HIGHER ORDER CONDITIONAL INFERENCE USING PARALLELS WITH APPROXIMATE BAYESIAN TECHNIQUES

BY JUAN ZHANG

A dissertation submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Statistics and Biostatistics

Written under the direction of

John E. Kolassa

and approved by

_____

_____

_____

_____

New Brunswick, New Jersey

October, 2008

ABSTRACT OF THE DISSERTATION

Higher order conditional inference using parallels with approximate Bayesian techniques

by JUAN ZHANG

Dissertation Director: John E. Kolassa

I consider parametric models with a scalar parameter of interest and multiple nuisance parameters. The likelihood ratio statistic is frequently used in statistical inference. The standard normal approximation to the likelihood ratio statistic generally has error of order $O(n^{-1/2})$, where n denotes the sample size. When n is small, the normal approximation may not be adequate to do accurate inference. In practice, the true error is more important than asymptotic order. The intention of this study is to find an approximation which is relatively easy to apply, but which is accurate under small sample size settings. Saddlepoint approximations are well-known for higher order accuracy properties and remarkably good relative error properties. There are several saddlepoint approximations. I look for one that is flexible in application while keeping a satisfactory convergence rate.

I evaluate, via Monte Carlo, the accuracies of several saddlepoint approximations, and of some classical methods, when these approximations are used to approximate p-values for hypotheses about a scalar parameter. Based on the results, I find that DiCiccio and Martin's (1993) approximations are interesting

and deserve more research. Approximations of DiCiccio and Martin (1993) involve exploiting the parallels between Bayesian and frequentist inference, and can be constructed from general log-likelihood functions with relatively easy calculation, while keep the accuracy property.

Two difficulties arise in the application of these approximations. One is the instability around a singularity. The other and far more significant is the construction of the prior density functions utilized in these approximations. These prior density functions are also called matching priors.

To make DiCiccio and Martin's (1993) approximations applicable in practice, I successfully resolve the above two problems. I remove the instability and fix the numerical difficulties in applying these approximations. The matching prior is the solution to a first order partial differential equation. The solution of this equation is non-trivial under the general parametrization. I use a procedure to solve the equation numerically given any initial conditions.

As a conclusion, I suggest the use of DiCiccio and Martin's (1993) approximations with the construction procedure and the correction that I provide in this thesis.

Acknowledgements

I am grateful to my thesis advisor Professor John E. Kolassa. In my first year at Rutgers, I took his course of Biostatistics–I. His intriguing teaching added lots of fun to my study and encouraged me to go further in statistical research. In the following years, when I started to do research with him, I met regularly almost every week, including countless email interactions, discussing statistics. Without his altruistic advice, my thesis can never be done.

I want to thank all the professors who taught and trained me as a statistician. Their intelligent teaching enlightened not only me but also my classmates interested in statistics. Especially, I would like to thank Professor Arthur Cohen and Professor William E. Strawderman for being in my committee. Both of them taught me and played very important roles through out my Ph.D. study. Professor Cohen gave me several chances to teach in college as an instructor and also provided me generous help in my career development. Professor Strawderman gave me patient assistance as one of the editors of *IMS-LNMS* 54, where I had the first part of this thesis published.

I want to express my sincere appreciation to Professor Linda J. Scheetz, College of Nursing, New York University, for the great opportunity of applying statistics she gave me, the helpful assistance in my Ph.D. training and the kind presence in my thesis committee.

I would like to thank my family and friends for their continuous supports.

The State University of New Jersey, for my trips to JSM 2007 at Salt Lake City UT and ENAR 2008 at Washington DC presenting partial work in this thesis.

Dedication

To my mom Wenfeng, my dad Dengdai, my husband Ye and my son Eric.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background

Tests of a simple null hypothesis, using the likelihood ratio statistic, are frequently employed to analyze data arising from a model with a single unknown parameter. The signed square root of the likelihood ratio statistic often has a distribution that is well-approximated by a standard normal distribution under the null hypothesis. In the presence of nuisance parameters, one can also construct a signed square root of the likelihood ratio statistic. Consider continuous random variables $\mathbf{X} = (X_1, \ldots, X_n)$ having joint density function that depends on unknown parameters $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_d)$. Suppose that $\boldsymbol{\omega} = (\psi, \boldsymbol{\lambda})$, where $\psi = \omega_1$ is the scalar parameter of interest and $\boldsymbol{\lambda} = (\omega_2, \ldots, \omega_d)$ is the vector of nuisance parameters. Let $\hat{\boldsymbol{\omega}} = (\hat{\psi}, \hat{\boldsymbol{\lambda}})$ be the maximum likelihood estimator of $\boldsymbol{\omega}$. For fixed ψ, let $\hat{\boldsymbol{\lambda}}_\psi$ be the constrained maximum likelihood estimator of $\boldsymbol{\lambda}$ satisfying

$$l^j(\psi, \hat{\boldsymbol{\lambda}}_\psi) = 0 \;\; \text{for all} \;\; j \geq 2,$$

where $l(\boldsymbol{\omega})$ is the log-likelihood function for $\boldsymbol{\omega}$ and $l^j(\boldsymbol{\omega}) = \dfrac{\partial}{\partial \omega_j} l(\boldsymbol{\omega})$. The likelihood ratio statistic for testing $\psi = \psi_0$ is $W = 2\{l(\hat{\boldsymbol{\omega}}) - l(\psi_0, \hat{\boldsymbol{\lambda}}_0)\}$, where $\hat{\boldsymbol{\lambda}}_0 = \hat{\boldsymbol{\lambda}}_{\psi_0}$. The signed root of the likelihood ratio statistic is

$$R = \text{sgn}(\hat{\psi} - \psi_0)W^{1/2}. \tag{1.1}$$

Denote by n the sample size. The standard normal approximation to the distribution of R typically has error of order $O(n^{-1/2})$, and R can be employed

to construct approximate confidence limits for the parameter of interest having coverage error of $O(n^{-1/2})$. In large sample settings, this approximation works well. However, in small sample situations, with 10 or 15 observations, the standard normal approximation may not be adequate for statistical inference. Hence, various authors developed modifications for R using saddlepoint approximation techniques. These modifications reduce the order of error in the standard normal approximation to the distribution of R.

The intention of this study is, with the parameter of interest as a scalar in the presence of nuisance parameters, to find an approximation that is relatively easy to apply, but which is more accurate when the sample size is small. The saddlepoint approximation is known for its better error properties. Consequently, it may lead to more accurate approximation to the distribution of R. There are several kinds of saddlepoint approximations and I would like to find the one that is flexible in application while keeping satisfactory convergence rate.

1.2   The contribution of this dissertation

The thesis is constituted by three parts. Each of them is innovative and can be reviewed separately. Meanwhile, they are closely related to each other. Integration of them can provide a satisfactory answer that achieves the goal of this study.

1.2.1   A comparison of various approximations

In the first part of the thesis, for p–values of hypothesis about a scalar parameter, I evaluate via Monte Carlo behaviors the accuracies of classical methods and several saddlepoint approximations, such as likelihood ratio test, the Wald test, Barndorff-Nielsen's approximations (1986), Sevirini's empirical adjustment (1999) and DiCiccio and Martin's approximations (1993). The likelihood ratio test and the Wald test are large sample methods, while the other methods under

comparison are saddlepoint approximations, which are well known for being accurate and relatively straightforward to implement (Butler, 2007; Kolassa, 2006; Reid, 2003).

Barndorff-Nielsen's approximation (1986) can have a relative error rate of order $O(n^{-3/2})$. Barndorff-Nielsen (1988, 1990, 1991) also considered using a variation on this approximation, of the same form as the univariate expansion of Lugannani and Rice (1980). Though these approximations are delicate and accurate, their application is restricted by the requirement of the calculation of an exact or approximate ancillary. For some situations it is hard or even impossible to construct this ancillary.

Severini (1999) proposed an approximation based on empirical covariances. Under some assumptions and model regularity properties, his approximation has the relative error rate of $O(n^{-1})$. Though the derivation of this approximation involves the same ancillary statistic as in Barndorff-Nielsen's approximation, its construction does not require the specification of the ancillary. A similar Lugannani and Rice (1980) format is also available for Sevirini's adjustment.

DiCiccio and Martin (1993) introduced approximations, which are also available without the specification of the ancillary. The derivation involves the Bayesian approach to constructing confidence limits considered by Welch and Peers (1963) and Peers (1965). The related prior density functions are called matching priors. The matching prior can be obtained by solving a corresponding first order partial differential equation. With this prior, the approximations of DiCiccio and Martin (1993) have relative error of order $O(n^{-1})$. Cox and Reid (1980) introduce the idea of parameter orthogonality. If the parameter of interest and the nuisance parameters are orthogonal, solving the first order partial differential equation is relatively easy. In some cases, where the parameters are not orthogonal, it may be more difficult to find a numerical solution.

The above mentioned approximations have never been systematically compared to each other as a group. I examine, as an example, the ratio of means of independent exponential distributed variables. I calculate via simulation the size of tests constructed as above, and then compare the results among different approximations. Approximations of DiCiccio and Martin (1993) were developed that reduce the order of error in the standard normal approximation to the distribution of the signed root of the likelihood ratio statistic. However, in our preliminary comparison, DiCiccio and Martin's (1993) approximations do not show their superiority. Furthermore, there are two difficulties in applying their approximations. One is the instability of the tail probability approximations around the conditional mean of the distribution approximated. The other is the numerical construction of the prior density function in the approximations. I find that Di-Ciccio and Martin's approximations are interesting and deserve more research. I will discuss them in detail and provide successful solutions to overcome those difficulties in the balance of the thesis.

1.2.2 Correcting instabilities near the mean

The second part of the thesis concerns the instability of DiCiccio and Martin's (1993) approximations observed from the accuracy comparison study. I provide a correction and remove the instability of the tail probability approximations around the conditional mean of the distribution approximated.

Saddlepoint approximations are useful for complicated survival and hazard functions for progressive diseases, mentioned by Huzurbazar and Huzurbazar (1999). Among examples presented by Huzurbazar and Huzurbazar (1999), some of the approximations that behave very well in the tail of the distribution show instability for ordinates near the conditional mean of the distribution approximated. Such instabilities can force the tail probability approximation out of the range $[0, 1]$, yielding an invalid approximation. This phenomenon also exists for

DiCiccio and Martin's (1993) approximations, as we can observe in the accuracy comparison study.

The approximations of DiCiccio and Martin (1993) can be expressed in Lugannani and Rice format

$$\Phi(R) + \phi(R)(R^{-1} - T^{-1}), \tag{1.2}$$

or in the Barndorff-Nielsen format

$$\Phi\{R + R^{-1}\log(T/R)\}, \tag{1.3}$$

where Φ is the standard normal distribution function, R is the signed root of the likelihood ratio statistic as defined in (1.1), and T is the quantity defined by DiCiccio and Martin (1993), and is given in (2.8).

Instability of the tail probability approximations may occur for ordinates near the conditional mean of the approximated distribution, where both R and T are approaching 0. The instability is not caused simply by computational imprecision. When both $R = 0$, and hence $T = 0$, the quantities $1/R - 1/T$ and $R^{-1}\log(T/R)$ are undefined. Symbolic computation packages, such as Mathematica and Maple, can not overcome the instability difficulties without additional analytical work. For all but the most trivial cases, R and T are calculated using iterative numerical solutions to nonlinear equations. For R near zero, then, the variation in $1/R - 1/T$ is driven primarily by error in the solution to the saddlepoint equation, and not by the arithmetic errors in floating point calculations.

I modify the techniques proposed by Yang and Kolassa (2002), to present a correction to repair numerical difficulties in applying DiCiccio and Martin's (1993) approximations around the conditional mean of the approximated distribution. This correction is necessary in order to calculate p-values. Statisticians routinely approximate probabilities for tails of size roughly .05 or .025. The endpoints of the intervals associated with these probabilities occasionally occur at or near the

conditional mean of the approximated distribution. This situation arises most often when constructing confidence intervals, and a search algorithm is employed to examine various candidates for the confidence interval endpoint. In these situations, formulae like those I present in chapter 3 are useful. The techniques of Yang and Kolassa (2002) and in chapter 3 can be applied similarly to general saddlepoint approximations with both Barndorff-Nielsen format and Lugannani and Rice format to remove the instability.

1.2.3   Numerical construction of matching priors

An open question in the application of matching prior approximations is to the provision of a procedure to find the matching priors. The derivation of DiCiccio and Martin's (1993) approximations uses Bayesian techniques, where a matching prior density function needs to be specified. The derivation of a matching prior relates to a first order partial differential equation. The solution of this first order partial differential equation is non-trivial when the parameters are not orthogonal. I use a procedure to solve the equation numerically given any initial conditions.

Cox and Reid (1987) introduced parameter orthogonality. It is defined with respect to the expected Fisher information matrix. The most direct statistical interpretation of orthogonality is that the relevant components of the score statistic are uncorrelated (Cox and Reid, 1987). In most cases one can parameterize the model so that parameters are orthogonal. However, finding such a transformation may be as difficult as constructing the matching prior without orthogonality. So it is of interest to consider the situation with general parametrization.

Matching priors were first proposed by Welch and Peers (1965) and Peers (1963). These priors have an important role in connecting the frequentist and the Bayesian approach. Under the frequentist paradigm, conditional inference can be complicated, while Bayesian techniques can simplify frequentist elimination of nuisance parameters. Determining a matching prior requires solving a first

order partial differential equation. Only in simple circumstances, such as when parameters are orthogonal, can the first order partial differential equation be solved analytically.

Levine and Casella (2003) proposed a procedure to solve the first order partial differential equation numerically under general parametrization, in models with a single nuisance parameter. They transformed the parameters into another parameter space and solved the first order partial differential equation in this transformed space. Then the procedure requires a transformation back to the original parameter space. Typically, the transformation between the two parameter spaces is nontrivial. Hence this procedure may be difficult to apply in practice. Also, no instructions on initial condition specification were given. Though the solution to the first order partial differential equation is not unique, it is unique when the initial conditions are given. Changes in the initial conditions lead to different choices of the matching prior, which can affect data analysis. Finding an appropriate prior may improve the inference results significantly, as we should indicate below in section 4.5.

I present a more practical way to solve for the matching priors, without the involvement of the back transformation described in Levine and Casella (2003). Levine and Casella (2003) did not give instructions on initial conditions. The procedure presented in this thesis can be suitable to all kinds of initial conditions. I apply this technique to a logistic regression model with a binary response and one explanatory variable. In this case, the two parameters are generally not orthogonal. I solve matching priors using the procedure presented and the approximations of DiCiccio and Martin (1993) to calculate p-values and confidence intervals for the unknown effect of the explanatory variable. The methods introduced in section 1.2.2 and addressed more thoroughly in chapter 3 are used to remove the singularity, so it is no longer an obstacle in the numerical computation. Then by choosing differential initial conditions one is able to improve the

performances of DiCiccio and Martin's approximations.

It should be pointed out that, the matching prior solving procedure presented in chapter 4 has many applications in addition to being used in DiCiccio and Martin's approximation.

Numerical examples have been examined throughout my work. I use R software in performing all related analyses. R is a system for statistical computation and graphics. It is free software distributed under a GNU-style copyleft, and an official part of the GNU project.

1.3   Outline

In chapter 2, I present an accuracy comparison through a numerical example. In chapter 3, I develop a way to remove the effect of the singularity around the mean of the distribution approximated. The concept of matching priors and a procedure of finding the matching prior are reviewed in chapter 4. Finally, conclusion is drawn in chapter 5. Some related proofs and derivations of formulas are given in the appendix.

Chapter 2

A Comparison of the Accuracy of Saddlepoint Conditional Cumulative Distribution Function Approximations

2.1 Introduction

For a model with a single unknown parameter, the signed square root of the likelihood ratio statistic R defined in (1.1) often has a distribution that is well-approximated by a standard normal distribution under the null hypothesis. In the presence of nuisance parameters, the distribution of the statistic R depends on the nuisance parameters. In large sample settings, the standard normal approximation works well. However, in small sample situations, with 10 or 15 observations, the standard normal approximation of R may not be adequate. Hence, various authors have developed modifications for R using saddlepoint approximation techniques. These modifications reduce the order of error in the standard normal approximation to the conditional distribution of R.

Barndorff-Nielsen (1986) proposed the modified directed signed root of the likelihood ratio statistic R^*. This statistic will be reviewed in the next section. The relative error in the standard normal approximation to the conditional distribution of R^* is of order $O(n^{-3/2})$. Barndorff-Nielsen (1988, 1990, 1991) also considered using a variation of this approximation, of the same form as the univariate expansion of Lugannani and Rice (1980). The drawback of these approximations is that they require the calculation of an exact or approximate ancillary, which in some situations is hard or impossible to construct. For the other approximations

studied in the following, no such ancillary needs to be specified, and hence the approximations are easier to apply in practice.

Severini (1999) proposed an approximation $\hat{R}^*$ to Barndorff-Nielsen's R^* based on empirical covariances. Under some assumptions and model regularity properties, $\hat{R}^*$ is distributed according to a standard normal distribution, with relative error $O_p(n^{-1})$, conditionally on the observed value of an ancillary statistic $\mathbf{A}$. However, the construction of this $\hat{R}^*$ does not require the specification of $\mathbf{A}$.

DiCiccio and Martin (1993) proposed an alternative quantity to R^*, denoted by R^+, that is also available without specification of $\mathbf{A}$. The derivation of R^+ involves the Bayesian approach to constructing confidence limits considered by Welch and Peers (1963) and Peers (1965). In the presence of nuisance parameters, Peers (1965) chose a prior density for the parameters to satisfy a partial differential equation. With this prior, the standard normal approximation to the conditional distribution of R^+ has relative error of order $O(n^{-1})$. If the parameter of interest and the nuisance parameter vector are orthogonal, solving the partial differential equation is easier and explicit solutions can be obtained. In the case that the parameters are not orthogonal, solving that equation numerically is problematic. Parameter orthogonality will be reviewed in the following section.

For a parameter of interest that is orthogonal to the nuisance parameter vector, Cox and Reid (1987) defined the signed root of the conditional likelihood ratio statistic $\overline{R}$. The standard normal approximation to the distribution of $\overline{R}$ has error of order $O(n^{-1/2})$. DiCiccio and Martin (1993) defined $\overline{R}^+$ similarly to the R^+ described above. The quantity $\overline{R}^+$ is available without specification of $\mathbf{A}$. The standard normal approximation to the distribution of $\overline{R}^+$ has relative error of order $O(n^{-1})$, both conditionally (on $\mathbf{A}$) and unconditionally. The use of $\overline{R}$ and its modifications is often effective in situations where there are many nuisance parameters. However, in such cases, the use of R and its modified versions can produce unsatisfactory results. DiCiccio, Field and Fraser (1990) present

examples.

The above variants on R have never been systematically compared to each other as a group. This chapter provides an accuracy comparison among the modifications stated above. Each of these approximations are used to generate an approximate one-sided p-value by approximating $P[R \geq r]$, for r the observed value of R. Approximate two-sided p-values are calculated by approximating $2\min(P[R \geq r], P[R < r])$. One and two-sided hypotheses tests of size α may be constructed by rejecting the null hypothesis when the p-value is less than α. Both the Barndorff-Nielson format approximation

$$\Phi\{R + R^{-1}\log(U/R)\} \tag{2.1}$$

and the Lugannani and Rice format approximation

$$\Phi(R) + \phi(R)(R^{-1} - U^{-1}) \tag{2.2}$$

are considered, where the variable U may vary for different modifications. I will examine, as an example, the ratio of means of independent exponentials. I calculate via simulation, the size of tests constructed as above, and compare the results among different approximations.

2.2   Methods compared

We first review several statistics whose marginal distributions are very close to standard normal.

Suppose $X_1, \ldots, X_n$ are independently identically distributed continuous random variables having joint density function that depends on unknown parameters $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_d)$. Suppose that $\boldsymbol{\omega} = (\psi, \boldsymbol{\lambda})$, where $\psi = \omega_1$ is a scalar parameter of interest and $\boldsymbol{\lambda} = (\omega_2, \ldots, \omega_d)$ is a nuisance parameter vector. Let $\hat{\boldsymbol{\omega}} = (\hat{\psi}, \hat{\boldsymbol{\lambda}})$ be the maximum likelihood estimator of $\boldsymbol{\omega}$, and for fixed ψ, let $\hat{\boldsymbol{\lambda}}_\psi$ be the constrained maximum likelihood estimator of $\boldsymbol{\lambda}$. The signed root likelihood ratio statistic R

is defined in (1.1). The standard normal approximation to the distribution of R typically has error of order $O(n^{-1/2})$, and R can be used to construct approximate confidence limits for ψ having coverage error of that order.

A form of conditional saddlepoint tail probability approximation was provided by Skovgaard (1987), who applied double saddlepoint techniques to the problem of approximating tail probabilities for conditional distributions when the data arise from a full exponential family. In this case the double saddlepoint distribution function approximation can be expressed in terms of the quantities in the joint density function. Skovgaard's double saddlepoint approximation to the conditional distribution function is of form (2.2), with U a Wald statistic. Here, we consider only models more complicated than canonical exponential families, and so will not apply this approximation.

2.2.1 Barndorff-Nielsen's modification

Barndorff-Nielsen (1986) proposed the modified signed root of the likelihood ratio statistic R^*, given by

$$R^* = R + R^{-1}\log(U/R), \tag{2.3}$$

where

$$U = \frac{|l_{\boldsymbol{\lambda};\hat{\boldsymbol{\omega}}}(\hat{\boldsymbol{\omega}}_\psi) \quad l_{;\hat{\boldsymbol{\omega}}}(\hat{\boldsymbol{\omega}}) - l_{;\hat{\boldsymbol{\omega}}}(\hat{\boldsymbol{\omega}}_\psi)|}{|j_{\boldsymbol{\lambda\lambda}}(\hat{\boldsymbol{\omega}}_\psi)|^{\frac{1}{2}}|j(\hat{\boldsymbol{\omega}})|^{\frac{1}{2}}}, \tag{2.4}$$

$$\hat{\boldsymbol{\omega}}_\psi = (\psi_0, \hat{\boldsymbol{\lambda}}_0),$$

and

$$j_{\boldsymbol{\lambda\lambda}}(\hat{\boldsymbol{\omega}}_\psi) = -l_{\boldsymbol{\lambda\lambda}}(\psi_0, \hat{\boldsymbol{\lambda}}_0), j(\hat{\boldsymbol{\omega}}) = -l_{\boldsymbol{\omega\omega}}(\hat{\boldsymbol{\omega}})$$

with $l_{\boldsymbol{\omega\omega}}(\boldsymbol{\omega})$ the matrix of second-order partial derivatives of $l(\boldsymbol{\omega}; \hat{\boldsymbol{\omega}}, \mathbf{A})$ taken with respect to $\boldsymbol{\omega}$, and $l_{\boldsymbol{\lambda\lambda}}(\boldsymbol{\omega})$ the submatrix of $l_{\boldsymbol{\omega\omega}}(\boldsymbol{\omega})$ corresponding to $\boldsymbol{\lambda}$. Here U represents an approximate conditional score statistic, which, in the multivariate normal case would exactly coincide with R. Outside the multivariate normal case, the correction to R in (2.3) measures the difference between R and U, and

hence is a measure of departure from normality. The quantity $l_{;\hat{\omega}}(\boldsymbol{\omega})$ is the $d \times 1$ vector of partial derivatives of $l(\boldsymbol{\omega}; \hat{\boldsymbol{\omega}}, \mathbf{A})$ taken with respect to $\hat{\boldsymbol{\omega}}$, and $l_{\boldsymbol{\lambda};\hat{\omega}}(\boldsymbol{\omega})$ is a $d \times (d-1)$ matrix of mixed second-order partial derivatives of $l(\psi, \boldsymbol{\lambda}; \hat{\boldsymbol{\omega}}, \mathbf{A})$ taken with respect to $\boldsymbol{\lambda}$ and $\hat{\boldsymbol{\omega}}$. The sign of U is the same as that of R and the resulting U is of the form

$$U = R + O_p(n^{-1/2}).$$

The relative error in the standard normal approximation to the conditional distribution of R^* is of order $O(n^{-3/2})$. The conditioning is on an exact or approximate ancillary statistic $\mathbf{A}$. The variable U is parametrization invariant and does not depend on $\boldsymbol{\lambda}$.

The value of ψ_0 satisfying $\Phi(R^*) = \alpha$ is an approximate upper $1-\alpha$ confidence limit which has relative coverage error of order $O(n^{-3/2})$ both conditionally and unconditionally. Barndorff-Nielsen (1988, 1990, 1991) also considered using the alternative to $\Phi(R^*)$ provided by the Lugannani and Rice format approximation (2.2).

Consider the exponential family model for a random vector $\mathbf{T}$ whose density evaluated at $\mathbf{t}$ is

$$f_{\mathbf{T}}(\mathbf{t}; \boldsymbol{\theta}) = \exp(\boldsymbol{\theta}^\top \mathbf{t} - \mathcal{H}_{\mathbf{T}}(\boldsymbol{\theta}) - \mathcal{G}(\mathbf{t})).$$

The random vector $\mathbf{T}$ is the sufficient statistic. Set $\tau(\boldsymbol{\theta}) = \mathrm{E}_{\boldsymbol{\theta}}[\mathbf{T}]$. In the presence of nuisance parameters, the calculation of U requires the specification of the ancillary $\mathbf{A}$. Barndorff-Nielsen (1980) suggested an approximate ancillary statistic for use in conditional inference. Kolassa (2006), in §8.4, presented this approximate ancillary $\mathbf{A}$ as

$$\mathbf{B}(\hat{\psi})(\mathbf{T} - \boldsymbol{\tau}(\hat{\psi}, \boldsymbol{\lambda}))^\top,$$

with $\boldsymbol{\lambda}$ held fixed, and

$$\mathbf{B}(\psi) = \left[\left(\frac{\partial \boldsymbol{\tau}}{\partial \psi} \right)^\perp \boldsymbol{\Sigma} \left(\frac{\partial \boldsymbol{\tau}}{\partial \psi} \right)^{\perp\top} \right]^{-\frac{1}{2}} \left(\frac{\partial \boldsymbol{\tau}}{\partial \psi} \right)^\perp,$$

where the superscript $^\perp$ represents the corresponding orthogonal vector. Suppose that $\boldsymbol{\theta}$ is scalar. Let

$$\tilde{l}(\theta; \hat{\theta}, \mathbf{a}) = \frac{l(\theta; \hat{\theta}, \mathbf{a})}{n}.$$

Then

$$F_{\hat{\Theta}|\mathbf{A}}(\hat{\theta}|\mathbf{a}; \theta) = \left\{ \Phi(\sqrt{n}\hat{w}) + \frac{\phi(\sqrt{n}\hat{w})}{\sqrt{n}} \left[\frac{1}{\hat{w}} - \frac{1}{\check{z}} \right] \right\} [1 + O_p(n^{-1})], \qquad (2.5)$$

with

$$\hat{w} = \frac{R}{\sqrt{n}}, \quad \check{z} = \frac{\tilde{l}^{;1}(\hat{\theta}; \hat{\theta}, \mathbf{a}) - \tilde{l}^{;1}(\theta; \hat{\theta}, \mathbf{a})}{\sqrt{j(\hat{\theta})}},$$

and the superscripts $;1$ on $\tilde{l}^{;1}$ represent differentiation of the likelihood with respect to $\hat{\theta}$, after reexpressing $\mathbf{t}$ in terms of $\hat{\theta}$ and $\mathbf{a}$. Here $\mathbf{a}$ is the observed value of $\mathbf{A}$, $F_{\hat{\Theta}|\mathbf{A}}(\hat{\theta}|\mathbf{a}; \theta)$ is the conditional cumulative distribution function, and $\Phi(\cdot)$ is the standard normal cumulative distribution function.

In the computation of Barndorff-Nielsen's R^*, the calculation of U requires the ancillary $\mathbf{A}$ to be specified, which may present difficulties in practice. In the following, several modifications will be reviewed that do not require the specification of $\mathbf{A}$.

2.2.2 An empirical adjustment

Severini (1999) proposed an approximation $\hat{R}^*$ to Barndorff-Nielsen's R^* based on empirical covariances. Recalling the formula of U (2.4), the key step is to approximate $l_{\boldsymbol{\lambda};\hat{\boldsymbol{\omega}}}(\hat{\boldsymbol{\omega}}_\psi)$ and $l_{;\hat{\boldsymbol{\omega}}}(\hat{\boldsymbol{\omega}}) - l_{;\hat{\boldsymbol{\omega}}}(\hat{\boldsymbol{\omega}}_\psi)$

Let $l^{(j)}(\boldsymbol{\omega})$ denote the log-likelihood function based on observation j alone. Denote

$$\hat{Q}(\boldsymbol{\omega}; \boldsymbol{\omega}_0) = \sum l^{(j)}(\boldsymbol{\omega}) l_{\boldsymbol{\omega}}^{(j)}(\boldsymbol{\omega}_0)^\top,$$

$$\hat{I}(\boldsymbol{\omega}; \boldsymbol{\omega}_0) = \sum l_{\boldsymbol{\omega}}^{(j)}(\boldsymbol{\omega}) l_{\boldsymbol{\omega}}^{(j)}(\boldsymbol{\omega}_0)^\top,$$

and

$$\hat{i} = \hat{I}(\hat{\boldsymbol{\omega}}; \hat{\boldsymbol{\omega}}).$$

The quantity $\boldsymbol{\omega}_0$ is any point in the parameter space. Then $l_{;\hat{\boldsymbol{\omega}}}(\hat{\boldsymbol{\omega}}) - l_{;\hat{\boldsymbol{\omega}}}(\boldsymbol{\omega})$ and $l_{\boldsymbol{\omega};\hat{\boldsymbol{\omega}}}(\boldsymbol{\omega})$ may be approximated by

$$\hat{l}_{;\hat{\boldsymbol{\omega}}}(\hat{\boldsymbol{\omega}}) - \hat{l}_{;\hat{\boldsymbol{\omega}}}(\boldsymbol{\omega}) = \{\hat{Q}(\hat{\boldsymbol{\omega}};\hat{\boldsymbol{\omega}}) - \hat{Q}(\boldsymbol{\omega};\hat{\boldsymbol{\omega}})\}\hat{i}(\hat{\boldsymbol{\omega}})^{-1}\hat{j}$$

and $\hat{l}_{\boldsymbol{\omega};\hat{\boldsymbol{\omega}}}(\boldsymbol{\omega}) = \hat{I}(\boldsymbol{\omega};\hat{\boldsymbol{\omega}})\hat{i}(\hat{\boldsymbol{\omega}})^{-1}\hat{j}$, where $\hat{j} = j(\hat{\boldsymbol{\omega}}) = -l_{\boldsymbol{\omega}\boldsymbol{\omega}}(\hat{\boldsymbol{\omega}})$.

Denote by $\hat{U}$ the approximation to the statistic U based on the above quantities, and then denote

$$\hat{R}^* = R + R^{-1}\log(\hat{U}/R).$$

The quantity $\hat{R}^*$ can be used in approximation (2.1). This represents a correction similar to that of (2.4), with expectations of quantities replaced by sample means. Under some assumptions plus model regularity properties, $\hat{R}^*$ is distributed according to a standard normal distribution, with error $O_p(n^{-1})$, conditionally on **a**, the observed value of the ancillary **A**. However, the construction of $\hat{R}^*$ does not require the specification of **A**. Again, the alternative approximation (2.2) is also available as $\Phi(R) + \phi(R)(R^{-1} - \hat{U}^{-1})$.

2.2.3 DiCiccio and Martin's modification

DiCiccio and Martin (1993) proposed an alternative variable to U, denoted by T, which is available without specification of the ancillary **A**. The modification for approximation (2.1) is

$$R^+ = R + R^{-1}\log(T/R), \tag{2.6}$$

where T is defined in (2.8). As with (2.4), the final term in R^+ represents the departure from normality; unlike (2.4), this measure represents the departure of the posterior of ψ from normality, and involves the prior distribution. Once again, one might use the alternative probability approximation (2.2) with T substituting the place of U. The replacement of T avoids the necessity of specifying **A** in calculating U and hence simplifies the calculations. The derivation of T involves

the Bayesian approach to constructing confidence limits considered by Welch and Peers (1963) and Peers (1965). When $\omega = \psi$, that is, when the entire parameter is scalar and there are no nuisance parameters, Welch and Peers (1963) showed that the appropriate choice is

$$\pi(\omega) \propto \{i(\omega)\}^{1/2},$$

the Jeffrey's prior, where

$$i(\omega) = \mathrm{E}\{-\,\mathrm{d}^2\,l(\omega)/\,\mathrm{d}\,\omega^2\}.$$

In the presence of nuisance parameters, Peers (1965) showed that $\pi(\boldsymbol{\omega})$ must be chosen to satisfy the partial differential equation

$$\sum_{j=1}^{d} i^{1j}(i^{11})^{-1/2}\frac{\partial}{\partial\omega^j}(\log\pi) + \sum_{j=1}^{d}\frac{\partial}{\partial\omega^j}\{i^{1j}(i^{11})^{-1/2}\} = 0, \qquad (2.7)$$

where

$$i_{jk}(\boldsymbol{\omega}) = \mathrm{E}\{-\partial^2 l(\boldsymbol{\omega})/\partial\omega^j\partial\omega^k\}$$

and (i^{jk}) is the $d \times d$ matrix inverse of (i_{jk}). The variable T is defined as

$$T = l_\psi(\psi_0, \hat{\boldsymbol{\lambda}}_0)\frac{|-l_{\boldsymbol{\lambda\lambda}}(\psi_0, \hat{\boldsymbol{\lambda}}_0)|^{1/2}\pi(\hat{\boldsymbol{\omega}})}{|-l_{\boldsymbol{\omega\omega}}(\hat{\boldsymbol{\omega}})|^{1/2}\pi(\psi_0, \hat{\boldsymbol{\lambda}}_0)}. \qquad (2.8)$$

Here $l_\psi(\boldsymbol{\omega}) = \partial l(\boldsymbol{\omega})/\partial\psi$, and $\pi(\boldsymbol{\omega})$ is a proper prior density for $\boldsymbol{\omega} = (\psi, \boldsymbol{\lambda})$ which satisfies the equation (2.7). Then the resulting approximation (2.2) is

$$\mathrm{P}(\psi \geq \psi_0 | X) = \Phi(R) + (R^{-1} - T^{-1})\phi(R) + O(n^{-3/2}),$$

where $T = U + O_p(n^{-1})$, and thus the approximation (2.1) to the conditional distribution of R given $\mathbf{A}$ based on $R + R^{-1}\log(T/R)$ has error of order $O(n^{-1})$. That is to say, DiCiccio and Martin's approximation (1993) approximates Barndorff-Nielsen's approximation (1980) of the distribution of R, conditionally given $\mathbf{A}$ as well as unconditionally. To error of the order $O_p(n^{-1})$, T is parametrization invariant under transformations $\boldsymbol{\omega} \mapsto \{\psi, \tau(\boldsymbol{\omega})\}$.

Parameter orthogonality makes solving the partial differential equation (2.7) easier. Orthogonality was introduced by Cox and Reid (1980). It is defined with respect to the expected Fisher information matrix. Define $\boldsymbol{\theta}_1$ to be orthogonal to $\boldsymbol{\theta}_2$ if the elements of the information matrix satisfy

$$i_{\theta_s \theta_t} = \frac{1}{n} \, \mathrm{E}\left(\frac{\partial l}{\partial \theta_s} \frac{\partial l}{\partial \theta_t}; \boldsymbol{\theta}\right) = \frac{1}{n} \, \mathrm{E}\left(-\frac{\partial^2 l}{\partial \theta_s \partial \theta_t}; \boldsymbol{\theta}\right) = 0 \qquad (2.9)$$

for $s = 1, \ldots, p_1, t = p_1 + 1, \ldots, p_1 + p_2$, where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$; $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are of length p_1 and p_2 respectively. If equation (2.9) is to hold for all $\boldsymbol{\theta}$ in the parameter space, then the parametrization is sometimes called globally orthogonal. If (2.9) holds at only one parameter value $\boldsymbol{\theta}_0$, then the vectors $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are said to be locally orthogonal at $\boldsymbol{\theta}_0$. The most direct statistical interpretation of (2.9) is that the relevant components of the score statistic are uncorrelated.

The definition of orthogonality can be extended to more than two sets of parameters, and in particular $\boldsymbol{\theta}$ is totally orthogonal if the information matrix is diagonal. In general, it is not possible to have total parameter orthogonality at all parameter values, but it is possible to obtain orthogonality of a scalar parameter of interest ψ to a set of nuisance parameters. If the parameter of interest and the nuisance parameter vector are orthogonal, solving the partial differential equation (2.7) is easier. The equation (2.7) reduces to

$$(i_{\psi\psi})^{-1/2} \frac{\partial}{\partial \psi}(\log \pi) + \frac{\partial}{\partial \psi}(i_{\psi\psi})^{-1/2} = 0, \qquad (2.10)$$

whose solutions are of the form (Tibshirani, 1989)

$$\pi(\psi, \boldsymbol{\lambda}) \propto \{i_{\psi\psi}(\psi, \boldsymbol{\lambda})\}^{1/2} g(\boldsymbol{\lambda}), \qquad (2.11)$$

where $g(\boldsymbol{\lambda})$ is arbitrary and the suggestive notation $i_{\psi\psi}(\psi, \boldsymbol{\lambda})$ is used in place of $i_{11}(\psi, \boldsymbol{\lambda})$. In some cases in which the parameters are not orthogonal, solving equation (2.7) numerically is problematic.

2.2.4 Conditional likelihood ratio statistic and its modification

For ψ and $\boldsymbol{\lambda}$ orthogonal, Cox and Reid (1987) defined the conditional likelihood ratio statistic for testing $\psi = \psi_0$ as

$$\overline{W} = 2\{\bar{l}(\bar{\psi}) - \bar{l}(\psi_0)\},$$

where $\bar{l}(\psi_0)$ is the conditional profile likelihood for ψ using $\hat{\boldsymbol{\lambda}}_\psi$ as the conditioning statistic and is defined as following.

$$\bar{l}(\psi) = l(\psi, \boldsymbol{\lambda}_\psi) - \frac{1}{2}\log|-l_{\boldsymbol{\lambda}\boldsymbol{\lambda}}(\psi, \hat{\boldsymbol{\lambda}}_\psi)|$$

and $\bar{\psi}$ is the point at which the function $\bar{l}(\psi)$ is maximized. The signed root of the conditional likelihood ratio statistic is

$$\overline{R} = \mathrm{sgn}(\bar{\psi} - \psi_0)\overline{W}^{1/2},$$

and the standard normal approximation to the distribution of $\overline{R}$ has error of order $O(n^{-1/2})$. Let

$$\overline{R}^+ = \overline{R} + \overline{R}^{\,-1}\log(\overline{T}/\overline{R}).$$

One may use approximations (2.1) and (2.2), say, $\Phi(\overline{R}^+)$ or $\Phi(\overline{R}) + \phi(\overline{R})(\overline{R}^{-1} - \overline{T}^{-1})$, where

$$\overline{T} = \bar{l}^{(1)}(\psi_0)\{-\bar{l}^{(2)}(\bar{\psi})\}^{-1/2}\frac{\pi(\bar{\psi}, \lambda_{\bar{\psi}})}{\pi(\psi_0, \lambda_0)},$$

and

$$\bar{l}^{(j)} = \frac{\mathrm{d}^j\,\bar{l}(\psi)}{\mathrm{d}\,\psi^j}, \quad j = 1, 2.$$

Those approximations have errors of order $O(n^{-1})$.

The use of $\overline{R}$ and its modifications is often effective in situations where there are many nuisance parameters. However, in such cases, the use of R and its modified versions can produce unsatisfactory results; see DiCiccio, Field and Fraser (1990) for examples.

2.3 Example: exponential samples with orthogonal interest and nuisance parameters

2.3.1 Settings

Let X and Y be exponential random variables with means μ and ν respectively; the ratio of the means ν/μ is the parameter of interest. The parameter transformation

$$\left\{\mu \to \frac{\lambda}{\sqrt{\psi}}, \ \nu \to \lambda\sqrt{\psi}\right\}$$

makes the two new parameters ψ and λ orthogonal. Then X and Y have expectations $\lambda\psi^{-\frac{1}{2}}$ and $\lambda\psi^{\frac{1}{2}}$, respectively.

Suppose there are n independent replications of (X, Y). Denote $\boldsymbol{\omega} = (\psi, \lambda)$. one can obtain the log-likelihood function as

$$l(\boldsymbol{\omega}) = -n\left[\frac{\psi\bar{x} + \bar{y}}{\lambda\sqrt{\psi}} + 2\log\lambda\right].$$

Each of the approximations in section 2 may be used to generate an approximate one-sided p-value by approximating $\mathrm{P}[R \geq r]$, for r the observed value of R. Approximate two-sided p-values may be calculated by approximating $2\min(\mathrm{P}[R \geq r], \mathrm{P}[R < r])$. One and two-sided hypotheses tests of size α may be constructed by rejecting the null hypothesis when the p-value is less than α. Both the Barndorff-Nielson format approximation (2.1) and the Lugannani and Rice format approximation (2.2) were considered. I calculate via simulation the size of tests constructed as above, and compare the results among different approximations.

Some of the approximations in section 2 require specific algebraic calculations. I present the related calculations in appendix A. Other applications are generic, and no specific algebraic calculations are needed.

2.3.2 Simulation results

Simulation procedure

For sample size $n = 10$,

(1) Generate 10 draws from the pair of $\{X, Y\}$, where X and Y both follow standard exponential distribution;

(2) Calculate one and two-sided p-values for each approximation;

(3) Compare the p-values to the α level, say 0.05; denote by q the number of miss coverages; if the p-value is less than 0.05, then $q = q + 1$;

(4) Repeat step (1) - (3) for s times and report the final value of q; let $q^* = (q/s) * 100$, the Type I error probability in percentage.

Approximations (2.1) and (2.2) have a removable singularity at $R = 0$. Consequently, these and similar formulae require care when evaluating near $R = 0$. Specifically, I find that (2.1) and (2.2) exhibit adequate numerical stability as long as $|R| > 10^{-4}$. Out of 1,000,000 simulated data sets, 60 presented R (or a modification of R) closer to zero. In these cases, for all but the most extreme conditioning events, the resulting conditional p-value is large enough as to not imply rejection of the null hypothesis, and so these simulated data sets were treated as not implying rejection of the null hypothesis. For more dedicated consideration, please refer to the paper of Yang and Kolassa (2002) and chapter 3 of this dissertation, where I give a linear approximation to DiCiccio and Martin's approximations to overcome the difficulties of the instability of the tail probability when the coordinates near the conditional mean of the approximated distribution.

Results

Table 2.1 and table 2.2 below report the Type I error probabilities (in percentage) of $s = 1,000,000$ rounds simulation. The quantities T_u and $\overline{T}_u$ are assumed with uniform prior densities.

Table 2.1: Type I error probability in percentage (BN)

Approximation	1-sided	2-sided
$\Phi(R)$	5.241	5.168
$\Phi(\overline{R})$	4.807	4.575
$\Phi(R + R^{-1}\log(U/R))$	5.046	4.760
$\Phi(R + R^{-1}\log(\hat{U}/R))$	5.018	4.882
$\Phi(R + R^{-1}\log(T/R))$	4.615	4.312
$\Phi(R + R^{-1}\log(T_u/R))$	11.017	6.828
$\Phi(\overline{R} + \overline{R}^{-1}\log(\overline{T}/\overline{R}))$	4.883	4.411
$\Phi(\overline{R} + \overline{R}^{-1}\log(\overline{T}_u/\overline{R}))$	11.723	7.249

From the simulation results in table 2.1 and table 2.2, one can see that for both the Barndorff-Nielsen format approximation (BN) and the Lugannani and Rice format approximation (LR), the empirical adjustment has best performance. Barndorff-Nielsen's modification has the best asymptotic error rate ($O_p(n^{-3/2})$ rather than $O_p(n^{-1})$), and hence one might expect that the best performance from this approximation. Instead one can observe the best performance from other modifications with worse asymptotic error.

Table 2.2: Type I error probability in percentage (LR)

Approximation	1-sided	2-sided
$\Phi(R)$	5.241	5.168
$\Phi(\overline{R})$	4.807	4.575
$\Phi(R) + \phi(R)(R^{-1} - U^{-1})$	5.046	4.760
$\Phi(R) + \phi(R)(R^{-1} - \hat{U}^{-1})$	5.017	4.881
$\Phi(R) + \phi(R)(R^{-1} - T^{-1})$	4.613	4.308
$\Phi(R) + \phi(R)(R^{-1} - T_u^{-1})$	11.274	6.943
$\Phi(\overline{R}) + \phi(\overline{R})(\overline{R}^{-1} - \overline{T}^{-1})$	4.878	4.403
$\Phi(\overline{R}) + \phi(\overline{R})(\overline{R}^{-1} - \overline{T}_u^{-1})$	12.190	7.510

One may also notice that the performance of DiCiccio and Martin's modification is not as good as expected. The importance of the choice of prior can be demonstrated by the poor performance of the approximations with the incorrect uniform priors. In appendix B, for the calculation of T, the prior is chosen as

$$\pi(\psi, \boldsymbol{\lambda}) = \frac{\sqrt{n}}{\sqrt{2}\psi}.$$

Recall that ψ and $\boldsymbol{\lambda}$ are orthogonal. The simplified partial differential equation (2.10) can be used with solution (2.11). In chapter 4, we will revisit the choice of prior and extend the case of orthogonal parametrization to general parametrization.

2.4 Research problems in the application of DiCiccio and Martin's approximations

The unexpected relatively worse performance, and some observed challenges in the application of DiCiccio and Martin's approximations, invoke our interest to do more research about them. From a theoretical point of view, DiCiccio and Martin's approximations can be superior to classical large sample methods and some saddlepoint approximations, since they have relatively higher convergence rate and avoid conditioning on an exact or approximate ancillary. However, when applying DiCiccio and Martin's approximations, two problems may restrict their use, as we observe from the comparison simulation process.

The first problem is the instability of the approximations for ordinates near the conditional mean of the distribution approximated. Such instabilities can force the tail probability approximation out of the range $[0, 1]$, yielding an invalid approximation. In chapter 3, I remove the instability and fix the numerical difficulties in applying DiCiccio and Martin's approximations.

The second problem is construction of the prior $\pi(\boldsymbol{\omega})$ and the choice of $\pi(\boldsymbol{\omega})$. In the example of the ratio of two exponential means, I transform the parameters to be orthogonal. In this case, the situation is relatively easy to handle. However, in general parametrization, finding solution to the partial differential equation that defines $\pi(\boldsymbol{\omega})$ is nontrivial. Also, based on our experience, different choices of $\pi(\boldsymbol{\omega})$ can give various performances. Then the choice of $\pi(\boldsymbol{\omega})$ is worth of study. However, before determine which prior to choose, one has to be able to obtain

these priors. In chapter 4, I successfully solve the problem of solving for $\pi(\boldsymbol{\omega})$ in general parametrization and give some suggestions in how to choose a good $\pi(\boldsymbol{\omega})$.

Among all the saddlepoint approximations considered in the accuracy comparison, we will put our interest on DiCiccio and Martin's approximations. Chapter 3 and chapter 4 provide ways in how to work with the above two problems. One can see that once the above two problems solved, DiCiccio and Martin's approximation can perform much better. It deserves to point out that, as first introduced in Yang and Kolassa (2002), the techniques used in chapter 3 can be applied analogically to saddlepoint approximations with Barndorff-Nielsen format and Lugannani and Rice format. Also, the matching prior solving procedure introduced in chapter 4 can be applied to many situations that require the specification of a matching prior, while the application in DiCiccio and Martin's approximations is only one of them.

Chapter 3

DiCiccio and Martin's Approximations for the Distribution of the Signed Root of Likelihood Ratio Statistic Near the Mean

3.1   Introduction

In this chapter, I remove the instability and fix numerical difficulties in applying DiCiccio and Martin's approximations (1993), for ordinates near the mean of the distribution approximated, both conditionally and unconditionally.

As a recall, DiCiccio and Martin (1993) proposed two approximations, which are

$$\Phi(R) + \phi(R)(R^{-1} - T^{-1}), \tag{3.1}$$

and

$$\Phi\{R + R^{-1}\log(T/R)\}, \tag{3.2}$$

where R is as above and T is defined in (2.8).

Instability of the tail probability may occur for the ordinates near the mean of the approximated distribution (conditionally or unconditionally), where both R and T are approaching 0. The instability is not caused simply by computational imprecision. When both $R = 0$, and hence $T = 0$, then $1/R - 1/T$ and $R^{-1}\log(T/R)$ are undefined. Therefore, symbolic computation packages such as Mathematica and Maple, where user can carry hundreds of digits floating point computation, can not overcome the instability difficulties without additional analytical work. For all but the most trivial cases, R and T are calculated using

interactive numerical solutions to nonlinear equations. For R near zero, then, the variation in $1/R - 1/T$ is driven primarily by error in the solutions to the saddle-point equations, and not by the arithmetic errors in floating point calculations. I will not attempt to address this instability through interpolation, since results would then depend heavily on points and methods used in interpolation.

In this chapter, one will see that the singularity for $1/R - 1/T$ or $R^{-1} \log(T/R)$ is removable. I present a linear approximation to repair numerical difficulties in applying (3.1) and (3.2). For (3.2), it always takes values within the range $[0, 1]$. Although even in the presence of instabilities (3.2) remains within $[0, 1]$, approximation (3.2) still has problems near the mean (both conditionally and unconditionally), since $\log(T/R)$ has a removable singularity at 0. Therefore, our correction is necessary in order to calculate significance tests. Statisticians routinely approximate probabilities for tails of size roughly .05 or .025. The endpoints of the intervals associated with these probabilities periodically occur at or near the mean of the approximated distribution. This situation arises most often when constructing confidence intervals, and a search algorithm is employed to examine various candidates for the confidence interval endpoint. In these situations, formulae like those I present are very useful. Some such situations are described in detail in section 3.3.

In section 3.2, I present a correction for the values of DiCiccio and Martin's approximation around the singularity. In section 3.3, examples are discussed.

3.2   Linear approximation near the mean

With $\mathbf{X}$ and $\boldsymbol{\lambda}$ fixed, R is a smooth invertible function of ψ. Hence $1/R - 1/T$ is a function of T. When applying equation (3.1) for certain values of ψ near the mean, R and T both take on the value 0, and hence $1/R - 1/T$ can not be evaluated naively. Since $T = 0$ if and only if $R = 0$, and $\dfrac{\mathrm{d}\,T}{\mathrm{d}\,R} \neq 0$, $1/R - 1/T$

may be approximated as a linear function in R, with error of size $O(R)$.

In order to determine this approximation, derivatives of $\boldsymbol{\omega}$ as a function of ψ_0 are required. In the following, the superscripts of l and $\boldsymbol{\omega}$ denote partial derivatives with respect to the corresponding components of the arguments. The subscript of ω denotes the coordinate whose derivative is taken.

For fixed ψ, let $\hat{\boldsymbol{\lambda}}_\psi$ be the constrained maximum likelihood estimator of $\boldsymbol{\lambda}$, satisfying

$$l^j(\psi, \hat{\lambda}_\psi) = 0 \quad \text{for all} \quad j \geq 2. \tag{3.3}$$

Let $\tilde{\boldsymbol{\omega}} = (\psi_0, \hat{\boldsymbol{\lambda}}_0)$ satisfy (3.3). Express $R(\psi_0) = \sqrt{2h(\psi_0)}$ for $h(\psi_0) = l(\hat{\boldsymbol{\omega}}) - l(\tilde{\boldsymbol{\omega}})$. Express $T(\psi_0) = b(\psi_0)[g(\psi_0)]^{1/2}$, where

$$b(\psi_0) = l_\psi(\tilde{\boldsymbol{\omega}}),$$

and

$$g(\psi_0) = \frac{|-l_{\boldsymbol{\lambda}\boldsymbol{\lambda}}(\tilde{\boldsymbol{\omega}})| \, \pi(\hat{\boldsymbol{\omega}})^2}{|-l_{\boldsymbol{\omega}\boldsymbol{\omega}}(\hat{\boldsymbol{\omega}})| \pi(\tilde{\boldsymbol{\omega}})^2}.$$

Expand $1/R(\psi_0) - 1/T(\psi_0)$ in ψ_0 about $\hat{\psi}$. Later one will see that

$$b(\hat{\psi}) = -h^1(\hat{\psi}) = 0, b^1(\hat{\psi}) = -h^{11}(\hat{\psi}), b^{11} = -h^{111}(\hat{\psi}), b^{111} = -h^{1111}(\hat{\psi}).$$

Also, noticing that $R(\psi_0)$ can be approximated by

$$-(\psi_0 - \hat{\psi})\sqrt{h^{11}(\hat{\psi})},$$

Replace $(\psi_0 - \hat{\psi})$ by

$$-[h^{11}(\hat{\psi})]^{-1/2}R(\psi_0).$$

Then we obtain the expansion of $1/R(\psi_0) - 1/T(\psi_0)$ in $R(\psi_0)$ about 0 as

$$\frac{1}{R(\psi_0)} - \frac{1}{T(\psi_0)} = -\frac{1}{2}g^1(\hat{\psi})[h^{11}(\hat{\psi})]^{1/2} - \frac{1}{3}h^{111}(\hat{\psi})[h^{11}(\hat{\psi})]^{-3/2}$$

$$- \{5[h^{11}(\hat{\psi})]^{-3}[h^{111}(\hat{\psi})]^2 + 6g^1(\hat{\psi})[h^{11}(\hat{\psi})]^{-1}h^{111}(\hat{\psi})$$

$$+ 9h^{11}(\hat{\psi})[g^1(\hat{\psi})]^2 - 3[h^{11}(\hat{\psi})]^{-2}h^{1111}(\hat{\psi})$$

$$- 6g^{11}(\hat{\psi})\}R(\psi_0)/24 + O([R(\psi_0)]^2). \tag{3.4}$$

One needs to evaluate (3.4) by calculating each term in the right hand side of the equation. Calculation details can be found in appendix B.

When no nuisance parameters are present, all the above can be simplified. Take a constant prior density, for example. In this case,

$$T(\psi_0) = \frac{l_\psi(\psi_0)}{[l_{\psi\psi}(\hat{\psi})]^{1/2}}.$$

Then

$$g(\psi_0) = [l_{\psi\psi}(\hat{\psi})]^{-1}, g'(\hat{\psi}) = 0, g''(\hat{\psi}) = 0.$$

Also, one can easily get

$$h'(\hat{\psi}) = 0, h''(\hat{\psi}) = -l''(\hat{\psi}), h^{(3)}(\hat{\psi}) = -l^{(3)}(\hat{\psi}), h^{(4)}(\hat{\psi}) = -l^{(4)}(\hat{\psi}).$$

Equation (3.4) reduces to

$$\frac{1}{R(\psi_0)} - \frac{1}{T(\psi_0)} = -\frac{1}{3}h^{(3)}(\hat{\psi})[h''(\hat{\psi})]^{-3/2} - \{5[h''(\hat{\psi})]^{-3}[h^{(3)}(\hat{\psi})]^2$$
$$- 3[h''(\hat{\psi})]^{-2}h^{(4)}(\hat{\psi})\}R(\psi_0)/24 + O([R(\psi_0)]^2). \qquad (3.5)$$

The foregoing corrections apply to (3.1). As we mentioned above, for (3.2), $R^{-1}\log(T/R)$ may not be evaluated when R and T both go to 0. We will not need another set of computation to remove the singularity in (3.2). Instead, I claim that the correction for (3.1) is also valid for (3.2). Noticing that $R \to 0$ implies $T \to 0$, and $\lim_{R \to 0} \frac{T}{R} = 1$, we have

$$\lim_{R \to 0} \frac{\frac{1}{R}\log(\frac{T}{R})}{\frac{1}{R} - \frac{1}{T}} = \lim_{R \to 0} \frac{\log T - \log R}{T - R}T = \lim_{R \to 0} \frac{T}{R} = 1.$$

Therefore, the correction for (3.2) is the same as what we have done for (3.1).

3.3 Examples

3.3.1 AIDS/HIV infection – convolution of exponential and gamma distributions

AIDS is a kind of progressive disease. It can be classified as five states of disease progression.

State 1: HIV-infected antibody negative.

State 2: HIV-infected antibody positive.

State 3: Pre-AIDS symptoms.

State 4: Full blown AIDS.

State 5: Death.

Waiting time in each state until transition to the next state can be modeled with a separate parametric distribution. A common assumption is that the waiting times are independent. The total waiting time from state 1 to state 3, which measures the waiting time to pre-AIDS symptoms, is always of great interest.

I present calculations specific to the example provided by Huzurbazar and Huzurbarzar (1999), who model the waiting time to pre-AIDS symptoms. The model was originally considered by Longini et al. (1989). Huzurbazar and Huzurbazar assumed that the first transition time has a distribution well-approximated by an exponential, and the second has a distribution well-approximated by a gamma. Denote the variable $X_1 \sim \exp(\lambda)$ with mean $1/\lambda$, and the variable $X_2 \sim \Gamma(\alpha, \beta)$ with mean α/β. Scale parameter, and the shape parameter, are the fitted values as reported by Huzurbazar and Huzurbazar (1999). To be more specific, $\lambda = 0.0348$, $\alpha = 3.490$ and $\beta = 0.214$. This example was also discussed by Yang and Kolassa (2002). Exact calculations of this nature with a non-integer shape parameter are quite difficult. Of interest is the calculation of percentage points of this distribution, including points near the center of the distribution.

The cumulant generating function for X is

$$K_X(\psi) = \log(\lambda) - \log(\lambda - \psi) + \alpha \log(\beta) - \alpha \log(\beta - \psi).$$

Then the log-likelihood function for X is

$$l(\psi;\ X) = \psi X - [\log(\lambda) - \log(\lambda - \psi) + \alpha \log(\beta) - \alpha \log(\beta - \psi)].$$

Figure 1. Comparison of modified and unmodified approximation, X=45.044.
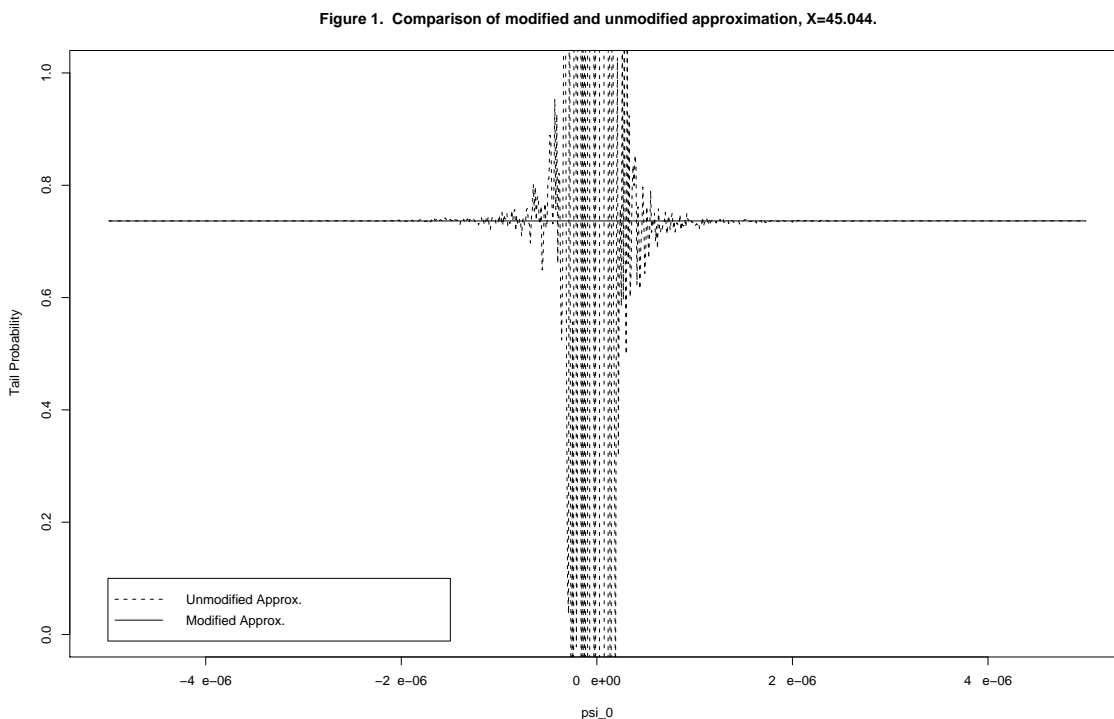
Figure 3.1: Figure 1

Let us take $X = 45.044$, which is close to the mean, and then the maximum likelihood estimator of ψ is very close to 0. This example features no nuisance parameters, and the equation (3.5) for the no nuisance parameter case discussed in section 3.2 can be used. For the uniform prior and the prior of the form $\pi(\psi) \propto \{i(\psi)\}^{1/2}$, approximation (3.2) and the one with correction (3.5) are presented in figure 3.1 and figure 3.2, respectively. Apparently in both cases, approximation (3.5) fixes the instability of the approximation (3.2) in a neighborhood of $\psi_0 = 0$,

Figure 2. Comparison of modified and unmodified approximation, X=45.044.
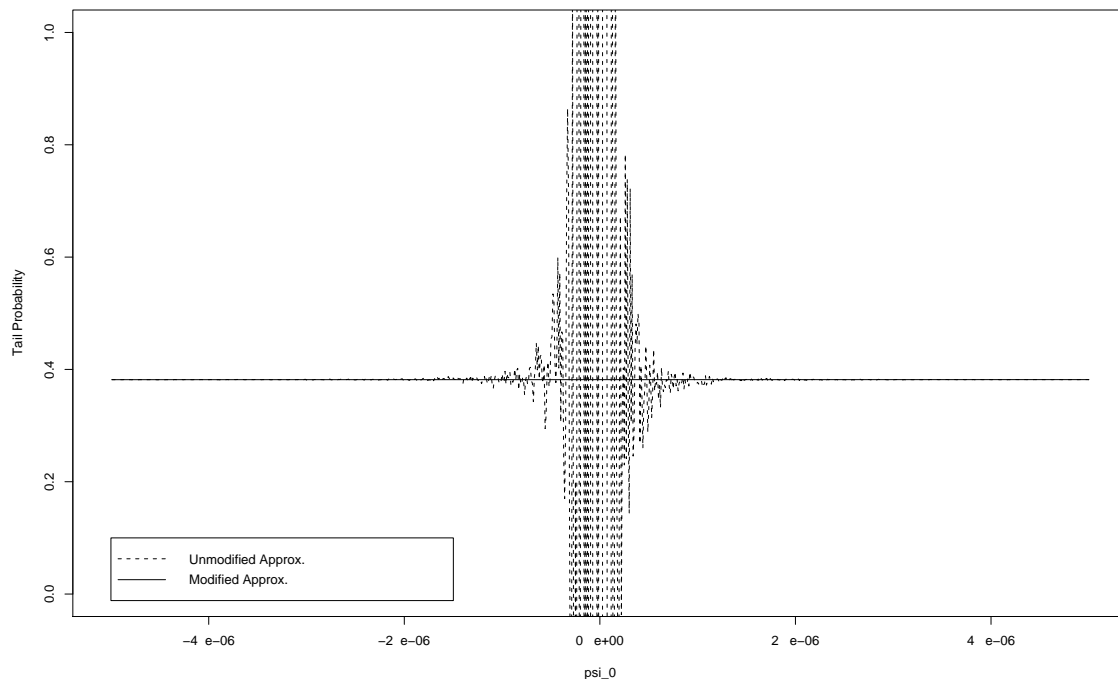
Figure 3.2: Figure 2

at which both T and R go to zero. Figure 3.3 shows the quantile function for the distribution in figure 3.1; a circle indicates the region in which instabilities in the unmodified approximation make inversion of the tail probability to obtain the quantile function impossible.

3.3.2   AIDS/HIV infection – confidence interval

With definitions as in example 3.3.1, let's take $\alpha = 0.04471, \beta = 1$ and $\lambda = 10$. At the place where $X = 0.14471$, the maximum likelihood estimator of ψ is almost 0. Under these settings, the tail probabilities without adjustment are obviously incorrect, which are negative infinity by (3.1) and 0 by (3.2). However, using the method specified in this paper, the corrected tail probability is 0.975, which corresponds to one of the endpoints of a 95% confidence interval. This simple example shows clearly the usefulness of our method in constructing confidence

Figure 3. Quantile Plot: Comparison of modified and unmodified approximation, X=45.044.
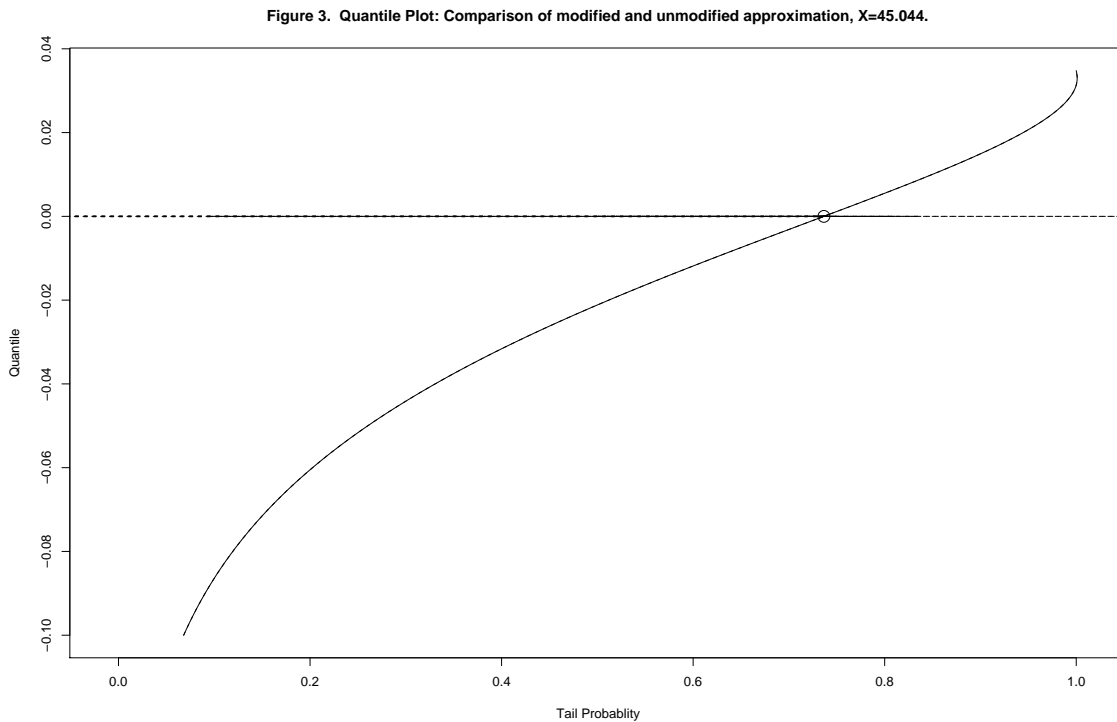
Figure 3.3: Figure 3

intervals.

3.3.3 Ratio of two exponential means

Let Y_1 and Y_2 be two exponential random variables with means $\lambda\psi^{-1/2}$ and $\lambda\psi^{1/2}$, respectively; the parameter of interest is the ratio of the means ψ, and λ is a nuisance parameter. Suppose one has n independent replications of (Y_1, Y_2), i.e. $(Y_{11}, \ldots, Y_{1n};\ Y_{21}, \ldots, Y_{2n})$. Then the log-likelihood function is

$$l = -2n \log \lambda - \lambda^{-1}\psi^{1/2} \sum_{i=1}^{n} Y_{1i} - \lambda^{-1}\psi^{-1/2} \sum_{i=1}^{n} Y_{2i}.$$

The maximum likelihood estimator of ψ is

$$\hat{\psi} = \frac{\sum_{i=1}^{n} Y_{2i}}{\sum_{i=1}^{n} Y_{1i}},$$

where both T and R are close to 0. For a simulated data set from the model described above with $n = 10$, $\sum_{i=1}^{n} Y_{1i} = 8.2773$, and $\sum_{i=1}^{n} Y_{1i} = 6.7485$, then

Figure 4. Comparison of modified and unmodified approximation

Figure 3.4: Figure 4

$\hat{\psi} = 0.8153$. The approximation (3.1) is unstable around that value. Taking a uniform prior, one can generate figure 3.4. Again, in the case in the presence of nuisance parameter, modification (3.4) fixes the instability of (3.1) around the singularity.

Chapter 4

A Practical Procedure to Find Matching Priors for Frequentist Inference

4.1 Introduction
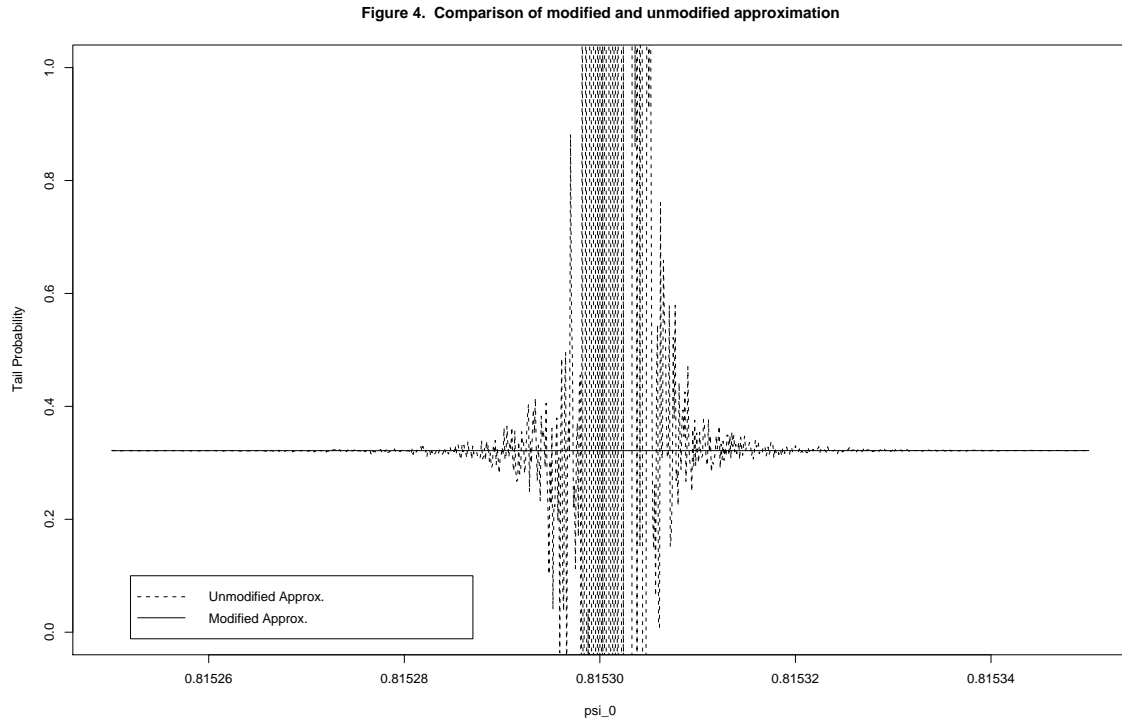
The prior density function $\pi(\cdot)$, satisfying (2.7) and used in (2.8) is known as a matching prior. We consider inference on a single scalar parameter in the presence of nuisance parameters. Under the frequentist paradigm, conditional inference can be complicated. Bayesian method can simplify frequentist elimination of nuisance parameters. Matching priors, first proposed by Welch and Peers (1963) and Peers (1965), can connect frequentist and Bayesian approaches. Determining a matching prior is equivalent to finding a solution of a first order partial differential equation. Only in simple circumstances, such as when parameters are orthogonal, can the partial differential equation be solved analytically. Levine and Casella (2003) note that "Unfortunately, except for these cases, the solution of the resulting partial differential equations becomes quite a hurdle; our only hope is to find numerical solutions to these partial differential equation."

In this chapter, we will see a practical way to solve for the matching priors, without the involvement of the back transformation described by Levine and Casella (2003). This procedure is easy to understand, can be implemented in R (R Development Core Team, 2007) and is suitable to all kinds of initial conditions.

The implementation of matching priors for the approximations proposed by DiCiccio and Martin (1993) is less complicated than other frequentist methods. DiCiccio and Martin's approximations are saddlepoint approximations that make

use of Bayesian–frequentist parallels. Our proposed implementation requires less computational effort compared to the iterative Metropolis-Hasting algorithm described by Levine and Casella (2003).

I end the introduction with a brief outline of this chapter. In section 4.2, I review the concepts of matching priors and discuss the circumstance when orthogonal parameters are presence. Existing analytical and numerical solutions are reviewed. In section 4.3, I present the procedure for solving matching priors, both analytically and numerically. Specification of initial condition is discussed. I also provide information of R software implementation of the solving procedure. The application of using matching priors conjuncted with DiCiccio and Martin's approximations is illustrated through examples in section 4.5. Different initial conditions are specified for obtaining various matching priors.

4.2   Matching priors

I consider parametric models with random variables $X_1, \ldots, X_n$ having joint density function that depends on the unknown parameter vector $\boldsymbol{\omega}$. Suppose $\boldsymbol{\omega}$ is of length d and $\boldsymbol{\omega} = (\omega^1, \omega^2, \ldots, \omega^d) = (\psi, \boldsymbol{\lambda})$ with $\psi = \omega^1$, the parameter of interest, and the nuisance parameter $\boldsymbol{\lambda} = (\omega^2, \ldots, \omega^d)$.

Matching priors were introduced by Welch and Peers (1963) and Peers (1965). In the following, denote the matching prior by $\pi(\cdot)$. Let $P_\pi(\cdot|X)$ be the posterior probability measure for ψ under prior $\pi(\cdot)$. The upper $(1 - \alpha)$ posterior quantile constructed on the basis of a prior density function $\pi(\psi)$ has the property that it is also the frequentist limit, such that

$$P_\pi\{\psi \le \psi^{(1-\alpha)}(\pi, X)|X\} = P_\psi\{\psi \le \psi^{(1-\alpha)}(\pi, X)\} = 1 - \alpha + O(n^{-1}).$$

When there are no nuisance parameters, Welch and Peers (1963) showed that the appropriate choice of $\pi(\omega)$ is $\pi(\omega) \propto \{i(\omega)\}^{1/2}$, where $i(\omega) = \mathrm{E}\{-\mathrm{d}^2\, l(\omega)/\mathrm{d}\,\omega^2\}$,

and $l(\cdot)$ is the log-likelihood function. In this case, matching priors can be easily obtained.

In the presence of nuisance parameters, Peers (1965) showed that $\pi(\boldsymbol{\omega})$ must be chosen to satisfy the partial differential equation

$$\sum_{j=1}^{d} i^{1j}(i^{11})^{-1/2}\frac{\partial}{\partial \omega^j}(\log \pi) + \sum_{j=1}^{d}\frac{\partial}{\partial \omega^j}\{i^{1j}(i^{11})^{-1/2}\} = 0, \qquad (4.1)$$

where

$$i_{jk}(\boldsymbol{\omega}) = \mathrm{E}\{-\partial^2 l(\boldsymbol{\omega})/\partial \omega^j \partial \omega^k\}$$

and (i^{jk}) is the $d \times d$ inverse matrix of (i_{jk}).

If the parameter of interest and the nuisance parameter vector are orthogonal, solving the partial differential equation (4.1) is relatively easy. Parameter orthogonality was introduced by Cox and Reid (1987) as we have reviewed in section 2.2.3. Orthogonality is defined with respect to the expected Fisher information matrix. The most direct statistical interpretation of parameter orthogonality is that the relevant components of the score statistic are uncorrelated. In general, it is possible to obtain orthogonality of a scalar parameter of interest to a set of nuisance parameters.

When the parameter of interest ψ is orthogonal to a set of nuisance parameters, equation (4.1) reduces to

$$(i_{\psi\psi})^{-1/2}\frac{\partial}{\partial \psi}(\log \pi) + \frac{\partial}{\partial \psi}(i_{\psi\psi})^{-1/2} = 0. \qquad (4.2)$$

Tibshirani (1989) showed that solutions were of the form

$$\pi(\psi, \boldsymbol{\lambda}) \propto \{i_{\psi\psi}(\psi, \boldsymbol{\lambda})\}^{1/2}g(\boldsymbol{\lambda}),$$

where $g(\boldsymbol{\lambda})$ is arbitrary, and the suggestive notation $i_{\psi\psi}(\psi, \boldsymbol{\lambda})$ is used in place of $i_{11}(\psi, \boldsymbol{\lambda})$.

However, choosing a parametrization to achieve parameter orthogonality is not always easy, and it can be hard in some cases. It is equivalently hard to obtain orthogonalization and to solve the partial differential equation (4.1) directly,

since the orthogonalization procedure also requires solutions to partial differential equations of form similar to (4.1). Staicu and Reid (2007) studied the use of matching priors with the approximations of DiCiccio and Martin (1993) under orthogonal parametrization, and showed that the Peers-Tibshirani class of matching priors is essentially unique. One can modify the arguments in this chapter to solve the partial differential equation that defines the orthogonality transformation, and attempt, using orthogonality, to narrow down the class of matching priors.

Levine and Casella (2003) propose a general procedure to solve the partial differential equation (4.1) numerically, in models with a single nuisance parameter. Firstly, they transform the parameters into another parameter space, solve the equation, and then transform back to the original parameter space. The numerical application of this procedure is not necessarily easy, and usually the transformation between the two parameter spaces is nontrivial. Levine and Casella (2003) implemented their procedure using Mathematica. They did not give instructions on initial condition specification, which is a necessary component to give specific solution in solving the partial differential equation. Sweeting (2005) introduced data-dependent priors that locally approximate the matching priors, and his procedure can deal with vector nuisance parameters.

4.3 Solving for matching priors with one nuisance parameter

In this section, a procedure to solve the partial differential equation (4.1) is given in general parametrization with one nuisance parameter. First, analytical form of the solutions is given, and then practical notes will be presented later in this section.

In the case that $d = 2$, equation (4.1) is reduced to

$$a(\psi, \lambda)z_\psi + b(\psi, \lambda)z_\lambda = d(\psi, \lambda), \tag{4.3}$$

where

$$z(\psi, \lambda) = \log\{\pi(\psi, \lambda)\},$$

$$a(\psi, \lambda) = \{i^{11}(\psi, \lambda)\}^{1/2},$$

$$b(\psi, \lambda) = i^{12}(\psi, \lambda)\{i^{11}(\psi, \lambda)\}^{-1/2},$$

and

$$d(\psi, \lambda) = -\left[\frac{\partial}{\partial\psi}\{i^{11}(\psi, \lambda)\}^{1/2} + \frac{\partial}{\partial\lambda}\{i^{12}(\psi, \lambda)\}\{i^{11}(\psi, \lambda)\}^{-1/2}\right].$$

The coefficient $a(\psi, \lambda)$ is a diagonal element of the inverse matrix of (i_{jk}), so $a(\psi, \lambda)$ can not be zero. Dividing both sides of (4.3) by $a(\psi, \lambda)$, one has

$$z_\psi + \frac{b(\psi, \lambda)}{a(\psi, \lambda)}z_\lambda = \frac{d(\psi, \lambda)}{a(\psi, \lambda)}.$$

This forces the coefficient of z_ψ to be 1, which simplifies the procedure of finding a solution.

To solve the equation (4.1), it suffices to solve the following ordinary differential equations system

$$\frac{\mathrm{d}\,\psi}{\mathrm{d}\,s} = 1, \frac{\mathrm{d}\,\lambda}{\mathrm{d}\,s} = \frac{b(\psi, \lambda)}{a(\psi, \lambda)}, \frac{\mathrm{d}\,z}{\mathrm{d}\,s} = \frac{d(\psi, \lambda)}{a(\psi, \lambda)}. \tag{4.4}$$

To be more specific with the solution, let us consider the initial conditions prescribed along an initial curve I. Suppose that I is given parametrically, in terms of a parameter ξ, as

$$\psi = \Psi(\xi), \quad \lambda = \Lambda(\xi).$$

Then evaluating $z(\psi, \lambda)$ at a point on I is equivalent to expressing z as a function of ξ,

$$z = Z(\xi) = z\{\Psi(\xi), \Lambda(\xi)\}. \tag{4.5}$$

Here, it is obvious to see that I can not be tangent to the direction $\left[1, \dfrac{b\{\Psi(\xi), \Lambda(\xi)\}}{a\{\Psi(\xi), \Lambda(\xi)\}}\right]$.

One then obtains

$$\psi = \psi(s, \xi), \quad \lambda = \lambda(s, \xi)$$

by simultaneously integrating the two equations defined by

$$\frac{d\psi}{ds} = 1, \quad \psi(s_0, \xi) = \Psi(\xi), \tag{4.6}$$

$$\frac{d\lambda}{ds} = \frac{b(\psi, \lambda)}{a(\psi, \lambda)}, \quad \lambda(s_0, \xi) = \Lambda(\xi). \tag{4.7}$$

From the third equation in (4.4), the initial condition is given by (4.5). Then one has,

$$\frac{dz}{ds} = \frac{d(\psi, \lambda)}{a(\psi, \lambda)}, \quad z(s_0, \xi) = Z(\xi). \tag{4.8}$$

Equation (4.8) can be integrated by quadrature, once equations (4.6) and (4.7) have been solved,

$$z(s, \xi) = Z(\xi) + \int_{s_0}^{s} \frac{d\{\psi(s', \xi), \lambda(s', \xi)\}}{a\{\psi(s', \xi), \lambda(s', \xi)\}} \, ds'. \tag{4.9}$$

These generate a surface in three dimensions, $Z(\psi, \lambda)$, that satisfies both the equation (4.3) and the initial condition. When there are no close form solutions for equations (4.6),(4.7) and (4.8), numerical solutions can be achieved. Rhee et al. (1986) presents more mathematical details.

In obtaining the solution formula (4.9) of $z(s, \xi)$, I avoid doing back transformation as described by Levine and Casella (2003). Noticing that if one wants to specify the value of a matching prior at a certain point, say (ψ^*, λ^*), one can directly specify s as ψ^* and ξ as λ^* in formula (4.9), and then the matching prior evaluated at (ψ^*, λ^*) can be achieved.

Without loss of generality, set the initial condition

$$\{\Psi(\xi), \Lambda(\xi), Z(\xi)\} = (0, \xi, -1).$$

With $\Psi(\xi) = 0$, one has $\psi = s$. The equations (4.7) and (4.8) can be simplified

as

$$\frac{\mathrm{d}\lambda}{\mathrm{d}s} = \frac{b(s,\lambda)}{a(s,\lambda)}, \quad \lambda(s_0,\xi) = \Lambda(\xi), \tag{4.10}$$

$$\frac{\mathrm{d}z}{\mathrm{d}s} = \frac{d(s,\lambda)}{a(s,\lambda)}, \quad z(s_0,\xi) = Z(\xi).$$

We will use R package `odesolve` by Setzer (2007) to solve equation (4.10) and get a numerical expression of $\lambda(\cdot)$ in s. The command `lsoda()` in `odesolve` package is designed to solve initial value problems for stiff or non-stiff systems of first order ordinary differential equations. It provides an interface to the Fortran ordinary differential equation solver of the same name, written by Hindmarsh (1983) and Petzold (1983). For (4.9), I do numerical integration using Simpson's Rule and employed the R function `sintegral()` in the `Bolstad` package by Curran (2005). Suppose z will be evaluated at (ψ^*, λ^*). Noticing that $\Lambda(\xi) = \xi$, choose the starting value as λ^* in solving (4.10), and then choose the upper integration limit as ψ^* in (4.9). The procedure is easy to perform if one has an ordinary differential equation solver, even if not using the solver provided by R package `odesolve`.

Based on the ordinary differential equation (4.6),

$$\psi = s + \Psi(\xi),$$

i.e. $s = \psi - \Psi(\xi)$. So s_0 must be chosen considering the range of ψ. If one chooses $\Psi(\xi) = 0$, then $\psi = s$. For the example in section 4.5.1, the parameter ψ is the ratio of two exponential means, and hence $\psi > 0$. Therefore, s_0 should be chosen as any positive value.

In the above the initial values is chosen as $\{\Psi(\xi), \Lambda(\xi), Z(\xi)\} = (0, \xi, -1)$. Now I will show that the numerical solving procedure is suitable to any initial values.

• Suppose the initial condition for the ordinary differential equation (4.7) is $\lambda(s_0, \xi) = \Lambda(\xi)$, for $\Lambda(\xi)$ an arbitrary known function rather than $\Lambda(\xi) = \xi$

as above. The solution formula of z is the same as stated in (4.9). When solving (4.7), the initial value should be chosen as $\Lambda(\lambda^*)$, no longer λ^*, if z is evaluated at (ψ^*, λ^*).

• If the initial condition of (4.6) is $\psi(s_0, \xi) = \Psi(\xi)$, then the solution from the equation (4.6) is $\psi = s + \Psi(\xi)$. Therefore, the equation (4.7) becomes,

$$\frac{d\lambda}{ds} = \frac{b\{s + \Psi(\xi), \lambda\}}{a\{s + \Psi(\xi), \lambda\}}.$$

Let $\tilde{s} = s + \Psi(\xi)$. By simple change of variables, (4.7) becomes

$$\frac{d\lambda}{d\tilde{s}} = \frac{b(\tilde{s}, \lambda)}{a(\tilde{s}, \lambda)}.$$

Equation (4.8) is

$$\frac{dz}{d\tilde{s}} = \frac{d[\psi\{\tilde{s} - \Psi(\xi), \xi\}, \lambda\{\tilde{s} - \Psi(\xi), \xi\}]}{a[\psi\{\tilde{s} - \Psi(\xi), \xi\}, \lambda\{\tilde{s} - \Psi(\xi), \xi\}]}$$

with $z\{\tilde{s}_0 - \Psi(\xi), \xi\} = Z(\xi)$, noticing that $\tilde{s}_0 = s_0 + \Psi(\xi)$. Then the solution of z is simply given by the following formula,

$$z(\tilde{s}, \xi) = Z(\xi) + \int_{s_0 - \Psi(\xi)}^{\tilde{s} - \Psi(\xi)} \frac{d\{\psi(s', \xi), \lambda(s', \xi)\}}{a\{\psi(s', \xi), \lambda(s', \xi)\}} \, ds'. \tag{4.11}$$

That is to say, the value of the prior on a certain point with the initial condition $\psi(s_0, \xi) = \Psi(\xi)$, is obtained by translating the interval of integration when $\Psi(\xi) = 0$ by $\Psi(\xi)$.

• Suppose the initial condition for (4.8) is $z(s_0, \xi) = Z(\xi)$ and $Z(\cdot)$ is a known function. This case is even simpler to deal with. One only needs to plug the value of $Z(\xi)$ into (4.9).

Therefore, the suggested numerical solving procedure is suitable to any initial values.

4.4 Solving for matching priors with multiple nuisance parameters

In section 4.3, the case of 2 model parameters is considered, i.e., both the parameter of interest and the nuisance parameter are scalars. With dimension 2, it is relatively easy to understand the first order partial differential equation solving procedure from the geometric point of view, since one can draw the initial conditions and the solution surface in a 3-dimensional space.

In this section, the solving procedure will be extended to higher dimension, while keeping the parameter of interest as a scalar. The procedure of the higher dimension is similar as the one of 2-dimensional model parameters. Once again, for more mathematical details about the solving procedure of a first order partial differential equation, please refer to Rhee et al. (1986).

Assume the dimension of the model parameters is d and write vectors as row vectors. Suppose the model parameter vector is $\boldsymbol{\omega} = (\psi, \lambda_1, \cdots, \lambda_{d-1})$, where ψ is the parameter of interest and $\lambda_1, \cdots, \lambda_{d-1}$ are the nuisance parameters. For the convenience of future description, denote $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_{d-1})$. The first order partial differential equation (4.1), which determines the matching prior density function, can be restated as

$$a(\psi, \boldsymbol{\lambda})z_\psi + \sum_{j=2}^{d} b_{j-1}(\psi, \boldsymbol{\lambda})z_{\lambda_{j-1}} = d(\psi, \boldsymbol{\lambda}), \tag{4.12}$$

where

$$z(\psi, \boldsymbol{\lambda}) = \log\{\pi(\psi, \boldsymbol{\lambda})\},$$

$$a(\psi, \boldsymbol{\lambda}) = \{i^{11}(\psi, \boldsymbol{\lambda})\}^{1/2},$$

$$b_{j-1}(\psi, \boldsymbol{\lambda}) = i^{1j}(\psi, \boldsymbol{\lambda})\{i^{11}(\psi, \boldsymbol{\lambda})\}^{-1/2}, \text{ where } j = 2, \cdots, d,$$

and

$$d(\psi, \boldsymbol{\lambda}) = -\left[\frac{\partial}{\partial\psi}\{i^{11}(\psi, \boldsymbol{\lambda})\}^{1/2} + \sum_{j=2}^{d} \frac{\partial}{\partial\lambda_{j-1}}\{i^{1j}(\psi, \boldsymbol{\lambda})\}\{i^{11}(\psi, \boldsymbol{\lambda})\}^{-1/2}\right].$$

The coefficient $a(\psi, \boldsymbol{\lambda})$ is a diagonal element of the inverse matrix of (i_{jk}), so $a(\psi, \boldsymbol{\lambda})$ can not be zero. Dividing both sides of (4.12) by $a(\psi, \boldsymbol{\lambda})$, one has

$$z_\psi + \sum_{j=2}^{d} \frac{b_{j-1}(\psi, \boldsymbol{\lambda})}{a(\psi, \boldsymbol{\lambda})} z_{\lambda_{j-1}} = \frac{d(\psi, \boldsymbol{\lambda})}{a(\psi, \boldsymbol{\lambda})}.$$

The same as for the case of $d = 2$, this forces the coefficient of z_ψ to be 1, which simplifies the procedure of finding a solution.

The solution $z(\psi, \boldsymbol{\lambda})$ is a hypersurface (dimension d) in a $(d+1)$-dimensional space. When $d = 2$, the solution $z(\psi, \boldsymbol{\lambda})$ is a 2-dimensional surface in a 3-dimensional space. Therefore, the solving procedure and the solution surface can be visualized with a single nuisance parameter.

To solve the equation (4.12), it suffices to solve the following ordinary differential equations system

$$\frac{d\psi}{ds} = 1, \frac{d\lambda_1}{ds} = \frac{b_1(\psi, \boldsymbol{\lambda})}{a(\psi, \boldsymbol{\lambda})}, \cdots, \frac{d\lambda_{d-1}}{ds} = \frac{b_{d-1}(\psi, \boldsymbol{\lambda})}{a(\psi, \boldsymbol{\lambda})}, \frac{dz}{ds} = \frac{d(\psi, \boldsymbol{\lambda})}{a(\psi, \boldsymbol{\lambda})}. \tag{4.13}$$

To be more specific with the solution, express the initial conditions parametrically, in terms of $\boldsymbol{\xi} = (\xi_1, \cdots, \xi_{d-1})$, as

$$\psi = \Psi(\boldsymbol{\xi}), \ \boldsymbol{\lambda} = \boldsymbol{\Lambda}(\boldsymbol{\xi}) = \{\Lambda_1(\boldsymbol{\xi}), \cdots, \Lambda_{d-1}(\boldsymbol{\xi})\}, \tag{4.14}$$

and

$$z = Z(\boldsymbol{\xi}) = z\{\Psi(\boldsymbol{\xi}), \boldsymbol{\Lambda}(\boldsymbol{\xi})\}. \tag{4.15}$$

Choose the initial conditions (4.14), $\{\Psi(\boldsymbol{\xi}), \Lambda_1(\boldsymbol{\xi}), \cdots, \Lambda_{d-1}(\boldsymbol{\xi})\}$, such that the vector

$$\left[1, \frac{b_1\{\Psi(\boldsymbol{\xi}), \boldsymbol{\Lambda}(\boldsymbol{\xi})\}}{a\{\Psi(\boldsymbol{\xi}), \boldsymbol{\Lambda}(\boldsymbol{\xi})\}}, \cdots, \frac{b_{d-1}\{\Psi(\boldsymbol{\xi}), \boldsymbol{\Lambda}(\boldsymbol{\xi})\}}{a\{\Psi(\boldsymbol{\xi}), \boldsymbol{\Lambda}(\boldsymbol{\xi})\}}\right]$$

can not lie in the space spanned by

$$\left\{\frac{\partial\Psi}{\partial\xi_i}, \frac{\partial\Lambda_1}{\partial\xi_i}, \cdots, \frac{\partial\Lambda_{d-1}}{\partial\xi_i}\right\}, \text{ where } i = 1, \cdots, d-1.$$

One then obtains

$$\psi = \psi(s, \boldsymbol{\xi}), \ \boldsymbol{\lambda} = \boldsymbol{\lambda}(s, \boldsymbol{\xi}) = \{\lambda_1(s, \boldsymbol{\xi}), \cdots, \lambda_{d-1}(s, \boldsymbol{\xi})\}$$

by simultaneously integrating a system of ordinary differential equations defined by

$$
\begin{aligned}
\frac{\mathrm{d}\,\psi}{\mathrm{d}\,s} &= 1, \quad \psi(s_0, \boldsymbol{\xi}) = \Psi(\boldsymbol{\xi}), \\
\frac{\mathrm{d}\,\lambda_1}{\mathrm{d}\,s} &= \frac{b_1(\psi, \boldsymbol{\lambda})}{a(\psi, \boldsymbol{\lambda})}, \quad \lambda_1(s_0, \boldsymbol{\xi}) = \Lambda_1(\boldsymbol{\xi}), \\
&\ \vdots \\
\frac{\mathrm{d}\,\lambda_{d-1}}{\mathrm{d}\,s} &= \frac{b_{d-1}(\psi, \boldsymbol{\lambda})}{a(\psi, \boldsymbol{\lambda})}, \quad \lambda_{d-1}(s_0, \boldsymbol{\xi}) = \Lambda_{d-1}(\boldsymbol{\xi}).
\end{aligned}
\tag{4.16}
$$

For the last equation of (4.13), the initial condition is given by (4.15). Then we have

$$
\frac{\mathrm{d}\,z}{\mathrm{d}\,s} = \frac{d(\psi, \boldsymbol{\lambda})}{a(\psi, \boldsymbol{\lambda})}, \quad z(s_0, \boldsymbol{\xi}) = Z(\boldsymbol{\xi}).
\tag{4.17}
$$

Once the equations system (4.16) has been solved, equation (4.17) has the solution

$$
z(s, \boldsymbol{\xi}) = Z(\boldsymbol{\xi}) + \int_{s_0}^{s} \frac{d\{\psi(s', \boldsymbol{\xi}), \lambda_1(s', \boldsymbol{\xi}), \cdots, \lambda_{d-1}(s', \boldsymbol{\xi})\}}{a\{\psi(s', \boldsymbol{\xi}), \lambda_1(s', \boldsymbol{\xi}), \cdots, \lambda_{d-1}(s', \boldsymbol{\xi})\}} \, \mathrm{d}\,s'.
\tag{4.18}
$$

These generate a hypersurface in a $(d+1)-$dimensional space, $Z(\psi, \boldsymbol{\lambda})$, that satisfies both the equation (4.12) and the initial condition. When there are no close form solutions for equations (4.16) and (4.17), numerical solutions can be achieved.

Again, no back transformation involved in obtaining the solution (4.18). If one wants to specify the value of a matching prior at a certain point, say $\boldsymbol{\omega}^* = (\psi^*, \lambda_1^*, \cdots, \lambda_{d-1}^*)$, then one can directly specify s as ψ^* and $\boldsymbol{\xi} = (\xi_1, \cdots, \xi_{d-1})$ as $(\lambda_1^*, \cdots, \lambda_{d-1}^*)$ in the formula (4.18). The matching prior evaluated at $\boldsymbol{\omega}^*$ can be achieved.

The key step in the implement of this procedure is simultaneously solving the system of ordinary differential equations (4.16). Fortunately, many mathematical/statistical software packages have these kinds of solvers for a system of ordinary differential equation. For example, the `lsoda` function in R software package `odesolve` has the ability to solve a system of ordinary differential equations. However, when $d > 2$, it can be computational intensive to implement the

procedure. Also, if the right hand sides of (4.16) do not have explicit expressions, numerical implementation may be more difficult.

As a conclusion, it is straight forward to extend the analytic formulae of the solving procedure of the partial differential equation from a single nuisance parameter to multiple nuisance parameters. It may not be easy for the numerical implementation of the procedure with $d > 2$, when there are no explicit expressions for the coefficients in the original first order partial differential equation (4.1).

4.5 Examples

4.5.1 Ratio of two exponential means

Let X and Y be exponential random variables with means μ and ν respectively; the ratio of the means, ν/μ, is the parameter of interest. The parameter transformation

$$\left(\mu \to \lambda\psi^{-\frac{1}{2}}, \; \nu \to \lambda\psi^{\frac{1}{2}}\right)$$

makes the two new parameters ψ and λ orthogonal. Then X and Y have expectations $\lambda\psi^{-\frac{1}{2}}$ and $\lambda\psi^{\frac{1}{2}}$, respectively.

Suppose there are n independent replications of (X, Y). Denote $\boldsymbol{\omega} = (\psi, \lambda)$. One can obtain the log-likelihood function as

$$l(\boldsymbol{\omega}) = -n\left\{\frac{\psi\bar{x} + \bar{y}}{\lambda\sqrt{\psi}} + 2\log\lambda\right\}.$$

Both approximations of the Barndorff-Nielson format (1.3) and the Lugannani and Rice format (1.2) are considered. Based on these approximations, p-values can be calculated. Approximations based on different prior density functions mentioned previously may be used to generate an approximate one-sided p-value by approximating $\mathrm{P}(R \geq r)$, for r the observed value of R. Approximate two-sided p-values may be calculated by approximating $2\min\{\mathrm{P}(R \geq r), \mathrm{P}(R < r)\}$.

One and two-sided hypotheses tests of size α may be constructed by rejecting the null hypothesis when the p-value is less than α. Table 4.1 reports type I error probabilities of the 1,000,000 rounds of simulation with $n = 10$.

In this example, the parameters ψ and λ are orthogonal. Using the simplified partial differential equation (4.2), $\pi(\psi, \lambda) = 1/\psi$ and $\pi(\psi, \lambda) = 1/(\psi\lambda)$ are both explicit solutions. Numerical solutions were also calculated. One of the initial condition is

$$\{\Psi(\xi), \Lambda(\xi), Z(\xi)\} = (0, \xi, -1).$$

The resulting matching prior corresponds to the the analytic solution $1/\psi$. Another numerically solved matching prior is based on the initial condition

$$\{\Psi(\xi), \Lambda(\xi), Z(\xi)\} = (0, \xi, -\log \xi),$$

which corresponds to the the analytic solution $1/(\psi\lambda)$. From Table 4.1, one can see that the numerical and analytic solutions give almost the same simulation results, which confirmed the validity of our numerical solution process.

Table 4.1: Ratio of two exponential means: type I error probability

	BN Format		LR Format	
Tests	1-sided	2-sided	1-sided	2-sided
Likelihood ratio test	0.0520	0.0526	0.0520	0.0526
I.C. $(0, \xi, -1)$	0.0456	0.0441	0.0456	0.0441
Analytic solution: $1/\psi$	0.0456	0.0441	0.0456	0.0441
I.C. $(0, \xi, -\log \xi)$	0.0499	0.0498	0.0499	0.0498
Analytic solution: $1/(\psi\lambda)$	0.0499	0.0498	0.0499	0.0498

[*]I.C. stands for initial condition.

[†]Results are based on 1,000,000 rounds of simulation with $n = 10$.

[‡]Tests are of nominal type I error 0.05.

Approximations (1.3) and (1.2) have a removable singularity at $R = 0$. Consequently, these and similar formulae require care when evaluating near $R = 0$. In these cases, for all but the most extreme conditioning events, the resulting conditional p-value is large enough as to not imply rejection of the null hypothesis,

and so these simulated data sets are treated as not implying rejection of the null hypothesis.

4.5.2 Logistic regression

In this example, I consider a logistic regression model with a binary response Y and only one explanatory variable X. Let ω_1 denote the unknown intercept and ω_2 denote the unknown effect of the explanatory variable. Suppose ω_2 is the parameter of interest and ω_1 is the nuisance parameter. I will solve matching priors and apply DiCiccio and Martin's approximations to do inference about ω_2. Levine and Casella (2003) considered a similar example.

Let Y_i be the response variable taking binary values with success probability as p_i, and X_i be the explanatory variable following uniform distribution $U(0, 1)$. Suppose there are n independent replications of (X_i, Y_i). Fit the model

$$\log\left(\frac{p_i}{1 - p_i}\right) = v_i'\boldsymbol{\omega} = \omega_1 + \omega_2 x_i,$$

where $v_i = (1, x_i)'$ and ω_2 is the parameter of interest. Inverting the equation, we have $p_i = (1 + e^{-v_i'\boldsymbol{\omega}})^{-1}$. One can obtain the log-likelihood function as

$$l(\boldsymbol{\omega}; x) = \sum_{i=1}^{n} y_i \log\left(\frac{p_i}{1 - p_i}\right) + \sum_{i=1}^{n} \log(1 - p_i).$$

The first derivative of the log-likelihood function is $V'(y - p)$, where V is the design matrix with v_i' in row i. The second derivative of log-likelihood function is $-V'WV$, where W is a diagonal matrix with diagonal elements $p_i(1 - p_i), i = 1, \cdots, n$.

Using sample size $n = 30$, generate data satisfying the logistic regression model with $\omega_1 = -1$, $\omega_2 = 0.5$, and the explanatory variable X following uniform distribution $U(0, 1)$. For the logistic regression model, generally the parameters ω_1 and ω_2 are not orthogonal. I use the numerical procedure described in §4.3 and study performances of different initial conditions. Table 4.2 contains type I

error probabilities for both one-sided and two-sided tests for approximations of both Barndorff-Nielson format and Lugannani and Rice format, based on 10,000 rounds of simulation.

As we mentioned previously, approximations (1.3) and (1.2) have a removable singularity when both R and T approaching 0. I deal with this singularity the same way as in §3.3.1.

In the following, I give some instructions on how to change the initial condition and how to choose favorable initial conditions. Initial condition $(0, \xi, -1)$ gives type I error probabilities larger than the nominal level 0.05; that is to say, it has the tendency to underestimate tail probabilities and reject the null hypothesis. I want to choose initial conditions to obtain a test whose type I error rate is closer to the nominal level. I adjust the initial condition when solving the partial differential equation (4.1), and use the Barndorff-Nielson format of the approximation. The quantity T in (2.8) is the only part in the approximation that relates to matching priors. For a one-sided test, when the probability is small and close to 0, R and T are negative. Making $\Phi\{R + R^{-1} \log(T/R)\}$ larger is equivalent to making T bigger. Also one may notice that $Z(\xi)$ is used only in equation (4.9). Suppose the initial condition is $\{\Psi(\xi), \Lambda(\xi), Z(\xi)\}$. Keep the first two components of the initial condition, $\Psi(\xi)$ and $\Lambda(\xi)$, unchanged, and only modify the third term, $Z(\xi)$. By doing so, the integral part in equation (4.9) is kept unchanged and z varies only with $Z(\xi)$. By changing $Z(\xi)$, I want to adjust T to be bigger. Because T is negative when reject a hypothesis, and matching priors appear in T as a ratio, one can construct a $Z(\cdot)$ such that the ratio, $\exp\{Z(\hat{\psi}, \hat{\lambda})\}/\exp\{Z(\psi_0, \hat{\lambda}_0)\}$, will be smaller than 1; recall that 1 is the value of the ratio when $Z(\xi) = -1$. Based on the above arguments, $Z(\cdot)$ function is constructed as $Z(\xi) = -\log\{(\xi + 1)^q + 1\}$, where q is a tuning parameter and leads $Z(\cdot)$ to an even function. As an even function, $Z(\xi)$ achieves its maximum value at -1, where -1 is the true value for the nuisance parameter when data

were simulated. We have constructed priors using knowledge of the true value of the nuisance parameter. Of course, in practice this knowledge is unavailable. One might instead use an estimator of the nuisance parameter in place of the true value.

When $Z(\xi)$ increases quickly, such as $q = 2$ in table 4.2, the type I error probability deviates far away from the nominal level in the other direction. If a more slowly increasing functions is used, the performance of type I error may be better.

Table 4.2: Logistic regression: type I error probability

Test	BN Format		LR Format	
	1-sided	2-sided	1-sided	2-sided
Likelihood ratio test	0.054	0.060	0.054	0.060
I.C. $(0, \xi, -1)$	0.052	0.057	0.052	0.057
I.C. $[0, \xi, -\log\{(\xi+1)^2 + 1\}]$	0.028	0.019	0.031	0.020
I.C. $[0, \xi, -\log\{(\xi+1)^{2/5} + 1\}]$	0.041	0.041	0.044	0.046
I.C. $[0, \xi, -\log\{(\xi+1)^{2/11} + 1\}]$	0.045	0.048	0.046	0.050

*I.C. stands for initial condition.

†Results are based on 10,000 rounds of simulation with $n = 30$.

‡Tests are of nominal type I error 0.05.

Unfortunately, with some choices of initial conditions, such as the last three listed in table 4.2, the Lugannani and Rice format approximation may fall outside the range of 0 and 1 in some cases. For example, the initial condition of $[0, \xi, -\log\{(\xi+1)^2 + 1\}]$ yielded 5 such probabilities out of 10,000 data sets. Those values are converted to 0 or 1 by $\min\{\max(p, 0), 1\}$, where p is the p-value that is outside 0 and 1.

For the parameter of interest ω_2, I calculate credible intervals using DiCiccio and Martin's approximation in Barndorff-Nielsen format. With initial condition $(0, \xi, -1)$, out of 1,000 generated data sets, there are 938 credible intervals covered the true value 0.5. With initial condition $[0, \xi, -\log\{(\xi+1)^{2/5} + 1\}]$, for the parameter of interest ω_2, there are 954 credible intervals covered the true value 0.5.

The above procedure was applied to a real data set from Hosmer and Lemeshow (2000, Table 1.1). The response variable is coronary heart disease indicator, y, and the explanatory variable is age, x. One hundred subjects were included in the study; i.e. $n = 100$. Fit the logistic regression model following the same definition as above, with ω_1 defined for the unknown intercept and ω_2 for the effect of age on heart disease status. Using initial condition $(0, \xi, -1)$ and Barndorff-Nielson format approximation, a two-sided testing p-values is 5.532326×10^{-8}, and five and ninety-five posterior percentiles are of 0.07 and 0.15 respectively.

4.6 Discussion

Matching priors are very interesting by themselves, when considered independently of DiCiccio and Martin's approximations. These priors have shown their usefulness as well as their well-known computational difficulty. In this section, I want to discuss the matching priors as solutions from a system of first order partial differential equation.

● **Solution from a system of first order partial differential equations**

As one may notice that the solution to (4.1) is not unique. Peers (1965) has proposed a way to narrow down the class of solutions, which is to use the same weight function for any component of $\boldsymbol{\omega}$. This is motivated by the fact that in the Bayesian framework it is the same prior density for all components of $\boldsymbol{\omega}$. Also, as noted in Peers (1965), if the confidence points for each component are all to be formally identical with the corresponding Bayesian probability points, it is essential that this be achieved by taking the same weight function for all components. To find a common prior density function is equivalent to find a common solution to the following system of d partial differential equations

$$\sum_{j=1}^{d} i^{ij} (i^{ii})^{-1/2} \frac{\partial}{\partial \omega^j} (\log \pi) + \sum_{j=1}^{d} \frac{\partial}{\partial \omega^j} \{ i^{ij} (i^{ii})^{-1/2} \} = 0 \quad i = 1, 2, \cdots, d. \quad (4.19)$$

As a reminder, here d is the number of components in model parameter $\boldsymbol{\omega}$. Equation (4.1) is the first equation ($i = 1$) in (4.19). A necessary and sufficient condition for the existence of a solution of (4.19) is that the equation

$$\mathrm{d}\,\pi(\boldsymbol{\omega}) = \pi_i(\boldsymbol{\omega})\,\mathrm{d}\,\omega_i$$

be completely integrable, where

$$\pi_i(\boldsymbol{\omega}) = \frac{\partial \pi(\boldsymbol{\omega})}{\partial \omega_i}, \quad i = 1, 2, \cdots, d.$$

The above will be so if and only if the integrability conditions

$$\frac{\partial \pi_i}{\partial \omega_j} = \frac{\partial \pi_j}{\partial \omega_i}, i \neq j$$

are satisfied.

Only in relatively special cases will the integrability conditions be satisfied. Peers (1965) comments on them. I studied the ratio of the means of two exponential distributed variables as an example, where a solution to (4.19) can be found as $1/(\psi\chi)$. The approximations based on the prior $1/(\psi\chi)$ yields the best performance among all DiCiccio and Martin–type approximations under consideration. However, for more complicated model, such as the Cox model, it is nearly impossible to find such a solution from a system of first order partial differential equations.

Chapter 5

Conclusion

Generally speaking, saddlepoint methods provide accurate approximations of tail probabilities, and consequently can be converted to confidence limits of the corresponding order. For some of those approximations, such as Barndorff-Nielsen's approximation (1986), it is necessary to conditional on an ancillary statistic to proceed with calculation. DiCiccio and Martin's approximations (1993) do not have such a difficulty. They can be constructed based on general form of log-likelihood functions, have more applications outside the canonical exponential family, and the calculation is relatively easy.

Similar to other saddlepoint approximations, DiCiccio and Martin's approximation may take two formats, which are the Barndorff-Nielson format

$$\Phi\{R + R^{-1}\log(T/R)\}$$

and the Lugannani and Rice format

$$\Phi(R) + \phi(R)(R^{-1} - T^{-1}),$$

where the variable T is defined in (2.8).

It is directly observed from the above two formats that they have a singularity when both R and T approach 0. This causes the unstable approximations to the tail probability around the conditional mean of the distribution approximated. Fortunately, this singularity for $1/R - 1/T$ or $R^{-1}\log(T/R)$ is removable. I propose a correction to overcome the instability difficulties. The correction around the singularity is necessary in order to calculate both the significance tests and

confidence intervals. In these situations, formulae like those I present in chapter 3 will be very useful.

The approximations of DiCiccio and Martin (1993) involve Bayesian techniques and there is always a prior density function needs to be specified. In the general parametrization, if the parameter of interest and the nuisance parameters are not orthogonal, solving the prior from a first order partial differential equation is nontrivial. Though one may try to obtain orthogonal parameters via parameter transformation as introduced in Cox and Reid (1980), sometimes it is hard to achieve such transformation. The prior density function here is also called as matching prior, which is first proposed by Welch and Peers (1963) and Peers (1965). I present a practical way to solve for the matching priors and the procedure can be suitable to all kinds of initial conditions. By choosing differential initial conditions one is able to improve the performances of DiCiccio and Martin's approximations.

Therefore, I suggest the use of DiCiccio and Martin's approximations with the construction procedure and correction that I provide in this dissertation.

References

[1] Barndorff-Nielsen, O. E. (1980) Conditionality resolutions. *Biometrika*, **67**, 293-310.

[2] Barndorff-Nielsen, O. E. (1986) Inference on full or partial parameters based on the standardized signed log likelihood ratio. *Biometrika*, **73**, 307-322.

[3] Barndorff-Nielsen, O. E. (1988) Discussion of the paper by N. Reid. *Statistical Science*, **3**, 228-229.

[4] Barndorff-Nielsen, O. E. (1990) Approximate interval probabilities. *Journal of the Royal Statistical Society Series B*, **52**, 485-496.

[5] Barndorff-Nielsen, O. E. (1991) Modified signed log likelihood raito. *Biometrika*, **78**, 557-563.

[6] Bickel, P. J. and Ghosh, J. K. (1990) A decomposition for the likelihood ratio statistic and the Bartlett correction - a Bayesian argument. *Annal of Statistics*, **18**, 1070-1090.

[7] Butler, R. W. (2007) *Saddlepoint Approximations with Applications*, Cambridge University Press, New York.

[8] Cox, D. R., and Reid, N. (1987) Parameter orthogonality and approximate conditional inference. *Journal of the Royal Statistical Society Series B*, **49**, 1-39.

[9] Curran, J. (2005) *Bolstad: Bolstad functions*. R package version 0.2-12.

[10] DiCiccio, T.J., Field, C. A. and Fraser, D. A. S. (1990) Approximations of marginal tail probabilities and inference for scalar parameters. *Biometrika*, **77**, 77-95.

[11] DiCiccio, T. J. and Martin, M. A. (1993) Simple modifications for signed roots of likelihood ratio statistics. *Journal of the Royal Statistical Society Series B*, **55**, No.1, 305-316.

[12] Hindmarsh, A. C. (1983) *Scientific Computing*. Amsterdam: North-Holland. Ed. Stepleman, R.W. et al.; Vol. 1 of IMACS Transactions on Scientific Computation.

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

[14] Huzurbazar, S., and Huzurbazar, A. V. (1999) Survival and hazard functions for progressive diseases using saddlepoint approximations. *Biometrics*, **55**, 198-203.

[15] Kolassa, J. E. (2006) *Series Approximation Methods in Statistics*, 3rd ed., Springer, New York.

[16] Levine, R. A. and Casella, G. (2003) Implementing matching priors for frequentist inference. *Biometrika*, **90**, 127-137.

[17] Lugannani, R., and Rice, S. (1980) Saddle point approximation for the distribution of the sum of independent random variables. *Advances in Applied Probability*, **12**, 475-490.

[18] Peers, H. W. (1965) On confidnece points and Bayesian probability points in the case of several parameters. *Journal of the Royal Statistical Society Series B*, **27**, 9-16.

[19] Petzold, L. R. (1983) Automatic selection of methods for solving stiff and nonstiff systems of ordinary differential equations. *SIAM: Journal on Scientific and Statistical Computing*, **4**, 136-48.

[20] R Development Core Team (2007) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

[21] Reid, N. (2003) Asymptotics and the theory of inference, *The Annals of Statistics*, **31**, 1695-1731.

[22] Rhee, H., Aris, R. and Amundson, N. (1986) *First-order Partial Differential Equations: Theory and Application of Single Equations*,Vol. 1. Englewood Cliffs, New Jersey: Prentice-Hall.

[23] Setzer, R. W. (2006). *odesolve: Solvers for Ordinary Differential Equations*. R package version 0.5-17.

[24] Severini, T. A. (1999) An empirical adjustment to the likelihood ratio statistic. *Biometrika*, **86**, 235-247.

[25] Skovgaard, I. M. (1987) Saddlepoint Expansions for Conditional Distributions. *Journal of Applied Probability*, **24**, 875-887.

[26] Staicu, A. and Reid, N. (2007) On the uniqueness of probability matching priors. Submitted manuscript.
http://www.utstat.toronto.edu/reid/research/staicu-reid.pdf

[27] Sweeting, T. J. (2005) On the implementation of local probability priors for interest parameters. *Biometrika*, **92**, 47-57.

[28] Tibshirani, R. (1989). Noninformative priors for one parameter of many. *Biometrika* **76**, 604-608.

[29] Welch, B. L. and Peers, H. W. (1963) On formulae for confidence points based on integrals of weighted likelihoods. *Journal of the Royal Statistical Society Series B*, **25**, 318-329.

[30] Yang, B. and Kolassa, J. E. (2002) Saddlepoint approximation for the distribution function near the mean. *Annals of the Institute of Statistical Mathematics*, **54**, No.4, 743-747.

[31] Zhang J. and Kolassa, J. E. (2007) A comparison of the accuracy of saddlepoint conditional cumulative distribution function approximations, *IMS-LNMS*, **54**, 250–259.

[32] Zhang J. and Kolassa, J. E. (2007) A practical procedure to find matching priors for frequentist inference. Technical report #2007-006 (10/24/2007), Department of Statistics and Biostatistics, Rutgers University.

[33] Zhang J. and Kolassa, J. E. (2008) Saddlepoint approximation for the distribution of the modified signed root of likelihood ratio statistics near the mean, *Communications in Statistics - Theory and Methods*, **37**, issue 2, 194–203.

Appendix A

Some algebraic calculations

A.1 Barndorff-Nielsen's modification

The expectations of the sufficient statistics $\mathbf{T} = (\overline{X}, \overline{Y})$ in the new parameterization are

$$\boldsymbol{\tau}(\psi, \lambda) = \left\{ \frac{\lambda}{\sqrt{\psi}}, \lambda\sqrt{\psi} \right\},$$

and

$$\frac{\mathrm{d}\,\boldsymbol{\tau}(\psi, \lambda)}{\mathrm{d}\,\psi} = \left\{ -\frac{\lambda}{2\psi^{\frac{3}{2}}}, \frac{\lambda}{2\psi^{\frac{1}{2}}} \right\}.$$

A vector perpendicular to this is

$$\left(\frac{\mathrm{d}\,\boldsymbol{\tau}(\psi, \lambda)}{\mathrm{d}\,\psi} \right)^{\perp} = \{\psi, 1\}.$$

The variance of the sample mean vector is

$$\boldsymbol{\Sigma}(\psi, \lambda) = \frac{1}{n} \begin{pmatrix} \frac{\lambda^2}{\psi} & 0 \\ 0 & \lambda^2 \psi \end{pmatrix}.$$

In our case,

$$\mathbf{B}(\psi) = \sqrt{n} \left\{ \frac{\sqrt{\psi}}{\sqrt{2}\lambda}, \frac{1}{\lambda\sqrt{2\psi}} \right\},$$

and

$$A = \sqrt{2n} \left(\frac{\sqrt{\overline{X}\,\overline{Y}}}{\lambda} - 1 \right).$$

Using Barndorff-Nielsen's formula (1990),

$$\tilde{l}(\psi; \hat{\psi}, a) = -\frac{|a + \sqrt{2n}|(\psi + \hat{\psi})}{\sqrt{2n\psi\hat{\psi}}} - 2\log\lambda,$$

and

$$\hat{\omega} = \text{sign}(\hat{\psi} - \psi)\psi^{1/4} \left| (a + \sqrt{2n})(\psi^{\frac{1}{2}} - \hat{\psi}^{\frac{1}{2}}) \right| \hat{\psi}^{-1/4} n^{-1/2}.$$

The negative of the second derivative of the log likelihood is

$$j(\psi) = \left| a + \sqrt{2n} \right| (3\psi - \hat{\psi}) \left(4\sqrt{2n\psi\hat{\psi}^5} \right)^{-1},$$

and the derivative of $\tilde{l}(\psi; \hat{\psi}, a)$ with respect to $\hat{\psi}$ is

$$\left| a + \sqrt{2n} \right| (\psi - \hat{\psi}) \left(2\sqrt{2n\psi\hat{\psi}^3} \right)^{-1}.$$

Then the quantity $\check{z}$ contributing to the tail probability approximation (2.5) is

$$\check{z} = -\sqrt{\left| a + \sqrt{2n} \right|} \, (\psi - \hat{\psi}) \left(2\sqrt{2n\psi\hat{\psi}} \right)^{-1}.$$

A.2 DiCiccio and Martin's modification

Based on the above, the information matrix is

$$i(\boldsymbol{\omega}) = \text{E}[-l''(\boldsymbol{\omega})] = n \begin{pmatrix} \frac{1}{2\psi^2} & 0 \\ 0 & 2/\lambda^2 \end{pmatrix}.$$

The maximum likelihood estimators are

$$\hat{\psi} = \frac{\overline{Y}}{\overline{X}} \quad \text{and} \quad \hat{\lambda} = \frac{\hat{\psi}\overline{X} + \overline{Y}}{2\hat{\psi}^{1/2}} = \sqrt{\overline{X}\,\overline{Y}}.$$

For fixed ψ, let $\hat{\lambda}_\psi$ be the constrained maximum likelihood of λ. Here,

$$\hat{\lambda}_\psi = \frac{\psi\overline{X} + \overline{Y}}{2\psi^{1/2}}.$$

If $\psi = \psi_0 = 1$, then

$$\hat{\lambda}_{\psi_0} = \hat{\lambda}_0 = \frac{\overline{X} + \overline{Y}}{2}.$$

In this case, the parameters are orthogonal. Using the simplified partial differential equation (4.2), choose $g(\lambda) = 1$, and hence

$$\pi(\psi, \lambda) = \frac{\sqrt{n}}{\sqrt{2}\,\psi}.$$

In addition to use the prior solved from equation (4.2), I also studied the outcome from a uniform prior, that is to say, the prior with a constant density, which is obviously not a solution to equation (4.2).

Appendix B

Evaluation of expansion (3.4)

One needs to evaluate (3.4). The functions g and h are expanded in ψ_0. Here the Einstein summation notation is employed, whereby an index repeated as a subscript and as a superscript implies summation over that index. First, one has

$$h(\psi_0) = l(\hat{\boldsymbol{\omega}}) - l(\tilde{\boldsymbol{\omega}}),$$

$$h^u(\psi) = -l^k(\tilde{\boldsymbol{\omega}})\tilde{\omega}_k^u,$$

$$h^{uv}(\psi_0) = -l^{jk}(\tilde{\boldsymbol{\omega}})\tilde{\omega}_k^u\tilde{\omega}_j^v - l^k(\tilde{\boldsymbol{\omega}})\tilde{\omega}_k^{uv},$$

$$h^{uvw}(\psi_0) = -l^{jkm}(\tilde{\boldsymbol{\omega}})\tilde{\omega}_k^u\tilde{\omega}_j^v\tilde{\omega}_m^w - l^{jk}(\tilde{\boldsymbol{\omega}})\tilde{\omega}_k^{uw}\tilde{\omega}_j^v[3] - l^k(\tilde{\boldsymbol{\omega}})\tilde{\omega}_k^{uvw};$$

$$h^{uvwt}(\psi_0) = -l^{jkmn}(\tilde{\boldsymbol{\omega}})\tilde{\omega}_k^u\tilde{\omega}_j^v\tilde{\omega}_m^w\tilde{\omega}_n^t - l^{jkm}(\tilde{\boldsymbol{\omega}})\tilde{\omega}_k^{ut}\tilde{\omega}_j^v\tilde{\omega}_m^w[6]$$

$$- l^{jk}(\tilde{\boldsymbol{\omega}})\tilde{\omega}_k^{uwt}\tilde{\omega}_j^v[4] - l^{jk}(\tilde{\boldsymbol{\omega}})\tilde{\omega}_k^{uw}\tilde{\omega}_j^{vt}[3] - l^k(\tilde{\boldsymbol{\omega}})\tilde{\omega}_k^{uvwt}.$$

Here integers in brackets denote permutations of the indices; for example,

$$l^{jk}(\tilde{\boldsymbol{\omega}})\tilde{\omega}_k^{uw}\tilde{\omega}_j^v[3] = l^{jk}(\tilde{\boldsymbol{\omega}})\tilde{\omega}_k^{uw}\tilde{\omega}_j^v + l^{jk}(\tilde{\boldsymbol{\omega}})\tilde{\omega}_k^{vw}\tilde{\omega}_j^u + l^{jk}(\tilde{\boldsymbol{\omega}})\tilde{\omega}_k^{uv}\tilde{\omega}_j^w$$

is the sum over the three terms of three indices. The bracket notation is simply a convenience to avoid listing explicitly all the terms with respect to the indices. Similarly, one obtains

$$b(\psi_0) = l^1(\tilde{\boldsymbol{\omega}}),$$

$$b^v(\psi_0) = l^{1j}(\tilde{\boldsymbol{\omega}})\tilde{\omega}_j^v,$$

$$b^{vw}(\psi_0) = l^{1jk}(\tilde{\boldsymbol{\omega}})\tilde{\omega}_k^w\tilde{\omega}_j^v + l^{1j}(\tilde{\boldsymbol{\omega}})\tilde{\omega}_j^{vw},$$

$$b^{vwt}(\psi_0) = l^{1jkm}(\tilde{\boldsymbol{\omega}})\tilde{\omega}_m^t\tilde{\omega}_k^w\tilde{\omega}_j^v + l^{1jk}(\tilde{\boldsymbol{\omega}})\tilde{\omega}_k^{wt}\tilde{\omega}_j^v + l^{1jk}(\tilde{\boldsymbol{\omega}})\tilde{\omega}_k^w\tilde{\omega}_j^{vt}$$

$$+ l^{1jk}(\tilde{\boldsymbol{\omega}})\tilde{\omega}_k^t\tilde{\omega}_j^{vw} + l^{1j}(\tilde{\boldsymbol{\omega}})\tilde{\omega}^{vwt}.$$

Evaluating all the derivatives at $\hat{\boldsymbol{\omega}} = (\hat{\psi}, \hat{\boldsymbol{\lambda}})$, one obtains

$$h(\hat{\psi}) = -l(\hat{\boldsymbol{\omega}}),$$

$$h^u(\hat{\psi}) = 0,$$

$$h^{uv}(\hat{\psi}) = -l^{jk}(\hat{\boldsymbol{\omega}})\hat{\omega}_k^u\hat{\omega}_j^v,$$

$$h^{uvw}(\hat{\psi}) = -l^{jkm}(\hat{\boldsymbol{\omega}})\hat{\omega}_k^u\hat{\omega}_j^v\hat{\omega}_m^w - l^{jk}(\hat{\boldsymbol{\omega}})\hat{\omega}_k^{uw}\hat{\omega}_j^v[3],$$

$$h^{uvwt}(\hat{\psi}) = -l^{jkmn}(\hat{\boldsymbol{\omega}})\hat{\omega}_k^u\hat{\omega}_j^v\hat{\omega}_m^w\hat{\omega}_n^t - l^{jkm}(\hat{\boldsymbol{\omega}})\hat{\omega}_k^{ut}\hat{\omega}_j^v\hat{\omega}_m^w[6]$$

$$- l^{jk}(\hat{\boldsymbol{\omega}})\hat{\omega}_k^{uwt}\hat{\omega}_j^v[4] - l^{jk}(\hat{\boldsymbol{\omega}})\hat{\omega}_k^{uw}\hat{\omega}_j^{vt}[3], \tag{B.1}$$

and

$$b(\hat{\psi}) = 0,$$

$$b^v(\hat{\psi}) = l^{1j}(\hat{\boldsymbol{\omega}})\hat{\omega}_j^v,$$

$$b^{vw}(\hat{\psi}) = l^{1jk}(\hat{\boldsymbol{\omega}})\hat{\omega}_k^w\hat{\omega}_j^v + l^{1j}(\hat{\boldsymbol{\omega}})\hat{\omega}_j^{vw},$$

$$b^{vwt}(\hat{\psi}) = l^{1jkm}(\hat{\boldsymbol{\omega}})\hat{\omega}_m^t\hat{\omega}_k^w\hat{\omega}_j^v + l^{1jk}(\hat{\boldsymbol{\omega}})\hat{\omega}_k^{wt}\hat{\omega}_j^v + l^{1jk}(\hat{\boldsymbol{\omega}})\hat{\omega}_k^w\hat{\omega}_j^{vt}$$

$$+ l^{1jk}(\hat{\boldsymbol{\omega}})\hat{\omega}_k^t\hat{\omega}_j^{vw} + l^{1j}(\hat{\boldsymbol{\omega}})\hat{\omega}^{vwt}.$$

Denote $G(\tilde{\boldsymbol{\omega}}) = \log(\det[-a^{j;p}(\tilde{\boldsymbol{\omega}})]) - 2\log(\pi(\tilde{\boldsymbol{\omega}}))$. One obtains

$$G^q(\psi_0) = a_{j;k}(\tilde{\boldsymbol{\omega}})a^{k;jq}(\tilde{\boldsymbol{\omega}}) - 2\pi^{-1}(\tilde{\boldsymbol{\omega}})\pi^q(\tilde{\boldsymbol{\omega}}),$$

$$G^{qr}(\psi_0) = a_{j;k}(\tilde{\boldsymbol{\omega}})a^{k;jqr}(\tilde{\boldsymbol{\omega}}) - a_{j;m}(\tilde{\boldsymbol{\omega}})a^{m;pr}(\tilde{\boldsymbol{\omega}})a_{p;k}(\tilde{\boldsymbol{\omega}})a^{k;jq}(\tilde{\boldsymbol{\omega}})$$

$$- 2\pi^{-1}(\tilde{\boldsymbol{\omega}})\pi^{qr}(\tilde{\boldsymbol{\omega}}) + 2\pi^{-2}(\tilde{\boldsymbol{\omega}})\pi^q(\tilde{\boldsymbol{\omega}})\pi^r(\tilde{\boldsymbol{\omega}}).$$

Converting $G(\tilde{\boldsymbol{\omega}})$ back to $g(\psi_0)$ and evaluating at $\hat{\boldsymbol{\omega}}$, one achieves

$$g(\hat{\psi}) = \exp(G(\hat{\boldsymbol{\omega}})),$$

$$g^u(\hat{\psi}) = g(\hat{\psi})G^k(\hat{\boldsymbol{\omega}})\hat{\omega}_k^u,$$

$$g^{uv}(\hat{\psi}) = g(\hat{\psi})\{[G^{jk}(\hat{\boldsymbol{\omega}}) + G^k(\hat{\boldsymbol{\omega}})G^j(\hat{\boldsymbol{\omega}})]\hat{\omega}_k^u\hat{\omega}_j^v + G^j(\hat{\boldsymbol{\omega}})\hat{\omega}_j^{uv}\}. \tag{B.2}$$

The derivatives of $\tilde{\boldsymbol{\omega}}$ are also needed. Recall that $\tilde{\boldsymbol{\omega}}$ satisfies $l^j(\tilde{\boldsymbol{\omega}}) = 0$ for all $j > 1$, and $\tilde{\omega}_j = \psi_0$ for $j = 1$. Then $l^{jk}(\tilde{\boldsymbol{\omega}})\tilde{\omega}_k^1 = 0$ for all $j > 1$, and $\delta^{jk}\tilde{\omega}_k^1 = j$ for

$j = 1$. Here δ^{jk} is defined as 1 if $j = k$, and 0 if $j \neq k$. Let

$$a^j(\boldsymbol{\omega}) = \begin{cases} l^j(\boldsymbol{\omega}) & \text{if } j > 1 \\ \omega_j & \text{if } j = 1. \end{cases} \tag{B.3}$$

For m indices $p, \ldots, s$, let

$$a^{j;p,\ldots,s}(\boldsymbol{\omega}) = \frac{\partial^m}{\partial \omega_p \ldots \omega_s} a^j(\boldsymbol{\omega}).$$

Let $a_{j;k}(\boldsymbol{\omega})$ be such that

$$a_{j;m}(\boldsymbol{\omega}) a^{m;k}(\boldsymbol{\omega}) = \delta^{jk}$$

for all j and k. Then evaluating (B.3) with $\boldsymbol{\omega} = \tilde{\boldsymbol{\omega}}(\psi_0, \omega_2, \ldots, \omega_d)$, and differentiating, one finds that $a^{j;p}(\tilde{\boldsymbol{\omega}})\tilde{\omega}_p^k = \delta^{jk}$, for $k = 1$. Hence

$$\tilde{\omega}_j^k = a_{j;m}(\tilde{\boldsymbol{\omega}})\delta^{mk},$$

$$\tilde{\omega}_j^{uv} = -a_{j;p}(\tilde{\boldsymbol{\omega}})a^{p;qr}(\tilde{\boldsymbol{\omega}})a_{q;k}(\tilde{\boldsymbol{\omega}})\tilde{\omega}_r^u \delta^{kv} = -a_{j;p}(\tilde{\boldsymbol{\omega}})a^{p;qr}(\tilde{\boldsymbol{\omega}})\tilde{\omega}_r^u \tilde{\omega}_q^v,$$

$$\tilde{\omega}_j^{uvw} = a_{j;s}(\tilde{\boldsymbol{\omega}})a^{s;tk}a_{t;p}a^{p;qr}(\tilde{\boldsymbol{\omega}})\tilde{\omega}_r^u \tilde{\omega}_q^v \tilde{\omega}_k^w - a_{j;p}(\tilde{\boldsymbol{\omega}})a^{p;qrs}(\tilde{\boldsymbol{\omega}})\tilde{\omega}_r^u \tilde{\omega}_q^v \tilde{\omega}_s^w$$

$$\quad - a_{j;p}(\tilde{\boldsymbol{\omega}})a^{p;qr}(\tilde{\boldsymbol{\omega}})\tilde{\omega}_r^{uw}\tilde{\omega}_q^v - a_{j;p}(\tilde{\boldsymbol{\omega}})a^{p;qr}(\tilde{\boldsymbol{\omega}})\tilde{\omega}_r^u \tilde{\omega}_q^{vw}$$

$$\quad = -a_{j;p}(\tilde{\boldsymbol{\omega}})a^{p;qrk}(\tilde{\boldsymbol{\omega}})\tilde{\omega}_r^u \tilde{\omega}_q^v \tilde{\omega}_k^w - a_{j;p}(\tilde{\boldsymbol{\omega}})a^{p;qr}(\tilde{\boldsymbol{\omega}})\tilde{\omega}_r^{uw}\tilde{\omega}_q^v[3]. \tag{B.4}$$

Then (3.4) may be calculated using (B.1), (B.2) and (B.4).

Vita

Juan Zhang

- **EDUCATION**

2000   B. S. in Mathematics, Fudan University, Shanghai, China.

2003   M. S. in Statistics, Fudan University, Shanghai, China.

2007   M. S. in Statistics, Rutgers University.

2008   Ph. D. in Statistics, Rutgers University.

- **PROFESSIONAL EXPERIENCE**

2003-2004   Fellowship, Department of Statistics and Biostatistics, Rutgers University.

2004-2007   Teaching assistant/Graduate assistant, Department of Statistics and Biostatistics, Rutgers University.

2005-2006   Assistant consultant, Office of Statistical Consulting, Department of Statistics and Biostatistics, Rutgers University.

2005-2007   Statistical associate. College of Nursing, Rutgers University.

Spring 2006   Instructor, Department of Statistics and Biostatistics, Rutgers University.

Spring 2007   Instructor, Department of Statistics and Biostatistics, Rutgers University.

Summer 2007   Intern, Department of Discovery and Preclinical, Sanofi-Aventis.

2007-2008   Contractor statistician, New York University, College of Nursing.

Spring 2008   Instructor, Department of Statistics and Biostatistics, Rutgers University.

- **PUBLICATIONS**

2002 The Error-Density Estimator of Nonparametric Regression in Continuous Time Processes, *Mathematica Applicata*, 2002, Vol. 15, No. 4, 62-66, Jia Shen and Juan Zhang.

2003 Uniform Convergence Rate of an Error-density Estimator of Nonparametric Regression in Continuous Time Processes, *Mathematica Applicata*, 2003, Vol. 16, suppl. 35-39, Juan Zhang, Jia Shen and Guodong Hui.

2007 Using Crash Scene Variables to Predict the Need for Trauma Center Care in Older Persons, *Research in Nursing & Health*, 2007, Vol. 30, 399-412, Linda J. Scheetz, Juan Zhang and John E. Kolassa.

2007 A Comparison of the Accuracy of Saddlepoint Conditional Cumulative Distribution Function Approximations, *IMS-LNMS*, 2007, Vol. 54, 250-259, Juan Zhang and John E. Kolassa.

2008 Saddlepoint Approximation for the Distribution of the Modified Signed Root of Likelihood Ratio Statistics near the Mean, *Communications in Statistics – Theory and Methods*, 2008, Vol. 37, issue 2, 194-203, Juan Zhang and John E. Kolassa.

2008 Evaluation of Injury Databases as a Preliminary Step to Developing a Triage Decision Rule, *Journal of Nursing Scholarship*, 2008; 40:2, 144C150, Linda J. Scheetz, Jane Zhang, John E. Kolassa, Patricia Allen, Michael Allen.