2014-01-01

Resampling-Based Multiple Comparisons For Generalized Linear Models

Josephine Sarpong Akosa
University of Texas at El Paso, jsakosa@gmail.com

RESAMPLING-BASED MULTIPLE COMPARISONS FOR GENERALIZED LINEAR MODELS

JOSEPHINE SARPONG AKOSA

Department of Mathematical Sciences

APPROVED:

Amy E. Wagler, Chair, Ph.D.

Joan G. Staniswalis, Ph.D.

Thompson Sarkodie-Gyan, Ph.D.

Charles H. Ambler, Ph.D.
Dean of the Graduate School

to the

AKOSA FAMILY

with love

RESAMPLING-BASED MULTIPLE COMPARISONS FOR GENERALIZED LINEAR MODELS

by

JOSEPHINE SARPONG AKOSA

THESIS

Presented to the Faculty of the Graduate School of

The University of Texas at El Paso

in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE

Department of Mathematical Sciences

THE UNIVERSITY OF TEXAS AT EL PASO

AUGUST 2014

Acknowledgements

I would first and foremost like to express my heart-felt gratitude to the Almighty God for His guidance, strength, knowledge and wisdom to undertake this study. My profound gratitude goes to my advisor, Dr. Amy E. Wagler, who in spite of her busy schedule was always available to assist and offer her help, support and encouragement. With her guidance, patience and advice, I have come this far. It has really been an honor and a blessing to have had her as an advisor.

I wish to also thank the other members of my committee: Dr. Joan G. Staniswalis from the Department of Mathematical Sciences and Dr. Thompson Sarkodie-Gyan from the Department of Electrical and Computer Engineering, both with the University of Texas at El Paso. Your immense help, time, constructive comments and support are very much appreciated.

I would further like to thank the Chair, Dr. Maria Christina Mariani and all the professors of the Department of Mathematical Sciences especially Dr. Panagis Moschopoulos, Dr. Ori Rosen, Dr. Xiaogang Su and Dr. Behzad Djafari-Rouhani for shaping me into the scholar I am now. Much appreciation also goes to the staff of the Department of Mathematical Sciences particularly, Maria Salayandia and Maria Barraza-Rios, I am grateful for your love and support.

Finally, to my dearest parents, family and friends, both in El Paso and Ghana, I say thank you. Your prayers, support and encouragement has made this study a reality. I am indeed grateful. "Nyame nhyira mo".

Abstract

Diverse applications in medical and epidemiological research routinely utilize generalized linear modeling to explain the relationship between the incidence of disease and particular risk factors. Researchers' interest in such models are estimated quantities from the model such as the response probabilities, the relative risks or the odds ratios and not the model itself. Often, the simultaneous estimation of these quantities or a subset of the quantities are warranted. The results are usually reported via confidence intervals at a pre-specified level of significance. Utilizing the usual 95% pointwise confidence intervals for the simultaneous inference inflates the risk of making type I errors. Several procedures have been proposed to control for multiplicity among a set of inferences, but no one solution is applicable for all situations.

This study develops and via simulation evaluate resampling-based multiple comparison procedures for contrasts of generalized linear model parameters. The main results of the study is a new algorithm for resampling-based multiple comparisons for quantities from generalized linear models that will in some situations outperform existing conservative procedures and in those situations be less data-wasteful.

Table of Contents

Appendix

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background and Motivation

Generalized linear models were formulated in the early 1970's as a way of unifying various statistical models including the linear regression model, the Poisson regression and the logistic regression models (Nelder and Wedderburn, 1972). These models have been found to be of substantial value to statisticians in diverse applications such as epidemiological and medical research. Generalized linear models can be useful in understanding the kind of behaviors or traits that influence the incidence of a particular disease or characteristic. For instance, logistic regression models can be employed to understand the relationship between the incidence of a disease and a set of possible risk factors while the log-linear models can be utilized to understand associations between a disease and predictor variables.

Normally, after building a generalized linear model (GLM), the primary interest centers on some estimated quantities from the model such as the response probabilities, the odds ratios or the relative risks and not on the model itself. Simultaneous estimation of a function of these quantities or a subset of these quantities are often warranted. The results are usually reported via confidence intervals or confidence bounds at a pre-specified level of significance for each inference. When the overall set of conclusions about the function of the estimated quantities are not warranted, then the use of one-at-a-time intervals is appropriate. More often than not, researchers' interest are rarely in the sets of conclusions involving a single inference. Researchers may particularly be interested in making inference about a subset of the quantities that lower or raise the risk of a particular outcome significantly. Thus, the overall sets of conclusions or a subset of conclusions are often warranted

leading to the problem of multiplicity. In such cases, if multiplicity is not accounted for, then the probability of concluding false non-coverage of the true function of the estimated quantities by the estimated intervals may be unduly large. For illustration, consider an eight-center clinical trail to compare a drug to placebo for curing a disease. The response variable in this case is binary, indicating whether the drug successfully cures the disease. Thus, a GLM can be employed for this analysis. The response variable in this example can be predicted using reference-coded explanatory variable specifying the center of the subject and the assigned treatment group. Here, every slope parameter is the log odds ratio for that level of the explanatory variable with respect to the reference variable. Suppose a researcher is interested in comparing the odds ratio for subjects in a specific treatment group across the centers. Confidence intervals for these quantities could be estimated using confidence intervals for linear combinations of the slope parameters. This will result in $\binom{8}{2}$ = 56 number of comparisons. Reasonable larger sets of simultaneous inference can even be specified for this example. Thus, the number of those comparisons that are simultaneously of interest to the researcher can become excessively large. In the aforementioned example, if the usual 95% confidence intervals are used for comparing the odds ratios, then the overall error rate for the estimation could be as high as 94%. It can easily be seen that with $\binom{k}{2}$ pairwise comparisons applied separately each at level α, the probability of concluding any false pairwise significance can be well in excess of α, which will be close to 1 for sufficiently large k.

Several procedures have been proposed in literature to deal with the problems associated with multiple comparisons in the linear model settings. Among them are the Bonferroni adjustments, the Šidák correction (1967) and the Kounias improvement (1968). However, these procedures are probability-based; making use of the mathematical properties of the hypotheses structure without taking into account the correlation structure of the test statistic. For highly correlated and non-normal test statistic, these methods could be very conservative, less powerful and waste data. A more powerful but computationally intensive procedure is the method of Hunter (1976) and Worsley (1982). This method takes into

account the correlation structure of the set of hypotheses. Though more powerful than the other probability-based methods, the Hunter-Worsley method is generally conservative.

Recent developments in this field include resampling-based procedures (Westfall, 2011, Westfall and Young, 1993). Resampling-based procedures are more flexible and allow for correlated inferences and high-dimensional comparisons. Despite the increased study on multiple comparison procedures in linear model settings, research on multiple comparison procedures for quantities of generalized linear models have received little attention aside the usual Bonferroni adjustments to confidence intervals.

Thus, the goal of this study is to develop and via simulation evaluate resampling-based multiple comparison procedures for contrasts of GLM parameters. The general attention of the study is focused on estimated quantities from a GLM, not the estimation of the model itself. Nevertheless, the performance of any of the multiple comparison procedures is greatly influenced by how well the estimated model fits the data. However, this study will assume the model fits the data well. Furthermore, before conducting any inferences on the parameters of interest, the parameters must first be estimated using a reliable estimation procedure. Maximum likelihood estimation is often employed. But due to the fact that these estimates may be biased at small sample sizes, it is sometimes not the most appropriate choice. Therefore, utilizing a procedure that corrects this bias is optimal. Thus, the study will additionally evaluate the influence of the estimation method on the multiple comparison procedure. With this established, the goal of the study is to develop and evaluate via simulation a less conservative resampling-based multiple comparison procedure for contrasts of generalized linear model parameters.

In the following chapters, we present an overview of the generalized linear models and simultaneous inference of contrasts of generalized linear model parameters. We also outline both the existing and our proposed method for multiple comparisons in generalized linear models. The simulation results of the proposed methods are presented and analyzed. Precisely, chapter two details the generalized linear models and quantities that are mostly of interest in such models. Estimation procedures for the generalized linear model pa-

rameters are additionally discussed. Further, existing methodologies used for simultaneous estimation of various functions resulting from generalized linear models are discussed in the chapter. In chapter three, we discuss resampling methods for generalized linear models and present an outline for our proposed multiple comparison procedure. Simulations are used to evaluate and compare the performance of the proposed method to existing conservative methods in chapter four. The method is applied to data from the 2009 National Health Interview Survey to analyze the impact of region and hayfever allergy status on the incidence of asthma-related emergency room (ER) visits for children in the United States. Finally, we discuss the results of the simulation study and provide some areas for future research in the concluding chapter.

Chapter 2

Literature Review

2.1 Generalized Linear Models

Statistical models describe the relationship between a set of random variables using mathematical equations. These random variables are usually classified as response (dependent) and explanatory (independent) variables. Examples of such statistical models include linear models, generalized linear models, mixed effects models and survival models. The use of a model for statistical inference generally depends on the characteristics of the response and the explanatory variables and how best the model can provide a reasonable approximation to reality. Explanatory variables can be quantitative or qualitative. Response variables on the other hand can be continuous, discrete (in the form of counts) or categorical. In cases where the response variable is categorical or discrete, the generalized linear models are often employed for the statistical inference.

Ordinary linear models are the most commonly used for modeling continuous response variables. These models are based on the assumptions of normality and homoscedasticity of the error terms. However, in most applications, these assumptions are violated. Nelder and Wedderburn (1972) therefore formulated the generalized linear models as a way of unifying various statistical models and allowing for flexibility in these assumptions.

Generalized linear models (GLMs) comprise of three components: random component, systematic component and link function. The random component consists of a response variable with independent observations from a distribution in the exponential class family. The systematic component specifies explanatory variables used in a linear predictor function. The linear predictor includes information about the explanatory variables into the

model. Lastly, the link function identifies a function of the mean response that relates the random and systematic components.

In general, a GLM can be expressed as:

$$\boldsymbol{\eta}_i = g\left[E\left(\mathbf{Y}_i|\mathbf{X}_i\right)\right] = \mathbf{X}_i\boldsymbol{\beta}; \quad i = 1, \ldots, n \tag{2.1}$$

where $\mathbf{X}_i$ is a vector of covariates corresponding to $\mathbf{Y}_i$, the response variables, $\boldsymbol{\beta}$ is a $p \times 1$ vector of regression parameters and $g(.)$ is a function which links the expected response $\boldsymbol{\mu}_i = E\left(\mathbf{Y}_i|\mathbf{X}_i\right)$ to $\boldsymbol{\eta}_i$. The link function must be monotone and differentiable. The choice of link function is dependent on the type of data. The link function is chosen such that its domain always matches the range of the distribution of the mean response. For instance, an identity link function is appropriate for continuous normal outcome. The link function should restrict the mean response to positive numbers and to the interval $[0, 1]$ for count outcomes and outcomes in the form of proportion respectively.

2.1.1 Estimation of the GLM parameters

Consider the generalized linear model

$$\boldsymbol{\eta}_i = \mathbf{x}_i'\boldsymbol{\beta}, \quad i = 1, \cdots, n; \tag{2.2}$$

where $\boldsymbol{\mu}_i = E(\mathbf{Y}_i \mid \mathbf{x}_i)$ is the expected response and $\mathbf{x}_i$ is an observed vector of covariates corresponding to the response variable $\mathbf{Y}_i$. Let $\widehat{\boldsymbol{\beta}} = \left(\widehat{\beta}_1, \widehat{\beta}_2, \cdots \widehat{\beta}_k\right)$ denote the maximum likelihood estimate (MLE) of $\boldsymbol{\beta}$. The MLE is asymptotically multivariate normal with mean vector $\boldsymbol{\beta}$ and covariance matrix $\mathbf{V} = \operatorname{cov}\left(\widehat{\boldsymbol{\beta}}\right) = \left(\mathbf{X}^T\mathbf{W}\mathbf{X}\right)^{-1}$, where $\mathbf{W}$ is the diagonal matrix with main-diagonal elements;

$$w_i = \frac{1}{\operatorname{Var}(Y_i)}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2. \tag{2.3}$$

This covariance matrix is estimated by $\widehat{\boldsymbol{V}} = \left(\mathbf{X}^T\widehat{\mathbf{W}}\mathbf{X}\right)^{-1}$, where $\widehat{\boldsymbol{W}}$ is $\boldsymbol{W}$ evaluated at $\widehat{\boldsymbol{\beta}}$.

In GLMs, the MLE is typically utilized for the estimation of the unknown parameters. Conversely, it is well known that the MLEs may be biased when the sample size or the

total Fisher information is small. Additionally, maximum likelihood estimation can yield infinite parameter estimates due to separability or quasi-separability in the data. Therefore, as an alternative to the MLE, the penalized maximum likelihood estimator (pMLE) was developed by Firth (1993) for the parameter estimations. The pMLE has an advantage of better approximating the true sampling distribution of the observations. The reason being that the pMLE "fixes" the asymptotic bias in the estimation *a priori* while the MLE requires the sample size to be very large to reduce the asymptotic bias. The pMLE also eliminates inestimable parameters due to separability or quasi-separability in the data. The pMLE bias correction has the effect of shrinking the MLE towards the origin.

2.1.2 The Penalized Maximum Likelihood Estimate

The asymptotic bias of the MLE, $\widehat{\boldsymbol{\beta}}$ in a regular model with p-dimensional parameter $\boldsymbol{\beta}$ can be written in the Taylor series expansion:

$$b(\boldsymbol{\beta}) = \frac{b_1(\boldsymbol{\beta})}{n} + \frac{b_2(\boldsymbol{\beta})}{n^2} + \cdots, \tag{2.4}$$

where n is the number of observations. Methods for bias-reduction in the MLE have been extensively studied in literature. Firth (1993) developed the pMLE as an alternative to the MLE. The pMLE procedure involves the introduction of an appropriate bias term into the score function. If the parameter of interest is the canonical parameter of an exponential class family, then this penalization is achieved via Jeffrey's invariant prior. This penalty yields estimates of the regression parameters that are unbiased in the first order. Normally, the maximum likelihood estimate is derived as a solution to the score function, $\nabla l(\boldsymbol{\beta}) = U(\boldsymbol{\beta}) = 0$, where $l(\boldsymbol{\beta}) = \log L(\boldsymbol{\beta})$ is the log likelihood function and $U(\boldsymbol{\beta}) = (U_1(\boldsymbol{\beta}), \cdots, U_k(\boldsymbol{\beta}))'$ is the usual vector of score functions for estimating $\boldsymbol{\beta}$. For convenient, the notation of Firth (1993) will be employed. Let

$$U_r(\boldsymbol{\beta}) = \frac{\partial l}{\partial \beta^r}, \qquad U_{rs}(\boldsymbol{\beta}) = \frac{\partial^2 l}{\partial \beta^r \partial \beta^s}, \tag{2.5}$$

where $\boldsymbol{\beta} = \left(\beta^1, \beta^2, \cdots, \beta^k\right)$ is the parameter vector. The joint null cummulants are $k_{r,s} = n^{-1}E\{U_r U_s\}$, $k_{r,s,t} = n^{-1}E\{U_r U_s U_t\}$ and $k_{r,st} = n^{-1}E\{U_r U_{st}\}$ for a p-dimensional vector

$\boldsymbol{\beta} = (\beta_1, \beta_2, \cdots, \beta_k)$ (McCullagh and Nelder, 1989). The first order bias of $\widehat{\boldsymbol{\beta}}_r$, the r^{th} parameter of a GLM is given by

$$b^r(\boldsymbol{\beta}) = \frac{-k^{r,s}k^{t,u}(k_{s,t,u} + k_{s,tu})}{2} \tag{2.6}$$

In matrix notation, the first order bias, $O(n^{-1})$ of $\widehat{\boldsymbol{\beta}}$ for a GLM with a canonical link reduces to the simple form $\mathbf{b} = \left(\mathbf{X}^T\widehat{\mathbf{W}}\mathbf{X}\right)^{-1}\mathbf{X}^T\boldsymbol{W\xi}$ where $\mathbf{W} = \text{cov}(\mathbf{Y})$; $\mathbf{Q} = \mathbf{X}\left(\mathbf{X}^T\widehat{\mathbf{W}}\mathbf{X}\right)^{-1}\mathbf{X}^T$ is the asymptotic covariance matrix of $\widehat{\eta}$, $\xi_i = -\frac{1}{2}\left(\frac{k_{3i}}{k_{2i}}\right)Q_{ii}$. The term applied to the log likelihood function for the bias-reduction is of the form $\frac{1}{2}\log|i(\boldsymbol{\beta})|$ where $i(\boldsymbol{\beta}) = \mathbf{X}^T\widehat{\mathbf{W}}\mathbf{X}$ is the Fisher information matrix and $|i(\boldsymbol{\beta})|^{\frac{1}{2}}$ is Jeffrey's invariant prior. Thus, the first order bias is removed by calculation of the posterior mode based on this prior. The pMLE is thus derived as a solution to the shifted score function; $U^*(\boldsymbol{\beta}) = U(\boldsymbol{\beta}) - i(\boldsymbol{\beta})b(\boldsymbol{\beta})$ set equal to 0. For a logistic regression with success probability $\gamma_i = \dfrac{\exp(\mathbf{X}_i\boldsymbol{\beta})}{1+\exp(\mathbf{X}_i\boldsymbol{\beta})}$, $(i = 1, 2, \ldots, n)$, maximum likelihood estimates are known to be biased away from the point $\beta = 0$. At this point, the determinant $|i(\beta)|$ is maximized, that is at $\gamma_i = 0.5$. The penalty applied to the log likelihood function therefore shrinks the parameter estimates towards $\beta = 0$.

The pMLE also has asymptotically multivariate normal with mean vector $\boldsymbol{\beta}$ and covariance matrix $\left(\mathbf{X}^T\mathbf{W}\mathbf{X}\right)^{-1}$, where the weight function is dependent on the estimate. Further details on this maximum likelihood estimation are discussed in Firth (1993).

2.2 The Multiple Comparison Problem

Multiplicity of data, hypotheses and analyses is a common problem in biomedical and epidemiological studies. Even so, there seems to be a lack of knowledge about statistical procedures for multiple testing or comparisons. The terms "multiple testing" and "multiple comparisons" are often used interchangeably to refer to the testing of more than one hypotheses. However, a principal distinction is that in multiple comparisons, inferences are based on confidence intervals whereas inferences are based on tests of hypotheses in multiple testing.

Generally, when testing a set of hypotheses, a first attempt may be to test each hypotheses separately at a pre-specified significance level α (an acceptable maximum probability of rejecting the null hypothesis when it is true, thus committing a type I error). Hence, as the number of hypotheses increases, the global type I error also increases. This may have serious consequences if the set of conclusions must be evaluated as a whole. Several procedures have been proposed to account for this problem but no one solution will be acceptable for all situations.

2.2.1 Type I Error Control

Let H_0 refers to a particular hypothesis and H_a refers to the alternative hypothesis. There are many such pairwise null/alternative hypotheses to be evaluated in multiple comparison applications. Let k_0 be the number of true hypotheses, R, the number of rejected hypotheses and V, the number of type I errors (false positives). There are different ways of defining the global type I error rates. Table 2.1 summarizes some of the ways of defining the different types of errors. The per-comparison error rate (PCER) is defined for each hypothesis as the

Table 2.1: Defining the different types of errors

	Decision		
	Null not rejected	Null rejected	Total
Null true	U	V	k_0
Null false	V	S	$k - k_0$
	$k - R$	R	k

the expected proportion of Type I errors. Thus, PCER $= \dfrac{E(V)}{k}$. Hence, it is evident that multiple α-level tests control the PCER at level α, but do not control the overall probability of erroneous significance conclusions at level α. As a result, an appropriate error rate for multiple testing settings is the familywise error rate (FWER); FWER $= P(V > 0)$. This is the probability that one or more values will be greater than α in a given family of inferences. The term "family" in this context refers to the whole set of hypotheses that

should be evaluated to achieve the goal of a study. In our settings, family refers to all pairwise comparisons of the regression parameters. Westfall and Young (1993) defined two kinds of FWER's: "FWEC, calculated under the complete null hypothesis (meaning that all H_i are true); and FWEP, computed under the partial null hypothesis (meaning that some subcollection of nulls, say $H_{j_1}, \cdots, H_{j_t}$, is true)". Mathematically,

$$\text{FWEC} = P(\text{reject at least one } H_i \mid \text{all } H_i \text{ are true}) \tag{2.7a}$$

$$\text{FWEP} = P(\text{reject at least one } H_i, i = j_1, \cdots, j_t \mid H_{j_1}, \cdots, H_{j_t} \text{ are true}) \tag{2.7b}$$

If a simultaneous test procedure produces FWEC $\leq \alpha$, then the procedure is said to control the FWER in the weak sense. Conversely, the FWER is controlled in the strong sense if FWEP $\leq \alpha$, regardless of which subject $j_1, \cdots j_t$ of hypotheses happens to be true, (Hochberg and Tamhane, 1987). Another type of the error rate is the per family error rate (PFER) defined as the expected number of false rejections in the family; PFER $= E(V)$. This error rate does not apply if the family size is infinite. A fourth type of error rate is the false discovery rate (FDR). This is defined as the expected proportion of type I errors among the rejected hypotheses; FDR $= E\left(\frac{V}{R} \mid R > 0\right) P(R > 0)$. Methods for dealing with multiple comparisons involve adjusting α such that the probability of making at least one false positive remains below the desired significance level. To control this problem, classical multiple comparison procedures aim to control the probability of committing any type I error in families of comparisons under simultaneous consideration. The study would therefore aim to control the familywise error rate, FWER in the strong sense.

2.3 Multiple Comparisons in GLMs

As previously indicated, the generalized linear models can be utilized to investigate the incidence of diseases and how the incidence is affected by factors such as age, gender, exposure to pollutants and any treatment procedures under study. The slope parameters or the regression parameters in such models are the odds ratios or relative risks of the explanatory

variables (such as age, gender) with respect to one another. Inferences for the odds ratio comparing subjects with particular risk factors for a particular response outcome are often desired. These comparisons result in a linear combination of the regression parameters. Mathematically, this linear combination can be expressed as $\boldsymbol{\theta} := \boldsymbol{C\beta}$ where $\boldsymbol{C} \in \mathbb{R}^{k,p}$ is a $k \times p$ matrix of constants. Suppose $\widehat{\boldsymbol{\beta}} \in \mathbb{R}^{p,1}$ is an estimate of $\boldsymbol{\beta}$ and $\boldsymbol{S} \in \mathbb{R}^{p,p}$ is a $p \times p$ matrix of covariance estimate of $\widehat{\boldsymbol{\beta}}$, $\text{cov}(\widehat{\boldsymbol{\beta}})$. Then, as stated earlier, $\widehat{\boldsymbol{\beta}}$ has an asymptotic multivariate normal distribution with mean vector $\boldsymbol{\beta}$ and covariance matrix $\boldsymbol{S}$. If $\widehat{\boldsymbol{\theta}}$ is an estimate of $\boldsymbol{\theta}$ such that $\widehat{\boldsymbol{\theta}} := \boldsymbol{C}\widehat{\boldsymbol{\beta}}$, then $\widehat{\boldsymbol{\theta}}$ also has asymptotic multivariate normal distribution with mean vector $\boldsymbol{\theta}$ and covariance matrix $\boldsymbol{S}^* := \boldsymbol{CSC}^T$.

Now, consider the global null hypotheses:

$$H_0 : \boldsymbol{\theta} = \boldsymbol{C\beta} = \boldsymbol{\theta_0}; \quad \boldsymbol{\theta_0} \in \mathbb{R}^{k,p}. \tag{2.8}$$

Under H_0, it follows that the k-dimensional test statistic

$$\boldsymbol{T} = \boldsymbol{D}^{-1/2}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta_0}\right) \overset{a}{\sim} MVN_k\left(\mathbf{0}, \mathbf{R}\right), \tag{2.9}$$

where $\boldsymbol{D} = \text{diag}(\boldsymbol{S}^*)$ and $\boldsymbol{R} \in \mathbb{R}^{k,k}$ is a $k \times k$ correlation matrix of the test statistic $\boldsymbol{T}$ defined as $\boldsymbol{R} = \boldsymbol{D}^{-1/2}\boldsymbol{S}^*\boldsymbol{D}^{-1/2}$. The set of global null hypotheses can be tested using standard global tests such as the F or the χ^2 test. However, such a test leading to the rejection of the null hypotheses gives no further indication about the nature of significance of the individual test statistics. Researchers on the other hand are rarely interested in the global null hypotheses. Thus, alternative test that includes the nature of significance of the individual tests are normally warranted. Researchers therefore seek procedures that test simultaneously the individual null hypotheses while maintaining the familywise error rate at the pre-specified significant level α.

For the simultaneous evaluation of the individual H_0 multiple contrast tests, $T = \max\{T_1, T_2, \ldots, T_k\}$ are often considered. This test leads to a *max-t* type of test statistic whose distribution under H_0 can be handled through the k-dimensional integral

$$P\left(\max_{i=1,\ldots,k} |T_i| < c_\alpha\right) \cong \int_{-\infty}^{c}\int_{-\infty}^{c}\cdots\int_{-\infty}^{c} \Phi_k\left(x_1, \cdots, x_k; \boldsymbol{R}, \nu\right) dx_1 dx_2 \cdots dx_k, \tag{2.10}$$

for some $c_\alpha \in \mathbb{R}$, where "Φ_k is the density function of either the limiting k-dimensional multivariate normal (with $\nu = \infty$ and the '$\approx$' operator) or the exact multivariate $c_k(\nu, \boldsymbol{R})$-distribution (with $\nu < \infty$ and the '=' operator)" (Hothorn et al., 2008). Unlike the global F and χ^2 tests, a test based on the maximum of the individual test statistics provides information on the nature of significance of the individual null hypotheses as well as the simultaneous confidence intervals.

In order to control the FWER at a pre-specified significant level α, multiple comparison procedures require the constant c_α, known as the critical value, for which

$$P\left(\max_{i=1,\ldots,k} |T_i| < c_\alpha\right) = 1 - \alpha. \tag{2.11}$$

The underlying factor for utilizing the maximum statistic to control FWER is simply that one or more of the test statistics will exceed c_α if and only if the maximum statistic exceeds c_α. For this c_α, the k intervals $\widehat{\theta}_i \pm c_\alpha \sqrt{S^*_{ii}}$ where S^*_{ii} is the ith diagonal element of $\boldsymbol{S}^*$ simultaneously contain their respective parameters θ_i with probability exactly $1 - \alpha$.

However, the nature of comparisons among the regression parameters make the joint distribution of the test statistics, T_i, complicated by the dependence among the T_i's. The integrals are typically high-dimensional and difficult to obtain. In some situations, these integrals are even unobtainable. Researchers are thus forced to settle with more conservative solutions which waste data. More recent developments in this field include the use of resampling methods to estimate the critical value (Westfall (2011); Westfall and Young (1993)). An advantage of resampling-based procedures is their efficiency in approximating the true distribution of the test statistics. Additionally, resampling-based procedures incorporate dependence structures and non-normal distributional properties otherwise not captured in other multiple comparison procedures. Thus, for highly correlated and non-normally distributed tests, this approach is considerably more powerful than other multiple comparison procedures. Conversely, the price for the gain of power is that the resampling-based procedures are computationally intensive. Details of some traditional approaches to multiple comparison procedures are reviewed by Hochberg and Tamhane (1987) and Hsu

(1996).

2.3.1 Existing Conservative Solutions

Statistical procedures designed to control the problems associated with multiple comparisons are called multiple comparison procedures (MCPs). There are many methods for multiple comparisons; these methods differ in how well they properly control the overall significance level and their relative power. Some existing procedures are discussed in the sections below.

Fisher's Least Significant Difference

In 1935, R. A. Fisher developed the first pairwise comparison procedure called the least significant difference (LSD). The procedure consists of two steps. The first step involves testing the null hypothesis by an α-level F-test. The procedure terminates without making detailed inferences on pairwise differences if the F-test is not significant; otherwise, each pairwise difference is tested by an α-level t-test in the second step. This procedure is known in literature as the Fisher's protected least significant difference test.

The LSD controls the FWER at level α under the null hypothesis due to the protection provided to the hypothesis by the preliminary F-test. Thus, the LSD controls the FWER in the weak sense − under the null hypothesis but not under all configurations.

Probability-based Corrections

The Bonferroni correction is a widely used method to deal with the problem of multiple comparisons. This method is simple and may be applied in any multiple comparison situation. It is based on the Bonferroni inequality:

$$P\left\{\bigcup_i (A_i)\right\} \leq \sum_i P\{A_i\} \tag{2.12}$$

where $A_i = Y_i > c$ relates to the event of committing a type I error, i.e, rejecting the null hypothesis when it is true. In our setting, A_i is the event that the interval i does not

contain its target $\mathbf{C}_i'\boldsymbol{\beta}$ where $\mathbf{C}_i$ is the ith row of the matrix of constants $\boldsymbol{C}$. For a family of hypotheses $\{H_i\}_{i=1,\ldots,k}$ with corresponding p-values $\{p_i\}_{i=1,\ldots,k}$, the Bonferroni correction states that rejecting all $p_i < \frac{\alpha}{k}$ will control the FWER at level α. Despite its simplicity and universality, the Bonferroni correction is very conservative and has low power at larger numbers of comparisons. Šidák adjustment (1967) is an improvement upon the Bonferroni's method which rejects the null hypothesis, H_i, when the p-value, $p_i < 1-(1-\alpha)^{1/k}$. Though this method is more powerful than the Bonferroni, the procedure requires the individual tests to be independent. Consequently, for correlated outcomes, this method could be very conservative. Kounias (1968) proved that

$$P\left\{\bigcup_i (A_i)\right\} \leq \sum_i P\{A_i\} - \max_k \left\{\sum_{i\neq k} P\left\{A_i \cap A_k\right\}\right\} \tag{2.13}$$

improved on the Bonferroni upper bound of $\sum_i P\{A_i\}$. Regardless of being powerful than the Bonferroni, the modified Bonferroni methods have a tendency to be very conservative. These procedures are probability-based; making use of the mathematical properties of the hypotheses structure without taking into account the correlation structure of the test statistics. Thus, for highly correlated and non-normal test statistics, these methods could be very conservative, less powerful and waste data.

Hunter-Worsley Method

Hunter (1976) improved on Kounias' upper bound by adapting it to use a maximum tree of a graph connecting the nodes A_i. Hunter's result was rediscovered by Worsley (1982) who applied it to diverse statistical problems. The method of Hunter and Worsley uses the inequality:

$$P\left\{\bigcup_i (A_i)\right\} \leq \sum_i P\{A_i\} - \max_{\tau\in T} \left\{\sum_{\tau} P\left\{A_i \cap A_k\right\}\right\} \tag{2.14}$$

where T is the set of all spanning trees. Though computationally intensive, the Hunter-Worsley method is more powerful than the Bonferroni and other improved Bonferroni methods. An advantage of this method over the Bonferroni correction and other related

methods is the fact that the Hunter-Worsley method takes into account the correlation or dependence structure of the test statistics. The Hunter-Worsley method improves upon the Bonferroni upper bound by the subtraction of the $p-1$ two-way overlap of the event of committing a type I error. When there are many higher overlap other than the two-way overlap removed by the subtraction, then the Hunter-Worsley method would also be very conservative.

2.3.2 Resampling Methods

Statisticians are gradually becoming comfortable with the use of resampling methods for inferential statistics. Some statisticians are of the opinion that resampling statistics will soon take over the traditional non parametric and parametric procedures. However, others argue that, whenever possible, model-based solutions should always be attempted alongside resampling-based methods. Resampling is a general term referring to statistical methodologies based on taking repeated random samples of observed data for approximating the true sampling distribution of a statistic using a computer-intensive simulation analysis. In this technique, the values of the observed variables are randomly re-allocated to treatment groups, and the test statistics are re-computed. These reallocation and re-computations are repeated a large number of times in order to approximate the sampling distribution of the desired statistic.

Resampling methods have advantage of analyzing larger classes of statistical problems. Additionally, resampling methods have the ability of incorporating distributional characteristics, thus making the tests more robust. The dependence structures in multiple comparison applications are naturally included in the resampling-based estimates, thereby improving the power of the tests. Despite the fact that unknown dependence structures and non-normal distributional characteristics are approximately included via resampling, they are often not accurately captured. On the other hand, they are neither captured accurately by other existing methods. In many cases, resampling provides a suitable and asymptotically valid method for incorporating these characteristics.

Examples of resampling methods include bootstrapping, jackknifing and permutation tests. In this study, our main attention will be focused on the bootstrap method. The basic idea of bootstrapping is to resample with replacement from the sample data set from which the sampling distribution can be approximated.

Westfall's (2011) Resampling-based multiplicity correction method

As earlier stated, though the modified Bonferroni methods are more powerful than the simple Bonferroni methods, they still tend to be conservative especially for high-dimensional comparisons. Furthermore, as noted earlier, these methods are based on the mathematical properties of the hypotheses structure, thereby not taking into account the dependence structure of the test statistics. An approach that uses the information of dependencies and distributional characteristics is given by resampling procedures. These procedures are more flexible and allow for correlated inferences and high-dimensional comparisons. Westfall (2011) outlined a resampling procedure for estimating the critical value c_α in equation (2.11) in linear model settings.

Consider the multivariate regression model

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{2.15}$$

where $\boldsymbol{Y}$ is an $n \times 1$ vector of response variables, $\boldsymbol{X}$ is a full rank $n \times p$ matrix of covariates, $\boldsymbol{\beta}$ is a $p \times 1$ vector of regression parameters and $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of random error terms. Next, assume the parameters of interest are a linear combination of the regression parameters such that $\theta_j = \mathbf{c}_j\boldsymbol{\beta}\mathbf{d}_j$ for constant vectors $\mathbf{c}_j$ and $\mathbf{d}_j$ where $\mathbf{d}_j$ identifies a particular endpoint. Suppose $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$ is the least-squares vector of parameter estimates and $S = \dfrac{\left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)^T\left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)}{n-1}$ is the unbiased pooled covariance matrix estimator. Then the test statistic T_j for testing these sets of hypotheses is given as:

$$T_j = \frac{\mathbf{c}_j'\widehat{\boldsymbol{\beta}}\mathbf{d}_j - \mathbf{c}_j'\boldsymbol{\beta}\boldsymbol{d}_j}{\left(\mathbf{d}_j\mathbf{S}\mathbf{d}_j'\right)^{1/2}\left\{\mathbf{c}_j(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}_j'\right\}^{1/2}} \tag{2.16}$$

It can be shown that the randomness in the distribution of the test statistic (2.16) depends solely on the random error term, ε_j (Westfall, 2011). Based on this fact, Westfall (2011) proposed bootstrapping the residuals to approximate the empirical joint sampling distribution of the test statistics and computing the critical point thereafter. The author thus proposed the following algorithm.

Westfall's multiplicity correction algorithm

Suppose the residuals of a linear model are independently and identically distributed from G, (i.e, $\varepsilon_j \overset{iid}{\sim} G$). Then,

- If G is known, then c_α can be estimated by simulating the residual vectors $\boldsymbol{\varepsilon}$ from G; computing the value of $\max |T_i|$ and repeating to obtain the estimated quantile $\widehat{c}_{\alpha,G}$.
- However, if G is unknown, estimate the distribution of G as the empirical distribution of the set of residuals $\{\widehat{\varepsilon}_i\}_{i=1,\ldots,n}$, where $\widehat{\varepsilon}$ is the ith row of the residual vector $\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}$.

Correlation between endpoints are incorporated in either way. Such simulated $\widehat{c}_\alpha$ is a consistent estimate of the true c_α; larger bootstrap sample size produces more accurate estimates of the true parameter c_α. Studies have shown that this kind of resampling scheme is effective in the multi-sample case (Fisher and Hall, 1990, Westfall and Young, 1993). Although the study of Westfall (2011) was for linear model settings, specifically the linear regression model, the author's techniques and methodologies are adaptable to the generalized linear model settings.

Chapter 3

Applied Techniques and Methodology

3.1 Resampling Methods for GLMs

After one has decided to estimate the critical value of equation (2.11) using a resampling-based procedure, the next question is what and how to resample. Methods for bootstrapping ordinary linear models have been shown in literature to approximate more accurately the sampling distribution of test statistics. The methods include residual and vector resampling. These methods are also applicable in the generalized linear model settings. In order to control the familywise error rate in multiple comparisons at a pre-specified significance level in linear models, Westfall (2011) recommends residual resampling for estimating the critical value. Reampling-based multiple comparison procedures for GLMs may be developed analogous to the residual resampling proposed by Westfall (2011). However, there are several different types of residuals defined for GLMs and one is faced with the challenge of which residuals to bootstrap. The ideal approach is to bootstrap from independent and identically distributed quantities, but no such residuals are easily obtainable for the class of GLMs. The Pearson residual is the most amenable residual, both theoretically and computationally, to a bootstrapping process (Moulton and Zeger, 1991). A simulation study by Moulton and Zeger (1991) showed that bootsrapping with the standardized Pearson residuals in GLMs consistently outperformed the use of the raw Pearson residuals. Thus in adapting the techniques and methodologies of Westfall (2011), the standardized Pearson residuals will be resampled.

3.2 Proposed Methodology

Consider the generalized linear model;

$$\boldsymbol{\eta} = g\left[E(\boldsymbol{Y}|\boldsymbol{X})\right] = \boldsymbol{X}\boldsymbol{\beta} \tag{3.1}$$

where $\mathbf{Y} \in \mathbb{R}^{n,1}$ is an $n \times 1$ vector of response variables, $\mathbf{X} \in \mathbb{R}^{n,p}$ is an $n \times p$ matrix of covariates and $\boldsymbol{\beta} \in \mathbb{R}^{p,1}$ is a $p \times 1$ vector of regression parameters. Let $\widehat{\boldsymbol{\beta}}_m = (\widehat{\beta}_{m1}, \widehat{\beta}_{m2}, \ldots, \widehat{\beta}_{mk})$ be the maximum likelihood estimates (MLEs) and $\widehat{\boldsymbol{\beta}}_p = (\widehat{\beta}_{p1}, \widehat{\beta}_{p2}, \ldots, \widehat{\beta}_{pk})$ be the penalized maximum likelihood estimates (pMLEs) both of $\boldsymbol{\beta}$. Then, $\widehat{\boldsymbol{\beta}}_m$ and $\widehat{\boldsymbol{\beta}}_p$ have an asymptotic multivariate normal distribution with mean vector $\boldsymbol{\beta}$ and covariance matrix $\mathbf{S} = \left(\boldsymbol{X}^T \boldsymbol{W_i} \boldsymbol{X}\right)^{-1}$ for $i = m$ or p since the weight matrix depends on the estimator utilized as discussed in sections 2.1.1 and 2.1.2 .

Quantities of interest in this model is a linear combination of the p-dimensional regression parameters, $\boldsymbol{\theta} := \boldsymbol{C}\boldsymbol{\beta}d$. Here, $\mathbf{C} \in \mathbb{R}^{k,p}$ is a full rank $k \times p$ matrix of constants, normally known as the contrast matrix and $d \in \mathbb{R}$ identifies a particular endpoint chosen *a priori*. Let $\boldsymbol{\theta}$ be estimated by $\widehat{\boldsymbol{\theta}} := \boldsymbol{C}\widehat{\boldsymbol{\beta}}d$, then as previously discussed, $\widehat{\boldsymbol{\theta}}$ has asymptotic multivariate normal distribution with mean vector $\boldsymbol{C}\boldsymbol{\beta}d$ and covariance matrix $\boldsymbol{S}^* := d\boldsymbol{C}\boldsymbol{S}\boldsymbol{C}^T d$.

The goal of the study as stated earlier is to adapt the techniques and methodologies of Westfall (2011) to develop and evaluate multiple comparison procedures for these contrasts, $\boldsymbol{C}\boldsymbol{\beta}d$ of the regression parameters. However, adaptation of this idea is complex due to the non-normal residuals.

The proposed bootstrap algorithm for the estimation of the critical value for the multiple comparison procedure for GLMs is then:

Algorithm for the proposed method

1. Estimate the model parameters from the data using a reliable estimation procedure

2. Estimate the Pearson standardized residuals, $\widehat{\varepsilon}$

3. Sample ε^* as a replacement sample from $\widehat{\varepsilon}$

4. Using ε^*, generate new set of response variables, $\mathbf{Y}^*$

5. Estimate $\boldsymbol{\beta}^*$ and $\mathbf{S}^*$ (bootstrap regression parameters and covariance matrix respectively) using $\mathbf{Y}^*$ and the original matrix of covariates

6. Compute

$$T_i\{\varepsilon^*\} = \frac{C_i\boldsymbol{\beta}^* d}{\sqrt{dC_i S^*_{ii} C_i^T d}}, \quad i = 1, \ldots k$$

7. Compute $\max |T_i\{\varepsilon^*\}|$ and store it as M

8. Repeat steps 3 to 7 N_B times (where N_B is the bootstrap sample size)

9. Compute the empirical $1-\alpha$ quantile of the N_B M values and call it $\widehat{c}_\alpha$

There are two sources of errors incorporated into the critical value; one which is due to the simulation error and the other due to sampling ε^* from $\widehat{\varepsilon}$. Again, the non-normal distribution of the residuals can make the distribution of the maximum test statistic very skewed. Thus, the right tail of the curve would be much longer than the left tail. The empirical $1-\alpha$ quantile will therefore be pulled out by the long right tail. This could result in estimates $\widehat{c}_\alpha$ that are not close to the true c_α. Thus, researchers need to be cautious in using the resampling-based procedures in estimating the empirical $1-\alpha$ quantile.

The existing conservative methods discussed in section 2.3.1 are all probability-based methods and thus, do not depend directly on the sample size or the parameter estimates. However, the resampling-based procedures depend directly on the sample size. There is therefore the need to check for the stability of the parameter estimates as unstable parameter estimates may have serious implications. For illustration, consider an example of a randomized clinical trial conducted at five centers to compare an active drug to placebo for treating fungal infections (See Agresti (2007), Example 5.3.3). The table of values for this example is shown in Table C.1 in Appendix C. In the example, centers 1 and 3 had

no success counts. Thus, the MLE estimates of the terms pertaining to their effects equal $-\infty$. The associated standard errors are very large. Therefore the usual MLE-based model predicting the success of the active drug given the treatment and center levels is not relaible and reasonable estimates are not obtained. This is due to separability or quasi-separability in the data. This is a very common issue with GLMs with multiple categorical predictors or with classification having several categories. When the data is separable, the MLE parameter estimates are not stable. Thus, it is not advisable to use the estimated model. Even if the estimates are reasonable, as discussed earlier, they may still perform poorly due to bias in the parameter estimates at small sample sizes. Hence, there is the need for stable parameter estimates in order to obtain meaningful and reasonable results when utilizing the resampling-based procedure. Investigators thus need to first check for the stability of the model before proceeding with the resampling-based procedures. Literature shows that the Bayesian approach to estimating the parameter estimates provides a finite and less biased estimates in cases for which the MLE is infinite (Firth, 1993).

Chapter 4

Results and Analysis

4.1 Simulation Settings

Literature shows that the existing multiple comparison procedures discussed in Section 2.3.1 conserves the nominal 5% significance level. Since a new algorithm has been proposed for multiple comparisons in generalized linear models, it is necessary to evaluate this new method's performance in comparison to other existing methods using relative efficiency calculations. Also, we need to investigate whether the proposed method conserves the nominal 5% familywise significance level. These evaluations can be attained via simulation.

All models are simulated as: $g\left[E(\boldsymbol{Y}_i|\boldsymbol{X}_i)\right] = \boldsymbol{X_i\beta}$ $(i = 1, \ldots n)$ where g is the link function, $\mathbf{Y}_i$ is the response, $\mathbf{X}_i$ is a vector of explanatory variables and $\boldsymbol{\beta}$ is the $p \times 1$ vector of regression parameters. Several settings for the sample size, the number and the type of explanatory variables and the number of regression parameters were considered. The focus was on logistic regression models, however, the proposed techniques and methodologies can be applied to other models in the class of generalized linear models.

To begin, the regression parameters were set to several fixed values and the sample size and number of explanatory variables were allowed to vary. For our study, both continuous and categorical explanatory variables and binary response variables were considered. Two cases were considered for the regression parameters:

- Set 1: case where all the $\beta = 1$,
- Set 2: case where the β alternate between 1 and 5.

For the set of continuous explanatory variables, $p = 5$ and $p = 10$ were considered, no in-

tercept term was included in the model fitting. All pairwise comparisons of the regression parameters were investigated. The set of continuous explanatory values, $X_i, i = 1, 2, \ldots, p$ of dimension n, where n is the sample size were generated as independent and identically distributed standard normal variables. In order to generate the response variables, Y_i's as binomial random variables, the success probabilities were calculated using the formula; $\gamma_i = \dfrac{\exp\{\mathbf{X}_i\boldsymbol{\beta}\}}{1+\exp\{\mathbf{X}_i\boldsymbol{\beta}\}}$, $i = 1, 2, \ldots, n$. The response variables were then simulated as binomial random variables with success probabilities γ_i. Next, the set of generated Y_i values and X data set were utilized to compute the estimated regression parameter values ($\widehat{\beta}_m$ and $\widehat{\beta}_p$), the estimated covariance matrix ($\boldsymbol{S}_m$ and $\boldsymbol{S}_p$) and the associated estimated standardized Pearson residuals ($\widehat{\boldsymbol{\varepsilon}}_m$ and $\widehat{\boldsymbol{\varepsilon}}_p$). The subscripts m and p indicate the estimation procedure utilized (mle and pmle respectively). The estimated standardized Pearson residuals were then resampled with replacement to obtain the new set of residuals, $\boldsymbol{\varepsilon}^*$. After, the bootstrapped residuals were utilized to generate the new set of response variables; calculating the bootstrapped success probabilities, $\gamma_i^* = \dfrac{\exp\{\boldsymbol{\varepsilon}^*\}}{1+\exp\{\boldsymbol{\varepsilon}^*\}}$ and generating Y_i^* as binomial random variables with success probability γ_i^*. The set of newly generated Y_i^* values and the original X data set were then utilized to compute bootstrapped regression parameters, β^*, and the associated bootstrapped covariance matrix, $\mathbf{S}^*$. Thereafter, the test statistic was computed as $T_i^* = \dfrac{\mathbf{C}_i\beta^* d}{\sqrt{d\mathbf{C}_i S_{ii}^* \mathbf{C}_i' d}}$, and the maximum test statistic was saved. The bootstrap process was repeated N_B times; saving the maximum test statistic each time. Here, C was chosen as the all-pairwise comparisons of Tukey (1953) and without loss of generality, d was assumed to be 1. At the end of all the bootstrap replications, the N_B maximum test statistic were sorted in an ascending order and the empirical critical value $\widehat{c}_\alpha$, was estimated by picking the 97.5th percentile. The entire simulation process was repeated r (simulation replication) times, each time finding the estimate of the empirical critical value. Afterwards, the coverage of $\boldsymbol{\theta}$ were analyzed and the error rates were computed. In the study, any individual error (i.e, any interval that failed to cover its true quantity) implied overall error.

Additionally, the set of categorical explanatory variables were generated as independent

and identically distributed (iid) binomial random variables with parameters $p-1$ and 0.5 success probability. In order to generate these categorical explanatory variables $T_i \overset{iid}{\sim} BIN(p-1, 0.5)$ was generated for $i = 1, \ldots, n$. Reference coding was employed in this set of explanatory variables. In this case, every slope parameter is the log odds ratio or the relative risk for that level of covariate with respect to the reference. To incorporate the reference coded indicator variables, whenever $T_i = 0$ then $X_i = (1, 0, 0, 0)$, if $T_j = j$ then the $j+1$ element of X_i is set to 1 and all other elements are set to 0. Once the explanatory variables were generated, the response variables were generated in an analogous manner to the case where the explanatory variables were continuous. The bootstrap procedure then followed as described above. Again, $p = 5$ and $p = 10$ were considered for the set of categorical explanatory variables.

For our simulation settings, $N_B = 10,000$ bootstrap sample size were considered. For the cases where $p = 10$, sample sizes, $n = 200,\ 300$ and 500 were considered. Sample sizes, $n = 100,\ 200,$ and 250 were considered for the cases where $p = 5$. In order to calculate the error rate, we considered $r = 1000$ simulation replications for the cases where $n = 300,\ p = 10$ and $r = 200$ for the cases where $n = 100,\ p = 5$ for the continuous explanatory variable settings whereas $r = 1000$ simulation replications were considered for the cases where $n = 100,\ p = 5$ and $r = 200$ for the cases where $n = 300,\ p = 10$ for the categorical explanatory variable settings. In these settings, both cases of the regression parameters were considered, that is, cases where the $\beta = 1$ and the β alternate between 1 and 5. Furthermore, in other to compare the performance of the methods relative to sample sizes, $r = 5$ simulation replications were considered for cases where $p = 10$ and $n = 200,\ 300$ and 500 and the case where $p = 5$ and $n = 100,\ 200$ and 250. This number of simulation replications was considered in order to check for the stability and consistency of the estimated empirical critical points. Both cases of β were again considered in these settings.

The relative efficiency of one method to the Bonferroni adjustment was then considered in all settings as the ratio of the Bonferroni's squared critical value to the method's

squared critical value. This ratio is approximately the ratio of sample sizes needed to achieve intervals of equal length. The ratio is directly proportional to the critical value; the optimal critical value automatically produces the most efficient procedure in terms of data wastage. Again, in order to investigate the influence of the estimation method on the resampling-based procedure, models utilizing both the MLE and pMLE were investigated in all the simulation settings. A pre-specified significance level $\alpha = 0.05$ was assumed for the familywise error rate in all the simulation settings. All simulations were performed in R statistical software package (R Core Team (2013)). The pMLE procedure is available in the *brglm* package in R.

4.2 Simulation Results

This section presents and analyzes the simulation results for the proposed methodology. Tables 4.1 and 4.2 display the average empirical simulated critical values for the continuous explanatory variables for cases where $n = 300,\ r = 1000$ and $n = 100,\ r = 200$ and for the categorical explanatory variables for cases where $n = 300,\ r = 200$ and $n = 100,\ r = 1000$. Thus, the average of all the empirical critical values were calculated across the number of simulation replications for each case. For instance, the first row of Table 4.1 represents the average of all the empirical critical values obtained for the 1000 simulation replications in the case where β was the set of all ones for the continuous explanatory variable settings. In the tables, c_{mle} is the empirical critical value based on the MLE, c_{pMLE} is the empirical critical value based on the pMLE, c_{hw} is the Hunter-Worsley based critical value and c_{bf} is the usual Bonferroni adjustment. Set specifies the case of regression parameters used where 1 is the case where all the $\beta = 1$ and 2 is the case where the β alternate between 1 and 5. Analyzes from these two tables indicate that the pMLE resampling-based procedure produces the smallest critical point in all settings; outperforming the other procedures.

Table 4.3 also contains the average relative efficiencies of all the 1000 simulation replications of each of the methods to the Bonferroni adjustment. Again, in this table, the pMLE

Table 4.1: Average simulated critical points for continuous explanatory variables

n	p	c_{mle}	c_{pMLE}	c_{hw}	c_{bf}	Set
300	10	3.33760	3.22480	3.27401	3.29377	1
300	10	3.33830	3.22470	3.27398	3.29377	2
100	5	2.84256	2.70165	2.83661	2.87407	1
100	5	2.84417	2.70103	2.83697	2.87407	2

c_{mle} = MLE-based critical value c_{pMLE} = pMLE-based critical value c_hw = Hunter-Worsley-based critical value c_{bf} = bonferroni critical value n = sample size p = number of parameters

Table 4.2: Average simulated critical points for categorical explanatory variables

n	p	c_{mle}	c_{pMLE}	c_{hw}	c_{bf}	Set
300	10	3.11932	2.98194	3.26118	3.29377	1
100	5	2.72363	2.60461	2.84352	2.87407	1
100	5	2.73711	2.61109	2.84348	2.87407	2

c_{mle} = MLE-based critical value c_{pMLE} = pMLE-based critical value c_hw = Hunter-Worsley-based critical value c_{bf} = bonferroni critical value n = sample size p = number of parameters

resampling-based procedure gives the highest relative efficiency values. For example, analyzes from the first row of the table indicate that utilizing the pMLE resampling-based procedure is 4% more efficient in terms of data than the Bonferroni adjustment while utilizing the Hunter-Worsley procedure is 1% efficient than the Bonferroni adjustment. However, it is more efficient utilizing the Bonferroni instead of the MLE resampling-based procedure in this scenario. Interestingly, with the exception of the cases where $n = 300$ and continuous explanatory variables are considered, utilizing the MLE resampling-based procedure is more efficient than utilizing the Bonferroni adjustment. This will be further discussed in the next chapter. Overall, utilizing the pMLE resampling-based method is data efficient

compared to the other methods in all the settings considered in Table 4.3.

Table 4.3: Average relative efficiency of a method to Bonferroni

Type of predictor	n	p	MLE	pMLE	HW	Set
Continuous	300	10	0.97399	1.04331	1.01211	1
Continuous	300	10	0.97358	1.04338	1.01213	2
Continuous	100	5	1.02242	1.13188	1.02660	1
Continuous	100	5	1.02127	1.13242	1.02633	2
Categorical	300	10	1.13452	1.24079	1.02013	1
Categorical	100	5	1.14965	1.23731	1.02161	1
Categorical	100	5	1.12410	1.22410	1.02164	2

To maintain the nominal 5% significant level, the error rates of multiple comparison procedures should be in the interval $0.05 \pm Z_{1-\alpha/2}\sqrt{\frac{(0.05)(0.95)}{r}}$, where $Z_{1-\alpha/2}\sqrt{\frac{(0.05)(0.95)}{r}}$ is the margin of error. For the setting of this study, $c_\alpha \in [0.0365, 0.0635]$ for the cases where 1000 simulation replications were considered and $c_\alpha \in [0.0198, 0.0802]$ for the cases where 200 simulation replications were considered. For this interval, any method that gives an error rate above the upper bound is said to be liberal and it is said to be very conservative if the error rates fall below the lower bound. Recall that researchers are always interested in methods that are less conservative. Tables 4.4a and 4.4b give the MLE-based and pMLE-based error rates respectively for the continuous explanatory variable scenario while Tables 4.5a and 4.5b give the MLE-based and pMLE-based error rates respectively for the categorical explanatory variable scenario. Liberal procedures are marked with an asterisk (*) and very conservative methods are marked with a plus (+) sign. Examining Tables 4.4a and 4.5a, it can be observed that utilizing the maximum likelihood estimation mostly produces a very conservative procedure for all the methods. In spite of this, the error rates indicate that the Hunter-Worsley method is less conservative in this case compared to the MLE resampling-based method and the Bonferroni adjustment. On the contrary, Tables 4.4b and 4.5b do not show the same patterns as seen in Tables 4.4a and 4.5a. Here, it can be noticed that when the penalized maximum likelihood estimation is utilized the

Table 4.4: Error rates for continuous explanatory variables scenario

(a) MLE-based error rates

n	p	MLE	HW	Bonferroni	Set
300	10	0.022^{+}	0.028^{+}	0.026^{+}	1
300	10	0.007^{+}	0.008^{+}	0.008^{+}	2
100	5	0.010^{+}	0.010^{+}	0.010^{+}	1
100	5	0.020	0.020	0.015^{+}	2

(b) pMLE-based error rates

n	p	pMLE	HW	Bonferroni	Set
300	10	0.022^{+}	0.019^{+}	0.019^{+}	1
300	10	0.052	0.048	0.047	2
100	5	0.010^{+}	0.010^{+}	0.010^{+}	1
100	5	0.080	0.070	0.070	2

+ indicates very conservative error rate

Table 4.5: Error rates for categorical explanatory variables scenario

(a) MLE-based error rates

n	p	MLE	HW	Bonferroni	Set
300	10	0.010^{+}	0.005^{+}	0.005^{+}	1
100	5	0.001^{+}	0.001^{+}	0.000^{+}	1
100	5	0.022^{+}	0.007^{+}	0.007^{+}	2

(b) pMLE-based error rates

n	p	pMLE	HW	Bonferroni	Set
300	10	0.020	0.000^{+}	0.000^{+}	1
100	5	0.004^{+}	0.001^{+}	0.001^{+}	1
100	5	0.128^{*}	0.089^{*}	0.083^{*}	2

+ indicates very conservative error rate

* indicates liberal error rate

pMLE resampling-based procedure produces a less conservative method in comparison to the Hunter-Worsley and the Bonferroni adjustment irrespective of which set of the regression parameters were employed. Surprisingly, for the categorical explanatory variable settings, the pMLE resampling-based method was liberal while the MLE resampling-based method was very conservative for the case where $p = 5$, $n = 100$ and β alternate between 1 and 5. The resampling-based procedures were however very conservative for the same settings with $\beta = 1$. Notwithstanding, the critical values and the relative efficiencies obtained for these settings for both sets of the β's are quite similar. These results are very astonishing and thus need further investigations. It should also be noted that due to the problems associated with dealing with GLMs with multiple categorical explanatory variables, the

error rates for the case where $p = 10$, $n = 300$ and β alternate between 1 and 5 for the categorical explanation variable settings could not be calculated. Estimation warnings and errors occurred for this setting during the simulation process.

Tables 4.6 and 4.7 show the comparisons of the performance of the methods in terms of the critical values and relative efficiencies with respect to the sample sizes and the number of parameters for both the continuous and categorical explanatory variables. As stated earlier, a total of 10,000 bootstrap replications and 5 simulation replications were considered for each case. The table reports the average of the desired quantities. Optimal critical values are marked with an asterisk (*) whereas optimal relative efficiency values are marked with a plus (+) sign. The results indicate that the pMLE resampling-based procedure has an improved performance for most of the settings. These results are also shown graphically in Figures 4.1 to 4.6 from pages 33 to 35.

Table 4.6: Comparisons of the performance of the methods with respect to sample size and number of parameters for continuous explanatory variables

n	p	n/p	c_{mle}	c_{pmle}	c_{hw}	c_{bf}	RE_{mle}	RE_{pmle}	RE_{hw}	Set
100	5	20	2.848	2.703*	2.836	2.874	1.024	1.135+	1.027	1
100	5	20	2.840	2.693*	2.838	2.874	1.025	1.139+	1.026	2
150	5	30	2.891	2.786*	2.819	2.851	0.976	1.058+	1.023	1
150	5	30	2.877	2.798*	2.817	2.851	0.982	1.038+	1.024	2
250	5	50	2.912	2.842	2.802*	2.833	0.946	0.994	1.022+	1
250	5	50	2.905	2.857	2.803*	2.833	0.951	0.983	1.021+	2
200	10	20	3.327	3.160*	3.289	3.311	0.990	1.098+	1.014	1
200	10	20	3.307	3.162*	3.290	3.311	1.002	1.097+	1.013	2
300	10	30	3.337	3.212*	3.274	3.294	0.974	1.052+	1.012	1
300	10	30	3.339	3.225*	3.274	3.294	0.973	1.043+	1.012	2
500	10	50	3.365	3.305	3.262*	3.280	0.950	0.985	1.011+	1
500	10	50	3.352	3.283	3.263*	3.280	0.958	0.999	1.011+	2

* indicates optimal critical value

+ indicates optimal relative efficiency

Table 4.7: Comparisons of the performance of the methods with respect to sample size and number of parameters for categorical explanatory variables

n	p	n/p	c_{mle}	c_{pmle}	c_{hw}	c_{bf}	RE_{mle}	RE_{pmle}	RE_{hw}	Set
100	5	20	2.765	2.639*	2.843	2.874	1.082	1.189+	1.022	1
100	5	20	2.774	2.655*	2.844	2.874	1.074	1.172+	1.021	2
150	5	30	2.807	2.707*	2.821	2.851	1.034	1.115+	1.021	1
150	5	30	2.819	2.711*	2.822	2.851	1.025	1.111+	1.021	2
250	5	50	2.838	2.753*	2.805	2.833	0.998	1.061+	1.020	1
250	5	50	2.873	2.832	2.806*	2.833	0.973	1.001	1.019+	2
200	10	20	3.080	2.912*	3.284	3.311	1.161	1.303+	1.017	1
200	10	20	3.064	2.920*	3.288	3.311	1.171	1.291+	1.014	2
300	10	30	3.157	3.002*	3.258	3.294	1.089	1.205+	1.022	1
300	10	30	3.222	3.085*	3.268	3.294	1.045	1.141+	1.016	2
500	10	50	3.261	3.163*	3.255	3.280	1.013	1.079+	1.016	1
500	10	50	3.287	3.197*	3.252	3.280	0.996	1.053+	1.017	2

* indicates optimal critical value

+ indicates optimal relative efficiency

4.3 Application

For illustrative purposes, the MLE and pMLE resampling-based multiple comparison procedures are applied to data from the 2009 National Health Interview Survey (NHIS). Information about childhood asthma and other health conditions affecting the youth of the United States (U.S.) is contained in this data. Our interest in this data is to analyze the impact of region and hayfever allergy status on the incidence of asthma-related emergency room (ER) visits for U.S. children. The response variable in this case is binary, indicating whether a child had visited ER due to an asthma attack in the past 12 months. Thus, a binomial GLM could be utilized for this analysis. The explanatory variables in the model are all indicator variables identifying the region of the United States (NE, MW, S, W),

where the child resides and whether the child is diagnosed with hayfever allergies. To incorporate the reference coding, x_1 is an indicator for the Mid West (MW), x_2 for the South (S), x_3 for the West (W), and the reference level is the Northeast (NE). x_4 is an indicator for a diagnosis of hayfever allergies. It is realistic in this scenario to make comparisons for children with and without diagnosed hayfever allergies across regions. The model can be expressed as

$$\text{logit}(\mu_i) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4,$$

where each x_i is an indicator variable for the specified condition. In this case, $\beta_1 - \beta_2$ is the log odds ratio for comparing subjects without hayfever allergies in the Midwest to subjects without hayfever allergies in the South. For this analysis, both the MLEs and the pMLEs are used for the parameter estimation of the model. Table 4.8 displays the estimated confidence intervals for the linear combinations of the parameters of interest. In the table, confidence intervals where the odds ratio include 1 are marked with an asterisk (*).

As stated earlier in chapter 3, unstable parameter estimates due to separability in the data could have serious implications on the resampling-based method especially when utilizing the maximum likelihood estimation. However, in this data example, the MLE resampling-based procedure yielded reasonable and stable parameter estimates and did not show any signs of separation. For this example, the MLE resampling-based critical value, $c_{\text{MLE}} = 2.647$, the pMLE resampling-based critical value, $c_{\text{pMLE}} = 2.410$ and the Hunter-Worsley critical point, $c_{\text{HW}} = 2.909$. The usual Bonferroni critical value, $c_{\text{bf}} = 2.934$. The relative efficiencies in this examples are $RE_{mle} = 1.228$, $RE_{pmle} = 1.482$ and $RE_{hw} = 1.018$. It can be concluded from the confidence intervals for the odds ratios that the factors were identified as significantly different across the competing intervals except:

- when the northeast (NE) and south (S) are compared for children with no diagnosed hayfever allergies

- when the south (S) and west (W) are compared for children with diagnosed hayfever allergies

The findings for this data example are consistent with the findings of Wagler and McCann (2014). It should be noted also that the critical points and relative efficiencies obtained for the data example are all within the range of the values observed in the simulations.

Table 4.8: Confidence intervals on odds ratio of childhood asthma

Region	Hayfever	log OR	MLE-based OR	MLE-based HW OR	pMLE-based OR	pMLE-based HW OR
NE vs MW	no	β_1	$(5.75, 12.50)$	$(5.54, 12.99)$	$(4.64, 9.15)$	$(4.32, 12.78)$
NE vs S	no	β_2	$(0.75, 1.39)^*$	$(0.73, 1.43)^*$	$(0.73, 1.27)^*$	$(0.69, 1.42)^*$
NE vs W	no	β_3	$(1.65, 3.19)$	$(1.60, 3.30)$	$(1.54, 2.77)$	$(1.45, 3.28)$
MW vs S	no	$\beta_1 - \beta_2$	$(6.04, 11.49)$	$(5.85, 11.86)$	$(5.10, 8.93)$	$(4.81, 11.68)$
MW vs W	no	$\beta_2 - \beta_3$	$(0.34, 0.57)$	$(0.34, 0.58)$	$(0.37, 0.59)$	$(0.35, 0.58)$
S vs W	no	$\beta_1 - \beta_3$	$(2.62, 5.19)$	$(2.54, 5.36)$	$(2.34, 4.25)$	$(2.20, 5.29)$
NE vs MW	yes	$\beta_1 + \beta_4$	$(1.27, 3.06)$	$(1.21, 3.19)$	$(1.29, 2.79)$	$(1.19, 3.14)$
NE vs S	yes	$\beta_2 + \beta_4$	$(0.16, 0.35)$	$(0.15, 0.36)$	$(0.20, 0.40)$	$(0.19, 0.36)$
NE vs W	yes	$\beta_3 + \beta_4$	$(0.36, 0.80)$	$(0.34, 0.83)$	$(0.42, 0.86)$	$(0.39, 0.83)$
MW vs S	yes	$\beta_1 - \beta_2 + \beta_4$	$(1.24, 3.02)$	$(1.18, 3.16)$	$(1.33, 2.91)$	$(1.22, 3.10)$
MW vs W	yes	$\beta_2 - \beta_3 + \beta_4$	$(0.07, 0.16)$	$(0.06, 0.16)$	$(0.09, 0.20)$	$(0.09, 0.16)$
S vs W	yes	$\beta_1 - \beta_3 + \beta_4$	$(0.54, 1.36)^*$	$(0.52, 1.42)^*$	$(0.61, 1.37)^*$	$(0.56, 1.39)^*$

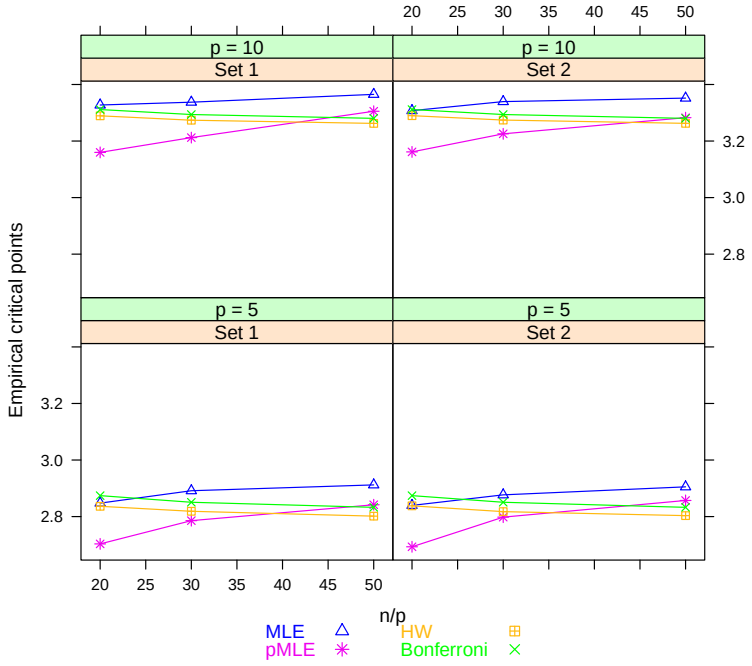

Figure 4.1: Empirical critical points for continuous explanatory variables

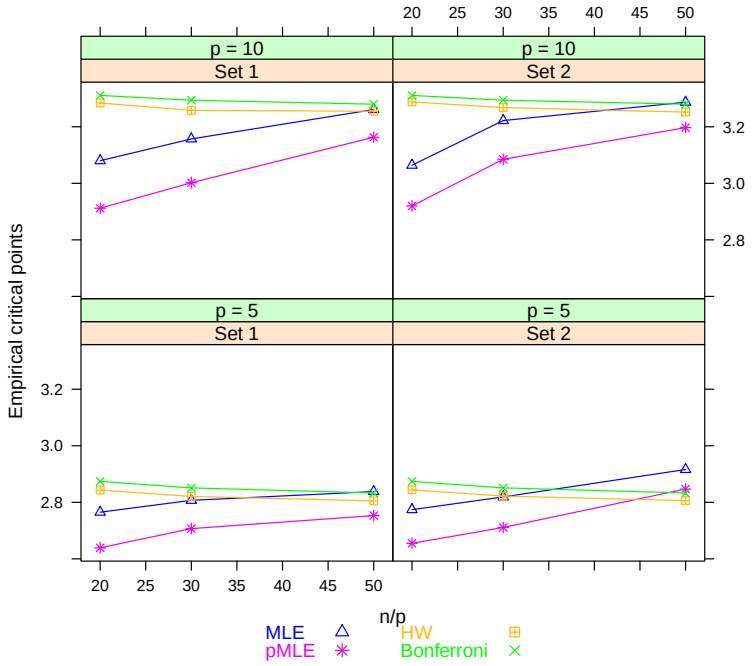

Figure 4.2: Empirical critical points for categorical explanatory variables

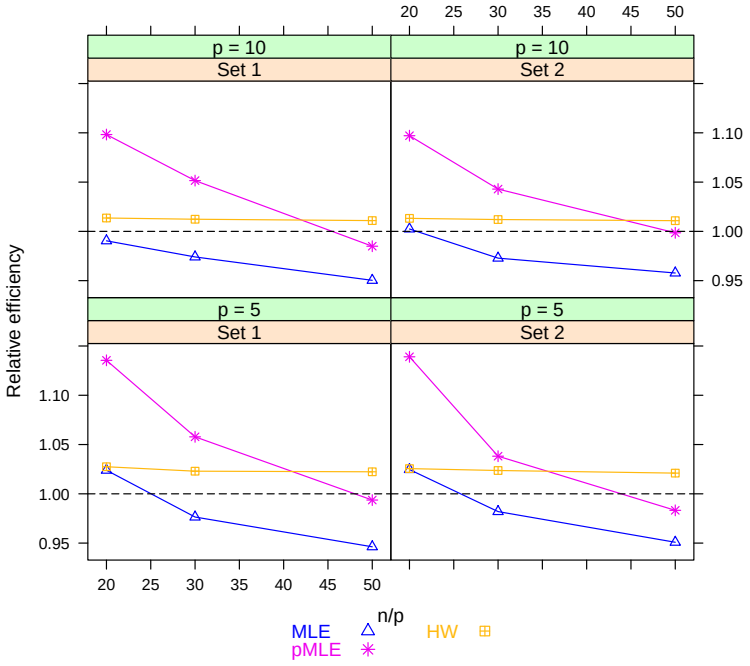

Figure 4.3: Relative efficiency of the various method to the Bonferroni for continuous explanatory variables

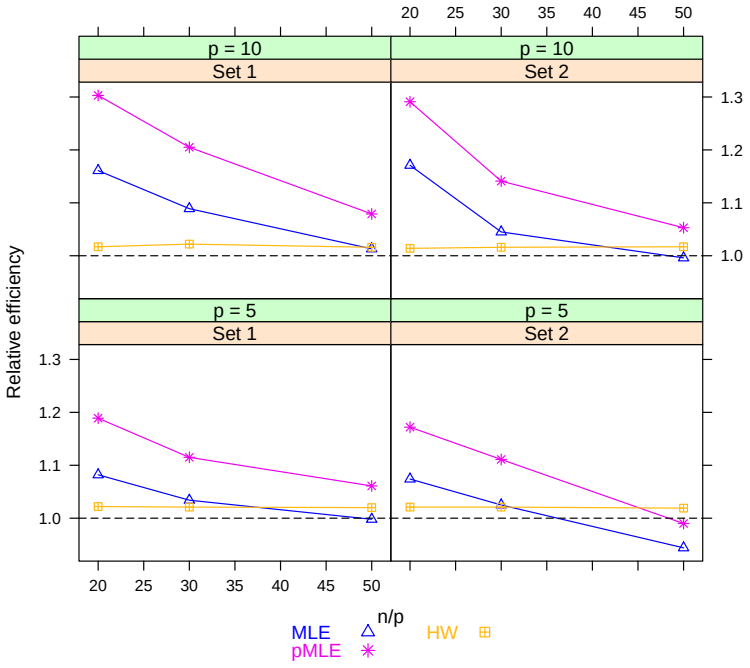

Figure 4.4: Relative efficiency of the various method to the Bonferroni for categorical explanatory variables

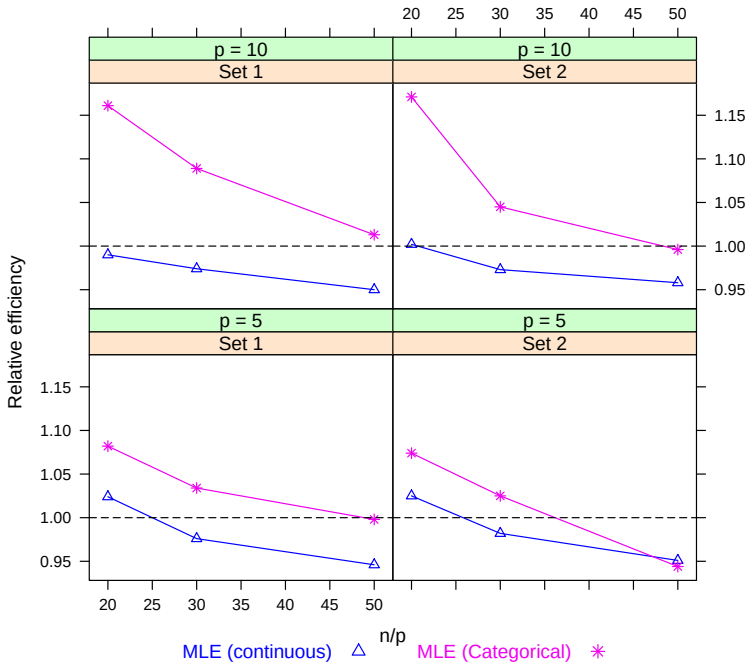

Figure 4.5: Comparison of relative efficiency with respect to MLE based on type of explanatory variable

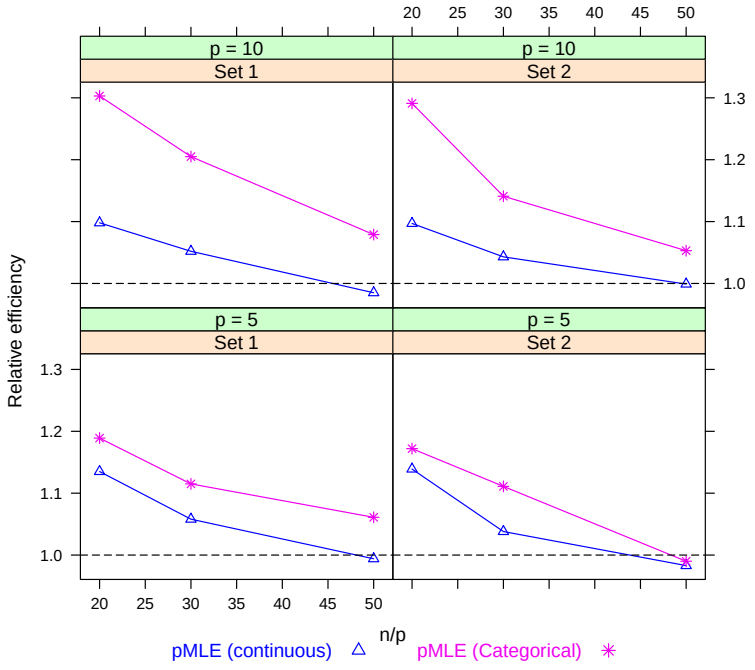

Figure 4.6: Comparison of relative efficiency with respect to pMLE based on type of explanatory variable

Chapter 5

Discussion and Conclusion

5.1 Discussion

In modern medical and epidemiological studies, several hypotheses may be of interest and tested as a result. In such cases, an appropriate procedure for controlling multiplicity that is applicable and feasible in the considered situation has to be identified. In this study, we focused on developing a resampling-based multiple comparison procedure for contrasts of generalized linear model parameters. The results of the simulation study to compare the performance of the proposed method to existing methods are discussed in this chapter.

Analysis of the results in Tables 4.6 and 4.7 and Figures 4.1 to 4.4, indicate that for the categorical explanatory variable settings, both the MLE and pMLE resampling-based procedures have improved performance over the Bonferroni and the Hunter-Worsley methods for almost all the cases with the pMLE resampling-based procedure showing the strongest improved performance. Conversely, the same conclusions cannot be made for the continuous explanatory variable settings. In the continuous explanatory variable settings, the pMLE resampling-based procedure performed better than all other methods for cases where the sample size to the number of regression parameters is small, $n/p \leq 30$ in this study. The MLE resampling-based procedure however performs poorly in almost all the cases. In cases where the resampling-based method performs well, the pMLE-based improves over all the other methods. For both the continuous and categorical explanatory variable settings, the improvement of the pMLE-based over the MLE-based is due to the bias correction in the penalized maximum likelihood estimation. However, the use of a Bayesian prior (Jeffrey's prior in the case of the pMLE) improves estimation of the critical point dramatically in

strictly categorical models. The effect is less dramatic with continuous predictors though the use of the priors still improves estimation of the critical point. This explains the trends in Figure 4.6 on page 35.

As illustrated earlier in the previous chapters, GLMs with multiple categorical predictors could be separable and/or biased at small sample sizes, leading to unreasonable parameter estimates and large standard errors. In such situations, estimation of critical points using MLE resampling-based procedures would be problematic since the procedures directly depend on the parameter estimates. This is exactly what was observed in the simulation settings for the MLE resampling-based procedure. Estimation warnings occurred when utilizing the MLE in the simulation due to convergence problems. This explains the vast discrepancies in the MLE-based critical values and the relative efficiencies for the continuous and categorical explanatory variables as can be observed in Figure 4.5 on page 35

It is more difficult to rationalize the patterns of the pMLE-based and the MLE-based methods and the patterns of the pMLE-based and the Hunter-Worsley methods. However, Efron (2012) justifies the improvement in the performance of the pMLE-based procedure over the MLE-based procedure. Let $\triangle(\beta)$ be the deviance difference between parametric bootstrap replications of MLE, $\widehat{\beta}$ and the parameter β, $\triangle(\beta) = \dfrac{D(\beta, \widehat{\beta}) - D(\widehat{\beta}, \beta)}{2}$. Then, using the pMLE is equivalent to reweighted parametric bootstrap with weights proportional to $\exp\{\triangle(\beta)\}$. We utilized the pMLE to reduce skewness and bias in the distribution of the *max-t* type test statistic. This reduction in skewness can be seen in Figure 5.1. The figure shows a plot of the *max-t* type test statistic of one simulation replication. From the figure, it can be seen that the pMLE estimates the true distribution well; reducing skewness and bias, thus leading to a smaller critical value compared to the MLE estimates. Efron(2012) provides the overall result that using the pMLE make $\triangle(\beta)$ approach zero of order $O_p(n^{-1/2})$. In a repeated sampling situation, skewness goes to zero as $n^{-1/2}$, making the distributions based on the pMLE converge faster at this rate to the actual distribution of the *max-t* type test statistic. However, there is a lot of sampling variability in the

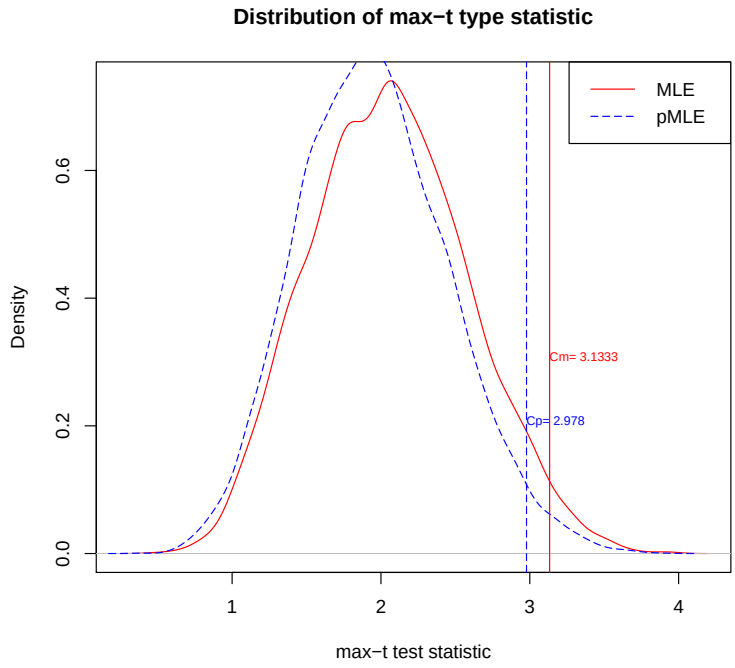

Figure 5.1: Plot of the distribution of the *max-t* type test statistic of one simulation replication

max-t distribution for the MLE and pMLE resampling-based procedures. This is due to the fact that these procedures are dependent on the estimators. In contrast, the Hunter-Worsley procedure is consistent due to its exact nature and it being less dependent on the estimators. Finally, recall that the Hunter-Worsley method bounds the union of the event of committing a type I error by the sum of the individual errors minus the maximum of the sum of the intersection of the $p-1$ consecutive event of committing a type I error. Thus, when n/p is small, then there are many more intersections of these events other than the 2-way overlap removed by the subtraction. This may explain the poor performance of the Hunter-Worsley in such cases.

5.2 Conclusion

Generalized linear models are often employed in epidemiological and medical research. Generally when these models are built, the primary interest centers on some estimated quantities of the model such as the odds ratios, the relative risks or the response probabilities. Various scenarios warrant the simultaneous estimation of these quantities. Most current simultaneous methods for controlling the family-wise error rate in estimating these quantities are probability-based and thus do not take into account the correlation and dependence structure of the test statistics. Additionally, the maximum likelihood estimation is known to produce bias estimates especially for categorical explanatory variables at small sample sizes and therefore it is not the most appropriate choice for parameter estimation. As a result, an appropriate method for simultaneous inference for multiple odds ratios, relative risks or response probabilities that account for the problems associated with the estimation of multiple categorical variables in the model and correlation structure of test statistics is warranted.

In this study, we extend a resampling-based method for simultaneous inference in linear models to simultaneous estimation of contrasts of GLM parameters. The method utilizes the penalized maximum likelihood estimation (pMLE) as an alternative to the maximum likelihood estimation (MLE) due to the bias nature and the small sample size problems associated with the MLE. The performance of the proposed resampling-based procedure is evaluated in comparison to other existing conservative methods using relative efficiency calculations. With regards to the estimation of the critical values, the pMLE-based procedure always produces a smaller critical value than the MLE-based procedure. The new method based on the penalized maximum likelihood estimation seems to outperform other existing methods for small sample size to number of parameters ratio for the case of continuous explanatory variables. However, it outperforms existing methods in almost all cases when the explanatory variable is categorical. This is due to the bias correction in the penalized maximum likelihood for the estimation of the parameters for cases where

the variables are categorical. It is established that the resampling-based procedure is dependent on the estimation procedure utilized. Thus, the proposed method based on the pMLE appears to improve over other existing methods in most situations and in those situations, it is less data-wasteful. However, due to the large sampling variability for the resampling-based procedure, we recommend utilizing the Hunter-Worsley procedure alongside the resampling-based procedure. The pMLE routines are available in R in the *brglm* package. The limitation to this method is that it is computationally intensive as compared to the other methods due to the resampling.

5.3 Recommendations for future work

In this study, categorical and continuous explanatory variables were considered separately in the simulations. It would be interesting to study the performance of the proposed method when both categorical and explanatory variables are considered simultaneously in the model.

Again, the use of the penalized maximum likelihood estimation improved the pMLE resampling based dramatically in the categorical explanatory variable scenario due to the fact that data separability is a very common issue in GLMs with multiple categorical predictors. Data separability is however not a common issue with continuous explanatory variable. Due to this, there is the need to consider reweighted parametric bootstrap for the continuous explanatory variable settings. This could eliminate the issue with the n/p ratio.

References

Agresti, Alan. 2007. *An introduction to categorical data analysis*. Wiley series in probability and mathematical statistics. Seco ed. Hoboken, NJ: Wiley-Interscience.

Bender, Ralf, and Stefan Lange. 2001. Adjusting for multiple testingwhen and how? *Journal of Clinical Epidemiology* 54 (4) (4): 343-9.

Bretz, Frank, and Hothorn, Ludwig A. 2003. Comparison of exact and resampling based multiple testing procedures. *Communications in Statistics: Simulation & Computation* 32 (2) (05): 461.

Division of Health, National Health Interview Survey (NHIS) public use data release NHIS survey description. Hyattsville, MD: Division of Health Interview Statistics, National Center for Health Statistics; 2009.

Efron, Bradley. 1979. Bootstrap methods: Another look at the jackknife. *The Annals of Statistics* 7 (1) (Jan.): 1-26, http://www.jstor.org/stable/2958830.

Efron, Bradley. 2012. Bayesian inference and the parametric bootstrap. *The Annals of Applied Statistics* 6 (4) (December): 1971-97, http://www.jstor.org/stable/41713502.

Efron, Bradley. 1997. The length heuristic for simultaneous hypothesis tests. *Biometrika* 84 (1) (Mar.): 143-57, http://www.jstor.org/stable/2337562.

Firth, David. 1993. Bias reduction of maximum likelihood estimates. *Biometrika* 80 (1) (Mar.): 27-38, http://www.jstor.org/stable/2336755.

Fisher, Nicholas I., and Peter Hall. 1990. On bootstrap hypothesis testing. *Australian Journal of Statistics* 32 (2): 177-90.

Ge, Youngchao, Sandrine Dudoit, and Terence P. Speed. 2003. Resampling-based multiple testing for microarray data analysis. *Test* 12 (1): 1-77.

Genz, Alan, and Frank Bretz. 2002. Comparison of methods for the computation of multivariate t probabilities. *Journal of Computational and Graphical Statistics* 11 (4) (Dec.): 950-71, http://www.jstor.org/stable/1391171.

Hochberg, Yosef, and Ajit C. Tamhane. 1987. *Multiple comparison procedures*. Wiley series in probability and mathematical statistics. New York: Wiley.

Hothorn, Torsten, Frank Bretz, and Peter Westfall. 2008. Simultaneous inference in general parametric models. *Biometrical Journal* 50 (3): 346-63.

Hsu, Jason. 1996. *Multiple comparisons*: Theory and methods CRC Press.

Hunter, David. 1976. An upper bound for the probability of a union. *Journal of Applied Probability* 13 (3) (Sep.): 597-603, http://www.jstor.org/stable/3212481.

Kounias, Eustratios G. 1968. Bounds for the probability of a union, with applications. *The Annals of Mathematical Statistics* 39 (6) (Dec.): 2154-8, http://www.jstor.org/stable/2239318.

McCann, Melinda, and Don Edwards. 1996. A path length inequality for the multivariate-t distribution, with applications to multiple comparisons. *Journal of the American Statistical Association* 91 (433) (Mar.): 211-6, http://www.jstor.org/stable/2291397.

McCullagh, P., and John A. Nelder. 1989. *Generalized linear models*. Monographs on statistics and applied probability. 2nd ed. Vol. 37. Boca Raton, Fla.: Chapman & Hall/CRC.

Moulton, Lawrence H., and Scott L. Zeger. 1991. Bootstrapping generalized linear models. *Computational Statistics & Data Analysis* 11 (1) (1): 53-63.

Nelder, J. A., and R. W. M. Wedderburn. 1972. Generalized linear models. *Journal of the Royal Statistical Society. Series A (General)* 135 (3): 370-84, http://www.jstor.org/stable/2344614.

Nichols, Thomas, and Satoru Hayasaka. 2003. Controlling the familywise error rate in functional neuroimaging: A comparative review. *Statistical Methods in Medical Research* 12 (5) (10): 419.

Olsson, Ulf. 2002. Generalized linear models. *An Applied Approach.Studentlitteratur, Lund* 18.

Pollard, Katherine S., and van der Laan, Mark J. 2003. Resampling-based multiple testing: Asymptotic control of type I error and applications to gene expression data.

R Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Serfling, Robert J. 2009. *Approximation theorems of mathematical statistics*. Vol. 162John Wiley & Sons.

Shaffer, Juliet Popper. 1995. Multiple hypothesis testing. *Annual Review of Psychology* 46 (1) (02): 561.

Šidák, Zbynek. 1967. Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association* 62 (318) (Jun.): 626-33, http://www.jstor.org/stable/2283989.

Tukey, John W. 1953. The Problem of Multiple Comparisons. Unpublished manuscript, Princeton University.

Wagler, Amy, and Melinda McCann. 2014. Improved simultaneous intervals for linear combinations of parameters from generalized linear models. *Journal of Statistical Computation and Simulation* (ahead-of-print): 1-19.

Wagler, Amy, and Melinda McCann. 2012. Bias-reduced simultaneous confidence bands on generalized linear models with restricted predictor variables. *Journal of Statistical Theory and Practice* 6 (2): 286-302.

Westfall, Peter H, and Young, Stanley S. 1993. *Resampling-based multiple testing: Examples and methods for p-value adjustment*. Vol. 279 John Wiley & Sons.

Westfall, Peter H. 2011. On using the bootstrap for multiple comparisons. *Journal of Biopharmaceutical Statistics* 21 (6) (Nov): 1187-205.

Worsley, K. J. 1982. An improved bonferroni inequality and applications. Biometrika 69 (2) (Aug.): 297-302, http://www.jstor.org/stable/2335402.

Yekutieli, Daniel, and Yoav Benjamini. 1999. Resampling-based false discovery rate controlling multiple test procedures for correlated test statistics. *Journal of Statistical Planning and Inference* 82 (1): 171-96.

Appendix A

R Codes for analyzing continuous explanatory variables

```
setwd("C:/Documents and Settings/jsakosa/Desktop/Thesis")
install.packages(c("brglm","multcomp","scatterplot3d","MVA",
                   "igraph","TSP","mvtnorm","arm"))
library(MASS)
library(brglm)
library(lattice)
library(multcomp)
library(TSP)
library(scatterplot3d )
library(MVA)
library(igraph)
library(mvtnorm)
library(arm)
# Create a new folder under the current working directory
dir.create("Results edit")
N=10000
reps=1000
k=10
ERROR.count <- matrix(NA,nrow=reps, ncol=6)
colnames(ERROR.count) <- c("MLE","hunter-worsley","bonferroni","n","k","set")
```

```
ERROR <- matrix(NA,nrow=reps, ncol=6)
colnames(ERROR.count) <- c("MLE","hunter-worsley","bonferroni","n","k","set")

pERROR.count <- matrix(NA,nrow=reps, ncol=6)
colnames(pERROR.count) <- c("pMLE","hunter-worsley","bonferroni","n","k","set")

pERROR <- matrix(NA,nrow=reps, ncol=6)
colnames(pERROR.count) <-c("pMLE","hunter-worsley","bonferroni","n","k","set")

#=========================================================================#
#Generating the random data                                               |
#n = the number of observations, k= number of parameters to be estimated.|
#~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~#
for (t in c(1,5)){
 for (k in c(5,10){
   for (n in c(200,300,500)){
     for (r in 1:reps) {
   ################### Generating the data ######################
       x.mat=matrix(NA, n, k)
       for (i in 1:k){
          x.mat[ , i] = matrix((rnorm(n,0,1)), n, 1)
       }
       beta <- matrix(rep(c(1,t),,k),nrow=k)
       eta <- x.mat%*%beta
       pi.resamp <- exp(eta)/(1+exp(eta))
       y <- rbinom(n,size=1,prob=pi.resamp)
       data <- data.frame(y,x.mat)
```

```
 ################### Computing beta hats ####################
 fit_mle <- tryCatch(glm(y~.+0,data=data,family="binomial"),error=function(e) e)
 if (inherits(fit_mle,"error")|sum(is.na(fit_mle$coefficients))>0) next
 fit_pmle <- tryCatch(brglm(y~.+0,data=data,family="binomial", br.maxit=500)
                             ,error=function(e) e)
 if (inherits(fit_pmle,"error")|sum(is.na(fit_mle$coefficients))>0) next
 beta_mle <-matrix(coef(fit_mle),ncol=1)
 beta_pmle <-matrix(coef(fit_pmle),ncol=1)
 infl=influence(fit_mle,do.coef=F)
 pr.mle<-infl$pear.res/(infl$sigma*sqrt(1-infl$hat))
 infl=influence(fit_pmle,do.coef=F)
 pr.pmle<-infl$pear.res/(infl$sigma*sqrt(1-infl$hat))
 var_mle <- vcov(fit_mle); var_pmle <- vcov(fit_pmle)

 ################### Resampling Method ####################
# FUNCTION resamp.boot() is used for resampling (bootstrapping)
resamp.boot <-function(data=data,r.mle=r.mle,coef.mle=coef.mle,
                         r.pmle=r.pmle,coef.pmle=coef.pmle){
               pr.b <- sample(r.mle,length(r.mle),replace=T)
               z <- as.matrix(data[-1])
               eta.boot <- pr.b
               pi.boot <- exp(eta.boot)/(1+exp(eta.boot))
               y.boot <- rbinom(n,size=1,prob=pi.boot)
               data.boot <- data.frame(y.boot,z)
               fit.b.mle <- glm(y.boot~.+0,data=data.boot,family="binomial")
               pr.b=sample(r.pmle,length(r.pmle),replace=T)
               z <- as.matrix(data[-1])
```

```
                    eta.boot<-pr.b
                    pi.boot<-exp(eta.boot)/(1+exp(eta.boot))
                    y.boot <- rbinom(n,size=1,prob=pi.boot)
                    data.boot<-data.frame(y.boot,z)
                    fit.b.pmle<-brglm(y.boot~.+0,data=data.boot,family="binomial")
                    beta_B<-list(beta.pmle=coef(fit.b.pmle),v_pmle=vcov(fit.b.pmle)
                                 ,beta.mle=coef(fit.b.mle),v_mle=vcov(fit.b.mle))
                  return(beta_B)
    }

#example call to new boot function
resamp.boot(data,pr.mle,beta_mle,pr.pmle,beta_pmle)
m<-1:k
names(m) <- paste("group",1:k,sep="")
contrast <- contrMat(m,type="Tukey",base=1)
tmax_pmle <- matrix(NA,nrow=N,ncol=1)
tmax_mle  <- matrix(NA,nrow=N,ncol=1)
cbeta <- contrast%*%as.matrix(beta)
for (i in 1:N){
   beta_B <- resamp.boot(data,pr.mle,beta_mle,pr.pmle,beta_pmle)
   cbeta_pmle<-contrast%*%as.matrix(beta_B[[1]])
   cbc_pmle <- contrast%*%as.matrix(beta_B[[2]])%*%t(contrast)
   t.pmle <- (cbeta_pmle)/(sqrt(diag(cbc_pmle)))
   cbeta_mle<-contrast%*%as.matrix(beta_B[[3]])
   cbc_mle <- contrast%*%as.matrix(beta_B[[4]])%*%t(contrast)
   t.mle <- (cbeta_mle)/(sqrt(diag(cbc_mle)))
   beta.final <- cbind(beta_B[[1]],beta_B[[3]])
   colnames(beta.final) <- c("beta_pmle.boot","beta_mle.boot")
```

```
   tmax_pmle[i,] <- max(abs(t.pmle))
   tmax_mle[i,]  <- max(abs(t.mle))
 }
#hist(tmax_pmle);hist(tmax_mle)
t.final <- cbind(tmax_pmle,tmax_mle)
colnames(t.final) <- c("tmax_pmle","tmax_mle")
write.table(data.frame(t.final), file = "./Results edit/tmax.csv",row.names=F,
            col.names=T,append=T,sep=",")

#################### Hunter-Worsley ######################
hw<-function(C,alpha,nu,V){
R=cov2cor(C%*%V%*%t(C))
acosR<-acos(R)
####### Step One: Find minimum spanning tree ####################
dis<-graph.adjacency(acosR,mode="max",weighted=TRUE)
mst <- minimum.spanning.tree(dis)
lay <- layout.reingold.tilford(dis, mode="all")
#plot(mst, layout=lay)
####### Step Two: Numerical Integration Method######
p<-nrow(C)
r=qr(C)$rank
intg<- function(d)
{
fx <- function(x){
B<-p*pf(((d*x)^(-2)-1)/(r-1), r-1, 1)
if (r==2) {for (i in 1:(p-1)) {
phi<-E(mst)$weight[i]
B<-B-max(-phi/pi+2*acos(x*d)/pi,0)}
```

```
}
else {
for (i in 1:(p-1)) {
phi<-E(mst)$weight[i]
gx<-function(w)
{acos(d*x*sqrt(1/(1-w)))*w^(r/2-2)}
#print(d)
gx<-Vectorize(gx)
if (((cos(phi/2)/(x*d))^2-1) >0 ) {
B<-B+phi/pi*pf(2*max(0,
                   (cos(phi/2)/(x*d))^2-1)/(r-2),r-2,2)-(r-2)/pi*(integrate(gx,
    0,1-1/(1+max(0,(cos(phi/2)/(x*d))^2-1)))$value) }
else B<-B+phi/pi*pf(2*max(0,(cos(phi/2)/(x*d))^2-1)/(r-2),r-2,2)
}
}
return (B*df(r*x^2, nu, r)*2*r*x)
}
fx<-Vectorize(fx)
gf <- function(k) {integrate(fx, lower=0, upper=1/k)$value}
#d=qt(1-.05,nu);d=sqrt(r*qf(1-alpha,r, nu))
return (gf(d)-alpha)}
######### Step Three:  Root Finding Algorithm ############
secant <- function(fun, x0, x1, tol=1e-4, niter=500){
for ( i in 1:niter ) {#fun=intg
x2 <- x1-fun(x1)*(x1-x0)/(fun(x1)-fun(x0))
if (abs(fun(x2)) < tol)
return(x2)
x0 <- x1
```

```
x1 <- x2
            #print(c(x0,x1,x2))
           }
           stop("exceeded allowed number of iteractions")
        }
        secant(intg, x0=qt(1-alpha/2,nu), x1=qt(1-alpha/(2*choose(k,2)),nu))
    }

################Parameter Setup################
################### resampling ####################
t.pmle <- quantile(tmax_pmle,0.975)
t.mle  <- quantile(tmax_mle,0.975)
out <-cbind(mle=t.mle,pmle=t.pmle)
################### HW ####################
V_hat<-beta_B[[2]]
alpha=0.05
nu=n-k
hw.quantile <- hw(contrast,alpha,nu,V_hat)
bf=qt(1-alpha/(2*choose(k,2)),nu)
hw=hw.quantile
out <- cbind(out,hw=hw.quantile,bf=bf)
out<-cbind(out,n=n,k=k,set=length(unique(beta)))
names <- cbind("MLE","pMLE","hunter worsley","bonferroni",
                      "sample size","k","set")
write.table(data.frame(out), file = "./Results edit/critical points.csv",
                    row.names=F, col.names=F,append=T,sep=",")
write.table(names,file="./Results edit/critical points-header.txt",
                    col.names=F,row.names=F)
```

```
print(out)

################### Relative Efficiency ####################
mle.rel_eff  <- (bf^2)/(t.mle^2)
pmle.rel_eff <- (bf^2)/(t.pmle^2)
hw.rel_eff   <- (bf^2)/(hw^2)
rel_eff <- cbind(mle.rel_eff=mle.rel_eff,pmle.rel_eff=pmle.rel_eff,
            hw.rel_eff=hw.rel_eff,n,set=length(unique(beta)))
names1 <- cbind("MLE rel efficiency", "pMLE rel efficiency",
              "hw rel efficiency",n,"set")
write.table(data.frame(rel_eff),file="./Results edit/relative efficiency.csv",
                    row.names=F, col.names=F,append=T,sep=",")
write.table(names1,file="./Results edit/relative efficiency-header.txt",
                     col.names=F,row.names=F)
print(rel_eff)

################### Error rates ####################
#browser()
cbeta <- contrast%*%as.matrix(beta)
cb_mle  <- contrast%*%beta_mle
cb_pmle <- contrast%*%beta_pmle
se_mle  <- sqrt(diag((contrast%*%var_mle%*%t(contrast))))
se_pmle <- sqrt(diag((contrast%*%var_pmle%*%t(contrast))))

############## Error rates using mle approach ###############
rmlel=(cb_mle-(t.mle*se_mle));     rmleu=(cb_mle+(t.mle*se_mle))
hmlel=(cb_mle-(hw*se_mle));        hmleu=(cb_mle+(hw*se_mle))
bmlel=(cb_mle-(bf*se_mle));        bmleu=(cb_mle+(bf*se_mle))
```

```
rmle.CI <- cbind(L=rmlel,U=rmleu)
hmle.CI <- cbind(L=hmlel,U=hmleu)
bmle.CI <- cbind(L=bmlel,U=bmleu)
mle.CI=cbind(rmle.CI,hmle.CI,bmle.CI)
colnames(mle.CI) <- c("mle.L","mle.U","hunter.L","hunter.U",
                              "bonferroni.L","bonferroni.U",)
write.table(mle.CI,file="./Results edit/MLE CI.csv",col.names=T,
                    row.names=F,append=T, sep=",")
rmle.error <- ifelse(cbeta<rmlel | cbeta>rmleu,1,0)
hmle.error <- ifelse(cbeta<hmlel | cbeta>hmleu,1,0)
bmle.error <- ifelse(cbeta<bmlel | cbeta>bmleu,1,0)
sum.rmle <- sum(rmle.error)
sum.hmle <- sum(hmle.error)
sum.bmle <- sum(bmle.error)
ERROR.count[r,1] <- sum.rmle
ERROR.count[r,2] <- sum.hmle
ERROR.count[r,3] <- sum.bmle
ERROR.count[r,4] <- n
ERROR.count[r,5] <- k
ERROR.count[r,6] <- length(unique(beta))
error.rmle <- ifelse(sum.rmle==0,0,1)
error.hmle <- ifelse(sum.hmle==0,0,1)
error.bmle <- ifelse(sum.bmle==0,0,1)
ERROR[r,1] <- error.rmle
ERROR[r,2] <- error.hmle
ERROR[r,3] <- error.bmle
ERROR[r,4] <- n
ERROR[r,5] <- k
```

```
ERROR[r,6] <- length(unique(beta))

############## Error rates using brglm approach ################
rpmlel=(cb_pmle-(t.pmle*se_pmle));    rpmleu=(cb_pmle+(t.pmle*se_pmle))
hpmlel=(cb_pmle-(hw*se_pmle));        hpmleu=(cb_pmle+(hw*se_pmle))
bpmlel=(cb_pmle-(bf*se_pmle));        bpmleu=(cb_pmle+(bf*se_pmle))
rpmle.CI <- cbind(L=rpmlel,U=rpmleu)
hpmle.CI <- cbind(L=hpmlel,U=hpmleu)
bpmle.CI <- cbind(L=bpmlel,U=bpmleu)
pmle.CI=cbind(rpmle.CI,hpmle.CI,bpmle.CI)
colnames(pmle.CI) <- c("pmle.L","pmle.U","hunter.L","hunter.U",
                               "bonferroni.L","bonferroni.U")
write.table(pmle.CI,file="./Results edit/pMLE CI.csv",col.names=T,
                    row.names=F,append=T, sep=",")
rpmle.error <- ifelse(cbeta<rpmlel | cbeta>rpmleu,1,0)
hpmle.error <- ifelse(cbeta<hpmlel | cbeta>hpmleu,1,0)
bpmle.error <- ifelse(cbeta<bpmlel | cbeta>bpmleu,1,0)
sum.rpmle <- sum(rpmle.error)
sum.hpmle <- sum(hpmle.error)
sum.bpmle <- sum(bpmle.error)
pERROR.count[r,1] <- sum.rpmle
pERROR.count[r,2] <- sum.hpmle
pERROR.count[r,3] <- sum.bpmle
pERROR.count[r,4] <- n
pERROR.count[r,5] <- k
pERROR.count[r,6] <- length(unique(beta))
error.rpmle <- ifelse(sum.rpmle==0,0,1)
error.hpmle <- ifelse(sum.hpmle==0,0,1)
```

```
error.bpmle <- ifelse(sum.bpmle==0,0,1)
pERROR[r,1] <- error.rpmle
pERROR[r,2] <- error.hpmle
pERROR[r,3] <- error.bpmle
pERROR[r,4] <- n
pERROR[r,5] <- k
pERROR[r,6] <- length(unique(beta))
   }
   ERROR.rate <- c((apply(ERROR[,1:3],2,sum)*(1/reps)),n,k,length(unique(beta)))
   print(ERROR)
   print(ERROR.rate)
   pERROR.rate <- c((apply(pERROR[,1:3],2,sum)*(1/reps)),n,k,length(unique(beta)))
   print(pERROR)
   print(pERROR.rate)
   write.table(ERROR.count,file="./Results edit/MLE Errors.csv",
               col.names=T,row.names=F,append=T, sep=",")
   write.table(t(data.frame(ERROR.rate)),file="./Results edit/MLE Error rates.csv",
               col.names=F,row.names=F,append=T, sep=",")
   write.table(pERROR.count,file="./Results edit/pMLE Errors.csv",
               col.names=T,row.names=F,append=T, sep=",")
   write.table(t(data.frame(pERROR.rate)),file="./Results edit/pMLE Error rates.csv",
               col.names=F,row.names=F,append=T, sep=",")
    }
  }
}
```

Appendix B

Generating data for categorical explanatory variables

```
k=10
x=rbinom(n,size=(k-1),prob=.5)
x.mat=matrix(NA, n, k)
x.mat[x==0,]=c(1,0,0,0,0,0,0,0,0,0)
x.mat[x==1,]=c(0,1,0,0,0,0,0,0,0,0)
x.mat[x==2,]=c(0,0,1,0,0,0,0,0,0,0)
x.mat[x==3,]=c(0,0,0,1,0,0,0,0,0,0)
x.mat[x==4,]=c(0,0,0,0,1,0,0,0,0,0)
x.mat[x==5,]=c(0,0,0,0,0,1,0,0,0,0)
x.mat[x==6,]=c(0,0,0,0,0,0,1,0,0,0)
x.mat[x==7,]=c(0,0,0,0,0,0,0,1,0,0)
x.mat[x==8,]=c(0,0,0,0,0,0,0,0,1,0)
x.mat[x==9,]=c(0,0,0,0,0,0,0,0,0,1)
```

Appendix C

Clinical Trial with Sparse Data

Table C.1: Clinical Trial Relating Treatment (X) to Response (Y) for Five Centers (Z), with XY and YZ Marginal Tables

		Response (Y)		YZ Marginal	
Center (Z)	Treatment (X)	Success	Failure	Success	Failure
1	Active Drug	0	5	0	14
	Placebo	0	9		
2	Active Drug	1	12	1	22
	Placebo	0	10		
3	Active Drug	0	7	0	12
	Placebo	0	5		
4	Active Drug	6	3	8	9
	Placebo	2	6		
5	Active Drug	5	9	7	21
	Placebo	2	12		
XY Marginal	Active Drug	12	36		
	Placebo	4	42		

Source: Agresti (2007)

Curriculum Vitae

Josephine Sarpong Akosa, the second child and first daughter to Kwaku Sarpong Akosa and Comfort Sarpong Akosa, was born in Kumasi, Ghana on March 9, 1988. She graduated from Holy Child School, Cape Coast, Ghana in the spring of 2006. In the fall of 2007, she was admitted to the Kwame Nkrumah University of Science and Technology (KNUST) where she pursued a bachelor's degree in Mathematics.

At KNUST, she was actively involved in extra-curriculum activities which enhanced her leadership skills. After graduating with first class honors, Miss Akosa was appointed to serve in the Mathematics Department, KNUST as a teaching assistant for the 2011/2012 academic year in fulfillment of her national service responsibility.

In the fall 2012, she began her Master's degree program in Statistics at the University of Texas at El Paso, Texas, United States. While pursuing her master's degree, she worked as a Graduate Teaching Assistant in the Mathematical Sciences Department until May 2014.

Miss Akosa is currently working as an intern as a Master's Research Assistant at the Center for Institutional Evaluation, Research and Planning, UTEP.

She plans to continue her studies in a Ph.D. Statistics program at the Oklahoma State University, Oklahoma, United States in Fall 2014.

Present address: 806 West Yandell Dr.
El Paso, Texas 79902