}\sqrt{p_{ni}(y,\theta)}\sum_{k=1}^{c}\frac{\partial l_{ni}}{\partial\theta_k}(y,\theta)h_k$$
$$=\frac{1}{2}\left[\frac{1}{4}\sqrt{p_{ni}(y,\theta+\tau h)}\left(\sum_{l=1}^{c}\frac{\partial l_{ni}}{\partial\theta_l}(y,\theta+\tau h)h_l\right)\left(\sum_{k=1}^{c}\frac{\partial l_{ni}}{\partial\theta_k}(y,\theta+\tau h)h_k\right)\right.$$
$$\left.+\frac{1}{2}\sqrt{p_{ni}(y,\theta+\tau h)}\left(\sum_{k=1}^{c}\sum_{l=1}^{c}\frac{\partial^2 l_{ni}}{\partial\theta_l\partial\theta_k}(y,\theta+\tau h)h_kh_l\right)\right]$$

for some $0\leq\tau\leq 1$. This implies that

$$\left(\sqrt{p_{ni}(y,\theta+h)}-\sqrt{p_{ni}(y,\theta)}-\frac{1}{2}\sqrt{p_{ni}(y,\theta)}\sum_{k=1}^{c}\frac{\partial l_{ni}}{\partial\theta_k}(y,\theta)h_k\right)^2$$
$$=\frac{1}{16}p_{ni}(y,\theta+\tau h)\left[\frac{1}{2}\left(\sum_{l=1}^{c}\frac{\partial l_{ni}}{\partial\theta_l}(y,\theta+\tau h)h_l\right)\left(\sum_{k=1}^{c}\frac{\partial l_{ni}}{\partial\theta_k}(y,\theta+\tau h)h_k\right)\right.$$
$$\left.+\left(\sum_{k=1}^{c}\sum_{l=1}^{c}\frac{\partial^2 l_{ni}}{\partial\theta_l\partial\theta_k}(y,\theta+\tau h)h_kh_l\right)\right]^2$$

$$=\frac{1}{16}p_{ni}(y,\theta+\tau h)\left[\frac{1}{4}\left(\sum_{l=1}^{c}\frac{\partial l_{ni}}{\partial\theta_l}(y,\theta+\tau h)h_l\right)^2\left(\sum_{k=1}^{c}\frac{\partial l_{ni}}{\partial\theta_k}(y,\theta+\tau h)h_k\right)^2\right.$$
$$+\left(\sum_{l=1}^{c}\frac{\partial l_{ni}}{\partial\theta_l}(y,\theta+\tau h)h_l\right)\left(\sum_{k=1}^{c}\frac{\partial l_{ni}}{\partial\theta_k}(y,\theta+\tau h)h_k\right)\left(\sum_{k=1}^{c}\sum_{l=1}^{c}\frac{\partial^2 l_{ni}}{\partial\theta_l\partial\theta_k}(y,\theta+\tau h)h_kh_l\right)$$
$$\left.+\left(\sum_{k=1}^{c}\sum_{l=1}^{c}\frac{\partial^2 l_{ni}}{\partial\theta_l\partial\theta_k}(y,\theta+\tau h)h_kh_l\right)^2\right]$$

for some $0\leq\tau\leq 1$. Substituting for $\frac{\partial l_{ni}}{\partial\theta_l}(y,\theta+\tau h)$ and $\frac{\partial^2 l_{ni}}{\partial\theta_l\partial\theta_k}(y,\theta+\tau h)$ and writing $\eta_{ni}(\theta+\tau h)=\eta_{ni}$, the preceding equation becomes

$$\left(\sqrt{p_{ni}(y,\theta+h)}-\sqrt{p_{ni}(y,\theta)}-\frac{1}{2}\sqrt{p_{ni}(y,\theta)}\sum_{k=1}^{c}\frac{\partial l_{ni}}{\partial\theta_k}(y,\theta)h_k\right)^2$$

$$=\frac{1}{16}p_{ni}(y,\theta+\tau h)\left[\frac{1}{4}\left[y-\varphi'(\eta_{ni})\right]^4\left(\sum_{l=1}^{c}\frac{\partial\eta_{ni}}{\partial\theta_l}h_l\right)^2\left(\sum_{k=1}^{c}\frac{\partial\eta_{ni}}{\partial\theta_k}h_k\right)^2\right.$$

$$+\left[y-\varphi'(\eta_{ni})\right]^2\left(\sum_{l=1}^{c}\sum_{k=1}^{c}\frac{\partial^2\eta_{ni}}{\partial\theta_k\partial\theta_l}h_kh_l\right)^2+\varphi''^2(\eta_{ni})\left(\sum_{l=1}^{c}\sum_{k=1}^{c}\frac{\partial\eta_{ni}}{\partial\theta_l}\frac{\partial\eta_{ni}}{\partial\theta_k}h_kh_l\right)^2$$

$$-2\varphi''^2(\eta_{ni})\left[y-\varphi'(\eta_{ni})\right]\left(\sum_{l=1}^{c}\sum_{k=1}^{c}\frac{\partial^2\eta_{ni}}{\partial\theta_k\partial\theta_l}\frac{\partial\eta_{ni}}{\partial\theta_l}\frac{\partial\eta_{ni}}{\partial\theta_k}h_k^2h_l^2\right)$$

$$+\left[y-\varphi'(\eta_{ni})\right]^3\left(\sum_{l=1}^{c}\frac{\partial\eta_{ni}}{\partial\theta_l}h_l\right)\left(\sum_{k=1}^{c}\frac{\partial\eta_{ni}}{\partial\theta_k}h_k\right)\left(\sum_{l=1}^{c}\sum_{k=1}^{c}\frac{\partial^2\eta_{ni}}{\partial\theta_k\partial\theta_l}h_kh_l\right)$$

$$\left.-\left[y-\varphi'(\eta_{ni})\right]^2\varphi''(\eta_{ni})\left(\sum_{l=1}^{c}\frac{\partial\eta_{ni}}{\partial\theta_l}h_l\right)\left(\sum_{k=1}^{c}\frac{\partial\eta_{ni}}{\partial\theta_k}h_k\right)\left(\sum_{l=1}^{c}\sum_{k=1}^{c}\frac{\partial\eta_{ni}}{\partial\theta_l}\frac{\partial\eta_{ni}}{\partial\theta_k}h_kh_l\right)\right]$$

for some $0\leq\tau\leq 1$.

Thus, by taking moments with respect to $p_{ni}(y,\theta+\tau h)$ and using the Schwarz inequality,

$$\int\left(\sqrt{p_{ni}(y,\theta+h)}-\sqrt{p_{ni}(y,\theta)}-\frac{1}{2}\sqrt{p_{ni}(y,\theta)}\sum_{k=1}^{c}\frac{\partial l_{ni}}{\partial\theta_k}(y,\theta)h_k\right)^2\mu(dy)\quad\leq\quad B\left\|h\right\|^4$$

where B serves as a bound for all terms involving the $\frac{\partial\eta_{ni}}{\partial\theta_l}$, $\frac{\partial^2\eta_i}{\partial\theta_k\partial\theta_l}$ and $\varphi(\eta_{ni})$. $\diamond$

Note that under conditions of Lemma 1, $\ell'_{ni,\theta}=\frac{\partial log\ p_{ni,\theta}}{\partial\theta}$.

The following lemma gives conditions under which a uniform q.m.d. array of exponential family densities can be shown to be LAN.

Lemma 2: $P_{ni,\theta}g_{ni,\theta}^2\leq\left\|h\right\|^2C$ and $P_{ni,\theta}g_{ni,\theta}^4\leq\left\|h\right\|^4D$, for some constants C and D, and for all $n\geq 1$ and $1\leq i\leq n$ where $g_{ni,\theta}=h^T\ell'_{ni,\theta}$.

Proof:

Note that

$$\begin{aligned}
\ell'_{ni,\theta} &= \frac{\partial\ log(p_{ni,\theta})}{\partial\theta} = \frac{\partial}{\partial\theta}\left[y\eta_{ni} - \varphi(\eta_{ni})\right] \\
&= \frac{\partial}{\partial\theta}\left[y\eta_{ni}(\theta) - \varphi(\eta_{ni}(\theta))\right] \\
&= \left[y\frac{\partial\eta_{ni}}{\partial\theta_1} - \varphi'(\eta_{ni})\frac{\partial\eta_{ni}}{\partial\theta_1}, ..., y\frac{\partial\eta_{ni}}{\partial\theta_c} - \varphi'(\eta_{ni})\frac{\partial\eta_{ni}}{\partial\theta_c}\right]^T \\
&= \left[(y - \varphi'(\eta_{ni}))\left(\frac{\partial\eta_{ni}}{\partial\theta_1}\right), ..., (y - \varphi'(\eta_{ni}))\left(\frac{\partial\eta_{ni}}{\partial\theta_c}\right)\right]^T \\
&= (y - \varphi'(\eta_{ni}))\left[\frac{\partial\eta_{ni}}{\partial\theta_1}, ..., \frac{\partial\eta_{ni}}{\partial\theta_c}\right]^T
\end{aligned}$$

Then, using the bounds specified in Lemma 1 and the Schwarz inequality,

$$\begin{aligned}
P_{i,\theta}\ g^2_{ni,\theta} &= P_{ni,\theta}\sum_{j=1}^{c}\sum_{k=1}^{c} h_j h_k \left(\ell'_{ni,\theta}\left(\ell'_{ni,\theta}\right)^T\right)_{jk} \\
&= P_{ni,\theta}\sum_{j=1}^{c}\sum_{k=1}^{c} h_j h_k [y - \varphi'(\eta_{ni})]^2 \left(\frac{\partial\eta_{ni}}{\partial\theta_j}\frac{\partial\eta_{ni}}{\partial\theta_k}\right) \\
&\leq C\left\|h\right\|^2
\end{aligned}$$

for some constant C for all n and i. Similarly, using the bounds specified in Lemma 1 and the Schwarz inequality,

$$
\begin{aligned}
P_{ni,\theta} g_{ni,\theta}^4 &= P_{ni,\theta} \sum_{j=1}^{c} \sum_{k=1}^{c} h_j h_k \left(\ell'_{ni,\theta} \left(\ell'_{ni,\theta} \right)^T \right)_{jk} \sum_{l=1}^{c} \sum_{m=1}^{c} h_l h_m \left(\ell'_{ni,\theta} \left(\ell'_{ni,\theta} \right)^T \right)_{lm} \\
&= P_{ni,\theta} \sum_{j=1}^{c} \sum_{k=1}^{c} h_j h_k [y - \varphi'(\eta_{ni})]^2 \left(\frac{\partial \eta_{ni}}{\partial \theta_j} \frac{\partial \eta_{ni}}{\partial \theta_k} \right) \sum_{l=1}^{c} \sum_{m=1}^{c} h_l h_m [y - \varphi'(\eta_{ni})]^2 \left(\frac{\partial \eta_{ni}}{\partial \theta_l} \frac{\partial \eta_{ni}}{\partial \theta_m} \right) \\
&\leq D \left\| h \right\|^4
\end{aligned}
$$

for some constant D for all n and i. $\diamond$

It follows from Reider (1994), Theorem 2.3.9, that an array of exponential family densities satisfying the conditions of Lemmas 1 and 2 is LAN. In particular, it suffices to establish that the following Lindeberg and differentiability conditions hold for such a uniform q.m.d. array of exponential family densities. In addition, it is presently only assumed that $\sqrt{n} I_{n,\theta}^{-1/2}$ is bounded. Here, $I_{n,\theta} = \sum_{i=1}^{n} I_{ni,\theta}$ is a $c \times c$ matrix, assumed to be positive definite, and $I_{ni,\theta} = P_{ni,\theta} \ell'_{ni,\theta} (\ell'_{ni,\theta})^T$ denotes the Fisher information matrix with $dP_{ni,\theta} = p_{ni,\theta}$.

Condition 1 (Lindeberg):

For all $t \in \mathbb{R}^c$ and for all $\varepsilon \in (0, \infty)$,

$$
\lim_{n \to \infty} \sum_{i=1}^{n} \int_{\left\{ \left| t^T I_{n,\theta}^{-1/2} \ell'_{ni,\theta} \right| > \varepsilon \right\}} \left(t^T I_{n,\theta}^{-1/2} \ell'_{ni,\theta} \right)^2 dP_{ni,\theta} = 0.
$$

Proof:

Note for a fixed $\varepsilon > 0$ and $t \in \mathbb{R}^c$,

$$
\begin{aligned}
&\sum_{i=1}^{n} \int_{\left\{ \left| t^T I_{n,\theta}^{-1/2} \ell'_{ni,\theta} \right| > \varepsilon \right\}} \left(t^T I_{n,\theta}^{-1/2} \ell'_{ni,\theta} \right)^2 dP_{ni,\theta} \\
&= \sum_{i=1}^{n} \int_{\left\{ \left| t^T \sqrt{n} I_{n,\theta}^{-1/2} \ell'_{ni,\theta} \right| > \sqrt{n} \varepsilon \right\}} \left(t^T \frac{1}{\sqrt{n}} \sqrt{n} I_{n,\theta}^{-1/2} \ell'_{ni,\theta} \right)^2 dP_{ni,\theta}.
\end{aligned}
$$

Let $h = \sqrt{n} I_{n,\theta}^{-1/2} t$ and note that h is bounded since $\sqrt{n} I_{n,\theta}^{-1/2}$ is bounded and t is fixed. Then the above expression becomes

$$\frac{1}{n} \sum_{i=1}^{n} \int_{\left\{ \left| h^T \ell'_{ni,\theta} \right| > \sqrt{n}\varepsilon \right\}} \left(h^T \ell'_{ni,\theta} \right)^2 dP_{ni,\theta}.$$

By Lemma 2, $P_{ni,\theta} g_{ni,\theta}^4$ with $g_{ni,\theta} = h^T \ell'_{ni,\theta}$, are uniformly bounded in n and i. Thus, the collection of random variables $g_{ni,\theta}^2$ is uniformly integrable (cf Billinglsey (1995), p.338). Thus, for sufficiently large n,

$$\int_{\left\{ g_{ni,\theta}^2 > n\varepsilon^2 \right\}} g_{ni,\theta}^2 dP_{ni,\theta}$$

is arbitrarily small for all such n and $1 \leq i \leq n$. Therefore,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \int_{\left\{ g_{ni,\theta}^2 > n\varepsilon^2 \right\}} g_{ni,\theta}^2 dP_{ni,\theta} = 0$$

for all $\varepsilon \in (0, \infty)$. Thus, the Lindeberg condition holds. $\diamond$

Condition 2 (Differentiability):

For all $b \in (0, \infty)$,

$$\lim_{n\to\infty} \sup_{\|t\| \leq b} \sum_{i=1}^{n} \int \left(\sqrt{p_{ni,\theta_n(t)}} - \sqrt{p_{ni,\theta}} \left(1 + \frac{1}{2} t^T I_{n,\theta}^{-1/2} \ell'_{ni,\theta} \right) \right)^2 dP_{ni,\theta} = 0$$

Proof:

The local parameter alternatives $\theta_n(t)$ about θ are given by $\theta_n(t) = \theta + I_{n,\theta}^{-1/2} t$. Let $\tilde{h} = \frac{1}{\sqrt{n}} \left(\sqrt{n} I_{n,\theta}^{-1/2} t \right)$ so that $\theta_n(t) = \theta + \tilde{h}$. Note that $\tilde{h} \to 0$ uniformly with $\|t\|$ bounded. Note also that

$$\begin{aligned} &\sup_{\|t\| \leq b} \sum_{i=1}^{n} \int \left(\sqrt{p_{ni,\theta_n(t)}} - \sqrt{p_{ni,\theta}} \left(1 + \frac{1}{2} t^T I_{n,\theta}^{-1/2} \ell'_{ni,\theta} \right) \right)^2 dP_{ni,\theta} \\ &\quad = \sup_{\|t\| \leq b} \sum_{i=1}^{n} \int \left(\sqrt{p_{ni,\theta+\tilde{h}}} - \sqrt{p_{ni,\theta}} - \frac{1}{2} \tilde{h}^T \ell'_{ni,\theta} \sqrt{p_{ni,\theta}} \right)^2 dP_{ni,\theta} = \sup_{\|t\| \leq b} \sum_{i=1}^{n} o\left(\left\| \tilde{h} \right\|^2 \right). \end{aligned}$$

The last equality comes from the uniform q.m.d. array of exponential family densities at θ as provided by Lemma 1. Also, $\left\|\tilde{h}\right\|^2 = O\left(\frac{1}{n}\right)$ since $\sqrt{n}I_{n,\theta}^{-1/2}$ is bounded and t is fixed, and thus the differentiability condition holds. $\diamond$

Conditions 1 and 2 imply the L_2 differentiability of the parametric array $(p_{ni,\theta}, 1 \leq i \leq n, n \geq 1)$ at a fixed θ (cf Reider (1994), Definition 2.3.8), and thus the desired log-likelihood expansion indicated below.

In particular, it is now assumed that $\frac{1}{n}I_{n,\theta} \to I_\theta$ where I_θ is positive definite, and hence $\sqrt{n}I_{n,\theta}^{-\frac{1}{2}} \to I_\theta^{-\frac{1}{2}}$. Let $h_n = \sqrt{n}I_{n,\theta}^{-\frac{1}{2}}t_n$ where $t_n \to t \in \mathbb{R}^c$. Thus, $h_n \to h \in \mathbb{R}^c$ where $h = I_\theta^{-\frac{1}{2}}t$. Thus, $t = I_\theta^{\frac{1}{2}}h$ so that $\|t\|^2 = h^T I_\theta h$ and $t^T I_{n,\theta}^{-\frac{1}{2}} = h^T I_\theta^{\frac{1}{2}}\left(\sqrt{n}I_\theta^{-\frac{1}{2}}\right)\left(\frac{1}{\sqrt{n}}\right)$. The local parameter alternatives $\theta_n(t_n)$ about θ are given by $\theta_n(t_n) = \theta + I_{n,\theta}^{-\frac{1}{2}}t_n$, and with the preceding identification for h_n, $\theta_n(t_n) = \theta + \frac{h_n}{\sqrt{n}}$. Accordingly, a L_2 differentiable array of exponential family densities is LAN by Theorem 2.3.9 of Reider (1994). That is,

$$\begin{aligned} log\frac{dP_{n,\theta+h_n/\sqrt{n}}}{dP_{n,\theta}} &= t^T I_{n,\theta}^{-\frac{1}{2}} \sum_i \ell'_{ni,\theta}(y_{ni}) - \frac{1}{2}\|t\|^2 + o_{P_{n,\theta}}(1) \\ &= h^T I_\theta^{\frac{1}{2}} \sqrt{n} I_{n,\theta}^{-\frac{1}{2}} \Delta_{n,\theta} - \frac{1}{2} h^T I_\theta h + o_{P_{n,\theta}}(1) \end{aligned}$$

where $\Delta_{n,\theta} = \frac{1}{\sqrt{n}}\sum_i \ell'_{ni,\theta}(y_{ni})$ are random vectors such that $\Delta_{n,\theta} \Rightarrow N_c(0, I_\theta)$ as $n \to \infty$ under $P_{n,\theta}$.

Based on LAN of the array of exponential family densities for the nonlinear regression models considered, the asymptotic non-central χ^2 distribution property of the sequence of log LRT statistics for testing $H_0 : s = 0$ versus $H_a : s \neq 0$ is provided by Theorem 1 below. The proof of Theorem 1 below follows from the general proof of Theorem 1 in Munasinghe (2010). In addition, the required $\sqrt{n}$-consistency of the MLEs for the proof of Theorem 1 below is provided under regularity conditions, including the convergence of the average Fisher information, by Wei (1998) (cf Theorem 4.1).

Theorem 1

Suppose that the conditions that were indicated under Lemma 1 and Lemma 2 hold at $\theta = (\beta, 0)$ for an array of exponential family densities, and that the $\frac{1}{n} I_n(\beta, 0)$ converge to $I(\beta, 0)$, say. Additionally, suppose that the matrices $\ell''_{ni,\theta}$ of second order partial derivatives are locally bounded. If the unrestricted and restricted (by $H_0 : s = 0$) MLEs $\hat{\theta}_n$ and $\hat{\theta}_{n,0}$, respectively, are $\sqrt{n}$-consistent under $(\beta, 0)$, then the sequence of log LRT statistics Λ_n for testing $H_0 : s = 0$ versus $H_a : s \neq 0$ converges under $(\beta, s/\sqrt{n})$ in distribution to a random variable distributed according to the $\chi^2(1, \delta)$ distribution with noncentrality parameter $\delta = (0, s)^T I(\beta, 0)(0, s)$.

Chapter 6

Conclusions and Future Research

The geometry for maximum likelihood estimation in exponential family nonlinear regression models was reviewed, and a geometrically motivated goodness of fit test statistic for exponential family regression based on the information metric was introduced. General alternative models for assessing lack of fit associated with the mean function for a proposed model was presented. The construction of such general alternatives was based on clusterings in the mean components. A maximin power clustering methodology utilized by Munasinghe (2010) for nonlinear regression with additive error was defined in the context of exponential family nonlinear regression models to determine suitable clusterings for assessing lack of fit.

Curved exponential family models and generalizations were discussed to exemplify the general discussion, including tests for generalized linear models and equal means testing for heterogenous Normal models. It is worthwhile to note that the constructed alternatives used for comparison with a proposed generalized linear model are not restricted to be generalized linear models, as is the usual case for testing hypotheses associated with generalized linear models. In addition, a saturated three parameter exponential family model was also developed which allows for equal means testing with unequal variances. A LRT and a test determined by comparing MLE parameter estimates, each based on clusters in the alterna-

tive model, were developed to test the equal means hypothesis in a heteroscedastic Normal model.

A simulation study was carried out for the logistic and Poisson regression models as particular examples of the generalized linear model. An investigation of comparative performance of the LRT, the deviance test and the goodness of fit test based on the information metric was carried out. For logistic regression, the standard Hosmer-Lemeshow test was also included in the simulations. Notably, the LRT had comparable power with that of the Hosmer-Lemeshow test under both m- and n-asymptotics, with superior power for constructed alternatives. The simulated power for the goodness of fit test statistic based on the information metric was very competitive with the LRT in the case of m-asymptotics, while the deviance test did not have comparable power. In the case of Poisson regression, the LRT was compared with the deviance and the goodness of fit test statistic based on the information metric. For constructed models, the LRT had superior power. However, the goodness of fit test outperformed the LRT and deviance for certain data generators differing from the constructed full model for Poisson regression.

A grid search was utilized to determine the MLE of parameters in the logistic and alternative models used for the simulation study. This holds also for the simulation study for the Poisson regression model. Specifically, for logistic regression, the densities were noted to become more and more deterministic as the natural parameters go to $\pm\infty$. This behavior motivated the use of a grid obtained by constraining the η and thus the β parameters in the model in such a way as to avoid parameter values that correspond to (nearly) deterministic densities. The usefulness of the bounded grid was illustrated via simulation using data generators that did not satisfy constraints imposed by the grid. In particular, the simulation without such constraints showed that the test statistics can achieve zero power for specified parameter settings. An analytical explanation of such phenomena is presented and is based on the information distance (squared) involving limits to infinity.

Simulated power for testing equal means in a heteroscedastic Normal model, using the LRT and a test determined by comparing MLE parameter estimates, was comparable for both tests using various data generators. Superior power was achieved by two group clustering strategy involving the limit of the S_1 statistic.

Asymptotic distributional results associated with the likelihood ratio statistics for one dimensional between cluster lack of fit were derived. The derivation of the asymptotic non-central chi-square distribution under local parameter alternatives was based in part on the corresponding array of distributions being LAN. To ensure LAN for exponential family nonlinear regression models, the key condition was shown to be the requirement that the array of exponential family distributions be uniformly differentiable in quadratic mean. L_2 differentiability of the parametric array of densities associated with the array of exponential family distributions was then established, and hence LAN. Based on LAN, the asymptotic non-central chi-square distribution for the sequence of log likelihood ratio statistics was concluded.

The following points are suggested topics for future research:

1. Development of tests in this work to more general exponential family regression models to further test the effectiveness of the proposed cluster based lack of fit tests, as well as the geometrically motivated goodness of fit test statistic for exponential family regression based on the information metric.

2. Extension of the proposed procedures to test nonlinear within-cluster lack of fit and also mixtures of between-cluster within-cluster lack of fit.

3. Development of computational approaches for higher dimensional predictor spaces to accommodate implementation of the clustering based tests, including the extension of ordered partitions to higher dimensions.

4. Asymptotic analysis of the distribution for the LRT for higher dimensional predictor spaces, and further asymptotic analysis of the information based distance function to assess lack of fit in general exponential family regression models. Application of noncentral χ^2 for power analysis.

Bibliography

Amari, S. I. (1982). Differential geometry of curved exponential families-curvatures and information loss. *The Annals of Statistics 10*, 357–385.

Amari, S. I. (1985). *Differential-Geometrical Methods in Statistics*. Lecture Notes in Statistics 28, Springer Verlag, New York.

Amari, S. I., O. E. Barndorff-Nielsen, K. R. E., S. L. Lauritzen, and C. R. Rao (1987). *Differential Geometry in Statistical Inference*. Institute of Mathematical Statistics, Hayward,CA.

Amari, S. I. and H. Nagaoka (2000). *Methods of Information Geometry*. American Mathematical Society, Translations of Mathematical Monographs Volume 191, Oxford University Press.

Archer, k. J., S. Lemeshow, and D. W. Hosmer (2007). Goodness of fit tests for logistic regression models when data are collected using a complex sampling design. *Computational Statistics and Data Analysis 51*, 4450–4464.

Billinglsey, P. (1995). *Probability and Measure, 3ed*. Wiley, New York.

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

Brown, L. D. (1986). *Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory*. Institute of Mathematical Statistics, Hayward,CA.

Brown, L. D., T. T. Cai, and H. H. Zhou (2010). Nonparametric regression in exponential families. *The Annals of Statistics 38*, 2005–2046.

Cameron, A. C. and F. A. G. Windmeijer (1997). Nonparametric regression in exponential families. *Journal of Econometrics 77*, 329–342.

Choi, H. S. and N. M. Kiefer (2011). Geometry of the log-likelihood ratio statistic in misspecified models. *Journal of Statistical Planning and Inference 141*, 2091–2099.

Christensen, R. (1989). Lack of fit tests based on near or exact replicates. *The Annals of Statistics 17*, 673–683.

Christensen, R. (1991). Small-sample characterizations of near replicate lack-of-fit tests. *Journal of the American Statistical Association 86*, 752–756.

Claeskens, G. and N. L. Hjort (2008). *Model Selection and Model Averaging*. Cambridge Press, New York.

Cook, R. D. and C. L. Tsai (1990). Diagnostics for assessing the accuracy of normal approximations in exponential family nonlinear models. *Journal of the American Statistical Association 85*, 770–777.

Cordeiro, G. M. (1983). Improved likelihood ratio statistics for generalized linear models. *Journal of the Royal Statistical Society B 45*, 404–413.

Cordeiro, G. M. and G. A. Paula (1989). Improved likelihood ratio statistics for exponential family nonlinear models. *Biometrika 76*, 93–100.

Dobson, A. J. and A. G. Barnett (2008). *An Introduction to Generalized Linear Models, 3ed*. Chapman and Hall/CRC, Boca Raton, FL.

Ferrari, S. L. P. and G. M. Cordeiro (1996). Corrected score tests for exponential family nonlinear models. *Statistics and Probability Letters 26*, 7–12.

Hardle, W., E. Mammen, and M. Muller (1998). Testing parametric versus semiparametric modeling in generalized linear models. *Journal of the American Statistical Association 93*, 1461–1474.

Hosmer, D. W., T. Hosmer, S. L. Cessie, and S. Lemeshow (1997). A comparison of goodness-of-fit tests for the logistic regression model. *Statistics in Medicine 16*, 965–980.

Hosmer, D. W. and S. Lemeshow (1980). A goodness-of-fit test for the multiple logistic regression model. *Communications in Statistics A 10*, 1043–1069.

Hosmer, D. W. and S. Lemeshow (2000). *Applied Logistic Regression, 2ed.* Wiley, New York.

Hosmer, D. W., S. Lemeshow, and J. Klar (1988). Goodness-of-fit testing for the logistic regression model when the estimated probabilities are small. *Biometrical Journal 11*, 911–924.

Kass, R. E. and P. W. Vos (1997). *Geometrical Foundations of Asymptotic Inference.* Wiley, New York.

Lemonte, A. J. (2011). Local power of some tests in exponential family nonlinear models. *Journal of Statistical Planning and Inference 141*, 1981–1989.

Lindsay, B. and J. Liu (2009). Model assessment tools for a model false world. *Statistical Science 24*, 303–318.

McCullagh, P. (2002). What is a statistical model? *The Annals of Statistics 30*, 1225–1310.

Miller, F. R., J. W. Neill, and B. W. Sherfey (1998). Maximin clusters for near-replicate regression lack of fit tests. *The Annals of Statistics 26*, 1411–1433.

Miller, F. R., J. W. Neill, and B. W. Sherfey (1999). Implementation of a maximin power clustering criterion to select near replicates for regression lack-of-fit tests. *Journal of the American Statistical Association 94*, 610–620.

Munasinghe, W. P. (2010). *Cluster Based Lack of Fit Tests for Nonlinear Regression Models.* Ph. D. thesis, Kansas State University.

Neill, J. W. and F. R. Miller (2003). Limit experiments, lack of fit tests and fuzzy clusterings. *Technical Report II-03-1, Department of Statistics, Kansas State University*.

Neill, J. W., F. R. Miller, and D. D. Brown (2000). Maximin clusters for near replicate multiresponse lack of fit tests. *Technical Report II-00-1, Department of Statistics, Kansas State University*.

Neill, J. W., F. R. Miller, and D. D. Brown (2002). Maximin clusters for nonreplicated multiresponse lack of fit tests. *Journal of Statistical Planning and Inference 102*, 359–375.

Pan, Z. and D. Y. Lin (2005). Goodness-of-fit methods for generalized linear mixed models. *Biometrics 61*, 1000–1009.

Pregibon, D. (1980). Goodness of link tests for generalized linear models. *Journal of the Royal Statistical Society C 29*, 15–24.

Reider, H. (1994). *Robust Asymptotic Statistics*. Springer-Verlag, New York.

Sarkar, S. K. and H. Midi (2010). Importance of assessing the model adequacy of binary logistic regression. *Journal of Applied Sciences 10*, 479–486.

Sellers, K. F. and G. Shmuelli (2010). A flexible regression model for count data. *The Annals of Applied Statistics 4*, 943–961.

Shao, J. (2003). *Mathematical Statistics, 2ed*. Springer-Verlag, New York.

Shmueli, G. (2010). To explain or to predict? *Statistical Science 25*, 289–310.

Stute, W. and L. X. Zhu (2002). Model checks for generalized linear models. *Scandinavian Journal of Statistics 29*, 535–545.

Su, J. Q. and L. J. Wei (1991). A lack-of-fit test for the mean function in a generalized linear model. *Journal of the American Statistical Association 86*, 420–426.

Tsiatis, A. A. (1980). A note on a goodness of fit test for the logistic regression model. *Biometrika 67*, 250–251.

Wei, B. C. (1998). *Exponential Family Nonlinear Models*. Lecture Notes in Statistics 130, Springer-Verlag, New York.

Wei, B. C., J. Q. Shi, W. K. Fung, and Y. Q. Hu (1998). Testing for varying dispersion in exponential family nonlinear models. *Annals of the Institute of Statistical Mathematics 50*, 277–294.

Zheng, B. (2000). Summarizing the goodness of fit of generalized linear models for longitudinal data. *Statistics in Medicine 19*, 1265–1275.

Appendix A

Generated Dial Settings for Logistic and Poisson models

-2.68667657	-2.48197830	-2.42802302	-2.13745227	-2.00318555
-1.95941877	-1.69872844	-1.59118778	-1.56786299	-1.50355328
-1.42593390	-1.31627511	-1.25794782	-1.16061046	-0.84807659
-0.72372107	-0.71894608	-0.58391733	-0.42173272	-0.16859679
0.01478212	0.05951435	0.21546714	0.25615356	0.34210246
0.45218965	0.45445007	0.53302232	0.63596014	0.83929070
0.86315950	0.93246935	1.05064720	1.25633511	1.25733595
1.26047923	1.28097711	1.30209739	1.50219093	1.58049510
1.92063971	1.99529495	2.28118479	2.29621865	2.46422383
2.46846361	2.49383085	2.53782106	2.58729063	2.62606696

Table A.1: *: Generated and sorted n = 50 Uniform [-3,3] values*

-2.95512624	-2.83734196	-2.68667657	-2.53651432	-2.4819783
-2.42802302	-2.34434466	-2.31650772	-2.15560175	-2.13745227
-2.07580989	-2.02323166	-2.00318555	-1.95941877	-1.8153121
-1.7979381	-1.69872844	-1.62537257	-1.59118778	-1.56786299
-1.50355328	-1.4259339	-1.38068114	-1.31627511	-1.25794782
-1.24070817	-1.16061046	-1.05677897	-0.9692665	-0.93217621
-0.91166776	-0.89128479	-0.84807659	-0.74177582	-0.72372107
-0.71894608	-0.60717476	-0.58391733	-0.42173272	-0.33716353
-0.30362781	-0.16859679	0.01478212	0.05951435	0.21546714
0.25615356	0.34210246	0.45218965	0.45445007	0.53302232
0.59122192	0.63596014	0.75428478	0.76121893	0.8392907
0.84006936	0.85191938	0.8631595	0.89617543	0.93246935
1.0506472	1.09889655	1.14392149	1.2120276	1.24452919
1.25633511	1.25733595	1.26047923	1.28097711	1.30209739
1.32421541	1.37130392	1.39563853	1.45556928	1.50219093
1.5804951	1.61730259	1.64826562	1.74307128	1.77939755
1.92063971	1.99529495	2.2659975	2.28118479	2.29621865
2.35097442	2.43089541	2.46422383	2.46846361	2.49383085
2.51545912	2.53782106	2.5491042	2.58643838	2.58729063
2.62606696	2.66733335	2.85818869	2.87399123	2.88522763

Table A.2: *: Generated and sorted n = 100 Uniform [-3,3] values*

Appendix B

Crisp Clustering for Logistic and Poisson Models

Point	(β_0, β_1) Coordinate	Crisp Clustering					
		$x_1 : x_{17}$ $x_{18} : x_{40}$ $x_{41} : x_{50}$	$x_1 : x_{17}$ $x_{18} : x_{36}$ $x_{37} : x_{50}$	$x_1 : x_{17}$ $x_{18} : x_{33}$ $x_{34} : x_{50}$	$x_1 : x_{17}$ $x_{18} : x_{32}$ $x_{33} : x_{50}$	$x_1 : x_{17}$ $x_{18} : x_{29}$ $x_{30} : x_{50}$	$x_1 : x_{17}$ $x_{18} : x_{28}$ $x_{29} : x_{50}$
1	(11,0)	x					
2	(5,-2)						x
3	(5,0)	x					
4	(5,2)		x				
5	(0,-3)		x				
6	(0,-2)	x					
7	(0,0)	x					
8	(0,2)	x					
9	(0,3)		x				
10	(-5,-2)		x				
11	(-5,0)	x					
12	(-5,2)						x
13	(-11,0)	x					
14	(-1,-1)			x			
15	(1,-1)	x				x	
16	(-1,1)	x				x	
17	(1,1)			x			
18	(-3,0)	x					
19	(3,0)	x					
20	(-2,0)	x					
21	(2,0)	x					

Table B.1: *Crisp clustering for logistic model at selected 21 points when n=50*

Point	(β_0, β_1) Coordinate	Crisp Clustering				
		$x_1 : x_{37}$ $x_{38} : x_{80}$ $x_{81} : x_{100}$	$x_1 : x_{37}$ $x_{38} : x_{66}$ $x_{67} : x_{100}$	$x_1 : x_{37}$ $x_{38} : x_{62}$ $x_{63} : x_{100}$	$x_1 : x_{37}$ $x_{38} : x_{57}$ $x_{58} : x_{100}$	$x_1 : x_{37}$ $x_{38} : x_{51}$ $x_{52} : x_{100}$
1	(11,0)	x				
2	(5,-2)					x
3	(5,0)	x				
4	(5,2)			x		
5	(0,-3)		x			
6	(0,-2)				x	
7	(0,0)	x				
8	(0,2)				x	
9	(0,3)		x			
10	(-5,-2)			x		
11	(-5,0)	x				
12	(-5,2)					x
13	(-11,0)	x				
14	(-1,-1)			x		
15	(1,-1)	x				
16	(-1,1)	x				
17	(1,1)			x		
18	(-3,0)	x				
19	(3,0)	x				
20	(-2,0)	x				
21	(2,0)	x				

Table B.2: *Crisp clustering for logistic model at selected 21 points when n=100*

Point	(β_0, β_1) Coordinate	Crisp Clustering			
		$x_1 : x_{17}$ $x_{18} : x_{40}$ $x_{41} : x_{50}$	$x_1 : x_{17}$ $x_{18} : x_{38}$ $x_{39} : x_{50}$	$x_1 : x_{17}$ $x_{18} : x_{36}$ $x_{37} : x_{50}$	$x_1 : x_{17}$ $x_{18} : x_{28}$ $x_{29} : x_{50}$
1	(11.5,0)	x			
2	(8.5,-1)		x		
3	(8.5,0)	x			
4	(8.5,1)				x
5	(6.5,-1.5)			x	
6	(6.5,0)	x			
7	(6.5,1.5)				x
8	(4.5,-1)		x		
9	(4.5,0)	x			
10	(4.5,1)				x
11	(1.5,0)	x			
12	(4.5,-0.5)	x			
13	(6.5,-0.5)	x			
14	(8.5,-0.5)	x			
15	(4.5,0.5)				x
16	(6.5,0.5)				x
17	(8.5,0.5)				x

Table B.3: *Crisp clustering for Poisson model at selected 17 points when n=50*

Point	(β_0, β_1) Coordinate	Crisp Clustering				
		$x_1 : x_{37}$ $x_{38} : x_{80}$ $x_{81} : x_{100}$	$x_1 : x_{37}$ $x_{38} : x_{75}$ $x_{76} : x_{100}$	$x_1 : x_{37}$ $x_{38} : x_{62}$ $x_{63} : x_{100}$	$x_1 : x_{37}$ $x_{38} : x_{63}$ $x_{64} : x_{100}$	$x_1 : x_{37}$ $x_{38} : x_{51}$ $x_{52} : x_{100}$
1	(11.5,0)	x				
2	(8.5,-1)				x	
3	(8.5,0)	x				
4	(8.5,1)					x
5	(6.5,-1.5)			x		
6	(6.5,0)	x				
7	(6.5,1.5)					x
8	(4.5,-1)				x	
9	(4.5,0)	x				
10	(4.5,1)					x
11	(1.5,0)	x				
12	(4.5,-0.5)		x			
13	(6.5,-0.5)		x			
14	(8.5,-0.5)		x			
15	(4.5,0.5)	x				
16	(6.5,0.5)	x				
17	(8.5,0.5)	x				

Table B.4: *Crisp clustering for Poisson model at selected 17 points when n=100*

Appendix C

R Code for Implementing the Power Calculations for Heterogeneous Normal Model

```
#################################################################
# Calculating the power of the LRT for 2 replicate case    #
#  based on Sufficient Statistics                          #
#################################################################
n<-50
set.seed(2250)
ki<-sort(runif(n,1,2))
si<-1/(ki^2)

l<-17
z<-c(rep(1,l),rep(0,n-l))
b<-sqrt(l/(n-l))
a<--1/b
c(a,b)
```

```
gammai<-c(rep(a,l),rep(b,n-l))

# initial values
m<-1
r<-1
alpha0<-seq(0,8,by=1)
B<-1000

LLRlist<-lapply(as.list(1:length(alpha0)),function(i)
{
alpha<-alpha0[i]
set.seed(722)
Z<- matrix(c(rnorm(n*B,0,1)),n,B,byrow=F)
likelihood<-lapply(as.list(1:B),function(j,x1)
# function calculating the log-likelihood ratio
{ x1<-m+ alpha*ki+r*ki*Z[,j]
  x2<-x1^2

  x1bar<-sum(si*x1)
  x2bar<-sum(si*x2)
  xthil<-sum(gammai*x1)

# parameters for the full model
  t2f<--n/(2*(x2bar-x1bar^2/sum(si)-xthil^2/sum(gammai^2/si)))
  tau1f<-x1bar/sum(si)
  t1f<--2*tau1f*t2f
  tau2f<-xthil/sum(gammai^2/si)
```

```
  sf<- -2*tau2f*t2f

  theta1f<-si*t1f+sf*gammai  # creating nx1 theta vectors
  theta2f<-si*t2f

  # log likelihood for full
  lf<-x1bar*t1f+x2bar*t2f+xthil*sf-sum(-(theta1f^2)/(4*theta2f)
                                +(1/2)*log(pi/(-theta2f)))

# parameters for null model
  t2n<--n/(2*(x2bar-x1bar^2/sum(si)))
  tau1n<-x1bar/sum(si)
  t1n<--2*tau1n*t2n
  tau2n<-0
  sn<-0

  theta1n<-si*t1n+sn*gammai
  theta2n<-si*t2n

  # log likelihood for null
  ln<-x1bar*t1n+x2bar*t2n+xthil*sn-sum(-(theta1n^2)/(4*theta2n)
                               +(1/2)*log(pi/(-theta2n)))

  # log likelihood ratio
  llr<--2*(ln-lf)
  llr
```

```
  S<-xthil^2/((-1/(2*t2f))*sum(gammai^2/si))
  c(llr,S)
}) # end of likelihood calculation function
}) # end of LLRlist calculation function

power.mat<-sapply(as.list(1:length(LLRlist)),function(j)
{
  mle.mat1<-LLRlist[[j]]
  mle.mat2<-matrix(0,B,2)
  power<-numeric(2)
  for (i in 1:2)
  {
  mle.mat2[,i]<-sapply(mle.mat1, function(x) x[i])
  power[i]<-sum(ifelse(mle.mat2[,i]>=qchisq(0.95,1),1,0))/B
  power
  }
  power
}) # end of power.mat calculation function

   Slim<-(n/(n-((sum(1/ki))^2/(sum(1/ki^2)))-((sum(gammai*ki))^2/
          (sum(gammai^2*ki^2)))))*((sum(gammai*ki))^2/sum(gammai^2*ki^2))
   print(Slim)

power<-cbind(alpha0,t(power.mat))
colnames(power)<-c("alpha","llr","S1")
power
```

Appendix D

R Code for Implementing the Power Calculations for Logistic Model

```
############################################################
# Function to calculate the overlapping subsets (covers) #
############################################################

cover<-function(n,m,k,l,xs)
{ # n- number of observations
  # xs -  generated x values
  c<-sapply(1:m,function(i)
  {
  # cover [k,l] divided in to m=5 cells
  # here we consider cells based on values c1:[-3,-1.8), c2:[-1.8,-0.6),...c5:[1.8,3]
  c<-ifelse(xs<(k+i*(l-k)/m) & xs>=(k+(i-1)*(l-k)/m),1,0)
}) # end of c calculation function

overlap<-sapply(1:m,function(i)
# function calculates the ovelapping subsets
```

```
{
  if (i%%2==0) rep(0,n)
  else if  (i==1) os<- c[,i]+c[,i+1]
  else if (i==m & i%%2==1) os<-c[,i-1]+c[,i]
  else if (i%%2==1)  os<-c[,i-1]+c[,i]+c[,i+1]
}) # end of overlap calculation function

# taking the odd columns in overlap
index<-seq(1,m, by=2)
z<-overlap[,index]  # matrix indicating the overlapping subsets for the selected cover
colnames(z)<-c("Fn1","Fn2","Fn3")

ca<-(sum(c[,2])+1)*(sum(c[,4])+1)  # cardinality of Kn,0
B<-matrix(rep(0,(n*(n-1))),n)  # matrix for the edge sets
out<-list(z,B,ca,c)
names(out)<-c("z","B","ca","c")
out
}  #end of cover function

#####################################################################
#Finding tangent space to the expectation surface at each x.      #
#logistic model f(x, beta0, beta1)=beta0+beta1*x) is used.        #
#####################################################################
tm<-function(xs,V)
{  # V - covariance matrix of xs
  B<-cbind(rep(1,n),xs)
  TM<-V%*%B%*%ginv(t(B)%*%V%*%B)%*%t(B)
```

```
  TM
}

#####################################################################
#Finding perpendicular projection operator for a vector space.
#####################################################################
ppo<-function(c,V)
{
# c is the matrix that we want to project onto
  Ppo<-c%*%ginv(t(c)%*%ginv(V)%*%c)%*%t(c)%*%ginv(V)
  Ppo
}

#####################################################################
#Finding basis set for the matrix based on output from ppo function.
#####################################################################
basis<-function(ppo.out)
{
  ba<-ppo.out%*%t(ppo.out)
  e<-eigen(ba,symmetric=T) # calculating eigen values and vectors
  e.vals<-e$values
  e.vects<-e$vectors
  n.evals<-e.vals[abs(e.vals)>10e-5] # avoid the prblem of decimal places
                                          used in R to get correct d.f.
  n1<-length(n.evals)
  Basis<-e.vects[,1:n1] # the final basis set
  Basis
```

```
}

##########################################
# selecting maximin power clusterings    #
##########################################
atoms<-function(z,ca,c1,c2,c3,c4,c5,V)
{ # z - overlapping subsets from cover function
  # ca - cardinality
  # c1-c5 - elements corresponding to 5 cells
  atom<-list(length=ca)
  a<-1    # index counter for number of atoms
  ZZa<-matrix(rep(0,nrow(z)*ncol(z)),nrow(z),ncol(z))
   # creating an empty crisp atom matrix
  for (i in 1:nrow(z))
    {
      ZZa[i,1]<-c1[i]
      ZZa[i,2]<-c3[i]
      ZZa[i,3]<-c5[i]
    }
    for (j in sum(c1):sum(c1+c2))
      # consider elements in first even cell correspond to 1st column
    {
      ZZa[j,1]<-1
      for (k in (j+1):sum(c1+c2))
        # consider elements in first even cell correspond to 2nd column
        ZZa[k,2]<-1
      if (j>sum(c1))
```

```
  # cleaning up extra 1's in second column before the required elements
{
  for (l in sum(c1):j)
  ZZa[l,2]<-0
}
for (m in sum(c1+c2+c3):sum(c1+c2+c3+c4))
  # consider elements in 2nd even cell corresponds to 2nd column
{
  ZZa[m,2]<-1
  for (o in (m+1):sum(c1+c2+c3+c4))
    # consider elements in 2nd even cell corresponds to 3rd column
    ZZa[o,3]<-1
  if (m>sum(c1+c2+c3))
    # cleaning up extra 1's in 3rd col before the required elements
  {
     for (p in sum(c1+c2+c3):m)
      ZZa[p,3]<-0
  }
  for (q in 1:nrow(z))
    # cleaning up extra 1's in second col after the required elements
  {
    if (sum(ZZa[q,])>1)
      ZZa[q,2]<-0
  }
  PZZ<-ppo(ZZa,V)-ppo(ppo(ZZa,V)%*%tm(xs,V),V)
  #gives perpendicular projection of tm(beta1, beta2) onto zza (crisp atoms)#
  PZZb<- basis(PZZ)#basis set for pzz, use for calculation of tau#
```

```
        cat(PZZb,file="Basis.txt","\n",append=TRUE) #stores the basis set in R#

        atom[[a]]<-ZZa
        a<-a+1
      } # end of m loop
    } # end of j loop
    atom
} # end of function atoms

################################################
# creating non-overlapping subset of odd cells #
################################################
zz<-function(c1,c2,c3,c4,c5)
{
  z1<-matrix(0,n,ncol(z))
  {
   if (sum(c1)<2)
   {
     z1[,1]<-c1+c2
     if (sum(c3)<2)
     {
       z1[,2]<- c3+c4
       z1[,3]<-c5
     }
     else
     {
       z1[,2]<-c3
```

```
      z1[,3]<- c4+c5
    }
  }
  else
  {
     z1[,1]<-c1
     z1[,2]<-c2+c3
     z1[,3]<-c4+c5
  }
  z1
 }
}

#######################################################
# calculating the positions of 1's in columns of z #
#######################################################
posi<-function(z,i)
{
 x<-z[,i]  # consider the columns of z
 k<-rep(0,length(x))  # vector to store the positions for each col of z
  for (j in 1:length(x))
  {
   if (x[j]!=0) # get the non-zero elements of the cols of z
   k[j]<-j   # assign those positions to k
  }
 a<-k[k!=0] # get the non-zero positions of non-zero elements of the cols of z
 {
```

```
    combn(a,2)  # find all the combinations taking 2 at a time
  }
}

#####################################################
# Function to get w values using correct x values #
#####################################################
# B is the edge set
w.values<-function(z,xs,B)
{
  for (a in 1:(nrow(z)-1))
  {
    if (sum(B[,a])>1)
    break
  }
  for (ab in 1:nrow(z))
  {
    if (B[ab,a]==1)
    break
  }
  for (ac in ab+1:nrow(z))
  {
    if (B[ac,a]==1)
    break
  }
  w<-(((xs[ab])-(xs[ac]))^2)  # weights for a particular element in edge set
  cat(w,file="weights.txt","\n",append=T)
```

```
}

#######################################
# Function to calculate the edge set  #
#######################################
edges<-function(z)
{
  m<-1
  B.list<-list()
  for (j in 1:ncol(z))
  {
    pos<-posi(z,j)  # combinations of positions of 1's in cols of z
    for (k in 1:ncol(pos))
    {
      B<-matrix(0,n,(n-1))
      pos1<-pos[,k]     # consider each combination seperately
        h<-pos1[1]    # choose the column in edge set to assign 1
        B[pos1,h]<-1  # assign 1's for 2 points in same cluster
        for (a in 1:pos1[1])  # assign 1's for individual clusterings
        {
          B[a,a]<-1
        }
        for (l in (pos1[1]):(pos1[2]))
        {
          if (sum(B[l,])==0) B[l,l]<-1
        }
        for (i in 1:nrow(B))
```

```
        {
          if (sum(B[i,])==0) B[i,(i-1)]<-1
        }
      weight<-w.values(z,xs,B)
      B.list[[m]]<-B
      m<-m+1 # index counter for number of edge sets
    }
  }
  B.list
}

#############################################################
# Function to calculate (w_Z_n)  that belongs to edge set #
#############################################################
wzn<-function(wi)
{
  for (i in 1:nrow(wi))
  {
    ww<-wi[i,]/sum(wi)
    cat(ww,file="totalweights.txt","\n",append=T)
  }
}

#####################################################################
# Function to calculate (l_Z_n)  and tau that belongs to each atom #
#####################################################################
# 'weight' W_Z_n values based on edge sets#
```

```
# vv's are the basis for each atom
lzn<-function(Basis,weight,edgeset,V)
{
  lzn1<-sapply(1:nrow(Basis),function(i)
  {
    tau<-0
    vv<-Basis[i,]
    tau.d<-sapply(1:length(edgeset),function(d) # calculate tau for each edgeset
     {
      B<-edgeset[[d]]
      x0<-ppo(B,V)%*%tm(xs,V)
      tau1<-weight[d]%*%(t(vv)%*%((ppo(B,V)-ppo(x0,V))%*%ginv(V))%*%vv)
      tau1
    })  #end of calculation of tau for each atom
    #calculate l_z_n before loop takes this to the next atom)
    tau<-sum(tau.d)
    lZn<-(t(vv)%*%vv)*(1/(tau))
    lZn
  })
}

################################################
#function to calculate g in fuzzy clusterings #
################################################
g<-function(x)
{
  if (x<=0) gx<-0
```

```
  if (x>0 & x<1) gx<-1/32*(-5*(2*x-1)^7+21*(2*x-1)^3+35*(2*x-1)+16)
  if (x>=1) gx<-1
  gx
}

#########################################################################
# Function to calculate the fuzzy clusterings for a given beta vector #
#########################################################################
zbeta<-function(z,beta,b1,b2)
{  # z- list of crisp clustering matrices
   # beta - point to find the fuzy clusters
   # b1,b2 - points of th crisp clusters
  znb<-list()
  n1<-length(b1)
  n2<-length(b2)
  if (beta[1]<=b1[1])
    for (j in 1:n2) znb[[j]]<-z[[1+(j-1)*n1]]
  if (beta[1]>b1[n1])
    for (j in 1:n2) znb[[j]]<-z[[j*n1]]
  if (beta[1]>b1[1] & beta[1]<=b1[n1])
  {
    for (i in 1:n1)
    {
      if (beta[1]>b1[i] & beta[1]<=b1[i+1])
      {
        for (j in 1:n2)
          znb[[j]]<-(1-g((beta[1]-b1[i])/(b1[i+1]-b1[i])))*z[[(i+(j-1)*n1)]]
```

```
                    +g((beta[1]-b1[i])/(b1[i+1]-b1[i]))*z[[(i+1+(j-1)*n1)]]
      }
    }
  }
  if (beta[2]<=b2[1]) zn<-znb[[1]]
  if (beta[2]>b2[n2]) zn<-znb[[n2]]
  if (beta[2]>b2[1] & beta[2]<=b2[n2])
  {
    for (j in 1:n2)
    {
      if (beta[2]>b2[j] & beta[2]<=b2[j+1])
          zn<-(1-g((beta[2]-b2[j])/(b2[j+1]-b2[j])))*znb[[j]]+g((beta[2]-b2[j])/
                                                  (b2[j+1]-b2[j]))*z[[j+1]]
    }
  }
  zn
}

#########################################################################
# function that calculates the fine grid for a given x range (b,d) #
#########################################################################
gridp<-function(delta1,delta2,b,d,dis)
{  # function that calculates the fine grid for a given x range (b,d)
   # b= lower limit of x
   # d= upper limit of x
   # dis = gap between grid points
c<--b
```

```
a<-1/(c+d)
b0bar<-seq(delta1,delta2,by=dis)
b1bar<-matrix(0,length(b0bar),length(b0bar))
grid<-list()
grid_bar<-list()
j<-1
for (i in 1:length(b0bar))
{
  b1bar[i,]<-seq(delta1-b0bar[i],delta2-b0bar[i],by=dis)
  for (k in 1:length(b1bar[i,]))
  {
    b1<-a*b1bar[i,k]
    b0<-b0bar[i]+c*a*b1bar[i,k]
    grid[[j]]<-c(b0,b1)
    grid_bar[[j]]<-c(b0bar[i],b1bar[i,k])
    j<-j+1
  }
}
grid
}

# Function used for Hosmer Lemeshow method

hosmerlem = function(y, yhat, g) {
  cutyhat = cut(yhat,
     breaks = quantile(yhat, probs=seq(0,
       1, 1/10)), include.lowest=TRUE)
```

```
  obs = xtabs(cbind(mi - y, y) ~ cutyhat)
  expect = xtabs(cbind(mi - yhat, yhat) ~ cutyhat)
  chisq = sum((obs - expect)^2/expect)
  P = 1 - pchisq(chisq, 10 - 2)
  return(c(chisq=chisq,p.value=P))
}

# End of set of functions in the program

##############################
# main body of the program  #
##############################
library(MASS)
betamatrix<-matrix(c(-1,1,-1,3,1,1,1,3),4,2,byrow=T)
ni<-50
n<-50 # number of samples (dial settings)
k<-0
l<-1
set.seed(2250)
x<-runif(n,k,l) # x vales from Uniform[0,1]
xs<-sort(x)  # sorted x values

m<-5 # number of clusters

cover.out<-cover(n,m,k,l,xs)
#cover.out
```

```
z<-cover.out[[1]]
ca<-cover.out[[3]]
B<-cover.out[[2]]
c<- cover.out[[4]]
c1<-c[,1]
c2<-c[,2]
c3<-c[,3]
c4<-c[,4]
c5<-c[,5]

maximin<-lapply(as.list(1:(nrow(betamatrix))),function(i)
{
# initial values
betavec<-betamatrix[i,]

# generating data

eta<-cbind(rep(1,n),xs)%*%betavec  # linear predictor
pii<-exp(eta)/(1+exp(eta)) # probabilities
set.seed(4650)
yi<-rbinom(n,ni,pii)
V<-diag(as.vector(ni*pii*(1-pii)))
dataset<-cbind(xs,pii,yi,ni)
colnames(dataset)<-c("xi","pi","yi","ni")
dataset    # generated data set

cat("number of crisp atoms=",ca,"\n")
```

```
atom<-atoms(z,ca,c1,c2,c3,c4,c5,V) # ordered atoms
#atom

nonoverlap<-zz(c1,c2,c3,c4,c5)  #creating non-overlapping subset of odd
                                      cells to apply edgeset function
#nonoverlap
edgeset<-edges(nonoverlap)  # calculation of edgesets
#edgeset
cat("number of edge sets=",length(edgeset),"\n")

wi<-as.matrix(read.table("weights.txt")) # output from w.values function  inside
                                                      the edges function
wzn(wi)      # running wzn function

weight<-as.matrix(read.table("totalweights.txt")) # read the ourput from wzn
Basis<- as.matrix(read.table("Basis.txt")) # read the output basis, from
                                                      atoms function
lzn.val<-lzn(Basis,weight,edgeset,V)     # caculating lzn values
lzn.val

lzn.max<-max(lzn.val)   # get the maximum of lzn values
lzn.max
position<-0          # finding the position of max(lzn)
for (j in 1:length(lzn.val))
  if (lzn.val[j]==lzn.max)  position<-j
position
```

```
maximin<-atom[[position]]    # the atom correspond to max(lzn)
cat("beta vector :",betavec,"\n")
cat("position of atom =",position,"\n")
cat("maximum lzn =",lzn.max,"\n")

file.remove("weights.txt")
file.remove("totalweights.txt")
file.remove("Basis.txt")

print(maximin)
})

# Calculating the power corresponding to one set of (beta0,beta1) values generating
                      data from constructed full model
grid2<-gridp(-12,12,0,1,0.5)
Z<-matrix(rep(c(1,0,1,0,1),c(6,50,22,50,22)),n,3)    # maximin[[3]] creating the
                                        maximin atom according to the beocat output
betav<-c(1,1)
s.vec<-c(0,2,4,8,12,16,20)
powerf<-sapply(1:length(s.vec),function(k)
{
  s<-s.vec[k]

  # proposed (null) model
  eta.n<-cbind(rep(1,n),xs)%*%betav  # linear predictor for proposed model (null)
      for (a in 1:length(eta.n)) {if (eta.n[a]>7.08e+02) eta.n[a]<-7.08e+02}
  pi.n<-exp(eta.n)/(1+exp(eta.n)) # probabilities
```

```
mu.n<-mi*pi.n
sigma2.n<-mi*pi.n*(1-pi.n)
V<-diag(as.vector(mi*pi.n*(1-pi.n)))

# altrenative (full) model
betai<-basis(ppo(Z,V)-ppo(ppo(Z,V)%*%tm(xs,V),V))
v<- betai/sigma2.n
eta.f<- eta.n + s*v
    for (a in 1:length(eta.f)) {if (eta.f[a]>7.08e+02) eta.f[a]<-7.08e+02}
pi.f<-exp(eta.f)/(1+exp(eta.f)) # probabilities
mu.f<-mi*pi.f
sigma2.f<-mi*pi.f*(1-pi.f)

#set.seed(6745)
B<-500
Y<-matrix(rbinom(n*B,mi,pi.f),n,B)
like<-sapply(1:B,function(j)
{
  yi<-Y[,j]

  mle<-sapply(1:(length(grid2)),function(i)  # finding beta which maximizes
                                               the likelihood
  {
    betavec<-grid2[[i]]
    eta.n<-cbind(rep(1,n),xs)%*%betavec
      for (a in 1:length(eta.n)) {if (eta.n[a]>7.08e+02) eta.n[a]<-7.08e+02}
    pi.n<-exp(eta.n)/(1+exp(eta.n)) # probabilities
```

```
mu.n<-mi*pi.n
sigma2.n<-mi*pi.n*(1-pi.n)
V<-diag(as.vector(mi*pi.n*(1-pi.n)))
betai<-basis(ppo(Z,V)-ppo(ppo(Z,V)%*%tm(xs,V),V))
v<- betai/sigma2.n
logl.n<- sum(yi*eta.n-mi*log(1+exp(eta.n)))

v.pos<-numeric(n); v.neg<-numeric(n); v.ze<-numeric(n)
for (i in 1:length(v))
{
  if (v[i]>0) { v.pos[i]<-v[i] ; v.neg[i]<-0; v.ze[i]<-0 }
  if (v[i]<0) { v.pos[i]<-0 ; v.neg[i]<-v[i]; v.ze[i]<-0 }
  if (v[i]==0) { v.pos[i]<-0 ; v.neg[i]<-0; v.ze[i]<-v[i]}
}
fs.pl<-sum((yi-mi)*v.pos)+sum(yi*v.neg)+sum((yi-mi*pi.n)*v.ze)  # positive
                                                          infinity limit of f(s)
fs.nl<- sum(yi*v.pos)+sum((yi-mi)*v.neg)+sum((yi-mi*pi.n)*v.ze) # negative
                                                          infinity limit of f(s)
if (fs.pl*fs.nl>0) print(grid2[[i]])  # non convergence points
else
{
  s1<--100; s2<-100
  eta.f1<- eta.n + s1*v
    for (a in 1:length(eta.f1)) {if (eta.f1[a]>7.08e+02) eta.f1[a]<-7.08e+02}
  eta.f2<- eta.n + s2*v
    for (a in 1:length(eta.f2)) {if (eta.f2[a]>7.08e+02) eta.f2[a]<-7.08e+02}
  pi.f1<-exp(eta.f1)/(1+exp(eta.f1))
```

```
pi.f2<-exp(eta.f2)/(1+exp(eta.f2))
fs1<-sum((yi-mi*pi.f1)*v)
fs2<-sum((yi-mi*pi.f2)*v)
if (fs1<0)                  # finding a positive f(s1)
{
  while (fs1<0)
  {
    s1<-s1-1
    eta.f1<- eta.n + s1*v
       for (a in 1:length(eta.f1)) {if (eta.f1[a]>7.08e+02) eta.f1[a]
                                          <-7.08e+02}
    pi.f1<-exp(eta.f1)/(1+exp(eta.f1))
    fs1<-sum((yi-mi*pi.f1)*v)
  }
}
if (fs2>0)                   # finding a negative f(s2)
{
  while (fs2>0)
  {
    s2<-s2+1
    eta.f2<- eta.n + s2*v
     for (a in 1:length(eta.f2)) {if (eta.f2[a]>7.08e+02) eta.f2[a]
                                                 <-7.08e+02}
    pi.f2<-exp(eta.f2)/(1+exp(eta.f2))
    fs2<-sum((yi-mi*pi.f2)*v)
  }
}
```

```
while (fs1>1e-5 & fs2<(-1e-5))
{
  s1.new<-(s1+s2)/2
  eta.f1<- eta.n + s1.new*v
    for (a in 1:length(eta.f1)) {if (eta.f1[a]>7.08e+02) eta.f1[a]
                                                    <-7.08e+02}
  pi.f1<-exp(eta.f1)/(1+exp(eta.f1))
  fs1<-sum((yi-mi*pi.f1)*v)
  if (fs1>0)
  {
    s1<-s1.new
  }
  if (fs1<0)
  {
    s2<-s1.new
    eta.f2<- eta.n + s2*v
     for (a in 1:length(eta.f2)) {if (eta.f2[a]>7.08e+02) eta.f2[a]
                                                    <-7.08e+02}
    pi.f2<-exp(eta.f2)/(1+exp(eta.f2))
    fs2<-sum((yi-mi*pi.f2)*v)
      eta.f1<- eta.n + s1*v
        for (a in 1:length(eta.f1)) {if (eta.f1[a]>7.08e+02) eta.f1[a]
                                                    <-7.08e+02}
      pi.f1<-exp(eta.f1)/(1+exp(eta.f1))
      fs1<-sum((yi-mi*pi.f1)*v)
  }
}
```

```
      logl.f<- sum(yi*eta.f1-mi*log(1+exp(eta.f1)))
    }
    cat(c(betavec,s1.new,logl.f,logl.n),file="result.txt","\n",append=T)
  })    # end of mle calculation function

  results<-as.matrix(read.table("result.txt")) # read the ourput

  file.remove("result.txt")
  maxl.f<-max(results[,4])
  maxl.n<-max(results[,5])

  posi.f <-match(maxl.f,results[,4])
  mle.f<-c(results[posi.f,1:3])
  mle.f

  posi.n<-match(maxl.n,results[,5])
  mle.n<-c(results[posi.n,1:2])
  mle.n

  cat(c(mle.f,mle.n,maxl.f,maxl.n),file="likelihood.txt","\n",append=T)

}) # end of like calculation function

likelihood<-as.matrix(read.table("likelihood.txt")) # read the ourput
file.remove("likelihood.txt")

llr<--2*(likelihood[,7]-likelihood[,6])
```

```
  power<-sum(ifelse(llr>=qchisq(0.95,1),1,0))/B
  power
  print(power)
}) # end of power calculation function

cbind(s.vec,powerf)
plot(s.vec,powerf,type="b")

# Calculating the maximin clustering matrix at 4 given (beta0,beta1) values

grid2<-gridp(-12,12,0,1,0.1)

z11<-matrix(rep(c(1,0,1,0,1),c(17,50,23,50,10)),n,3)   # beta(-1,1)
z12<-matrix(rep(c(1,0,1,0,1),c(11,50,17,50,22)),n,3)  # beta(-1,3)
z21<-matrix(rep(c(1,0,1,0,1),c(6,50,22,50,22)),n,3)    # beta(1,1)
z22<-matrix(rep(c(1,0,1,0,1),c(6,50,22,50,22)),n,3)   # beta(1,3)
z<-list(z11,z21,z12,z22)

b1<-c(-1,1)
b2<-c(1,3)

fuz<-list(NULL)
fuzzy<-lapply(1:(length(grid2)),function(i)
{
  fuz<-zbeta(z,grid2[[i]],b1,b2)
  fuz
})
```

```
# R-code to Implement Power Calculations with Rogue Data Generator with
                    Hosmer-Lemeshow data generator
source("functions.r")

library(MASS)

mi<-50
n<-50 # number of samples (dial settings)
k<--3
l<-3
set.seed(2250)
x<-runif(n,k,l) # x vales from Uniform[0,1]
xs<-sort(x)  # sorted x values

grid2<-gridp(-12,12,-3,3,0.5)

Z<-matrix(rep(c(1,0,1,0,1),c(6,50,22,50,22)),n,3)    # maximin[[3]] creating the
                                        maximin atom according to the beocat output

J<-c(0.01,0.02,0.03,0.04,0.05,0.06)

powerf<-sapply(1:length(J),function(k)
{
  J1<-J[k]
  b1<-(log(0.95/0.05)-log(J1/(1-J1)))/6
```

```
b2<-(2*log(0.95/0.05)-4.5*b1)/6.75
b0<-log(0.95/0.05)-3*b1-9*b2
betav<-c(b0,b1)

# data generator
eta.n<-cbind(rep(1,n),xs)%*%betav+b2*xs^2  # linear predictor for proposed model
  for (a in 1:length(eta.n)) {if (eta.n[a]>7.08e+02) eta.n[a]<-7.08e+02}
pi.n<-exp(eta.n)/(1+exp(eta.n)) # probabilities

#set.seed(56479)
B<-500
Y<-matrix(rbinom(n*B,mi,pi.n),n,B)
like<-sapply(1:B,function(j)
{
  yi<-Y[,j]

  mle<-sapply(1:(length(grid2)),function(i)  # finding beta which maximizes
                                                    the likelihood
  {
    betavec<-grid2[[i]]
    eta.n<-cbind(rep(1,n),xs)%*%betavec
      for (a in 1:length(eta.n)) {if (eta.n[a]>7.08e+02) eta.n[a]<-7.08e+02}
    pi.n<-exp(eta.n)/(1+exp(eta.n)) # probabilities
    mu.n<-mi*pi.n
    sigma2.n<-mi*pi.n*(1-pi.n)
    V<-diag(as.vector(mi*pi.n*(1-pi.n)))
    betai<-basis(ppo(Z,V)-ppo(ppo(Z,V)%*%tm(xs,V),V))
```

```
v<- betai/sigma2.n
logl.n<- sum(yi*eta.n-mi*log(1+exp(eta.n)))

v.pos<-numeric(n); v.neg<-numeric(n); v.ze<-numeric(n)
for (i in 1:length(v))
{
  if (v[i]>0) { v.pos[i]<-v[i] ; v.neg[i]<-0; v.ze[i]<-0 }
  if (v[i]<0) { v.pos[i]<-0 ; v.neg[i]<-v[i]; v.ze[i]<-0 }
  if (v[i]==0) { v.pos[i]<-0 ; v.neg[i]<-0; v.ze[i]<-v[i]}
}
fs.pl<-sum((yi-mi)*v.pos)+sum(yi*v.neg)+sum((yi-mi*pi.n)*v.ze)  # positive
                                                        infinity limit of f(s)
fs.nl<- sum(yi*v.pos)+sum((yi-mi)*v.neg)+sum((yi-mi*pi.n)*v.ze) # negative
                                                        infinity limit of f(s)

if (fs.pl*fs.nl>0) print(grid2[[i]])  # non convergence points
else
{
  s1<--100; s2<-100
  eta.f1<- eta.n + s1*v
    for (a in 1:length(eta.f1)) {if (eta.f1[a]>7.08e+02) eta.f1[a]<-7.08e+02}
  eta.f2<- eta.n + s2*v
    for (a in 1:length(eta.f2)) {if (eta.f2[a]>7.08e+02) eta.f2[a]<-7.08e+02}
  pi.f1<-exp(eta.f1)/(1+exp(eta.f1))
  pi.f2<-exp(eta.f2)/(1+exp(eta.f2))
  fs1<-sum((yi-mi*pi.f1)*v)
  fs2<-sum((yi-mi*pi.f2)*v)
```

```
if (fs1<0)                  # finding a positive f(s1)
{
  while (fs1<0)
  {
    s1<-s1-1
    eta.f1<- eta.n + s1*v
       for (a in 1:length(eta.f1)) {if (eta.f1[a]>7.08e+02) eta.f1[a]
                                                    <-7.08e+02}
    pi.f1<-exp(eta.f1)/(1+exp(eta.f1))
    fs1<-sum((yi-mi*pi.f1)*v)
  }
}
if (fs2>0)                   # finding a negative f(s2)
{
  while (fs2>0)
  {
    s2<-s2+1
    eta.f2<- eta.n + s2*v
     for (a in 1:length(eta.f2)) {if (eta.f2[a]>7.08e+02) eta.f2[a]
                                                    <-7.08e+02}
    pi.f2<-exp(eta.f2)/(1+exp(eta.f2))
    fs2<-sum((yi-mi*pi.f2)*v)
  }
}
while (fs1>1e-5 & fs2<(-1e-5))
{
  s1.new<-(s1+s2)/2
```

```
    eta.f1<- eta.n + s1.new*v
      for (a in 1:length(eta.f1)) {if (eta.f1[a]>7.08e+02) eta.f1[a]
                                                  <-7.08e+02}
    pi.f1<-exp(eta.f1)/(1+exp(eta.f1))
    fs1<-sum((yi-mi*pi.f1)*v)
    if (fs1>0)
    {
      s1<-s1.new
    }
    if (fs1<0)
    {
      s2<-s1.new
      eta.f2<- eta.n + s2*v
       for (a in 1:length(eta.f2)) {if (eta.f2[a]>7.08e+02) eta.f2[a]
                                                    <-7.08e+02}
      pi.f2<-exp(eta.f2)/(1+exp(eta.f2))
      fs2<-sum((yi-mi*pi.f2)*v)
        eta.f1<- eta.n + s1*v
          for (a in 1:length(eta.f1)) {if (eta.f1[a]>7.08e+02) eta.f1[a]
                                                  <-7.08e+02}
        pi.f1<-exp(eta.f1)/(1+exp(eta.f1))
        fs1<-sum((yi-mi*pi.f1)*v)
    }
  }
  logl.f<- sum(yi*eta.f1-mi*log(1+exp(eta.f1)))
}
cat(c(betavec,s1.new,logl.f,logl.n),file="result.txt","\n",append=T)
```

```
  })    # end of mle calculation function

  results<-as.matrix(read.table("result.txt")) # read the ourput

  file.remove("result.txt")
  maxl.f<-max(results[,4])
  maxl.n<-max(results[,5])

  posi.f <-match(maxl.f,results[,4])
  mle.f<-c(results[posi.f,1:3])
  mle.f

  posi.n<-match(maxl.n,results[,5])
  mle.n<-c(results[posi.n,1:2])
  mle.n

  cat(c(mle.f,mle.n,maxl.f,maxl.n),file="likelihood.txt","\n",append=T)
  cat(c(mle.f,mle.n,maxl.f,maxl.n),file="like.txt","\n",append=T)
}) # end of like calculation function

likelihood<-as.matrix(read.table("likelihood.txt")) # read the ourput
file.remove("likelihood.txt")

llr<--2*(likelihood[,7]-likelihood[,6])
power<-sum(ifelse(llr>=qchisq(0.95,1),1,0))/B
power
print(power)
```

```
}) # end of power calculation function

plot(J,powerf,type="b")
cbind(J,powerf)



# Part of the code used to calculate Wald-Shao Statistic

    s<-mle.f[3]; b0<-mle.f[1]; b1<-mle.f[2]

  eta.n<-cbind(rep(1,n),xs)%*%c(b0,b1)  # linear predictor for proposed model (null)
    for (a in 1:length(eta.n)) {if (eta.n[a]>7.08e+02) eta.n[a]<-7.08e+02}
  pi.n<-exp(eta.n)/(1+exp(eta.n)) # probabilities
  mu.n<-mi*pi.n
  sigma2.n<-mi*pi.n*(1-pi.n)
  V<-diag(as.vector(mi*pi.n*(1-pi.n)))
  betai<-basis(ppo(Z,V)-ppo(ppo(Z,V)%*%tm(xs,V),V))
  vi<- betai/sigma2.n

  logl<-expression(yi*(b0+b1*xs+s*vi)-mi*log(1+exp(b0+b1*xs+s*vi)))

  dls<- eval(D(logl,"s"))   # logl wrt s
  dlb0<-eval(D(logl,"b0"))  # logl wrt b0
  dlb1<-eval(D(logl,"b1"))  # logl wrt b1

  d2ls<- sum(eval(D(D(logl,"s"),"s")))  # logl wrt s & s
  d2lb0<-sum(eval(D(D(logl,"b0"),"b0"))) # logl wrt b0 & b0
```

```
  d2lb1<-sum(eval(D(D(logl,"b1"),"b1")))  # logl wrt b1 & b1
  d2lsb0<-sum(eval(D(D(logl,"s"),"b0"))) # logl wrt s & b0
  d2lsb1<-sum(eval(D(D(logl,"s"),"b1")))   # logl wrt s & b1
  d2lb0b1<-sum(eval(D(D(logl,"b0"),"b1")))  # logl wrt b0 & b1

    info<--matrix(c(d2lb0,d2lb0b1,d2lsb0,d2lb0b1,d2lb1,d2lsb1,d2lsb0,d2lsb1,
                d2ls),3,3,byrow=T)

    H.theta<-c(0,0,1)
    wald.stat<-s*ginv(t(H.theta)%*%ginv(info)%*%(H.theta))%*% s

# Part of the code used to calculate Wald-Boos Statistic

    s<-mle.f[3]; b0<-mle.f[1]; b1<-mle.f[2]

  eta.n<-cbind(rep(1,n),xs)%*%c(b0,b1)  # linear predictor for proposed model (null)
    for (a in 1:length(eta.n)) {if (eta.n[a]>7.08e+02) eta.n[a]<-7.08e+02}
  pi.n<-exp(eta.n)/(1+exp(eta.n)) # probabilities
  mu.n<-mi*pi.n
  sigma2.n<-mi*pi.n*(1-pi.n)
  V<-diag(as.vector(mi*pi.n*(1-pi.n)))
  betai<-basis(ppo(Z,V)-ppo(ppo(Z,V)%*%tm(xs,V),V))
  vi<- betai/sigma2.n


  logl<-expression(yi*(b0+b1*xs+s*vi)-mi*log(1+exp(b0+b1*xs+s*vi)))
```

```
  dls<- eval(D(logl,"s"))   # logl wrt s
  dlb0<-eval(D(logl,"b0"))  # logl wrt b0
  dlb1<-eval(D(logl,"b1"))  # logl wrt b1

  d2ls<- sum(eval(D(D(logl,"s"),"s")))  # logl wrt s & s
  d2lb0<-sum(eval(D(D(logl,"b0"),"b0"))) # logl wrt b0 & b0
  d2lb1<-sum(eval(D(D(logl,"b1"),"b1")))  # logl wrt b1 & b1
  d2lsb0<-sum(eval(D(D(logl,"s"),"b0"))) # logl wrt s & b0
  d2lsb1<-sum(eval(D(D(logl,"s"),"b1")))   # logl wrt s & b1
  d2lb0b1<-sum(eval(D(D(logl,"b0"),"b1")))  # logl wrt b0 & b1

    score<-cbind(dlb0,dlb1,dls)
    info<--matrix(c(d2lb0,d2lb0b1,d2lsb0,d2lb0b1,d2lb1,d2lsb1,d2lsb0,d2lsb1,
                d2ls),3,3,byrow=T)

    D<- t(score)%*%score
    v.hat<- (ginv(info))%*%D%*%(ginv(info))
    H.theta<-c(0,0,1)
    wald.stat<-s*ginv(t(H.theta)%*%v.hat%*%(H.theta))%*% s

# Part of the code used to calculate S Statistic

   b0<-mle.n[1]; b1<-mle.n[2]
   eta.n<-cbind(rep(1,n),xs)%*%c(b0,b1)  # linear predictor for proposed model (null)
   pi.n<-exp(eta.n)/(1+exp(eta.n)) # probabilities
   mu.n<-mi*pi.n
   sigma2.n<-mi*pi.n*(1-pi.n)
```

```
   sy<- sum((yi-mu.n)^2/sigma2.n)

# Part of the code used to calculate S-diff Statistic

   # calculating sy for full model
   s<-mle.f[3]; b0<-mle.f[1]; b1<-mle.f[2]
   eta.n<-cbind(rep(1,n),xs)%*%c(b0,b1)  # linear predictor for proposed model (null)
    for (a in 1:length(eta.n)) {if (eta.n[a]>7.08e+02) eta.n[a]<-7.08e+02}
   pi.n<-exp(eta.n)/(1+exp(eta.n)) # probabilities
   mu.n<-mi*pi.n
   sigma2.n<-mi*pi.n*(1-pi.n)
   V<-diag(as.vector(mi*pi.n*(1-pi.n)))
   betai<-basis(ppo(Z,V)-ppo(ppo(Z,V)%*%tm(xs,V),V))
   vi<- betai/sigma2.n
      eta.f<- eta.n + s*vi
        for (a in 1:length(eta.f)) {if (eta.f[a]>7.08e+02) eta.f[a]<-7.08e+02}
      pi.f<-exp(eta.f)/(1+exp(eta.f)) # probabilities
      mu.f<-mi*pi.f
      sigma2.f<-mi*pi.f*(1-pi.f)

   # calculating sy for null model
   b00<-mle.n[1]; b11<-mle.n[2]
   eta.nn<-cbind(rep(1,n),xs)%*%c(b00,b11)  # linear predictor for proposed model
     for (a in 1:length(eta.nn)) {if (eta.nn[a]>7.08e+02) eta.nn[a]<-7.08e+02}
   pi.nn<-exp(eta.nn)/(1+exp(eta.nn)) # probabilities
   mu.nn<-mi*pi.nn
   sigma2.nn<-mi*pi.nn*(1-pi.nn)
```

```
    sy.f<- sum((yi-mu.f)^2/sigma2.f)
    sy.n<- sum((yi-mu.nn)^2/sigma2.nn)

# Part of the program to calculate the power of HL test

  set.seed(357)
  B<-500
  Y<-matrix(rbinom(n*B,mi,pi.f),n,B)
  like<-sapply(1:B,function(j)
  {
    yi<-Y[,j]
      y<-cbind(yi,(mi-yi))
      mod<- glm(y~xs,family="binomial")
      fit<-mi*fitted(mod)
      out<-hosmerlem(y<-yi, yhat<-fit,10)
      cat(c(out),file="likelihood.txt","\n",append=T)
})
  likelihood<-as.matrix(read.table("likelihood.txt")) # read the ourput
  file.remove("likelihood.txt")
  p<-likelihood[,2]

  powerhl<-sum(ifelse(p<0.05,1,0))/B

# Part of the program to calculate the power of Deviance test

set.seed(357)
```

```
B<-500
Y<-matrix(rbinom(n*B,mi,pi.f),n,B)
like<-sapply(1:B,function(j)
{
  yi<-Y[,j]
    y<-cbind(yi,(mi-yi))
    mod<- glm(y~xs,family="binomial")
    anov<-anova(mod,test="Chisq")
    res.dev<-mod$deviance
    p<-1-pchisq(res.dev,mod$df.residual)


  cat(c(res.dev,p),file="likelihood.txt","\n",append=T)
 # cat(c(dev,p.val,res.dev,p),file="like.txt","\n",append=T)
}) # end of like calculation function

likelihood<-as.matrix(read.table("likelihood.txt")) # read the ourput
file.remove("likelihood.txt")
p<-likelihood[,2]

powerd<-sum(ifelse(p<0.05,1,0))/B
```

Appendix E

R Code for Implementing the Power Calculations for Poisson Model

```
source("functions.r")


n<-50 # number of samples (dial settings)
k<--3
l<-3
set.seed(2250)
x<-runif(n,k,l) # x vales from Uniform[0,1]
xs<-sort(x)  # sorted x values


grid2<-gridp(1,12,k,l,1)


Z<-matrix(rep(c(1,0,1,0,1),c(17,n,23,n,10)),n,3)    #  maximin[[156]]
betav<-c(5,-0.5)
s.vec<-seq(0,10,by=2)
powerf<-sapply(1:length(s.vec),function(k)
```

```
{
  s<-s.vec[k]

  # proposed (null) model
  eta.n<-cbind(rep(1,n),xs)%*%betav  #linear predictor for proposed model
  mu.n<-exp(eta.n) # mean
  V<-diag(as.vector(mu.n))

  # altrenative (full) model
  betai<-sqrt(n)*(basis(ppo(Z,V)-ppo(ppo(Z,V)%*%tm(xs,V),V)))
  v<- betai/mu.n
  eta.f<- eta.n + s*v
  mu.f<-exp(eta.f)

 # set.seed(6745)
  B<-500
  Y<-matrix(rpois(n*B,mu.f),n,B)
  like<-sapply(1:B,function(j)
  {
    yi<-Y[,j]

    mle<-sapply(1:(length(grid2)),function(i)  #finding beta which maximizes
                                                    the likelihood
    {
      betavec<- grid2[[i]]
      eta.n<-cbind(rep(1,n),xs)%*%betav  #linear predictor for proposed model
      mu.n<-exp(eta.n) # mean
```

```
V<-diag(as.vector(mu.n))
betai<-sqrt(n)*(basis(ppo(Z,V)-ppo(ppo(Z,V)%*%tm(xs,V),V)))
v<- betai/mu.n
logl.n<- sum(yi*eta.n-mu.n)



  s1<--1e+2; s2<-1e+2
  eta.f1<- eta.n + s1*v
     for (a in 1:length(eta.f1)) {if (eta.f1[a]>7.08e+02) eta.f1[a]
                                        <-7.08e+02}
  eta.f2<- eta.n + s2*v
    for (a in 1:length(eta.f2)) {if (eta.f2[a]>7.08e+02) eta.f2[a]
                                        <-7.08e+02}
  mu.f1<-exp(eta.f1)
  mu.f2<-exp(eta.f2)
  fs1<-sum((yi-mu.f1)*v)
  fs2<-sum((yi-mu.f2)*v)
  if (fs1<0)                 # finding a positive f(s1)
  {
    while (fs1<0)
    {
      s1<-s1-1
      eta.f1<- eta.n + s1*v
         for (a in 1:length(eta.f1)) {if (eta.f1[a]>7.08e+02)
                                             eta.f1[a]<-7.08e+02}
      mu.f1<-exp(eta.f1)
      fs1<-sum((yi-mu.f1)*v)
```

```
  }
}
if (fs2>0)                    # finding a negative f(s2)
{
  while (fs2>0)
  {
    s2<-s2+1
    eta.f2<- eta.n + s2*v
     for (a in 1:length(eta.f2)) {if (eta.f2[a]>7.08e+02)
                                          eta.f2[a]<-7.08e+02}
    mu.f2<-exp(eta.f2)
    fs2<-sum((yi-mu.f2)*v)
  }
}
s1.new<-(s1+s2)/2
while (fs1>1e-5 & fs2<(-1e-5))
{
  s1.new<-(s1+s2)/2
  eta.f1<- eta.n + s1.new*v
    for (a in 1:length(eta.f1)) {if (eta.f1[a]>7.08e+02)
                                            eta.f1[a]<-7.08e+02}
  mu.f1<-exp(eta.f1)
  fs1<-sum((yi-mu.f1)*v)
  if (fs1>0)
  {
    s1<-s1.new
  }
```

```
      if (fs1<0)
      {
        s2<-s1.new
        eta.f2<- eta.n + s2*v
         for (a in 1:length(eta.f2)) {if (eta.f2[a]>7.08e+02)
                                                  eta.f2[a]<-7.08e+02}
        mu.f2<-exp(eta.f2)
        fs2<-sum((yi-mu.f2)*v)
          eta.f1<- eta.n + s1*v
          mu.f1<-exp(eta.f1)
          fs1<-sum((yi-mu.f1)*v)
      }
    }
    eta.f<- eta.n + s1.new*v
    logl.f<- sum(yi*eta.f-exp(eta.f))
  cat(c(betavec,s1.new,logl.f,logl.n),file="result.txt","\n",append=T)
})    # end of mle calculation function

results<-as.matrix(read.table("result.txt")) # read the ourput

file.remove("result.txt")
maxl.f<-max(results[,4])
maxl.n<-max(results[,5])

posi.f <-match(maxl.f,results[,4])
mle.f<-c(results[posi.f,1:3])
mle.f
```

```
    posi.n<-match(maxl.n,results[,5])
    mle.n<-c(results[posi.n,1:2])
    mle.n

    cat(c(mle.f,mle.n,maxl.f,maxl.n),file="likelihood.txt","\n",append=T)

  }) # end of like calculation function

  likelihood<-as.matrix(read.table("likelihood.txt")) # read the ourput
  file.remove("likelihood.txt")

  llr<--2*(likelihood[,7]-likelihood[,6])
  power<-sum(ifelse(llr>=qchisq(0.95,1),1,0))/B
  power
  print(power)
}) # end of power calculation function

cbind(s.vec,powerf)
plot(s.vec,powerf,type="b")
```