# INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

**The quality of this reproduction is dependent upon the quality of the copy submitted.** Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

# An Evaluation of Saddlepoint Approximations in the Generalized Linear Model

by

Robert William Platt

A dissertation submitted in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

University of Washington

1996

Approved by Norman Breslow
(Chairperson of Supervisory Committee)

Program Authorized
to Offer Degree Biostatistics

Date 13 September 1996

UMI Number: 9716898

---

UMI Microform 9716898

---

UMI
300 North Zeeb Road
Ann Arbor, MI 48103

Signature Robert

Date 09/16/96

University of Washington

Abstract

# An Evaluation of Saddlepoint Approximations in the Generalized Linear Model

by Robert William Platt

Chairperson of Supervisory Committee: *Professor Norman E. Breslow*
*Department of Biostatistics*

Higher order asymptotic methods based on the saddlepoint approximation to a density provide fast and accurate approximate inference in a variety of situations. The double saddlepoint approximation to the exact conditional density and distribution function in particular is useful in generalized linear model problems which are common in biostatistical applications. These methods give approximations to exact inference that are in general accurate and computationally much less intense than the exact methods. In this work we fully review the theory of the saddlepoint approximation from first principles and then develop and evaluate the approximation in two common discrete problems. Log odds ratio regression in stratified case control studies is an example of a problem where the number of nuisance parameters can increase with sample size, and conditioning on sufficient statistics for the nuisance parameters is essential for asymptotically unbiased inference. Investigation of asymptotic issues and numerical simulation demonstrate that for this model, the saddlepoint approximation accurately approximates exact conditional inference both for point estimation and confidence intervals. For the test for trend in a sequence of binomial random variables, numerical evidence points to the saddlepoint method being an effective approximation to conditional inference in almost all situations. The third part of the work involves applying saddlepoint inference to tests of higher order interaction in logistic regression. We implement the saddlepoint approximation for this problem and through simulation demonstrate that the properties of the saddlepoint approximation are significantly better than those of the commonly used unconditional methods. Finally, we examine and extend the theory for saddlepoint inference in the generalized linear model.

# TABLE OF CONTENTS

**List of Tables** v

**List of Figures** vii

**Chapter 1: Introduction** 1

1.1 Motivation . . . . . . . . . . . . . . . . . . . . . . . . . . . . 1

1.2 Literature Review . . . . . . . . . . . . . . . . . . . . . . . . . 3

1.3 Outline of the Dissertation . . . . . . . . . . . . . . . . . . . . 5

**Chapter 2: The Saddlepoint Approximation** 7

2.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . 7

2.2 The Edgeworth Approximations . . . . . . . . . . . . . . . . . . 7

2.2.1 Univariate Densities . . . . . . . . . . . . . . . . . . . . . 8

2.2.2 Multivariate Densities . . . . . . . . . . . . . . . . . . . . 10

2.3 The Saddlepoint Expansions for Densities . . . . . . . . . . . . . 12

2.3.1 The Exponential Tilt . . . . . . . . . . . . . . . . . . . . 13

2.3.2 The Saddlepoint Approximation For Univariate Densities . . . 15

2.3.3 Multivariate Saddlepoint Expansion . . . . . . . . . . . . . 17

2.3.4 Double Saddlepoint Approximation to the Conditional Density 18

2.4 Approximate Tail Probabilities . . . . . . . . . . . . . . . . . . 19

2.4.1 The Tail Probability – Univariate Densities . . . . . . . . . . 19

2.4.2 Tail Probability – Complex Variable Derivation . . . . . . . . 24

2.4.3 Double Saddlepoint Approximation – Tail Probability . . . . . 28

**Chapter 3: Applications to Generalized Linear Models** 34
3.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 34
3.2 The Density and the Modified Profile Likelihood . . . . . . . . . . . . 34
3.3 The Approximate Tail Probability . . . . . . . . . . . . . . . . . . . . 36
  3.3.1 The Sequential Saddlepoint Approximation . . . . . . . . . . . 41
3.4 Applied Work in the Exponential Family and GLM . . . . . . . . . . 42
  3.4.1 Theoretical Work in the Exponential Family and GLM . . . . 43

**Chapter 4: Application to Log Odds Ratio Regression** 45
4.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 45
4.2 The Model . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 46
  4.2.1 Unconditional Inference . . . . . . . . . . . . . . . . . . . . 47
  4.2.2 Conditional Inference . . . . . . . . . . . . . . . . . . . . . 48
  4.2.3 Alternative Methods . . . . . . . . . . . . . . . . . . . . . . 49
  4.2.4 The Double Saddlepoint Approximation . . . . . . . . . . . . 50
  4.2.5 The Direct Saddlepoint Approximation . . . . . . . . . . . . . 52
  4.2.6 Log Odds Ratio Regression . . . . . . . . . . . . . . . . . . . 53
4.3 Asymptotic Bias and Variance of the Estimators . . . . . . . . . . . . 55
  4.3.1 Asymptotic Model I . . . . . . . . . . . . . . . . . . . . . . 55
  4.3.2 Asymptotic Model II . . . . . . . . . . . . . . . . . . . . . . 59
4.4 Small Sample Performance of the Approximations . . . . . . . . . . . 68
  4.4.1 Design of the Simulation Study . . . . . . . . . . . . . . . . . 71
  4.4.2 Results of the Simulation Study . . . . . . . . . . . . . . . . 73

**Chapter 5: Application to the Test for Trend in a Sequence of Binomial Variables** 82
5.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 82

5.2 The Model . . . . . . . . . . . . . . . . . . . . . . . . . . . . 83
  5.2.1 Unconditional Inference . . . . . . . . . . . . . . . . . . . . 83
  5.2.2 Conditional Inference . . . . . . . . . . . . . . . . . . . . . 84
  5.2.3 Double Saddlepoint Inference . . . . . . . . . . . . . . . . . 85
  5.2.4 Direct Saddlepoint Inference . . . . . . . . . . . . . . . . . 86
5.3 Asymptotic Distribution Theory of the Estimators . . . . . . . . . . . 87
5.4 Small Sample Performance of the Methods . . . . . . . . . . . . . . 88
  5.4.1 Design of the Simulation Studies . . . . . . . . . . . . . . . 89
  5.4.2 Results of the Simulation Studies . . . . . . . . . . . . . . . 90

**Chapter 6: High Order Interaction in Logistic Regression – A Case Study 99**
6.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . 99
6.2 The Model . . . . . . . . . . . . . . . . . . . . . . . . . . . . 100
  6.2.1 Unconditional Inference . . . . . . . . . . . . . . . . . . . . 100
  6.2.2 Conditional Inference . . . . . . . . . . . . . . . . . . . . . 101
  6.2.3 Double Saddlepoint Inference . . . . . . . . . . . . . . . . . 102
6.3 Data Analysis . . . . . . . . . . . . . . . . . . . . . . . . . . . 103
  6.3.1 The Data . . . . . . . . . . . . . . . . . . . . . . . . . . . 103
  6.3.2 Analysis . . . . . . . . . . . . . . . . . . . . . . . . . . . 104
  6.3.3 Results . . . . . . . . . . . . . . . . . . . . . . . . . . . . 106
6.4 Simulation Study . . . . . . . . . . . . . . . . . . . . . . . . . 108
  6.4.1 Design of Simulation Study . . . . . . . . . . . . . . . . . . 108
  6.4.2 Results of Simulation Study . . . . . . . . . . . . . . . . . . 109

**Chapter 7: Discussion of General Results in the GLM 113**
7.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . 113

7.2 Previous Theoretical Work . . . . . . . . . . . . . . . . . . . . . . 113
7.3 Original Work . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 114
7.4 Conclusions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 120

**Chapter 8: Conclusions** **122**
8.1 Summary of the Dissertation . . . . . . . . . . . . . . . . . . . . . 122
8.2 Future Work . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 124
8.3 Conclusions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 126

**Bibliography** **127**

**Appendix A: Some Theory of Complex Variables** **134**

**Appendix B: Profile Likelihood Modifier Derivatives – Log Odds Ratio Regression** **138**

# LIST OF TABLES

4.1 Ille-et-Vilaine Data – Results of Saddlepoint Analysis – Log-Odds Ratio Regression . . . 54
4.2 Ille-et-Vilaine Data – Results of Saddlepoint Analysis - Common OR 54
4.3 Asymptotic Means of Double Saddlepoint Maximum Likelihood Estimator - OR = 1.2 and 1.5 . . . 69
4.4 Asymptotic Means of Double Saddlepoint Maximum Likelihood Estimator - OR = 2.0 and 5.0 . . . 70
4.5 Simulation Parameters . . . 72
4.6 Simulation Design . . . 74
4.7 Log Odds Ratio Simulation Results – Bias and Variance . . . 75
4.8 Log Odds Ratio Simulation Results – 95% Coverage Probability . . . 76
4.9 Log Odds Ratio Simulation Results – Misclassification . . . 77
4.10 Log Odds Ratio Simulation Results – Mean Relative Errors of 95% Confidence Bounds . . . 78
4.11 Log Odds Ratio Simulation Results – Maximum Relative Errors of 95% Confidence Bounds . . . 79

5.1 Data on Alcohol Consumption and Malformation . . . 83
5.2 Malformation Data – Analysis . . . 87
5.3 Binomial Denominators For Trend Test Simulations . . . 89
5.4 Trend Simulations – Mean Estimates of Δ . . . 91
5.5 Trend Simulations – Mean Estimates of Δ . . . 92

5.6 Trend Simulations – Results for Dataset 1 . . . . . . . . . . . . . . . . 93
5.7 Trend Simulations – Results for Dataset 2 . . . . . . . . . . . . . . . . 94
5.8 Trend Simulations – Results for Dataset 3 . . . . . . . . . . . . . . . . 95
5.9 Trend Simulations – Results for Dataset 4 . . . . . . . . . . . . . . . . 96

6.1 APOE Data Summary . . . . . . . . . . . . . . . . . . . . . . . . . . 104
6.2 Results of APOE Analysis – Main Effects . . . . . . . . . . . . . . . . . 106
6.3 Results of APOE Analysis – Interactions . . . . . . . . . . . . . . . . . 106
6.4 Results of APOE Analysis – APOE as a Factor . . . . . . . . . . . . . 107
6.5 Power of Tests for Four-Way Interaction . . . . . . . . . . . . . . . . . 110
6.6 Agreement Between Tests of Four-Way Interaction . . . . . . . . . . . 111
6.7 Coverage Probabilities – Four-Way Interaction Parameter . . . . . . . 112

# LIST OF FIGURES

# ACKNOWLEDGMENTS

I wish to express gratitude to my advisor, Norman Breslow, for his insights and guidance. It was a pleasure to work with him. I also thank David Yanez and Steve Self for sitting on my reading committee and for their service on my committee. I am indebted to Jon Wellner who taught me statistical theory in such a skillful manner and served on my committee, and all the others who taught me courses during my years at Washington.

I would like to thank the ARCS Foundation of Seattle for their generous support of my graduate work. The ARCS Fellowship eased my financial burden considerably over the past three years.

My fellow students at the UW were a tremendous help and influence on me. Carl De Moor, Rich Holubkov, Chuck Spiekerman, Kim Siegmund, Jim Lynch, Peter Gilbert, Jim Lymp, Adam Olshen and a host of others made my time at Washington enjoyable and productive.

Finally I would like to thank my family. My parents have always given me the encouragement and opportunity to do my best. I want to especially thank my wife Jennifer. Her support and encouragement through my entire graduate career was invaluable.

# DEDICATION

**To Jennifer**

# Chapter 1

# INTRODUCTION

## *1.1 Motivation*

Inference in the exponential family in the presence of nuisance parameters is one of the most common problems in applied statistics. Ordinarily, standard first order asymptotic methods such as likelihood ratio/deviance tests, Wald (maximum likelihood) and score tests are used in these cases. For many problems, however, estimation of the nuisance parameters has an effect on estimation of the parameter of interest, and in situations with large numbers of nuisance parameters can lead to biased results. Exact conditional inference which conditions on appropriate sufficient statistics so as to eliminate nuisance parameters can be done and often is necessary for accurate inference. As computing becomes better and better, it becomes more feasible to do larger problems using the exact conditional distribution. However, higher-order asymptotic methods have the potential to provide a quick and accurate approximation to the exact conditional results which takes only slightly more effort than the standard first-order approximations. They can be easily programmed using standard packages like S-Plus and SAS and provide a quick comparison method to both assess the validity of the unconditional results and determine whether resorting to exact conditional analysis is required. Pierce and Peters (1992) suggest that compared to exact inference, "adequate asymptotic methods are less inhibiting, encouraging more thorough data analysis, and can often lead to clearer inferential summaries".

### Example 1.1 Log Odds Ratio Regression

In retrospective studies, examining the relationship between disease incidence and exposure to a potential risk factor, the parameter of interest is the odds ratio in a 2 × 2 table of individuals classified as diseased vs. non-diseased and exposed vs. non-exposed. To reduce the effects of confounding, the sample is usually divided into strata on the basis of baseline variables thought to be related to the disease and/or exposure under study (e.g., age). This leads to a sequence of 2 × 2 tables. Separate estimates of the odds ratio can be obtained for each table, and these estimates can be combined in a variety of ways (Breslow 1976; Gart 1962; Mantel and Haenszel 1959; McCullagh 1984; Zelen 1971) to yield overall estimates of the relative risk of disease in the exposed vs. non-exposed groups. The conditional maximum likelihood estimator of the odds ratio has been shown to be consistent even when the number of tables increases with sample size but is difficult to implement, while the unconditional maximum likelihood estimator is biased (Breslow, 1981). The saddlepoint approximation to the conditional likelihood may provide a nearly unbiased estimator which is easy to calculate using basic software.

### Example 1.2 Test for Trend in a $2 \times k$ Table

Often in the examination of a sequence of binomial probabilities, one wishes to test for the possibility of an increasing trend in the probabilities. This trend test can be done unconditionally using logistic regression, or conditionally on the marginal totals using conditional logistic regression. This latter method is often required when the binomial denominators are badly unbalanced (Agresti 1992; Graubard and Korn 1987). However, when the binomial denominators are unbalanced and some are quite large, the conditional analysis can be very difficult to do, as it requires enumeration of the entire reference distribution of the conditional statistic and hence calculation of a substantial number of binomial coefficients.

### Example 1.3 High Order Interactions in Logistic Regression

Genetic researchers and others are often interested in testing for high-order interactions in logistic regression. For example, in a study of Alzheimer's disease and the apolipoprotein-E gene, Jarvik et al (1996) studied the interaction of age, sex, family history of disease and genetic markers in a logistic model for a case-control study. The data for this study were quite sparse. Commonly used unconditional methods performed poorly and different asymptotic tests gave entirely different results. This situation will be considered using saddlepoint approximations and exact conditional logistic regression to determine if the saddlepoint corrections improve on first-order asymptotics.

**Example 1.4 Analysis of Large and Sparse $I \times J$ Tables**

Log-linear models are used to test for interaction in $I \times J$ tables. When data in the table are sparse and the table dimensions are large, the inference on interaction parameters from unconditional log-linear models and other first-order methods (such as Pearson's $\chi^2$) is in general poor, while exact conditional inference, which accounts for the sparseness in the data, can be extremely cumbersome for large tables. Work on examples of this type of dataset has been examined by Pierce and Peters (1992).

The thrust of this work will be to do a systematic analysis of higher-order asymptotic methods to determine how well they work in certain practical situations. Preliminary work by several authors has shown that in certain cases with many nuisance parameters the higher-order asymptotic inference can be poor. In other cases with many nuisance parameters, and yet in other cases where the number of nuisance parameters is relatively small but the data distribution is such that the information matrix for the parameter of interest from the unconditional likelihood is poorly conditioned, higher order methods appear to perform quite well.

## *1.2 Literature Review*

The saddlepoint approximation was first described by Daniels (1954). This work developed an approximation to the density and distribution of the mean of a set of random variables using an approximation to the inversion integral of the characteristic function. This approximation can be derived as an extension of the Edgeworth expansion. The saddlepoint approximation to a density and tail probability provided an accurate approximation, even in the tails of the density where many approximate densities break down.

Little work was done in this area until the 1970s and 1980s. Barndorff-Nielsen and Cox (1979) proposed the double saddlepoint approximation to conditional densities, derived by approximating both the full and marginal densities with saddlepoint approximations. A series of papers by Barndorff-Nielsen continued to explore this approximation both on and off the exponential family. Barndorff-Nielsen (1980, 1983) developed the modified profile likelihood. The modified profile likelihood is a generalization of the saddlepoint approximation which in the exponential family is the saddlepoint approximate likelihood but off the exponential family is an approximate likelihood requiring alternative justification. This provided an efficient way of approximating conditional densities.

Lugannani and Rice (1980) proposed a tail probability approximation which enabled calculation of the approximate cumulative distribution function in the univariate case. Skovgaard (1987) generalized this result to the double saddlepoint approximation. An alternate formula for the tail probability for the single and double saddlepoint approximation arises from the work of Barndorff-Nielsen (1986, 1990) and Jensen (1986).

Davison (1988) first applied these methods explicitly to the generalized linear model, considering the tail probability approximation of Skovgaard (1987) in the generalized linear model and noting that several features of the saddlepoint approximations correspond to easily obtainable and standard likelihood quantities in the generalized linear model, including the maximum likelihood estimate. Bedrick and Hill (1992) continued this work, examining three applied problems and comparing the exact to approximate densities. Fraser, Reid and Wong (1991) suggested the sequential saddlepoint approximation, which evaluated the tail probability approximations at the maximum modified profile likelihood estimate rather than at the maximum likelihood estimate. Pierce and Peters (1992) applied the tail probability method of Barndorff-Nielsen to the exponential family and problems of the sort being considered for this dissertation. Levin (1990) and Levin and Kong (1990) considered estimation using the modified profile likelihood in logistic regression models and proposed a correction to the modified likelihood to improve performance of the methods by better centering the approximate density and likelihood about the origin.

## *1.3 Outline of the Dissertation*

This dissertation will examine in some detail the double saddlepoint approximation in the exponential family/generalized linear model. Chapter 2 is a review of the saddlepoint approximation and higher order asymptotics, and chapter 3 reviews the development of applications of these asymptotics to the generalized linear model. The bulk of the work will examine three specific models where unconditional inference is commonly used but can give poor or erroneous results, while exact conditional inference is inconvenient. I will implement the saddlepoint approximation in a variety of forms and examine in these three models the theoretical properties of the likelihood approximation and approximate tail probabilities. The examination will consist primarily of simulation studies and numerical results comparing the saddlepoint approximate inference to exact conditional inference and unconditional inference.

The first model under consideration will be the model from Example 1.1, the study of the odds ratio in a set of $2 \times 2$ tables. For this model, the small sample properties of both the likelihood and the tail probability approximation will be derived. Since estimation of the odds ratio is a primary issue for this model, here we also examine the asymptotic bias and variance of estimators of the odds ratio derived from exact conditional, unconditional and saddlepoint methods. Asymptotics are done using two frameworks – where the number of nuisance parameters increases with sample size, and where it stays fixed. This will form chapter 4.

The second and third models are Examples 1.2 and 1.3, the trend test (chapter 5) and high-order interactions in logistic regression (chapter 6). For these two models the primary work is implementation and small sample comparison via simulation. Parameter estimation is of limited interest in both of these problems, but for the trend test parameter estimates are compared to exact and unconditional estimates.

Chapter 7 examines general asymptotic theory for the saddlepoint approximations in the exponential family in the case where the number of nuisance parameters increases with sample size. The work by Pierce and Peters (1992), which shows that the inference is poor at extreme values of the sufficient statistics, suggests that a theorem may in fact not be possible except under very restrictive conditions. However, this work will examine this possibility.

## Chapter 2

# THE SADDLEPOINT APPROXIMATION

### *2.1 Introduction*

In this chapter the background work for the saddlepoint approximation and the modified profile likelihood will be examined in detail. We start by reviewing the Edgeworth expansion. The next section reviews the tilted Edgeworth expansion or saddlepoint expansion and presents results for univariate and multivariate densities based on this expansion. From there, we develop approximations to the tail probability based on these expansions, and methods for coping with lattice or discrete data.

This review is based primarily on work by Daniels (1954, 1987), Barndorff-Nielsen (1980, 1983, 1986, 1990), Barndorff-Nielsen and Cox (1979, 1989), Skovgaard (1987), and a review by Reid (1988). A complete reference for much of the development of saddlepoint approximations and other expansions can be found in Kolassa (1994).

### *2.2 The Edgeworth Approximations*

Edgeworth expansions for densities date to the turn of the century. These expansions generalize the normal approximation for the density to more terms in a Taylor series. The development here is an expansion on the development by Barndorff-Nielsen and Cox (1989, Ch. 2). We first develop the univariate Edgeworth expansion, and then after defining some terms, develop the multivariate Edgeworth expansion for the density and distribution function of the mean of a set of random variables.

### *2.2.1 Univariate Densities*

**Theorem 2.2.1** *Given $X_1, \ldots, X_n$ independent and identically distributed with expectation $E(X_i) = \mu$, variance $Var(X_i) = \sigma^2$, and cumulants $\kappa_r$, expansions for the density and distribution of*

$$S_n^* = \frac{\Sigma(X_i - \mu)}{\sigma\sqrt{n}}$$

*are given by*

$$f_{S_n^*}(s) = \phi(s)\left\{1 + \frac{\rho_3 H_3(s)}{6\sqrt{n}} + \frac{\rho_4 H_4(s)}{24n} + \frac{\rho_3^2 H_6(s)}{72n} + O(n^{-3/2})\right\}$$

*and*

$$F_{S_n^*}(s) = \Phi(s) - \phi(s)\left\{1 + \frac{\rho_3 H_2(s)}{6\sqrt{n}} + \frac{\rho_4 H_3(s)}{24n} + \frac{\rho_3^2 H_5(s)}{72n} + O(n^{-3/2})\right\},$$

*where $\rho_r = \kappa_r/\sigma^r$ and $H_i(s)$ is the ith Hermite polynomial (defined below) evaluated at s.*

*Proof*. First note that the cumulant generating function for $S_n^*$ is given by

$$K_{S_n^*}(t) = -\sqrt{n}\mu t/\sigma + nK_X\left(\frac{t}{\sigma\sqrt{n}}\right). \tag{2.1}$$

Now expand the cumulant generating function for $X$ around 0:

$$K_X(t) = \mu t + 1/2\sigma^2 t^2 + 1/6\kappa_3 t^3 + 1/24\kappa_4 t^4 + \ldots, \tag{2.2}$$

then

$$K_{S_n^*}(t) = 1/2t^2 + \rho_3 t^3/(6\sqrt{n}) + \rho_4 t^4/(24n) + O(n^{-3/2}). \tag{2.3}$$

Hence, the moment generating function is given by

$$\begin{aligned} M_{S_n^*}(t) &= e^{\frac{1}{2}t^2} \exp(\rho_3 t^3/(6\sqrt{n}))\exp(\rho_4 t^4/(24n))\exp(O(n^{-3/2})) \\ &= e^{\frac{1}{2}t^2}\left\{1 + \rho_3 t^3/(6\sqrt{n}) + \rho_3^2 t^6/(72n) + \ldots)(1 + \rho_4 t^4/(24n) + \ldots\right\} \\ &= e^{\frac{1}{2}t^2}\left\{1 + \rho_3 t^3/(6\sqrt{n}) + \rho_4 t^4/(24n) + \rho_3^2 t^6/(72n) + O(n^{-3/2})\right\}. \end{aligned}$$

We invert this to derive an approximation for the density and distribution. To do this we need to introduce the Hermite polynomials. These are a set of orthogonal polynomials defined by the equation

$$\phi(x)H_r(x) = (-1)^r \frac{\partial^r \phi}{\partial x^r}. \tag{2.4}$$

Noting that

$$\begin{aligned}\int_{-\infty}^{\infty} e^{tx} \frac{\partial^r}{\partial x^r}\phi(x)dx &= e^{tx}\frac{\partial^{r-1}}{\partial x^{r-1}}\phi(x)|_{-\infty}^{\infty} - t\int_{-\infty}^{\infty} e^{tx}\frac{\partial^{r-1}}{\partial x^{r-1}}\phi(x)dx \\ &= -t\int_{-\infty}^{\infty} e^{tx}\frac{\partial^{r-1}}{\partial x^{r-1}}\phi(x)dx\end{aligned}$$

and that $H_0(x) = 1$, we derive that

$$\int_{-\infty}^{\infty} e^{tx} H_r(x)\phi(x)dx = (-1)^r t^r e^{\frac{1}{2}t^2}, \tag{2.5}$$

and use this and the uniqueness of moment generating functions to derive the density expansion:

$$f_{S_n^*}(s) = \phi(s)\left\{1 + \frac{\rho_3 H_3(s)}{6\sqrt{n}} + \frac{\rho_4 H_4(s)}{24n} + \frac{\rho_3^2 H_6(s)}{72n} + O(n^{-3/2})\right\}. \tag{2.6}$$

Term by term integration of (2.6) and the fact that

$$\int_{-\infty}^{s} H_r(x)\phi(x)dx = \phi(s)H_{r-1}(s), \tag{2.7}$$

which can be derived from the above equations about Hermite polynomials gives

$$F_{S_n^*}(s) = \Phi(s) - \phi(s)\left\{1 + \frac{\rho_3 H_2(s)}{6\sqrt{n}} + \frac{\rho_4 H_3(s)}{24n} + \frac{\rho_3^2 H_5(s)}{72n} + O(n^{-3/2})\right\}. \tag{2.8}$$

This completes the proof.

There are several problems with the Edgeworth expansion that make it impractical for use in statistics. The main problem is the quality of the approximation in the tails. The key is that the expansion terms and hence the error of the approximation depend on $s$. There exists a function $A$ such that

$$|f_{S_n^*}(s)/\widehat{f_{S_n^*}}(s) - 1| \leq A(s)/n \tag{2.9}$$

where $\widehat{fs_n^*}$ is the expansion in (2.6) up to terms of order $1/n$. However, in the tails this $A(s)$, which depends on $s$ through the Hermite polynomials, can become intolerably large for the Edgeworth expansion (Barndorff-Nielsen and Cox (1989)). The saddlepoint approximation avoids this problem by re-centering the approximation to perform better in the tails.

### *2.2.2 Multivariate Densities*

The multivariate Edgeworth expansion is considerably more complicated than the univariate expansion. The exposition below is based on that of Barndorff-Nielsen and Cox (1989), and takes advantage of their use of tensor notation. To discuss the multivariate Edgeworth expansion, it is important to understand the multivariate Hermite polynomials. The following section briefly explains some of the issues with these polynomials.

### *Digression – Multivariate Hermite Polynomials*

The multivariate Hermite polynomials are derived in the same manner as the univariate ones but the multivariate aspect adds some complications. The derivation is explained in some detail in Barndorff-Nielsen and Cox (1989). We first need a bit of notation to deal with the problem.

**Notation 2.2.1** *In general, $i, j, k$ etc. refer to indices running from 1 to some integer $m$. Also, when an index appears both as subscript and superscript in an expression, this implies summation over that index.*

The multivariate covariant Hermite polynomials are written as $h_{i_1,\dots,i_k}(x;\Sigma)$ where $x = (x_1, \dots x_m)^T$ is a vector and $\Sigma$ is the covariance matrix for a multivariate normal random variable with distribution $N_m(0,\Sigma)$. The polynomials are defined by the following equation:

$$\phi_m(x;\Sigma)h_{i_1,\dots,i_k}(x;\Sigma) = (-1)^k \frac{\partial}{\partial x^{i_1}} \cdots \frac{\partial}{\partial x_{i_k}} \phi_m(x;\Sigma) \qquad (2.10)$$

where $\phi_m(x;\Sigma)$ denotes the normal density of dimension $m$ with mean 0 and covariance matrix $\Sigma$. This is similar to the equation that defines the univariate Hermite polynomials. The following also applies in the multivariate case:

$$\int e^{\sum_{j=1}^{m} t_j x_j} h_{i_1,\ldots,i_k}(x;\Sigma)\phi_m(x;\Sigma)dx = t_{i_1}\ldots t_{i_k}\exp(-1/2t^T\Sigma t) \tag{2.11}$$

where $\Sigma_{ij}$ denotes the $i,j$th element of $\Sigma$. Let $\Sigma^{ij}$ denote the $i,j$th element of $\Sigma^{-1}$ and define $x_i^* = \sum_j \Sigma^{ij} x_j$. The first three Hermite polynomials are given by the following equations:

$$\begin{aligned} h_i(x;\Sigma) &= x_i^*, \\ h_{ij}(x;\Sigma) &= x_i^* x_j^* - \Sigma^{ij}, \\ h_{ijk}(x;\Sigma) &= x_i^* x_j^* x_k^* - \Sigma^{ij} x_k^* - \Sigma^{ik} x_j^* - \Sigma^{jk} x_i^*. \end{aligned}$$

We will use (2.11) in the same manner as in the univariate case to derive the Edgeworth expansion for the multivariate case.

### *The Expansion*

The Edgeworth expansion in the multivariate case is derived in the same manner as the univariate case. Let $X_1,\ldots,X_n$ be iid random variables with mean $\mu = (\mu_1,\ldots\mu_m)$ and cumulants $\kappa^{i_1,\ldots,i_k} = \partial K(t)/\partial t_{i_1}\ldots\partial t_{i_k}|_{t=0}$ where $K$ is the cumulant generating function. Let $S_n^* = \sum(X_i - \mu)/\sqrt{n}$ (note that we do not standardize the variance here as in the univariate case).

The cumulant generating function of $S_n^*$ is

$$\begin{aligned} K_{S_n^*}(t) &= nK_{X-\mu}(t/\sqrt{n}) \\ &= \sum_{\nu=2}^{\infty}\frac{1}{\nu!}\kappa^{i_1,\ldots,i_\nu} t_{i_1}\ldots t_{i_\nu} n^{-\nu/2+1}. \end{aligned} \tag{2.12}$$

The expansion proceeds exactly as in the univariate case, first converting to the moment generating function and then using equation (2.11) to derive an expansion

$$\begin{aligned} f_{S_n^*}(s) &= \phi_m(s;\kappa)\{1 + (1/6\sqrt{n})\kappa^{i,j,k}h_{ijk}(s;\kappa) + (1/24n)\kappa^{i,j,k,l}h_{ijkl}(s;\kappa) + \\ &\quad (1/72n)\kappa^{i,j,k}\kappa^{l,p,q}h_{ijklpq}(s;\kappa)\} + O(n^{-3/2}). \qquad (2.13) \end{aligned}$$

It is important to remember that the terms of the form $\kappa^i h_i$ are complicated sums of polynomials. For example, if $m = 2$ the term $\kappa^{i,j,k}h_{ijk}(x;\kappa)$ becomes

$$\kappa^{1,1,1}h_{111}(x;\kappa) + 3\kappa^{1,1,2}h_{112}(x;\kappa) + 3\kappa^{1,2,2}h_{122}(x;\kappa) + \kappa^{2,2,2}h_{222}(x;\kappa).$$

These sums get large quickly. It is convenient to write the expansion in the following form:

$$f_{S_n^*}(s) = \phi_m(s;\kappa)\left\{1 + Q_3(s,\kappa^{(3)})/\sqrt{n} + Q_4(s,\kappa^{(4)})/n + \ldots\right\}, \qquad (2.14)$$

where $Q_i(x,\kappa^{(i)})$ is a polynomial of degree $i$ containing cumulants up to order $i$.

## *2.3 The Saddlepoint Expansions for Densities*

The saddlepoint approximation was first derived as such by Daniels (1954). Daniels' paper derived the saddlepoint approximation using two methods.

The first method is the one which gives the approximation its name. This method for approximating the density of the mean involves writing the density as the inversion integral (inverse Fourier transform) of the characteristic function. The contour of integration of this integral can be deformed so that it passes through a saddlepoint. For large $n$, the integrand is negligible outside the immediate neighbourhood of the saddlepoint, and an expression for the density can be derived by expanding about the saddlepoint.

The second method, which we present here, does not involve complex analysis and is developed as a generalization of the Edgeworth expansion. This method is due to Daniels (1954), and is based on work by Cramér (1938) and Khinchin (1949). This method involved 'tilted' Edgeworth expansions, where the Edgeworth expansion is performed on a density made up of the original one multiplied by an exponential function. We consider a density which arises as the product of the density of interest and an exponential form. An Edgeworth expansion of this modified density is taken, and then after choice of appropriate constants, the expanded modified density is multiplied back by the exponential form. This gives an expansion for the original density which is identical to the saddlepoint expansion.

### *2.3.1 The Exponential Tilt*

The exponential tilt was first proposed by Esscher (1932). This is a modification of a density to give an exponential family density. We give here an expanded exposition of the work in Barndorff-Nielsen and Cox (1989, Ch. 4). Let $X_1, \ldots, X_n$ be iid random variables from a density $f$ with moment generating function $M(\lambda)$ and cumulant generating function $K(\lambda)$. Define the exponential tilt of $f$ as the probability distribution with density

$$f_X(x;\lambda) = \exp(\lambda x - K(\lambda)) f_X(x), \tag{2.15}$$

where $\lambda$ is a constant. Observe that

$$\begin{aligned} E(X;\lambda) &= \int x f_X(x) \exp(\lambda x) dx e^{-K(\lambda)} \\ &= \frac{\int x f_X(x) \exp(\lambda x) dx}{M(\lambda)} \\ &= \frac{\int \frac{\partial}{\partial \lambda} f_X(x) \exp(\lambda x) dx}{M(\lambda)} \\ &= \frac{\frac{\partial}{\partial \lambda} M(\lambda)}{M(\lambda)} \\ &= K'(\lambda) \end{aligned}$$

and

$$Var(X;\lambda) = \int (x - K'(\lambda))^2 f_X(x) \exp(\lambda x) dx \, e^{-K(\lambda)}$$

$$\begin{aligned}
&= \frac{\int x^2 f_X(x)\exp(\lambda x)dx}{M(\lambda)} - (\frac{\frac{\partial}{\partial\lambda}M(\lambda)}{M(\lambda)})^2 \\
&= \frac{\frac{\partial^2}{\partial\lambda^2}M(\lambda)}{M(\lambda)} - (\frac{\frac{\partial}{\partial\lambda}M(\lambda)}{M(\lambda)})^2 \\
&= \frac{M''(\lambda)M(\lambda) - [M'(\lambda)]^2}{M(\lambda)^2} \\
&= K''(\lambda).
\end{aligned}$$

**Lemma 2.3.1** *Let* $X_1, \ldots, X_n$ *be iid random variables from* $f_X$ *above. If* $S_n = \sum_{i=1}^n X_i$ *then*

$$f_{S_n}(s;\lambda) = \exp(\lambda s - nK(\lambda))f_{S_n}(s). \tag{2.16}$$

*Proof.* Consider $f_X(x;\lambda)$ and derive the moment generating function.

$$\begin{aligned}
M(t;\lambda) &= \int e^{(t+\lambda)x - K(\lambda)} f_X(x)dx \\
&= \int e^{(t+\lambda)x - K(t+\lambda)} f_X(x)dx e^{K(t+\lambda)-K(\lambda)} \\
&= e^{K(t+\lambda)-K(\lambda)} \\
&= \frac{M(t+\lambda)}{M(\lambda)}.
\end{aligned}$$

This implies that the moment generating function for $f_{S_n}(s;\lambda)$ is $[M(t+\lambda)/M(\lambda)]^n$. Hence to show that the density $f_{S_n}(s;\lambda)$ is given by the right hand side of equation (2.16) we simply need to show that $[M(t+\lambda)/M(\lambda)]^n$ is the moment generating function for the density on the right hand side of equation (2.16). Now the moment generating function for the right hand side density is

$$\begin{aligned}
\int e^{ts} e^{\lambda s - nK(\lambda)} f_{S_n}(s)ds &= \int e^{(t+\lambda)s - nK(\lambda)} f_{S_n}(s)ds \\
&= e^{-nK(\lambda)} \int e^{(t+\lambda)s} f_{S_n}(s)ds \\
&= \frac{M(t+\lambda)^n}{M(\lambda)^n}, \qquad (2.17)
\end{aligned}$$

which thus implies that the two densities are identical.

Now, $E(S_n;\lambda) = nE(X;\lambda) = nK'(\lambda)$ and $Var(S;\lambda) = nVar(X;\lambda) = nK''(\lambda)$. Another important fact is that the cumulant generating function for the tilted distribution is $K^*(t) = K(t+\lambda) - K(\lambda)$. We can now use the Edgeworth approximation for the tilted distribution to determine the saddlepoint approximation.

### *2.3.2 The Saddlepoint Approximation For Univariate Densities*

Let $S_n^* = S_n - nK'(\lambda)/n\sqrt{K''(\lambda)}$ Then using the Edgeworth expansion derived in equation (2.6), we get that

$$f_{S_n^*}(s^*;\lambda) = \phi(s^*)\left\{1 + \frac{\rho_3^* H_3(s^*)}{6\sqrt{n}} + \frac{\rho_4^* H_4(s^*)}{24n} + \frac{\rho_3^{*2} H_6(s^*)}{72n} + O(n^{-3/2})\right\}, \quad (2.18)$$

where $s^* = (s - nK'(\lambda))/n\sqrt{K''(\lambda)}$. In this case

$$\rho_j^* = \frac{K^{*(j)}(t)|_{t=0}}{K''(\lambda)^{j/2}} = \frac{K^{(j)}(\lambda)}{K''(\lambda)^{j/2}}.$$

Transforming from $S_n^*$ to $S_n$, we see that

$$f_{S_n}(s;\lambda) = \frac{\phi(s*)}{(nK''(\lambda))^{1/2}}\left\{1 + \frac{\rho_3^* H_3(s^*)}{6\sqrt{n}} + \frac{\rho_4^* H_4(s^*)}{24n} + \frac{\rho_3^{*2} H_6(s^*)}{72n} + O(n^{-3/2})\right\}. \quad (2.19)$$

Hence, a valid expansion for $f_{S_n}(s)$ is

$$e^{nK(\lambda)-\lambda s}\frac{\phi(s^*)}{(nK''(\lambda))^{1/2}}\left\{1 + \frac{\rho_3^* H_3(s^*)}{6\sqrt{n}} + \frac{\rho_4^* H_4(s^*)}{24n} + \frac{\rho_3^{*2} H_6(s^*)}{72n} + O(n^{-3/2})\right\}.$$

The key to the saddlepoint approximation comes here. The three Hermite polynomials in the expansion are given by

$$H_3(z) = z^3 - 3z,$$

$$H_4(z) = z^4 - 6z^2 + 3,$$

and

$$H_6(z) = z^6 - 15z^4 + 45z^2 - 15.$$

In the above expansion, $\lambda$ is arbitrary. To center the expansion about a different point (and enable us to evaluate the Hermite polynomials at zero) we take $\lambda$ such that $s = nK'(\lambda)$. The $\lambda$ that solves this equation is the saddlepoint in the complex plane in the sense that Daniels (1954) first developed it. We label the saddlepoint $\hat{\lambda}$. This choice of $\lambda$ makes $s^* = 0$ and, since this means we must only evaluate the Hermite polynomials at one point (and $H_3(0) = 0$, eliminating the $n^{-1/2}$ term). Hence this simplifies expansion (2.19) to

$$\begin{aligned} f_{S_n}(s) &= e^{nK(\hat{\lambda})-\hat{\lambda}s}\left\{\frac{1}{nK''(\hat{\lambda})}\right\}^{1/2}\phi(0)\left\{1+\frac{3\rho_4^*}{24n}-\frac{15\rho_3^{*2}}{72n}+O(n^{-3/2})\right\} \\ &= e^{nK(\hat{\lambda})-\hat{\lambda}s}\left\{\frac{1}{2\pi nK''(\hat{\lambda})}\right\}^{1/2}\left\{1+\frac{3\rho_4^*}{24n}-\frac{15\rho_3^{*2}}{72n}+O(n^{-3/2})\right\}. \end{aligned}$$

This gives an approximation

$$f_{S_n}(s) \approx e^{nK(\hat{\lambda})-\hat{\lambda}s}\left\{\frac{1}{2\pi nK''(\hat{\lambda})}\right\}^{1/2}, \tag{2.20}$$

which has error $O(n^{-1})$ for any $s$. This is a significant improvement over the Edgeworth approximation because the error in the tails is significantly reduced. In several important classes of densities, the relative error is $O(n^{-1})$ uniformly in $s$. It is important that $\hat{\lambda}$ is defined as a function of $s$ through the saddlepoint equation and hence changes depending on the value of $s$.

There is a notational complexity to understand here. In the above equations, $\hat{\lambda}$ is a function of $s$, but one can also conceive of a random variable $\hat{\lambda} = \hat{\lambda}(S_n)$ which is a function of $S_n$, the sum of the observations, given by solving the saddlepoint equation at the random point $S_n$. We will endeavor to clearly state the interpretation of $\hat{\lambda}$ which is being used in the rest of this dissertation.

A simple change of variables from $S_n$ to $\bar{X}$ gives

$$f_{\bar{X}}(\bar{x}) \approx e^{n(K(\hat{\lambda})-\hat{\lambda}\bar{x})}\left\{\frac{n}{2\pi K''(\hat{\lambda})}\right\}^{1/2}. \tag{2.21}$$

In general, the saddlepoint density approximations do not integrate to 1. They can be made to by multiplication by an appropriate constant $c_n$. In three cases, the normal, gamma and inverse Gaussian distributions, correction by this constant gives back the original density.

### *2.3.3 Multivariate Saddlepoint Expansion*

Let $X_1, \ldots, X_n$ be an iid sample of m-dimensional random variables from a density $f_X(x)$ and let $S_n = \sum X_i$. Note that $X_1 = (X_{11}, \ldots, X_{1m})^T$. Define $f_X(x;\lambda) = \exp(\lambda x - K(\lambda)) f_X(x)$. Some of the results that were derived in the univariate case still apply:

$$f_{S_n}(s;\lambda) = \exp(\lambda s - nK(\lambda)) f_{S_n}(s),$$

$$E(S_n;\lambda) = nK'(\lambda),$$

and

$$Var(S_n;\lambda) = nK''(\lambda).$$

Note that $K'(\lambda)$ is a vector and $K''(\lambda)$ is a matrix.

Again let $S_n^* = (S_n - nK'(\lambda))/\sqrt{n}$. We apply the Edgeworth expansion as in the univariate case to see that

$$f_{S_n^*}(s^*;\lambda) = \phi(s^*;K''(\lambda))\left\{1 + Q_3(s^*;\kappa^{(3)}) + Q_4(s^*;\kappa^{(4)}) + \ldots\right\}, \tag{2.22}$$

and if we let $\hat{\lambda}$ solve

$$K'(\lambda) = s/n \tag{2.23}$$

so that $s^* = 0$, we can derive (in the same manner as the univariate expansion)

$$f_{S_n}(s) = \frac{\exp(nK(\hat{\lambda}) - \hat{\lambda}s)}{(2\pi)^{m/2}\left\{n|K''(\hat{\lambda})|\right\}^{1/2}}\left\{1 + Q_4(0;\kappa^{(4)})/n + Q_6(0;\kappa^{(6)})/n^2 + \ldots\right\}. \tag{2.24}$$

When we evaluate the Edgeworth expansion at the saddlepoint, $Q_3(s^*;\kappa^{(3)}) = 0$, in the same way as it was in the univariate case.

### *2.3.4 Double Saddlepoint Approximation to the Conditional Density*

Conditional densities are in general complicated due to calculation of the marginal density for the conditioning variables. Barndorff-Nielsen and Cox (1979) proposed a method of approximating a conditional density based on saddlepoint approximations to the two parts of the conditional density.

Let $X_1, \ldots, X_n$ be a sample of $n$ iid vector valued random variables from a density $f$. Denote the components of each $X_i$ by $X_{i1}, \ldots, X_{im}$. Let $S_n = \sum X_i = (S_{\cdot 1}, \ldots, S_{\cdot m})$. We are interested in the density of $S_{\cdot 1}$ given $S_{\cdot 2}, \ldots, S_{\cdot m}$. When both densities are known, this density takes the form

$$f_{S_{\cdot 1}}(s_1 | s_2, \ldots, s_m) = \frac{f_{S_{\cdot 1},\ldots,S_{\cdot m}}(s_1, \ldots, s_m)}{f_{S_{\cdot 2},\ldots,S_{\cdot m}}(s_2, \ldots, s_m)}. \tag{2.25}$$

One could approximate this density by calculating cumulants and means for this conditional density. This method is known as the direct saddlepoint approximation for the conditional density. However, it is quite complicated and can lead to difficult computations, because in general, calculating the cumulant generating function and solving the saddlepoint equation for these densities is non-trivial. It is however fairly accurate.

To avoid these complications, Barndorff-Nielsen and Cox (1979) proposed to approximate each of the densities separately and use the ratio of these expansions as an expansion for the conditional distribution. This is known as the double saddlepoint approximation. The density in the numerator is given in equation (2.24). To derive the denominator density we need to define a few terms. Let $S_{\cdot 1} = \sum_i X_{i1}$ and $S_{\cdot 2} = \sum_i (X_{i2}, \ldots, X_{im})$. Define $\hat{\lambda}_{20}$ to be the $(m-1)$-dimensional saddlepoint found by setting $K'(0, \lambda_{20}) = S_{\cdot 2}/n$. Then the denominator density becomes:

$$f_{S_{\cdot 2}}(s_2) = \frac{\exp\left\{nK(\hat{\lambda}_{20}) - \hat{\lambda}_{20}s_2\right\}}{(2\pi)^{(m-1)/2}|K''(\hat{\lambda}_{20})|^{1/2}} \left\{1 + Q_4(0;\kappa^{(4)})/n + Q_6(0;\kappa^{(6)})/n^2 + \ldots\right\}. \tag{2.26}$$

The double saddlepoint expansion is the ratio of these two expansions,

$$
\begin{aligned}
&f_{S_{\cdot 1}}(s_1|S_{\cdot 2} = s_2) = \\
&= \frac{\exp\left[n\left\{K(\hat{\lambda}) - K(0,\hat{\lambda}_{20})\right\} - \hat{\lambda}s + \hat{\lambda}_{20}s_2\right]}{(2\pi)^{1/2}(|K''(\hat{\lambda})|/|K''(\hat{\lambda}_{20})|)^{1/2}} \frac{\left\{1 + Q_4(0,\kappa^{(4)})/n + \ldots\right\}}{\left\{1 + Q_4(0,\kappa^{(4)}_{20})/n + \ldots\right\}}. \qquad (2.27)
\end{aligned}
$$

The subscript on $\kappa^{(4)}_{20}$ indicates that it is cumulants based on the marginal distribution of $S_{\cdot 2}$. It is trivial to see that the ratio of the errors simplifies to

$$1 + \frac{1}{n}\left\{Q_4(0,\kappa^{(4)}) - Q_4(0,\kappa^{(4)}_{20})\right\} + O(n^{-3/2}).$$

The double saddlepoint approximation for the conditional density is given by the leading term of (2.27). Again, a normalizing constant will in general be needed to ensure that the double saddlepoint approximation integrates to 1.

## 2.4 *Approximate Tail Probabilities*

The saddlepoint approximation to the density and the double saddlepoint approximation to conditional densities provide accurate approximate likelihoods, but direct integration for tail probabilities, test statistics and confidence limits is cumbersome. Several authors, beginning with Daniels (1954), have examined alternatives to direct integration of the saddlepoint approximate density which give accurate tail probability approximations.

### *2.4.1 The Tail Probability – Univariate Densities*

An approximation to the cumulative distribution function or tail probability for the saddlepoint approximation to the density can be derived in two ways. We will present them both here. The first involves only real variables, while the second uses complex variables and generalizes more readily to the multivariate and double saddlepoint cases. The first expansion is due to Daniels (1954) based on the exponential tilt methods, and the second is due to Lugannani and Rice (1980). A review of these results with comments can be found in Daniels (1987).

It is not convenient to simply integrate the saddlepoint density approximation term by term as was done with the Edgeworth approximation. We wish to examine the integral

$$P(S_n \leq s_0) = \int_{-\infty}^{s_0} e^{nK(\hat{\lambda})-\hat{\lambda}s} \left\{ \frac{1}{2\pi nK''(\hat{\lambda})} \right\}^{1/2} ds. \tag{2.28}$$

We derive a value for this integral by a series of changes of variable. The first change of variables is from $s$ to $\hat{\lambda}$. These two variables are connected by the saddlepoint equation and hence the Jacobian is

$$\frac{\partial s}{\partial \hat{\lambda}} = nK''(\hat{\lambda}).$$

The integral becomes

$$P(S_n \leq s_0) = \int_{-\infty}^{\hat{\lambda}_0} e^{n(K(\hat{\lambda})-\hat{\lambda}K'(\hat{\lambda}))} \left\{ \frac{1}{2\pi nK''(\hat{\lambda})} \right\}^{1/2} nK''(\hat{\lambda})d\hat{\lambda}, \tag{2.29}$$

where $\hat{\lambda}_0$ is defined by the saddlepoint equation $s_0 = nK'(\hat{\lambda}_0)$.

The second change of variable is to $q = \text{sign}(\hat{\lambda})\{2(\hat{\lambda}K'(\hat{\lambda}) - K(\hat{\lambda})\}^{1/2}$. This has Jacobian given by $\partial\hat{\lambda}/\partial q$. Now,

$$\partial q/\partial\hat{\lambda} = \text{sign}(\hat{\lambda})\{2(\hat{\lambda}K'(\hat{\lambda}) - K(\hat{\lambda})\}^{-1/2}(\frac{1}{2}) \left\{2(K'(\hat{\lambda}) + \hat{\lambda}K''(\hat{\lambda}) - K'(\hat{\lambda})\right\},$$

which simplifies to $\hat{\lambda}K''(\hat{\lambda})/q$. The Jacobian is thus $q/(\hat{\lambda}K''(\hat{\lambda}))$.

This makes our probability become:

$$\begin{aligned} P(S_n \leq s_0) &= \int_{-\infty}^{q_0} e^{-\frac{n}{2}q^2} \left\{ \frac{1}{2\pi nK''(\hat{\lambda})} \right\}^{1/2} nK''(\hat{\lambda}) \frac{q}{\hat{\lambda}K''(\hat{\lambda})} dq \\ &= \int_{-\infty}^{q_0} e^{-\frac{n}{2}q^2} \left\{ \frac{n}{2\pi nK''(\hat{\lambda})} \right\}^{1/2} \frac{q}{\hat{\lambda}} dq \\ &= \int_{-\infty}^{q_0} \frac{q}{\hat{\lambda}K''(\hat{\lambda})^{1/2}} \phi(q; n^{-1}) dq, \end{aligned}$$

where $\phi(x;\tau)$ is the density of a normal random variable with mean 0 and variance $\tau$.

Now, define $f(q) = q/(\hat{\lambda}K''(\hat{\lambda})^{1/2})$. To continue, we use a lemma due to Temme (1982) which enables an expansion of the integral of the product of a real-valued function and a normal density.

**Lemma 2.4.1** *Given $f$ a real-valued function of a real variable which vanishes at $\infty$, let*

$$Q_\tau(\gamma) = \int_{-\infty}^{\gamma} f(x)\phi(x;\tau^{-1})dx.$$

*Then*

$$Q_\tau(\gamma) = Q_\tau(\infty)\Phi(\gamma\sqrt{\tau}) - \frac{f(\gamma)-f(0)}{\gamma}\tau^{-1}\phi(\gamma;\tau^{-1})\{1+O(\tau^{-1})\}$$

*Proof.* We first demonstrate that $Q_\tau(\infty) = f(0)\{1+O(\tau^{-1})\}$.

$$\begin{aligned} Q_\tau(\infty) &= \int_{-\infty}^{\infty}\left[f(0)+f'(0)x+f''(0)x^2/2+\ldots\right]\phi(x;\tau^{-1})dx \\ &= f(0)+f'(0)E(X)+f''(0)E(X^2)/2+\ldots, \qquad (2.30) \end{aligned}$$

where $X \sim N(0,\tau^{-1})$. Since $X$ is normally distributed with mean 0, all the odd-powered expectations are 0, and the integral can be expressed as

$$\begin{aligned} & \sum_{i=0}^{\infty} f^{(2i)}(0)\frac{E(X^{2i})}{(2i)!} \\ = & \sum_{i=0}^{\infty} f^{(2i)}(0)\frac{(2i)!}{i!2^i(2i)!} \\ = & \sum_{i=0}^{\infty} f^{(2i)}(0)\frac{1\cdot 3\cdot 5\cdots(i-1)}{\tau^i(2i)!} \\ = & \sum_{i=0}^{\infty} f^{(2i)}(0)(2/\tau)^i\frac{\Gamma(i+1/2)}{(2i)!\Gamma(1/2)}, \end{aligned}$$

which implies that $Q_\tau(\infty) = f(0)\{1+O(\tau^{-1})\}$ as stated.

Now, we expand the original integral using repeated integration by parts. Write

$$Q_\tau(\gamma) = f(0)\Phi(\gamma\sqrt{\tau}) + \int_{-\infty}^{\gamma}\frac{f(x)-f(0)}{x}x\phi(x;\tau^{-1})dx. \qquad (2.31)$$

Now, integrate by parts with $u = (f(x) - f(0))/x$ and $dv = x\phi(x; \tau^{-1})dx$. This gives (using the identity derived earlier for $Q_\tau(\infty)$)

$$\begin{aligned} Q_\tau(\gamma) &= f(0)\Phi(\gamma\sqrt{\tau}) - \tau^{-1}\phi(\gamma; \tau^{-1})\frac{f(\gamma) - f(0)}{\gamma} \\ &+ \tau^{-1}\int_{-\infty}^{\gamma} \frac{\partial}{\partial x}\left\{\frac{f(x) - f(0)}{x}\right\}\phi(x; \tau^{-1})dx. \end{aligned} \tag{2.32}$$

Repeating the integration by parts gives

$$Q_\tau(\gamma) = \Phi(\gamma\sqrt{\tau})\sum_{i=0}^{\infty}\frac{f^{(i)}(0)}{\tau^i} - \phi(\gamma; \tau^{-1})\sum_{i=0}^{\infty}\frac{f^{(i)}(\gamma) - f^{(i)}(0)}{\tau^{i+1}},$$

where

$$f^{(i)}(x) = \frac{\partial}{\partial x}\frac{f^{(i-1)}(x) - f^{(i-1)}(0)}{x}$$

and $f^{(0)}(x) = f(x)$. Hence, we can write

$$Q_\tau(\gamma) = \Phi(\gamma\sqrt{\tau})f(0) - \tau^{-1}\phi(\gamma; \tau^{-1})\left\{\frac{f(\gamma) - f(0)}{\gamma}\right\} + O(\tau^{-1}), \tag{2.33}$$

and when we replace $f(0)$ with $Q_\tau(\infty) + O(\tau^{-1})$ we get the desired result.

We apply this lemma to the integral in question with

$$f(q) = \frac{q}{\hat{\lambda}K''(\hat{\lambda})^{1/2}}.$$

This gives the following approximation for the cumulative probability:

$$\begin{aligned} P(S_n \leq s_0) &\doteq \Phi(q_0\sqrt{n}) - \frac{1}{n}\phi(q_0; n^{-1})\frac{f(q_0) - 1}{q_0} + O(n^{-1}) \\ &= \Phi(q_0\sqrt{n}) - \phi(q_0\sqrt{n})\left\{\frac{q_0}{\hat{\lambda}K''(\hat{\lambda})^{1/2}\sqrt{n}q_0} - \frac{1}{\sqrt{n}q_0}\right\} \end{aligned}$$

which simplifies to

$$\Phi(q_0\sqrt{n}) + \phi(q_0\sqrt{n})\left\{\frac{1}{\sqrt{n}q_0} - \frac{1}{\sqrt{n}\hat{\lambda}K''(\hat{\lambda})^{1/2}}\right\} \tag{2.34}$$

where $q_0 = \text{sign}(\hat{\lambda}_0)[2(\hat{\lambda}_0K'(\hat{\lambda}_0) - K(\hat{\lambda}_0))]^{1/2}$ and $s_0 = nK'(\hat{\lambda}_0)$. This is the desired approximation. With this we can approximate the tail probability with error $O(n^{-1})$. The upper tail probability $P(S_n \geq s_0)$ has a similar form

$$1 - \Phi(q_0\sqrt{n}) - \phi(q_0\sqrt{n})\left\{\frac{1}{\sqrt{n}q_0} - \frac{1}{\sqrt{n}\hat{\lambda}K''(\hat{\lambda})^{1/2}}\right\}. \tag{2.35}$$

Complex analysis is required to derive approximations to the cumulative distribution function for the double saddlepoint approximation. We give two background results first, and then derive again the tail probability for the univariate saddlepoint expansion, and then for the double saddlepoint approximation. In both cases, this derivation allows modification for the use of discrete or lattice random variables, which are of primary interest in the sequel.

We first give Abel's lemma, which is actually only used in the real case (the proof can be found in appendix A):

**Lemma 2.4.1** *Take $v$ a nonincreasing function on $[A, B]$ that is bounded above. Define*

$$F(x) = \int_A^x f(u)du,$$

*and let $F$ be bounded in the interval so that $h \leq F(x) \leq H$ for $x$ in $[A, B]$, then*

$$hv(A) \leq \int_A^B v(u)f(u)du \leq Hv(A). \tag{2.36}$$

Abel's lemma is used in the proof of Watson's lemma, which is here stated in its simplest form (again see appendix A for a proof):

**Lemma 2.4.1** *Let*

$$I = \int_0^Z e^{-az} z^m f(z)dz, \tag{2.37}$$

*where $f$ is a possibly complex valued function analytic in a neighborhood of $z = 0$ and $a$ is large and positive. The path of integration is along the real axis. Assume $I$ exists for some $\alpha \leq a$. The result is that*

$$I = \sum_{j=0}^{\infty} \frac{f^{(j)}(0)}{j!} \frac{(m+j)!}{a^{m+j+1}} + O(e^{-aX}), \tag{2.38}$$

*where $X$ is within the radius of convergence of $f$.*

Watson's lemma can be used in other forms, in particular ones that will be more useful in our applications. First, a change of variable from $z$ to $u = z^2$ gives

$$\int_0^Z \exp(-\frac{1}{2}nz^2)f(z)dz \sim \frac{1}{2}\sum_{j=0}^{\infty} \frac{f^{(j)}(0)\Gamma[\frac{j+1}{2}]}{(j)!(\frac{n}{2})^{\frac{j+1}{2}}}, \tag{2.39}$$

and then summing both a positive and negative part (the odd terms in the sum cancel) gives

$$\int_{-A}^B \exp(-\frac{1}{2}nz^2)f(z)dz \sim \sum_{j=0}^{\infty} \frac{f^{(2j)}(0)\Gamma[j+\frac{1}{2}]}{(2j)!(\frac{n}{2})^{j+\frac{1}{2}}}, \tag{2.40}$$

and secondly, with another change of variable (this time from real $z$ to a complex variable)

$$\frac{1}{i}\int_{\tilde{w}-iA}^{\tilde{w}+iB} \exp(-\frac{1}{2}n(w-\tilde{w})^2)f(w)dw \sim \sum_{j=0}^{\infty} \frac{(-1)^j f^{(2j)}(\tilde{w})\Gamma[j+\frac{1}{2}]}{(2j)!(\frac{n}{2})^{j+\frac{1}{2}}}, \tag{2.41}$$

or

$$\frac{1}{i}(\frac{n}{2\pi})^{1/2}\int_{\tilde{w}-iA}^{\tilde{w}+iB} \exp(-\frac{1}{2}n(w-\tilde{w})^2)f(w)dw \sim \sum_{j=0}^{\infty} \frac{(-1)^j f^{(2j)}(\tilde{w})}{j!(2n)^j}. \tag{2.42}$$

Each of these transformations comes from an appropriate choice of $f$ and of the range of integration. This final form will come in handy in the next two derivations.

### *2.4.2 Tail Probability - Complex Variable Derivation*

The derivation of an approximate tail probability based on saddlepoint expansions is due to Lugannani and Rice (1980). We begin by writing the exact tail probability (based on the real density and the characteristic function inversion formula, not the approximation) as

$$\begin{aligned}\int_{-\infty}^{s_0} f_{S_n}(s)ds &= \frac{1}{2\pi i}\int_{-\infty}^{s_0}\int_{c-i\infty}^{c+i\infty} e^{nK(\lambda)-\lambda s}d\lambda ds \\ &= \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{nK(\lambda)-\lambda s_0}\frac{(\lambda-1)d\lambda}{\lambda},\end{aligned}$$

where $c$ is an arbitrary small real number chosen such that the integral does not pass through the origin. This makes the upper tail probability

$$\int_{s_0}^{-\infty} f_{S_n}(s)ds\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{nK(\lambda)-\lambda s_0}\frac{d\lambda}{\lambda}, \tag{2.43}$$

which is easier to work with. Define $\hat{\lambda}$ to solve $K'(\hat{\lambda}) = s_0/n$. We make a change of variables from $\hat{\lambda}$ to $w$ by the equation

$$w - \tilde{w} = (\lambda - \hat{\lambda})\sqrt{\frac{2(h(\lambda) - h(\hat{\lambda}))}{(\lambda - \hat{\lambda})^2}},$$

with $h(\lambda) = K(\lambda) - \lambda s_0/n$ and $\tilde{w} = \sqrt{2(\hat{\lambda} s_0/n - K(\hat{\lambda}))}$ so that $-\frac{1}{2}\tilde{w}^2 = h(\hat{\lambda})$. We then simply substitute in (2.43) to get

$$\frac{1}{2\pi i}\exp(-\frac{n}{2}\tilde{w}^2)\int_{\tilde{w}-i\infty}^{\tilde{w}+i\infty}\exp\left\{\frac{n}{2}(w-\tilde{w})^2\right\}\frac{1}{\lambda}\frac{\partial\lambda}{\partial w}\partial w. \tag{2.44}$$

Define the function $g(\cdot)$ such that

$$g(w) = (1/\lambda)\frac{\partial\lambda}{\partial w}. \tag{2.45}$$

This function has a singularity at $w = 0$. We perform the integration in two parts, splitting $g(w)$ into $g_1(w) = (1/\lambda)(\partial\lambda/\partial w) - (1/w)$ and $g_2(w) = (1/w)$. The second integral is exact:

$$\begin{aligned}
&\frac{1}{2\pi i}\exp(-\frac{n}{2}\tilde{w}^2)\int_{\tilde{w}-i\infty}^{\tilde{w}+i\infty}\exp\left\{\frac{n}{2}(w-\tilde{w})^2\right\}\frac{\partial w}{w}\\
= &\frac{1}{2\pi i}\int_{\tilde{w}-i\infty}^{\tilde{w}+i\infty}\exp(\frac{n}{2}w^2 - w\sqrt{n}\tilde{w})\frac{\partial w}{w}\\
= &\frac{\sqrt{n}}{2\pi i}\int_{\tilde{w}-i\infty}^{\tilde{w}+i\infty}\int_{\sqrt{n}\tilde{w}}^{\infty}\exp(\frac{n}{2}w^2 - w\sqrt{n}u)\partial u\partial w\\
= &\int_{\sqrt{n}\tilde{w}}^{\infty}\frac{\sqrt{n}}{2\pi i}\int_{-i\infty}^{+i\infty}\exp(\frac{n}{2}w^2 - w\sqrt{n}u)\partial w\partial u\\
= &\int_{\sqrt{n}\tilde{w}}^{\infty}\frac{1}{2\pi i}\int_{-\infty}^{+\infty}\exp(\frac{n}{2}v^2 - iuv)\partial v\partial u\\
= &1 - \Phi(\sqrt{n}\tilde{w}).
\end{aligned} \tag{2.46}$$

To approximate the first integral we use Watson's Lemma in the form (2.42). We first have to show that $g_1$ is analytic. We do this by showing that it has a power series representation. Now,

$$g_1(w) = \frac{1}{\lambda}\frac{\partial\lambda}{\partial w} - \frac{1}{w},$$

which implies that

$$wg_1(w) = \frac{w}{\lambda}\frac{\partial\lambda}{\partial w} - 1.$$

This function is analytic and hence has a power series representation $wg(w) = \sum_0^\infty a_i w^i$. This means that $g_1(w)$ can be written $\sum_0^\infty a_i w^{i-1}$. If we can show that $a_0 = 0$, then we can write $g_1(w) = \sum_0^\infty a_{i+1} w^i$ and hence it would have a power series representation and be analytic. Now, $a_0$ is $wg_1(w)$ evaluated at 0. Since $\lambda = 0$ when $w = 0$,

$$\lim_{w\to 0} \frac{w}{\lambda}\frac{\partial\lambda}{\partial w} = 1,$$

and hence, the value at 0 is 0 and $a_0 = 0$. Hence, $g_1$ is analytic.

Note that since $\frac{1}{2}(w - \tilde{w})^2 = (h(\lambda) - h(\hat{\lambda}))$ we can differentiate twice to determine that

$$(w - \tilde{w})\frac{\partial^2 w}{\partial\lambda^2} + (\frac{\partial w}{\partial\lambda})^2 = h''(\lambda).$$

Now $h''(\lambda) = K''(\lambda)$ by simple differentiation, implying that at $\tilde{w}$ (and hence at $\hat{\lambda}$)

$$\frac{\partial\lambda}{\partial w} = \frac{1}{\sqrt{K''(\hat{\lambda})}}.$$

Watson's lemma enables us to write the first integral as the sum

$$\begin{aligned} \frac{\exp(-\frac{n}{2}\tilde{w}^2)}{\sqrt{2\pi n}} g_1(\tilde{w}) + O(n^{-1}) &= \frac{\phi(\sqrt{n}\tilde{w})}{\sqrt{n}} g_1(\tilde{w}) + O(n^{-1}) \\ &= \phi(\sqrt{n}\tilde{w})\left(\frac{1}{\sqrt{n}\tilde{z}} - \frac{1}{\sqrt{n}\tilde{w}}\right) + O(n^{-1}), \quad (2.47) \end{aligned}$$

where $\tilde{z} = \hat{\lambda}\sqrt{K''(\hat{\lambda})}$ (it is equal to the corresponding term in (2.34) and $\tilde{w}$ is defined above and is equal to $q_0$ in (2.34).

The overall approximation is hence the same as that derived earlier,

$$P(S \geq s_0) = 1 - \Phi(\sqrt{n}\tilde{w}) + \phi(\sqrt{n}\tilde{w})\left(\frac{1}{\sqrt{n}\tilde{z}} - \frac{1}{\sqrt{n}\tilde{w}}\right) + O(n^{-1}). \quad (2.48)$$

*Lattice or Discrete Variables*

In the discrete case we have a situation similar to the continuous case.

$$\begin{aligned}
\sum_{s_0}^{\infty} fs_n(s) &= \frac{1}{2\pi i}\sum_{s_0}^{\infty}\int_{c-i\infty}^{c+i\infty} e^{nK(\lambda)-\lambda s}d\lambda \\
&= \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{nK(\lambda)}(\sum_{s_0}^{\infty} e^{-\lambda s})d\lambda \\
&= \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{nK(\lambda)}(e^{-\lambda s_0}\sum_{i=0}^{\infty} e^{-\lambda i})d\lambda \\
&= \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{nK(\lambda)-\lambda s_0}\frac{d\lambda}{1-\exp(-\lambda)}. \qquad (2.49)
\end{aligned}$$

An alternate derivation begins with a continuity correction, which rearranges (2.49) as follows:

$$\sum_{s_0}^{\infty} fs_n(s) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{nK(\lambda)-\lambda(s_0-1/2)}\frac{d\lambda}{2\sinh(\frac{1}{2}\lambda)}. \qquad (2.50)$$

I will derive the first case. Begin with the integral form

$$\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{nK(\lambda)-\lambda s_0}\frac{d\lambda}{1-\exp(-\lambda)}.$$

We make the same change of variable as earlier, giving the following new integral form:

$$\frac{1}{2\pi i}\exp(-\frac{n}{2}\tilde{w}^2)\int_{\tilde{w}-i\infty}^{\tilde{w}+i\infty}\exp\left\{\frac{n}{2}(w-\tilde{w})^2\right\}\frac{1}{(1-\exp(-\lambda))}\frac{\partial\lambda}{\partial w}\partial w. \qquad (2.51)$$

where $h$ and $w$ and $\tilde{w}$ are defined as before. We now proceed with the same procedures as before. The second case is derived identically with a few slight changes in the final results. Define $g(w)$ as in (2.45). Again, there is a singularity which must be accounted for. We do so by splitting the integral into two parts, as before. The first part of the integral thus can be evaluated approximately using Watson's lemma and the second integrates exactly. Both of these derivations follow the previous ones identically, and the final result is nearly the same:

$$1-\Phi(\sqrt{n}\tilde{w})+\phi(\sqrt{n}\tilde{w})\left(\frac{1}{\sqrt{n}\tilde{z}}-\frac{1}{\sqrt{n}\tilde{w}}\right)+O(n^{-1}) \qquad (2.52)$$

where $\tilde{z} = (1-\exp(\hat{\lambda}))\sqrt{K''(\hat{\lambda})}$ and $\tilde{w}$ is defined in the same way as in the continuous case.

### *2.4.3 Double Saddlepoint Approximation - Tail Probability*

The tail probability for the double saddlepoint approximation is of particular interest to researchers using generalized linear models. Typically the performance of unconditional approximations to conditional inference is weakest for calculation of test statistics and tail probabilities. The saddlepoint approximation to the tail probability improves the error from $O(n^{-1/2})$ to $O(n^{-1})$.

The saddlepoint tail probability could be calculated using numerical integration of the density, but this is a complicated integral and an approximate tail probability is desirable. Skovgaard (1987) gives the following derivation of the tail probability, based on a generalization of the above work due to Daniels (1954, 1987) and Lugananni and Rice (1980). The derivation of the tail probability in the conditional distribution using the double saddlepoint approximation is quite difficult. It relies on results from theory of complex variables.

We are interested in the tail probability and work with the upper tail for convenience, results for the lower tail being parallel (the derivations proceed identically). The probability in question is

$$\int_{s_{01}}^{\infty} f_{S_{\cdot 1}}(s_{\cdot 1}|s_{\cdot 2},\ldots,s_{\cdot m})ds_{\cdot 1} \tag{2.53}$$

which can be expressed as

$$\int_{s_{01}}^{\infty} \frac{f_{S_{\cdot 1},\ldots,S_{\cdot m}}(s_{\cdot 1},\ldots,s_{\cdot m})}{f_{S_{\cdot 2},\ldots,S_{\cdot m}}(s_{\cdot 2},\ldots,s_{\cdot m})}ds_{\cdot 1} = \frac{1}{f_{S_{\cdot 2},\ldots,S_{\cdot m}}(s_{\cdot 2},\ldots,s_{\cdot m})}\int_{s_{01}}^{\infty} f_{S_{\cdot 1},\ldots,S_{\cdot m}}(s_{\cdot 1},\ldots,s_{\cdot m})ds_{\cdot 1}. \tag{2.54}$$

Hence, we need to evaluate the second part of this expression,

$$\int_{s_{01}}^{\infty} f_{S_{\cdot 1},\ldots,S_{\cdot m}}(s_{\cdot 1},\ldots,s_{\cdot m})ds_{\cdot 1}.$$

This is done by writing the density as the inversion of a characteristic function, as mentioned above. For this derivation, we will assume that $S_{.2}$ is one-dimensional. The generalization to multivariate results is simple and will be mentioned later. The integral

$$
\begin{aligned}
& (\frac{1}{2\pi i})^2 \int_{s_{01}}^{\infty} \int_{c-i\infty}^{c+i\infty} \int_{-i\infty}^{i\infty} \exp\left\{nK(\lambda_1,\lambda_2) - \lambda_1 s_{.1} - \lambda_2 s_{.2}\right\} d\lambda_1 d\lambda_2 ds_{.1} \\
= \; & (\frac{1}{2\pi i})^2 \int_{c-i\infty}^{c+i\infty} \int_{-i\infty}^{i\infty} \exp\left\{nK(\lambda_1,\lambda_2) - \lambda_2 s_{.2}\right\} d\lambda_2 d\lambda_1 \int_{s_{01}}^{\infty} \exp(-\lambda_1 s_{.1}) ds_{.1} \\
= \; & (\frac{1}{2\pi i})^2 \int_{c-i\infty}^{c+i\infty} \int_{-i\infty}^{i\infty} \exp\left\{nK(\lambda_1,\lambda_2) - \lambda_2 s_{.2}\right\} \exp(-\lambda_1 s_{01}) \frac{d\lambda_2 d\lambda_1}{\lambda_1} \\
= \; & (\frac{1}{2\pi i})^2 \int_{c-i\infty}^{c+i\infty} \int_{-i\infty}^{i\infty} \exp\left\{nK(\lambda_1,\lambda_2) - \lambda_1 s_{01} - \lambda_2 s_{.2}\right\} \frac{d\lambda_2 d\lambda_1}{\lambda_1} \\
= \; & (\frac{1}{2\pi i}) \int_{c-i\infty}^{c+i\infty} (\frac{1}{2\pi i}) \int_{i\infty}^{i\infty} \exp\left\{nK(\lambda_1,\lambda_2) - \lambda_2 s_{.2}\right\} d\lambda_2 \exp(-\lambda_1 s_{01}) \frac{d\lambda_1}{\lambda_1}. \quad (2.55)
\end{aligned}
$$

Note that the integral ranges over $c \pm i\infty$ - it is necessary to avoid the singularity at $\lambda_1 = 0$, so the contour of integration is shifted an arbitrary small quantity $c$. The inner integral in (2.55) looks like the ordinary inversion formula for the density of $S_{.2}$, except that $K(\lambda_1, \lambda_2)$ should be $K(\lambda_2)$. Let $\hat{\lambda}_2(\lambda_1)$ solve the equation $K'_2(\lambda_1, \hat{\lambda}_2(\lambda_1)) = s_{.2}/n$ and then we have a situation where the saddlepoint approximation can be applied. For example, $\hat{\lambda}_2(0)$ solves $K'_2(0, \hat{\lambda}_2(0)) = s_{.2}/n$ This means that we can approximate

$$
(\frac{1}{2\pi i}) \int_{-i\infty}^{i\infty} \exp\left\{nK(\lambda_1,\lambda_2) - \lambda_2 s_{.2}\right\} d\lambda_2 \approx \sqrt{\frac{1}{2n\pi}} \frac{\exp\left\{nK(\lambda_1, \hat{\lambda}_2(\lambda_1)) - \hat{\lambda}_2(\lambda_1) s_{.2}\right\}}{\sqrt{K''_{22}(\lambda_1, \hat{\lambda}_2(\lambda_1))}} \quad (2.56)
$$

which yields as an overall approximation

$$
\begin{aligned}
& (\frac{1}{2\pi i}) \int_{c-i\infty}^{c+i\infty} \sqrt{\frac{1}{2n\pi}} \frac{\exp\left\{nK(\lambda_1, \hat{\lambda}_2(\lambda_1)) - \hat{\lambda}_2(\lambda_1) s_{.2}\right\}}{\sqrt{K''_{22}(\lambda_1, \hat{\lambda}_2(\lambda_1))}} \exp(-\lambda_1 s_{01}) \frac{d\lambda_1}{\lambda_1} \\
= \; & (\frac{1}{2\pi i}) \int_{c-i\infty}^{c+i\infty} \sqrt{\frac{1}{2n\pi}} \frac{\exp\left\{nK(\lambda_1, \hat{\lambda}_2(\lambda_1)) - \hat{\lambda}_2(\lambda_1) s_{.2} - \lambda_1 s_{01}\right\}}{\sqrt{K''_{22}(\lambda_1, \hat{\lambda}_2(\lambda_1))}} \frac{d\lambda_1}{\lambda_1}. \quad (2.57)
\end{aligned}
$$

Now let

$$
h(\lambda_1) = K(\lambda_1, \hat{\lambda}_2(\lambda_1)) - \hat{\lambda}_2(\lambda_1) s_{.2}/n - \lambda_1 s_{01}/n.
$$

Differentiating $K'_2(\lambda_1, \hat{\lambda}_2(\lambda_1)) = s_{\cdot 2}/n$ gives

$$\partial\hat{\lambda}_2(\lambda_1)/\partial\lambda_1 = -K''_{12}(\lambda_1, \hat{\lambda}_2(\lambda_1))/K''_{22}(\lambda_1, \hat{\lambda}_2(\lambda_1)),$$

and that the first two derivatives of $h$ are

$$\begin{aligned} h'(\lambda_1) &= K'_1(\lambda_1, \hat{\lambda}_2(\lambda_1)) + \\ &\quad \left\{K'_2(\lambda_1, \hat{\lambda}_2(\lambda_1)) - s_{\cdot 2}/n\right\}\left\{-K''_{12}(\lambda_1, \hat{\lambda}_2(\lambda_1))/K''_{22}(\lambda_1, \hat{\lambda}_2(\lambda_1))\right\} - s_{01}/n \\ &= K'_1(\lambda_1, \hat{\lambda}_2(\lambda_1)) - s_{01}/n \qquad (2.58) \end{aligned}$$

and

$$h''(\lambda_1) = K''_{11}(\lambda_1, \hat{\lambda}_2(\lambda_1)) - \frac{[K''_{12}(\lambda_1, \hat{\lambda}_2(\lambda_1)]^2}{K''_{22}(\lambda_1, \hat{\lambda}_2(\lambda_1))}. \qquad (2.59)$$

Now we rewrite the integral of interest as

$$(\frac{1}{2\pi i})\int_{c-i\infty}^{c+i\infty} \sqrt{\frac{1}{2n\pi}} \frac{\exp(nh(\lambda_1))}{\sqrt{K''_{22}(\lambda_1, \hat{\lambda}_2(\lambda_1))}} \frac{d\lambda_1}{\lambda_1}. \qquad (2.60)$$

It is now necessary to approximate this expression using another saddlepoint approximation. Let $\hat{\lambda}_1$ solve $h'(\lambda_1) = 0$. We do this by first defining a new variable $w$. Let $w$ be such that $\frac{1}{2}(w - \tilde{w})^2 = h(\lambda_1) - h(\hat{\lambda}_1)$ , where $\text{sign}(\tilde{w}) = \text{sign}(\hat{\lambda}_1)$. This means that when $\lambda_1 = 0$, $w = 0$, and $w = \tilde{w}$ when $\lambda_1 = \hat{\lambda}_1$. Now, choose $c = \hat{\lambda}_1$ in (2.60) and then we can rewrite

$$\begin{aligned} &(\frac{1}{2\pi i})\int_{c-i\infty}^{c+i\infty} \sqrt{\frac{1}{2n\pi}} \frac{\exp\{nh(\lambda_1)\}}{\sqrt{K''_{22}(\lambda_1, \hat{\lambda}_2(\lambda_1))}} \frac{d\lambda_1}{\lambda_1} \\ &= (\frac{1}{2\pi i})\sqrt{\frac{1}{2n\pi}} \exp(nh(\hat{\lambda}_1)) \int_{\hat{\lambda}_1-i\infty}^{\hat{\lambda}_1+i\infty} \frac{\exp\left\{nh(\lambda_1) - nh(\hat{\lambda}_1)\right\}}{\sqrt{K''_{22}(\lambda_1, \hat{\lambda}_2(\lambda_1))}} \frac{d\lambda_1}{\lambda_1} \\ &= (\frac{1}{2\pi i})\sqrt{\frac{1}{2n\pi}} \exp\left\{-\frac{n}{2}\tilde{w}^2 + nh(0)\right\} \int_{\tilde{w}-i\infty}^{\tilde{w}+i\infty} \frac{\exp\left\{\frac{n}{2}(w - \tilde{w})^2\right\}}{\sqrt{K''_{22}(\lambda_1, \hat{\lambda}_2(\lambda_1))}} \frac{\partial\lambda_1}{\partial w} \frac{w}{\lambda_1} \frac{\partial w}{w} \end{aligned} \qquad (2.61)$$

Define $g(w) = (1/\sqrt{K''_{22}(\lambda_1, \hat{\lambda}_2(\lambda_1))}(\partial\lambda_1/\partial w)(w/\lambda_1)$. Then we can rewrite (2.61) as

$$(\frac{1}{2\pi i})\sqrt{\frac{1}{2n\pi}} \exp(-\frac{n}{2}\tilde{w}^2 + nh(0)) \int_{\tilde{w}-i\infty}^{\tilde{w}+i\infty} \exp(\frac{n}{2}(w - \tilde{w})^2) g(w) \frac{\partial w}{w}.$$

Parallel to the development of the single saddlepoint approximation, define

$$
\begin{aligned}
g_1(w) &= \frac{g(w)-g(0)}{w} \\
&= \frac{1}{\sqrt{K''_{22}(\lambda_1,\hat\lambda_2(\lambda_1))}}\frac{\partial\lambda_1}{\partial w}\frac{1}{\lambda_1}-\frac{g(0)}{w}, \qquad (2.62)
\end{aligned}
$$

and $g_2(w) = g(0)/w$. The integration proceeds in the same manner as before, approximating the first part with Watson's lemma and integrating the second part exactly. We can rewrite the whole integral

$$
\begin{aligned}
&(\frac{1}{2\pi i})\sqrt{\frac{1}{2n\pi}}\exp\left\{-\frac{n}{2}\tilde w^2+nh(0)\right\} \\
&\times\int_{\tilde w-i\infty}^{\tilde w+i\infty}\exp\left\{\frac{n}{2}(w-\tilde w)^2\right\}\{g_1(w)+g_2(w)\}\partial w \\
&= \sqrt{\frac{1}{2n\pi}}\exp(nh(0))g(0)(\frac{1}{2\pi i})\exp(-\frac{n}{2}\tilde w^2) \\
&\times\int_{\tilde w-i\infty}^{\tilde w+i\infty}\exp\left\{\frac{n}{2}(w-\tilde w)^2\right\}\left\{\frac{g_1(w)}{g(0)}+\frac{1}{w}\right\}\partial w. \qquad (2.63)
\end{aligned}
$$

Note that $g(0) = K''_{22}(0,\lambda_2(0))^{-1/2}$ which means that the first part of this expression is exactly the saddlepoint approximation for the marginal distribution, and hence divides out when we condition on the appropriate margins. The integral (2.63) can be shown to have a form similar to the univariate case. For the first part, we demonstrate that $g_1$ is analytic in the same manner as in the univariate case. Consider $wg_1(w)$. This is again analytic. By an identical argument, we can show that $g_1$ is analytic and apply Watson's lemma in the same form as before. This gives a part of the approximation as follows:

$$
\begin{aligned}
\frac{\exp(-\frac{n}{2}\tilde w^2)}{\sqrt{2\pi n}}\frac{g_1(\tilde w)}{g(0)}+O(n^{-1}) &= \frac{\phi(\sqrt{n}\tilde w)}{\sqrt{n}}\frac{g_1(\tilde w)}{g(0)}+O(n^{-1}) \\
&= \phi(\sqrt{n}\tilde w)\left(\frac{1}{\sqrt{n}\tilde z}-\frac{1}{\sqrt{n}\tilde w}\right)+O(n^{-1}). \qquad (2.64)
\end{aligned}
$$

where $\tilde z = \tilde w g(0)/g(\tilde w)$. $g(\tilde w) = \tilde w/\hat\lambda_1|K''(\hat\lambda_1,\hat\lambda_2(\hat\lambda_1))|^{-1/2}$, so

$$
\tilde z = \hat\lambda_1\sqrt{\frac{|K''(\hat\lambda_1,\hat\lambda_2(\hat\lambda_1))|}{K''_{22}(0,\hat\lambda_2(0))}}
$$

and $\tilde{w}$ is defined above. The second part can be integrated exactly, giving

$$1 - \Phi(\sqrt{n}\tilde{w}).$$

This leaves the overall approximation as

$$1 - \Phi(\sqrt{n}\tilde{w}) + \phi(\sqrt{n}\tilde{w})\left(\frac{1}{\sqrt{n}\tilde{z}} - \frac{1}{\sqrt{n}\tilde{w}}\right) + O(n^{-1}), \tag{2.65}$$

where $\tilde{w}$ and $\tilde{z}$ are defined above.

This result extends to conditioning on multivariate random variables. There are no complications in the proof – it is just a matter of notation and the fact that the lower right corner of the second derivative of the cumulant generating function becomes a matrix, and $\tilde{z}$ becomes

$$\tilde{z} = \hat{\lambda}_1 \sqrt{\frac{|K''(\hat{\lambda}_1, \hat{\lambda}_2(\hat{\lambda}_1))|}{|K''_{22}(0, \hat{\lambda}_2(0))|}}.$$

*Lattice or Discrete Case*

Skovgaard (1987) extended the development of the double saddlepoint tail probability approximation to the discrete case, which is of essential interest for the problems under consideration for this dissertation. The derivation parallels the double saddlepoint in the continuous case, but uses the same summation that was used in the univariate discrete case. The result is as one would expect. One can choose to apply a continuity correction, replacing $s_{.1}$ in the equations by $s_{.1} - 1/2$.

The tail probability retains the same form as in (2.65) but since the outer integral changes to a summation, $\tilde{z}$ becomes

$$\tilde{z} = (1 - \exp(-\hat{\lambda}_1)) \sqrt{\frac{|K''(\hat{\lambda}_1, \hat{\lambda}_2(\hat{\lambda}_1))|}{|K''_{22}(0, \hat{\lambda}_2(0))|}}, \tag{2.66}$$

if the continuity correction is not used. If a continuity correction is used, the change is to

$$(2\sinh(\frac{1}{2}\hat{\lambda}_1)) \sqrt{\frac{|K''(\hat{\lambda}_1, \hat{\lambda}_2(\hat{\lambda}_1))|}{|K''_{22}(0, \hat{\lambda}_2(0))|}}. \tag{2.67}$$

Corresponding results for the lower tail can be derived as well. Without the continuity correction, $\tilde{z}$ becomes

$$\tilde{z} = (1 - \exp(\hat{\lambda}_1))\sqrt{\frac{|K''(\hat{\lambda}_1, \hat{\lambda}_2(\hat{\lambda}_1))|}{|K''_{22}(0, \hat{\lambda}_2(0))|}}.$$

If the continuity correction is implemented, $\tilde{z}$ remains as in (2.67) but the perturbation of the sufficient statistic is changed to reflect the lower tail, changing $s_{\cdot 1}$ to $s_{\cdot 1} - 1/2$.

# Chapter 3

# APPLICATIONS TO GENERALIZED LINEAR MODELS

## *3.1 Introduction*

This chapter reviews the implementation of the saddlepoint approximations in the generalized linear model. Several facts about the GLM are convenient in using the saddlepoint approximation and these will be demonstrated below.

## *3.2 The Density and the Modified Profile Likelihood*

The generalized linear model (McCullagh and Nelder, 1989) is a class of models which extends the form of the linear model to observations with non-normal error components.

Consider data of the form $Y_1, \ldots, Y_n$ with covariates $X_1, \ldots, X_n$ where $Y_i$ is a scalar and $X_i = (X_{i1}, \ldots, X_{ip})^T$ is a $p$-dimensional vector. The data are said to follow a generalized linear model (GLM) when the following requirements are satisfied:

1. The components of $Y$ have independent distributions from a family of densities of the form

   $$f_Y(y_i; \theta) = \exp\left[\{y_i\theta_i - b(\theta_i)\}/a(\phi) + c(y, \phi)\right]$$

   for some functions $a$, $b$ and $c$. Note that $E(Y_i) = \mu_i = b'(\theta_i)$, since $b(\cdot)$ is the cumulant generating function.

2. There exists a monotonic differentiable function $g$ (called the link function) such that $g(\mu_i) = \eta_i$ where $\eta_i = \sum_j X_{ij}\beta_j$ for some $\beta_j,\ j = 1, \ldots, p$.

The link function $g$ is called the canonical link function if $g(\mu) = \theta$ (which implies that $\theta_i = \eta_i$), which enables the density to be written as a function of the $\beta$ as

$$f_Y(y;\beta) = \exp\left[\{y\eta - b(\theta(\beta))\}/a(\phi) + c(y,\phi)\right],$$

or

$$f_Y(y;\beta) = \exp\left\{\left(\beta^T t(y) - b(\theta(\beta))\right)/a(\phi) + c(y,\phi)\right\},$$

where $t(y) = y^T x$. In the rest of the chapter, we assume that $\phi = 1$ and that $a(\phi) = 1$. This simplifies the density and the approximations considerably; however, the case of general $\phi(\cdot)$ and $a(\cdot)$ does not add complexity to anything except the written form of the approximation.

Daniels (1954) first noted that when the data follow an exponential family model, the form of the saddlepoint approximation is simplified. Consider data $Y_1, \ldots, Y_n$ come from a density of the form:

$$f_Y(y;\theta) = \exp\left\{\theta^T t(y) - b(\theta) - d(y)\right\}, \tag{3.1}$$

where $T = t(Y)$ is the vector of minimal sufficient statistics. Here, $K(\lambda) = b(\lambda+\theta) - b(\theta)$ and hence $\hat{\lambda} = \hat{\theta} - \theta$ where $\hat{\theta}$ is the ordinary maximum likelihood estimate. The saddlepoint approximation to the density of $T$ at $T = t(Y)$ is given by

$$\begin{aligned} f_T(t;\theta) &= \frac{1}{(2\pi)^{m/2}|K''(\hat{\lambda})|^{1/2}} \exp\left[n\left\{b(\hat{\theta} - \theta + \theta) - b(\theta)\right\} - (\hat{\theta} - \theta)^T t\right] \\ &= \frac{1}{(2\pi)^{m/2}|I(\hat{\theta})|^{1/2}} L(\theta)/L(\hat{\theta}). \end{aligned} \tag{3.2}$$

The double saddlepoint approximation in the exponential family has a similar form. Letting $T_1$ be the sufficient statistic for $\theta_1$ and $T_2$ the sufficient statistic for the nuisance parameter $\theta_2$, we can develop the following approximation:

$$f_{T_1}(t_1;\theta_1|T_2 = t_2) = \frac{|I_{22}(\theta_1, \hat{\theta}_2(\theta_1))|^{1/2}}{(2\pi|I(\hat{\theta})|)^{1/2}} L(\theta_1, \hat{\theta}_2(\theta_1))/L(\hat{\theta}) \tag{3.3}$$

where $\hat{\theta}_2(\theta_1)$ is the maximum likelihood estimator of $\theta_2$ for a fixed value of $\theta_1$ and the information matrix is partitioned into blocks as

$$\begin{bmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{bmatrix}.$$

This approximation gives rise to an approximate likelihood, which has been named the modified profile likelihood by Barndorff-Nielsen(1983) (see also the adjusted profile likelihood of Cox and Reid(1987)). This approximate likelihood has the form

$$L_{MP}(\theta_1; t_1 | T_2 = t_2) = |I_{22}(\theta_1, \hat{\theta}_2(\theta_1))|^{1/2} L(\theta_1, \hat{\theta}_2(\theta_1)), \tag{3.4}$$

where the second term is the profile likelihood for $\theta_1$. This approximate likelihood has been shown to approximate the true likelihood for models both on and off the exponential family, and has given rise to a body of work on higher-order asymptotics in general (Barndorff-Nielsen 1980, 1983, 1986). For the exponential family, the results of chapter 2 imply directly that

$$L_{MP}(\theta_1; t_1 | T_2 = t_2)(1 + O(n^{-1})) = L_C(\theta_1; t_1 | T_2 = t_2) \tag{3.5}$$

where $L_C$ is the conditional likelihood for $\theta_1$ given $T_2$.

## 3.3 The Approximate Tail Probability

The tail probability approximation in the univariate case has a simple form as well. Recall that (compare to (2.48) - note that we incorporate the $\sqrt{n}$ into $w$ and $z$ by letting $w = \sqrt{n}\tilde{w}$ and $z = \sqrt{n}\tilde{z}$)

$$P(T \leq t | \theta_0) \doteq \Phi(w) + \phi(w)\,(1/w - 1/z)\,. \tag{3.6}$$

In the exponential family, several favorable things occur. As derived above, $\hat{\lambda} = \hat{\theta} - \theta_0$, and

$$nK''(\hat{\lambda}) = I(\hat{\theta})$$

Also, in (2.48),

$$\begin{aligned}\sqrt{n}\tilde{w} &= \sqrt{2(\hat{\lambda}t - K(\hat{\lambda}))} \\ &= \operatorname{sign}(\hat{\theta})(2(l(\hat{\theta}) - l(\theta_0)))^{1/2} \\ &= \operatorname{sign}(\hat{\phi})\sqrt{2\log\Lambda}, \end{aligned} \tag{3.7}$$

and

$$\begin{aligned} z = \sqrt{n}\tilde{z} &= \hat{\lambda}\sqrt{nK''(\hat{\lambda})} \\ &= \left(\hat{\theta} - \theta_0\right) I(\hat{\theta}) \\ &= \frac{\hat{\theta} - \theta_0}{SE(\hat{\theta})}, \end{aligned} \tag{3.8}$$

where $2\log\Lambda$ is the standard LR statistic and $z$ is the square root of the Wald statistic for testing $\theta = \theta_0$.

In the exponential family, the double saddlepoint approximation to the tail probability also has a convenient form. The key fact additional is that

$$\begin{aligned} |K''(\hat{\lambda}_1, \hat{\lambda}_2(\lambda_1))| &= |I(\hat{\theta})| \\ &= |I_{11\cdot 2}(\hat{\theta})||I_{22}(\hat{\theta})|. \end{aligned} \tag{3.9}$$

This means that

$$\begin{aligned} \sqrt{n}\tilde{z} &= \sqrt{n}\hat{\lambda}_1\sqrt{\frac{|K''(\hat{\lambda}_1, \hat{\lambda}_2(\hat{\lambda}_1))|}{K''_{22}(0, \hat{\lambda}_2(0))}} \\ &= (\hat{\theta}_1 - \theta_{10})\sqrt{\frac{|I_{11\cdot 2}(\hat{\theta})||I_{22}(\hat{\theta})|}{|I_{22}(\theta_{10}, \hat{\theta}_2(\theta_{10})|}}. \end{aligned} \tag{3.10}$$

This tail probability is slightly different from the univariate – see (3.6) and (2.65) – here $\tilde{z} = z\rho$ where $z$ is the square root of the Wald statistic for the single degree of freedom test of $\theta_1 = \theta_{10}$ and

$$\rho = \left[\frac{|I_{22}(\hat{\theta})|}{|I_{22}(\theta_{10}, \hat{\theta}_2(\theta_1))|}\right]^{1/2}, \tag{3.11}$$

the ratio of the determinant of the block of the information matrix corresponding to the nuisance parameters in the full model to that in the reduced model. Again,

$$w = \operatorname{sign}(\hat{\phi})\sqrt{2\log\Lambda} \tag{3.12}$$

the signed square root of the likelihood ratio statistic for testing $\theta_1 = \theta_{10}$.

Hence the Skovgaard tail probability approximation takes the form

$$P(T_1 \leq t|T_2, \theta_{10}) \doteq \Phi(w) + \phi(w)\left(1/w - 1/(z\rho)\right). \tag{3.13}$$

Barndorff-Nielsen(1986) developed a different approximation for the tail probability based on direct integration of the double saddlepoint approximation. Pierce and Peters(1992) applied this work to the GLM. This approximation has a slightly simpler form than the one given in (3.13), and is given by

$$P(T_1 \leq t|T_2, \theta_{10}) \doteq \Phi\left[w + \frac{1}{w}\left\{\log\rho - \log(w/z)\right\}\right] \tag{3.14}$$

where the terms $w$, $z$ and $\rho$ are given above.

To derive this approximation, one expands not the integral of the actual density, as is done in section 2.4, but the integral of the density based on Barndorff-Nielsen's modified profile likelihood. First, it is useful to note that

$$|I(\hat{\theta})| = |I_{11\cdot 2}(\hat{\theta})||I_{22}(\hat{\theta})|, \tag{3.15}$$

so that the double saddlepoint approximation to the density can be written as

$$\frac{1}{|I_{11\cdot 2}(\hat{\theta})|^{1/2}} \frac{L_P(\theta_1)}{\rho L_P(\hat{\theta}_1)}. \tag{3.16}$$

From here, the approach is to make a change of variables first to the directed log-likelihood ratio statistic $w$, and then to the modified likelihood ratio statistic $w^* = w + \frac{1}{w}(\log\rho - \log(w/z))$ in (3.14). A rough derivation can be found in Pierce and Peters (1992) and a more general and detailed exposition in Barndorff-Nielsen (1990).

It is straightforward to demonstrate (as noted by Pierce and Peters (1992)) that the right hand sides of (3.14) and (3.13) are asymptotically equivalent as $n \to \infty$ . Taking the approximation

$$\begin{aligned} P(T_1 \leq t | T_2, \theta_{10}) &\doteq \Phi\left[w + \frac{1}{w}\{\log\rho - \log(w/z)\}\right] \\ &= \Phi\left\{w + \frac{1}{w}\log(\rho z/w)\right\} \\ &\doteq \Phi\left\{w - \frac{1}{w}(\frac{w}{z\rho} - 1)\right\} \\ &= \Phi\{w + 1/(z\rho) + 1/w\}, \end{aligned} \tag{3.17}$$

and expanding about $w$ gives

$$\Phi(w) + \phi(w)(1/w - 1/(z\rho)) + \phi'(w)(1/w - 1/(z\rho))^2/2 + \ldots. \tag{3.18}$$

The third term of this expansion is of order $1/n$. Hence, the two approximations are asymptotically equivalent. Approximation (3.14) is easier to use, and showed superior performance in limited work by Pierce and Peters.

It is interesting to consider this approximation in terms of $w$, which is just the unconditional likelihood ratio statistic, and two adjustment terms. They define

$$NP = \frac{1}{w}\log\rho \tag{3.19}$$

which they consider to be adjusting for the nuisance parameters and

$$INF = \frac{1}{w}\log(w/z). \tag{3.20}$$

which Pierce and Peters interpret as adjusting for the information difference between the conditional and unconditional models. Pierce and Peters (1992) proposed this division of the adjustment, and investigated the effects of nuisance parameters on the two adjustments, and gave general comments about the effects of these two quantities on the approximation.

When $\hat{\lambda} = 0$, that is when $\hat{\theta} = \theta$, the approximations in (3.14) and (3.13) are infinite. Two methods to avoid this problem have been proposed. The first, due to Skovgaard (1987) and Davison (1988) replaces the formulae with

$$P(T_1 \leq t | T_2, \theta_{10}) \doteq \frac{1}{2} + \frac{\lambda_1/6\gamma^2 + \mathrm{tr}(\lambda_2\lambda_3)/2}{\gamma(2\pi n)^{1/2}} \tag{3.21}$$

where $\gamma = [I_{11} - I_{12}I_{22}^{-1}I_{21}]^{1/2}$, $I_{ij}$ is the $i,j$ block of the information matrix, $\lambda_1 = \sum_{ijk} K'''(\hat{\theta})_{ijk}v_iv_jv_k$, $\lambda_2$ is the $p \times p$ ($p$ is the number of parameters) matrix with $I_{22}^{-1}$ in the lower right hand corner, and $\lambda_3$ is the $p \times p$ matrix with elements $\sum_k K'''(\hat{\theta})_{ijk}v_k$. $v$ is the $1 \times p$ vector $(1, -I_{12}I_{22}^{-1})$. The second, suggested by Bedrick and Hill (1992), used either the continuity corrected value or the uncorrected value, depending on which gives a smaller approximate normally distributed statistic. This approach appears to work well in practice (Bedrick and Hill, 1992), and is much easier to implement than the first approach above.

In all the above cases, the tail probability result given is valid only for continuous data. The results change only slightly, however, for discrete data, whether counts or more general lattice data because the tail probability, instead of being an integral, is a sum. The methods of sections 2.4.2 and 2.4.3 are applied, and one can take advantage of the fact that $\hat{\lambda} = \hat{\theta} - \theta$. To convert the expansion to one useful for discrete data we must modify $z$ in equations (3.13) and (3.14) by replacing $(\hat{\theta}_1 - \theta_1)$ with $1 - \exp(\hat{\theta}_1 - \theta_1)$. If one wishes to utilize a continuity correction, the sufficient statistic is perturbed by $\pm 1/2$ as needed, and $(\hat{\theta}_1 - \theta_1)$ is replaced by $\exp(\frac{1}{2}(\hat{\theta}_1 - \theta_1)) - \exp(-\frac{1}{2}(\hat{\theta}_1 - \theta_1))$.

In all the above cases, the approximation is valid only for the canonical link. Approximations have been derived for non-canonical links (Pierce and Peters (1992)) but not examined in any detail. These approximations involve using sample space derivatives (Pierce and Peters (1992)) of the log-likelihood.

### *3.3.1 The Sequential Saddlepoint Approximation*

Fraser, Reid, and Wong (1991) proposed an alternative to (3.13) and (3.14). They suggested evaluating the two tail probability approximations using the modified profile likelihood (3.4) in place of the unconditional likelihood. The Wald and likelihood ratio statistics are based on the modified profile likelihood. Recall that the modified profile likelihood is given by

$$L_{MP}(\theta_1; t_1 | T_2 = t_2) = |I_{22}(\theta_1, \hat{\theta}_2(\theta_1))|^{1/2} L(\theta_1, \hat{\theta}_2(\theta_1)). \tag{3.22}$$

Maximization of this likelihood gives rise to $\hat{\theta}_{1MP}$, the maximum modified profile likelihood estimator. A Wald statistic

$$z_{MP} = (\hat{\theta}_{1MP} - \theta_1)\sqrt{I_{MP}(\hat{\theta}_{1MP})}, \tag{3.23}$$

and a likelihood ratio statistic

$$w_{MP} = \sqrt{2(\log L_{MP}(\hat{\theta}_{1MP}) - \log L_{MP}(\theta_1))}, \tag{3.24}$$

can be calculated, where $I_{MP}$ is the information for $\theta_1$ from the modified profile likelihood. These two statistics can be used as if they were from a univariate approximation to give the sequential saddlepoint tail probability approximations

$$P(T_1 \le t | T_2, \theta_{10}) \doteq \Phi(w_{MP}) + \phi(w_{MP}) \left( \frac{1}{w_{MP}} - \frac{1}{z_{MP}} \right), \tag{3.25}$$

and

$$P(T_1 \le t | T_2, \theta_{10}) \doteq \Phi \left\{ w_{MP} - \frac{1}{w_{MP}} \log(w_{MP}/z_{MP}) \right\}. \tag{3.26}$$

The modified profile likelihood depends only on $\beta$, so there is no need to use $\rho$. Empirical work in the gamma distribution showed that for several situations, this method improved the empirical performance of the methods.

## *3.4 Applied Work in the Exponential Family and GLM*

There has been some research done regarding the implementation of these approximations in the generalized linear model. Davison (1988) first applied the work of Skovgaard (1988) and Daniels (1988) to the GLM. Pierce and Peters (1992) examined the performance of the approximations in more detail. They examined the test of independence in a 4x4 contingency table with ordered categories. They calculated one-sided p-values and confidence intervals for the MLE for the range of possible outcomes for tables with all marginal totals being 1, 2 and 4. They examined both the Davison and Barndorff-Nielsen approaches to the tail probability and the use of the sequential saddlepoint approximation. For the sequence of $2 \times 2$ tables (Example 1.1), they examined sets of 10 and 40 tables, with 2 controls and 1 case in each table. They plotted the tail probability values using both the Barndorff-Nielsen and Davison approaches. Barndorff-Nielsen and Cox (1994) plotted the likelihood function for the same setup with 10 tables and marginals of 1 and 2.

Bedrick and Hill (1992) examined three models from an empirical perspective. For Bartlett's test of no three way interaction in a $2 \times 2 \times 2$ table, and twenty published data sets, they calculated the test statistic and p-value for all possible outcomes with the given marginals using Davison's saddlepoint approximation, the exact conditional distribution, and a normal approximation. They repeated this for a monotonic trend test. For the test of a common odds ratio in a set of two by two tables, they examined a dataset from a matched case-control study analyzed in Breslow and Day (1980). Here, they created 25 randomly generated datasets for each of several different configurations ranging from 1:1 to 1:4 matching and from 10 to 40 tables. Again, they studied the difference between the exact conditional tail probability and the saddlepoint approximation using Davison's approach.

Both Pierce and Peters and Bedrick and Hill found that the saddlepoint and modified profile likelihood inference broke down when the number of nuisance parameters became large relative to the number of data points. For example, with 10 tables, the saddlepoint methods were quite good in the midrange of the possible data, but poor near the extremes. With 40 tables, the results broke down considerably. In contrast to the work of Fraser, Reid and Wong, Pierce and Peters found little benefit in using the sequential saddlepoint approximation. Both pairs of authors pointed to the need for a more systematic study.

All of these Authors examined the use of continuity corrections to the saddlepoint approximation. In both cases, the results were somewhat equivocal, in that the continuity corrected saddlepoint approximation did not always perform better than the uncorrected in approximating the conditional results. Pierce and Peters' work suggested that the continuity correction is of limited benefit given the complications that arise using it. Bedrick and Hill, on the other hand, found that continuity corrected saddlepoint approximations in general performed better in their applied work. This problem will be considered later in the dissertation, in particular in the simulation studies.

### *3.4.1 Theoretical Work in the Exponential Family and GLM*

In an unpublished technical report, Barndorff-Nielsen (1994) proved a result regarding asymptotics where the number of parameters increases with the sample size in the gamma distribution. This result is as follows: Let $\exp Y_{i1}, \ldots, \exp Y_{im}$, $i = 1, \ldots, q$ have the gamma density

$$f_{Y_{ij}}(y_{ij}) = \frac{\alpha_i^\beta}{\Gamma(\beta)} \exp(-\alpha_i e^{y_{ij}}) \exp(y_{ij}\beta).$$

The unconditional maximum profile likelihood estimator $\hat{\beta}$ has the same asymptotic distribution as the maximum exact conditional likelihood estimator $\hat{\beta}_C$ if $q/m \to 0$, while the maximum modified profile likelihood estimator $\hat{\beta}_{MP}$ has the same asymptotic distribution as $\hat{\beta}_C$ if $q/m^3 \to 0$. This means that if one has ten sets of observations (ten nuisance parameters) the inference from the modified profile likelihood has the same properties with ten observations per set as the profile likelihood would have with 1000 observations per set.

This result demonstrates that at least in one model, the maximum likelihood estimator from the modified profile likelihood is better in the asymptotic sense at approximating the conditional MLE. It would be ideal to generalize this result to Example 1.1, but it is not likely that this is possible, due to the vastly different performance of the approximations at different values of the parameter of interest. It is possible to examine the performance of the estimator at values of the parameter near the origin, and this will be examined in chapter 4. It is also possible to extend the above result to a wider class of distributions which includes the gamma distribution, and this will be done in chapter 7. For practical purposes, these distributions are of limited use to biostatisticians, but the problem and its solution suggest future work.

# Chapter 4

# APPLICATION TO LOG ODDS RATIO REGRESSION

## *4.1 Introduction*

Case control studies often involve grouping or cross-classification of the data into homogeneous strata. When confounding is present, it is beneficial to split the data into sets such that for each set the cases and controls are balanced with regard to the confounders. This gives rise to a set of 2 × 2 tables.

In a case control study, it is not possible to directly estimate the relative risk of disease given exposure, which is the parameter of primary interest. However, it is possible to estimate the odds ratio of disease given exposure, which, in the rare disease setting approximates the relative risk. In a single, non-stratified study, Fisher's exact test is easily performed to test for odds ratios different from 1. In the case of many tables, a basic analysis uses the method of Mantel and Haenszel (1959). An alternative method is to use logistic regression on the binomial variates. Breslow (1976, 1981) developed an analysis of these data using exact conditional inference and the non-central hypergeometric distribution which also allows for modeling of the dependence of the logarithm of the odds ratio on confounders. This relies on conditioning on the observed level of exposure in cases and controls in each table. It was also demonstrated that ordinary logistic regression provides badly biased results in sparse data, and even inconsistent results for some asymptotic situations. For these data, a method of analysis which would provide estimates near those given by the exact methods without the need for cumbersome exact methods would be ideal.

This chapter will examine the saddlepoint methods applied to log odds ratio regression to determine if they fill such a role. We consider two uses of the saddlepoint methods. Following Breslow (1982), Breslow and Cologne (1986) and Levin (1990) we examine the maximum likelihood estimator that arises from the modified profile likelihood, and its properties with respect to consistency and asymptotic variance, and then following Pierce and Peters (1992) and Bedrick and Hill (1992) we examine the properties of the tail probability approximation and confidence limits for the log odds ratio.

**Example** Consider the data on oesophageal cancer in the French department of Ille-et-Vilaine analyzed by Breslow and Day (1980). These data consisted of 200 cases of oesophageal cancer diagnosed between January 1972 and April 1974 and 775 adult male controls who provided enough data for analysis. Both cases and controls were administered an interview which determined their alcohol and tobacco consumption as well as their food intake. The data were split into six groups stratified by age, and the alcohol consumption variable was dichotomized. This gave rise to six $2 \times 2$ tables which were analyzed using exact conditional inference and unconditional logistic regression.

## *4.2 The Model*

Zelen (1971) first suggested a regression model for the odds ratio in a sequence of $2 \times 2$ tables. We first derive the model for a common odds ratio, and later generalize to the problem of odds ratio regression.

To rigorously formulate the problem, consider $k$ strata or pairs of binomials. Assume that in each matched set there are $n_k$ cases and $m_k$ controls. Let $X_k$ and $Y_k$ be binomial, with denominators $n_k, m_k$ and success probabilities $p_{1k}, p_{0k}$ for $k = 1, \ldots, K$.

To start, we assume a common odds ratio $\psi = \exp(\beta)$ in the tables, which gives $\text{logit}(p_{0k}) = \alpha_k$, $\text{logit}(p_{1k}) = \alpha_k + \beta$ for some $\alpha_k$. This gives rise to the following binomial likelihood:

$$L(\beta, \alpha) = \prod_{k=1}^{K} \frac{e^{\alpha_k y_k}}{(1+e^{\alpha_k})^{m_k}} \frac{e^{(\alpha_k+\beta)x_k}}{(1+e^{\alpha_k+\beta})^{n_k}}, \tag{4.1}$$

where $\alpha = (\alpha_1, \ldots, \alpha_K)^T$ The $\alpha_k$ can be thought of as nuisance parameters – we are interested in $\beta$.

### *4.2.1 Unconditional Inference*

Inference about $\beta$ can be done using the above likelihood unconditionally (Breslow, 1981). However, the likelihood depends on the estimation of the $\alpha_k$, and when there are many tables with few observations, the standard logistic regression methods yield biased results. To use this likelihood, we first solve

$$\dot{l}_{\alpha_k}(\beta, \alpha) = \frac{\partial l}{\partial \alpha_k} = 0 \tag{4.2}$$

where $\dot{l}_{\alpha_k}$ refers to the derivative of the log likelihood function with respect to $\alpha_k$, which gives a profile likelihood for $\beta$ as follows:

$$L_P(\beta) \propto \prod_{k=1}^{K} \frac{e^{\beta x_k}}{(1+e^{\hat{\alpha}_k(\beta)+\beta})^{n_k}}. \tag{4.3}$$

We write $\hat{\alpha}_k = \hat{\alpha}_k(\beta) = \hat{\alpha}(\beta, m_k, n_k, t_k)$ to denote the maximum likelihood estimator of $\alpha_k$ for a given value of $\beta$ and the $k$th set of given values of the marginal totals and $\hat{\alpha} = (\hat{\alpha}_1, \ldots, \hat{\alpha}_K)^T$. This likelihood is easily maximized for $\beta$ using standard generalized linear model procedures. It gives rise to several test statistics for testing the hypothesis $H_0 : \beta \leq \beta_0$, including of primary interest the log-likelihood ratio test statistic

$$2 \log \Lambda = (\hat{\beta} - \beta) \sum_k x_k - \sum_k n_k \log \left( \frac{1 + e^{\hat{\alpha}_k(\beta)+\beta}}{1 + e^{\hat{\alpha}_k(\beta_0)+\beta_0}} \right) \tag{4.4}$$

and the Wald statistic

$$W = (\hat{\beta} - \beta_0)^2 I_{11 \cdot 2}(\hat{\beta}), \tag{4.5}$$

where

$$I_{11 \cdot 2} = \sum_k n_k \left(\frac{\partial \alpha_k}{\partial \beta} + 1\right) \frac{e^{\alpha_k+\beta}}{(1+e^{\alpha+\beta})^2},$$

which is equal to $I_{11} - I_{12}I^{22}I_{21}$; $I$ is the overall information for $\beta$ and $\alpha$.

### *4.2.2 Conditional Inference*

Breslow (1976, 1981) noted the bias in asymptotic methods for analysis of this problem in the presence of a number of nuisance parameters which increases with sample size. He proposed (1976) a model which conditions on the statistics for the $\alpha_k$ which are the marginal totals of the $2 \times 2$ tables and uses the exact conditional distribution of the sufficient statistic for $\beta$ given the other sufficient statistics.

The sufficient statistics for $\beta$ and $\alpha_k$ are $\sum x_k$ and $t_k = x_k + y_k$ respectively. The exact conditional distribution of $x_k$ given $n_k$, $m_k$ and $t_k$ is non-central hypergeometric

$$p(X_k = x | n_k, m_k, t_k) = \frac{\binom{n_k}{x}\binom{m_k}{t_k - x} e^{\beta x}}{\sum_{u=0}^{t_k} \binom{n_k}{u}\binom{m_k}{t_k - u} e^{\beta u}}. \tag{4.6}$$

Inference can be based on the hypergeometric likelihood that arises from this probability. Maximization of this likelihood, while more difficult than maximization of (4.3), is feasible using a variety of computer packages (e.g., SAS, StatXact, EGRET).

Tail probabilities and hence test statistics and confidence intervals for this distribution are usually calculated directly, rather than asymptotically, using the equations

$$P(X \in R) = \sum_{R} \prod \frac{\binom{n_k}{x}\binom{m_k}{t_k - x} e^{\beta x}}{\sum_{u=0}^{t_k} \binom{n_k}{u}\binom{m_k}{t_k - u} e^{\beta u}}, \tag{4.7}$$

where $R$ is a set contained in the conditional sample space.

This equation can be solved using the network algorithms of Mehta, Patel and Gray(1985). Approximate probabilities can be derived with a normal approximation to the exact density, using only the mean and variance of the exact density. This can be done with either exact mean and variance or approximate values (Breslow and Cologne, 1986). However, the computational difficulty of repeatedly solving high-order polynomials prevents the use of the exact methods for a substantial number of tables, while the normal approximation to the exact distribution does not provide accurate confidence limits.

### *4.2.3 Alternative Methods*

Mantel and Haenszel (1959) developed a method for estimating the odds ratio using a simple formula of the ratio of sums of cross-table products

$$\hat{\beta}_{MH} = \log(\hat{\psi}_{MH}) = \log\left\{\frac{\sum_k x_k(m_k - t_k + x_k)/N_k}{\sum_k(n_k - x_k)(t_k - x_k)/N_k}\right\}. \tag{4.8}$$

A variety of forms for estimators of the variance of this estimator have been developed with varying properties. An estimate with good properties under both asymptotic situations of primary interest was developed independently by Robins, Breslow and Greenland (1986) and by Phillips and Holland (1987). Letting $R_k = x_k(m_k - t_k + x_k)/N_k$, and $R = \sum_k R_k$, we have

$$\begin{aligned} &Var(\hat{\beta}_{MH}) \doteq \\ &\quad \frac{1}{R^2}\sum_k \frac{1}{{N_k}^2}\left\{x_k(m_k - t_k + x_k) + \hat{\psi}_{MH}(n_k - x_k)(t_k - x_k)\right\} \\ &\times \quad \left\{x_k + (m_k - t_k + x_k) + \hat{\psi}_{MH}((n_k - x_k) + (t_k - x_k))\right\}. \end{aligned} \tag{4.9}$$

McCullagh (1984) developed a set of approximations to the mean and variance of the noncentral hypergeometric distribution. Let $X$ be a noncentral hypergeometric random variable with log odds ratio parameter $\beta$, sample sizes $n$ and $m$,total number of successes $t$ and total sample size $N = m+n$. The approximate mean $\tilde{\mu} = \hat{E}(X : \beta)$ and variance $\tilde{v} = \widehat{Var}(X : \beta)$ of $X$ are defined to satisfy the equations

$$\tilde{\mu}(m - t + \tilde{\mu}) + \tilde{v} = e^{\beta}\left\{(t - \mu)(n - \mu) + \tilde{v}\right\}, \tag{4.10}$$

and

$$\tilde{v} = (\frac{N}{N-1})\left\{\frac{1}{\tilde{\mu}} + \frac{1}{t - \tilde{\mu}} + \frac{1}{n - \tilde{\mu}} + \frac{1}{m - t + \tilde{\mu}}\right\}^{-1}. \tag{4.11}$$

These equations can be used on each table to approximate the mean, and then an estimator for the odds ratio parameters can be derived using Newton-Raphson algorithms for generalized linear models. Breslow and Cologne (1986) examined this estimator and its performance in some detail.

### *4.2.4 The Double Saddlepoint Approximation*

The saddlepoint approximations and the modified profile likelihood for this model are based on the unconditional likelihood and tail probabilities. The modified profile likelihood (3.4) takes the form

$$L_{MP}(\beta|t_1,\ldots,t_k) \propto \prod_{k=1}^{K}\left\{\frac{e^{(\hat{\alpha}_k+\beta)X_k}}{(1+e^{\hat{\alpha}_k+\beta})^{n_k}}\right\}|I_\alpha(\beta,\hat{\alpha}(\beta))|^{1/2}, \tag{4.12}$$

where $\hat{\alpha}_k$ depends on $\beta$ and $t_k$ (recall that $\hat{\alpha} = (\hat{\alpha}_1,\ldots,\hat{\alpha}_K)^T$). The estimator $\hat{\alpha}_k$ solves the maximum likelihood equation

$$\dot{l}_{\alpha_k}(\beta,\alpha) = \frac{\partial l}{\partial \alpha_k} = 0,$$

which is a commonly known quadratic equation. It is worth noting again that (4.12) is the unconditional likelihood multiplied by a modification term that corrects for the conditioning. The modifier is

$$|I_\alpha(\beta,\hat{\alpha}(\beta))|^{1/2}. \tag{4.13}$$

The lower right block of the information matrix, $I_\alpha(\beta,\alpha)$ is a $K \times K$ diagonal matrix with components

$$\begin{aligned} -\frac{\partial^2}{\partial\alpha_k^2}K(\beta,\alpha) &= \frac{\partial^2}{\partial\alpha_k^2}\sum_k\left\{n_k\log(1+e^{\alpha_k+\beta}) + m_k\log(1+e^{\alpha_k})\right\} \\ &= \frac{n_k e^{\alpha_k+\beta}}{(1+e^{\alpha_k+\beta})^2} + \frac{m_k e^{\alpha_k}}{(1+e^{\alpha_k})^2}, \end{aligned} \tag{4.14}$$

which makes the modifying term in the likelihood

$$|I_\alpha(\beta,\hat{\alpha}(\beta))|^{1/2} = \prod_k\left\{\frac{n_k e^{\hat{\alpha}_k+\beta}}{(1+e^{\hat{\alpha}_k+\beta})^2} + \frac{m_k e^{\hat{\alpha}_k}}{(1+e^{\hat{\alpha}_k})^2}\right\}^{1/2}. \tag{4.15}$$

The maximum modified profile likelihood estimator can be determined by maximizing (4.12).

To maximize (4.12) it is useful to derive the score statistic and information matrix. Levin (1990) derived the score and information statistics for grouped binomial data, and we have independently verified this result for this specific model. The score statistic is given by

$$\begin{aligned}\dot{l}_{MP}(\beta) &= \dot{l}_P(\beta) + \frac{1}{2}\frac{\partial}{\partial\beta}\left\{\sum_k \log\left(\frac{m_k e^{\hat{\alpha}_k}}{(1+e^{\hat{\alpha}_k})^2} + \frac{n_k e^{\hat{\alpha}_k+\beta}}{(1+e^{\hat{\alpha}_k+\beta})^2}\right)\right\} \\ &= \sum_k x_k - \sum_k \frac{n_k e^{\hat{\alpha}_k+\beta}}{1+e^{\hat{\alpha}_k+\beta}} \\ &+ \frac{1}{2}\sum_k \frac{\partial}{\partial\beta}\log\left(\frac{m_k e^{\hat{\alpha}_k}}{(1+e^{\hat{\alpha}_k})^2} + \frac{n_k e^{\hat{\alpha}_k+\beta}}{(1+e^{\hat{\alpha}_k+\beta})^2}\right) \qquad (4.16)\end{aligned}$$

and the information statistic is given by

$$\begin{aligned}I_{MP}(\hat{\beta}) &= I_P(\hat{\beta}) - \frac{1}{2}\frac{\partial^2}{\partial\beta^2}\left\{\sum_k \log\left(\frac{m_k e^{\hat{\alpha}_k}}{(1+e^{\hat{\alpha}_k})^2} + \frac{n_k e^{\hat{\alpha}_k+\beta}}{(1+e^{\hat{\alpha}_k+\beta})^2}\right)\right\} \\ &= \sum_k \frac{n_k e^{\hat{\alpha}_k+\beta}}{(1+e^{\hat{\alpha}_k+\beta})^2} \\ &- \frac{1}{2}\sum_k \frac{\partial^2}{\partial\beta^2}\log\left(\frac{m_k e^{\hat{\alpha}_k}}{(1+e^{\hat{\alpha}_k})^2} + \frac{n_k e^{\hat{\alpha}_k+\beta}}{(1+e^{\hat{\alpha}_k+\beta})^2}\right), \qquad (4.17)\end{aligned}$$

where $\dot{l}_P(\beta)$ and $I_P(\hat{\beta})$ refer to the score and information statistics for $\beta$ from the profile (unconditional) likelihood.

The tail probability approximations take on the forms of equations (3.13) and (3.14), with

$$\begin{aligned}w &= \text{sign}(\hat{\beta}-\beta)(l_P(\hat{\beta}) - l_P(\beta))^{1/2} \\ &= \text{sign}(\hat{\beta}-\beta)\left\{(\hat{\beta}-\beta)\sum_k x_k - \sum_k \log\left(\frac{1+e^{\hat{\alpha}_k(\beta)+\beta}}{1+e^{\hat{\alpha}_k(\hat{\beta})+\hat{\beta}}}\right)\right\}^{1/2}, \qquad (4.18)\end{aligned}$$

$$z = \left\{1 - \exp(-\hat{\beta}+\beta)\right\} I(\hat{\beta})_{11.2}, \qquad (4.19)$$

and

$$\rho = \left\{\frac{|I(\hat{\beta},\hat{\alpha}(\hat{\beta}))_{22}|}{|I(\beta,\hat{\alpha}(\beta))_{22}|}\right\}^{1/2}, \qquad (4.20)$$

where $I_{22}$ is the component of the information matrix corresponding to the $\alpha_k$. This matrix is diagonal, with diagonal components

$$I(\beta,\hat{\alpha}(\beta))_{kk} = n_k \frac{e^{\hat{\alpha}_k(\beta)+\beta}}{(1+e^{\hat{\alpha}_k(\hat{\beta})+\hat{\beta}})^2} + m_k \frac{e^{\hat{\alpha}_k(\beta)}}{(1+e^{\hat{\alpha}_k(\hat{\beta})})^2}. \qquad (4.21)$$

Recall that it is necessary to use slightly different formulae for discrete data. A continuity correction can be implemented by making appropriate corrections to the sufficient statistic $\sum_k x_k$ (perturbing by $\pm 1/2$) and $z$ (replacing $(1 - \exp(-\hat{\beta} + \beta))$ with $\exp((\hat{\beta} - \beta)/2) - \exp((-\hat{\beta} + \beta)/2))$.

When $\hat{\beta} - \beta$ is near zero, two approaches are possible. One can use equation (3.21), which is relatively easy to implement in this situation as the $I_{22}$ block of the information matrix is diagonal. However, the alternative method proposed by Bedrick and Hill, which simply chooses whether or not to use a continuity correction based on nearness to zero, also works well. In our simulations the method of Bedrick and Hill was used.

### *The Sequential Saddlepoint Approximation*

The sequential saddlepoint approximation can be easily evaluated for this model by evaluating (3.13) and (3.14) using $\hat{\beta}_{MP}$ rather than $\hat{\beta}$ and $L_{MP}$ rather than $L_P$. Since $L_{MP}$ depends only on $\beta$, the equations take on the form

$$P(T_1 \leq t | T_2, \theta_{10}) \doteq \Phi(w_{MP}) + \phi(w_{MP})\,(1/w_{MP} - 1/z_{MP})$$

and

$$P(T \leq t) \doteq \Phi \left\{ w_{MP} + \frac{1}{w_{MP}}(\log(w_{MP}/z_{MP})) \right\},$$

where $w_{MP}$ and $z_{MP}$ are the likelihood ratio and Wald statistics based on the modified profile likelihood. In this case, since it is relatively straightforward to calculate the modified profile likelihood, these tail probabilities are easy to calculate.

### *4.2.5 The Direct Saddlepoint Approximation*

An alternative approach to the double saddlepoint approximation is the direct saddlepoint approximation. In this case, we use a single saddlepoint approximation to the conditional distribution of $X$ given $\{T_k$, $m_k$, and $n_k\}$.

This approximation simply takes a saddlepoint approximation to equation (4.7). This is an exponential family probability distribution, with cumulant generating function

$$K(\lambda) = \sum_k \log \left\{ \sum_{u=0}^{t_k} \binom{n_k}{u} \binom{m_k}{t_k - u} e^{(\beta+\lambda)u} \right\} - \log \left\{ \sum_{u=0}^{t_k} \binom{n_k}{u} \binom{m_k}{t_k - u} e^{\beta u} \right\}. \quad (4.22)$$

The saddlepoint $\hat{\lambda} = \hat{\beta}_C - \beta$ where $\hat{\beta}_C$ is the exact conditional MLE. Unfortunately, the complications of performing exact inference for this distribution are due to calculation of the cumulant generating function (CGF). This means that point and interval estimation are not particularly less complicated than in the exact case. However, testing only requires evaluation of the CGF at $\hat{\beta}_C$ and at 0, which reduces complexity of test statistics and their distributions considerably.

### *4.2.6 Log Odds Ratio Regression*

The problem of estimating the odds ratio can also be extended to regression analysis of the log odds ratio, where the relationship of the odds ratio to table-level covariates is of interest. This is done, for example, when the tables are stratified by age and one wishes to examine the dependence of the odds ratio on age. This is done by simply replacing $\beta$ with $\beta^{*T} Z_k$, where $\beta^*$ is now a vector of parameters, and $Z_k$ is a vector of table-level covariates. Under this model, the full likelihood becomes

$$L(\beta^*, \alpha) = \prod_{k=1}^{K} \frac{e^{\alpha_k y_k}}{(1 + e^{\alpha_k})^{m_k}} \frac{e^{(\alpha_k + \beta^{*T} z_k) x_k}}{(1 + e^{\alpha_k + \beta^{*T} z_k})^{n_k}}. \quad (4.23)$$

This model contains the simpler problem of estimating the common odds ratio, when $\beta^*$ is a scalar and $Z$ is one. Hence, in the rest of the chapter, we refer directly to the odds ratio regression model except when the common odds ratio is specifically of interest.

**Example (cont'd)** Tables 4.1 and 4.2 present two analyses of the Ille-et-Vilaine data. In both analyses, the parameter of interest is the odds ratio for oesophageal

Table 4.1: Ille-et-Vilaine Data – Results of Saddlepoint Analysis - Log-Odds Ratio Regression

| | **Exact** | **Uncond.** | **McCull.** | **SP (Corr.)** |
|---|---|---|---|---|
| Intercept | 1.7026 | 1.7047 | 1.7027 | 1.7027 |
| SE(Int.) | 0.1998 | 0.2009 | 0.2002 | 0.1998 |
| Age | -0.1255 | -0.1260 | -0.1255 | -0.1255 |
| SE(Age) | 0.1878 | 0.1890 | 0.1883 | 0.1878 |

Table 4.2: Ille-et-Vilaine Data – Results of Saddlepoint Analysis - Common OR

| | **Exact** | **Uncond.** | **McCull.** | **SP (Corr.)** |
|---|---|---|---|---|
| Intercept | 1.6584 | 1.6699 | 1.6584 | 1.6584 |
| SE(Int.) | 0.1888 | 0.1899 | 0.1891 | 0.1890 |
| 95% CI | (1.273,2.049) | (1.300,2.044) | (1.273,2.049) | (1.282,2.029) |

cancer with alcohol consumption (dichotomized into high and low). Table 4.1 gives an analysis using log odds ratio regression where the odds ratio for a given table is of the form $\beta_0 + \beta_1 * (Z - 3.5)$ where $Z$ corresponds to age group, $Z = 1, \ldots, 6$. Table 4.2 presents an analysis of the same data using a common odds ratio for alcohol consumption. For the common odds ratio problem, tail probability approximations enable the calculation of 95% confidence intervals. In both cases, the saddlepoint inference results are essentially identical to the exact conditional inference results. Unconditional inference gives biased results, while the McCullagh estimator gives good results for point estimation but poor confidence interval estimates.

## *4.3 Asymptotic Bias and Variance of the Estimators*

There are two asymptotic frameworks in which the problem of estimating odds ratios is typically considered (Breslow, 1981). The first is the case of a fixed number of tables or strata and an increasing sample size in each table. This is the standard framework of most asymptotic theory and the estimators in general have good properties (consistency, asymptotic normality). The second framework is the case of a bounded sample size per table and an increasing number of tables. This situation occurs in several settings, in particular where the stratification is over families or other groups where increasing the sample size is done through addition of strata rather than through addition of subjects to individual strata.

### *4.3.1 Asymptotic Model I*

The first asymptotic model to consider is where the sample size in each table $N_k = m_k + n_k$ increases to infinity and the number of tables is constant. Under this asymptotic framework, the number of nuisance parameters is fixed and does not increase with sample size. Hence, as the sample size grows estimation of the nuisance parameters has a smaller effect on the estimation of the parameter of interest. Given this, provided the ratio of the number of cases to the number of subjects in each table $n_k/N_k$ converges to some $\lambda_k$ as $N_k$ and $n_k$ increase to infinity, the estimators based on unconditional maximum likelihood and on the conditional maximum likelihood are both consistent.

One can also examine the asymptotic variance of the estimators to determine performance. Let $\beta_0$ denote the true value of $\beta$. The unconditional ML estimator is best asymptotically normal (Gart, 1962) with asymptotic variance

$$Var^A(\hat{\beta}_U) = \frac{1}{\sum_k \left\{1/(n_k \frac{e^{\alpha+\beta_0}}{(1+e^{\alpha+\beta_0})^2}) + 1/(m_k \frac{e^{\alpha}}{(1+e^{\alpha})^2})\right\}^{-1}} \tag{4.24}$$

The conditional ML estimator is also asymptotically normal with variance

$$Var^A(\hat{\psi}_C) = \frac{1}{\sum_k Var_k(X : \beta_0)} \tag{4.25}$$

where $Var_k(X : \beta)$ denotes the variance of the exact conditional distribution for a given table (Andersen, 1970). Breslow (1981) demonstrated that these two estimators were asymptotically equivalent.

The saddlepoint estimator is derived by solving the score equation

$$\dot{l}_{MP}(\beta) = \sum_k x_k - \sum_k \frac{n_k e^{\hat{\alpha}_k+\beta}}{1+e^{\hat{\alpha}_k+\beta}} + \frac{1}{2}\sum_k \frac{\partial}{\partial\beta}\log\left\{\frac{m_k e^{\hat{\alpha}_k}}{(1+e^{\hat{\alpha}_k})^2} + \frac{n_k e^{\hat{\alpha}_k+\beta}}{(1+e^{\hat{\alpha}_k+\beta})^2}\right\} = 0.$$

The correction term

$$\frac{1}{2}\sum_k \frac{\partial}{\partial\beta}\log\left\{\frac{m_k e^{\hat{\alpha}_k}}{(1+e^{\hat{\alpha}_k})^2} + \frac{n_k e^{\hat{\alpha}_k+\beta}}{(1+e^{\hat{\alpha}_k+\beta})^2}\right\}$$

is equal to

$$\frac{1}{2}\sum_k \frac{\frac{m_k e^{\hat{\alpha}_k}(1-e^{\hat{\alpha}_k})\frac{\partial\hat{\alpha}_k}{\partial\beta}}{(1+e^{\hat{\alpha}_k})^3} + \frac{n_k e^{\hat{\alpha}_k+\beta}(1-e^{\hat{\alpha}_k+\beta})(\frac{\partial\hat{\alpha}_k}{\partial\beta}+1)}{(1+e^{\hat{\alpha}_k+\beta})^3}}{\frac{m_k e^{\hat{\alpha}_k}}{(1+e^{\hat{\alpha}_k})^2} + \frac{n_k e^{\hat{\alpha}_k+\beta}}{(1+e^{\hat{\alpha}_k+\beta})^2}}.$$

Now, $e^{\hat{\alpha}_k}/(1+e^{\hat{\alpha}_k}) \rightarrow e^{\alpha}/(1+e^{\alpha})$ and $e^{\hat{\alpha}_k+\beta}/(1+\epsilon^{\hat{\alpha}_k+\beta}) \rightarrow e^{\alpha+\beta}/(1+e^{\alpha+\beta})$ provided that $n_k/N_k \rightarrow \lambda_k$ for some $\lambda_k$. Hence,

$$\left(\frac{1}{\sqrt{M}}\right)\frac{1}{2}\sum_k \frac{\partial}{\partial\beta}\log\left\{\frac{m_k e^{\hat{\alpha}_k}}{(1+e^{\hat{\alpha}_k})^2} + \frac{n_k e^{\hat{\alpha}_k+\beta}}{(1+e^{\hat{\alpha}_k+\beta})^2}\right\}$$

is $o_p(1)$ where $M$ is the total sample size $\sum_k N_k$. This implies that the saddlepoint score equation is asymptotically equivalent to

$$\sum_k x_k - \sum_k \frac{n_k e^{\hat{\alpha}_k+\beta}}{1+e^{\hat{\alpha}_k+\beta}} = 0,$$

the score equation for the unconditional estimator. Using this we will demonstrate that

$$\left|\hat{\beta}_{MP} - \hat{\beta}_U\right| = o_p(1) \tag{4.26}$$

and hence that $\hat{\beta}_{MP}$ and $\hat{\beta}_U$ are asymptotically equivalent. First note that

$$\begin{aligned} \dot{l}_{MP}(\beta) &= \sum_k x_k - \sum_k \frac{n_k e^{\hat{\alpha}_k+\beta}}{1+e^{\hat{\alpha}_k+\beta}} + \frac{1}{2}\sum_k \frac{\partial}{\partial\beta}\log\left\{\frac{m_k e^{\hat{\alpha}_k}}{(1+e^{\hat{\alpha}_k})^2} + \frac{n_k e^{\hat{\alpha}_k+\beta}}{(1+e^{\hat{\alpha}_k+\beta})^2}\right\} \\ &= \dot{l}_P(\beta) + \frac{1}{2}\sum_k \frac{\partial}{\partial\beta}\log\left\{\frac{m_k e^{\hat{\alpha}_k}}{(1+e^{\hat{\alpha}_k})^2} + \frac{n_k e^{\hat{\alpha}_k+\beta}}{(1+e^{\hat{\alpha}_k+\beta})^2}\right\} \qquad (4.27) \end{aligned}$$

and hence that

$$\frac{1}{\sqrt{M}}\dot{l}_{MP}(\beta) = \frac{1}{\sqrt{M}}\dot{l}_P(\beta) + o_p(1). \tag{4.28}$$

Further,

$$\ddot{l}_{MP}(\beta) = \ddot{l}_P(\beta) + \frac{1}{2}\sum_k \frac{\partial^2}{\partial\beta^2}\log\left\{\frac{m_k e^{\hat{\alpha}_k}}{(1+e^{\hat{\alpha}_k})^2} + \frac{n_k e^{\hat{\alpha}_k+\beta}}{(1+e^{\hat{\alpha}_k+\beta})^2}\right\} \tag{4.29}$$

and

$$\frac{1}{M}\ddot{l}_{MP}(\beta) = \frac{1}{M}\ddot{l}_P(\beta) + o_p(1) \tag{4.30}$$

(it is easy to verify that the second derivative of the saddlepoint correction is $o_p(M^{-1})$ – see appendix B). Finally,

$$\frac{1}{M}\frac{\partial^3}{\partial\beta^3}l_{MP}(\beta) = \frac{1}{M}\frac{\partial^3}{\partial\beta^3}l_P(\beta) + o_p(1) \tag{4.31}$$

(see appendix B for details).

We first demonstrate consistency of $\hat{\beta}_{MP}$. First define $\dot{l}_P^{kj}(\beta)$ such that

$$\begin{aligned} \dot{l}_P(\beta) &= \sum_k x_k - \sum_k \frac{n_k e^{\hat{\alpha}_k+\beta}}{1+e^{\hat{\alpha}_k+\beta}} \\ &= \sum_{k=1}^{K}\sum_{j=1}^{n_k}\left\{x_{kj} - \frac{n_k e^{\hat{\alpha}_k+\beta}}{1+e^{\hat{\alpha}_k+\beta}}\right\} \\ &= \sum_{k=1}^{K}\sum_{j=1}^{n_k}\dot{l}_P^{kj}(\beta). \qquad (4.32) \end{aligned}$$

By the central limit theorem,

$$\sum_{k=1}^{K} \frac{1}{n_k} \sum_{j=1}^{n_k} \dot{l}_P^{kj}(\beta) \rightarrow^{as} \sum_{k=1}^{K} E_\beta \left\{ \dot{l}_P^{kj}(\beta) \right\}. \tag{4.33}$$

Now, $\dot{l}_P^{kj}(\beta)$ and hence $E_\beta \left\{ \dot{l}_P^{kj}(\beta) \right\}$ are monotone decreasing in $\beta$, and $\hat{\beta}_P$ is consistent, so for any $\varepsilon > 0$ such that $\beta_0 + \varepsilon$ is a continuity point of the distribution, (using (4.33))

$$P(\hat{\beta}_P \leq \beta_0 + \varepsilon) = P\left\{ \sum_{k=1}^{K} \frac{1}{n_k} \sum_{j=1}^{n_k} \dot{l}_P^{kj}(\beta_0 + \varepsilon) \leq 0 \right\} \rightarrow 1, \tag{4.34}$$

and for $\varepsilon > 0$ such that $\beta_0 - \varepsilon$ is a continuity point of the distribution,

$$P(\hat{\beta}_P \leq \beta_0 - \varepsilon) = P\left\{ \sum_{k=1}^{K} \frac{1}{n_k} \sum_{j=1}^{n_k} \dot{l}_P^{kj}(\beta_0 - \varepsilon) \leq 0 \right\} \rightarrow 0. \tag{4.35}$$

Further,

$$\begin{aligned} P(\hat{\beta}_{MP} \leq \beta_0 + \varepsilon) &= P\left\{ \dot{l}_{SP}(\beta_0 + \varepsilon) \leq 0 \right\} \\ &= P\left\{ \sum_{k=1}^{K} \frac{1}{n_k} \sum_{j=1}^{n_k} \dot{l}_P^{kj}(\beta_0 + \varepsilon) + o_p(1) \leq 0 \right\} \rightarrow 1, \end{aligned} \tag{4.36}$$

and similarly,

$$P(\hat{\beta}_{MP} \leq \beta_0 - \varepsilon) = P\left\{ \sum_{k=1}^{K} \frac{1}{n_k} \sum_{j=1}^{n_k} \dot{l}_P^{kj}(\beta_0 - \varepsilon) + o_p(1) \leq 0 \right\} \rightarrow 0. \tag{4.37}$$

To examine asymptotic equivalence, we expand

$$\dot{l}_{MP}(\hat{\beta}_{MP}) = \dot{l}_{MP}(\beta) + (\hat{\beta}_{MP} - \beta)\ddot{l}_{MP}(\beta) + (1/2)(\hat{\beta}_{MP} - \beta)^2 \frac{\partial^3}{\partial \beta^3} l_{MP}(\beta^*) \tag{4.38}$$

where $\beta^*$ is in between $\beta$ and $\hat{\beta}_{MP}$. The left hand side of equation (4.38) is by definition equal to zero, so we rearrange to write

$$\begin{aligned} &\sqrt{M}(\hat{\beta}_{MP} - \beta) \\ &= \frac{(1/\sqrt{M})\dot{l}_{MP}(\beta)}{-(1/M)\ddot{l}_{MP}(\beta)} + \frac{(1/(2M))(\hat{\beta}_{MP} - \beta)\frac{\partial^3}{\partial \beta^3} l_{MP}(\beta^*)}{-(1/M)\ddot{l}_{MP}(\beta)} \\ &= \frac{(1/\sqrt{M})\dot{l}_P(\beta) + o_p(1)}{-(1/M)\ddot{l}_P(\beta) + o_p(1)} + \frac{(1/(2M))(\hat{\beta}_{MP} - \beta)\frac{\partial^3}{\partial \beta^3} l_P(\beta^*) + o_p(1)}{-(1/M)\ddot{l}_P(\beta) + o_p(1)}. \end{aligned} \tag{4.39}$$

Standard results on exponential families (Lehmann, 1983, Ch. 6) show that

$$(1/\sqrt{M})\dot{l}_P(\beta) \rightarrow N(0, I_P(\beta)),$$

that $\ddot{l}_P(\beta)$ is strictly negative, that

$$-(1/M)\ddot{l}_P(\beta) \rightarrow I_P(\beta),$$

and that

$$\frac{\partial^3}{\partial\beta^3} l_P(\beta^*)$$

is bounded in probability. The implication of this and the consistency of $\hat{\beta}_{MP}$ is that

$$\sqrt{M}(\hat{\beta}_{MP} - \beta) \rightarrow N(0, \frac{1}{I_P(\beta)}), \tag{4.40}$$

and further, that

$$\begin{aligned} &\sqrt{M}(\hat{\beta}_{MP} - \hat{\beta}_P) \\ = \;& \frac{(1/\sqrt{M})\dot{l}_P(\beta) + o_p(1)}{-(1/M)\ddot{l}_P(\beta) + o_p(1)} + o_p(1) - \frac{(1/\sqrt{M})\dot{l}_P(\beta)}{-(1/M)\ddot{l}_P(\beta)} + o_p(1) \\ = \;& o_p(1). \end{aligned} \tag{4.41}$$

Hence, since all three estimators are consistent and have the same asymptotic variance (and hence are asymptotically equivalent) under this asymptotic framework, the unconditional ML estimator is the most useful estimator. It is consistent and asymptotically efficient, and is the simplest estimator to calculate.

### *4.3.2 Asymptotic Model II*

The second asymptotic framework is more complicated. Here, the number of nuisance parameters increases linearly with sample size. We assume that the sample sizes in each stratum are uniformly bounded under this model, which means that there is a finite limit to the number of marginal configurations (values for the marginal totals). Let $I$ denote the number of marginal configurations and suppose that there are $K_i$

of each type, $i = 1, \ldots, I$. We use a two-step method (Breslow, 1981) to arrive at the limiting distributions of the statistics. We first determine the limiting frequencies $\pi_i = \lim_{K\to\infty}(K_i/K)$ of each type of configuration, and then we determine the limiting distribution by using the fact that conditional on the marginal totals and the $K_i$ the random variables $X_{ij}$ giving rise to the tables are independently hypergeometrically distributed.

*Properties of Current Estimators*

Breslow (1981) and Breslow and Cologne (1986) gave several results pertaining to the unconditional estimator, the Mantel-Haenszel estimator and the estimator based on the McCullagh approximation. They will be summarized here.

Breslow (1981), using a Taylor expansion of the true expectation and the unconditional expectation, showed that the unconditional estimator is badly biased. For the situation where the number of cases and controls are fixed at $n$ and $m$ per table respectively (making the total number of observations per table $n + m = N$), it was demonstrated that to first order, for $\beta$ near zero,

$$\beta_{UC} \sim \frac{N}{N-1}\beta \tag{4.42}$$

where $\beta_{UC}$ is the limiting value of the unconditional estimator amd $\beta$ is the true value. This is a fairly serious bias, especially considering that $\beta$ is the logarithm of the odds ratio, and hence, for example, when $m = n = 1$, the odds ratio estimator converges to the square of the odds ratio. Breslow and Cologne (1986) used similar techniques to show that to first order, the McCullagh approximate estimator converges to a limiting value $\beta_{MC}$ where for small $\beta$

$$\beta_{MC} \sim \beta. \tag{4.43}$$

*The Double Saddlepoint MLE*

The asymptotic bias of the double saddlepoint estimator can be determined by applying the arguments above. Define

$$E_k^{SP}(X;\beta) = \frac{n_k e^{\hat{\alpha}_k+\beta}}{1+e^{\hat{\alpha}_k+\beta}} - \frac{1}{2}\frac{\partial}{\partial\beta}\log\left\{\frac{m_k e^{\hat{\alpha}_k}}{(1+e^{\hat{\alpha}_k})^2} + \frac{n_k e^{\hat{\alpha}_k+\beta}}{(1+e^{\hat{\alpha}_k+\beta})^2}\right\}. \tag{4.44}$$

The saddlepoint estimator $\hat{\beta}_{MP}$ satisfies

$$\sum_k x_k = \sum_k E_k^{SP}(X;\hat{\beta}_{MP}). \tag{4.45}$$

Given that the number of possible marginal totals is fixed at $I$, we can rewrite equation (4.45) as

$$\sum_{i=1}^{I}\sum_{j=1}^{K_i} x_{ij} = \sum_{i=1}^{I} K_i E_i^{SP}(X;\hat{\beta}_{MP}) \tag{4.46}$$

where $E_i^{SP}(X;\beta)$ is indexed by $i$ instead of $k$ (it only depends on the marginal totals) and $\sum_j x_{ij}$ refers to the sum of the values of $X$ for the $J$ tables with the $i$th marginal configuration. Rearranging equation (4.46) gives

$$\sum_{i=1}^{I}\frac{K_i}{K}\left(\frac{1}{K_i}\sum_{j=1}^{K_i} x_{ij}\right) = \sum_{i=1}^{I}\frac{K_i}{K}E_i^{SP}(X;\hat{\beta}_{MP}). \tag{4.47}$$

Taking limits on both sides and noting that the expression in parentheses on the left hand side is the observed value of the mean of $K_i$ independent identically distributed random variables (which converges to $E_i(X;\beta)$, the expectation of a non-central hypergeometric random variable with the $i$th marginal configuration) and that $K_i/K$ converges to $\pi_i$ by assumption, we have a limiting equation

$$\sum_{i=0}^{I}\pi_i E_i(X;\beta) = \sum_{i=0}^{I}\pi_i E_i^{SP}(X;\beta_{MP}) \tag{4.48}$$

where $\beta_{MP}$ is a limit point of the $\hat{\beta}_{MP}$. It is not generally true that $\beta_{MP} = \beta$, since $E_i$ and $E_i^{SP}$ are not equal for $\beta$ different from zero.

Levin (1990) suggests a correction to expression (4.44) and to the saddlepoint likelihood, score statistic and information statistic, to improve the performance near

$\beta = 0$. This correction replaces the $\frac{1}{2}$ in equations (4.12), (4.16), (4.17) and (4.44) with $\frac{1}{2}(n_k + m_k)/(n_k + m_k - 1)$. Levin's correction has the effect that the modified profile likelihood and its first derivative agree with those of the exact conditional likelihood. This does improve performance of the estimator for small $\beta$, as will be demonstrated numerically later.

By expanding both sides of equation (4.48) in a Taylor expansion about $\beta = 0$ we can examine the asymptotic bias as a function of $\beta$. Levin's correction to the likelihood leads to the following result on the limiting value of the saddlepoint estimator.

**Theorem 4.3.1** *Consider a single $\beta$ (a common odds ratio throughout the tables). The saddlepoint maximum likelihood estimator from the corrected likelihood converges to a locally unique limit point $\beta_{MP}$ where for $\beta \to 0$,*

$$\beta_{MP} = \beta + o(\beta)$$

*Proof.* For the proof the index of summation will be $i$ running over the possible marginal configurations. The estimator $\hat{\alpha}_i(\beta)$ refers to the solution to equation (4.2) for the $i$th marginal configuration. The asymptotic mean of $X$ for the $i$th marginal configuration is given by

$$E_i^A(X;\beta) = \frac{n_i e^{\hat{\alpha}_i(\beta)+\beta}}{1 + e^{\hat{\alpha}_i(\beta)+\beta}}.$$

Thus,

$$E_i^{SP}(X;\beta) = E_i^A(X;\beta) - \frac{1}{2}\frac{\partial}{\partial\beta}\log\left\{\frac{m_i e^{\hat{\alpha}_i(\beta)}}{(1+e^{\hat{\alpha}_i(\beta)})^2} + \frac{n_i e^{\hat{\alpha}_i(\beta)}}{(1+e^{\hat{\alpha}_i(\beta)})^2}\right\}. \tag{4.49}$$

We know that

$$\sum \pi_i E_i^A(X;\beta) = \sum \pi_i \left\{\frac{n_i t_i}{N_i} + \frac{n_i m_i t_i (N_i - t_i)}{N_i^3}\beta + \ldots\right\}, \tag{4.50}$$

(Breslow, 1981), which means that to expand $E_i^{SP}(X;\beta)$ about zero we need only to derive the expansion of the second term. Expanding the right hand side of equation

(4.48) gives

$$\sum_{i=0}^{I} \pi_i E_i^{SP}(X;\beta_{MP}) = \sum_{i=0}^{I} \pi_i \left\{ E_i^{SP}(X;0) + \frac{\partial}{\partial\beta} E_i^{SP}(X;\beta)\Big|_{\beta=0} \beta_{MP} + \ldots \right\}. \tag{4.51}$$

Substantial algebra is involved in taking derivatives of the expectation. Several basic facts enable simplification of the derivatives at zero. First of all,

$$\begin{aligned} E_i^{SP}(X;0) &= E_i^{A}(X;0) \\ &- \frac{1}{2}\frac{N_i}{N_i-1}\frac{\partial}{\partial\beta}\log\left\{\frac{m_i e^{\hat{\alpha}_i(\beta)}}{(1+e^{\hat{\alpha}_i(\beta)})^2} + \frac{n_i e^{\hat{\alpha}_i(\beta)+\beta}}{(1+e^{\hat{\alpha}_i(\beta)+\beta})^2}\right\}\Bigg|_{\beta=0}. \end{aligned} \tag{4.52}$$

Now, at $\beta = 0$, we can solve the equation $\dot{l}_\alpha(\alpha, 0) = 0$ to determine $\hat{\alpha}_i(0)$. We have

$$\dot{l}_\alpha(\alpha,\beta) = t_i - m_i\frac{\exp(\alpha)}{1+\exp(\alpha)} - n_i\frac{\exp(\alpha+\beta)}{1+\exp(\alpha+\beta)}. \tag{4.53}$$

Setting the right hand side of equation (4.53) to zero and rearranging, we derive

$$t_i(1+e^{\hat{\alpha}_i(\beta)+\beta})(1+e^{\hat{\alpha}_i(\beta)}) = m_i e^{\hat{\alpha}_i(\beta)}(1+e^{\hat{\alpha}_i(\beta)+\beta}) + n_i e^{\hat{\alpha}_i(\beta)+\beta}(1+e^{\hat{\alpha}_i(\beta)}), \tag{4.54}$$

which is quadratic in $e^{\hat{\alpha}_i(\beta)}$ and hence easily solvable. At $\beta = 0$, the equation reduces to

$$t_i - m_i\frac{e^{\hat{\alpha}_i(0)}}{1+e^{\hat{\alpha}_i(0)}} - n_i\frac{e^{\hat{\alpha}_i(0)}}{1+e^{\hat{\alpha}_i(0)}} = 0, \tag{4.55}$$

which is easily simplified to

$$e^{\hat{\alpha}_i(0)} = \frac{t_i}{N_i - t_i}, \tag{4.56}$$

where $N_i = m_i + n_i$, implying that when $\beta = 0$ the best unconditional estimate for $p_0 = p_1 = e^{\hat{\alpha}_i(0)}/(1+e^{\hat{\alpha}_i(0)})$ is just the total number of exposures $t_i$ divided by the total number of subjects $N_i$, which agrees with the conditional estimate. Implicit differentiation gives

$$\frac{\partial\hat{\alpha}_i}{\partial\beta} = \frac{-n_i e^{\hat{\alpha}_i(\beta)+\beta}(1+e^{\hat{\alpha}_i(\beta)})^2}{n_i e^{\hat{\alpha}_i(\beta)+\beta}(1+e^{\hat{\alpha}_i(\beta)})^2 + m_i e^{\hat{\alpha}_i(\beta)}(1+e^{\hat{\alpha}_i(\beta)+\beta})^2}, \tag{4.57}$$

which simplifies to $-n_i/N_i$ at $\beta = 0$. A second implicit differentiation gives an expression for $\partial^2\hat{\alpha}_i/\partial\beta^2$, which is excessively complicated but is equal to $n_i m_i(2t_i - N_i)/N_i^3$ at $\beta = 0$. The details of this derivation and the derivation of the third derivative $\partial^3\hat{\alpha}_i/\partial\beta^3$ (for use later) can be found in appendix B.

The saddlepoint correction term and its first and second derivatives at $\beta = 0$ must be evaluated as well. The saddlepoint correction term in the score equation is equal to

$$\frac{1}{2}\frac{N_i}{N_i - 1}\frac{\partial}{\partial\beta}\log\left\{\frac{m_i e^{\hat{\alpha}_i(\beta)}}{(1+e^{\hat{\alpha}_i(\beta)})^2} + \frac{n_i e^{\hat{\alpha}_i(\beta)+\beta}}{(1+e^{\hat{\alpha}_i(\beta)+\beta})^2}\right\}. \tag{4.58}$$

Define

$$M_i(\beta) = \frac{m_i e^{\hat{\alpha}_i(\beta)}}{(1+e^{\hat{\alpha}_i(\beta)})^2} + \frac{n_i e^{\hat{\alpha}_i(\beta)+\beta}}{(1+e^{\hat{\alpha}_i(\beta)+\beta})^2}. \tag{4.59}$$

At $\beta = 0$, $M_i(0)$ simplifies to

$$m_i\frac{t_i(N_i - t_i)^2}{(N_i - t_i)N_i^2} + n_i\frac{t_i(N_i - t_i)^2}{(N_i - t_i)N_i^2} = \frac{t_i(N_i - t_i)}{N_i}. \tag{4.60}$$

The first derivative of $M_i$ can easily be calculated to be

$$m_i\frac{e^{\hat{\alpha}_i(\beta)}(1 - e^{\hat{\alpha}_i(\beta)})\frac{\partial\hat{\alpha}_i}{\partial\beta}}{(1+e^{\hat{\alpha}_i(\beta)})^3} + n_i\frac{e^{\hat{\alpha}_i(\beta)+\beta}(1 - e^{\hat{\alpha}_i(\beta)+\beta})(\frac{\partial\hat{\alpha}_i}{\partial\beta} + 1)}{(1+e^{\hat{\alpha}_i(\beta)+\beta})^3} \tag{4.61}$$

(see appendix B for details). At $\beta = 0$, the expression becomes

$$\frac{e^{\hat{\alpha}_i(0)}(1 - e^{\hat{\alpha}_i(0)})}{(1+e^{\hat{\alpha}_i(0)})^3}\left\{m_i\left.\frac{\partial\hat{\alpha}_i}{\partial\beta}\right|_{\beta=0} + n_i\left(1 + \left.\frac{\partial\hat{\alpha}_i}{\partial\beta}\right|_{\beta=0}\right)\right\}, \tag{4.62}$$

which simplifies to

$$\left.\frac{\partial M_i}{\partial\beta}\right|_{\beta=0} = \frac{e^{\hat{\alpha}_i(0)}(1 - e^{\hat{\alpha}_i(0)})}{(1+e^{\hat{\alpha}_i(0)})^3}\left\{\frac{-m_i n_i}{N_i} + \frac{n_i(N_i - n_i)}{N_i}\right\} = 0. \tag{4.63}$$

The second derivative of this correction term is very complicated and involves differentiating several products and ratios of functions of $m_i$,$n_i$,$N_i$, $t_i$, $t_i$, and $\beta$ with respect to $\beta$. At $\beta = 0$, this derivative is equal to

$$\left.\frac{\partial^2 M_i}{\partial\beta^2}\right|_{\beta=0} = \frac{-2m_i n_i t_i^2(N_i - t_i)^2}{N_i^5}. \tag{4.64}$$

From this, we can calculate $E_i^{SP}(X;0)$ and $\partial/\partial\beta E_i^{SP}(X;\beta)|_{\beta=0}$.

$$\begin{aligned}
E_i^{SP}(X;0) &= E_i^A(X;0) - \frac{1}{2}\frac{N_i}{N_i-1}\frac{\partial}{\partial\beta}\log M_i(\beta)\bigg|_{\beta=0} \\
&= E_i^A(X;0) - \frac{1}{2}\frac{N_i}{N_i-1}\frac{1}{M_i(0)}\frac{\partial}{\partial\beta}M_i(\beta)\bigg|_{\beta=0} \\
&= \frac{n_i t_i}{N_i} \qquad (4.65)
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial}{\partial\beta}E_i^{SP}(X;\beta)\bigg|_{\beta=0} &= \frac{\partial}{\partial\beta}E_i^A(X;\beta)\bigg|_{\beta=0} - \frac{1}{2}\frac{N_i}{N_i-1}\frac{\partial^2}{\partial\beta^2}\log M_i(\beta)\bigg|_{\beta=0} \\
&= \frac{\partial}{\partial\beta}E_i^A(X;\beta)\bigg|_{\beta=0} - \frac{1}{2}\frac{N_i}{N_i-1}\left\{\frac{M_i''(0)M_i(0)-[M_i'(0)]^2}{[M_i(0)]^2}\right\} \\
&= \frac{n_i m_i t_i(N_i-t_i)}{N_i^3} - \frac{1}{2}\frac{N_i}{N_i-1}\frac{M_i''(0)}{M_i(0)} \\
&= \frac{n_i m_i t_i(N_i-t_i)}{N_i^3} - \frac{1}{2}\frac{N_i}{N_i-1}\frac{-2m_i n_i t_i^2(N_i-t_i)^2}{N_i^5}\frac{N_i}{t_i(N_i-t_i)} \\
&= \frac{n_i m_i t_i(N_i-t_i)}{N_i^3} + \frac{1}{2}\frac{N_i}{N_i-1}\frac{2m_i t_i n_i(N_i-t_i)}{N_i^4} \\
&= \frac{n_i m_i t_i(N_i-t_i)}{N_i^3} + \frac{m_i t_i n_i(N_i-t_i)}{N_i^3(N_i-1)} \\
&= \frac{n_i m_i t_i(N_i-t_i)}{N_i^2(N_i-1)}. \qquad (4.66)
\end{aligned}$$

Note that this derivative is positive – this implies that in a neighborhood of $\beta = 0$, the solution to equation (4.48) is unique.

We can now combine equations (4.65) and (4.66) with equations (16) and (17) of Breslow(1981) to rewrite equation (4.48) as

$$\begin{aligned}
&\sum \pi_i\left\{\frac{n_i t_i}{N_i} + \frac{n_i m_i t_i(N_i-t_i)}{N_i^2(N_i-1)}\beta + \dots\right\} \\
&= \sum \pi_i\left\{E_i^{SP}(X;0) + \frac{\partial}{\partial\beta}E_i^{SP}(X;\beta)\bigg|_{\beta=0}\beta + \dots\right\} \\
&= \sum \pi_i\left\{\frac{n_i t_i}{N_i} + \frac{n_i m_i t_i(N_i-t_i)}{N_i^2(N_i-1)}\beta_{MP} + \dots\right\}. \qquad (4.67)
\end{aligned}$$

Now, when $\beta = 0$, $\beta_{MP} = 0$, so $\beta_{MP} \to 0$ as $\beta \to 0$. We can rewrite equation (4.67) as

$$\sum \pi_i \left\{ \frac{n_i m_i t_i (N_i - t_i)}{N_i^2 (N_i - 1)} \beta + \ldots \right\} = \sum \pi_i \left\{ \frac{n_i m_i t_i (N_i - t_i)}{N_i^2 (N_i - 1)} \beta_{MP} + \ldots \right\}. \tag{4.68}$$

Dividing both sides by $\beta$ and noting that all higher terms converge to zero as $\beta \to 0$, we have, to first order,

$$\lim_{\beta \to 0} \frac{\beta_{MP}}{\beta} = 1. \tag{4.69}$$

This concludes the proof. It should be noted that without Levin's (1990) correction, the saddlepoint estimator would be asymptotically biased to first order even as $\beta \to 0$.

Consider now the case with a fixed $m$ and $n$. The $o(\beta^2)$ term is determined in part by the second derivative of the left hand side of (4.48). The coefficient of $\beta^2$ in the expansion of the expectation from the conditional model is the third central moment of the central hypergeometric distribution, given by

$$\frac{nm(m-n)t(N-t)(N-2t)}{N^3(N-1)(N-2)}, \tag{4.70}$$

(McCullagh and Nelder, 1984, Ch. 7) and the coefficient of $\beta_{MP}^2$ in the expansion of the saddlepoint mean is given by

$$\frac{nm(m-n)t(N-t)(N-2t)}{N^5} - \frac{1}{2}\frac{N}{N-1} H(m,n,t), \tag{4.71}$$

where $H(m,n,t)$ is approximately given by

$$\frac{mn\,(m-n)\,(N-13t)}{N^3\,(N-t)} \tag{4.72}$$

(see appendix B for details). This makes the coefficient of $\beta_{MP}^2$ in the expansion of the saddlepoint (to order $N^{-2}$)

$$\frac{m\,(m-n)\,n\,(N^4 - 11N^3t - 2N^4t + 8N^3t^2)}{2\,(N-1)\,N^5\,(N-t)}. \tag{4.73}$$

It is apparent that for balanced samples with $m = n$, the coefficient of $\beta^2$ on the left hand side of equation (4.48) and the coefficient of $\beta_{MP}^2$ on the right hand side are both zero, and hence that $\beta_{MP} = \beta + o(\beta^2)$. When the samples are unbalanced, the saddlepoint limit $\beta_{MP}$ agrees with $\beta$ to $o(n^{-1})$ as shown above.

Another way to understand equation (4.48) is to solve it numerically in representative situations. To do so we consider the situation where the number of cases and controls in each stratum is fixed at $n$ and $m$ respectively. This means that the $I$ possible marginal configurations are entirely determined by the marginal total number of exposed subjects $t$ which runs from 0 to $m + n = N$, and that $\hat{\alpha}(\beta, m_k, n_k, t_k)$ depends only on $t$ and $\beta$. The estimator $\hat{\alpha}_t(\beta)$ refers to the solution of equation (4.2) for a given $t$ and $\beta$. Equation (4.48) can be rewritten as

$$\sum_{t=0}^{N} \pi_t E_t(X) = \sum_{t=0}^{N} \pi_t E_t^{SP}(X; \beta_{MP}), \tag{4.74}$$

where $N = m + n$ and

$$\begin{aligned} & E_t^{SP}(X;\beta) \\ & \quad = \frac{n e^{\hat{\alpha}_t(\beta)+\beta}}{1 + e^{\hat{\alpha}_t(\beta)+\beta}} \\ & \quad - \frac{1}{2} \frac{N}{N-1} \frac{\partial}{\partial \beta} \log \left\{ \frac{m e^{\hat{\alpha}_t(\beta)}}{(1 + e^{\hat{\alpha}_t(\beta)})^2} + \frac{n e^{\hat{\alpha}_t(\beta)+\beta}}{(1 + e^{\hat{\alpha}_t(\beta)+\beta})^2} \right\}. \end{aligned} \tag{4.75}$$

We also assume a constant odds ratio and baseline probability, which allows us to write

$$\pi_t = \sum_{u=0}^{N} \binom{n}{u} \binom{m}{n-u} p_1^u (1 - p_1)^{n-u} p_0^{t-u} (1 - p_0)^{m-t+u},$$

$p_0 = e^{\alpha_t}/(1 + e^{\alpha_t})$ and $p_1 = e^{\alpha_t+\beta}/(1 + e^{\alpha_t+\beta})$. Results are shown in tables 4.3 and 4.4 of numerical solutions of equation (4.74) for four different true odds ratios and a variety of sample sizes with $p_0 = 0.3$. These results were derived using the Newton-Raphson algorithm. These results are given in terms of the odds ratio $\psi = \exp(\beta)$ for comparison to earlier work by Breslow and Cologne. Results are presented both with the actual saddlepoint likelihood and with Levin's corrected likelihood.

It is readily apparent from this table that the saddlepoint maximum likelihood estimator of the odds ratio is nearly asymptotically unbiased except in cases where the odds ratio is large and the sample size per table is very small. Provided the odds ratio is lower than 5 or the sample sizes for each binomial trial are larger than 2, the estimator is essentially asymptotically unbiased. For very small samples, Levin's correction definitely improves the asymptotic performance.

*The Direct Saddlepoint Approximation*

The direct saddlepoint approximation uses the single saddlepoint approximation to the hypergeometric distribution. This is the limiting distribution for this asymptotic model, so the basic results of section 2.3.2 apply here. Hence, the error in the density approximation is $O(M^{-3/2})$ where $M$ refers to the total sample size $\sum_k m_k + n_k$, as compared to $O(M^{-1/2})$ for the normal approximation to the exact conditional density, and compared to biased estimates for other methods. This is a highly accurate density and tail probability approximation. However, given that implementation of this method is only substantially simpler than exact methods for testing hypotheses of the form $\beta = 0$ it is of limited use.

### 4.4 Small Sample Performance of the Approximations

It has been demonstrated numerically and analytically to first order above that asymptotically, the maximum likelihood estimators of the log odds ratio parameters based on the double saddlepoint likelihood are essentially asymptotically unbiased except in the case of extremely sparse data and odds ratios that are far from 1. However, the main benefit of the estimators is in small samples, where asymptotic theory does not really play a role. In addition, study of the performance in small samples enables examination of the tail probability approximations and the confidence limits. It is these approximations which should demonstrate the greatest improvement of the saddlepoint approximate methods over the unconditional methods, as they enable adjustment for skewness and kurtosis of the conditional likelihood.

Table 4.3: Asymptotic Means of Double Saddlepoint Maximum Likelihood Estimator - OR = 1.2 and 1.5

| n | m | OR=1.2 | | | |
|---|---|---|---|---|---|
| | | **Uncond.** | **McCull.** | **Corr.** | **Uncorr.** |
| 1 | 1 | 1.4400 | 1.2000 | 1.1995 | 1.2747 |
| 1 | 2 | 1.3167 | 1.2009 | 1.2010 | 1.2289 |
| 2 | 2 | 1.2752 | 1.2000 | 1.1999 | 1.2146 |
| 1 | 4 | 1.2575 | 1.2005 | 1.2007 | 1.2099 |
| 2 | 4 | 1.2452 | 1.2002 | 1.2003 | 1.2066 |
| 4 | 4 | 1.2317 | 1.2000 | 1.2000 | 1.2035 |
| 1 | 8 | 1.2285 | 1.2002 | 1.2003 | 1.2031 |
| 2 | 8 | 1.2251 | 1.2001 | 1.2002 | 1.2024 |
| 4 | 8 | 1.2203 | 1.2000 | 1.2001 | 1.2016 |
| 8 | 8 | 1.2147 | 1.2000 | 1.2000 | 1.2009 |

| n | m | OR=1.5 | | | |
|---|---|---|---|---|---|
| | | **Uncond.** | **McCull.** | **Corr.** | **Uncorr.** |
| 1 | 1 | 2.2500 | 1.5000 | 1.4938 | 1.7101 |
| 1 | 2 | 1.8522 | 1.5059 | 1.5047 | 1.5843 |
| 2 | 2 | 1.7181 | 1.5005 | 1.4992 | 1.5406 |
| 1 | 4 | 1.6702 | 1.5032 | 1.5039 | 1.5301 |
| 2 | 4 | 1.6309 | 1.5012 | 1.5013 | 1.5190 |
| 4 | 4 | 1.5899 | 1.5001 | 1.4999 | 1.5096 |
| 1 | 8 | 1.5835 | 1.5011 | 1.5016 | 1.5095 |
| 2 | 8 | 1.5727 | 1.5007 | 1.5009 | 1.5073 |
| 4 | 8 | 1.5580 | 1.5003 | 1.5003 | 1.5047 |
| 8 | 8 | 1.5413 | 1.5000 | 1.5000 | 1.5024 |

Table 4.4: Asymptotic Means of Double Saddlepoint Maximum Likelihood Estimator - OR = 2.0 and 5.0

| n | m | **OR=2.0** | | | |
|---|---|---|---|---|---|
| | | **Uncond.** | **McCull.** | **Corr.** | **Uncorr.** |
| 1 | 1 | 4.0000 | 2.0000 | 1.9600 | 2.4694 |
| 1 | 2 | 2.8969 | 2.0241 | 2.0107 | 2.2000 |
| 2 | 2 | 2.5289 | 2.0041 | 1.9957 | 2.0922 |
| 1 | 4 | 2.4210 | 2.0129 | 2.0126 | 2.0744 |
| 2 | 4 | 2.3148 | 2.0049 | 2.0035 | 2.0449 |
| 4 | 4 | 2.3148 | 2.0049 | 1.9995 | 2.0221 |
| 1 | 8 | 2.2118 | 2.0011 | 2.0052 | 2.0238 |
| 2 | 8 | 2.2024 | 2.0042 | 2.0029 | 2.0177 |
| 4 | 8 | 2.1368 | 2.0011 | 2.0010 | 2.0111 |
| 8 | 8 | 2.0962 | 2.0003 | 2.0000 | 2.0055 |

| n | m | **OR=5.0** | | | |
|---|---|---|---|---|---|
| | | **Uncond.** | **McCull.** | **Corr.** | **Uncorr.** |
| 1 | 1 | 25.0000 | 5.0000 | 4.0000 | 6.7600 |
| 1 | 2 | 12.8085 | 5.3320 | 4.8095 | 6.0560 |
| 2 | 2 | 9.1125 | 5.1818 | 4.9057 | 5.5394 |
| 1 | 4 | 8.2019 | 5.1895 | 5.0325 | 5.4452 |
| 2 | 4 | 7.2222 | 5.0911 | 4.9917 | 5.2582 |
| 4 | 4 | 6.4590 | 5.0449 | 4.9967 | 5.1401 |
| 1 | 8 | 6.3690 | 5.0541 | 5.0134 | 5.1348 |
| 2 | 8 | 6.1617 | 5.0354 | 5.0035 | 5.0980 |
| 4 | 8 | 5.8977 | 5.0211 | 5.0023 | 5.0655 |
| 8 | 8 | 5.6336 | 5.0113 | 5.0021 | 5.0370 |

The most straightforward way of examining the performance of the tail probabilities is through a simulation study. In this section, we give the design and results of a simulation study which examines coverage probabilities for confidence limits, power of hypothesis tests and relative error of the confidence limits (we compute this final statistic since the objective of the saddlepoint method is to approximate the conditional method, and hence the interesting statistic is how close the limits are to the exact ones). We also calculate the mean and estimated variance of the maximum likelihood estimators for the three methods and compare them.

It is of course well known that the exact conditional confidence limits are conservative, because their discreteness requires that the confidence coefficient for the interval be at least the pre-specified level. Since the saddlepoint limits are approximating the exact ones, we expect them to also be conservative.

The programs for these simulations were written in Fortran and in C. The network algorithm program for the exact conditional analysis is due to Thomas and Gart (1992). Uniform and Binomial random variables were generated using Fortran routines due to Wickman and Hill (Applied Statistics Alg. 183, 1982).

### *4.4.1 Design of the Simulation Study*

So as to perform confidence limit calculations, we simulate data with a common odds ratio. For simplicity, we assume that the number of tables is constant from repetition to repetition, and that the number of cases and controls in each table is fixed. There are several parameters that can be varied in the generation of data which may influence the performance of the methods. Chief among these are the baseline probability, the odds ratio, the number of tables, the number of exposed subjects per table and the number of unexposed subjects per table.

To effectively determine if and how any of these parameters affect the performance, we use a $2^{5-1}$ factorial design for the study, where we choose high and low levels of each parameter and use a design of 16 simulation runs to examine main effects and first

Table 4.5: Simulation Parameters

| **Factor** | **Parameter** | **Low Value** | **High Value** |
|---|---|---|---|
| 1 | Baseline P | 0.2 | 0.5 |
| 2 | Odds Ratio | 1.5 | 5.0 |
| 3 | No. Tables | 5 | 10 |
| 4 | No. Exposed | 5 | 10 |
| 5 | No. Unexposed | 5 | 10 |

order interactions. The parameter values for the simulation study are given in table 4.5. These parameter values give a range of table sizes and numbers of tables, and allow for examination of sparse data situations. The study is limited somewhat by the exact inference technique. While the network algorithm programs are reasonably fast in computing odds ratio estimators, they are slower in doing confidence limits, and for table sizes above 15 it is not practical to use them in a simulation study. Values of the simulation parameters are chosen to account for both moderate and extreme levels of the odds ratio and baseline probability, and two different small values for the numbers of tables and subjects.

The simulation design is a standard $2^{5-1}$ factorial design (Montgomery, 1991, Ch. 11). The study design is given in table 4.6. Provided the 3, 4 and 5-way interactions are negligible, we can derive estimates of the main effects and two-way interactions from this design.

The data were generated unconditionally (this is how they occur in practice) from a binomial model, where

$$P(Y = 1|X = x) = \frac{p_0\psi^x}{(1 + p_0 - p_0\psi)^x} \tag{4.76}$$

where $Y$ is a variable indicating disease status, $p_0$ is the baseline probability, $\psi$ is the odds ratio, and $X$ is an indicator variable for exposure status ($X = 1$ if exposed, 0 if not). Each dataset was composed of either 5 or 10 tables, with 5 or 10 exposed (unexposed) subjects. One thousand datasets were generated for each parameter combination.

Analysis of the data involved calculation of a point estimate, its estimated variance, and confidence intervals (95% and 99%) for $\beta$. Several methods were used to calculate estimates. Exact conditional inference and unconditional inference were both used to calculate estimates and tail probabilities. For the saddlepoint approximations the modified profile likelihood (as corrected by Levin) was used for the estimates, while formula (3.14) was used in several forms to calculate the saddlepoint tail probabilities - the standard formula (3.14) with and without continuity correction, and the sequential saddlepoint formula with and without continuity correction.

### *4.4.2 Results of the Simulation Study*

The complete results of the simulation study are presented in tables 4.7 through 4.11. Table 4.7 presents the mean estimate of the log odds ratio using the saddlepoint method, the unconditional method and exact conditional inference, as well as the variance of each of these estimates. The continuity corrected saddlepoint and the sequential saddlepoint methods provide the same inference here as the saddlepoint method. Table 4.8 gives the coverage probability for a 95% confidence interval for the odds ratio. Table 4.9 gives the number of times (out of 1000) that the confidence interval from each of the four approximations led to a different conclusion from that of the exact conditional confidence interval. Tables 4.10 and 4.11 give the mean and maximum relative error of the confidence limits (both low and high). The relative error is the absolute value of the difference between the exact and approximate limits divided by the exact limit,

$$RE(p) = \frac{|p - p_c|}{p_c}.$$

Table 4.6: Simulation Design

| | **Factor** | | | | |
|---|---|---|---|---|---|
| Run | 1 | 2 | 3 | 4 | 5 |
| 1 | L | L | L | L | L |
| 2 | H | H | L | L | L |
| 3 | L | L | H | H | L |
| 4 | H | H | H | H | L |
| 5 | H | L | L | L | H |
| 6 | L | H | L | L | H |
| 7 | H | L | H | H | H |
| 8 | L | H | H | H | H |
| 9 | H | L | H | L | L |
| 10 | L | H | H | L | L |
| 11 | H | L | L | H | L |
| 12 | L | H | L | H | L |
| 13 | L | L | H | L | H |
| 14 | H | H | H | L | H |
| 15 | L | L | L | H | H |
| 16 | H | H | L | H | H |

Table 4.7: Log Odds Ratio Simulation Results - Bias and Variance

| **Run** | **Mean Log Odds Ratio** | | | **Variance** | | |
|---|---|---|---|---|---|---|
| | Ex | SP | Uncond. | Ex | SP | Uncond. |
| 1 | 0.4253 | 0.4238 | 0.4726 | 0.2631 | 0.2603 | 0.3245 |
| 2 | 1.7007 | 1.6926 | 1.9118 | 0.2857 | 0.2750 | 0.3740 |
| 3 | 0.4173 | 0.4165 | 0.4448 | 0.0340 | 0.0339 | 0.0386 |
| 4 | 1.6191 | 1.6206 | 1.7589 | 0.0255 | 0.0256 | 0.0312 |
| 5 | 0.4133 | 0.4130 | 0.4431 | 0.0649 | 0.0648 | 0.0747 |
| 6 | 1.6454 | 1.6471 | 1.7905 | 0.0967 | 0.0968 | 0.1194 |
| 7 | 0.4175 | 0.4175 | 0.4398 | 0.0068 | 0.0068 | 0.0075 |
| 8 | 1.6395 | 1.6391 | 1.7345 | 0.0116 | 0.0115 | 0.0132 |
| 9 | 0.3930 | 0.3930 | 0.4378 | 0.0279 | 0.0278 | 0.0347 |
| 10 | 1.6551 | 1.6515 | 1.8607 | 0.0516 | 0.0507 | 0.0679 |
| 11 | 0.4423 | 0.4424 | 0.4756 | 0.0634 | 0.0634 | 0.0733 |
| 12 | 1.6725 | 1.6681 | 1.7966 | 0.1408 | 0.1384 | 0.1639 |
| 13 | 0.4012 | 0.4016 | 0.4324 | 0.0287 | 0.0288 | 0.0333 |
| 14 | 1.6610 | 1.6559 | 1.7774 | 0.0410 | 0.0403 | 0.0470 |
| 15 | 0.3962 | 0.3956 | 0.4405 | 0.0573 | 0.0570 | 0.0708 |
| 16 | 1.6244 | 1.6232 | 1.7172 | 0.0577 | 0.0574 | 0.0654 |

Table 4.8: Log Odds Ratio Simulation Results - 95% Coverage Probability

| **Run** | **95% Coverage Probability** | | | | |
|---|---|---|---|---|---|
| | Cond | SP | Uncond. | CC-SP | SQ-SP |
| 1 | 0.9829 | 0.9840 | 0.9458 | 0.9809 | 0.9840 |
| 2 | 0.9799 | 0.9778 | 0.9305 | 0.9789 | 0.9799 |
| 3 | 0.9750 | 0.9750 | 0.9440 | 0.9720 | 0.9750 |
| 4 | 0.9730 | 0.9680 | 0.9310 | 0.9680 | 0.9840 |
| 5 | 0.9800 | 0.9790 | 0.9350 | 0.9800 | 0.9780 |
| 6 | 0.9710 | 0.9690 | 0.9300 | 0.9700 | 0.9680 |
| 7 | 0.9690 | 0.9610 | 0.9510 | 0.9690 | 0.9640 |
| 8 | 0.9630 | 0.9560 | 0.9190 | 0.9610 | 0.9620 |
| 9 | 0.9650 | 0.9650 | 0.9310 | 0.9630 | 0.9650 |
| 10 | 0.9580 | 0.9550 | 0.8970 | 0.9580 | 0.9550 |
| 11 | 0.9740 | 0.9740 | 0.9350 | 0.9730 | 0.9700 |
| 12 | 0.9789 | 0.9789 | 0.9277 | 0.9779 | 0.9739 |
| 13 | 0.9640 | 0.9630 | 0.9420 | 0.9590 | 0.9630 |
| 14 | 0.9690 | 0.9650 | 0.9280 | 0.9680 | 0.9620 |
| 15 | 0.9740 | 0.9740 | 0.9280 | 0.9690 | 0.9710 |
| 16 | 0.9730 | 0.9730 | 0.9310 | 0.9740 | 0.9710 |

Table 4.9: Log Odds Ratio Simulation Results – Misclassification

| **Run** | **Number of Errors** | | | |
|---|---|---|---|---|
| | SP | Uncond. | CC-SP | SQ-SP |
| 1 | 1 | 37 | 2 | 1 |
| 2 | 2 | 49 | 1 | 0 |
| 3 | 0 | 31 | 3 | 2 |
| 4 | 5 | 46 | 5 | 4 |
| 5 | 1 | 45 | 0 | 2 |
| 6 | 2 | 41 | 1 | 3 |
| 7 | 8 | 18 | 0 | 5 |
| 8 | 7 | 48 | 2 | 1 |
| 9 | 2 | 34 | 2 | 0 |
| 10 | 7 | 65 | 4 | 3 |
| 11 | 0 | 39 | 1 | 4 |
| 12 | 2 | 51 | 1 | 5 |
| 13 | 3 | 22 | 5 | 1 |
| 14 | 6 | 43 | 3 | 9 |
| 15 | 0 | 46 | 5 | 3 |
| 16 | 2 | 42 | 1 | 2 |

Table 4.10: Log Odds Ratio Simulation Results – Mean Relative Errors of 95% Confidence Bounds

| **Run** | **Mean Relative Error** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | SP | | Uncond. | | CC-SP | | SQ-SP | |
| | Low | Hi | Low | Hi | Low | Hi | Low | Hi |
| 1 | 0.0133 | 0.0164 | 0.2258 | 0.1240 | 0.0049 | 0.0093 | 0.0030 | 0.0106 |
| 2 | 0.0061 | 0.0262 | 0.3965 | 0.0918 | 0.0072 | 0.0208 | 0.0023 | 0.0387 |
| 3 | 0.0061 | 0.0262 | 0.0872 | 0.0356 | 0.0011 | 0.0020 | 0.0003 | 0.0027 |
| 4 | 0.0010 | 0.0120 | 0.1940 | 0.1143 | 0.0038 | 0.0105 | 0.0013 | 0.0024 |
| 5 | 0.0043 | 0.0064 | 0.1176 | 0.0540 | 0.0013 | 0.0030 | 0.0004 | 0.0016 |
| 6 | 0.0023 | 0.0175 | 0.2604 | 0.0686 | 0.0038 | 0.0119 | 0.0013 | 0.0049 |
| 7 | 0.0009 | 0.0027 | 0.0481 | 0.0127 | 0.0003 | 0.0016 | 0.0001 | 0.0014 |
| 8 | 0.0005 | 0.0085 | 0.1343 | 0.0692 | 0.0019 | 0.0048 | 0.0002 | 0.0017 |
| 9 | 0.0051 | 0.0101 | 0.0851 | 0.0405 | 0.0021 | 0.0060 | 0.0003 | 0.0006 |
| 10 | 0.0016 | 0.0194 | 0.2811 | 0.1825 | 0.0065 | 0.0137 | 0.0012 | 0.0118 |
| 11 | 0.0042 | 0.0073 | 0.1216 | 0.0487 | 0.0013 | 0.0036 | 0.0007 | 0.0007 |
| 12 | 0.0024 | 0.0082 | 0.2626 | 0.0440 | 0.0036 | 0.0071 | 0.0018 | 0.0146 |
| 13 | 0.0033 | 0.0062 | 0.0864 | 0.0314 | 0.0011 | 0.0031 | 0.0006 | 0.0007 |
| 14 | 0.0004 | 0.0041 | 0.1927 | 0.0503 | 0.0027 | 0.0025 | 0.0023 | 0.0113 |
| 15 | 0.0069 | 0.0105 | 0.1105 | 0.0417 | 0.0026 | 0.0052 | 0.0008 | 0.0036 |
| 16 | 0.0009 | 0.0054 | 0.1861 | 0.0225 | 0.0021 | 0.0035 | 0.0005 | 0.0042 |

Table 4.11: Log Odds Ratio Simulation Results – Maximum Relative Errors of 95% Confidence Bounds

| **Run** | **Maximum Relative Error** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | SP | | Uncond. | | CC-SP | | SQ-SP | |
| | Low | Hi | Low | Hi | Low | Hi | Low | Hi |
| 1 | 0.0696 | 0.0851 | 1.3095 | 0.5794 | 0.0741 | 0.0746 | 0.0741 | 0.1734 |
| 2 | 0.0160 | 0.1806 | 1.3633 | 0.6569 | 0.0384 | 0.3387 | 0.0226 | 0.3930 |
| 3 | 0.0160 | 0.1806 | 0.2571 | 0.1152 | 0.0058 | 0.0062 | 0.0053 | 0.0206 |
| 4 | 0.0059 | 0.0300 | 0.3940 | 0.3211 | 0.0108 | 0.4908 | 0.0058 | 0.0336 |
| 5 | 0.0102 | 0.0155 | 0.3129 | 0.1787 | 0.0063 | 0.0106 | 0.0069 | 0.0213 |
| 6 | 0.0127 | 0.0833 | 1.0050 | 0.4257 | 0.0205 | 0.0927 | 0.0406 | 0.2053 |
| 7 | 0.0023 | 0.4363 | 0.1039 | 0.0532 | 0.0014 | 0.4026 | 0.0003 | 0.4646 |
| 8 | 0.0020 | 0.4615 | 0.2745 | 0.1603 | 0.0041 | 0.4431 | 0.0030 | 0.4820 |
| 9 | 0.0132 | 0.0230 | 0.3663 | 0.2469 | 0.0093 | 0.0173 | 0.0033 | 0.0175 |
| 10 | 0.0167 | 0.1397 | 0.9819 | 0.7606 | 0.0244 | 0.0846 | 0.0542 | 0.4276 |
| 11 | 0.0087 | 0.0185 | 0.3333 | 0.1738 | 0.0058 | 0.0145 | 0.0103 | 0.0160 |
| 12 | 0.0131 | 0.0861 | 0.8223 | 0.5124 | 0.0146 | 0.0735 | 0.0225 | 0.1473 |
| 13 | 0.0060 | 0.0120 | 0.2143 | 0.1398 | 0.0139 | 0.0083 | 0.0080 | 0.0045 |
| 14 | 0.0047 | 0.0354 | 0.5005 | 0.2220 | 0.0067 | 0.0262 | 0.0183 | 0.1033 |
| 15 | 0.0156 | 0.0220 | 0.4478 | 0.2464 | 0.0096 | 0.0161 | 0.0431 | 0.0798 |
| 16 | 0.0029 | 0.0318 | 0.4149 | 0.1739 | 0.0060 | 0.0311 | 0.0082 | 0.0421 |

We do not present here the results for 99% confidence intervals as the conclusions are essentially the same.

Examination of these results does show that the unconditional methods perform very poorly in general. Overall, the saddlepoint methods perform very well in approximating exact conditional inference. The mean of the estimated log odds ratio is identical to that using exact conditional inference to at least one and often two or more decimal places, especially when the true log odds ratio is small. The variance of the estimated log-odds ratio is close to that from exact inference and usually lower than that from unconditional inference.

The coverage probability of the confidence intervals is essentially exactly the same as that from the exact conditional inference. Both methods are equivalently conservative. Also, the three different saddlepoint methods rarely gave different conclusions than did exact inference.

A more detailed investigation shows that in general, the sequential saddlepoint and the continuity corrected saddlepoint perform slightly better than the ordinary saddlepoint approximation. The coverage probability for each of the three methods is always quite close to that of conditional inference, but the relative error of the confidence bounds, particularly for the high end of the interval, is in general lower for the two more complicated methods. The continuity corrected saddlepoint method tended to have higher relative errors on the low end of the interval, while the relative error from the sequential saddlepoint method was nearly uniformly lower than that from the ordinary saddlepoint method. Both these methods are more complicated to implement than the ordinary approximation: the sequential saddlepoint because score and information statistics are required from the modified profile likelihood instead of the profile likelihood; and the continuity corrected saddlepoint because it requires evaluation of scores and likelihoods at perturbed values of the sufficient statistic. It is likely that in practice the additional complications of these two methods are not warranted given the only slight increase in accuracy.

An analysis of the simulation results using ANOVA methods for the $2^{5-1}$ design provides some insight into the performance of the saddlepoint methods. The difference between the approximate and exact estimators, and the relative error of the confidence limits for the saddlepoint approximation appear to be related to the magnitude of the odds ratio, but this result is not statistically significant.

It should be noted that the relative error of the approximations at the upper end of the confidence interval is fairly high on average. This is due to the fact that occasionally the upper confidence limit is extremely high (sometimes over 10000). This is not really a limitation of the method, because in practice, researchers are likely to disregard limits this high as having little practical value. The lower end of the confidence interval, which is typically of most interest because researchers are interested in whether or not the odds ratio differs from zero, shows different results. The three different saddlepoint methods rarely differ from the conditional method by more than 1 or 2 percent. This indicates that for practical purposes any of the three methods would work well as approximate conditional inference.

## Chapter 5

# APPLICATION TO THE TEST FOR TREND IN A SEQUENCE OF BINOMIAL VARIABLES

### *5.1 Introduction*

The test for trend in a sequence of binomial proportions is a statistical problem of particular use to epidemiologists. It is common in dose-response models and other problems where the response is binomial and the probability of success is expected to increase or decrease with some other variable, but due to the nature of the covariate (the covariate arises from grouping into arbitrary strata, or the covariate is ordinal, for example) a specific relationship is not assumed other than an increasing trend.

**Example** Graubard and Korn (1987) used as an example data on maternal drinking and congenital malformations. Of interest in this study was whether the probability of a malformation increased with average number of alcoholic drinks per day. Table 5.1 gives the results of a study of 32574 women, with the average alcohol consumption and the presence/absence of malformation. To determine whether there is an increasing trend in the presence of malformation given alcohol consumption, we can perform logistic regression on an appropriately chosen trend variable. Here a reasonable choice is to use a variable giving the midpoint of each category. The usual analysis for these data is unconditional maximum likelihood analysis. However, when data are skewed as these are, "exact" conditional maximum likelihood analysis is generally done, because the asymptotic assumptions for unconditional analysis are not satisfied. For these data, the conditional analysis is difficult to do because it

Table 5.1: Data on Alcohol Consumption and Malformation

| **Malfor-<br>mation** | **Alcohol Consumption** (avg. drinks per day) 0 | <1 | 1-2 | 3-5 | $\geq 6$ |
|---|---|---|---|---|---|
| Absent | 17,066 | 14,464 | 788 | 126 | 37 |
| Present | 48 | 38 | 5 | 1 | 1 |

requires enumerating a large sample space and calculating large numbers of binomial coefficients. An approximation to the conditional analysis would be useful here, to provide accurate yet quick inference.

## 5.2 *The Model*

The model for this problem is a simple logistic regression of the binomial variates on the trend covariate $W$. Let $Y_i, i = 1, \ldots, N$ be binomial with parameters $n_i$ and $p_i$ where $\text{logit}(p_i) = \mu + w_i\Delta$ and $w_1 < w_2 < \ldots < w_N$. The overall likelihood for the data is just the binomial likelihood,

$$L(\Delta, \mu|y) \propto \prod_i \frac{e^{(\mu+w_i\Delta)y_i}}{(1+e^{\mu+w_i\Delta})^{n_i}}, \tag{5.1}$$

$T_1 = \sum Y_i w_i$ is the minimal sufficient statistic for $\Delta$, and $T_2 = \sum Y_i$ is the minimal sufficient statistic for $\mu$. Inference on $\Delta$ is based on the distribution of $T_1$.

### 5.2.1 *Unconditional Inference*

Unconditional inference on $\Delta$ is based again on the profile likelihood which is obtained by maximizing expression (5.1) over all possible values of $\mu$ for a fixed $\Delta$. This gives an unconditional likelihood as follows:

$$L_P(\Delta|\mu = \hat{\mu}) \propto \frac{e^{t_1\Delta}}{\prod_i(1+e^{\hat{\mu}+w_i\Delta})^{n_i}}. \tag{5.2}$$

This likelihood is easily maximized using standard GLM procedures. The likelihood ratio statistic

$$\begin{aligned} 2\log\Lambda &= 2\left\{l_P(\hat{\Delta}) - l_P(\Delta)\right\} \\ &= t_1(\hat{\Delta} - \Delta) - \sum_i n_i \log\left(\frac{1 + e^{\hat{\mu}+w_i\hat{\Delta}}}{1 + e^{\hat{\mu}+w_i\Delta}}\right), \end{aligned} \tag{5.3}$$

and the Wald statistic

$$\frac{(\hat{\Delta} - \Delta)^2}{I_P(\hat{\Delta})}, \tag{5.4}$$

where

$$I_P(\Delta) = \sum_i w_i^2 n_i \frac{e^{\hat{\mu}(\Delta)+w_i\Delta}}{(1 + e^{\hat{\mu}(\Delta)+w_i\Delta})^2}, \tag{5.5}$$

both have approximately the $\chi^2$ distribution with one degree of freedom provided the conditions in section 5.3 hold.

### *5.2.2 Conditional Inference*

The exact conditional likelihood is the likelihood for $\Delta$ conditional on $T_2$, the sufficient statistic for $\mu$. The probability mass function is

$$P_C(T_1 = t_1 | T_2 = t_2) = \frac{C(t_1, t_2)e^{t_1\Delta}}{\sum_u C(u, t_2)e^{u\Delta}}, \tag{5.6}$$

where

$$C(t_1, t_2) = \sum_{y:T_1=t_1, T_2=t_2} \prod_{i=1}^{N} \binom{n_i}{y_i}. \tag{5.7}$$

In general, inference is done using enumeration of the probabilities from this likelihood. This involves calculating large numbers of binomial coefficients $\binom{n_i}{y_i}$ and enumerating a large sample space, which can be very cumbersome computationally.

An exact tail probability (for use in a test of $\Delta = \Delta_0$) is given by summation of the appropriate probabilities,

$$P(T_1 > t_1 | T_2 = t_2, \Delta) = \sum_{s>t_1} \frac{C(s, t_2)e^{t_1\Delta}}{\sum_u C(u, t_2)e^{u\Delta}}. \tag{5.8}$$

This is again a quantity which involves many binomial coefficients and is cumbersome to calculate.

### *5.2.3 Double Saddlepoint Inference*

The double saddlepoint approximate likelihood for this model is again a modification of the unconditional result.

$$L_{MP}(\Delta; T_1|T_2 = t_2) = \frac{e^{t_2\hat{\mu}+t_1\Delta}}{\prod_i(1+e^{\hat{\mu}+w_i\Delta})^{n_i}}\sqrt{I_\mu(\Delta, \hat{\mu}(\Delta))}, \tag{5.9}$$

where

$$I_\mu(\Delta, \mu) = \sum_i n_i \frac{e^{\mu+w_i\Delta}}{(1+e^{\mu+w_i\Delta})^2} \tag{5.10}$$

and $\hat{\mu} = \hat{\mu}(\Delta)$ solves the score equation

$$\dot{l}_\mu = t_2 - \sum_i n_i \frac{e^{\mu+w_i\Delta}}{1+e^{\mu+w_i\Delta}} = 0.$$

To maximize the approximate likelihood (5.9) it is again necessary to derive the score and information statistics. This is a second special case of Levin (1990). The score for $\Delta$ is given by

$$\dot{l}_{SP}(\Delta) = t_1 - \sum_i \frac{n_i w_i e^{y_i(\hat{\mu}+w_i\Delta)}}{1+e^{\hat{\mu}+w_i\Delta})^{n_i}} + \frac{1}{2}\frac{\partial}{\partial\Delta}\log\left\{\sum_i n_i \frac{e^{\hat{\mu}+w_i\Delta}}{(1+e^{\hat{\mu}+w_i\Delta})^2}\right\}, \tag{5.11}$$

and the information by

$$I_{SP}(\Delta) = \sum_i n_i w_i^2 \frac{e^{\hat{\mu}+w_i\Delta}}{(1+e^{\hat{\mu}+w_i\Delta})^2} - \frac{1}{2}\frac{\partial^2}{\partial\Delta^2}\log\left\{\sum_i n_i \frac{e^{\hat{\mu}+w_i\Delta}}{(1+e^{\hat{\mu}+w_i\Delta})^2}\right\}. \tag{5.12}$$

The approximate tail probability from the Davison derivation is again given by (3.13) and from the Pierce and Peters derivation is given by (3.14) where $w$ is the signed square root of the unconditional likelihood ratio statistic (5.3), $z$ is the square root of the Wald statistic (5.4) and $\rho$ is the ratio of the determinant of the information matrix for $\mu$ from the full model to the determinant of the information matrix for $\mu$ from the reduced model. This is a discrete data problem, so we use the correction for discreteness given in chapter 3 and replace $\hat{\Delta} - \Delta$ in the Wald statistic by $1 - \exp(-(\hat{\Delta} - \Delta))$. The continuity corrected saddlepoint replaces $\hat{\Delta} - \Delta$ with $\exp(-(\hat{\Delta} - \Delta)/2) - \exp((\hat{\Delta} - \Delta)/2)$ if the continuity correction is used.

*Sequential Saddlepoint Inference*

The sequential saddlepoint method is easily implemented here. The tail probabilities are given by (3.25) and (3.26) where now, $w_{MP}$ and $z_{MP}$ are the likelihood ratio and Wald statistics from the modified profile likelihood (5.9). For the purposes of the simulation study we use formula (3.26).

### 5.2.4 Direct Saddlepoint Inference

Direct saddlepoint inference again uses a single saddlepoint approximation to the exact conditional density (5.6). This is an exponential family density, which we can rewrite in the form

$$P_C(T_1 = t_1 | T_2 = t_2) = \exp\left[t_1\Delta - \log\left\{\sum_u C(u, t_2)e^{u\Delta}\right\} + \log\left\{C(t_1, t_2)\right\}\right]. \quad (5.13)$$

Hence, the cumulant generating function for the exact conditional density is given by

$$K(\lambda) = \log\left\{\sum_u C(u, t_2)e^{u(\lambda)_i}\right\}, \quad (5.14)$$

and (using the methods of chapter 3) the saddlepoint $\hat{\lambda} = \hat{\Delta}_C - \Delta$ where $\hat{\Delta}_C$ is the MLE for $\Delta$ under the conditional likelihood. The direct saddlepoint approximation to the density is thus

$$P_{DSP}(T_1 = t_1 | T_2 = t_2) \doteq \frac{1}{\sqrt{2\pi}} \frac{L_C(\Delta)}{L_C(\hat{\Delta}_C)}, \quad (5.15)$$

where

$$L_C(\Delta) \propto \exp\left[t_1\Delta - \log\left\{\sum_u C(u, t_2)e^{u\Delta}\right\}\right]. \quad (5.16)$$

For point estimation and confidence intervals, this likelihood presents the same computational problem that the exact conditional model does. However, for testing $\Delta = 0$ (which is of primary interest for this model), there is less difficulty, because the exact likelihood only needs to be evaluated at the MLE and at $\Delta = 0$.

Table 5.2: Malformation Data - Analysis

| **Method** | $\hat{\Delta}$ | **P-value** |
|---:|:---:|:---:|
| Conditional | 0.3175 | 0.017 |
| Unconditional | 0.3175 | 0.005 |
| Saddlepoint | 0.3165 | 0.019 |

**Example (contd.)** Table 5.2 presents an analysis of the data on congenital malformations given at the beginning of the chapter. All three point estimators of $\Delta$ are essentially identical, while the p-value for the one-sided test of $\Delta > 0$ using unconditional inference is substantially lower than that using conditional inference and saddlepoint inference, and the saddlepoint inference value is reasonably close to the value from conditional inference. Of note is the fact that unconditional inference gives a different conclusion from the other two methods at the 0.001 significance level.

## *5.3 Asymptotic Distribution Theory of the Estimators*

The unconditional maximum likelihood estimator of $\Delta$ is consistent and asymptotically normal provided that the $n_i$ all increase to infinity with $M = \sum_i n_i$ (Fahrmeir and Tutz, 1994, Ch. 2). However, if one of the $p_i = e^{\mu+\Delta w_i}/(1+e^{\mu+\Delta w_i})$ is very small or one of the $n_i$ increases at a slow rate the convergence can be very slow. Under these conditions, the exact conditional maximum likelihood estimator is asymptotically equivalent to the unconditional MLE. The maximum modified profile likelihood (or saddlepoint) estimator solves the equation

$$\dot{l}(\Delta) = t_1 - \sum_i \frac{n_i w_i e^{y_i(\hat{\mu}+w_i\Delta)}}{1 + e^{\hat{\mu}+w_i\Delta})^{n_i}} + \frac{1}{2}\frac{\partial}{\partial\Delta}\log\left\{\sum_i n_i \frac{e^{\hat{\mu}+w_i\Delta}}{(1+e^{\hat{\mu}+w_i\Delta})^2}\right\} = 0. \qquad (5.17)$$

Now,

$$\frac{\partial}{\partial\Delta}\log\left\{\sum_i n_i \frac{e^{\hat{\mu}+w_i\Delta}}{(1+e^{\hat{\mu}+w_i\Delta})^2}\right\} = \frac{\sum_i n_i \frac{e^{\hat{\mu}+w_i\Delta}\{w_i+(\partial\hat{\mu}/\partial\Delta)\}}{1+e^{\hat{\mu}+w_i\Delta}}\left\{\frac{1}{(1+e^{\hat{\mu}+w_i\Delta})^2}-\frac{2}{(1+e^{\hat{\mu}+w_i\Delta})^3}\right\}}{\sum_i n_i \frac{e^{\hat{\mu}+w_i\Delta}}{(1+e^{\hat{\mu}+w_i\Delta})^2}}, \tag{5.18}$$

and hence, provided that the $n_i$ all increase to infinity with $M = \sum_i n_i$,

$$\frac{1}{M}\frac{\partial}{\partial\Delta}\log\left\{\sum_i n_i \frac{e^{\hat{\mu}+w_i\Delta}}{(1+e^{\hat{\mu}+w_i\Delta})^2}\right\} = o_p(1). \tag{5.19}$$

An argument identical to that of chapter 4 can be used to demonstrate that $\hat{\Delta}_{SP}$ is consistent and asymptotically equivalent to $\hat{\Delta}_P$.

## *5.4 Small Sample Performance of the Methods*

A variety of methods exist for comparing the performance of the saddlepoint and unconditional inference methods to exact conditional inference. One is to enumerate the entire exact conditional distribution and compare the value of the tail probability from the exact distribution to the value of the tail probability from the approximate distribution. This method was used (Bedrick and Hill, 1993) to demonstrate that under many circumstances, the tail probability formula in equation (3.13) is very accurate in approximating the entire conditional distribution.

An alternate approach, and the one we will use in this chapter, is to examine the methods via simulation. This has significant advantages over the enumeration method. The primary advantage is that while the enumeration method examines all possible values of the conditional distribution, a simulation study with well-chosen parameter values will concentrate on areas of the parameter space that are likely to occur in practice. The approximate methods are most likely to fail for extreme values of the conditional tail probability (0.001 or lower), but if these values are unlikely to ever be observed, using these values does not give an accurate impression of the methods' performance.

Table 5.3: Binomial Denominators For Trend Test Simulations

| **Dataset** | **Denominators** |
|---|---|
| 1 | 10,3,7,5 |
| 2 | 36,3,3,2 |
| 3 | 25,7,16 |
| 4 | 95,1,1,2,2,4 |

Hence, to expand on the work of Bedrick and Hill, we undertake a simulation study of both saddlepoint tail probability approximations and the unconditional approximation.

### *5.4.1 Design of the Simulation Studies*

The key to the performance of the methods in the trend test is the distribution of the binomial denominators. We have chosen four representative sets of denominators from a set of 20 used by Bedrick and Hill (1993) taken from an analysis of the trend test due to Graubard and Korn (1987). The binomial denominators from the four chosen datasets are given in table 5.3. These sets are chosen to represent situations where the data are slightly unbalanced, moderately unbalanced and extremely unbalanced. The third and fourth sets of denominators typify a situation where Bedrick and Hill (1992) found poor performance by the saddlepoint approximation, where the binomial denominators are high at the ends of the range of the data but low in the middle.

For each dataset, we examine three levels of baseline probability and trend parameter. The baseline probabilities used in the simulation were 0.10, 0.30, 0.50, and the trend parameters are 0, 0.5 and 1.0, which give a relatively wide range of the effects of the parameters on performance of the methods. There were 1000 datasets generated for each simulation. The trend scores chosen for each dataset are simple ordinal scores.

The data were generated unconditionally (this is how they occur in practice) from the binomial model,

$$P(Y=1|W=w) = \frac{p_0 e^{w\Delta}}{(1+p_0-p_0 e^{w\Delta})}, \tag{5.20}$$

where $Y$ is a variable indicating disease status, $p_0$ is the baseline probability and $\Delta$ is the odds ratio. Analyses of the data were done using exact conditional one-sided tail probabilities, unconditional tail probability approximations using the log-likelihood ratio statistic and formula 3.14 for saddlepoint tail probabilities, again both with and without continuity corrections and using the double saddlepoint and sequential saddlepoint approximations.

The result of primary interest is the p-value for testing $\Delta = 0$ vs. $\Delta > 0$. To determine the performance of the methods, we report both the mean difference between the approximate and exact p-values as well as the estimated power of the test for detecting a trend parameter of the given size. Again, we expect the exact conditional tests to be conservative.

Following Bedrick and Hill, certain results were excluded from the analysis to avoid roundoff errors. When the exact tail probability was less than 0.001, the result was discarded. When the estimated $\hat{\Delta}$ was zero, Bedrick and Hill's method of avoiding the singularity was used. This method is as follows: when the non-continuity corrected value was undefined because $\hat{\Delta}$ was zero the tail probability was replaced with the corrected one. When the continuity corrected tail probability was undefined it was replaced with the uncorrected one.

### *5.4.2 Results of the Simulation Studies*

The simulation results are summarized in tables 5.4 through 5.9. Tables 5.4 and 5.5 give the mean of the estimator of the trend parameter and of the estimator of the variance of the trend parameter over the set of simulations for the unconditional and

Table 5.4: Trend Simulations – Mean Estimates of $\Delta$

| | **Estimated $\Delta$** | | **Estimated Variance** | |
|---|---|---|---|---|
| **Dataset** | **SP** | **Uncond.** | **SP** | **Uncond.** |
| 1 | 0.0318 | -0.0109 | 0.3740 | 0.3571 |
| | 0.5583 | 0.5348 | 0.3191 | 0.2919 |
| | 1.1162 | 1.1413 | 0.3184 | 0.3262 |
| | -0.0334 | -0.0422 | 0.1717 | 0.1666 |
| | 0.5456 | 0.5496 | 0.1564 | 0.1565 |
| | 1.1073 | 1.1156 | 0.2670 | 0.2712 |
| | -0.0157 | -0.0134 | 0.1298 | 0.1281 |
| | 0.5462 | 0.5530 | 0.1821 | 0.1839 |
| | 1.0430 | 1.0489 | 0.3501 | 0.3534 |
| 2 | 0.2336 | 0.2217 | 0.3434 | 0.3366 |
| | 0.5605 | 0.5533 | 0.2610 | 0.2582 |
| | 1.0933 | 1.0930 | 0.2683 | 0.2691 |
| | -0.0104 | -0.0133 | 0.2307 | 0.2278 |
| | 0.5347 | 0.5339 | 0.2084 | 0.2082 |
| | 0.9337 | 0.9340 | 0.4503 | 0.4506 |
| | -0.0104 | -0.0100 | 0.1948 | 0.1933 |
| | 0.5234 | 0.5238 | 0.2765 | 0.2766 |
| | 0.9337 | 0.9340 | 0.4503 | 0.4506 |

Table 5.5: Trend Simulations – Mean Estimates of $\Delta$

| | **Estimated $\Delta$** | | **Estimated Variance** | |
|---|---|---|---|---|
| **Dataset** | **SP** | **Uncond.** | **SP** | **Uncond.** |
| 3 | 0.0228 | -0.0153 | 0.3554 | 0.3471 |
| | 0.5076 | 0.4774 | 0.2510 | 0.2456 |
| | 1.0644 | 1.0571 | 0.2131 | 0.2041 |
| | -0.0153 | -0.0221 | 0.1345 | 0.1327 |
| | 0.5273 | 0.5269 | 0.1180 | 0.1178 |
| | 1.0725 | 1.0790 | 0.1500 | 0.1520 |
| | -0.0004 | 0.0005 | 0.1075 | 0.1069 |
| | 0.5389 | 0.5432 | 0.1310 | 0.1319 |
| | 1.0994 | 1.1046 | 0.2362 | 0.2384 |
| 4 | 0.1294 | 0.1277 | 0.0603 | 0.0600 |
| | 0.5000 | 0.4996 | 0.0391 | 0.0391 |
| | 0.9510 | 0.9513 | 0.0779 | 0.0780 |
| | 0.0077 | 0.0073 | 0.0401 | 0.0400 |
| | 0.4956 | 0.4957 | 0.0508 | 0.0508 |
| | 0.7117 | 0.7117 | 0.0853 | 0.0854 |
| | -0.0080 | -0.0080 | 0.0335 | 0.0335 |
| | 0.4115 | 0.4116 | 0.0639 | 0.0640 |
| | 0.5304 | 0.5304 | 0.0849 | 0.0849 |

Table 5.6: Trend Simulations - Results for Dataset 1

| | | **Mean Rel. Error** | | | |
|---|---|---|---|---|---|
| **Baseline** | **Trend Par.** | Uncond. | SP | SQ-SP | CC-SP |
| 0.1 | 0.0 | 0.3857 | 0.0330 | 0.0567 | 0.0419 |
| 0.1 | 0.5 | 0.4169 | 0.0258 | 0.0390 | 0.0248 |
| 0.1 | 1.0 | 0.5010 | 0.0266 | 0.0292 | 0.0089 |
| 0.3 | 0.0 | 0.2803 | 0.0111 | 0.0160 | 0.0057 |
| 0.3 | 0.5 | 0.3654 | 0.0129 | 0.0122 | 0.0030 |
| 0.3 | 1.0 | 0.4989 | 0.0207 | 0.0179 | 0.0068 |
| 0.5 | 0.0 | 0.2578 | 0.0066 | 0.0108 | 0.0018 |
| 0.5 | 0.5 | 0.3831 | 0.0150 | 0.0175 | 0.0053 |
| 0.5 | 1.0 | 0.5156 | 0.0382 | 0.0361 | 0.0131 |

| | | **Max. Rel. Error** | | | |
|---|---|---|---|---|---|
| **Baseline** | **Trend Par.** | Uncond. | SP | SQ-SP | CC-SP |
| 0.1 | 0.0 | 0.7685 | 0.1300 | 0.1201 | 0.1434 |
| 0.1 | 0.5 | 0.8158 | 0.1300 | 0.1778 | 0.1434 |
| 0.1 | 1.0 | 0.8158 | 0.0846 | 0.1778 | 0.1434 |
| 0.3 | 0.0 | 0.7685 | 0.1300 | 0.1257 | 0.1434 |
| 0.3 | 0.5 | 0.7307 | 0.0785 | 0.1258 | 0.0528 |
| 0.3 | 1.0 | 0.7769 | 0.1115 | 0.1257 | 0.0372 |
| 0.5 | 0.0 | 0.6834 | 0.0305 | 0.0371 | 0.0281 |
| 0.5 | 0.5 | 0.7769 | 0.1646 | 0.1637 | 0.0621 |
| 0.5 | 1.0 | 0.7769 | 0.1886 | 0.1777 | 0.1082 |

Table 5.7: Trend Simulations – Results for Dataset 2

| | | **Mean Rel. Error** | | | |
|---|---|---|---|---|---|
| **Baseline** | **Trend Par.** | Uncond. | SP | SQ-SP | CC-SP |
| 0.1 | 0.0 | 0.2644 | 0.0453 | 0.0438 | 0.0508 |
| 0.1 | 0.5 | 0.3202 | 0.0505 | 0.0480 | 0.0527 |
| 0.1 | 1.0 | 0.4394 | 0.0461 | 0.0442 | 0.0477 |
| 0.3 | 0.0 | 0.2467 | 0.0129 | 0.0104 | 0.0106 |
| 0.3 | 0.5 | 0.3546 | 0.0151 | 0.0122 | 0.0110 |
| 0.3 | 1.0 | 0.5101 | 0.0280 | 0.0254 | 0.0299 |
| 0.5 | 0.0 | 0.3031 | 0.0115 | 0.0077 | 0.0097 |
| 0.5 | 0.5 | 0.4038 | 0.0186 | 0.0149 | 0.0178 |
| 0.5 | 1.0 | 0.5101 | 0.0280 | 0.0254 | 0.0299 |

| | | **Max. Rel. Error** | | | |
|---|---|---|---|---|---|
| **Baseline** | **Trend Par.** | Uncond. | SP | SQ-SP | CC-SP |
| 0.1 | 0.0 | 0.7673 | 0.1767 | 0.1746 | 0.1862 |
| 0.1 | 0.5 | 0.7673 | 0.1767 | 0.1746 | 0.1862 |
| 0.1 | 1.0 | 0.7673 | 0.1767 | 0.1746 | 0.1862 |
| 0.3 | 0.0 | 0.6600 | 0.1207 | 0.1212 | 0.1074 |
| 0.3 | 0.5 | 0.7837 | 0.0669 | 0.1074 | 0.1074 |
| 0.3 | 1.0 | 0.7837 | 0.0560 | 0.0523 | 0.0946 |
| 0.5 | 0.0 | 0.7571 | 0.0560 | 0.0523 | 0.0906 |
| 0.5 | 0.5 | 0.7571 | 0.0560 | 0.0523 | 0.0906 |
| 0.5 | 1.0 | 0.7837 | 0.0560 | 0.0523 | 0.0946 |

Table 5.8: Trend Simulations – Results for Dataset 3

| | | **Mean Rel. Error** | | | |
|---|---|---|---|---|---|
| **Baseline** | **Trend Par.** | Uncond. | SP | SQ-SP | CC-SP |
| 0.1 | 0.0 | 0.3080 | 0.0379 | 0.0493 | 0.0418 |
| 0.1 | 0.5 | 0.3526 | 0.0225 | 0.0359 | 0.0199 |
| 0.1 | 1.0 | 0.4259 | 0.0127 | 0.0146 | 0.0052 |
| 0.3 | 0.0 | 0.2367 | 0.0043 | 0.0102 | 0.0015 |
| 0.3 | 0.5 | 0.3075 | 0.0050 | 0.0044 | 0.0007 |
| 0.3 | 1.0 | 0.4105 | 0.0072 | 0.0039 | 0.0011 |
| 0.5 | 0.0 | 0.2249 | 0.0034 | 0.0073 | 0.0011 |
| 0.5 | 0.5 | 0.3249 | 0.0061 | 0.0090 | 0.0010 |
| 0.5 | 1.0 | 0.4523 | 0.0138 | 0.0118 | 0.0034 |

| | | **Max. Rel. Error** | | | |
|---|---|---|---|---|---|
| **Baseline** | **Trend Par.** | Uncond. | SP | SQ-SP | CC-SP |
| 0.1 | 0.0 | 0.6738 | 0.0942 | 0.1690 | 0.1491 |
| 0.1 | 0.5 | 0.6756 | 0.2312 | 0.1768 | 0.1491 |
| 0.1 | 1.0 | 0.6756 | 0.2312 | 0.1768 | 0.1215 |
| 0.3 | 0.0 | 0.5402 | 0.0497 | 0.0479 | 0.0834 |
| 0.3 | 0.5 | 0.5359 | 0.0153 | 0.0322 | 0.0059 |
| 0.3 | 1.0 | 0.5887 | 0.0197 | 0.0184 | 0.0048 |
| 0.5 | 0.0 | 0.4891 | 0.0090 | 0.0259 | 0.0059 |
| 0.5 | 0.5 | 0.6723 | 0.0362 | 0.0380 | 0.0231 |
| 0.5 | 1.0 | 0.7306 | 0.1570 | 0.1584 | 0.0626 |

Table 5.9: Trend Simulations – Results for Dataset 4

| | | **Mean Rel. Error** | | | |
|---|---|---|---|---|---|
| **Baseline** | **Trend Par.** | Uncond. | SP | SQ-SP | CC-SP |
| 0.1 | 0.0 | 0.1273 | 0.0861 | 0.0865 | 0.0884 |
| 0.1 | 0.5 | 0.1929 | 0.0527 | 0.0536 | 0.0545 |
| 0.1 | 1.0 | 0.3061 | 0.0290 | 0.0300 | 0.0314 |
| 0.3 | 0.0 | 0.1114 | 0.0143 | 0.0135 | 0.0134 |
| 0.3 | 0.5 | 0.2996 | 0.0167 | 0.0148 | 0.0145 |
| 0.3 | 1.0 | 0.4172 | 0.0259 | 0.0248 | 0.0210 |
| 0.5 | 0.0 | 0.1604 | 0.0128 | 0.0111 | 0.0117 |
| 0.5 | 0.5 | 0.3432 | 0.0350 | 0.0340 | 0.0280 |
| 0.5 | 1.0 | 0.4277 | 0.0560 | 0.0548 | 0.0350 |

| | | **Max. Rel. Error** | | | |
|---|---|---|---|---|---|
| **Baseline** | **Trend Par.** | Uncond. | SP | SQ-SP | CC-SP |
| 0.1 | 0.0 | 0.4179 | 0.5045 | 0.5160 | 0.4843 |
| 0.1 | 0.5 | 0.3707 | 0.3958 | 0.3908 | 0.3811 |
| 0.1 | 1.0 | 0.3707 | 0.0570 | 0.0616 | 0.0672 |
| 0.3 | 0.0 | 0.4632 | 0.0542 | 0.0488 | 0.0464 |
| 0.3 | 0.5 | 0.5505 | 0.0661 | 0.0642 | 0.0568 |
| 0.3 | 1.0 | 0.5578 | 0.0661 | 0.0642 | 0.0568 |
| 0.5 | 0.0 | 0.4991 | 0.0762 | 0.0740 | 0.0692 |
| 0.5 | 0.5 | 0.5578 | 0.0867 | 0.0838 | 0.0659 |
| 0.5 | 1.0 | 0.5505 | 0.0890 | 0.0864 | 0.0692 |

saddlepoint methods. Tables 5.6 through 5.9 give the mean and maximum relative error of the p-value for testing $H_0 : \Delta = 0$, where the relative error of a p-value (relative to the exact conditional value $p_c$) is given by

$$RE(p) = \frac{|p - p_c|}{p_c}.$$

Tables 5.4 and 5.5 demonstrate that the saddlepoint estimator is only very slightly less biased than the unconditional maximum likelihood estimator for these problems. This is likely due to the fact that the magnitude of the correction to the likelihood is quite small because there is only one nuisance parameter. In chapter 4, where there were many nuisance parameters, the correction (and the bias in the unconditional estimate) were far more substantial. For estimation of $\Delta$, which is typically not the result of major interest in trend test problems, both methods perform reasonably well.

Tables 5.6 through 5.9 show that the three saddlepoint methods perform quite well in approximating exact conditional inference on the trend parameter for typical sizes of the trend parameter. The maximum relative error in these data is rarely more than 10%. It is quite high for the fourth set of marginal totals, but this occurs in data of the form $\{x, 0, 0, 0, 0, 1\}$ where $x$ is between 5 and 15. For these data it is unlikely that a trend test would be performed, because of the extreme sparseness. However, if one is interested in a trend test for data of this extreme sparse nature, it is necessary to use exact conditional methods. Removing datasets of this type makes the maximum relative error for the four methods $\{0.3291, 0.1905, 0.1792, 0.2017\}$ for the first trial and $\{0.3707, 0.1742, 0.1690, 0.1732\}$ for the second, which confirms that the saddlepoint methods perform well under normal circumstances. The continuity corrected saddlepoint and the sequential saddlepoint methods perform slightly better than the uncorrected methods in most cases, but this is not without exception. This fact, and the fact that implementation of the sequential and continuity corrected saddlepoints is significantly more complicated than implementation of the uncorrected saddlepoint, indicates that it is probably reasonable to use the uncorrected method in most instances.

It is interesting to note that the maximum relative error is the same for many trials with the same marginal totals. This occurs simply because the dataset which generates the maximum relative error for a given marginal configuration is generated by the simulation with several different parameter configurations.

Overall, the saddlepoint method appears from this simulation study to be extremely accurate at approximating exact conditional inference in the trend test. Unconditional maximum likelihood inference should be used with great caution in these problems, as the relative error of the tail probability is on average quite high and as high as 70% in some cases.

## Chapter 6

# HIGH ORDER INTERACTION IN LOGISTIC REGRESSION – A CASE STUDY

### *6.1 Introduction*

Jarvik et al (1996) reported on the use of the $\varepsilon 4$ allele of the apolipoprotein E locus (APOE) in predicting progression to Alzheimer disease (AD). Their case-control study examined APOE and a variety of other predictors and their effects on development of AD. One aspect of this study used a logistic regression analysis to predict disease status using four predictors – age, sex, presence of family history of disease, and APOE genotype. Of interest were main effects as well as two, three and four-way interactions. Analysis was done using ordinary unconditional logistic regression.

The estimators and test statistics given by unconditional logistic regression are consistent only when the information matrix is not extremely sparse – specifically, it is necessary that all the cell totals increase with the sample size (Fahrmeir and Tutz, 1994, Ch. 2). When the data are sparse and higher order interactions are of interest, as was the case in the APOE study, it is quite likely that unconditional logistic regression will break down. The obvious solution is to test for interactions conditionally on the sufficient statistics for other parameters. Conditional logistic regression is an available procedure, but it is extremely slow and cumbersome in general situations, and can also give conservative results when the conditioning reference set is very discrete. Also, conditional logistic regression is operationally limited in the size of problems it can handle.

Double saddlepoint inference may provide an alternative here that approximates the conditional distribution quickly and accurately. This chapter will examine the use of double saddlepoint inference in this situation. We first develop the model (including an extension of the saddlepoint method to consider hypotheses on multiple parameters in discrete data), and then proceed to analysis of the APOE data and a simulation study to determine the effectiveness of the saddlepoint approximation.

## 6.2 The Model

Logistic regression can be viewed as an extension of the model used for the trend test (chapter 5). For $n$ subjects, let $Y_i$, $i = 1, \dots, n$ indicate case or control status, and let $X_i = X_{i1}, \dots, X_{ip}$ be a vector of predictors. The variables $X_i$ may include interactions between other predictors. The logistic regression model uses a logit link and

$$P(Y_i = 1 | X_i = x_i) = \frac{e^{\beta^T x_i}}{1 + e^{\beta^T x_i}}. \tag{6.1}$$

This leads to the likelihood

$$L(\beta) \propto \prod_{i=1}^{n} \frac{(e^{\beta^T x_i})^{y_i}}{(1 + e^{\beta^T x_i})} \tag{6.2}$$

with log-likelihood

$$l(\beta) = \sum_{i=1}^{n} (\beta^T x_i) y_i - \log(1 + e^{\beta^T x_i}) + C \tag{6.3}$$

for some constant $C$. It should be noted that this is not the correct likelihood for case control data, but that under certain assumptions this likelihood provides the correct parameter estimates for the true likelihood.

### 6.2.1 Unconditional Inference

Unconditional inference is performed using ordinary generalized linear model techniques. The likelihood and log-likelihood are given in equations (6.2) and (6.3), the score vector has components

$$\dot{l}_j = \sum_{i=1}^{n} x_{ij}y_i - \frac{x_{ij}e^{\beta x_i}}{1+e^{\beta x_i}}, \tag{6.4}$$

and the information matrix has components

$$I_{jk} = \sum_{i=1}^{n} \frac{x_{ij}x_{ik}e^{\beta x_i}}{(1+e^{\beta x_i})^2}. \tag{6.5}$$

The Newton-Raphson method with the above score and information statistics are used to derive $\hat{\beta}$, the maximum likelihood estimator.

The important issue in this problem is of course testing the hypothesis that $\beta$ is fixed for some components of the vector $\beta$, or more generally, $H_0 : \beta_j = \beta_{j0}$, $j = q, \ldots, p$. This means that the first $q-1$ components of $\beta$ are unrestricted while the last $p-q$ are fixed at specified values. To test this hypothesis, the most often used statistics are the log-likelihood ratio statistic,

$$2\log\Lambda = 2\left\{l_P(\hat{\beta}) - l_P(\tilde{\beta})\right\} \tag{6.6}$$

and the Wald statistic

$$\frac{(\hat{\beta}-\tilde{\beta})^2}{I_P(\hat{\beta})} \tag{6.7}$$

where $\hat{\beta}$ refers to the unrestricted maximum likelihood estimator of $\beta$, and $\tilde{\beta}$ refers to the maximum likelihood estimator under the null hypothesis. Both the likelihood ratio statistic and the Wald statistic are asymptotically $\chi^2$ with $r$ degrees of freedom, where $r$ is the number of parameters $\beta_j$ set fixed under the null hypothesis. When $r$ is 1, the signed square root of both statistics are asymptotically standard normal.

### *6.2.2 Conditional Inference*

Conditional logistic regression can be performed on these data by conditioning on the sufficient statistics for the nuisance parameters. It is possible to develop a conditional likelihood for the data based on non-central hypergeometric densities. Tail probabilities can be calculated from this likelihood, and confidence limits for the parameters can be derived using bisection or other root-finding routines. However, when the $n$ are large (as in this problem) this distribution is in fact not useful computationally in practice. Analysis of the APOE data of Jarvik et al (1996) was attempted using these methods and the program LogXact (Cytel Software), which does exact conditional estimation and testing, but it was not successful.

### *6.2.3 Double Saddlepoint Inference*

Double saddlepoint inference for a one-parameter hypothesis is just a generalization of the trend test.

Consider the situation where we wish to make inference on $\beta^* = (\beta_q, \ldots, \beta_p)^T$ conditional on the sufficient statistics for $\beta_1, \ldots, \beta_{q-1}$. The modified profile likelihood for $\beta_1$ is given by

$$L_{MP}(\beta^*) = L_P(\beta^*)\sqrt{|I_{11}(\beta)|} \tag{6.8}$$

where $L_P$ is the ordinary multiple logistic regression profile likelihood, and $I_{11}$ is the component of the information matrix corresponding to $\beta_1, \ldots, \beta_{q-1}$. Since testing is of primary interest in this problem and the determinant of the matrix $I_{11}$ is a complicated function of $\beta$, we will not investigate maximization of this likelihood. The primary interest of this chapter will be to examine the performance of the tail probability and test statistics.

When $q = p$ and $\beta^* = \beta_p$ is one-dimensional, the tail probability approximation as derived by Pierce and Peters (1992) takes the form

$$P(T_{\leq} t) \doteq \Phi\left[w + \frac{1}{w}\{\log\rho - \log(w/z)\}\right] \tag{6.9}$$

where

$$w = \operatorname{sign}(\hat{\beta}_p - \beta_{p0})\sqrt{2\log\Lambda} \tag{6.10}$$

and

$$z = \sqrt{W_n} \tag{6.11}$$

where $2\log\Lambda$ is the standard likelihood ratio statistic for testing $\beta_p = \beta_{p0}$. The Wald statistic for testing $\beta_p = \beta_{p0}$ is

$$W_n = (\hat{\beta}_p - \beta_{p0})^2 I_{22\cdot 1}(\hat{\beta}_p).$$

and

$$\rho = [|I_{11}(\hat{\beta})|/|I_{11}(\hat{\beta}_1(\beta_{p0}), \beta_{p0})|]^{1/2}. \tag{6.12}$$

*Extension to Hypotheses on More Than One Parameter*

Tail probability approximations for saddlepoint inference on discrete data as developed by Skovgaard (1987) and Pierce and Peters (1992) can only accommodate single-degree of freedom hypotheses. The extension of this to hypotheses on more than one parameter poses some theoretical problems. The Wald and likelihood ratio statistics both generalize to $\chi^2$ statistics for multiple degree of freedom tests, and we will examine their performance on multi-parameter hypotheses.

## 6.3 Data Analysis

In this section we present an analysis of the Jarvik et al (1996) data to determine if the saddlepoint methods provide different conclusions from the unconditional likelihood methods used in the original work.

### 6.3.1 The Data

The data consist of 237 cases and 310 controls, ranging in age from 48 to 94 years with a mean age of 79. Two cases and two controls were less than 60 years of age. The variables considered for analysis are presence of family history of memory problems, APOE genotype (five categories), age and sex. Family history was not available for 33 cases and 40 controls. APOE genotype was unavailable for four cases and six controls. Also, the $\varepsilon 2\varepsilon 2$ genotype only occurred in four controls. Eliminating all

Table 6.1: APOE Data Summary

| | **Cases** | **Controls** |
|---|---|---|
| Total | 267 | 310 |
| Male | 87 | 117 |
| (mean age) | 79.7 | 79.4 |
| Female | 150 | 193 |
| (mean agè) | 79.4 | 78.6 |
| Fam. Hist + | 85 | 57 |
| Fam. Hist − | 119 | 213 |
| Fam. Hist. Missing | 33 | 40 |
| APOE: $\varepsilon2\varepsilon2$ | 0 | 4 |
| $\varepsilon2\varepsilon3$ | 8 | 38 |
| $\varepsilon2\varepsilon4$ | 10 | 15 |
| $\varepsilon3\varepsilon3$ | 89 | 182 |
| $\varepsilon3\varepsilon4$ | 96 | 63 |
| $\varepsilon4\varepsilon4$ | 30 | 2 |
| APOE missing | 4 | 6 |

these subjects from the sample leaves a sample size of 465, with 202 cases and 263 controls. A summary of the entire dataset (including subjects eliminated from the analysis) is given in table 6.1.

### *6.3.2 Analysis*

Of interest in this particular problem are the interactions between the four factors in table 6.1. The most important of these interactions are the interaction between family history and genotype and the four-way interaction between family history, genotype, gender and age. If family history is a predictor of Alzheimer disease in $\varepsilon 4+$ subjects then it is likely that there is another genetic factor involved in the disease.

For simplicity of computation we have dichotomized APOE into the presence or absence of the $\varepsilon 4$ haplotype. A more detailed analysis can be done, but proves exceedingly complicated for high-order interactions. Because of the computational difficulties with the four-way interaction (computer programs for determinants and matrix inverses have difficulty with $40 \times 40$ matrices), we take a two-tiered approach to the problem.

First, to examine the fourth-order interaction, we dichotomize the APOE genotype variable in a model containing all lower-order terms (first order interactions). Second, to examine the more important second-order interaction between family history and genotype, we examine the more detailed model with genotype as a factor with five levels.

For each case, we consider forwards stepwise regression, building a model starting with the individual factors and working up to high-order interactions. To test the significance of an individual factor, we compare the null model to the model with the factor. To test a two-way interaction, we test the model with the two individual factors against the model with the individual factors and the interaction term. To test the four-way interaction we compare a model with all first, second and third order terms to the full model.

We report four test statistics and their p-values. Unconditional analysis is done using directed likelihood ratio and Wald statistics, and saddlepoint inference is done with and without the continuity correction. It is worth noting that the continuity correction does not affect the saddlepoint method in the way one would expect, uniformly moving the test statistic away from the boundary. This is because the continuity corrected saddlepoint statistic is a complex function of the corrected likelihood ratio statistic, the corrected Wald statistic, and $\rho$, and increases or decreases depending on the relative correction in all of the above statistics. This means that

Table 6.2: Results of APOE Analysis – Main Effects

| Factor | | Lik.Rat. | Wald | SP | CC-SP |
|---|---|---|---|---|---|
| G | $Z$ | 0.3365 | 0.3330 | 0.2094 | 1.0590 |
| | $p$ | 0.3683 | 0.3696 | 0.4171 | 0.1448 |
| F | $Z$ | -4.8287 | -4.7638 | -4.9312 | -4.8509 |
| | $p$ | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| A | $Z$ | 1.3737 | 1.3654 | 1.3609 | 1.4022 |
| | $p$ | 0.0848 | 0.0868 | 0.0804 | 0.0906 |
| E | $Z$ | -2.1595 | -2.1569 | -2.2559 | -2.1861 |
| | $p$ | 0.0154 | 0.0120 | 0.0144 | 0.0167 |

the p-value may be smaller or larger depending on these relative corrections.

### *6.3.3 Results*

Tables 6.2 and 6.3 present the results of the basic analysis. In the tables, family history is labeled F, gender is labeled G, APOE type is labeled E, and age is labeled A.

Table 6.3: Results of APOE Analysis – Interactions

| Factor | | Lik.Rat. | Wald | SP | CC-SP |
|---|---|---|---|---|---|
| E*F | $Z$ | 1.8352 | 1.8160 | 1.8157 | 1.8771 |
| | $p$ | 0.0332 | 0.0347 | 0.0303 | 0.0384 |
| E*S*A*F | $Z$ | 1.0816 | 1.0709 | 0.9360 | 1.0415 |
| | $p$ | 0.1397 | 0.1421 | 0.1746 | 0.1488 |

Table 6.4: Results of APOE Analysis – APOE as a Factor

| Factor | | Lik.Rat. | Wald |
|---|---|---|---|
| E | $Z^2$ | 8.8135 | 7.3423 |
| | *df* | 4 | 4 |
| | $p$ | 0.0000 | 0.0000 |
| F*E | $Z^2$ | -7.9324 | -5.6672 |
| | *df* | 4 | 4 |
| | $p$ | 0.0941 | 0.2254 |

From these tables, it appears that saddlepoint inference is usually more conservative than unconditional inference. This is not unexpected as saddlepoint inference approximates conditional inference which is known to be conservative. Fortunately, for this dataset and this analysis, the three analysis methods almost always provided the same conclusions. There were some odd results for the continuity corrected saddlepoint method when the test statistic was near zero, but the uncorrected saddlepoint approximation provided reasonable results in almost all cases. The problem with the continuity corrected saddlepoint near the origin has to do with the fact that the continuity correction can have a large effect on the p-value and also can cause the test statistic to change sign.

The conclusion that can be drawn from this first analysis is that controlling for all other factors, the four-way interaction is not statistically significant when genotype is considered as a dichotomous variable, but that interaction between family history and genotype is statistically significant when genotype is considered as a dichotomous variable. The implication of this is that the genetic factor (APOE genotype) may not be simply presence or absence of the $\varepsilon 4$ haplotype, and more complicated analysis is necessary.

The more complex analysis, retaining APOE genotype as a factor, gave the results in table 6.4. The two tests done here are first for the main effect of genotype on disease status and second for the interaction between family history and genotype. Genotype is highly significant here, as it was previously, but when the interaction is between the full factor form of the genotype and family history, it is not statistically significant when controlling only for family history and genotype. This example illustrates that when the data are sparse, the significance level of the two unconditional tests can vary wildly.

In conclusion, this dataset, though extremely sparse and having a poorly conditioned information matrix, is not an example where the saddlepoint inference provides results that are significantly different from those found using unconditional inference. This is a good result for researchers, in that the fact that the saddlepoint corrections are small indicates that the unconditional results can be trusted to be reasonable. The conditions on accuracy of the saddlepoint approximations in estimating the conditional tail probability are weaker than those on the unconditional tail probabilities.

## *6.4 Simulation Study*

We simulate data similar to the APOE data to determine the performance of unconditional and saddlepoint methods on this sort of data and to determine if there are situations where the saddlepoint methods provide different conclusions than do unconditional methods and should be used in place of unconditional methods. Computational problems preclude the use of exact conditional logistic regression methods for this problem.

### *6.4.1 Design of Simulation Study*

The simulation is designed to reflect the problems with the APOE dataset. However, in the interests of computing speed, the model is slightly simplified to reduce complexity without disturbing the character of the data.

To ensure a reasonably large dataset, we simulated 200 subjects. The simulation examined four predictor variables and their interactions, family history of Alzheimer Disease (coded 0 for no, 1 for yes), sex, age, and APOE genotype. For simplicity (as in the data analysis), we assume that APOE is a binary variable rather than four binary variables. This reduces the dimensionality of the problem without reducing much of the complexity. The data were generated so that approximately half of the sample were male, half female, and that the mean age was 80.0, near the mean age for the sample in the Jarvik et al (1996) study. A positive family history occurred in on average 30% of simulated subjects, and the APOE genotype was $\epsilon 4+$ in 30% of the subjects. Correlation among the covariates was moderate, ranging from odds ratios of 1.07 to 1.33, and this was not considered in the simulation study. The covariates were generated independently of one another according to these distributions using a random number generator. The disease variable was simulated using a binomial random number generator, with probability of disease equal to

$$P(AD|X=x) = \frac{e^{\sum \beta_i x_i}}{(1+e^{\sum \beta_i x_i})} \tag{6.13}$$

where the $\beta$ were chosen randomly from the set 0.0, ±0.2, ±0.5, ±0.7, except for the fourth-order interaction parameter. Runs were performed with this parameter ranging from 0 to 1.0 by 0.2, so as to determine the power curve for each of the tests. 500 simulation runs were performed at each level of the interaction parameter.

Analysis was done using unconditional methods, both the likelihood ratio and Wald statistics, and uncorrected saddlepoint approximations.

### *6.4.2 Results of Simulation Study*

The results of the simulation study are summarized in tables 6.5 through 6.7. Table 6.5 is a summary of the power of the test at each of the tested levels of the interaction parameter. The power at $\beta = 0$ is the simulated $\alpha$ level for the test. Table 6.6 is

Table 6.5: Power of Tests for Four-Way Interaction

| | $\beta$ | | | | | |
|---|---|---|---|---|---|---|
| Test | 0 | 0.2 | 0.4 | 0.6 | 0.8 | 1.0 |
| LR | 0.044 | 0.043 | 0.072 | 0.044 | 0.059 | 0.094 |
| W | 0.046 | 0.046 | 0.060 | 0.044 | 0.057 | 0.087 |
| SP | 0.050 | 0.043 | 0.059 | 0.048 | 0.047 | 0.074 |

a 3 × 3 × 5 table which gives the percentage of the simulations for which a pair of test statistics gave the same conclusion. Table 6.7 is a summary of the coverage probability for 95% confidence intervals for the four-way interaction parameter $\beta$.

Several simulated datasets, particularly for higher values of the interaction parameter, resulted in estimates reaching boundary points. These datasets were removed from the analysis.

Some conclusions are immediately apparent from these data. The coverage probability of the 95% confidence intervals for the two unconditional tests are quite poor, ranging from 75% to 86%. The saddlepoint confidence interval performed better, having coverage probability close to the nominal 95% level. For confidence limits in this situation, it appears that saddlepoint methods are more accurate than unconditional methods.

The power of the hypothesis tests for detecting an interaction is fairly poor on all counts. All three tests had the correct size, but the power curves do not increase with the value of the parameter, even with the number of sampling runs increased to 2000. It should be noted that in most of the cases where there is a decline in power, the decline is within the range that could occur due to sampling error.

Finally, Pierce and Peters (1992) proposed the saddlepoint method as a method of differentiating between effectiveness of the two unconditional methods. Table 6.6

Table 6.6: Agreement Between Tests of Four-Way Interaction

| 1 |  | LR | W | SP |
|---|---|---|---|---|
|  | LR | X | 0.997 | 0.983 |
|  | W | - | X | 0.985 |
|  | SP | - | - | X |

| 2 |  | LR | W | SP |
|---|---|---|---|---|
|  | LR | X | 0.995 | 0.989 |
|  | W | - | X | 0.994 |
|  | SP | - | - | X |

| 3 |  | LR | W | SP |
|---|---|---|---|---|
|  | LR | X | 0.989 | 0.980 |
|  | W | - | X | 0.991 |
|  | SP | - | - | X |

| 4 |  | LR | W | SP |
|---|---|---|---|---|
|  | LR | X | 0.991 | 0.981 |
|  | W | - | X | 0.990 |
|  | SP | - | - | X |

| 5 |  | LR | W | SP |
|---|---|---|---|---|
|  | LR | X | 0.996 | 0.992 |
|  | W | - | X | 0.995 |
|  | SP | - | - | X |

| 6 |  | LR | W | SP |
|---|---|---|---|---|
|  | LR | X | 0.993 | 0.987 |
|  | W | - | X | 0.993 |
|  | SP | - | - | X |

Table 6.7: Coverage Probabilities – Four-Way Interaction Parameter

| $\beta$ | **Method** | | |
|---|---|---|---|
| | LR | Wald | SP |
| **0.0** | 0.7844 | 0.7725 | 0.9858 |
| **0.2** | 0.8311 | 0.8147 | 0.9292 |
| **0.4** | 0.8412 | 0.8353 | 0.9412 |
| **0.6** | 0.8432 | 0.8571 | 0.9861 |
| **0.8** | 0.8534 | 0.8647 | 0.9774 |
| **1.0** | 0.7823 | 0.7661 | 0.8266 |

gives the percentage of the simulation runs for which the various tests agree. Overall, the saddlepoint results tend to agree with the Wald results more often than they do with the likelihood ratio test results.

In conclusion, we find that the saddlepoint test performs as well or better than the two unconditional tests on these data. The unconditional confidence intervals have far lower than nominal coverage probability, while the saddlepoint intervals are in general close to the nominal value. While for the data given (200 observations with 16 predictors) all the tests are under-powered to detect a difference, the saddlepoint tests have reasonable significance levels and power that is close to that of the unconditional tests.

Without conditional inference for comparison it is not possible to say how well the saddlepoint methods approximate conditional methods. However, it can be said that the saddlepoint methods do provide reasonable inference, and based on these simulations, better inference than do the unconditional methods.

## Chapter 7

# DISCUSSION OF GENERAL RESULTS IN THE GLM

### 7.1 *Introduction*

In this section we review general results on the saddlepoint approximations in the generalized linear model and expand briefly on these results.

Of interest is the accuracy of the saddlepoint approximations for an arbitrary exponential family model where the number of parameters increases with sample size. We propose a general theorem in this direction, and report a result due to Barndorff-Nielsen (1994) and a generalization of this result, as well as conjectures regarding a proof of the general theorem.

### 7.2 *Previous Theoretical Work*

We state here two results, one due to Portnoy (1988) and the other due to Barndorff-Nielsen (1994), regarding inference when the number of nuisance parameters increases with sample size.

Portnoy (1988) proved the following:

**Theorem 7.2.1** *Let $X_1, \ldots, X_n$ be i.i.d. from an exponential family with density*

$$f_X(x) = \exp\left\{\theta_n^t x - b(\theta_n) - d(x)\right\}$$

*where $\theta_n$ and $X_i$ are p-dimensional vectors and p increases with n (meaning that the dimensionality of $b'$ increases with n). Let $\gamma_1(\theta)$ be the minimum eigenvalue of $c''(\theta)$ and let $\gamma_2(\theta)$ be the maximum eigenvalue. Define $B_n$ and $b_n$ by $B_n = \gamma_2^{1/2}(\theta_n)/\gamma_1(\theta_n)$*

*and* $b_n = \sqrt{p/n}B_n$. *Let* $\theta_n$ *satisfy* $B_n = O(1)$ *and let* $\|a\|$ *denote the Euclidean norm of* $a$. *Suppose that*

$$\sup\left\{\left|E_\theta(a^TX)^3\right| : \|\theta - \theta_n\| \leq (1.2)b_n, \|a_n = 1\|\right\}$$
$$\leq (0.1)\sqrt{n/p}\gamma_1^2(\theta_n)/\gamma_2^{1/2}(\theta_n).$$

*Then the maximum likelihood estimators* $\hat{\theta}_n$ *are consistent, i.e.,*

$$\left\|\theta_n - \hat{\theta}_n\right\| = O_p(\sqrt{p/n}). \tag{7.1}$$

The conclusion that can be drawn is that in cases where $p$ increases with $n$, the maximum likelihood estimator for the entire parameter $\theta$ (and hence the maximum profile likelihood estimator for a component of $\theta$) is consistent if $\sqrt{p/n}$ converges to zero as $n$ increases to infinity. Portnoy's paper also contained results regarding asymptotic normality under similar conditions.

Barndorff-Nielsen (1994) proved the following result: Let $\exp Y_{i1}, \ldots, \exp Y_{im}$, $i = 1, \ldots, q$ (so that the total number of observations is $mq$), have the gamma density

$$f_{Y_{ij}}(y_{ij}) = \frac{\alpha_i^\beta}{\Gamma(\beta)}\exp(-\alpha_i e^{y_{ij}})\exp(y_{ij}\beta)$$

(or $Y$ ~log gamma). The unconditional maximum profile likelihood estimator $\hat{\beta}$ has the same asymptotic distribution as the maximum exact conditional likelihood estimator $\hat{\beta}_C$ if $q/m \rightarrow 0$, while the maximum modified profile likelihood estimator $\hat{\beta}_{MP}$ has the same asymptotic distribution as $\hat{\beta}_C$ if $q/m^3 \rightarrow 0$.

## 7.3 Original Work

When the data are from the exponential family, we can write the exact conditional likelihood for $\theta_1$ conditional on the sufficient statistics for $\theta_2$ as the ratio of two likelihoods,

$$\frac{L(\theta)}{L_{\text{marg}}(\theta|T_2)} \tag{7.2}$$

which can be written using basic saddlepoint methods as

$$\frac{|I_{22}(\theta_1, \hat{\theta}_2(\theta_1))|^{1/2}}{|I(\hat{\theta})|^{1/2}} \frac{L_P(\theta_1)}{L_P(\hat{\theta}_1)} \left\{1 + Q_4(0, \kappa_{20}^{(4)})\right\} \tag{7.3}$$

$$\sim \ |I_{22}(\theta_1, \hat{\theta}_2(\theta_1))|^{1/2} L_P(\theta_1) \left\{1 + Q_4(0, \kappa_{20}^{(4)})\right\} \tag{7.4}$$

where $Q_4(0, \kappa_{20}^{(4)})$ is a polynomial of degree four containing log likelihood derivatives and information quantities of order four, all of which are functions of $\theta_1$ and $\hat{\theta}_2(\theta_1)$. To make conclusions regarding the general double saddlepoint approximation under the many nuisance parameter problem, one can investigate these polynomials and their properties with many nuisance parameters. This is difficult unless the polynomial $Q_4(\cdot)$ has special properties, as it does in the following situations.

The result of Barndorff-Nielsen can be generalized in a straightforward fashion to a class of distributions due to Bar-Lev and Reiser (1982).

Consider the set of distributions with exponential family density

$$\exp\left\{\theta_1 t_1(x) + \theta_2 t_2(x) + b(\theta) - c(x)\right\}, \tag{7.5}$$

where $\theta_1$ and $\theta_2$ are scalars. Let

$$E_\theta(T) = (\eta_1, \eta_2),$$

and assume that

$$\theta_2 = -\theta_1 * \psi\prime(\eta_2), \tag{7.6}$$

where $\psi$ is some known function and $\psi\prime(\eta_2)$ denotes the derivative of $\psi$ with respect to $\eta_2$. This family includes the gamma, normal inverse gaussian distribution and one-one transformations of these densities. It has been conjectured that those are the only densities in this family. The binomial distribution and the Poisson distribution, which are two of the distributions of greatest interest in conditional inference, are not members of this family. However, Barndorff-Nielsen's (1994) result can readily be generalized to distributions in this family as will be demonstrated below.

**Theorem 7.3.1** *Let* $X_{i1}, \ldots, X_{im}$*,* $i = 1, \ldots, p$ *be* $p$ *samples of size* $m$ *(for a total of* $mp$ *observations) from the normal, inverse gaussian or gamma distribution or a one-to-one transformation of one of these densities. These densities are of the form (7.5) with parameters* $\theta_1$ *and* $\theta_{2i}$*, where (7.6) holds for* $\theta_1$*,* $\theta_{2i}$ *and* $\eta_{2i} = E_\theta(X_{ij})$*. Then, if* $\hat{\theta}$ *is the unconditional maximum likelihood estimator for* $\theta_1$*,* $\hat{\theta}_{1MP}$ *the saddlepoint maximum modified profile likelihood estimator for* $\theta_1$ *and* $\hat{\theta}_{1C}$ *the conditional likelihood estimator for* $\theta$*, the following is true:* $\hat{\theta}_{1MP}$ *is asymptotically equivalent to* $\hat{\theta}_{1C}$ *if* $p/m^3 \to 0$*, while* $\hat{\theta}$ *is asymptotically equivalent to* $\hat{\theta}_{1C}$ *if* $p/m \to 0$*.*

It is worth noting that the condition on the unconditional maximum likelihood estimator is much stronger than that on the maximum modified profile likelihood estimator. Also it is assumed that the conditional maximum likelihood estimator has "gold-standard" properties in that it provides the best inference in an asymptotic sense.

*Proof.* First consider one sample $(X_1, \ldots, X_m)$ of size $m$ from a density of the form (7.5). By (7.6),

$$\theta_2 = -\theta_1 \psi'(\eta_2).$$

Now, the log-likelihood for $\theta_1$ and $\theta_2$ is

$$l(\theta_1, \theta_2 | X = x) = \theta_1 t_{1\cdot} + \theta_2 t_{2\cdot} + mb(\theta_1, \theta_2) \tag{7.7}$$

where $t_{j\cdot} = \sum_{i=1}^m t_j(x_i)$. and to maximize over $\theta_2$, we differentiate (7.7) and set it to zero and derive

$$\frac{t_{2\cdot}}{m} = \frac{\partial b(\theta_1, \theta_2)}{\partial \theta_2}, \tag{7.8}$$

and hence

$$\frac{t_{2\cdot}}{m} = \hat{\eta}_2(\theta_1), \tag{7.9}$$

and by substitution into (7.6),

$$\hat{\theta}_2(\theta_1) = -\theta_1 \psi'(t_{2\cdot}/m). \tag{7.10}$$

Now, this means that the log-profile-likelihood can be written as

$$\theta_1 t_{1\cdot} - \theta_1 \psi'(t_{2\cdot}/m) t_2 + mc\{\theta_1, -\theta_1 \psi'(t_{2\cdot}/m)\}. \tag{7.11}$$

Bar-Lev and Reiser (1982) showed that for this class of models, there exists a function $M$ such that

$$b(\theta_1, -\theta_1 \psi'(t_{2\cdot}/m)) = \theta_1 \psi'(t_{2\cdot}/m)\hat{\eta}(\theta_1) - \theta_1 \psi(\hat{\eta}_2) - M(\theta_1). \tag{7.12}$$

Substituting (7.12) into (7.11) and rearranging, the profile log-likelihood can be written as

$$\theta_1 Z - mM(\theta_1), \tag{7.13}$$

where

$$\begin{aligned} Z &= t_{1\cdot} - m\psi(\hat{\eta}_2) - \psi'(t_{2\cdot}/n)(t_{2\cdot} - m\hat{\eta}_2(\theta_1)) \\ &= t_1 - m\psi(t_{2\cdot}/m). \end{aligned}$$

Bar-lev and Reiser (1982) demonstrated that $Z$ and $t_{2\cdot}$ are independent and that the density of $Z$ is of the form

$$f_Z(z) = d(z) \exp\left[\theta_1 z - \{mM(\theta_1) - M(m\theta_1)\}\right]. \tag{7.14}$$

The conditional likelihood for $\theta_1$ given $T_{2\cdot}$ can be determined from (7.14) to be

$$L(\theta_1 | t_2) = \exp\left[\theta_1 z - \{mM(\theta_1) - M(m\theta_1)\}\right]. \tag{7.15}$$

Note that this is an exponential family density which means that the derivative of the log-likelihood with respect to $\theta_1$ is positive.

Now, assume we have $p$ independent samples from the above distribution, with parameters $\theta_1, \theta_{2i}$ for $i = 1, \ldots, p$. Let $Z_i$ denote the statistic $Z$ above from the $i$th sample, and let $T_{ji}$ denote the statistic $T_{j\cdot}$ from the $i$th sample for $J = 1, 2$. The overall profile log-likelihood is

$$l_P(\theta_1) = \sum_i \theta_1 z_i - pmM(\theta_1), \tag{7.16}$$

and the conditional log-likelihood is

$$l_C(\theta_1) = \sum_i \theta_1 z_i - p\left[mM(\theta_1) - M(m\theta_1)\right]. \tag{7.17}$$

To evaluate the modified profile (saddlepoint) likelihood, we must derive

$$\prod_{i=1}^{p} \log\left(\frac{\partial^2 b(\theta_1, \theta_2(\theta_1))}{\partial\theta_2^2}\right)^{1/2},$$

but for an individual sample, $\theta_{2i} = -\theta_1\psi'(\eta_{2i})$, and hence

$$\left\{\frac{\partial\eta_{2i}}{\partial\theta_{2i}}\right\}^{-1} = -\theta_1\psi''(\nu_{2i}),$$

meaning that

$$\frac{\partial\eta_{2i}}{\partial\theta_{2i}} = \frac{-1}{\theta_1\psi''(\hat{\nu}_{2i})} = \frac{-1}{\theta_1\psi''(t_{2i}/m)}. \tag{7.18}$$

We have

$$\eta_{2i} = E(T_{2i}) = -\frac{\partial b(\theta_1, \theta_2)}{\partial\theta_2}, \tag{7.19}$$

and combining (7.18) and (7.19) and rearranging, we can derive that

$$\frac{\partial^2 b(\theta_1, \theta_2(\theta_1))}{\partial\theta_2^2} = \frac{1}{\theta_1\psi''(t_{2i}/m)}, \tag{7.20}$$

and note that $\psi''(t_{2i}/m)$ does not depend on $\theta_1$ and will drop out of the likelihood for $\theta_1$. The profile likelihood modifier is hence proportional to

$$\prod_{i=1}^{p}\left(\frac{1}{\theta_1}\right)^{1/2} = \log\left(\frac{1}{\theta_1}\right)^{p/2}. \tag{7.21}$$

This means that we can write the modified profile log likelihood as

$$l_{MP}(\theta_1) = \sum_i \theta_1 z_i - pm\left\{M(\theta_1) + \left(\frac{1}{2\theta_1}\right)\right\}. \tag{7.22}$$

The corresponding score functions for the three likelihoods can be written

$$\dot{l}_P(\theta_1) = pm\left\{\bar{z} - M'(\theta_1)\right\}, \tag{7.23}$$

$$\dot{l}_C(\theta_1) = pm\left\{\bar{z} - M'(\theta_1) + M'(m\theta_1)\right\}, \tag{7.24}$$

and

$$\dot{l}_{MP}(\theta_1) = pm\left\{\bar{z} - M'(\theta_1) - \frac{1}{2m\theta_1}\right\}. \tag{7.25}$$

Bar-Lev and Reiser (1982) proved that $M$ is infinitely differentiable and $M''(\theta_1) > 0$ for all $\theta_1$. They also showed that for the densities mentioned in the statement of the proof, $M'(\theta_1)$ is expressible as an expansion in powers of $1/(\theta_1)$ with leading term $-1/(2\theta)$. This means that the conditional score (7.24) can be rewritten

$$\dot{l}_C(\theta_1) = pm\left\{\bar{z} - M'(\theta_1) - \frac{1}{2m\theta_1} + \sum_{j=2}^{\infty}\frac{a_j}{(m\theta_1)^j}\right\}, \tag{7.26}$$

for some $a_j$, $j = 2, \ldots, \infty$.

To examine asymptotic properties of $\hat{\theta}_{1P}$ we first assume that $m$ increases to infinity with $p$. We write

$$\begin{aligned}
\frac{1}{\sqrt{pm}}\left\{\dot{l}_C(\theta_1) - \dot{l}_P(\theta_1)\right\} &= \sqrt{pm}\left\{-\frac{1}{2m\theta_1} + \sum_{j=2}^{\infty}\frac{a_j}{(m\theta_1)^j}\right\} \\
&= -\sqrt{\frac{p}{m}}\left\{\frac{1}{2\theta_1} + \sum_{j=2}^{\infty}\frac{a_j}{m^{j-1}\theta_1^j}\right\} \\
&= o_p(\sqrt{\frac{p}{m}}).
\end{aligned} \tag{7.27}$$

Hence, we can apply the same consistency argument as in section 4.3.1 to show that $\hat{\theta}_{1P}$ has the same limit as $\hat{\theta}_{1C}$ provided that $\sqrt{p/m} \to 0$ as $p$ and $m$ increase to infinity. Further,

$$\begin{aligned}
\frac{1}{pm}\left\{\ddot{l}_C(\theta_1) - \ddot{l}_P(\theta_1)\right\} &= M''(m\theta_1) \\
&= \frac{1}{2(m\theta_1)^2} - \sum_{j=2}^{\infty}\frac{ja_j}{(m\theta_1)^{j+1}} \\
&= \sqrt{\frac{p}{m}}\left\{\frac{1}{2\sqrt{pm^3}\theta_1^2} - \sum_{j=2}^{\infty}\frac{ja_j}{\sqrt{pm}m^j\theta_1^{j+1}}\right\} \\
&= o_p(\sqrt{\frac{p}{m}}).
\end{aligned} \tag{7.28}$$

Hence,

$$
\begin{aligned}
\sqrt{pm}\left|\hat{\theta}_{1P}-\hat{\theta}_{1C}\right| &= \frac{\frac{1}{\sqrt{pm}}\dot{l}_C(\theta_1)}{\frac{1}{pm}\ddot{l}_C(\theta_1)}-\frac{\frac{1}{\sqrt{pm}}\dot{l}_P(\theta_1)}{\frac{1}{pm}\ddot{l}_P(\theta_1)}+o_p(1) \\
&= \frac{\frac{1}{\sqrt{pm}}\dot{l}_C(\theta_1)}{\frac{1}{pm}\ddot{l}_C(\theta_1)}-\frac{\frac{1}{\sqrt{pm}}\dot{l}_C(\theta_1)+o_p(\sqrt{\frac{p}{m}})}{\frac{1}{pm}\ddot{l}_C(\theta_1)o_p(\sqrt{\frac{p}{m}})}+o_p(1), \qquad (7.29)
\end{aligned}
$$

which demonstrates that $\hat{\theta}_{1P}$ is asymptotically equivalent to $\hat{\theta}_{1C}$ provided that

$$\sqrt{p/m} \to 0$$

as $m$ and $p$ increase to infinity.

For $\hat{\theta}_{1MP}$ a similar argument is used to show that

$$\frac{1}{\sqrt{pm}}\left\{\dot{l}_C(\theta_1)-\dot{l}_{MP}(\theta_1)\right\}=o_p(\sqrt{\frac{p}{m^3}}), \qquad (7.30)$$

and that

$$\frac{1}{pm}\left\{\ddot{l}_C(\theta_1)-\ddot{l}_{MP}(\theta_1)\right\}=o_p(\sqrt{\frac{p}{m^3}}). \qquad (7.31)$$

From this we conclude in a similar fashion that

$$
\begin{aligned}
\sqrt{pm}\left|\hat{\theta}_{1MP}-\hat{\theta}_{1C}\right| &= \frac{\frac{1}{\sqrt{pm}}\dot{l}_C(\theta_1)}{\frac{1}{pm}\ddot{l}_C(\theta_1)}-\frac{\frac{1}{\sqrt{pm}}\dot{l}_{MP}(\theta_1)}{\frac{1}{pm}\ddot{l}_{MP}(\theta_1)}+o_p(1) \\
&= \frac{\frac{1}{\sqrt{pm}}\dot{l}_C(\theta_1)}{\frac{1}{pm}\ddot{l}_C(\theta_1)}-\frac{\frac{1}{\sqrt{pm}}\dot{l}_C(\theta_1)+o_p(\sqrt{\frac{p}{m^3}})}{\frac{1}{pm}\ddot{l}_C(\theta_1)o_p(\sqrt{\frac{p}{m^3}})}+o_p(1), \qquad (7.32)
\end{aligned}
$$

which demonstrates that $\hat{\theta}_{1MP}$ is asymptotically equivalent to $\hat{\theta}_{1C}$ provided that

$$\sqrt{p/m^3} \to 0,$$

as $m$ and $p$ increase to infinity. This concludes the proof.

## *7.4 Conclusions*

This is an extremely difficult theoretical area. The results given above are relatively insignificant in terms of practical use, but do demonstrate that in certain situations where the number of nuisance parameters increases with sample size, the saddlepoint/modified profile likelihood inference methods provide better inference than do profile likelihood methods. Unfortunately, the result relies on a rather unique characteristic of the distributions mentioned. It is likely to be difficult to generalize. A useful and more practical result would demonstrate something similar to the above in distributions such as the binomial and Poisson distribution or in general exponential family models.

# Chapter 8

# CONCLUSIONS

## *8.1 Summary of the Dissertation*

This dissertation has examined the development and use of saddlepoint approximations, in particular double saddlepoint approximations, in generalized linear models. We have reviewed the extensive literature on the topic, developed the approximations in detail in two classical conditional inference problems, applied the approximations to an interesting conditional inference problem, and discussed and extended the theory in the area.

The literature of saddlepoint approximations and expansions is extensive but complex and often contains errors. Chapters 2 and 3 are a complete and detailed development of the approximations in the general case and in the generalized linear model. The key aspects of the development are the saddlepoint or modified profile likelihood,

$$L(\theta_1|T_2 = t_2) = L_P(\theta_1)\left|I_{22}(\theta_1, \hat{\theta}_2(\theta_2))\right|^{1/2} \tag{8.1}$$

due to Barndorff-Nielsen(1980, 1983) and the approximate tail probability

$$P(T_1 \le t_1|T_2 = t_2) = \Phi\left\{w + \frac{1}{w}(\log\rho - \log(w/z))\right\} \tag{8.2}$$

due to Barndorff-Nielsen (1986) and Pierce and Peters (1992). The simple nature of the approximations in the generalized linear model makes them particularly useful in these problems.

The problem of log odds ratio regression for stratified case control studies has been studied extensively. Exact conditional inference is the "gold standard" for this inference, and is implemented in a variety of software packages (Mehta, Patel and

Gray 1985, Thomas and Gart 1992). However, a quick and efficient approximate method for this problem is desirable. In chapter 4 we have demonstrated analytically and numerically that the maximum modified profile likelihood estimator is asymptotically consistent to first order for small values of the parameter, and that it is extremely close to the exact conditional maximum likelihood estimator in many situations. We have implemented the saddlepoint approximation for the tail probability and generated confidence intervals both with and without the continuity correction, and also the sequential saddlepoint approximation (Fraser, Reid and Wong, 1991). Through simulation studies, we have examined confidence intervals for the estimator, and demonstrated numerically that the confidence limits generated by the saddlepoint method are very close to the exact conditional confidence limits for a variety of practical situations. The lower confidence limit in particular is virtually identical to the exact conditional limit, and inference on the log odds ratio based on the confidence limits is almost always the same based on the two methods (in 16000 simulation runs, there were only 0.3% (48/16000) where the various saddlepoint approximations and conditional inference disagreed on the 95% confidence interval). It appears from the simulation studies in this dissertation that the continuity correction and the sequential saddlepoint approximation provide slightly improved inference over the uncorrected saddlepoint approximation in this problem, but the improvement is only marginal, and the methods are both more difficult to implement than the uncorrected double saddlepoint approximation. Overall, for this problem, for even extremely sparse data with high odds ratios, the saddlepoint approximations are accurate approximations to the exact conditional method, while the unconditional methods do not provide the same effectiveness.

The test for trend in a sequence of binomial random variables is another classic example where exact conditional inference is considered the gold standard. Bedrick and Hill (1992) examined one form of the tail probability saddlepoint approximation in this problem. In chapter 5 we have examined the saddlepoint approximation for this problem both in terms of estimation and hypothesis testing. We develop a method for maximum modified profile likelihood estimation and for confidence intervals. Simulation studies demonstrate that the maximum modified profile likelihood estimator is similarly biased to the unconditional maximum likelihood estimator, and that it offers little improvement, since neither of them are extremely biased, even in sparse data. On the other hand, for hypothesis testing, the unconditional methods performed poorly, with substantial error, while the saddlepoint methods (uncorrected, corrected and the sequential saddlepoint) all perform extremely well in approximating exact conditional inference.

For the test of high-order interaction in logistic regression in a real dataset on Alzheimer disease and the APOE genotype, little difference was found between the saddlepoint approximations and the unconditional inference results. This indicates that the size of the dataset is large enough to offset problems with conditioning. In smaller datasets of a similar type, it is likely that conditioning will be necessary and that the saddlepoint results and the unconditional results will differ more substantially. Simulation results demonstrate that on average the saddlepoint methods provide reasonable results that are as good as or better than the unconditional results.

The theory of the saddlepoint approximation and the modified profile likelihood is extremely complex and difficult to deal with. Chapter 7 discussed some of the theoretical results on the situation where the number of nuisance parameters increases with sample size. A small extension of an example by Barndorff-Nielsen (1995) to a theorem on the convergence of modified profile likelihood estimators in a class of distributions was proved.

## *8.2 Future Work*

There are many opportunities for future work in this area. It is likely to continue to be an area of great research interest for years to come. Both theoretical and applied statistical challenges are available.

From the applied perspective there is an obvious need for examination of other areas where exact conditional inference is used. Mentioned only in the introduction, but prominent in the literature on conditional inference, is the log-linear model and conditional Poisson regression. This is a problem easily addressed using saddlepoint methods, and the preliminary work done by Pierce and Peters (1992) on this problem showed promising results. Other logistic regression problems, including that of estimation in the general conditional logistic regression model, are also interesting areas to pursue the saddlepoint approximations.

Many interesting problems off the exponential family may also be addressed by saddlepoint approximations. Saddlepoint approximations have recently been suggested for estimating equation problems (Ronchetti and Welsh, 1994) and for general marginal densities (Gatto and Ronchetti, 1996). An exploration of these methods and their uses in practical problems such as generalized estimating equations would be a useful pursuit.

Pierce and Peters (1992) suggested that problems off the exponential family may provide a great deal of interesting research for saddlepoint researchers, because for some of these problems, there is no obvious conditional distribution for comparison. Curved exponential families, including models such as non-linear regression, are an example where work has been done on saddlepoint approximations and other higher-order asymptotics.

Finally, implementation of these methods in commonly used software is essential for their practical use and popular acceptance. The ideal use for these methods is as a secondary test in unconditional inference. When doing unconditional logistic regression, researchers often ignore sparseness in the data and the possible need for conditional methods.

From a theoretical perspective, there is much to be done. A generalization of theorem 7.3.1 to a more practical set of distributions would be a significant development in the theory of saddlepoint approximations. However, this has demonstrated itself to be a very difficult problem, and is unlikely to be solved generally in the near future. Other theoretical issues that are of interest relate to the performance of the confidence limits and test statistics in asymptotic frameworks involving many nuisance parameters. This was not examined in detail in this dissertation and represents a significant theoretical problem for the future.

## *8.3 Conclusions*

This dissertation demonstrates that there are many cases where saddlepoint approximations can be used as a replacement for exact conditional inference. However, there will always be cases where saddlepoint inference will fail, and we have discovered some of these and established loose criteria for their occurrence. There are some important ways in which these methods can be put to use. Saddlepoint methods, which can be easily programmed using ordinary unconditional likelihood methods for generalized linear models, could provide a check on the unconditional methods – if the two methods differed considerably in their answers, the researcher could either report the saddlepoint result (it has been demonstrated in several cases to be a reasonable approximation of the exact conditional result) or use the fact that the two results differ as a reason to consult exact conditional inference software. While faster and more powerful computers are making it more and more convenient to use exact conditional inference, it is still more cumbersome and there are still problems of sufficient size that the exact procedures are infeasible. The saddlepoint method provides both a backup to the unconditional inference and a reasonable approximation to the conditional inference in problems where it is infeasible.

# BIBLIOGRAPHY

[1] Agresti, Alan (1992) "A survey of exact inference for contingency tables" *Statistical Science* **7** pp. 131-153.

[2] Andersen, E.B. (1970) "Asymptotic properties of conditional maximum likelihood estimators" *JRSS B* **32** pp. 283-301.

[3] Bar-Lev, S. and Reiser, B.(1983) "An exponential subfamily which admits UMPU tests based on a single test statistic"*Ann. Stat.* **11** pp. 753-769.

[4] Barndorff-Nielsen, O. (1980) "Conditionality resolutions" *Biometrika* **67** pp. 293-310.

[5] Barndorff-Nielsen, O. (1983) "On a formula for the distribution of the maximum likelihood estimator" *Biometrika* **70** pp. 343-365.

[6] Barndorff-Nielsen, O. (1984) "On conditionality resolution and the likelihood ratio for curved exponential models" *Scand. J. Statist.* **11** pp. 157-170.

[7] Barndorff-Nielsen, O. (1986) "Inference on full or partial parameters based on the standardized signed log likelihood ratio" *Biometrika* **73** pp. 307-322.

[8] Barndorff-Nielsen, O. (1990) "A note on the standardized signed likelihood ratio" *Scand. J. Statist.* **17** pp. 157-160.

[9] Barndorff-Nielsen, O. (1994) "Two index asymptotics" Unpublished technical report, Aarhus University.

[10] Barndorff-Nielsen, O. and Cox, D.R. (1979) "Edgeworth and saddle-point approximations with statistical applications", *JRSS B* **41** pp. 279-312.

[11] Barndorff-Nielsen, O. and Cox, D.R. (1989) *Asymptotic Techniques For Use In Statistics* New York: Chapman and Hall

[12] Barndorff-Nielsen, O. and Cox, D.R. (1994) *Inference and Asymptotics*, New York: Chapman and Hall.

[13] Bedrick, E.J. and Hill, J.R (1992) "An empirical assessment of saddlepoint approximations for testing a logistic regression parameter", *Biometrics* **48** pp. 529-544.

[14] Bickel, P.J., Klaassen, C.A.J., Ritov, Y. and Wellner, J.A. (1993) *Efficient and Adaptive Estimation for Semiparametric Models* Baltimore:Johns Hopkins University Press

[15] Blaesild, P. and Jensen, J.L. (1985) "Saddlepoint formulas for reproductive exponential models" *Scand. J. Statist.* **12** pp. 193-202.

[16] Booth, J.G. and Butler, R.W. (1990) "Randomization distributions and saddlepoint approximations in generalized linear models" *Biometrika* **77** pp. 797-796.

[17] Breslow, N. (1976) "Regression analysis of the log odds ratio: a method for retrospective studies" *Biometrics* **32** pp. 409-416.

[18] Breslow, N. (1981) "Odds ratio estimators when the data are sparse" *Biometrika* **68** pp. 73-84.

[19] Breslow, N. (1982) "Covariance adjustment of relative-risk estimates in matched studies" *Biometrics* **38** pp. 661-672.

[20] Breslow, N. and Cologne, J. (1986) "Methods of estimation in log odds ratio regression models" *Biometrics* **42** pp. 949-954.

[21] Breslow, N.E. and Day, N.E. (1980) *Statistical Methods in Cancer Research:***1**, *Case-Control Studies*, Lyon:IARC

[22] Cramér, H. (1938) "Sur un nouveau théorème-limite de la théorie des probabilités" *Actualités Scientifiques et Industrielles* No. 736, Paris:Hermann & Cie.

[23] Daniels, H.E. (1954) "Saddlepoint approximations in statistics" *Ann. Math. Statist.* **25** pp. 631-649.

[24] Daniels, H.E. (1987) "Tail probability approximations" *International Statistical Review* **55** pp. 37-48.

[25] Davison, A.C. (1988) "Approximate conditional inference in generalized linear models" *JRSS B* **50** pp. 445-461.

[26] Esscher, F. (1932) "On the probability function in the collective theory of risk" *Scand. Aktuarietidsskr.* **15** pp. 175-195.

[27] Fahrmeir, L. and Tutz, G. (1991)*Multivariate Statistical Modelling Based on Generalized Linear Models* New York:Springer-Verlag.

[28] Fraser, D.A.S., Reid, N., and Wong, A. (1991) "Exponential linear models: A two-pass procedure for saddlepoint approximation" *JRSS B* **53** pp. 483-492.

[29] Gart, J.J. (1962) "On the combination of relative risks" *Biometrics* **18** pp. 601-610.

[30] Gart, J.J. (1971) "The comparison of proportions: A review of significance tests, confidence intervals and adjustments for stratification" *Int. Statist. Rev.* **39** pp. 148-169.

[31] Gatto, R. and Ronchetti, E. "General saddlepoint approximations of marginal densities and tail probabilities" *J. Amer. Stat. Assn.* **91** pp. 666-673.

[32] Graubard, B.I. and Korn, E.L (1987) "Choice of column scores for testing independence in ordered $2 \times K$ tables" *Biometrics* **43** pp. 471-476.

[33] Jarvik, G.P., Larson, E.B., Goddard, K., Shellenberg, G.D. and Wijsman, E.M. (1996) "Influence of apolipoprotein E genotype on the transmission of alzheimer disease in a community-based sample" *Am. J. Hum. Genet.* **58** pp. 191-200.

[34] Jensen, J.L. (1986) "Similar tests and the standardized likelihood ratio statistic" *Biometrika* **73**, pp. 567-572.

[35] Jeffreys, H.S. (1962) *Asymptotic Approximations* Oxford:Clarendon Press

[36] Khinchin, A.I. (1949)*Mathematical Foundations of Statistical Mechanics* New York:Dover Publications Inc.

[37] Kolassa, J.E. (1994) *Series Approximation Methods in Statistics*, New York: Springer-Verlag.

[38] Lehmann, E.L. (1983) *Theory of Point Estimation* Belmont, CA: Wadsworth & Brooks/Cole

[39] Levin, B. (1990) "Saddlepoint corrections in logistic analysis" *Biometrika* **77** pp. 275-285.

[40] Levin. B. and Kong, F. (1990) "Bartlett's bias correction is a saddlepoint correction" *Biometrika* **77** pp. 219-221.

[41] Liao, J. (1992) "An algorithm for the mean and variance of the noncentral hypergeometric distribution" *Biometrics* **48** pp. 889-892.

[42] Lugannani, R. and Rice, S.O. (1980) "Saddlepoint approximation for the distribution of the sum of independent random variables" *Adv. Appl. Prob.* **12** pp. 475-490.

[43] Mantel, N. and Haenszel, W (1959) "Statistical aspects of the analysis of data from retrospective studies of disease", *J. Nat. Cancer Inst.* **22** pp. 719-748.

[44] McCullagh, P. (1984) "On the elimination of nuisance parameters in the proportional odds model" *JRSS B* **46** pp. 250-256.

[45] McCullagh, P. and Nelder, J.A.(1989) *Generalized Linear Models*, 3rd. Ed., New York:Chapman and Hall.

[46] Mehta, C.R., Patel, N.R. and Gray, R. (1985) "Computing an exact confidence interval for the common odds ratio in several 2 by 2 contingency tables" *J. Amer. Statist. Assoc.* **80** pp. 969-973.

[47] Montgomery, D.C. (1991) *Design and Analysis of Experiments*, New York:Wiley

[48] Pierce, D. A. and Peters, D. (1992) "Practical use of higher order asymptotics for multiparameter exponential families" *JRSS B* **54** pp. 701-737.

[49] Phillips, A., and Holland, P.W. (1987) "Estimators of the variance of the Mantel-Haenszel log-odds-ratio estimate" *Biometrics* **43** pp. 425-431.

[50] Portnoy, S. "Asymptotic behavior of likelihood methods for exponential families when the number of parameters tends to infinity" *Ann. Stat.* **16** pp. 356-366.

[51] Reid, N. (1988) "Saddlepoint methods and statistical inference" *Statistical Science* **3** pp. 213-238.

[52] Robins, J., Breslow, N.E. and Greenland, S. (1986) "Estimators of the Mantel-Haenszel variance consistent in both sparse data and large-strata limiting models" *Biometrics* **42** pp. 311-323.

[53] Ronchetti, E. and Welsh, A.H. (1994) "Empirical saddlepoint approximations for multivariate M-estimators" *JRSS B* **56** pp. 313-326.

[54] Self, S.G., Longton, G., Kopecky, K.J. and Liang, K.Y. (1991) " On estimating HLA/disease association with application to a study of aplastic anemia" *Biometrics* **47** pp. 53-61.

[55] Skovgaard, I.M. (1987) "Saddlepoint approximations for conditional distributions" *J. Appl. Prob.* **24** pp. 875-887.

[56] Temme, N.M. (1982) "The uniform asymptotic expansion of a class of integrals related to cumulative distribution functions" *SIAM J. Math. Anal.* **13** pp. 239-253.

[57] Thomas, D.G. and Gart, J.J. (1992) "Improved and extended exact and asymptotic methods for the combination of 2×x2 tables" *Computers and Biomedical Research*, **25** pp.75-84.

[58] Wichmann, B.A. and Hill, I.D. (1982) "An efficient and portable pseudo-random number generator (Algorithm 183)" *Applied Statistics* **31** pp. 188-190.

[59] Wolfram Research Inc. (1993) *Mathematica 2.2*

[60] Zelen M (1971) "The analysis of several 2 × 2 tables" *Biometrika* **58** pp. 129-137.

# Appendix A

# SOME THEORY OF COMPLEX VARIABLES

This appendix contains proofs of Abel's lemma and Watson's lemma, both of which are used in chapter 2 in the derivation of the saddlepoint approximations.

**Lemma A.0.1 (Abel's Lemma)** *Take $v$ a nonincreasing function on $[A, B]$ that is bounded above. Define*

$$F(x) = \int_A^x f(u)du$$

*and let $F$ be bounded in the interval so that $h \le F(x) \le H$ for $x$ in $[A, B]$, then*

$$hv(A) \le \int_A^B v(u)f(u)du \le Hv(A). \tag{A.1}$$

*Proof.* Consider the integral

$$I = \int_A^B v(u)f(u)du = \int_A^B v(u)dF(u). \tag{A.2}$$

Using integration by parts we can see that

$$\begin{aligned} I &= v(x)F(x)|_A^B - \int_A^B F(u)dv(u) \\ &= v(B)F(B) - v(A)F(A) - \int_A^B F(u)dv(u) \\ &= v(B)F(B) - \int_A^B F(u)dv(u) \qquad \text{(A.3)} \end{aligned}$$

because $F(A) = 0$. Now, $v$ is non-increasing, so all steps are $\le 0$. Hence, the following are true:

$$\begin{aligned} I &\le v(B)H - H\int_A^B dv(u) \\ &= v(A)H \qquad \text{(A.4)} \end{aligned}$$

and

$$\begin{aligned} I &\geq v(B)h - h\int_A^B dv(u) \\ &= v(A)h \end{aligned} \tag{A.5}$$

which is the result claimed.

**Lemma A.0.1 (Watson's Lemma)** *Let*

$$I = \int_0^Z e^{-az} z^m f(z) dz \tag{A.6}$$

*where $f$ is a possibly complex valued function analytic in a neighborhood of $z = 0$ and $a$ is large and positive. The path of integration is along the real axis. Assume $I$ exists for some $\alpha \leq a$. The result is that*

$$I = \sum_{j=0}^{\infty} \frac{f^{(j)}(0)}{j!} \frac{(m+j)!}{a^{m+j+1}} + O(e^{-aX}) \tag{A.7}$$

*where $X$ is within the radius of convergence of $f$.*

*Proof.* Since $f$ is analytic in some neighborhood of 0, we can take

$$f(z) = \sum_{j=0}^{\infty} \frac{f^{(j)}(0)}{j!} z^n$$

for $|z| \leq R$, where $R$ is the radius of convergence. If $Z$ lies in the circle of convergence, then we can just use the expansion and integrate term by term. Consider however the case $|Z| \geq R$. Take a fixed $X$ so that $0 < X < R$. The basic idea of the proof is to show that the integral is equal to the integral on $[0, X]$ which can be done term by term, in an expansion in powers of $a^{-1}$. When we use this in the saddlepoint approximation, we set $a = n$, which is really what we want. Hence we split up

$$I = \int_0^X e^{-az} z^m f(z) dz + \int_X^Z e^{-az} z^m f(z) dz.$$

For $z \leq X$ we can write

$$f(z) = \sum_{j=0}^{n} \frac{f^{(j)}(0)}{j!} z^n + R_n \tag{A.8}$$

where $R_n$ is bounded, say $|R_n| < Mz^{n+1}$ (by Taylor's theorem). We can then write

$$\int_0^X e^{-az} z^m f(z) dz = \int_0^X e^{-az} z^m (\sum_{j=0}^{n} \frac{f^{(j)}(0)}{j!} z^j + R_n) dz. \tag{A.9}$$

Consider the first term of the right hand side. These integrals are hard to do compared to the same ones, taken from 0 to $\infty$. However, we know that for $u \geq 0$ and $k \geq 0$, $(1+u)^k \leq e^{ku}$ (if $k < 0$ then $(1+u)^k \leq 1$). Hence, putting $z = X(1+u)$,

$$\begin{aligned}
\int_0^X e^{-az} z^k dz &= e^{-aX} X^{k+1} \int_0^\infty e^{-aXu} (1+u)^k du \\
&\leq e^{-aX} X^{k+1} \int_0^\infty e^{(-aX+k)u} du \\
&= \frac{e^{-aX} X^{k+1}}{aX - k}.
\end{aligned}$$

This is $o(e^{-aX})$ for large $a$, and hence we can write

$$\begin{aligned}
&\int_0^X e^{-az} z^m \sum_{j=0}^{n} \frac{f^{(j)}(0)}{j!} z^j dz \\
&= \int_0^\infty e^{-az} z^m \sum_{j=0}^{n} \frac{f^{(j)}(0)}{j!} z^j dz + o(e^{-aX}) \\
&= \sum_{j=0}^{n} \frac{f^{(j)}(0)}{j!} \frac{(m+j)!}{a^{m+j+1}} + o(e^{-aX}).
\end{aligned}$$

Now we need to show that the term $\int_0^X e^{-az} z^m R_n dz$ is also small for large $a$. This is easy because it is equal to

$$\begin{aligned}
\int_0^X e^{-az} z^m \theta_1 M z^{n+1} dz &= \int_0^\infty e^{-az} z^m \theta_2 M z^{n+1} dz \\
&= \frac{\theta_2 M (m+n+1)!}{a^{m+n+2}}
\end{aligned}$$

where $0 \leq |\theta_2| \leq |\theta_1| \leq 1$. Hence, this is $o(a^{-m+n+1})$.

The final term we have to control is $\int_X^Z e^{-az} z^m f(z) dz$. We use Abel's lemma to do so. Let $0 < \alpha < a$, and then define for $X \leq Z_1 \leq Z$,

$$F(Z_1) = \int_X^{Z_1} e^{-\alpha z} z^m f(z) dz.$$

$|F(Z_1)|$ has an upper bound of $N$ (which must exist since the integral exists). Hence, with $v(z) = e^{-(a-\alpha)z}$ we define $f(z) = u(z) + iw(z)$ and get that

$$\left|\int_X^Z e^{-az} z^m f(z) dz\right|^2 = \left|\int_X^Z e^{-az} z^m u(z) dz\right|^2 + \left|\int_X^Z e^{-az} z^m (iw(z)) dz\right|^2 . \text{(A.10)}$$

We apply Abel's lemma to the root of each of the RHS integrals (which are now real) which gives that each one is $\leq Ne^{-(a-\alpha)X}$ and take square roots giving

$$\left|\int_X^Z e^{-az} z^m f(z) dz\right| < \sqrt{2} N e^{-(a-\alpha)X}. \tag{A.11}$$

This term is hence $O(e^{-aX})$.

This concludes the proof of Watson's Lemma.

## Appendix B

# PROFILE LIKELIHOOD MODIFIER DERIVATIVES – LOG ODDS RATIO REGRESSION

Here we give the algebraic details of the derivation of the second and third derivatives of the profile likelihood modifier for the log odds ratio regression problem (chapter 4). In the appendix we drop the subscript for the marginal configuration and derive the results for a single table. $\hat{\alpha}(\beta)$ and $M(\beta)$ refer to the functions $\hat{\alpha}_i$ and $M_i$ for an arbitrary set of margins $m$, $n$, and $t$.

First we derive the derivatives of $\hat{\alpha}(\beta)$ with respect to $\beta$. $\hat{\alpha}$ is defined by equation (4.53) and as derived in chapter 4 satisfies

$$\exp(\hat{\alpha}(0)) = \frac{t}{N-t}.$$

We implicitly differentiate equation (4.53) to get $\partial\hat{\alpha}/\partial\beta$. Recall that

$$\begin{aligned} & t(1+e^{\hat{\alpha}(\beta)+\beta})(1+e^{\hat{\alpha}(\beta)}) = \\ & \qquad m e^{\hat{\alpha}(\beta)}(1+e^{\hat{\alpha}(\beta)+\beta}) + n e^{\hat{\alpha}(\beta)+\beta}(1+e^{\hat{\alpha}(\beta)+\beta}) \end{aligned}$$

or

$$t = m\frac{e^{\hat{\alpha}(\beta)}}{1+e^{\hat{\alpha}(\beta)}} + n\frac{e^{\hat{\alpha}(\beta)+\beta}}{1+e^{\hat{\alpha}(\beta)+\beta}}. \tag{B.1}$$

Differentiating both sides of (B.1) gives

$$0 = m\frac{\partial\hat{\alpha}}{\partial\beta}\frac{e^{\hat{\alpha}(\beta)}}{(1+e^{\hat{\alpha}(\beta)})^2} + n(\frac{\partial\hat{\alpha}}{\partial\beta}+1)\frac{e^{\hat{\alpha}(\beta)+\beta}}{(1+e^{\hat{\alpha}(\beta)+\beta})^2}$$

or

$$\frac{\partial\hat{\alpha}}{\partial\beta} = \frac{-n\frac{e^{\hat{\alpha}(\beta)+\beta}}{(1+e^{\hat{\alpha}(\beta)+\beta})^2}}{m\frac{e^{\hat{\alpha}(\beta)}}{(1+e^{\hat{\alpha}(\beta)})^2} + n\frac{e^{\hat{\alpha}(\beta)+\beta}}{(1+e^{\hat{\alpha}(\beta)+\beta})^2}}. \tag{B.2}$$

At $\beta = 0$, simplification of (B.2) to $-n/N$ is clear.

The second derivative of $\hat{\alpha}$ with respect to $\beta$ is found by differentiating (B.2). This is tedious, but we note that the right hand side of (B.2) is of the form $-A/(A+B)$ and take advantage of this fact. Using this notation,

$$\begin{aligned}\frac{\partial A}{\partial \beta} &= ne^{\hat{\alpha}(\beta)+\beta}(\frac{\partial \hat{\alpha}}{\partial \beta}+1)(1+e^{\hat{\alpha}(\beta)})^2 + 2ne^{\hat{\alpha}(\beta)+\beta}e^{\hat{\alpha}(\beta)}(1+e^{\hat{\alpha}(\beta)})\frac{\partial \hat{\alpha}}{\partial \beta} \\ &= ne^{\hat{\alpha}(\beta)+\beta}(1+e^{\hat{\alpha}(\beta)})\left\{(1+e^{\hat{\alpha}(\beta)})(\frac{\partial \hat{\alpha}}{\partial \beta}+1) + 2e^{\hat{\alpha}(\beta)}\frac{\partial \hat{\alpha}}{\partial \beta}\right\}\end{aligned}$$

and

$$\begin{aligned}\frac{\partial B}{\partial \beta} &= me^{\hat{\alpha}(\beta)}\frac{\partial \hat{\alpha}}{\partial \beta}(1+e^{\hat{\alpha}(\beta)+\beta})^2 + 2me^{\hat{\alpha}(\beta)}e^{\hat{\alpha}(\beta)+\beta}(1+e^{\hat{\alpha}(\beta)+\beta})(\frac{\partial \hat{\alpha}}{\partial \beta}+1) \\ &= me^{\hat{\alpha}(\beta)}(1+e^{\hat{\alpha}(\beta)+\beta})\left\{(1+e^{\hat{\alpha}(\beta)+\beta})\frac{\partial \hat{\alpha}}{\partial \beta} + 2e^{\hat{\alpha}(\beta)+\beta}(\frac{\partial \hat{\alpha}}{\partial \beta}+1)\right\}.\end{aligned}$$

The derivative is thus equal to

$$\frac{\partial^2 \hat{\alpha}}{\partial \beta^2} = \frac{-\frac{\partial A}{\partial \beta} * (A+B) - (-A) * (\frac{\partial B}{\partial \beta} + \frac{\partial A}{\partial \beta})}{(A+B)^2}.$$

We are only interested in this derivative at $\beta = 0$, so we evaluate all the components at $\beta = 0$, and see that

$$A = \frac{ntN^2}{(N-t)^3},$$

$$B = \frac{mtN^2}{(N-t)^3},$$

$$\left.\frac{\partial A}{\partial \beta}\right|_{\beta=0} = \frac{ntN}{(N-t)^3}\left\{\frac{mN-2nt}{N(N-t)}\right\},$$

and

$$\left.\frac{\partial B}{\partial \beta}\right|_{\beta=0} = \frac{mtN}{(N-t)^3}\left\{\frac{-nN+2mt}{N(N-t)}\right\}$$

so that

$$\left.\frac{\partial A}{\partial \beta}\right|_{\beta=0} + \left.\frac{\partial B}{\partial \beta}\right|_{\beta=0} = \frac{2t^2N}{(N-t)^3}(m-n).$$

Combining gives

$$\begin{aligned}\left.\frac{\partial^2\alpha}{\partial\beta^2}\right|_{\beta=0} &= \frac{\frac{t^2N^3}{(N-t)^6}\left\{mn\,(2t-N)\right\}}{\frac{t^2N^6}{(N-t)^6}} \\ &= \frac{mn(2t-N)}{N^3}. \qquad \text{(B.3)}\end{aligned}$$

It is also necessary to derive the third derivative of $\alpha$ with respect to $\beta$ at $\beta = 0$. We evaluate

$$\begin{aligned}\frac{\partial^2 A}{\partial\beta^2} &= ne^{\hat{\alpha}(\beta)+\beta}(1+e^{\hat{\alpha}(\beta)})\left\{(1+e^{\hat{\alpha}(\beta)})(\frac{\partial\hat{\alpha}}{\partial\beta}+1)+2e^{\hat{\alpha}(\beta)}\frac{\partial\hat{\alpha}}{\partial\beta}\right\}(\frac{\partial\hat{\alpha}}{\partial\beta}+1) \\ &+ ne^{\hat{\alpha}(\beta)+\beta}e^{\hat{\alpha}(\beta)}\left\{(1+e^{\hat{\alpha}(\beta)})(\frac{\partial\hat{\alpha}}{\partial\beta}1)+2e^{\hat{\alpha}(\beta)}\frac{\partial\hat{\alpha}}{\partial\beta}\right\}\frac{\partial\hat{\alpha}}{\partial\beta} \\ &+ ne^{\hat{\alpha}(\beta)+\beta}(1+e^{\hat{\alpha}(\beta)})\left\{e^{\hat{\alpha}(\beta)}\frac{\partial\hat{\alpha}}{\partial\beta}(\frac{\partial\hat{\alpha}}{\partial\beta}+1)+(1+e^{\hat{\alpha}(\beta)})\frac{\partial^2\hat{\alpha}}{\partial\beta^2}\right. \\ &+ \left.2e^{\hat{\alpha}(\beta)}\left(\frac{\partial\hat{\alpha}}{\partial\beta}\right)^2+2e^{\hat{\alpha}(\beta)}\frac{\partial^2\hat{\alpha}}{\partial\beta^2}\right\} \qquad \text{(B.4)}\end{aligned}$$

and

$$\begin{aligned}\frac{\partial^2 B}{\partial\beta^2} &= me^{\hat{\alpha}(\beta)}(1+e^{\hat{\alpha}(\beta)+\beta})\left\{(1+e^{\hat{\alpha}(\beta)+\beta})\frac{\partial\hat{\alpha}}{\partial\beta}+2e^{\hat{\alpha}(\beta)+\beta}(\frac{\partial\hat{\alpha}}{\partial\beta}+1)\right\}\frac{\partial\hat{\alpha}}{\partial\beta} \\ &+ me^{\hat{\alpha}(\beta)}e^{\hat{\alpha}(\beta)+\beta}\left\{(1+e^{\hat{\alpha}(\beta)+\beta})\frac{\partial\hat{\alpha}}{\partial\beta}+2e^{\hat{\alpha}(\beta)+\beta}(\frac{\partial\hat{\alpha}}{\partial\beta}+1)\right\}(\frac{\partial\hat{\alpha}}{\partial\beta}+1) \\ &+ me^{\hat{\alpha}(\beta)}(1+e^{\hat{\alpha}(\beta)+\beta})\left\{e^{\hat{\alpha}(\beta)+\beta}\frac{\partial\hat{\alpha}}{\partial\beta}(\frac{\partial\hat{\alpha}}{\partial\beta}+1)+(1+e^{\hat{\alpha}(\beta)+\beta}))\frac{\partial^2\hat{\alpha}}{\partial\beta^2}\right. \\ &+ \left.2e^{\hat{\alpha}(\beta)+\beta}\left(\frac{\partial\hat{\alpha}}{\partial\beta}\right)^2+2e^{\hat{\alpha}(\beta)+\beta}\frac{\partial^2\hat{\alpha}}{\partial\beta^2}\right\}. \qquad \text{(B.5)}\end{aligned}$$

Mathematica software (Wolfram Research Inc. 1993) was used to evaluate the above derivatives at $\beta = 0$ and combine them to derive the third derivative. The third derivative is equal to

$$\begin{aligned}&\frac{\partial^3\hat{\alpha}}{\partial\beta^3} \\ &= \left\{\frac{1}{N^6\,(N-t)}\right\}\bigl\{n\,\bigl(m^2nN^3+12m^2\,N^4-19mnN^4-m^2nN^4+8n^2N^4\end{aligned}$$

$$\begin{aligned}
- & \; m^2N^5 + 2mn^3Nt - 2m^3N^2t + 3m^2nN^2t + 4mn^2N^2t - 2n^3N^2t - 12m^2N^3t \\
+ & \; 113mnN^3t + m^2nN^3t - 2n^2N^3t + m^2N^4t - 2mn^3t^2 + 2m^3Nt^2 - 6m^2Nt^2 \\
- & \; 8mn^2Nt^2 + 2n^3Nt^2 + 2mnN^2t^2 - 6n^2N^2t^2 - 2n^3N^2t^2 - 2n^2N^3t^2 + 4m^2nt^3 \\
+ & \; 4mn^2t^3 + 4mnNt^3 + 4n^3Nt^3 + 4n^2N^2t^3 - 2n^3t^4 - 2n^2Nt^4\Big)\Big\}.
\end{aligned}$$

If we assume that $m$ and $n$ are of the same order of magnitude as $N$, the highest order terms are all equivalent to

$$\frac{nm^2N^5}{N^6(N-t)} \sim N$$

and hence the entire derivative is of order $N$ at $\beta = 0$.

We next calculate the derivatives of the modified profile likelihood correction term in (4.75) with respect to $\beta$ at $\beta = 0$. The multiplier is equal to

$$\frac{1}{2}\frac{N}{N-1}\frac{\partial}{\partial\beta}\log(M(\beta)) \tag{B.6}$$

where

$$M(\beta) = \frac{me^{\hat{\alpha}(\beta)}}{(1+e^{\hat{\alpha}(\beta)})^2} + \frac{ne^{\hat{\alpha}(\beta)+\beta}}{(1+e^{\hat{\alpha}(\beta)+\beta})^2}. \tag{B.7}$$

Using the fact that $e^{\hat{\alpha}}/(1+e^{\hat{\alpha}}) = t/N$ when $\beta = 0$, we see that

$$\begin{aligned}
M(0) &= \frac{mt(N-t)}{N^2} + \frac{nt(N-t)}{N^2} \\
&= \frac{(m+n)t(N-t)}{N^2} \\
&= \frac{t(N-t)}{N}.
\end{aligned} \tag{B.8}$$

Also,

$$\begin{aligned}
\frac{\partial M}{\partial\beta} &= \frac{m(e^{\hat{\alpha}(\beta)}(1+e^{\hat{\alpha}(\beta)})^2\frac{\partial\hat{\alpha}}{\partial\beta} - 2e^{2\hat{\alpha}(\beta)}(1+e^{\hat{\alpha}(\beta)})\frac{\partial\hat{\alpha}}{\partial\beta})}{(1+e^{\hat{\alpha}(\beta)})^4} \\
&+ \frac{n(e^{\hat{\alpha}(\beta)+\beta}(1+e^{\hat{\alpha}(\beta)+\beta})^2(\frac{\partial\hat{\alpha}}{\partial\beta}+1) - (2e^{2(\hat{\alpha}(\beta)+\beta)}(1+e^{\hat{\alpha}(\beta)+\beta})(\frac{\partial\hat{\alpha}}{\partial\beta}+1))}{(1+e^{\hat{\alpha}(\beta)+\beta})^4} \\
&= \frac{me^{\hat{\alpha}(\beta)}(1-e^{\hat{\alpha}(\beta)})\frac{\partial\hat{\alpha}}{\partial\beta}}{(1+e^{\hat{\alpha}(\beta)})^3} + \frac{ne^{\hat{\alpha}(\beta)+\beta}(1-e^{\hat{\alpha}(\beta)+\beta})(\frac{\partial\hat{\alpha}}{\partial\beta}+1)}{(1+e^{\hat{\alpha}(\beta)+\beta})^3}.
\end{aligned} \tag{B.9}$$

At $\beta = 0$, we substitute $t/N - t$ for $e^{\hat{\alpha}(0)}$ and $-n/N$ for $\partial\alpha/\partial\beta$ to get

$$\left.\frac{\partial M}{\partial\beta}\right|_{\beta=0} = \frac{mt(N-2t)(-n)(N-t)}{N^4} + \frac{nt(N-2t)m(N-t)}{N^4} \tag{B.10}$$

which is equal to zero.

To calculate the second derivative of $M$ with respect to $\beta$ we require considerable algebra, and the derivatives of the numerator and denominator of each of the two fractions that make up the right hand side of (B.9). We do this one at a time, for convenience. Let the right hand side of (B.9) be written as $t1/b1 + t2/b2$. We write

$$\frac{\partial t1}{\partial\beta} = me^{\hat{\alpha}(\beta)}\left\{(1-2e^{\hat{\alpha}(\beta)})(\frac{\partial\alpha}{\partial\beta})^2 + (1-e^{\hat{\alpha}(\beta)})\frac{\partial^2\alpha}{\partial\beta^2}\right\}$$

which at $\beta = 0$ is equal to

$$m\frac{t}{N-t}\left\{\frac{n^2(N-3t)}{N^2(N-t)} - \frac{nm(N-2t)^2}{N^3(N-t)}\right\}$$

and

$$\frac{\partial t2}{\partial\beta} = ne^{\hat{\alpha}(\beta)+\beta}\left\{(1-2e^{\hat{\alpha}(\beta)+\beta})(\frac{\partial\alpha}{\partial\beta}+1)^2 + (1-e^{\hat{\alpha}(\beta)+\beta})\frac{\partial^2\alpha}{\partial\beta^2}\right\}$$

which at $\beta = 0$ is equal to

$$n\frac{t}{N-t}\left\{\frac{m^2(N-3t)}{N^2(N-t)} - \frac{nm(N-2t)^2}{N^3(N-t)}\right\}.$$

Further,

$$\frac{\partial b1}{\partial\beta} = 3(1+e^{\hat{\alpha}(\beta)})^2 e^{\hat{\alpha}(\beta)}\frac{\partial\alpha}{\partial\beta}$$

is equal to

$$\frac{-3Nnt}{(N-t)^3}$$

at $\beta = 0$ and

$$\frac{\partial b1}{\partial\beta} = 3(1+e^{\hat{\alpha}(\beta)+\beta})^2 e^{\hat{\alpha}(\beta)+\beta}(\frac{\partial\alpha}{\partial\beta}+1)$$

is equal to

$$\frac{-3Nmt}{(N-t)^3}$$

at $\beta = 0$. For completeness, $t1 = -t2 = -mnt(N-2t)/(n(N-t)^2)$ and $b1 = b2 = (N/(N-t))^3$. Combining,

$$\left.\frac{\partial^2 M}{\partial \beta^2}\right|_{\beta=0} = \frac{-2mnt^2(N-t)^2}{N^5}. \tag{B.11}$$

Under asymptotic model I,

$$\frac{1}{N}\frac{\partial^2}{\partial \beta^2}\log M(\beta) = \frac{1}{N}\left\{\frac{M''(\beta)}{M(\beta)} - \left(\frac{M'(\beta)}{M(\beta)}\right)^2\right\}. \tag{B.12}$$

The first term of the right hand side is

$$\frac{1}{N}\frac{M''(\beta)}{M(\beta)} \tag{B.13}$$

which can be seen to be $o_p(1)$ by virtue of the fact that as $N_k$ increases to infinity, $t1$ and $t2$ are $o_p(n_k)$ and $o_p(m_k)$ respectively, that $b1$ and $b2$ are both $o_p(1)$, and that $M(\beta)$ is $o_p(N_k)$.

To calculate $\partial^3 M/\partial\beta^3$ we require the above derivatives as well as $\partial^2 t1/\partial\beta^2$, $\partial^2 t2/\partial\beta^2$, $\partial^2 b1/\partial\beta^2$, and $\partial^2 b2/\partial\beta^2$. Mathematica Software was again used to assist calculation of these derivatives. First,

$$\frac{\partial^2 b1}{\partial\beta^2} = \frac{-3nt\,(mN - nN - 2mt - nt)}{N(N-t)^3} \tag{B.14}$$

and

$$\frac{\partial^2 b2}{\partial\beta^2} = \frac{3mt\,(mN - nN + mt + 2nt)}{N(N-t)^3} \tag{B.15}$$

Also,

$$\frac{\partial^2 t1}{\partial\beta^2} = \frac{mn^2t\,(mN^2 - nN^2 - 4mNt + 3nNt + 4mt^2)}{N^4(N-t)^2} \tag{B.16}$$

and

$$\frac{\partial^2 t2}{\partial\beta^2} = \frac{m^2nt\,(mN^2 - nN^2 - 3mNt + 4nNt - 4nt^2)}{N^4(N-t)^2}. \tag{B.17}$$

Combining and simplifying gives

$$\left.\frac{\partial^3 M}{\partial\beta^3}\right|_{\beta=0} = \frac{mn\,(m-n)\,t\,(N^4 - 13N^3t + 51N^2t^2 - 69Nt^3 + 30t^4)}{N^7} \tag{B.18}$$

and hence that

$$\left.\frac{\partial^3 M}{\partial\beta^3}\right|_{\beta=0} /(M(\beta)) = \frac{mn\,(m-n)\,(N^4 - 13N^3t + 51N^2t^2 - 69Nt^3 + 30t^4)}{N^6\,(N-t)}. \tag{B.19}$$

Again, for the purposes of asymptotic model I, it is easy to verify that

$$\frac{1}{N}\frac{\partial^3}{\partial\beta^3}\log M(\beta)$$

is $o_p(1)$.

# VITA

Robert Platt was born on September 9, 1968 in Winnipeg, Manitoba. Robert is the oldest son of Craig and Virginia Platt. He graduated from Kelvin High School in Winnipeg, Manitoba in 1986. He earned a Bachelor of Science degree in Mathematics from McGill University, Montreal, Quebec in 1990 and a Master of Science degree in Statistics from the University of Manitoba , Winnipeg, Manitoba in 1993, a Master of Science degree in Biostatistics from the University of Washington, Seattle, Washington in 1994, and has been in the Ph.D. program in Biostatistics at the University of Washington, Seattle, Washington since 1994. Upon graduation Robert will pursue a career as assistant professor in the departments of Biostatistics and Pediatrics at McGill University and the Montreal Children's Hospital Research Institute, Montreal, Quebec.