AN ABSTRACT OF THE THESIS OF

Benjamin Lyons for the degree of Doctor of Philosophy in Statistics presented on October 14, 1997. Title: Applying Higher Order Asymptotics to Mixed Linear Models.

Redacted for privacy

Abstract approved. _____

Dawn Peters

Mixed linear models are a time honored method of analyzing correlated data. However, there is still no method of calculating exact confidence intervals or p-values for an arbitrary parameter in any mixed linear model. Instead, researchers must use either specialized approximate and exact tests that have been developed for particular models or rely on likelihood based approximate tests and confidence intervals which may be unreliable in problems with small sample sizes. This thesis develops procedures to improve small sample likelihood based inference in these important models.

The first manuscript develops I.M. Skovgaard's modified directed likelihood for mixed linear models and shows how it is a general, accurate, and easy to apply method of improving inference in mixed linear models. In the second manuscript, O.E. Barndorff-Nielsen's approximate modified profile likelihood is applied to mixed linear models. This modified profile likelihood is a sensible generalization of the commonly used residual likelihood and can be applied if either a fixed or a covariance parameter is of interest. The final manuscript discusses how the design of a mixed linear model effects the accuracy of Skovgaard's modified likelihood and suggests a useful decomposition of that statistic.

Applying Higher Order Asymptotics to Mixed Linear Models

by

Benjamin Lyons

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented October 14, 1997
Commencement June 1998

Doctor of Philosophy thesis of Benjamin Lyons presented on October 14, 1997

APPROVED:

Redacted for privacy

Major Professor, representing Statistics

Redacted for privacy

Chair of Department of Statistics

Redacted for privacy

Dean of the Graduate School

I understand that my thesis will become part of the permanent collection at Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

Redacted for privacy

Benjamin Lyons, Author

ACKNOWLEDGMENTS

Much of the credit for this thesis belongs to Dawn Peters. I am indebted to Dr. Peters for suggesting this interesting topic, sharing her knowledge of higher order asymptotics and patiently reviewing numerous manuscripts. I would also like to thank Dr. Donald Pierce for many useful suggestions and comments.

I am grateful to Dr. David Birkes for sharing his understanding of mixed linear model theory and his contributions to the third manuscript. I would also like to thank Dr. Cliff Pereira for many useful discussions about this work and mixed linear models in general. Finally, this thesis would not have been possible without the use of the department's computing facilities and the technical assistance of Anne Russ and Russ Ruby.

Dr. Justus Seely and Dr. Scott Urquhart helped me a great deal by offering me funding these last six years and I especially would like to thank Dr. Pereira and Dr. Urquhart for making my assistantships worthwhile. These practical experiences have greatly helped me develop as a statistician.

Of course Kathleen Purdy should take credit for anything I do and I would like to thank her for reviewing the manuscripts and putting up with me. Finally, I'd like to thank my parents, Judy and David Lyons, and my sister, Dayna Lyons, for supporting my education over these many years and seeing me through to the twenty second grade.

CONTRIBUTION OF AUTHORS

Dr. Dawn Peters proposed applying higher order methods to these models. She was involved in the development and editing of each manuscript. Dr. Dave Birkes assisted in the development of the results in the third manuscript and also edited that manuscript.

TABLE OF CONTENTS

List of Figures

List of Tables

In memory of Gerry Lyons and Moez Ali-Mohammed,
two men who did not deserve the end they got.

Applying Higher Order Asymptotics to Mixed Linear Models

1. Introduction

1.1 Introduction

Researchers have a variety of methods available to obtain p-values and confidence intervals in Gaussian mixed linear models. In addition to exact and approximate methods based on sums of squares, commercial optimization programs that give maximum likelihood estimates have made likelihood based tests widely available. Likelihood based tests such as the likelihood ratio test (LRT) offer a general method of making inference about any parameter in any mixed model. However, like some sums of squares based methods, these likelihood based tests are approximate and can be misleading when the sample size is small compared to the complexity of the model. This thesis presents methods of adjusting the likelihood to improve small sample inference. Based on recent results in higher order asymptotics, the methods are straight forward to implement in any model with current commercial software.

Specifically, the thesis shows how to apply three higher order asymptotic procedures for making inference in Gaussian mixed linear models: the modified directed deviance statistic introduced by I.M. Skovgaard (1996), the modified profile likelihood, defined by O.E. Barndorff-Nielsen (1983), and Barndoff-Nielsen's approximate modified profile likelihood (1994b). Although our primary concern is to develop practical methods for improving small sample likelihood inference, more fundamental issues of inference will be addressed as well.

After defining the mixed linear model and discussing available tests and confidence interval procedures, this chapter ends with a more technical discussion that places the Gaussian linear mixed model in the context of exponential families. The

second and third manuscript chapters concern Skovgaard's modified directed deviance

statistic, $\tilde{r}$, and Barndorff-Nielsen's approximate modified profile likelihood, $L_{mp}^{\dagger}$,

respectively. The final manuscript chapter discusses a useful decomposition of $\tilde{r}$ and

some technical notes on its accuracy.

1.2 The Gaussian Mixed Linear Model

Encompassing both linear regression and analysis of variance procedures, the

standard Gaussian linear model is arguably the most widely used statistical method. In

these models an $n \times 1$ response vector Y is modeled as a linear function of unknown

slope and intercept parameters and a normally distributed random error term with

independent components. In matrix form,

$$Y = X\beta + \varepsilon$$

where X is an $n \times p$ matrix of constants, β is an unknown vector of parameters and the

error term ε is distributed multivariate normal with 0 mean and covariance matrix $\sigma^2 I$.

In many applications the assumption of independence is not appropriate. In a

repeated measures model Y consists of multiple measurements taken over time on a

group of subjects so that the inter-subject measurements are dependent. For this case and

others a generalization of the standard linear model is useful:

$$Y = X\beta + \varepsilon \text{ and } \varepsilon \sim N_n(0, V(\rho)),$$

where $V(\rho)$ is a symmetric positive definite covariance matrix. Researchers identify

these models by the form of the covariance matrix; e.g., variance component models

(Searle, Casella, McCulloch 1992), random coefficient models (Longford 1993), and

multivariate analysis of variance models (Johnson and Wichern 1982).

Our main interest lies with the common case of Y consisting of m independent

sub-vectors, i.e., V is block diagonal. This common type of covariance matrix arises in

nested variance component models and repeated measures models. Since Y consists of

sets of independent vectors, the asymptotic properties of $\hat{\tau}$ and $L_{mp}^\dagger$ are straight forward to derive, i.e., their behavior stabilizes as the number of independents units, m, increases. Through out we will assume that $V(\rho)$ is block diagonal; Miller (1977), Jiang (1996) and Cressie and Laihari (1993) discuss the behavior of asymptotic tests when V is not block diagonal.

1.3 Inference in Gaussian Mixed Linear Models

Statisticians have proposed a variety of methods to obtain tests and confidence intervals for mixed models that for convenience we will categorize as follows: exact procedures, specialized approximate procedures, and approximate procedures based on the likelihood. Of these methods, only the likelihood based tests are available for inference about any parameter in any mixed linear model. Without a general method, researchers may choose to only fit models with exact tests even if other models offer a better fit. The likelihood ratio test (LRT), unlike the Wald test, has the important property of parameterization invariance so we will focus on adjusting it in hopes of obtaining a more accurate and general inference tool.

Exact procedures have been developed for two types of models: balanced variance component models and multivariate analysis of variance models with no missing data. To obtain tests and confidence intervals when exact methods are not available, statisticians have developed several methods that extend the exact F-tests. Separate methods were developed for inference about covariance parameters and fixed effects.

1.3.1 Covariance Parameter Inference

For variance component models, Mathew and Sinha (1988) and Seely and El-Bassiouni (1983) describe exact optimal F-tests for variance components. These tests are optimal in the sense that they are most powerful among either invariant or unbiased tests

(Lehmann 1986). However, in contrast to likelihood based tests, these procedures are invalidated by the loss of even one observation or the addition of a continuous fixed covariate. Furthermore even in the "ideal" balanced case, there is no method for constructing exact confidence intervals for linear combinations of variance components (Burdick and Graybill 1992).

For inference about variance components in unbalanced designs, Satterthwaite (1941) introduced an approximate F-test. Although quite accurate in simple unbalanced models, in models with several variance components the test is not unique; one can construct several sensible approximate F-tests for the same parameter. In some cases more accurate specialized approximations based on sums of squares exist (Ting et. al. 1991) but they too are not well defined in unbalanced models. The analyst then must choose between several approximate procedures which may give contradictory results.

Both the LRT and the Wald are well defined in variance component models but they may sometimes be inaccurate for small or moderate sample inference. Many researchers prefer using a penalized likelihood called the residual, or restricted, likelihood for making inference on covariance parameters. Introduced by Patterson and Thompson (1973) the residual likelihood eliminates the nuisance fixed effects and gives unbiased estimates of the variance components in balanced models. Li, Birkes, and Thomas (1996) present a numerical study comparing the residual likelihood ratio to another exact, but not optimal, test in a one way analysis of variance model and find that it performs quite well. In chapter three we show that this residual likelihood is a special case of this approximate modified profile likelihood.

1.3.2 *Fixed Effect Inference*

Currently, exact methods are available for most parameters of interest in multivariate repeated measures without missing data and balanced variance component

models. However, in models with unbalanced data, missing data, or covariates, researchers have to rely on approximate methods. One popular approximate method is the generalized least squares (GLS) t-test. Analogous to the t-test constructed in the standard linear model, the GLS t is actually a modification of the Wald statistic that need not follow a t distribution under the null hypothesis.

Calculation of the GLS t and F statistics requires two stages: one must first estimate $\widehat{\rho}$ and then calculate the GLS estimates of β as

$$\widehat{\beta} = (X'V^{-1}(\widehat{\rho})X)^{-1}X'V^{-1}(\widehat{\rho})Y.$$

The covariance matrix of $\widehat{\beta}$ is then estimated with $X'V^{-1}(\widehat{\rho})X$. A test statistic analogous to the simple t-test for testing $\beta_i = \beta_{oi}$ is given by

$$\text{GLS } t = \frac{\widehat{\beta_i} - \beta_{oi}}{SE(\widehat{\beta_i})}.$$

The variance of $\widehat{\beta_i}$ is estimated with the ith diagonal element of $X'V^{-1}(\widehat{\rho})X$. If the MLE is used to estimate ρ then the GLS t reduces to the Wald test with the expected (Fisher) information used to estimate the standard error. If other consistent estimators of ρ are used, such as REML estimates, then in large samples the GLS t generally follows a normal distribution. For small samples, however, it is often compared to a t distribution the degrees of freedom of which are calculated in a variety of ways.

The estimator of the standard error of $\widehat{\beta_i}$ given above usually underestimates the true standard error (Kackar and Harville 1984) so the p-values obtained from the GLS t are too small and the confidence intervals too narrow when the GLS t is compared to a normal distribution. To remedy this, McLean, Sanders, and Stroup (1991) argue that the GLS t should be compared to a t distribution, rather than a normal, with degrees of freedom determined by a Satterthwaite approximation. Although this ignores the correlation of $SE(\widehat{\beta_i})$ and $\widehat{\beta_i}$, the resulting test is exact for many balanced variance component models. The SAS® Institute implements this method with its MIXED procedure. Kackar and Harville suggest using a Taylor series approximation to reduce

the first order bias of the standard error estimates. However, neither this adjustment nor the approximate degrees of freedom method appear to be invariant to reparameterizations of V.

The GLS statistic requires estimates of ρ. Using ML estimates gives the standard Wald test but many researchers prefer to use residual maximum likelihood (REML) estimates since they are believed to be less biased and yield exact t-tests in some balanced models. Another point of contention is whether the statistic should be compared to a t distribution or a normal and when using a t distribution how should the degrees of freedom be determined. Generally, the LRT is considered superior to the Wald test. As the two manuscript chapters will demonstrate the LRT, adjusted for small sample accuracy, can perform much better than the GLS t.

For making inference about fixed effects in multivariate repeated measures models Hotelling's T gives exact confidence intervals and p-values. When there is missing data, Pillai and Samson (1954) and McKeon (1971) provide approximate extensions of Hotelling's T statistic. S.P. Wright (1994) reported that these tests are superior to Wald type tests with REML variance estimates. However, using these tests requires fitting an unstructured covariance matrix in lieu of another more parsimonious repeated measures model, such as, an autoregressive model, that may offer a better fit. We consider a multivariate repeated measures problem in chapter two.

Specialized tests exists in particular models for inference about some parameters but the LRT is defined in *all* models for *any* parameter of interest. Recent advances in software have made maximum likelihood estimates available for a wide variety of mixed models so that the LRT is now a feasible alternative to the specialized non-likelihood based approximations. However, some numerical studies (Lyons and Peters 1996 and Wright 1994) indicate that these specialized approximations can be more accurate than both the likelihood ratio test and the Wald tests for some models. In practice researchers too often resort to fitting models that offer exact or specialized approximate tests even if

these models fail to reflect the experimental design, do not answer the question of interest, or fit the data poorly. Furthermore, applying these procedures requires specialized knowledge and judgment which distracts researchers from their principle task of fitting and interpreting useful models.

By employing likelihood ratio tests that are adjusted to be at least as reliable as the specialized approximations researchers could fit models without considering the existence of exact or specialized approximate tests. This would also place mixed linear model analyses in a general context since most of the specialized methods above are peculiar to mixed linear models while the likelihood ratio test is applied successfully in many contexts, e.g., generalized linear models. In the next two chapters we will develop these adjustments. First though, we need to place mixed models in the context of exponential families.

1.4 Exponential Families and the Gaussian Mixed Linear Model

Some authors view mixed models in the general context of random effects or even as extensions of standard linear models. However, to understand how to apply higher order methods, the Gaussian mixed linear model is best viewed as a member of a regular or curved exponential family. It is convenient to adopt notation similar to that of Barndorff-Nielsen and Cox (1994).

Let $\theta = (\beta, \rho)$ be partitioned into a parameter of interest, ψ, and a nuisance parameter, χ. We will write the likelihood of θ as $L(\theta)$ and its log as $l(\theta)$. Differentiation will be denoted with a sub-scripted / followed by the parameter, e.g., the score $U(\theta) = l_{/\theta}$ and the observed information $j(\theta) = -l(\theta)_{/\theta\theta}$. Evaluation of a quantity at the maximum likelihood estimates, $(\widehat{\psi}, \widehat{\chi})$, or at the restricted maximum likelihood estimates , $(\psi, \widetilde{\chi})$, is denoted by a superior $\frown$ and $\sim$, respectively; e.g., the profile log

likelihood for ψ, $l_p(\psi) = l(\psi, \widetilde{\chi})$, can be written as $\widetilde{l}$. For a matrix X, $R(X)$ will denote the range space, $r(X)$ its rank, $tr(X)$ its trace and $|X|$ its determinant.

Suppose that the random vector Y has density f indexed by the parameter ω of dimension p. The density $f(Y; \theta)$ follows an exponential family if it is of the form:

$$f(Y; \theta) = h(Y)exp(\omega(\theta)' t(y) - \kappa(\omega(\theta))) \qquad (1)$$

relative to some dominating measure. Here $t(y)$, the canonical statistic, and $\omega(\theta)$, the canonical parameter, are vectors of dimension k. The smallest natural number k for which (1) holds is called the order of this (k, p) exponential family. The statistic $t(y)$ is a minimal sufficient statistic (see Lehmann 1986).

If $k = p$ then f is generally a member of a regular exponential family; if $k \geq p$ then f is a member of a curved exponential family. The binomial, gamma, poison and normal densities all belong to regular exponential families. Most balanced variance component models are members of a regular exponential family while the densities of unbalanced models generally belong to curved families. If $k \leq p$ then the model is not useful since it is over-parameterized.

To see that the density of Y in a mixed model follows an exponential family note that

$$lnf(Y; \beta, \rho) = -\tfrac{1}{2}((Y - X\beta)'V^{-1}(\rho)(Y - X\beta) + ln|V(\rho)|).$$

Hence if we let v_{ij} denote the (i, j) element of $V^{-1}(\rho)$ we see that,

$$lnf(Y; \beta, \rho) =$$
$$-\tfrac{1}{2}(\sum_{i=1}^{n} \sum_{i<j}^{n} v_{ij}(Y - X\beta)'(Y - X\beta) + ln|V(\rho)|).$$

The minimal representation, and hence, the order will depend on the form of $V^{-1}(\rho)$.

As an example consider one of the simplest mixed models, the one-way random variance component model arising when k_i measurements are taken on each of a clusters. Consider the $n \times 1$ vector Y:

$$Y = 1_n\mu + \varepsilon$$

where $\varepsilon \sim N_n(0, V(\tau, \chi))$ and

$$V(\tau, \chi) = \tau I_n + \chi(\overset{a}{\underset{i=1}{\oplus}} 1_{k_i} 1'_{k_i}).$$

By inspection we see that

$$V^{-1} = \frac{1}{\tau} P_o + \sum_{i=1}^{s} \frac{1}{\tau + k_i \chi} P_i$$

where $P_o = \overset{S}{\underset{i=1}{\oplus}} (I_{k_i} - \frac{1}{k_i} 1_{k_i} 1'_{k_i})$ and P_i is a block diagonal matrix with the ith block

given by $\frac{1}{k_i} 1_{k_i} 1'_{k_i}$. This implies that

$$(Y - \mu 1_n)' V^{-1}(\tau, \chi)(Y - \mu 1_n)$$

$$= \frac{1}{\tau} Y' P_o Y + \sum_{i=1}^{s} \frac{1}{\tau + k_i \chi} Y' P_i Y - \frac{2\mu}{\tau} 1'_n P_o Y$$

$$- \sum_{i=1}^{s} \frac{2\mu}{\tau + k_i \chi} 1'_n P_i Y + \mu^2 1'_n V^{-1} 1'_n$$

$$= \frac{1}{\tau} Y' P_o Y + \sum_{i=1}^{s} \frac{1}{\tau + k_i \chi} Y' P_i Y - \sum_{i=1}^{s} \frac{2\mu}{\tau + k_i \chi} 1'_{k_i} Y + \mu^2 1'_n V^{-1} 1'_n.$$

The last step follows since $1'_n P_o = 0$.

The order of the family depends on the number of distinct k_is. If $k_i = c$ for all i
then the model is said to be balanced and the density forms a (3,3) exponential family. If
there are two distinct k_is the density forms a (5,3) exponential family. Generally, w
distinct k_is leads to a $(1 + 2*w, 3)$ exponential family. A one-way analysis of variance
model with two distinct k_is then differs substantially from a model with one distinct k_i.
In fact from the point of view of finding optimal tests or applying higher order methods
the densities of these two models differ in a more fundamental way than the densities of a
Binomial and Poisson, both of which follow a regular exponential family. This one-way
model will be considered again in chapter two.

The sufficiency of the MLEs in the regular exponential family case not only
leads to a convenient reduction of the data but to a compact approximation with third
order relative error for the density of the MLE: the p* formula given by Barndorff-
Nielsen (Banrdorff-Nielsen and Cox 1989, ch. 9)

$$\text{p*}(\hat{\theta}; \theta) = c(\theta)|\hat{j}|^{-\frac{1}{2}} e^{l - \hat{l}}.$$

Here $c(\theta)$ is a norming constant that insures p* integrates to 1. For regular exponential families the p* formula leads to both the modified profile likelihood, L_{mp}, and r^* statistic.

Since in curved exponential families $\widehat{\theta}$ is not sufficient, Barndorff-Nielsen and Cox (1994 ch. 7) present a second order relative error approximation for the density of $\widehat{\theta}$ conditioned on an approximate ancillary statistic A:

$$\text{p*}(\widehat{\theta};\theta|A) = c(\theta,A)|\widehat{j}|^{-\frac{1}{2}}e^{l-\widehat{l}}.$$

Only the norming constant requires the specification of the ancillary statistic. This decrease in precision from third to second order for curved families occurs in the statistics discussed in chapters two and three as well. In practice, though, the order is only a rough indicator of accuracy since the actual magnitude of the error depends on unknown parameter values.

Sufficient conditions for the density of a mixed linear model to follow a regular family are given in Chapter 4 but generally they follow curved families. The fact that the MLEs alone are not sufficient in the curved case makes applying higher order methods difficult since an ancillary statistic must be specified. Neither the score nor the directed deviance ancillary are unique and calculating either r^* or L_{mp} while accounting for an arbitrary ancillary statistic is difficult. Instead, Barndorff-Nielsen and Cox (1994) suggest using methods that are stable in the sense that asymptotically they give the same inference that would have been obtained by conditioning on any reasonable ancillary. In doing this a degree of precision is lost but the approximations are relatively easy to calculate and can greatly improve on existing methods. Finding stable higher order methods for mixed linear models is the main focus of this thesis.

Chapter 2
A Simple Higher Order Asymptotic Test for Mixed Linear Models

Benjamin Lyons and Dawn Peters

2. A Simple Higher Order Asymptotic Test for Mixed Linear Models

2.1 Abstract

The introduction of software to calculate maximum likelihood estimates for mixed linear models has made likelihood estimation a practical alternative to methods based on sums of squares. Likelihood based tests and confidence intervals, however, may be misleading in problems with small sample sizes. This paper discusses an adjusted version of the directed likelihood statistic for mixed models that can improve the accuracy of the likelihood ratio test for any one parameter hypothesis. Introduced in general form by Skovgaard (1996), in mixed models this statistic has a simple compact form so that it is easy to obtain with existing software. Several simulations studies indicate this statistic is more accurate than several specialized procedures that have been advocated. Skovgaard's statistic, however, is available in mixed models where specialized exact and approximate procedures have not been developed so that it offers an automatic method of improving inference in many Gaussian mixed linear models.

2.2 Introduction

Gaussian mixed linear models are often used to model correlated data and a variety of exact and approximate methods have been introduced to obtain p-values and confidence intervals in some common models. Exact and approximate procedures based on sums of squares are available in balanced and unbalanced variance component models and similar approximations have been developed for multivariate analysis of variance models. With the introduction of commercial programs to find maximum likelihood estimates, tests based on either the full or residual maximum likelihood estimates are now widely available. However, recent simulation studies, e.g., Wright (1994) and Wright and Wolfinger (1997), indicate that first order likelihood based methods may be

misleading. Furthermore, these likelihood based procedures are less reliable than some specialized approximations that are not widely available.

For moderate sample sizes researchers are then faced with a choice between unreliable likelihood based methods and specialized approximations that are not always available. This hinders the application of non-traditional correlation structures, such as the heterogenous covariance models discussed by Wolfinger (1996), where specialized approximations have not been developed. These more parsimonious models may be ignored in favor of correlation structures that admit reliable tests. Skovgaard (1996) introduced a modified directed likelihood statistic, called $\tilde{r}$, for curved exponential families that offers a practical solution to this problem. The $\tilde{r}$ statistic can substantially improve the accuracy of the standard likelihood based methods, is available for most models, and is straight forward to calculate with existing software.

We show below that $\tilde{r}$ has a compact closed form depending on full and restricted maximum likelihood estimates, observed and expected information, and other quantities based on simple derivatives of the covariance matrix. It is not substantially more complex then the wide variety of specialized approximations currently in use. Simulations given below indicate that this statistic is a marked improvement on first order likelihood methods and is often superior to specialized approximations when they exist.

We adopt notation similar to that used by Barndorff-Nielsen and Cox (1994) and partition the parameter vector θ into a parameter of interest, ψ, and a nuisance parameter, χ. The likelihood of θ will be written as $L(\theta)$ and its log as $l(\theta)$. Differentiation with respect to θ will be denoted by a sub-scripted parameter, e.g., the score $U(\theta)$ can be written as l_θ and the observed information $j(\theta)$ as $- l(\theta)_{\theta\theta}$. The expected information is denoted by $i(\theta) = E(j(\theta))$. Evaluation of a quantity depending on θ at the full maximum likelihood estimates , $(\hat{\psi}, \hat{\chi})$, or at the restricted maximum likelihood estimates, $(\psi, \tilde{\chi})$, is denoted by a superior $\hat{\ }$ and $\tilde{\ }$, respectively; e.g., the profile log

likelihood for ψ is denoted by $\tilde{l}$. For a matrix X, $R(X)$ will denote the range space, $r(X)$ its rank, $tr(X)$ its trace and $|X|$ its determinant.

In section 2 we review some existing procedures. Section 3 gives some background regarding Skovgaard's $\tilde{r}$ and the related statistic, Barndorff-Nielsen's r^*. Section 4 gives a compact form for $\tilde{r}$ in Gaussian mixed linear models. Section 5 gives simulation results and some technical details are presented in the appendix.

2.3 Inference in Mixed Linear Models

Of interest here are models of the form:

$$Y_n = X\beta + \varepsilon \text{ and } \varepsilon \sim N_n(0, V(\rho)), \tag{1}$$

where X is an n by p design matrix and V is a covariance matrix that is positive definite. Either the p-dimensional fixed effect vector β or the k dimensional covariance parameter vector ρ may be of interest. We are primarily concerned with cases, such as repeated measures models, where Y consists of m independent subsets and hence $V(\rho)$ is block diagonal and the likelihood based approximations improve as m gets large. It is easy to see that the density of Y belongs to an exponential family that may be either regular or curved.

In balanced variance component models a variety of exact F-tests can be formed using sums of squares for many, but not all, fixed and covariance parameters of interest. Besides being exact and easy to calculate, these tests are often optimal (Mathew and Sinha 1988). Even in this ideal case, however, exact methods do not exist for important parameters, such as the "between" variance component in one-way random analysis of variance models and certain fixed effect contrasts in balanced split plot models (Milliken and Johnson 1984 ch. 17). Similarly for many fixed effect hypotheses in multivariate analysis of variance (MANOVA) models, Hotelling's T is exact and optimal.

These exact methods can be extended to unbalanced models yielding tests that are usually approximate. In unbalanced variance component models, approximate F-tests can be calculated using the Satterthwaite procedure. Although these tests may perform well in simple models they often fail in more complex models especially when the numerator and denominator of the approximate F statistic are not independent (Tietjen 1974). Furthermore, in more complex unbalanced models there are usually several approximate F-tests and no method of choosing one over the other. In multivariate analysis of variance models with missing data, approximate tests for fixed effects based on Hotelling's T have also been developed and may be quite effective (see Pillai and Samson 1954 and Wright 1994).

These approximate methods do not extend easily to models with less traditional correlation structures. In these and other useful models, specialized approximate or exact methods have not been developed and only methods based on the full or residual likelihood are available. The lack of exact or specialized approximate methods may hinder the fitting of more parsimonious repeated measures models or random slope models. Skovgaard's directed likelihood statistic can be applied to one dimensional interest parameters in most of these non-traditional models.

2.4 Skovgaard's Modified Directed Likelihood

Developed for curved exponential families as an alternative to Barndorff-Nielsen's (1986) r^* statistic, Skovgaard's $\tilde{r}$, is simpler to calculate than r^* but slightly less accurate. Both statistics are modified versions of the directed likelihood statistics and they are identical in regular exponential families.

If the log likelihood ratio statistic for the single parameter of interest ψ is

$$W(\psi_0) = 2(l(\hat{\psi}, \hat{\chi}; Y) - l(\psi_0, \tilde{\chi}; Y)).$$

then the directed likelihood statistic (Barndorff-Nielsen and Cox 1994) for testing $\psi = \psi_0$ is

$$r(\psi_o) = \text{sgn}(\widehat{\psi} - \psi_0) \sqrt{W(\psi_0)}.$$

Quite generally, under the null hypothesis $r(\psi_o)$ is asymptotically standard normal with absolute first order error, $O_p(n^{-\frac{1}{2}})$, in moderate deviation regions and $O_p(1)$ error in large deviation regions.

To improve the accuracy of this approximation, Barndorff-Nielsen (1986) introduced the modified directed likelihood statistic r^*:

$$r^*(\psi_o) = r(\psi_o) - \frac{1}{r(\psi_o)}\log(\frac{r(\psi_o)}{u(\psi_o)}).$$

The statistic $u(\psi_o)$, defined below, is a function of the observed and expected information and two derivatives taken with respect to the data, i.e., sample space derivatives. This statistic is standard normal with relative third order error, $O_p(n^{-\frac{3}{2}})$, in moderate deviation regions and second order error, $O_p(n^{-1})$, in large deviation regions. Numerical studies, see Pierce and Peters (1992), indicate that r^* can be very accurate in small sample problems. Unfortunately, $u(\psi_o)$ is often difficult to calculate in curved exponential families since it depends on the specification of an ancillary statistic.

The improvement in the order of the error does not fully capture the benefits of using r^* or $\widetilde{r}$ instead of r. The accuracy of r and the Wald statistic decrease as models become more complex since they fail to account for nuisance parameters as the modified statistics do. For instance, when testing covariance parameter in mixed linear models, r^* and $\widetilde{r}$, like the residual likelihood, correct for the nuisance fixed effects. Unlike the residual likelihood, however, they also correct for nuisance covariance parameters. Similarly, when testing fixed effects, where there is no equivalent to the residual likelihood, r^* and $\widetilde{r}$ adjust for both nuisance fixed and covariance parameters. Furthermore, unlike the Wald test, the modified and unmodified directed likelihood

statistics are invariant to interest respecting reparameterizations (Barndorff-Nielsen and Cox 1994 pg. 11).

The statistic $u(\psi_o)$ is a function of the full MLEs, the restricted MLEs, the observed information and two sample space derivatives:

$$u(\psi_0) \;=\; [\,(\widetilde{l}_{\theta;\widehat{\theta}})^{-1}\,(\widehat{l}_{;\widehat{\theta}} - \widetilde{l}_{;\widehat{\theta}})\,]_\psi \,|\widehat{j}|^{-\frac{1}{2}}\,|\widetilde{l}_{\theta;\widehat{\theta}}|\,|\widetilde{j}_{\chi\chi}|^{-\frac{1}{2}},$$

where $\widetilde{j}_{\chi\chi}$ is the nuisance parameter block of the observed information. The sample space derivatives, $l_{\theta;\widehat{\theta}}$ and $l_{;\widehat{\theta}}$, are taken with respect to the data, specifically with respect to $\widehat{\theta}$. In the curved exponential family setting, an approximate ancillary statistic A can be found so that $\widehat{\theta}$ and A are jointly sufficient (see Barndorff-Nielsen and Cox 1994). The sample space derivatives, taken with A held fixed, are written as

$$l_{\theta;\widehat{\theta}} \;=\; \tfrac{\partial^2}{\partial\theta\partial\widehat{\theta}} l(\theta;\widehat{\theta}, A) \text{ and } l_{;\widehat{\theta}} \;=\; \tfrac{\partial}{\partial\widehat{\theta}} l(\theta;\widehat{\theta}, A).$$

In regular exponential families $\widehat{\theta}$ itself is sufficient and calculating the sample space derivatives is usually elementary. In curved exponential families, however, several sensible ancillary statistics could be specified, each giving a different version of r^*.

Skovgaard avoids specifying an ancillary statistic by approximating the sample space derivatives:

$$\widetilde{l}_{\theta;\widehat{\theta}} \;\overset{\dagger}{=}\; \widehat{S i}^{-1}\widehat{j} \text{ and } \widehat{l}_{;\widehat{\theta}} - \widetilde{l}_{;\widehat{\theta}} \;\overset{\dagger}{=}\; q\widehat{i}^{-1}\widehat{j}$$

where $\overset{\dagger}{=}$ indicates equality to second order. The terms S and q are given by

$$S \;=\; Cov_{\theta_1}(\,l_\theta(\theta_1),\, l_\theta(\theta_2))\Big|_{\theta_1=\widehat{\theta},\,\theta_2=\widetilde{\theta}} \text{ and }$$

$$q \;=\; Cov_{\theta_1}(\,l_\theta(\theta_1),\, l(\theta_1) - l(\theta_2))\Big|_{\theta_1=\widehat{\theta},\,\theta_2=\widetilde{\theta}}.$$

Here, evaluation at $\widehat{\theta}$ and $\widetilde{\theta}$ occurs after determining the covariance of the two terms. Applying S and q to the u statistic gives Skovgaard's $\widetilde{r}$. Skovgaard shows that $\widetilde{r}$ is standard normal with relative second order error in moderate deviation regions and first order error in large deviation regions. If the density of Y belongs to a regular exponential family, $\widetilde{r} = r^*$ and the higher order of accuracy is achieved. This is useful in the mixed model setting where it is often difficult to determine if the density belongs to a regular

family. Although $\tilde{r}$ is less accurate than r^* it is much easier to calculate and has a very compact form in mixed linear models.

2.5 Calculating Skovgaard's $\tilde{r}$ in a Gaussian Mixed Linear Model

In mixed linear models the score and the deviance drop are functions of quadratic and linear functions of Y, a multivariate normal random vector. In this case, using well known results concerning the covariance of quadratic and linear forms and multivariate random normal variables (Mathai and Provost 1992 ch. 3), it is elementary to show that S and q have a compact form. For convenience we will write $V_{\rho_j}^{-1} = \frac{\partial}{\partial \rho_j} V^{-1}$ and

$$V_{\rho_j} = \frac{\partial}{\partial \rho_j} V$$

We show in the appendix for the parameterization given in (1) that S and q have the following forms. By partitioning S

$$S = \begin{pmatrix} S_{\beta\beta} & S_{\beta\rho} \\ S_{\rho\beta} & S_{\rho\rho} \end{pmatrix}$$

it can be shown that the components are

$$S_{\beta\beta} = X'\tilde{V}^{-1}X,$$
$$[S_{\rho\rho}]_{ij} = \tfrac{1}{2}tr(\hat{V}_{\rho_i}\tilde{V}^{-1}\tilde{V}_{\rho_j}\tilde{V}^{-1}), \text{ for } i,j = 1, \ldots, k,$$
$$S_{\rho\beta} = 0,$$

and each of the k columns of $S_{\beta\rho}$ are given by

$$S_{\beta\rho_i} = X'\tilde{V}_{\rho_i}^{-1}X(\tilde{\beta} - \hat{\beta}), \text{ for } i = 1, \ldots, k.$$

Similarly, partition the vector q as

$$q = \begin{pmatrix} q_\beta \\ q_\rho \end{pmatrix}$$

and it can be shown that its components are

$$q_\beta = X'\tilde{V}^{-1}X(\hat{\beta} - \tilde{\beta})$$

with each of the k elements of q_ρ given by

$$[q_\rho]_j = -\tfrac{1}{2} tr(\hat{V}_{\rho_j}(\hat{V}^{-1} - \tilde{V}^{-1})) \text{ for } j = 1, \ldots, k.$$

Given full and restricted maximum likelihood estimates, the components of S and q are

then easy to calculate with existing software when V_{ρ_j} has a closed form. In certain

common models, such as balanced variance component models, the formulas for S and q

simplify.

Note that in all mixed models $S_{\beta\beta} = X'\widetilde{V}^{-1}X$, the fixed effect block of the

expected and observed information matrix. If the covariance matrix V is a linear function

of ρ, e.g., variance component models, then $S_{\rho\rho} = \widetilde{i}_{\rho\rho}$. To see this note that

$V_{\rho_j}^{-1} = - V^{-1} V_{\rho_j} V^{-1}$ (Graybill 1984 ch. 10) so that for known symmetric matrices

Q_i,

$$V = \sum_{i=1}^{k} \rho_i Q_i \Rightarrow$$

$$[S_{\rho\rho}]_{ij} = \tfrac{1}{2} tr(Q_i \widetilde{V}^{-1} Q_j \widetilde{V}^{-1}) = [\widetilde{i}_{\rho\rho}]_{ij} \text{ and}$$

$$[q_\rho]_j = - \tfrac{1}{2} tr(Q_j(\widehat{V}^{-1} - \widetilde{V}^{-1})).$$

Another interesting simplification occurs when interest lies with ρ. Clearly, if for

fixed ρ, $\widetilde{\beta} = \widehat{\beta}$ then $S_{\beta\rho} = 0$ and $q_\beta = 0$. Using the terminology of Lindsey (1996),

β is estimation orthogonal to ρ. In Gaussian mixed models this is implied by Zyskind's

(1967) condition: $R(V(\rho)X) \subset R(X)$ for all ρ.

2.6 Numerical Examples

We employ three models to illustrate the accuracy of Skovgaard's statistic: a

balanced variance component model, a repeated measures model with an unstructured

covariance matrix, and a repeated measures models with a heterogenous toeplitz

covariance matrix. In all three cases, exact tests do not exist for some parameters but

specialized approximations have been developed in the first two examples. The

simulations compare the empirical rejection rates given by $r, \widetilde{r}$, tests based on residual

maximum likelihood (REML) estimates, and several specialized approximate tests.

Results show that in each case $\widetilde{r}$ is a significant improvement on the likelihood based

approximations. In examples where they exist, the non-likelihood based tests are more accurate than the unmodified likelihood based tests but generally inferior to $\tilde{r}$

2.6.1 *Balanced Variance Component Model*

Consider the one-way balanced variance component model with normal errors:

$$Y_{ij} = \mu + d_i + \varepsilon_{ij} \text{ for } i = 1,\ldots,s, \text{ and } j = 1,\ldots,k,$$

where d and ε are independent normally distributed random vectors. The mean and covariance of the multivariate normal vector Y of, dimension n, are given by $E(Y) = \mu 1_n$ and $Var(Y) = \sigma_\varepsilon^2 I_n + \sigma_u^2 DD'$ where D is the $n \times s$ balanced classification matrix for the random effect d. In order for the covariance matrix of Y to be positive definite the eigenvalues of $Var(Y)$ must be positive, i.e., $\sigma_\varepsilon^2 > 0$ and $\sigma_\varepsilon^2 + k\sigma_u^2 > 0$. These are the only constraints placed on the variance components.

In this simple model, there is no exact method of obtaining confidence intervals on the "between" variance component, σ_u^2. A specialized approximation based on the Cornish-Fisher expansion, introduced by Ting et. al. (1991), is known to work well and is calculated here. The signed square root of the directed residual likelihood, REML r, is also considered. Simulation results given in Table 2.1 were obtained by using the MATLAB programming language with $s = 10, k = 2, \mu = 1$, and $\sigma_\varepsilon^2 = 1$. Rejection rates are given for testing $\sigma_u^2 = 0, 0.5$, and 1 against upper and lower alternatives in Table 2.1. Other simulations with $k = 3$ and 4 gave similar results.

The density of Y in this three parameter model is a member of a regular exponential family, hence for moderate sample sizes one might expect first order methods, such as r and REML r, to perform well. The results summarized in Table 2.1 indicate that these tests are very misleading especially for the smaller nominal levels. Skovgaard's $\tilde{r}$ statistic, equivalent to r^* here, however, performs quite well at all levels. Ting's test is exact when $\sigma_u^2 = 0$. In other cases it gives satisfactory results when the

nominal levels are large but is inferior to $\tilde{r}$ at the smaller nominal levels. The errors for r

and REML r are very asymmetric since both tests are too conservative for the lower

alternative and too liberal for the upper alternative. Ting's procedure and $\tilde{r}$ correct this

asymmetry. The Wald tests based on the full and residual likelihood are inferior to the

both directed likelihoods and so are omitted.

Table 2.1

Empirical rejection rates in 100,000 repetitions for testing $\sigma_u^2 = 0$, .5, and 1 for the one-way balanced variance component models.

σ_u^2	Test	Upper				Lower			
		10%	5%	1%	.1%	10%	5%	1%	.1%
0	r	8.316	4.114	.856	.114	14.864	8.326	2.114	.306
	REML r	10.445	5.365	1.168	.145	10.913	5.614	1.196	.135
	Ting	10.069	5.031	1.029	.131	10.126	5.004	.995	.106
	r^*	9.894	4.924	.999	.126	10.355	5.151	1.038	.111
.5	r	6.042	2.870	.546	.045	17.478	10.044	2.625	.351
	REML r	8.899	4.219	.815	.082	12.290	6.440	1.399	.160
	Ting	9.595	4.703	.846	.071	10.152	5.159	1.050	.129
	r^*	9.810	4.968	1.009	.111	10.117	5.084	.991	.110
1	r	5.477	2.535	.440	.034	18.498	10.607	2.796	.421
	REML r	7.934	3.838	.704	.062	12.734	6.738	1.460	.169
	Ting	9.621	4.684	.819	.065	10.205	5.124	1.064	.115
	r^*	9.738	4.889	.966	.089	10.234	5.115	1.042	.114

2.6.2 Repeated Measures Model

Gaussian mixed models are commonly applied to repeated measures data. To

illustrate the accuracy of $\tilde{r}$ we employ data presented by Pothoff and Roy (1964) which

consists of growth measurements of 16 boys and 11 girls at ages 8, 10, 12, and 14. For

this simulation study, the mean will be modeled as linear in age with a gender by age interaction:

$$E(Y_{ij}) = \mu + \beta_1 gender_i + \beta_2 age_j + \beta_3 (age \times gen)_{ij}$$

$$i = 1, 2 \text{ and } j = 1, 2, 3, 4.$$

In the first simulation, we fit an unstructured covariance matrix and empirical rejection rates are given for $\tilde{r}$. and four commonly used procedures. In the second simulation we fit a heterogenous toeplitz covariance matrix and compares the empirical rejection rates given by $\tilde{r}$, a Wald test and r. Both simulations were conducted in SAS® using its mixed model optimization procedure. The nuisance parameters were set equal to their MLEs for the Pothoff and Roy data.

In the first simulation an unstructured covariance matrix is fit where for each subject

$$V_k = \begin{pmatrix} \sigma_{11} & \sigma_{21} & \sigma_{31} & \sigma_{41} \\ \sigma_{21} & \sigma_{22} & \sigma_{32} & \sigma_{42} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \sigma_{43} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44} \end{pmatrix}, \quad k = 1, \ldots, 27.$$

Empirical rejection rates are compared for r, a Wald statistic using REML estimates of the covariance parameters, $\tilde{r}$ and two specialized approximations: the Hotelling-Lawly-Pillai-Samson trace (Pillai and Samson 1959) and the Hotelling-Lawley-McKeon (McKeon 1971) trace statistic. As suggested by Wright (1994) we employed REML estimates to calculate the HLM and HLPS statistics.

Table 2.2 displays the levels for testing $\beta_3 = -1$ against a two sided alternative. Errors from the stated level were nearly identical for the lower and upper alternatives. Both the REML Wald and the directed likelihood reject far too often. When the stated level is lower than 2.5% the REML Wald and the directed likelihood, both compared to a standard normal, reject over twice as often as they should. The directed likelihood is slightly less liberal than the REML Wald but still far inferior to the HLPS, HLM, and $\tilde{r}$ statistic. The two specialized approximate procedures are more conservative than the

likelihood based tests. However, their performance varies with the stated level and so it is difficult to choose between them. As in the other simulations, $\tilde{r}$ is more accurate than its competitors at nearly every level. Simulations for testing $\beta_3 = 1$ and 0 gave similar results.

Table 2.2

Empirical two sided rejection rates for testing $\beta_3 = -1$ with an unstructured covariance matrix given by 100,000 simulations.

β_3	Test	Level	10%	5%	2.5%	1%	.1%
-1	r		13.449	7.384	4.133	1.871	.255
	REML $Wald$		14.679	8.492	5.170	2.750	.594
	HLPS		11.913	5.814	2.701	0.881	.034
	HLM		11.502	5.979	3.069	1.253	.112
	$\tilde{r}$		10.193	5.181	2.587	1.020	.106

In our final example, we will evaluate Skovgaard's $\tilde{r}$ statistic in a model where no specialized approximations exist. The same means model as above is employed but the more parsimonious heterogenous toeplitz covariance matrix (Wolfinger 1996) is fit. Here the covariance matrix for each subject is

$$
V_k = \begin{pmatrix}
\sigma_1^2 & \sigma_1\sigma_2\rho_1 & \sigma_1\sigma_3\rho_2 & \sigma_1\sigma_4\rho_3 \\
\sigma_1\sigma_2\rho_1 & \sigma_2^2 & \sigma_2\sigma_3\rho_1 & \sigma_2\sigma_4\rho_2 \\
\sigma_1\sigma_3\rho_2 & \sigma_2\sigma_3\rho_1 & \sigma_3^2 & \sigma_3\sigma_4\rho_1 \\
\sigma_1\sigma_4\rho_3 & \sigma_2\sigma_4\rho_2 & \sigma_3\sigma_4\rho_1 & \sigma_4^2
\end{pmatrix}, k = 1, ..., 27.
$$

Skovgaard's $\tilde{r}$ is compared to the directed likelihood and the REML Wald. The empirical rejection rates are given in Table 2.3. In three cases the likelihood failed to converge and in two other cased $\tilde{r}$ could not be calculated since $\tilde{j}$ was nonsingular.

Even in this more parsimonious model the REML Wald and the directed likelihood, compared to a standard normal, reject the null hypothesis of $\beta_3 = -1$ far too often. The modified directed likelihood is substantially more accurate than either r or the REML Wald. It gives relatively highly accurate levels for the larger nominal levels and good accuracy at the lower nominal levels. Simulations for testing $\beta_3 = 0$ and 1 gave similar results.

Table 2.3

Empirical two sided rejection rates for testing $\beta_3 = -1$ with a heterogenous toeplitz covariance matrix given by 100,000 simulations.

β_3	Test	Level	10%	5%	2.5%	1%	.1%
-1	r		12.316	6.658	3.571	1.602	.200
	REML $Wald$		12.389	6.978	3.983	1.969	.357
	$\tilde{r}$		10.503	5.335	2.730	1.194	.146

2.7 Conclusion

Skovgaard's $\tilde{r}$ statistic is easy to apply to many useful Gaussian mixed models using existing software. The compact and general form of the adjustment makes it as simple to implement as many specialized approximations. For testing both fixed or covariance parameters, simulations indicate that $\tilde{r}$ gives substantially more accurate levels than standard likelihood based approximations and is generally superior to available specialized approximations. Skovgaard's statistic then offers a general and automatic method of improving likelihood based inference in mixed models. Further work remains to be done, however, to determine for which models and sample sizes $\tilde{r}$ gives adequate accuracy and power. It would also be useful to investigate the relationship between $\tilde{r}$ and the residual likelihood. For instance, if interest lies in the covariance parameters

$\tilde{r}$ could be constructed directly from the residual likelihood. It might then be useful to expand on Barndorff-Nielsen, Peters and Pierce's (1994) discussion of the relationship between r^* and Barndorff-Nielsen's modified profile likelihood.

2.8 References

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

Barndorff-Nielsen, O.E. and Cox, D.R. (1994). *Inference and Asymptotics.* Chapman and Hall, New York.

Barndorff-Nielsen, O.E., Pierce, D.A., and Peters, D. (1994). Modified directed likelihood: decomposition of the correction term. Unpublished manuscript.

Graybill, F.A. (1983). *Matrices with Applications to Statistics.* Wadsworth Publishing, Belmont, California.

Lindsey, J.K. (1996). *Parametric Statistical Inference.* Oxford University Press, Oxford.

Mathai, A.M. and Provost, S.B. (1992). *Quadratic Forms in Random Variables: Theory and Application.* Marcel Dekker, New York.

Mathew, T. and Sinha, S.B. (1988). Optimum tests for fixed effects and variance components in balanced models. *Journal of the American Statistical Association,* 83, 133-135.

McKeon, J. (1974). F approximations to the distribution of Hotelling's T_o^2. *Biometrika,* 61, 381-383.

Milliken, G.A. and Johnson, D.E. (1984). *Analysis of Messy Data. Volume 1: Designed Experiments.* Van Nostrand Rheinhold Company, New York.

Pierce, D.A. and Peters, D. (1992). Practical use of higher order asymptotics for multiparameter exponential families. *Journal of the Royal Statistical Society Series B,* 54, 701-737.

Pillai, K.C. and Samson. P. (1959). On Hotelling's generalization of T^2. *Biometrika,* 46, 160-171.

Pothoff, R.F. and Roy, S.N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika*, 51, 313-326.

Skovgaard, I.M. (1996). An explicit large-deviation approximation to one-parameter tests. *Journal of the Bernoulli Society*, 2, 145-165.

Tietjen, G.L. (1974). Exact and approximate tests for unbalanced random effects designs. *Biometrics*, 30, 573-581.

Ting, N., Burdick, R.K., Graybill, F.A., Jeyaratnam, S., and Lu T.-F.C. (1990). Confidence intervals on linear combinations of variance components that are unrestricted in sign. *Journal of Statistical Computation and Simulation*, 35, 135-141.

Wolfinger, R.D. (1996). Heterogenous variance-covariance structures for repeated measures. *Journal of Agricultural, Biological, and Environmental Statistics*, 1, 205 - 230.

Wright, S.P. (1994). Adjusted F tests for repeated measures with the MIXED procedure. 328 SMC-Statistics Department, University of Tennessee.

Wright, S.P. and Wolfinger, R.D. (1997). Repeated measures analysis using mixed models: some simulation results. *Modeling Longitudinal and Spatially Correlated Data, Methods, Applications and Future Directions* (Des TG Groggier, D.R. Brillinger, J.D. Giggle, E. Russek-Cohen, W.G. Warren, R.D. Wolfinger). Lecture Notes in Statistics 122. Springier, New York.

Zyskind, G. (1967). On canonical forms, non-negative covariance matrices and best and simple least square linear estimators in linear models. *Annals of Mathematical Statistics*, 38, 1092-1109.

2.9 Appendix

First we show that $S(\psi)$ has form the given in section 4.

Claim 1: $S(\psi) = \begin{pmatrix} S_{\beta\beta} & S_{\beta\rho} \\ S_{\rho\beta} & S_{\rho\rho} \end{pmatrix}$ where

$$S_{\beta\beta} = X'\widetilde{V}^{-1}X,$$

$$S_{\rho\beta} = 0,$$

$$S_{\rho_r\rho_l} = \tfrac{1}{2}tr(\widehat{V}_{\rho_r}\widetilde{V}^{-1}\widetilde{V}_{\rho_l}\widetilde{V}^{-1})$$

and for $r, l = 1, \ldots, k$ each column of $S_{\beta\rho}$ is given by

$$S_{\beta\rho_r} = X'\widetilde{V}_{\rho_r}^{-1}X(\widetilde{\beta} - \widehat{\beta}) \quad \text{for } r = 1, \ldots, k.$$

Proof: The score equations are proportional to

$$[U_\beta(\beta, \rho)]_i \propto Y'V^{-1}X_i - \sum_{j=1}^{p} X_i'V^{-1}X_j\beta_j \text{ for } i = 1, \ldots, p \text{ and}$$

$$[U_\rho(\beta, \rho)]_j \propto -\tfrac{1}{2}((Y - X\beta)'V_{\rho_j}^{-1}(Y - X\beta) + ln|V|_{\rho_j})$$

$$\text{for } j = 1, \ldots, k,$$

where $ln|V|_{\rho_j} = \frac{\partial}{\partial\rho_j} ln|V|$. In both cases the score statistics are functions of linear and quadratic forms of Y. The covariance of terms like these are well known (Mathai and Provost 1992 ch. 3) and the result follows directly. $\square$

Claim 2: $q = \begin{pmatrix} q_\beta \\ q_\rho \end{pmatrix}$

with components given by $q_\beta = X'\widetilde{V}^{-1}X(\widehat{\beta} - \widetilde{\beta})$ and q_ρ where the k elements of this vector are given by

$$[q_\rho]_j = -\tfrac{1}{2} tr(\widehat{V}_{\rho_j}(\widehat{V}^{-1} - \widetilde{V}^{-1})) \text{ for } j = 1, \ldots, k.$$

Proof: Note that the deviance drop

$$\widehat{l} - \widetilde{l} = -\tfrac{1}{2}((Y - X\widehat{\beta})'\widehat{V}^{-1}(Y - X\widehat{\beta}) + ln|\widehat{V}|)$$

$$+ \tfrac{1}{2}((Y - X\widetilde{\beta})'\widetilde{V}^{-1}(Y - X\widetilde{\beta}) + ln|\widetilde{V}|).$$

Like the score, this is a function of quadratic and linear forms of Y. $\square$

Chapter 3

A General Modified Profile Likelihood for Gaussian Mixed Linear Models

Benjamin Lyons and Dawn Peters

3. A General Modified Profile Likelihood for Gaussian Mixed Linear Models

3.1 Abstract

In mixed linear models researchers often use a penalized version of the profile likelihood to make inference about covariance parameters since this residual likelihood does not involve the nuisance fixed effects. This paper presents Barndorff-Nielsen's approximate modified profile likelihood, $L_{mp}^{\dagger}$, as a generalization of the residual likelihood, that can correct the profile likelihood for either fixed or covariance nuisance parameters. This adjusted profile likelihood is a second order approximation to Barndorff-Nielsen's modified profile likelihood, L_{mp}. Straightforward to obtain with commercial software, the approximate modified profile likelihood corrects the bias in the score of the profile likelihood and equals L_{mp} when the density is a member of a regular exponential family.

3.2 Introduction

The introduction of software to obtain maximum and residual maximum likelihood (REML) estimates in mixed models has made inference based on the profile likelihood a viable alternative to more traditional methods that depend on sums of squares. Unfortunately, in many problems inference based on the profile likelihood is hindered by small sample sizes and large numbers of nuisance parameters which make the profile likelihood misleading. Hence, many researchers use the residual likelihood for inference concerning covariance parameters since it accounts for the nuisance fixed effects. This paper presents a generalization of the residual likelihood that is easy to construct for an arbitrary parameter of interest in any Gaussian mixed model.

Barndorff-Nielsen's approximate modified profile likelihood, $L_{mp}^{\dagger}$, is a second order approximation to Barndorff-Nielsen's modified profile likelihood, L_{mp}, which often

approximates the conditional or marginal likelihood appropriate for inference to third order (Barndorff-Nielsen and Cox 1994). It is straightforward to calculate requiring only restricted MLEs and derivatives of the covariance matrix and offers a sensible method of correcting the profile likelihood for both fixed and covariance nuisance parameters. The score of the $L_{mp}^{\dagger}$ statistic is asymptotically unbiased so $L_{mp}^{\dagger}$ behaves more like a true likelihood than the unadjusted profile likelihood. In regular exponential families, $L_{mp}^{\dagger}$ is equal to L_{mp}.

Of interest here are models of the following form:

$$Y_n = X\beta + \varepsilon \text{ and } \varepsilon \sim N_n(0, V(\rho)) \tag{1}$$

where X is an n by p design matrix and V is a positive definite covariance matrix. Here either the p-dimensional fixed effects vector β or the k dimensional covariance parameter vector ρ may be of interest. We are primarily concerned with cases, such as repeated measures models, where Y consists of m independent subsets, and hence, $V(\rho)$ is block diagonal and the approximations improve as m gets large. It is easy to see that the density of Y belongs to an exponential family that may be either regular or curved.

When the entire vector ρ is of interest, Patterson and Thompson (1971) suggest maximizing the residual likelihood, that is, the likelihood constructed from the density of $Q'Y$ where Q is an n by $n - p$ matrix that satisfies

$$QQ' = I_n - X(X'X)^- X'.$$

The density of $Q'Y$ depends only on ρ. In balanced variance component models the residual maximum likelihood (REML) estimates are, unlike the standard maximum likelihood estimates, equal to the unbiased ANOVA estimates. Numerical studies conducted by Swallow and Monahan (1984) indicate that the bias reduction extends to unbalanced variance component models as well.

Harville (1974) showed that the residual likelihood is equal to the product of the profile likelihood for ρ and a penalty term consisting of the fixed effect block of the observed information:

$$L_{REML}(\rho) = L_p(\rho)|X'V^{-1}(\rho)X|^{-\frac{1}{2}}.$$

Because the fixed effects and the covariance parameters are information orthogonal, i.e.,

$$E\left(-\frac{\partial lnL(\beta,\rho)}{\partial\beta\partial\rho}\right) = 0,$$

the residual likelihood also equals the approximate conditional likelihood introduced by Cox and Reid (1987).

In contrast to the profile likelihood which can yield maximum likelihood estimates with large bias and has a biased score function, the residual likelihood yields variance components estimates with reduced bias and has a score function which is unbiased. Hence the residual likelihood behaves more like a "real" likelihood than the unadjusted profile likelihood. The residual likelihood, however, is defined only when ρ is the parameter of interest and β is the nuisance parameter vector. It can not be applied if, as is often the case, interest lies in the fixed effects. If only some of the components of ρ are of interest, the residual likelihood fails to account for the remaining variance component nuisance parameters. The approximate modified profile likelihood, defined below, is applicable for an arbitrary interest parameter and reduces to the residual likelihood if the entire vector ρ is of interest. Hence, Barndorff-Nielsen's approximate modified profile likelihood generalizes the residual likelihood to other sets of interest parameters.

We adopt notation similar to that used by Barndorff-Nielsen and Cox (1994) and partition the parameter vector θ into a parameter of interest, ψ, and a nuisance parameter, χ. The likelihood of θ will be written as $L(\theta)$ and its log as $l(\theta)$. Differentiation with respect to θ will be denoted by a / followed by a sub-scripted parameter, e.g., the score $U(\theta)$ can be written as $l_{/\theta}$ and the observed information $j(\theta)$ as $-l(\theta)_{/\theta\theta}$. The expected

information is denoted by $i(\theta) = E(j(\theta))$. Evaluation of a quantity dependent on θ at the maximum likelihood estimates $(\widehat{\psi}, \widehat{\chi})$ or at the restricted maximum likelihood estimates $(\psi, \widetilde{\chi}_\psi)$ is denoted by a superior $\frown$ and $\sim$, respectively; e.g., the profile log likelihood for ψ, is denoted by $\widetilde{l}$, i.e., $\widetilde{l} = l_p(\psi) = l(\psi, \widetilde{\chi})$. For a matrix X, $R(X)$ will denote the range space, $r(X)$ its rank, $tr(X)$ its trace, and $|X|$ its determinant.

In section 2 both Barndorff-Nielsen's modified profile likelihood, L_{mp}, and his second order approximation to it are presented. In mixed models this approximate modified profile likelihood has a compact form requiring only calculation of the restricted MLEs and derivatives of the covariance matrix with respect to ρ. The specific adjusted likelihoods for covariance and fixed parameters are discussed in sections 3 and 4, respectively. In section 5 we present some numerical examples and the appendix contains some technical results.

3.3 The Modified Profile Likelihood

The modified profile likelihood for ψ (Barndorff-Nielsen 1983) is

$$L_{mp}(\psi) = D(\psi)|\widetilde{j}_{\chi\chi}|^{-\frac{1}{2}}L_p(\psi) \tag{2}$$

where $\widetilde{j}_{\chi\chi}$ is the nuisance parameter block of the observed information and

$$D(\psi) = |\tfrac{\partial \widetilde{\chi}_\psi}{\partial \widehat{\chi}}|^{-1},$$

where $\widetilde{\chi}_\psi$ is the MLE of χ when ψ is held fixed. $D(\psi)$ is equal (Barndorff-Nielsen and Cox 1994) to the ratio of the determinants of the observed information and the mixed sample space derivative $\widetilde{l}_{/\chi;\widehat{\chi}}$:

$$D(\psi) = \frac{|\widetilde{j}_{\chi\chi}|}{|\widetilde{l}_{/\chi;\widehat{\chi}}|}. \tag{3}$$

The sample space derivative, $l_{/\chi;\widehat{\chi}} = \frac{\partial}{\partial \chi}\frac{\partial}{\partial \widehat{\chi}}l$, is the derivative of the log likelihood taken with respect to the parameters χ and with respect to the data through $\widehat{\chi}$. In curved exponential families it is possible to choose an approximately ancillary statistic A such that $(\widehat{\psi}, \widehat{\chi}, A)$ is sufficient. Hence, the mixed sample space derivative is well defined for

a particular A. In practice, however, these derivatives can be difficult to calculate outside of regular exponential families where the MLEs alone are sufficient.

Barndorff-Nielsen's modified profile likelihood has several favorable properties. When the joint likelihood can be factored into either a conditional or a marginal likelihood that isolates ψ in a reasonable way then the $L_{mp}(\psi)$ approximates that likelihood to third order. Also, unlike the unadjusted profile likelihood, the score of $L_{mp}(\psi)$ is asymptotically unbiased (Barndorff-Nielsen 1994a). In contrast to the Cox-Reid approximate conditional likelihood (Cox and Reid 1987), $L_{mp}(\psi)$ is invariant to interest respecting reparameterizations (Barndorff-Nielsen and Cox 1994). That is, reparameterizations of the form $(\psi,\chi) \leftrightarrow (\phi,\lambda)$ where $\phi = \phi(\psi)$ is a monotone function of ψ, and $\lambda = \lambda(\psi,\chi)$. Generally, however, outside of regular exponential families, calculation of $L_{mp}(\psi)$ requires the specification of an approximate ancillary statistic and in curved families there may be several reasonable ancillary statistic. This complicates the calculation of the sample space derivative in (2) for mixed models since outside of a few important cases, e.g., balanced variance component models, the densities in mixed linear models belong to curved exponential families rather than regular exponential families.

Barndorff-Nielsen (1994b) gave a second order approximation to the sample space derivative in (3) for curved exponential families. The approximation depends on writing the density of Y in canonical form with canonical parameter $\eta(\theta)$, canonical statistic T and cumulant generating function $K(\theta)$:

$$l(\theta) = \sum_{i=1}^{m} \eta_i(\theta)T_i(Y) - K(\theta).$$

Ignoring factors that do not depend on ψ, Barndorff-Nielsen showed that to second order

$$|\tilde{l}_{/\chi;\hat{\chi}}| = |\widetilde{\eta}'_{/\chi}\widehat{K}_{/\eta\eta}\,\widetilde{\eta}_{/\chi}|. \tag{4}$$

Approximating $|\tilde{l}_{/\chi;\hat{\chi}}|$ with $|\widetilde{\eta}'_{/\chi}\widehat{K}_{/\eta\eta}\,\widetilde{\eta}_{/\chi}|$ in (4) gives Barndorff-Nielsen's approximate modified profile likelihood:

Page 35

$$L_{mp}^\dagger(\psi) = D^\dagger(\psi)\,|\tilde{j}_{\chi\chi}|^{-\frac{1}{2}}L_p(\psi), \tag{5}$$

where $D^\dagger(\psi) = |\widehat{\eta}'_{/\chi}\widehat{K}_{/\eta\eta}\,\widetilde{\eta}_{/\chi}|^{-1}|\tilde{j}_{\chi\chi}|$. In regular exponential families the relation in (4) is exact so that $L_{mp}^\dagger(\psi)$ equals $L_{mp}(\psi)$. This is convenient in mixed linear models since it is often a nuisance to determine if the density of Y in a given model is a member of a regular or curved exponential family. Like L_{mp}, $L_{mp}^\dagger(\psi)$ is invariant to interest respecting parameterizations and does not require parameter orthogonality. Furthermore, as shown in the appendix, it behaves more like a genuine likelihood since the score is asymptotically unbiased.

Since writing mixed model densities in canonical form is sometimes difficult, it is convenient to note that $\widehat{\eta}'_{/\chi}\widehat{K}_{/\eta\eta}\,\widetilde{\eta}_{/\chi}$ is equal to the covariance of the score evaluated at the full and restricted MLEs:

$$S_{\chi\chi}(\psi) = Cov_{\theta_1}\left(U_\chi(\theta_1),\,U_\chi(\theta_2)\right)\Big|_{\theta_1=\widehat{\theta},\theta_2=\widetilde{\theta}}. \tag{6}$$

In a mixed model setting where interest lies with a component of β or of ρ, $S_{\chi\chi}$ has a compact representation depending only on X, V, and derivatives of V with respect to each component of ρ. Skovgaard (1996) proposed a similar approximation to the sample space derivative: the nuisance parameter block of $S\,\widehat{i}^{-1}\widehat{j}$ where S is

$$S(\psi) = Cov_{\theta_1}\left(U(\theta_1),\,U(\theta_2)\right)\Big|_{\theta_1=\widehat{\theta},\theta_2=\widetilde{\theta}}.$$

When the interest parameter ψ is either a component of β or ρ we show in the appendix that S has the following compact form:

$$S(\psi) = \begin{pmatrix} S_{\beta\beta} & S_{\beta\rho} \\ S_{\beta\rho} & S_{\rho\rho} \end{pmatrix} \text{ where}$$

$$S_{\beta\beta} = X'\widetilde{V}^{-1}X,$$

$$S_{\rho\beta} = 0,$$

$$S_{\rho_r\rho_l} = \tfrac{1}{2}tr(\widehat{V}_{/\rho_r}\widetilde{V}^{-1}\widetilde{V}_{/\rho_l}\widetilde{V}^{-1}),\ \text{for } r,l = 1,\ldots,k,$$

and the rth column of $S_{\beta\rho}$ given by

$$S_{\beta\rho_r} = X'\widetilde{V}_{/\rho_r}^{-1}X(\widetilde{\beta} - \widehat{\beta}) \quad \text{for } r = 1,\ldots,k.$$

The matrix $S_{\chi\chi}(\psi)$ is then given by the nuisance parameter block of S. If a component of either ρ or β is of interest then neither $S_{\beta\rho}$ nor $S_{\beta\rho}$ appear in the determinant of $S_{\chi\chi}$ and can be ignored.

The fixed effect portion of S, $S_{\beta\beta}$, is equal to the fixed effect block of the observed and expected information matrices:

$$S_{\beta\beta} = \widetilde{i}_{\beta\beta} = \widetilde{j}_{\beta\beta}.$$

The covariance parameter block changes substantially with the form of V. When V is linear in ρ, as in variance component problems, this block is equal to the expected information evaluated at the restricted MLEs. To see this note that

$$V(\rho) = \sum_{i=1}^{k}\rho_i Q_i$$

so that $V_{/\rho_i} = Q_i$ and

$$S_{\rho_r\rho_l} = \tfrac{1}{2}tr(\widetilde{V}^{-1}Q_i\widetilde{V}^{-1}Q_l).$$

Elementary methods show that this is equal to $\widetilde{i}_{\rho\rho}$ (see Searle, Casella, and McCulloch 1992).

$L_{mp}^{\dagger}$ differs from L_{mp} when evaluated at the full maximum likelihood estimate losing a compelling property. When evaluated at the full MLE the sample space derivative is equal to the observed information (Barndorff-Nielsen and Cox 1994), $\widehat{l}_{/\theta;\widehat{\theta}} = \widehat{j}$, hence $D(\widehat{\psi}) = 1$. It is clear from (6), however, that

$$S_{\chi\chi}(\widehat{\psi}) = \widehat{\eta}'_{/\chi}\widehat{K}_{/\eta}\widehat{\eta}_{/\chi} = \widehat{i}_{\chi\chi}$$

so that $D^{\dagger}(\widehat{\psi}) = 1$ only when the nuisance parameter blocks of the observed and expected information are equal. Skovgaard's approximation remedies this by multiplying S by $\widehat{i}^{-1}\widehat{j}$. However, since these two approximations are equal if the density follows a regular exponential family and are equal to second order generally, we will employ Barndorff-Nielsen's simpler approximation.

3.4 Inference About Covariance Parameters

If the entire vector of covariance parameters ρ is of interest in the model defined by (1) then because of parameter orthogonality, $|\frac{\partial \tilde{\beta}_\rho}{\partial \beta}| = 1$ to second order (Barndorff-Nielsen and McCullagh 1993), so that the residual likelihood is a second order approximation to $L_{mp}(\rho)$. For this case it is clear that the approximation $D^\dagger(\rho) = 1$ so that the approximate modified profile likelihood is exactly equal to the residual likelihood. Furthermore, the residual, modified profile, and approximate modified likelihoods are exactly equal if $\tilde{\beta}_\rho = \hat{\beta}$ which implies that $|\frac{\partial \hat{\beta}_\rho}{\partial \beta}| = 1$ and hence $D(\rho) = D^\dagger(\rho) = 1$. In mixed models, $\tilde{\beta}_\rho = \hat{\beta}$ is equivalent to Zyskind's condition (Zyskind 1967): $R(V^{-1}(\rho)X) \subset R(X) \; \forall \rho$.

By itself Zyskind's condition, which holds in most balanced variance component models as well as a variety of repeated measures settings, does not imply that the density of Y belongs to a member of a regular exponential family. In this special case, the MLEs are not sufficient but the ancillary statistic does not enter into the calculation of $L_{mp}(\rho)$. Generally, however, $L_{mp}(\rho)$ depends on the specification of an ancillary and so may differ from the residual likelihood.

If only a subset of the covariance parameters are of interest the residual likelihood fails to account for the covariance nuisance parameters. Suppose that $\rho = (\rho_1, \rho_2)$ and interest lies with ρ_1 so that the nuisance parameter $\chi = (\beta, \rho_2)$. The adjustment is then made up of the familiar determinant of the fixed effect block of the observed information and an adjustment for the covariance nuisance parameters:

$$lnL_{mp}^\dagger(\rho_1) = lnL(\rho_1) - \tfrac{1}{2}ln|X'V^{-1}(\tilde{\rho})X|$$
$$+ \tfrac{1}{2}ln|\tilde{j}_{\rho_2\rho_2|\beta}| - ln|S_{\rho_2\rho_2}|$$

where $|\tilde{j}_{\rho_2\rho_2|\beta}|$ is the observed information for ρ_2 adjusted for β, i.e.,

$$\lfloor \tilde{j}_{\rho_2\rho_2|\beta} \rfloor \;=\; \lfloor \tilde{j}_{\,\rho_2\rho_2} - \tilde{j}_{\,\rho_2\beta}(\tilde{j}_{\beta\beta})^{-1}\tilde{j}_{\,\beta\rho_2} \rfloor.$$

Again, if Zyskind's condition holds this term simplifies since it is easy to show that

$$\tilde{j}_{\rho_2\beta} \;=\; 0.$$

The extra adjustment for the nuisance covariance parameters insures that the score of $L^{\dagger}_{mp}(\rho_1)$ is asymptotically unbiased. In cases where the covariance parameter of interest is a canonical parameter in a regular exponential family the modified profile likelihood is a third order approximation to the conditional likelihood. For inference on canonical parameters in a balanced variance component model, the extra adjustment turns out to be a function of the data only, so that both the profile residual likelihood and the modified profile likelihood are proportional to the conditional likelihood. In some cases though, the extra adjustment may be appreciable.

3.4.1 *Balanced Nested Variance Component Model*

Here we consider models of the form

$$Y_n \;=\; \mu 1_n \;+\; \varepsilon, \quad \varepsilon \;\sim\; N_n(0, V(\tau,\chi)),$$

where the covariance matrix $V(\tau,\chi) = \tau I_n + \chi M$ with $M = \bigoplus_{i=1}^{a} J_k$, J_k is a $k \times k$ matrix of 1s, and the variance of $\hat{\mu}$ is given by $(\tau + k\chi)$. The variable a represents the number of clusters and k the number of elements in each cluster. It is easy to see that

$$V^{-1}(\tau,\chi) = \tfrac{1}{\tau}(I_n - Q) + \tfrac{1}{(\tau+k\chi)}Q,$$

where $Q = \tfrac{1}{ak}M$ is the orthogonal projection operator on $R(M)$. Suppose that interest lies with $\tau + k\chi$. A convenient parameterization employs the canonical interest parameter $\psi = (\tau + k\chi)^{-1}$, and nuisance parameter $\chi = (\chi_1, \chi_2) = (\tau^{-1}, \mu)$, where τ and μ are canonical and mean parameters respectively. In this case ψ and (τ^{-1}, μ) are estimation orthogonal since for fixed ψ

$$\widetilde{\mu} \ = \widehat{\mu} \ = \ \tfrac{1}{ak}1_n'Y \ \text{and}$$

$$\widetilde{\tau}^{-1} \ = \widehat{\tau}^{-1} \ = \ \frac{n-a}{(Y-\widehat{\mu}1_n)'(I_n-Q)(Y-\widehat{\mu}1_n)} \ \cdot$$

The profile log likelihood is given by

$$l_p(\psi) \ \propto \ -\tfrac{\psi}{2}Y'N_xQN_xY \ + \ \tfrac{a}{2}ln\psi$$

where $N_x \ = (I \ - \ \tfrac{1}{ak}1_n1_n')$ is the orthogonal projection operator on $R(1_n)^{\perp}$.

The statistic $(\widehat{\mu}, Y'N_xQN_xY, Y'N_x(I \ - \ Q)N_xY)$ is a complete and sufficient

statistics in the full model so the conditional likelihood given by the density of

$$Y'N_xQN_xY \,|\, \widehat{\mu}, \ Y'N_x(I \ - \ Q)N_xY$$

can be used for inference. The term $\psi Y'N_xQN_xY$ is distributed as a chi-square random

variable with $a-1$ degrees of freedom so that $\psi Y'N_xQN_xY$ is ancillary when ψ is

held fixed. Noting that $(\widehat{\mu}, Y'N_x(I \ - \ Q)N_xY)$ form a complete sufficient statistic in

the model with ψ held fixed we see by Basu's theorem (Lehmann 1986) that the

conditional likelihood is given by

$$l_c(\psi) \ \propto \ -\tfrac{\psi}{2}(Y'N_xQN_xY) \ + \ \tfrac{(a-1)}{2}ln(\psi) \ = l_p(\psi) - \tfrac{1}{2}ln(\psi)$$

which differs from the profile likelihood by a factor $-\tfrac{1}{2}ln\psi$.

Noting that $R(1_n) \ \subset \ R(M)$ and that the restricted REML estimate of χ_1

coincides with the restricted maximum likelihood estimate, one can show that the residual

log likelihood is given by

$$l_{PREML}(\psi) \ = l_p(\psi) \ - \ \tfrac{1}{2}ln|1_n'\widetilde{V}^{-1}1_n|$$

$$\propto \ l_p(\psi) \ - \ \tfrac{1}{2}ln(\psi)$$

since $r(1_n'Q1_n) \ = \ r(1_n'1_n)$. Since the restricted MLEs of the nuisance parameters do

not depend on ψ, the modified profile likelihood is equal to the Cox Reid profile

likelihood, i.e., $|\frac{\partial \tilde{\chi}_\psi}{\partial \chi}| = 1$. Noting further that $j_{\chi_1\mu} = 0$, $\tilde{j}_{\chi_1\chi_1}$ is a function of χ_1 alone, and that $\tilde{\chi}_1 = \hat{\chi}_1$ we see that

$$l_{mp}(\psi) = l_p(\psi) - \tfrac{1}{2}ln|1_n'\tilde{V}^{-1}1_n| - \tfrac{1}{2}ln|\tilde{j}_{\chi_1\chi_1}|$$

$$\propto l_p(\psi) - \tfrac{1}{2}ln\,\psi.$$

Hence both the residual likelihood and the modified profile likelihood correct the profile so that it is proportional to the conditional likelihood. This result generalizes readily to other variance component models where Zyskind's condition holds.

3.4.2 Bivariate Normal

Suppose that m pairs of data follow a bivariate normal distribution and hence have the following density:

$$\begin{pmatrix} Y_{1i} \\ Y_{2i} \end{pmatrix} \sim N_2\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \Sigma_i\right) \text{ where } \Sigma_i = \begin{pmatrix} \sigma^2 & \chi \\ \chi & \tau^2 \end{pmatrix}, i = 1 \text{ to } m.$$

For this case the density of Y follows a regular exponential family. Suppose that the parameter of interest is

$$\psi = \frac{\chi}{\sigma\tau - \chi^2}$$

which is a canonical parameter, and that the nuisance parameters are τ^2, σ^2, μ_1 and μ_2, which are mean parameters in this case. Since $\chi \geq 0$ is equivalent to $\psi \geq 0$ the former can be tested through a canonical parameter. The restricted MLEs for the nuisance parameters are

$$\tilde{\sigma}^2 = \frac{\sum_{i=1}^{m}(y_{1i} - \bar{y})^2}{m}, \quad \tilde{\tau}^2 = \frac{\sum_{i=1}^{m}(y_{2i} - \bar{y})^2}{m}, \quad \hat{\mu}_1 = \frac{\sum_{i=1}^{m}y_{1i}}{m}, \text{ and } \hat{\mu}_2 = \frac{\sum_{i=1}^{m}y_{2i}}{m}.$$

Given the estimates of τ^2 and σ^2, the parameter χ is a nonlinear function of ψ and the adjustments are:

$$|S_{\rho_r\rho_r}| = -\tfrac{1}{16}\frac{n^2(\chi^2(\psi)+\widetilde{\sigma}\widetilde{\tau})}{(\chi^2(\psi)-\widetilde{\sigma}\widetilde{\tau})^3}, \quad |X'V^{-1}X| = (\tfrac{n}{2})^2\frac{1}{(\chi^2(\psi)-\widetilde{\sigma}\widetilde{\tau})}, \quad \text{and}$$

$$|j_{\rho_r\rho_r}| = -(n\widetilde{\sigma}^4 + n\chi^4(\psi) - 4\widetilde{\sigma}^2\chi^2(\psi) + 4\widetilde{\sigma}^3\widetilde{\tau}) + (-n\widetilde{\tau}^4 +$$
$$4\widetilde{\tau}^2\chi^2(\psi) - 4\widetilde{\tau}^3\widetilde{\sigma} + n\widetilde{\sigma}^4 - 4\widetilde{\sigma}^2\chi^2(\psi)$$
$$+ 4\widetilde{\sigma}^3\widetilde{\tau})\,(\chi^2(\psi) - \widetilde{\sigma}\widetilde{\tau})^{-2}.$$

The log of each of these terms depends on ψ so, unlike the previous example, the covariance nuisance parameter adjustment is a function of ψ.

The density formed from the residual likelihood also is a member of a regular exponential family with ψ a canonical parameter. The restricted residual maximum likelihood estimates for the nuisance parameters τ and σ are given by

$$\widetilde{\sigma}^2_{REML} = \frac{\sum_{i=1}^{m}(y_{1i}-\overline{y})^2}{m-1} \quad \text{and} \quad \widetilde{\tau}^2_{REML} = \frac{\sum_{i=1}^{m}(y_{2i}-\overline{y})^2}{m-1},$$

so that for moderately large m the fixed effect adjustment, $|X'\widetilde{V}^{-1}X|$, of the modified profile likelihood will be very close to the residual likelihood's adjustment. Unlike the previous example, however, the covariance parameter adjustment term $\tfrac{1}{2}ln|\widetilde{j}_{\rho_2\rho_2|\beta}| - ln|S_{\rho_2\rho_2}|$ is a function of ψ so there may be an appreciable difference between the residual and modified profile likelihoods. Since the modified profile likelihood approximates the conditional likelihood for ψ to third order any difference between the modified profile and the residual profile likelihood may be important.

3.5 Inference About Fixed Effects

For fixed effects, Barndorff-Nielsen's approximate modified profile likelihood compliments the residual likelihood by adjusting for the distorting effects that the covariance and other fixed effects nuisance parameters have on the profile likelihood. Suppose interest lies in a parameter μ defined as

$$E(Y) = A\mu + X_r\beta_r.$$

Hence the nuisance parameter vector $\chi = (\beta_r, \rho)$. The approximate modified profile likelihood is then given by

$$L^{\dagger}_{mp}(\beta_1) = |S_{xx}| \, |\tilde{j}_{xx}|^{\frac{1}{2}} L_p(\beta_1).$$

$$= |X_r'V^{-1}X_r|^{-\frac{1}{2}} |S_{\rho\rho}|^{-1} |\tilde{j}_{\rho\rho|\beta_r}|^{\frac{1}{2}} L_p(\beta_1),$$

where $|\tilde{j}_{\rho\rho|\beta_r}| = |\tilde{j}_{\rho\rho} - \tilde{j}_{\rho\beta_r}(X_r'V^{-1}(\tilde{\rho})X_r)^{-1}\tilde{j}_{\beta\rho_2}|.$

Unlike the covariance parameters, the profile score of a single fixed effect is always asymptotically unbiased. Cox and Reid (1992) noted this fact for mean parameters in regular exponential families and a short proof in the appendix shows that this holds for all mixed models. The unbiasedness of the profile score also implies that the adjustment is $o_p(1)$ rather than $O_p(1)$ and so may be less important. Despite this, for small samples the profile likelihood may still be misleading especially if the covariance matrix is complex. Lyons and Peters (1996) present simulation studies that indicate that the unadjusted likelihood ratio test may be very misleading, rejecting too often. Another numerical study of tests based on the Wald statistic presented by Wright (1994) indicates that the Wald test is also too liberal.

The problem with the profile likelihood is analogous to the problem of inference about mean parameters outlined in Example 8.2 of Barndorff-Nielsen and Cox (1994). They show that the modified profile likelihood for an orthogonal contrast in a standard linear model is a much closer approximation to the likelihood based on the exact t-statistic than the unadjusted profile likelihood. Below we show that this result also holds for orthogonal fixed effect contrasts in some mixed linear models, such as, balanced variance component models. A sketch of the proof is given in the appendix.

***Theorem* 3.5.1** : Suppose that the mixed model satisfies the following two conditions insuring that the density of Y follows a regular exponential family:

(i) The covariance matrix can be rewritten as $V(\rho) = \sum_{i=1}^{k} \lambda_i(\rho) Q_i$

where the matrices Q_i are known symmetric, idempotent, non-negative definite, pair-wise orthogonal and $\sum_{i=1}^{k} Q_i = I$. Also, $\lambda(\rho)$ is a one to one continuous and differentiable function of ρ.

(ii) Zyskind's condition: $R(Q_i X) \subset R(X) \, \forall \, i$.

Suppose also that the scalar fixed effect parameter of interest is μ where

$$X\beta = A\mu + X_r \beta_r \, .$$

Together with conditions (i) and (ii) the next two conditions insure that there is an exact t-test for testing $\mu = \mu_o$:

(iii) $A'X_r = 0$.

(iv) $R(A) \subset R(Q_i)$ for some i.

Then for testing $\mu = \mu_o$ the statistic

$$t = \frac{\hat{\mu} - \mu_o}{SE(\hat{\mu})},$$

where $SE(\hat{\mu})$ is, in variance component models, the usual ANOVA standard error for $\hat{\mu}$, follows a t distribution exactly with $r(Q_i) - r(Q_i X_r) - 1$ degrees of freedom. Also, the profile likelihood, modified profile likelihood, normed so that they have maximum 1, and the likelihood based on the t statistic are given by

$$L_p(\mu) = \left(1 + \frac{t^2}{r(Q_i) - r(Q_i Px) - 1}\right)^{-\frac{r(Q_i)}{2}},$$

$$L_{mp}(\mu) = \left(1 + \frac{t^2}{r(Q_i) - r(Q_i P_x) - 1}\right)^{\frac{-r(Q_i) + r(Q_i X_r) + 1}{2}},$$

and $L_t(\mu) = \left(1 + \frac{t^2}{r(Q_i) - r(Q_i P_x) - 1}\right)^{\frac{-r(Q_i) + r(Q_i X_r)}{2}}$. $\square$

Since in this case the density of Y follows a regular exponential family the approximate modified profile likelihood equals the modified profile likelihood. In hierarchical mixed linear models, such as split plot models, the rank of $Q_i X_r$ is the

number of fixed effect nuisance parameters on the same "level" as μ. The profile likelihood errs by not accounting for these variables while the modified profile likelihood is nearly identical to the likelihood induced by the t statistic. This result is illustrated by the following two split plot examples.

3.5.1 *Fixed Whole Plot Effect in a Balanced Split-Plot*

Suppose we wish to make inference concerning an orthogonal whole plot contrast μ in a balanced split plot model. To apply the above theorem this model can be written as

$$Y_n = A\mu + X_1\beta_1 + X_2\beta_2 + \varepsilon$$

$$\varepsilon \sim N_n(0, V(\tau, \chi))$$

where $A'[X_1 X_2] = 0$, β_1 has dimension p_1, and β_2 has dimension p_2. The covariance matrix $V(\tau, \chi) = \tau I_n + \chi M$, where $M = \bigoplus_{i=1}^{a} J_k$ and J_k is a $k \times k$ matrix of 1s. Hence, V can be written as,

$$V(\tau, \chi) = \tau(I_n - Q) + (k\chi + \tau)Q$$

where $Q = \frac{1}{k}M$ is the orthogonal projection operator on $R(M)$ and $(I_n - Q)$ is the orthogonal projection operator on $R(M)^\perp$. It is easy to see that

$$V^{-1}(\tau, \chi) = \frac{1}{\tau}(I_n - Q) + \frac{1}{(k\chi + \tau)}Q.$$

Since A is a whole plot effect it is easy to show that $R(A) \subset R(Q)$ which gives $QA = A$. Also, without loss of generality, we can assume that $R(X_1) \subset R(Q)$ so that $r(QX_1) = p_1$ and that $R(X_2) \subset R(Q)^\perp$ so that $r(QX_2) = 0$. Hence, the profile likelihood, modified profile likelihood and the likelihood induced by the t statistic, normalized so that the maximum of each is 1, are

$$L_p(\mu) = \left(1 + \frac{t^2}{a - p_1 - 1}\right)^{-\frac{a}{2}},$$

$$L_{mp}(\mu) = \left(1 + \frac{t^2}{a - p_1 - 1}\right)^{-\frac{a}{2} + \frac{p_1}{2} + \frac{1}{2}},$$

$$\text{and } L_t(\mu) = \left(1 + \frac{t^2}{a-p_1-1}\right)^{-\frac{a}{2}+\frac{p_1}{2}}.$$

If the number of fixed effects at the whole plot level, p_1, is large compared to the number of whole plot units, $a = r(Q)$, then the unadjusted profile likelihood will differ substantially from the likelihood given by the t statistic since it in effect ignores the nuisance fixed effects. On the other hand the modified profile likelihood accounts for these effects and is nearly identical to $L_t(\mu)$.

3.5.2 *Fixed Sub Plot Effect in a Balanced Split-Plot*

If instead μ is an orthogonal subplot contrast where $(I - Q)A = A$ then the normed profile likelihood, modified profile likelihood and the likelihood induced by the t statistic are given by

$$L_p(\mu) = \left(1 + \frac{t^2}{a(k-1)-p_2-1}\right)^{-\frac{a(k-1)}{2}},$$

$$L_{mp}(\mu) = \left(1 + \frac{t^2}{a(k-1)-p_2-1}\right)^{-\frac{a(k-1)}{2}+\frac{p_2}{2}+\frac{1}{2}},$$

$$\text{and } L_t(\mu) = \left(1 + \frac{t^2}{a(k-1)-p_2-1}\right)^{-\frac{a(k-1)}{2}+\frac{p_2}{2}},$$

where $r(X_2) = p_2$. The difference between the modified and unadjusted profile likelihood will usually be less marked here then in the whole plot case since $\frac{a(k-1)}{2}$ is usually large relative to p_2.

3.6 Numerical Example

In a one parameter model, a plot of its profile likelihood neatly summarizes the information available for that single parameter. As we have seen in the example above, in more complex models with many nuisance parameters the profile may be misleading because it assumes that the nuisance parameters are known; that is, it fails to account for the uncertainty of the nuisance parameter estimates. The following repeated measures

example illustrates how the adjustment modifies the profile likelihood to account for these nuisance parameters.

Pothoff and Roy (1964) presented a repeated measures example consisting of the growth measurements of 16 boys and 11 girls taken at ages 8,10,12, and 14. For this example the mean is modeled as linear in age with a gender by age interaction:

$$E(Y_{ij}) = \alpha_o + \alpha_1 gen_i + \alpha_2 age_j + \beta(age \times gen)_{ij}$$

$$i = 1,2 \text{ and } j = 1,2,3,4.$$

For inference about the interaction coefficient β, various variance structures can be fit to account for the within subject correlation. A saturated model is given by the unstructured covariance matrix:

$$V_i^{un} = \begin{pmatrix} \sigma_{11} & \sigma_{21} & \sigma_{31} & \sigma_{41} \\ \sigma_{21} & \sigma_{22} & \sigma_{32} & \sigma_{42} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \sigma_{43} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44} \end{pmatrix}.$$

A variety of other more parsimonious covariance matrices have been proposed and can be fit with commercial software. For any covariance matrix, the nuisance parameters may distort the profile likelihood of β giving a profile that is too narrow.

Figures 3.1 and 3.2 show the profile and approximate modified profile likelihoods, normed so that they have maximum one, for the unstructured covariance structured above, as well as the more parsimonious heterogenous toeplitz structure (see Wolfinger 1996):

$$V_i = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_1 & \sigma_1\sigma_3\rho_2 & \sigma_1\sigma_4\rho_3 \\ \sigma_1\sigma_2\rho_1 & \sigma_2^2 & \sigma_2\sigma_3\rho_1 & \sigma_2\sigma_4\rho_1 \\ \sigma_1\sigma_3\rho_2 & \sigma_2\sigma_3\rho_1 & \sigma_3^2 & \sigma_3\sigma_4\rho_1 \\ \sigma_1\sigma_4\rho_3 & \sigma_3\sigma_4\rho_1 & \sigma_3\sigma_4\rho_1 & \sigma_4^2 \end{pmatrix}.$$

Figure 3.1
Profile and Approximate Modified Profile Likelihood for β:
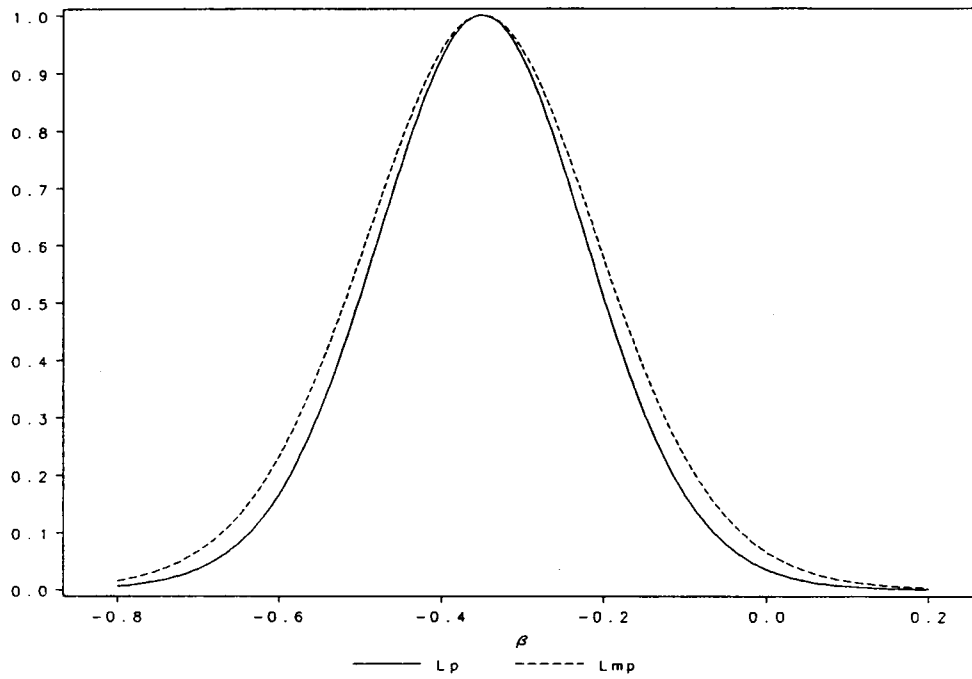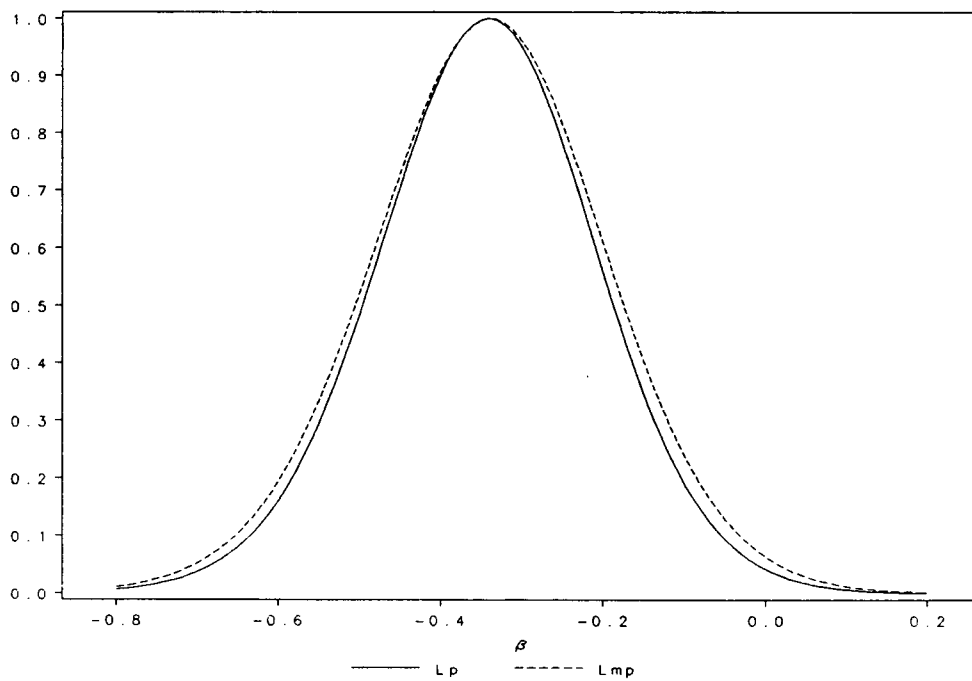### Unstructrued Covariance Matrix

Figure 3.2
Profile and Approximate Modified Profile Likelihood for β:
Heterogenous Toeplitz Covariance Matrix

In both cases the profile and approximate modified likelihoods give nearly identical point estimates of β. However, the approximate modified profile likelihood is wider than the profile likelihood especially in the unstructured case and when β is far from the its MLE, i.e., the larger deviation regions. The profile likelihood is indeed too narrow. Simulations presented in Lyons and Peters (1996) showed that the unmodified profile likelihood gives confidence intervals that are too narrow in both models. It fails egregiously in the more complex unstructured model and at the higher nominal levels of coverage, that is, where the profile and approximate modified profile likelihood differ the most. The appreciable difference between the profile and approximate modified profile likelihoods in this case then indicates potential problems in using the LRT and other first order methods, such as, the Wald test, to construct confidence intervals and p-values.

3.7 Conclusion

The approximate modified profile likelihood, $L_{mp}^{\dagger}(\psi)$, constructed for an arbitrary interest parameter in a complex linear model sensibly corrects for the nuisance parameters giving an approximately unbiased score. Furthermore, it is often a good approximation to what is generally regarded as the correct likelihood. The adjustment improves profile likelihood based inference since it takes into account the uncertainty associated with the nuisance parameter.

For inference about a single parameter, plotting the profile likelihood and the approximate modified profile likelihood may indicate whether the nuisance parameters are having a substantially adverse effect on the likelihood ratio test. For more formal inference about an individual parameter, such as, the construction of confidence intervals and p-values, the modified directed deviance statistic (see Lyons and Peters 1996 and Skovgaard 1996) may be more appropriate. This statistic is known to be standard normal to second order in moderate deviation regions and first order in large deviation regions.

To obtain p-values and confidence regions for several parameters at a time further

research is required to determine the efficacy of a deviance drop based on the

approximate modified profile likelihood.

3.8 References

Barndorff-Nielsen O.E. (1983). On a formula for the distribution of the maximum
likelihood estimator. *Biometrika,* 70, 343-365.

Barndorff-Nielsen, O.E. (1994a). Adjusted version of profile likelihood and directed
likelihood, and extended likelihood. *Journal of the Royal Statistical Society Series
B,* 56, 125-140.

Barndorff-Nielsen, O.E. (1994b). Stable and invariant adjusted profile likelihood and
directed likelihood for curved exponential models. Technical Report, Department of
Statistics. University of Aarhus, Aarhus, Denmark.

Barndorff-Nielsen, O.E. and Cox, D.R. (1994). *Inference and Asymptotics.* Chapman and
Hall, New York.

Barndorff-Nielsen, O.E. and McCullagh, P. (1991). A note on the relation between
modified profile likelihood and the Cox-Reid adjusted profile likelihood.
Biometrika, 80, 321-328.

Cox, D. R. and Reid, N. (1987). Parameter orthoganality and approximate conditional
inference. *Journal of the Royal Statistical Society Series B,* 49, 1-18.

Cox, D. R. and Reid, N. (1992). A note on the difference between profile and modified
profile likelihood. *Biometrika,* 79, 408-411.

Lehmann, E.L. (1986). *Testing Statistical Hypotheses.* Wiley, New York.

Lyons, B. and Peters, D. (1996). A modified directed deviance statistic for mixed linear
models. Technical Report 164, Department of Statistics, Oregon State University.

Mathai, A.M. and Provost, S.B. (1992). *Quadratic Forms in Random Variables: Theory
and Application.* Marcel Dekker, New York.

McCullagh, P. and Tibshirani, R. (1990). A simple method for the adjustment of profile
likelihoods. *Journal of the Royal Statistical Society Series B,* 52, 325-344.

Patterson, H.R. and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, 58, 545-554.

Pothoff, R.F. and Roy, S.N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika*, 51, 313-326.

Searle, S.R., Casella, G., and McCulloch, C.E. (1992). *Variance Components*. John-Wiley and Sons, New York.

Seely, J. (1977). Minimal sufficient statistics and completeness for multivariate normal families. *Sankhya A*, 39, 170-185.

Skovgaard, I.M. (1996). An explicit large-deviation approximation to one-parameter tests. *Journal of the Bernoulli Society*, 2, 145 -165.

Swallow, W.H., and Monahan, J.F (1984). Monte Carlo comparisons of ANOVA, MIVQUE, REML and ML estimators of variance components. *Technometrics*, 26, 47-57.

Wolfinger, R.D. (1996). Heterogenous Variance-Covariance Structures for Repeated Measures. *Journal of Agricultural, Biological, and Environmental Statistics*, 1, 205 - 230.

Wright, S.P. (1994). Adjusted F tests for repeated measures with the MIXED procedure. 328 SMC-Statistics Department, University of Tennessee.

Zyskind, G. (1967). On canonical forms, non-negative covariance matrices and best and simple least square linear estimators in linear models. *Annals of Mathematical Statistics*, 38, 1092-1109.

3.9 Appendix

First we develop the form of $S(\psi)$.

Claim 1: $S(\psi) = \begin{pmatrix} S_{\beta\beta} & S_{\beta\rho} \\ S_{\rho\beta} & S_{\rho\rho} \end{pmatrix}$ where

$$S_{\beta\beta} = X'\widetilde{V}^{-1}X,$$
$$S_{\rho\beta} = 0,$$
$$S_{\rho_r\rho_l} = \tfrac{1}{2}tr(\widehat{V}_{/\rho_r}\widetilde{V}^{-1}\widetilde{V}_{/\rho_l}\widetilde{V}^{-1}), \text{ for } r, l = 1, ..., k$$

and the rth column of $S_{\beta\rho}$ is given by

$$S_{\beta\rho_r} = X'\widetilde{V}^{-1}_{/\rho_r}X(\tilde{\beta} - \hat{\beta}) \text{ for } r = 1, ..., k.$$

Proof: The score equations are given by

$$[U_\beta(\beta, \rho)]_i = Y'V^{-1}X_i - \sum_{j=1}^{p} X_i'V^{-1}X_j\beta_j \text{ for } i = 1, ..., p \text{ and}$$

$$[U_\rho(\beta, \rho)]_j = -\tfrac{1}{2}((Y - X\beta)'V^{-1}_{/\rho_j}(Y - X\beta) + ln|V|_{/\rho_j})$$

$$\text{for } j = 1, ..., k.$$

The score statistics are made up of linear and quadratic forms of Y. The covariance of linear and quadratic terms like these are well known (see Mathai and Provost 1992) and the result follows directly. □

Claim 2: The score statistic of the approximate modified profile likelihood is asymptotically unbiased.

Proof: The bias in the profile score statistic for a parameter of interest ψ can be written as (see Barndorff-Nielsen and Cox 1994)

$$E(l_{p/\psi}(\psi)) \overset{\circ}{=} -\tfrac{1}{2}tr(i^{-1}_{\chi\chi}(\tau_{\chi\chi,\psi} + \tau_{\chi,\chi,\psi}))$$

where $\overset{\circ}{=}$ implies equality to first order. The terms $\tau_{\chi\chi,\psi}$ and $\tau_{\chi,\chi,\psi}$ are functions of joint moments of derivatives of the log likelihood:

$$\tau_{\chi\chi,\psi} = \nu_{\chi\chi,\psi} - \nu_{\chi\chi,\chi}i^{-1}_{\chi\chi}i_{\chi\psi} \text{ and } \tau_{\chi,\chi,\psi} = \nu_{\chi,\chi,\psi} - \nu_{\chi,\chi,\chi}i^{-1}_{\chi\chi}i_{\chi\psi},$$

$$\text{for } \nu_{\chi_i,\chi_j,\chi_k} = E(l_{\chi_i}l_{\chi_j}l_{\chi_k}) \text{ and } \nu_{\chi_i\chi_j,\chi_k} = E(l_{\chi_i\chi_j}l_{\chi_k}).$$

Barndorff-Nielsen (1994a) showed that

$$-ln|\widetilde{l}_{x;\widehat{x}}| + ln|\widehat{j}_{xx}| + \tfrac{1}{2}ln|\widetilde{j}_{xx}| - \tfrac{1}{2}ln|\widehat{j}_{xx}| \stackrel{\mathrm{o}}{=}$$
$$-\tfrac{1}{2}(\psi - \widehat{\psi})tr(\widehat{j}_{xx}^{-1}(\widehat{\tau}_{xx,\psi} + \widehat{\tau}_{x,x,\psi})),$$

where the left hand side is the adjustment term of the modified profile likelihood. Using the results of Barndorff-Nielsen (1994b) it is straightforward to show using a first order Taylor expansion that

$$-ln|S_{xx}| + ln|\widehat{i}| \stackrel{\dagger}{=} -ln|\widetilde{l}_{x;\widehat{x}}| + ln|\widehat{j}_{xx}| - ln(\tfrac{|\widehat{i}|}{|i|}) + C(Y)$$

where $\stackrel{\dagger}{=}$ denotes equality to second order and $C(Y)$ is an $O_p(1)$ function of the data. Clearly, then up to an additive constant

$$-ln|S_{xx}| + ln|\widehat{i}| + \tfrac{1}{2}ln|\widetilde{j}_{xx}| - \tfrac{1}{2}ln|\widehat{j}_{xx}| \stackrel{\mathrm{o}}{=}$$
$$-\tfrac{1}{2}(\widehat{\psi} - \psi)tr(\widehat{j}_{xx}^{-1}(\widehat{\tau}_{xx,\psi} + \widehat{\tau}_{x,x,\psi})).$$

Taking derivatives we see that the score of the approximate modified profile likelihood, $L_{mp}^{\dagger}(\psi)$ is also asymptotically unbiased.

Claim 3: The score of the unadjusted profile likelihood is asymptotically unbiased if the parameter of interest is a fixed effect.

Proof: Suppose that the interest parameter is $\psi = \beta_1$ so that the nuisance parameters are given by $\chi = (\beta_2, \rho)$. We will only consider the case where β_1 is a scalar. Then, as given above

$$E(l_{p/\beta}(\psi)) \stackrel{\mathrm{o}}{=} -\tfrac{1}{2}tr(i_{xx}^{-1}(\tau_{xx,\beta_1} + \tau_{x,x,\beta_1})).$$

By the orthogonality of β_2 and ρ it is clear that $i_{\beta\rho} = 0$ and $i_{\beta\rho}^{-1} = 0$ so that

$$E(l_{p/\beta}(\psi)) \stackrel{\mathrm{o}}{=} -\tfrac{1}{2}tr(i_{\beta_2\beta_2}^{-1}(\tau_{\beta_2\beta_2,\beta_1} + \tau_{\beta_2,\beta_2,\beta_1})) - \tfrac{1}{2}tr(i_{\rho\rho}^{-1}(\tau_{\rho\rho,\beta_1} + \tau_{\rho,\rho,\beta_1})),$$

where the τ terms are given above. Then, using elementary facts concerning the covariance and quadratic forms of Y (Mathai and Provost 1992), it can be shown that

$$\nu_{\beta_i\beta_j,\rho_k} = \nu_{\beta_i,\rho_i\rho_j} = \nu_{\beta_i,\rho_i,\rho_j} = 0 \text{ and } \nu_{\beta_i,\beta_j,\beta_k} = \nu_{\beta_i\beta_j,\beta_k} = 0.$$

From this it follows that $E(l_{p/\beta}(\psi)) \stackrel{\mathrm{o}}{=} 0$. $\square$

Proof of Theorem 3.5.1: Recall the assumptions

(i) $V(\rho) = \sum_{i=1}^{k}\lambda_i(\rho)Q_i$, $\sum_{i=1}^{k}Q_i = 1$ and the Q_i's are symmetric, idempotent and pair-wise orthogonal. Also, $\lambda(\rho)$ is a one to one continuous and differentiable function of ρ.

(ii) $R(Q_i[A\ X_r]) \subset R([A\ X_r])$ for $i = 1$ to k (Zyskind's condition).

(iii) $A'X_r = 0$.

(iv) $R(A) \subset R(Q_i)$ for some i.

For convenience assume that $R(A) \subset R(Q_1)$ which implies that $Q_jA = 0$ $\forall j \neq 1$. The proof has two parts. In the first part we show that the profile likelihood has the form given in the theorem. This is shown primarily by noting that μ is estimation orthogonal to $(\beta, \{\lambda_j\}_{j\neq1})$; i.e., for any fixed μ, $\widetilde{\beta_r} = \widehat{\beta}_r$ and $\widetilde{\lambda}_j = \widehat{\lambda}_j$ for $j \neq 1$. In the second part of the proof we turn our attention to the adjustment term.

First for fixed β and μ the restricted MLE of λ_j is given by

$$\widetilde{\lambda}_j(\mu,\beta) = \frac{(Y-A\mu-X_r\beta_r)'Q_j(Y-A\mu-X_r\beta_r)}{r(Q_j)}$$

which implies that the profile likelihood for (μ,β_r) is

$$L_p(\mu,\beta_r) \propto |\widetilde{V}|^{-\frac{1}{2}} \propto \prod_{j=1}^{k}(1 + \frac{\widetilde{\lambda}_j(\mu,\beta)-\widehat{\lambda}_j(\mu,\beta)}{\widehat{\lambda}_j(\mu,\beta)})^{-\frac{r(Q_j)}{2}}.$$

The last step follows since the k distinct eigenvalues of V are given by λ_j with multiplicity $r(Q_j)$. Next we show that Zyskind's condition and the orthogonality of A and X imply that when interest lies with μ then

$$\prod_{j=2}^{k}(1 + \frac{\widetilde{\lambda}_j(\mu,\widetilde{\beta})-\widehat{\lambda}_j(\widehat{\mu},\widehat{\beta})}{\widehat{\lambda}_j(\widehat{\mu},\widehat{\beta})})^{-\frac{r(Q_j)}{2}}$$

is a function of the data alone.

For fixed μ it can be shown using conditions $(ii) - (iv)$ that

$$R(Q_iX_r) \subset R(X_r)\ \forall i$$

so that Zyskind's condition holds in the sub model with μ held fixed. With $A'X_r = 0$ this implies that for fixed μ, $\widetilde{\beta}_r = \widehat{\beta}_r$. By inspection of the form of $\widetilde{\lambda}(\mu, \beta)$ and the fact that $Q_j A = 0$ for $j \neq 1$ it follows that

$$L_p(\mu) \propto (1 + \frac{\widetilde{\lambda}_1 - \widehat{\lambda}_1}{\widehat{\lambda}_1})^{-\frac{r(Q_1)}{2}}.$$

Next we show that this quantity is related to the usual t statistic.

Observe that

$$\widetilde{\lambda}_1 - \widehat{\lambda}_1 = ((Y - A\mu - X_r\widetilde{\beta}_r)'Q_1(Y - A\mu - X_r\widetilde{\beta}_r)$$
$$- (Y - A\widehat{\mu} - X_r\widehat{\beta}_r)'Q_1(Y - A\widehat{\mu} - X_r\widehat{\beta}_r))/r(Q_1).$$

Zyskind's condition in the full model implies that $(\widehat{\mu}, \widehat{\beta}_r)' = (X'X)^{-1}X'Y$ and we have already shown that $\widetilde{\beta}_r = \widehat{\beta}_r$. By noting that $Q_1 A = A$ and that $X_r'A = 0$ it can be easily shown that

$$\widetilde{\lambda}_1 - \widehat{\lambda}_1 = (A'A)(\widehat{\mu} - \mu)^2 / r(Q_1),$$

where $(A'A)$ is a scalar. Also, the standard error estimate, obtained from REML estimates, or from sums of squares in balanced variance component models, of the variance of $\widehat{\mu}$ is given by,

$$Var(\widehat{\mu}) = \frac{(A'A)^{-1}(Y - A\widehat{\mu} - X_r\widehat{\beta}_r)'Q_1(Y - A\widehat{\mu} - X_r\widehat{\beta}_r)}{(r(Q_1) - r(Q_1 X_r) - 1)} = \frac{\widehat{\lambda}_1(A'A)r(Q_1)}{(r(Q_1) - r(Q_1 X_r) - 1)}$$

Hence,

$$\left(\frac{\widetilde{\lambda}_1 - \widehat{\lambda}_1}{\widehat{\lambda}_1} \right) = \frac{(\widehat{\mu} - \mu)^2}{(r(Q_1) - r(Q_1 X_r) - 1)Var(\widehat{\mu})}$$

which gives

$$L_p(\mu) \propto (1 + \frac{t^2}{df})^{-\frac{r(Q_1)}{2}}$$

where $df = r(Q_1) - r(Q_1 X_r) - 1$ and

$$t = \frac{(\widehat{\mu} - \mu)}{SE(\widehat{\mu})}, \text{ and } SE(\widehat{\mu}) = Var(\widehat{\mu})^{\frac{1}{2}}.$$

To see that t follows a t distribution with df degrees of freedom we first introduce a reparameterization of X_r that will be useful here as well as for calculating the adjustment.

Because of conditions (ii) and (iii) it is possible to find X_1 and X_2 such that

$$R(X_1) + R(X_2) = R(X_r),$$

$$R(X_1) \subset R(Q_1) \text{ and}$$

$$R(X_2) \subset R(Q_1)^{\perp}$$

from which it follows that $X_1' X_2 = 0$. Without loss of generality we can then reparameterize:

$$X_r \beta = X_1 \beta_1 + X_2 \beta_2.$$

This gives

$$(Y - A\widehat{\mu} - X_r\widehat{\beta}_r)'Q_1 (Y - A\widehat{\mu} - X_r\widehat{\beta}_r)$$
$$= Y'(I - P_A - P_{x_1} - P_{x_2})' Q_1 (I - P_A - P_{x_1} - P_{x_2})Y$$

where $P_{x_1} = X_1(X_1'X_2)^{-1}X_1'$ is the orthogonal projection operator on $R(X_1)$ and P_A and P_{x_2} are defined analogously. This term then reduces to the typical ANOVA Type I sums of squares in balanced variance component models and elementary methods show that it is a chi-squared random variable with degrees of freedom $r(Q_1) - r(Q_1 X_r) - 1$. Noting that

$$Cov(\widehat{\mu}, Q_1(I - P_A - P_{x_1} - P_{x_2})Y) = 0$$

we have the result that t follows a t − distribution with df degrees of freedom.

Turning to the adjustment of the modified profile likelihood, equal to the approximate modified profile likelihood in these models, recall that the modified profile likelihood can be written as

$$L_{mp}(\mu) \propto \frac{L_p(\mu)}{L_p(\widehat{\mu})} \left| \widetilde{j}_{xx} \right|^{\frac{1}{2}} \left| S_{xx} \right|^{-1} \left| \widehat{j}_{xx} \right|^{-\frac{1}{2}} \left| \widehat{i}_{xx} \right|.$$

By (i) and (iv) we see that for all $j \neq 1$

$$Q_j(Y - X_r\widetilde{\beta}_r - A\widehat{\mu}) = 0$$

which implies that $\widetilde{j}_{\lambda,\beta} = 0$. The form of $j_{\lambda\lambda}$ simplifies further since $Q_i Q_j = 0$ for $i \neq j$ and the form of $\widetilde{\lambda}$ given above implies that

$$\widetilde{i}_{\lambda\lambda} = \widetilde{j}_{\lambda\lambda} = \mathrm{diag}(\frac{r(Q_1)}{\widetilde{\lambda}_1^2}, \ldots, \frac{r(Q_k)}{\widetilde{\lambda}_k^2})$$

which in turn gives

$$|\widetilde{i}_{\lambda\lambda}| = \prod_{j=1}^{k}(\frac{r(Q_j)}{\widetilde{\lambda}_j^2}).$$

The linearity of V in ρ implies that $S_{\lambda\lambda} = \widetilde{i}_{\lambda\lambda}$ and hence

$$|\widetilde{j}_{xx}|^{\frac{1}{2}}|S_{xx}|^{-1} = |X_r'\widetilde{V}^{-1}X_r|^{-\frac{1}{2}}|\widetilde{j}_{\lambda\lambda}|^{\frac{1}{2}}|\widetilde{i}_{\lambda\lambda}|^{-1}$$

$$= |X_r'\widetilde{V}^{-1}X_r|^{-\frac{1}{2}}|\widetilde{i}_{\lambda\lambda}|^{-\frac{1}{2}}.$$

The assumption that $X_r = [X_1 X_2]$ yields

$$|X_r'V^{-1}X_r| = |X_1'X_1/\lambda_1||X_2'(\sum_{j\neq1}\lambda_j Q_j)X_2|.$$

Recalling from above that $\widetilde{\lambda}_j = \widehat{\lambda}_j$ we have that,

$$L_{mp}(\mu) \propto \frac{L_p(\mu)}{L_p(\widehat{\mu})}|\widetilde{j}_{xx}|^{\frac{1}{2}}|S_{xx}|^{-1}|\widehat{j}_{xx}|^{-\frac{1}{2}}|\widehat{i}_{xx}|$$

$$= \frac{L_p(\mu)}{L_p(\widehat{\mu})}|X_1'X_1/\widetilde{\lambda}_1|^{-\frac{1}{2}}(\frac{r(Q_1)}{\widetilde{\lambda}_1^2})^{-\frac{1}{2}}|X_1'X_1/\widehat{\lambda}_i|^{\frac{1}{2}}(\frac{r(Q_1)}{\widehat{\lambda}_1^2})^{\frac{1}{2}}$$

$$\propto (\frac{\widetilde{\lambda}_1}{\widehat{\lambda}_1})^{\frac{p_1+1}{2}} = (1 + \frac{\widetilde{\lambda}_1-\widehat{\lambda}_1}{\widehat{\lambda}_1})^{\frac{p_1+1}{2}}$$

where $p_1 = r(X_1) = r(Q_1X_1) = r(Q_1X_r)$. This gives the desired result that

$$L_{mp}(\mu) \propto \left(1 + \frac{t^2}{r(Q_1)-r(Q_iX_r)-1}\right)^{\frac{-r(Q_1)+r(Q_1X_r)+1}{2}}.$$

Furthermore, by inspecting the score of $L_{mp}(\mu)$, it is clear that like the profile, the modified profile likelihood is maximized at $\widehat{\mu}$ which is the uniformly best linear unbiased estimator in this model. $\square$

Chapter 4

Decomposition of the $\tilde{\tau}$ Statistic for Curved Exponential Families with Application to the

Gaussian Mixed Linear Models

Benjamin Lyons, Dawn Peters and David Birkes

4. Decomposition of the $\tilde{r}$ Statistic for Curved Exponential Families with Applications to Gaussian Mixed Linear Models

4.1 Abstract

Lyons and Peters (1997a and 1997b) demonstrate that Skovgaard's modified directed likelihood statistic, $\tilde{r}$, and Barndorff-Nielsen's approximate modified profile likelihood, $L^\dagger_{mp}$, are simple methods of improving small sample inference in curved exponential families, such as, Gaussian mixed linear models. This paper discusses a decomposition of the adjustment that forms Skovgaard's $\tilde{r}$ statistic. Like the decomposition of r^* presented for canonical parameters in regular exponential families by Pierce and Peters (1992) and, more generally, in Barndorff-Nielsen, Pierce and Peters (1994), this decomposition incorporates $L^\dagger_{mp}$'s nuisance parameter adjustment. Simplifications of this adjustment are presented for the most common curved exponential family: the Gaussian mixed linear model. These simplifications may help researchers gauge the accuracy of the modified directed deviance.

4.2 Introduction

For regular exponential families and transformation models, Barndorff-Nielsen's modified profile likelihood, L_{mp}, and modified directed likelihood, r^*, can greatly improve small sample inference. Unfortunately, in curved exponential families these statistics are difficult to calculate since they require the specification of an ancillary statistic. To remedy this, Skovgaard (1996) introduced another modified directed likelihood, $\tilde{r}$, that is standard normal to second order and, as shown by Lyons and Peters (1997a), is easy to calculate and very accurate in Gaussian mixed linear models. Barndorff-Nielsen (1994a), using techniques similar to Skovgaard's, introduced an approximate modified profile likelihood, $L^\dagger_{mp}$, for curved exponential families. Lyons and Peters (1997b) showed that $L^\dagger_{mp}$ is simple to construct for Gaussian mixed linear

models and is a sensible generalization of the residual likelihood introduced by Patterson and Thompson (1971).

Our concern here is to investigate the relationship between Barndorff-Nielsen's $L_{mp}^{\dagger}$ and Skovgaard's $\tilde{r}$ and suggest a decomposition of $\tilde{r}$'s adjustment term similar to the decomposition of r^* presented by Pierce and Peters (1992) and Barndorff-Nielsen, Pierce and Peters (1994). By examining the magnitude of the different adjustment terms, researchers may be able to determine if $\tilde{r}$ will be accurate enough for inference. Since $\tilde{r}$ is particularly useful for Gaussian mixed linear models, we derive some simplifications of the adjustment in this setting. We hope that the simplifications outlined will serve as the foundation for a more thorough understanding of $\tilde{r}$'s accuracy. Specifically, it may help determine which designs and models can be analyzed effectively using higher order methods.

We adopt notation similar to that used by Barndorff-Nielsen and Cox (1994) and partition the parameter vector θ into a parameter of interest, ψ, and a nuisance parameter χ. The likelihood of θ will be written as $L(\theta)$ and its log as $l(\theta)$. Differentiation with respect to θ will be denoted by a $/$ followed by a sub-scripted parameter, e.g., the score $U(\theta)$ will be written as $l_{/\theta}$ and the observed information $j(\theta)$ as $-l(\theta)_{/\theta\theta}$. The expected information is denoted by $i(\theta) = E(j(\theta))$. Evaluation of a quantity dependent on θ at the maximum likelihood estimates $(\widehat{\psi}, \widehat{\chi})$ or at the constrained maximum likelihood estimates $(\psi, \widetilde{\chi})$ is denoted by a superior $\frown$ and $\sim$, respectively; e.g., the profile log likelihood for ψ is denoted by $\widetilde{l}$, i.e., $\widetilde{l} = l_p(\psi) = l(\psi, \widetilde{\chi})$. For a matrix X, $R(X)$ will denote the range space, $r(X)$ its rank, $tr(X)$ its trace, and $|X|$ its determinant.

In section 2 we present $\tilde{r}$ and $L_{mp}^{\dagger}$. In section 3 we develop the decomposition of the adjustment and in section 4 we apply this decomposition to Gaussian mixed linear models. Additional technical results are discussed in the appendix.

4.3 Modified Directed Likelihood

The modified directed likelihood corrects the directed likelihood: the signed square root of the likelihood ratio statistic r. For testing a single parameter $\psi = \psi_o$ in the presence of a nuisance parameter χ,

$$r(\psi_o) = \text{sgn}(\widehat{\psi} - \psi_o)(2(l(\widehat{\psi}, \widehat{\chi}) - l(\psi_o, \widehat{\chi})))^{\frac{1}{2}}$$

which, under the null hypothesis, is standard normal with absolute first order, i.e., $O_p(n^{-\frac{1}{2}})$, error. To improve its accuracy Barndorff-Nielsen (1986) proposed adjusting r, obtaining the r^* statistic:

$$r^*(\psi_o) = r(\psi_o) - \frac{1}{r(\psi_o)} log(\frac{r(\psi_o)}{u(\psi_o)}).$$

Barndorff-Nielsen's r^* statistic is standard normal with relative third order, $O_p(n^{-\frac{3}{2}})$, error in moderate deviation regions and second order, $O_p(n^{-1})$, error in large deviation regions. The statistic $u(\psi_o)$ is a function of the full MLEs, the constrained MLEs, the observed information, and two sample space derivatives:

$$u(\psi_o) = [\,(\widetilde{l}_{/\theta;\widehat{\theta}})^{-1}\,(\widehat{l}_{/;\widehat{\theta}} - \widetilde{l}_{/;\widehat{\theta}})\,]_\psi\, |\widehat{j}|^{-\frac{1}{2}}\,|\widetilde{l}_{/\theta;\widehat{\theta}}|\,|\widetilde{j}_{xx}|^{-\frac{1}{2}},$$

where $[\,(\widetilde{l}_{/\theta;\widehat{\theta}})^{-1}\,(\widehat{l}_{/;\widehat{\theta}} - \widetilde{l}_{/;\widehat{\theta}})\,]_\psi$ is the element of this vector corresponding to ψ.

In the curved exponential family setting, an approximate ancillary statistic A can be found so that $\widehat{\theta}$ and A are jointly sufficient (see Barndorff-Nielsen and Cox 1994 ch. 7). The derivatives $l_{/\theta;\widehat{\theta}}$ and $l_{/;\widehat{\theta}}$ are taken with respect to the data, i.e., with respect to $\widehat{\theta}$ while A is held fixed. The sample space derivatives can then be written as

$$l_{/\theta;\widehat{\theta}} = \frac{\partial^2}{\partial\theta\partial\widehat{\theta}}l(\theta;\widehat{\theta}, A) \text{ and } l_{/;\widehat{\theta}} = \frac{\partial}{\partial\widehat{\theta}}l(\theta;\widehat{\theta}, A),$$

the derivatives taken with the ancillary statistic held fixed. Numerical studies, e.g., Pierce and Peters (1992) and Lyons and Peters (1997a) have shown that r^* gives quite accurate p-values and confidence intervals for a variety of models. Unfortunately, the sample space derivatives are difficult to calculate and require the specification of an ancillary statistic.

By accounting for nuisance parameters, the modified profile likelihood generally provides improved inference about ψ compared to the profile likelihood. Like r^*, it is a function of the observed information and sample space derivatives:

$$L_{mp}(\psi) = |\tilde{l}_{/\chi;\hat{\chi}}|^{-1} |\tilde{j}_{\chi\chi}|^{\frac{1}{2}} L_p(\psi_o),$$

where $\tilde{l}_{/\chi;\hat{\chi}}$ and $\tilde{j}_{\chi\chi}$ are the nuisance parameter blocks of $\tilde{l}_{/\theta;\hat{\theta}}$ and $\tilde{j}$ respectively. In many cases L_{mp} is a third order approximation to the conditional or marginal likelihood of the parameter of interest. In all models the score of $L_{mp}(\psi)$, unlike the unadjusted profile likelihood score, is asymptotically unbiased (Barndorff-Nielsen 1994b). Hence, L_{mp} behaves more like a "real" likelihood than the unadjusted profile likelihood.

In practice, however, calculating either r^* or L_{mp} is difficult. Constructing the sample space derivatives requires the specification of an exact or approximate ancillary statistic. It is technically difficult to take the derivative with respect to the MLE holding the ancillary statistic fixed and in curved exponential models there are several possible approximate ancillary statistics and hence neither r^* nor L_{mp} are uniquely defined. For curved exponential families both Barndorff-Nielsen (1994b) and Skovgaard (1996) have proposed approximations to some sample space derivatives that do not require the specification of an ancillary statistic. Rather, both approximations depend on the covariance of the score statistic and the deviance drop and are relatively simple to calculate.

Skovgaard proposed approximating the sample space derivatives with

$$\tilde{l}_{/\theta;\hat{\theta}} \stackrel{\dagger}{=} S\,\widehat{\tilde{j}\tilde{i}}^{-1} \text{ and } \widehat{l}_{/\hat{\theta}} - \tilde{l}_{/\hat{\theta}} \stackrel{\dagger}{=} q\widehat{\tilde{j}\tilde{i}}^{-1},$$

where $\stackrel{\dagger}{=}$ indicates equality to second order. The terms S and q are given by

$$S = Cov_{\theta_1}(\,l_\theta(\theta_1),\ l_\theta(\theta_2))\Big|_{\theta_1=\hat{\theta},\theta_2=\tilde{\theta}} \text{ and }$$

$$q = Cov_{\theta_1}(\,l_\theta(\theta_1),\ l\,(\theta_1) - l(\theta_2))\Big|_{\theta_1=\hat{\theta},\theta_2=\tilde{\theta}}.$$

Applying these approximations to u gives Skovgaard's $\tilde{r}$ statistic which is approximately standard normal with second order relative error in moderate deviation regions and first

order relative error in large deviation regions. Using S itself to approximate $\tilde{l}_{l\theta;\hat\theta}$ in the formula for L_{mp} gives Barndorff-Nielsen's approximate modified profile likelihood, $L^\dagger_{mp}$, a second order approximation to the modified profile likelihood. For Gaussian mixed linear models, Lyons and Peters (1997a and b) show that S and q have compact forms so that $\tilde{r}$ and $L^\dagger_{mp}$ are easy to calculate with existing software.

In regular exponential families, no ancillary statistic is necessary since the MLEs alone are sufficient. In this case, both $\tilde{r}$ and $L^\dagger_{mp}$ reduce to r^* and L_{mp}, and so, $\tilde{r}$ has third order accuracy. In curved exponential families, however, the accuracy of $\tilde{r}$ depends on properties of the ancillary statistic (see Skovgaard 1996). Our decomposition will attempt to account for this.

4.4 Decomposition

Pierce and Peters (1992) explained that in problems with nuisance parameters it is useful to decompose the higher order adjustment, $u(\psi)$, into two parts: one analogous to the adjustment needed in one parameter models and the other pertaining to the nuisance parameters. Our decomposition is similar but in addition highlights a component due to the ancillary statistic.

Using Skovgaard's approximation we can approximate u with u_s:

$$u_s(\psi) = [S^{-1} q]_\psi \, |\tilde{j}_{xx}|^{-\frac{1}{2}} |\hat{j}|^{-\frac{1}{2}} |S_{xx}| |S_{\psi\psi} - S_{\psi x}(S_{xx})^{-1} S_{x\psi}| |\hat{i}|^{-1} |\hat{j}|,$$

here $[S^{-1} q]_\psi$ is the element of the vector corresponding to the interest parameter. To obtain the nuisance parameter adjustment, note that the approximate modified profile likelihood can be written as

$$L^\dagger_{mp}(\psi) = D^\dagger(\psi) \, |\tilde{j}_{xx}|^{-\frac{1}{2}} L_p(\psi),$$

where $D^\dagger(\psi) = \dfrac{|\tilde{j}_{xx}|}{|S_{xx}|}$. Multiplying $L^\dagger_{mp}(\psi)$ by the ψ independent factor, $|\hat{j}_{xx}|^{\frac{1}{2}}$, gives

$$L^\dagger_{mp}(\psi) \propto C^\dagger(\psi) \, L_p(\psi),$$

where

$$C^\dagger(\psi) = D^\dagger(\psi)\{ \frac{|\widehat{j}_{xx}|}{|\widehat{j}_{xx}|} \}^{\frac{1}{2}}.$$

In a regular exponential family, the nuisance parameter adjustment $C^\dagger(\widehat{\psi}) = 1$ since $S_{xx}(\widehat{\psi}) = \widehat{i}$ and the observed and expected information are equal. This is not the case in curved exponential families where $C^\dagger(\widehat{\psi}) \neq 1$. A method for preserving this equality is discussed below.

The other adjustment in regular exponential families is often called the information adjustment since it tends to be large when the adjusted information is small. If no nuisance parameters were present, it would constitute the entire adjustment. Denoted by $\widetilde{u}$ this adjustment is given by

$$\widetilde{u}(\psi) = [S^{-1} q]_\psi |\widehat{j}_{\psi\psi|\chi}|^{-\frac{1}{2}} | S_{\psi\psi|\chi} | \quad \text{where}$$

$$S_{\psi\psi|\chi} = S_{\psi\psi} - S_{\psi x}(S_{\chi\chi})^{-1}S_{\chi\psi} \quad \text{and} \quad j_{\psi\psi|\chi} = j_{\psi\psi} - j_{\psi x}(j_{\chi\chi})^{-1}j_{\chi\psi}.$$

In regular exponential families r^* is

$$r^*(\psi) = r(\psi) - \frac{1}{r(\psi)}ln(C^\dagger(\psi)) + \frac{1}{r(\psi)}ln(\frac{\widetilde{u}(\psi)}{r(\psi)}).$$

Calculating $\widetilde{r}$ in curved exponential families, however, requires an additional function of $|\widehat{i}|^{-1}|\widehat{j}|$:

$$\widetilde{r}(\psi) = r(\psi) - \frac{1}{r(\psi)}ln(C^\dagger(\psi)) + \frac{1}{r(\psi)}ln(\frac{\widetilde{u}(\psi)}{r(\psi)}) + \frac{1}{r(\psi)}ln(\frac{|\widehat{j}|}{|\widehat{i}|}).$$

Large values of the final term above would suggest that the ancillary statistic is influential and may affect the accuracy of the $\widetilde{r}$ statistic.

In one parameter families the Efron-Hinkley, or affine ancillary statistic is defined as

$$a = (\frac{\widehat{j}}{\widehat{i}} - 1)/\widehat{\gamma},$$

where γ is a function of the expected information, the expectation of the square of the observed information, and the expectation of the product of the score and the observed information (see Barndorff-Nielsen and Cox 1994). This statistic is standard normal to first order and hence approximately ancillary. In the one parameter setting, large values of $ln(\frac{|\widehat{j}|}{|\widehat{i}|})$ corresponds to large values of this ancillary statistic.

It is convenient to decompose the ancillary adjustment into two terms: one dependent on the nuisance parameters and another that would be present in a one parameter family. That is, write

$$\frac{1}{r(\psi)} ln\left(\frac{|\widehat{j}|}{|\widehat{i}|}\right) = -\frac{1}{r(\psi)} ln\left(\frac{|\widehat{i}_{xx}|}{|\widehat{j}_{xx}|}\right) - \frac{1}{r(\psi)} ln\left(\frac{|\widehat{j}_{\psi\psi|x}|}{|\widehat{i}_{\psi\psi|x}|}\right).$$

For $\psi = \widehat{\psi}$:

$$C^{\dagger}(\widehat{\psi}) \frac{|\widehat{i}_{xx}|}{|\widehat{j}_{xx}|} = 1$$

so that the first adjustment and the nuisance parameter term cancel in this case. The other adjustment term, the ratio $|\widehat{j}_{\psi\psi|x}||\widehat{i}_{\psi\psi|x}|^{-1}$, would be present in one parameter models. Generally, this ratio would be used to form the affine ancillary statistic in the model defined by the profile likelihood.

4.5 Decomposition for Gaussian Mixed Linear Models

Gaussian mixed models are the most commonly used models where the density often follows a curved exponential family and Skovgaard's $\widetilde{r}$ statistic is particularly useful in this setting. Of interest here are models of the form:

$$Y_n = X\beta + \varepsilon \text{ and } \varepsilon \sim N_n(0, V(\rho)), \tag{1}$$

where X is a known n by p design matrix and V is a positive definite covariance matrix that depends on the unknown covariance parameter vector ρ of dimension k. Either the fixed effects vector β, dimension p, or ρ may be of interest. Since only one dimensional hypotheses of β are considered, we can assume without loss of generality that X has full column rank. We are primarily concerned with cases, such as, repeated measures models, where Y consists of m independent subsets, and hence $V(\rho)$, is block diagonal and approximations improve as m gets large. It is easy to see that the density of Y belongs to an exponential family that may be either regular or curved. If this density belongs to a regular exponential family then $\widetilde{r} = r^*$ and hence its accuracy may improve a great deal.

The accuracy of $\hat{r}$ depends explicitly on the type of mixed model being fit. Most balanced variance component models belong to regular exponential families as do balanced multivariate analysis of variance models. Many unbalanced variance component models and repeated measures models are, however, curved. Sufficient conditions for the density of Y to follow a regular exponential family are given in the appendix. Unlike the Wald and the directed likelihood which are first order standard normal regardless of the design, the accuracy of $\hat{r}$ improves in "nice" mixed models. Our aim in giving the simplifications below then is to help lay a foundation for a more systematic study of when $\hat{r}$ will give accurate inference outside of regular exponential families.

In Gaussian mixed models Lyons and Peters (1997a) show that S and q have a compact form if either a component of ρ or β is of interest. Below $V_{/\rho_i} = \frac{\partial}{\partial \rho_i} V$. Writing

$$ S = \begin{pmatrix} S_{\beta\beta} & S_{\beta\rho} \\ S_{\rho\beta} & S_{\rho\rho} \end{pmatrix}; $$

the component matrices have the form:

$$ S_{\beta\beta} = X'\tilde{V}^{-1}X, $$

$$ [S_{\rho\rho}]_{ij} = \tfrac{1}{2}tr(\hat{V}_{/\rho_i} \tilde{V}^{-1} \tilde{V}_{/\rho_j} \tilde{V}^{-1}) \text{ for } i,j = 1,\ldots,k, $$

$$ S_{\rho\beta} = 0, $$

and the ith column of $S_{\beta\rho}$ is given by

$$ S_{\beta\rho_i} = X'\tilde{V}_{/\rho_i}^{-1}X(\tilde{\beta} - \hat{\beta}), \ i = 1, \ldots, k. $$

Similarly, $q = \begin{pmatrix} q_\beta \\ q_\rho \end{pmatrix}$ where $q_\beta = X'\tilde{V}^{-1}X(\hat{\beta} - \tilde{\beta})$ and the jth elements of q_ρ are given by

$$ [q_\rho]_j = -\tfrac{1}{2} tr(\hat{V}_{\rho_j}(\hat{V}^{-1} - \tilde{V}^{-1})) \text{ for } j = 1 \ldots, k. $$

The simplifications of the adjustment will be presented separately for ρ and β.

4.5.1 *Inference About a Covariance Parameter*

Suppose that the scalar $\psi = \rho_1$ is of interest and $\chi = (\beta, \rho_2)$ so noting that

$S_{\beta\beta} = \tilde{j}_{\beta\beta}$ and that $S_{\beta\rho} = 0$, the nuisance parameter adjustment can be written:

$$ln(C^\dagger(\psi)) = \tfrac{1}{2}ln(\frac{|X'\hat{V}_\rho^{-1}X|}{|X'\tilde{V}_\rho^{-1}X|}) + ln(\frac{|\tilde{j}_{\rho_2\rho_2|\beta}|^{\frac{1}{2}}|\hat{j}_{\rho_2\rho_2|\beta}|^{\frac{1}{2}}}{|S_{\rho_2\rho_2}|}),$$

where $j_{\rho_2\rho_2|\beta} = j_{\rho_2\rho_2} - j_{\rho_2\beta}(X'V_\rho^{-1}X)^{-1}j_{\beta\rho_2}$. The first term corrects for the fixed effects

while the second term corrects for the nuisance covariance parameters. Also, the two

parts of the ancillary adjustment term also simplify since $i_{\beta\rho} = 0$:

$$ln(\frac{|\hat{i}_{xx}|}{|\hat{j}_{xx}|}) = ln(\frac{|\hat{i}_{\rho_2\rho_2}|}{|\hat{j}_{\rho_2\rho_2|\beta}|}) \text{ and}$$

$$i_{\rho_1\rho_1|(\beta,\rho_2)} = i_{\rho_1\rho_1} - i_{\rho_1\rho_2}(i_{\rho_2\rho_2})^{-1}i_{\rho_2\rho_1} = i_{\rho_1\rho_1|\rho_2}.$$

Other notable simplifications occur when either Zyskind's condition (Zyskind

1967) holds or the covariance matrix belongs to a commutative quadratic subspace. As is

shown in the appendix, together these conditions imply that the density of Y belongs to a

regular exponential family and that $\hat{r} = r^*$.

In the appendix we show that for fixed ρ, Zyskind's condition ,

$R(V_\rho^{-1}X) \subset R(X) \forall \rho$, implies that β is estimation orthogonal (see Lindsey 1996):

$\tilde{\beta} = \hat{\beta}$. This in turn implies that $\tilde{j}_{\rho_2\beta} = 0$ so that the adjusted and unadjusted

information matrices for ρ_2 are equal: $\tilde{j}_{\rho_2\rho_2|\beta} = \tilde{j}_{\rho_2\rho_2}$. This simplifies both the nuisance

parameter and ancillary adjustments.

Zyskind's condition holds not only in balanced variance component models but in

many repeated measures models where there is no missing data and a saturated means

model is fit to the fixed effects. Other interesting simplifications occur when the

covariance matrix belongs to a commutative quadratic subspace (CQS).

In balanced variance component models the covariance matrix can usually be

written as

$$V(\rho) = \sum_{i=1}^{k} \lambda_i(\rho) Q_i$$

where the matrices Q_i are known symmetric, idempotent, non-negative definite, and pair-wise orthogonal matrices that sum to the identity and $\lambda(\rho)$ is a one to one continuous function of ρ. When these conditions hold, $V(\rho)$ belongs to a CQS (Seely 1977). This condition does not depend on the form of the fixed effect design matrix X and so may be satisfied without Zyskind's condition. For example, the CQS condition may be satisfied in balanced incomplete block models and split plot models with covariates.

If interest lies in the parameter λ_1 so that $\chi = (\beta, \{\lambda_d\}_{d \neq 1})$, then the information adjustment $\widetilde{u}$ simplifies. Noting that

$$V^{-1}(\rho) = \sum_{i=1}^{k} \tfrac{1}{\lambda_i} Q_i$$

it is easy to see that the covariance parameter block of the observed information is given by $j_{\lambda_1 \lambda_i} = 0$ for $i \neq 1$ and

$$\widehat{j}_{\lambda_i \lambda_i} = \frac{(Y - X\beta)' Q_i (Y - X\beta)}{\lambda_i^3} - \frac{1}{2} \frac{r(Q_i)}{\lambda_i^2}$$

and hence $i_{\lambda\lambda}$ is a diagonal matrix with $(i,i)th$ element given by

$$i_{\lambda_i \lambda_i} = \frac{r(Q_i)}{\lambda_i^2}.$$

The MLE for λ_i is

$$\widehat{\lambda}_i = \frac{(Y - X\widehat{\beta})' Q_i (Y - X\widehat{\beta})}{r(Q_i)}.$$

Hence, we see that $\widehat{i}_{\lambda\lambda} = \widehat{j}_{\lambda\lambda}$ even though the density of Y may not be regular. Since V is linear in λ, we also have $S_{\lambda\lambda} = i_{\lambda\lambda}$. The parameterization orthoganality of λ_1 and $\{\lambda_d\}_{d \neq 1}$ as well as the fact that $S_{\lambda_1 \beta} = 0$ implies that $S_{\lambda_1 \lambda_1 | \chi} = S_{\lambda_1 \lambda_1}$. Finally, the q_ρ term can be written as

$$[q_\rho]_1 = - \frac{r(Q_1)}{2} \left(\tfrac{1}{\widehat{\lambda}_1} - \tfrac{1}{\lambda_1} \right) \text{ so that}$$

$$\widetilde{u} = - \frac{r(Q_i)}{2} \left(\tfrac{1}{\widehat{\lambda}_1} - \tfrac{1}{\lambda_1} \right) |\widehat{j}_{\lambda_1 \lambda_1 | \chi}|^{-\frac{1}{2}}.$$

To illustrate the decomposition when interest lies in a covariance parameter, we will employ a balanced analysis of variance or covariance model with sub-sampling. Suppose that k samples are taken from a experimental units and that the response vector, with length $n = a \times k$, follows the Gaussian mixed linear model:

$$Y_n = X\beta + \varepsilon, \quad \varepsilon \sim N_n(0, V(\tau, \chi)).$$

Assume the covariance matrix $V(\tau, \chi) = \tau I_n + \chi M$ where $M = \overset{a}{\underset{i=1}{\oplus}} J_k$, and J_k is a $k \times k$ matrix of ones. So V can be rewritten as

$$V(\tau, \chi) = \tau(I_n - Q) + (k\chi + \tau)Q,$$

where $Q = \frac{1}{k}M$ is the orthogonal projection operator on $R(M)$ and $I_n - Q$ is the orthogonal projection operator on $R(M)^{\perp}$. Clearly, V belongs to a commutative quadratic subspace. Since the treatments are applied to the experimental units, it is easy to see that $R(X) \subset R(M)$ regardless of the form of X, for instance the columns of X could be either continuous or discrete.

Suppose that interest lies with the sub-sampling variance component τ. In this case $(I_n - Q)X = 0$. This implies that $\tilde{j}_{\beta\lambda_1} = 0$ and with since $j_{\lambda\lambda}$ is diagonal this fact implies estimation orthoganality (see Barndorff-Nielsen and Cox 1994 pg. 99); that is, for fixed λ_1, $\tilde{\lambda}_d = \hat{\lambda}_d$ for $d \neq 1$ and $\hat{\beta} = \tilde{\beta}$. This leads to the entire nuisance parameter adjustment disappearing.

Since V is linear in λ, $S_{\lambda_d\lambda_d} = i_{\lambda_d\lambda_d}$ so that

$$ln\left(\frac{|\tilde{j}_{\lambda_d\lambda_d}|^{\frac{1}{2}} |\hat{j}_{\lambda_d\lambda_d}|^{\frac{1}{2}}}{|S_{\lambda_d\lambda_d}|} \right) = 0.$$

The fixed effect portion of the nuisance parameter adjustment disappears since

$$X'\tilde{V}^{-1}X = X'(\sum_{d\neq i}\frac{1}{\tilde{\lambda}_d}Q_d)X = X'(\sum_{d\neq i}\frac{1}{\hat{\lambda}_d}Q_d)X = X'\hat{V}^{-1}X.$$

Finally, the orthogonality implies that $i_{\lambda_1\lambda_1|\chi} = i_{\lambda_1\lambda_1}$ and $j_{\lambda_1\lambda_1|\chi} = j_{\lambda_1\lambda_1}$ so that one portion of the ancillary adjustment disappears, specifically $|\hat{i}_{\lambda_1\lambda_1|\chi}| = |\hat{j}_{\lambda_1\lambda_1|\chi}|$.

Thus, there is no nuisance parameter adjustment and

$$\tilde{r}(\tau) = r(\tau) + \tfrac{1}{r(\tau)} ln(\tfrac{\tilde{u}}{r(\tau)}) + \tfrac{1}{r(\tau)} ln(\frac{|\hat{i}_{\lambda_d \lambda_d}|}{|\hat{j}_{\lambda_d \lambda_d |\beta}|})$$

where $\tilde{u} = -\tfrac{a(k-1)}{2}(\tfrac{1}{\tilde{\tau}} - \tfrac{1}{\tilde{\tau}})\tfrac{1}{\tilde{\tau}^2}$. If Zyskind's condition holds, i.e., the density belongs to a regular exponential family, then the ancillary adjustment disappears entirely:

$$r^*(\tau) = r(\tau) - \tfrac{1}{r(\tau)} ln(\tfrac{r(\tau)}{\tilde{u}}).$$

In this case, r^* has the same form regardless of the number of fixed effects in the model and the discussion above can be extended for an arbitrary number of random effects as well.

4.5.2 Inference About a Fixed Parameter

Suppose that interest lies with a one dimensional component β_1 where $X\beta = X_1\beta_1 + X_2\beta_2$ so that $\chi = (\beta_2, \rho)$. The nuisance parameter adjustment reduces to

$$lnC^\dagger(\psi) = \tfrac{1}{2}ln(\frac{|X_2' \hat{V}_\rho^{-1} X_2|}{|X_2' \tilde{V}_\rho^{-1} X_2|}) + ln(\frac{|\hat{j}_{\rho\rho|\beta_2}|^{\frac{1}{2}} |\tilde{j}_{\rho\rho|\beta_2}|^{\frac{1}{2}}}{|S_{\rho\rho}|})$$

where $\hat{j}_{\rho\rho|\beta_2} = \hat{j}_{\rho\rho} - \hat{j}_{\rho\beta_2}(X_2' V_\rho^{-1} X_2)\hat{j}_{\beta_2\rho}$. The ancillary adjustment simplifies in a similar manner.

The most interesting simplification, however, lies in the information adjustment. Using results given by Graybill (1983), it is easy to show that

$$S^{-1} q = \begin{pmatrix} S_{\beta\beta}^{-1}q_\beta - S_{\beta\beta}^{-1}S_{\beta\rho} S_{\rho\rho}^{-1}q_\rho \\ S_{\rho\rho}^{-1}q_\rho \end{pmatrix}$$

and that $S_{\beta\beta}^{-1}q_\beta = (\hat{\beta} - \tilde{\beta})$ and hence,

$$[S^{-1} q]_{\beta_1} = (\hat{\beta}_1 - \tilde{\beta}_1) - [S_{\beta\beta}^{-1}S_{\beta\rho} S_{\rho\rho}^{-1}q_\rho]_{\beta_1}.$$

It is common to estimate the variance of $\hat{\beta}$ with $(X'\hat{V}^{-1}X)^{-1}$ so that an estimate of the standard error of $\hat{\beta}_1$ is given by $(\hat{i}_{\beta_1\beta_1|\beta_2})^{-\frac{1}{2}}$. We see then that $\tilde{u}$ can be written as a function of this popular version of this Wald statistic:

$$\tilde{u} = (Wald - [S_{\beta\beta}^{-1}S_{\beta\rho} S_{\rho\rho}^{-1}q_\rho]_{\beta_1}(\hat{i}_{\beta_1\beta_1|\beta_2}|^{\frac{1}{2}})|\hat{i}_{\beta_1\beta_1|\beta_2}|^{-\frac{1}{2}}|\hat{j}_{\beta_1\beta_1|\chi}|^{-\frac{1}{2}}|\tilde{i}_{\beta_1\beta_1|\beta_2}|$$

where $Wald = \frac{(\hat{\beta}_1 - \tilde{\beta}_1)}{|\hat{i}_{\beta_1\beta_1|\beta_2}|^{-\frac{1}{2}}}$.

Pierce and Peters (1992) found it convenient to write $\widetilde{u}$ as a function of the Wald statistic since $ln(\frac{Wald}{r})$ will be small when the likelihood is quadratic in the interest parameter. Below we show that the term $[S_{\beta\beta}^{-1}S_{\beta\rho}\,S_{\rho\rho}^{-1}q_{\rho}]_{\beta_1}|\widehat{i}_{\beta_1\beta_1|\beta_2}|^{\frac{1}{2}}$ is generally $O_p(n^{-\frac{1}{2}})$ while the Wald term is $O_p(1)$. In some special cases this adjustment simplifies as well. The Wald statistic and r are easy to obtain from existing mixed model software so their ratio offers a simple way of gauging $\widetilde{u}$'s size.

Recall that $\widehat{\rho}-\widetilde{\rho}=O_p(n^{-\frac{1}{2}})$, in moderate deviation regions, and hence so are the elements of the difference $\widehat{V}^{-1}-\widetilde{V}^{-1}$ which implies that q_{β} is $O_p(n^{\frac{1}{2}})$. Similarly, $\widehat{\beta}-\widetilde{\beta}=O_p(n^{-\frac{1}{2}})$ so that $S_{\beta\rho}$ is $O_p(n^{\frac{1}{2}})$. Noting that each element of $\widehat{i}$, $S_{\beta\beta}$, and $S_{\rho\rho}$ are $O_p(n)$ we find that the entire term $[S_{\beta\beta}^{-1}S_{\beta\rho}\,S_{\rho\rho}^{-1}q_{\rho}]_{\beta_1}|\widehat{i}_{\beta_1\beta_1|\beta_2}|^{\frac{1}{2}}$ is $O_p(n^{-\frac{1}{2}})$. In addition, if β_1 and β_2 are orthogonal with respect to the observed information, i.e.,

$$X_1'\widetilde{V}^{-1}(\rho)X_2 = 0, \forall \rho,$$

then $\widehat{\rho}-\widetilde{\rho}=O_p(n^{-1})$ and $\widehat{\beta}-\widetilde{\beta}=O_p(n^{-1})$ (see Cox and Reid (1987)) so that $[S_{\beta\beta}^{-1}S_{\beta\rho}\,S_{\rho\rho}^{-1}q_{\rho}]_{\beta_1}|\widehat{i}_{\beta_1\beta_1|\beta_2}|^{\frac{1}{2}}$ is $O_p(n^{-\frac{3}{2}})$.

This entire term simplifies in the special case $\widehat{\beta}_2=\widetilde{\beta}_2$ which, as shown in the appendix, is implied by Zyskind's condition holding in the sub-model and design orthogonality, i.e., $R(V_{\rho}^{-1}X_2)\subset R(X_2)$ and $X_1'X_2=0$. This estimation orthoganality implies that for each of the k columns of $S_{\beta\rho}$

$$S_{\beta\rho_i} = X'\widetilde{V}_{\rho_i}^{-1}X(\widetilde{\beta}-\widehat{\beta})$$
$$= (\widetilde{\beta}_1-\widehat{\beta}_1)\begin{pmatrix}X_1'\widetilde{V}_{/\rho_i}^{-1}X_1\\X_2'\widetilde{V}_{/\rho_i}^{-1}X_1\end{pmatrix},$$

effectively removing $\widetilde{\beta}_2$ from the adjustment. Noting that $\frac{1}{|\widehat{\beta}_1-\widetilde{\beta}_1|}S_{\beta\rho}$ is $O_p(n)$ we see

$$\widetilde{u} = \frac{(\widehat{\beta}_1-\widetilde{\beta}_1)}{|\widehat{i}_{\beta_1\beta_1|\beta_2}|^{-\frac{1}{2}}}(1-[S_{\beta\beta}^{-1}S_{\beta\rho}\,S_{\rho\rho}^{-1}q_{\rho}]_{\beta_1}) = Wald(1-O_p(n^{-\frac{1}{2}})).$$

Furthermore, as shown in the appendix, the remainder term is actually $\frac{1}{r}ln(|1-O_p(n^{-1})|)$, in moderate deviation regions, since

$$R(V^{-1}X_2) \subset R(X_2) \text{ and } X_1'X_2 = 0$$

implies that $X_1'V^{-1}(\rho)X_2 = 0, \ \forall \rho$.

In regular exponential families, β is generally a mean parameter and the likelihood is quadratic in β. Hence, as Pierce and Peters (1992) noted, $\frac{1}{r}ln(\frac{Wald}{r})$ should be small for moderate sample sizes. In this cases the ratio of Wald and r may be an accurate guide to the magnitude of the information adjustment. The appendix contains a more technical discussion of conditions that guarantee estimation and information orthoganality of β_1 and (β_2, ρ).

To illustrate the decomposition in the fixed parameter cases, we will employ a balanced analysis of covariance model with sub-sampling. Suppose there are a blocks of size k so that for $n = a \times k$

$$Y_n = X\beta + \varepsilon, \ \varepsilon \sim N(0, V(\tau, \chi))$$

the covariance matrix $V(\tau, \chi)$ having the same form as in example 1.

Suppose that interest lies in β_1 :

$$E(Y_n) = X_1\beta_1 + X_2\beta_2$$

where X_1 is a covariate that varies with either each sub-sample or with each experimental unit. Suppose further that X_2 is a balanced classification matrix representing treatments measured on each experimental unit. Hence, Zyskind's condition holds in the sub-model: $R(X_2) \subset R(Q) \Rightarrow R(V^{-1}X_2) = R(QX_2) = R(X_2)$. If we assume further that $X_1'X_2 = 0$ then for fixed β_1, $\tilde{\beta}_2 = \hat{\beta}_2$ and $\tilde{u}$ can be simplified:

$$\tilde{u} = \frac{(\hat{\beta}_1 - \tilde{\beta}_1)}{|\hat{i}_{\beta_1\beta_1|\beta_2}|^{-\frac{1}{2}}} (1 - O_p(n^{-1})).$$

The magnitude of the information adjustment is well approximated in this case with the ratio of the Wald statistic to r.

4.6 Conclusion

The greatest difficulty in applying asymptotic methods, with or without higher order adjustment, is determining if one has enough data to obtain accurate p-values and confidence intervals. The decomposition given in section 3 may help researchers use $\hat{r}$ in curved exponential families. By gauging the magnitude of the adjustment terms they can diagnose accuracy problems. In section 4 we showed how these adjustment terms simplify in the case of the most common curved exponential families, the Gaussian mixed linear model, and indicated that the same conditions that result in Y being a member of a regular exponential family leads to simplifications of the adjustment term.

4.7 References

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

Barndorff-Nielsen, O.E. (1994a). Adjusted version of profile likelihood and directed likelihood, and extended likelihood. *Journal of the Royal Statistical Society Series B*, 56, 125-140.

Barndorff-Nielsen, O.E. (1994b). Stable and invariant adjusted profile likelihood and directed likelihood for curved exponential models. Technical Report, Department of Statistics. University of Aarhus, Aarhus, Denmark.

Barndorff-Nielsen, O.E. and Cox, D.R. (1994). *Inference and Asymptotics*. Chapman and Hall, London.

Barndorff-Nielsen, O.E., Pierce, D.A., and Peters, D.A. (1994). Modified directed likelihood: decomposition of the correction term. Unpublished manuscript.

Cox, D. R. and Reid, N. (1987). Parameter orthoganality and approximate conditional inference. *Journal of the Royal Statistical Society Series B*, 49, 1-18.

Graybill, F.A. (1983). *Matrices with Applications to Statistics*. Wadsworth Publishing, Belmont, California.

Lindsey, J.K. (1996). *Parametric Statistical Inference*. Oxford University Press, Oxford.

Lyons, B. and Peters, D. (1997a). A modified directed deviance statistic for mixed linear models. Unpublished manuscript.

Lyons, B. and Peters, D. (1997b). Modified profile likelihood for Gaussian mixed linear models. Unpublished manuscript.

Seely, J. (1977). Minimal sufficient statistics and completeness for multivariate normal families. *Sankhya A,* 39, 170-185.

Skovgaard, I.M. (1996). An explicit large-deviation approximation to one-parameter tests. *Journal of the Bernoulli Society,* 2, 145 -165.

Zyskind, G. (1967). On canonical forms, non-negative covariance matrices and best and simple least square linear estimators in linear models. *Annals of Mathematical Statistics,* 38, 1092-1109.

4.8 Appendix

Claims 1 and 2 below are similar, but less general, to results given by Seely (1977) and Zyskind (1967) respectively. We feel some redundancy is warranted, however, since our methods differ slightly and our motivation differs a great deal from that of those authors. Specifically, we would like to place these results in the context of likelihood based inference and we wish to determine when $\widetilde{r} = r^*$ or when the adjustment terms simplify in some useful way. Claims 3 and 4 are consequences of Claims 1 and 2.

Again, we are concerned with the model $Y_n = X\beta + \varepsilon$ where $\varepsilon \sim N_n(0, V(\rho))$. Here $V(\rho)$ depends on the k dimensional vector ρ and is positive definite for all ρ. The p dimensional vector $\beta \in R^p$ and X is assumed to have full column rank. The first claim establishes sufficient conditions for the density of Y to be a member of a regular exponential family implying that $\widetilde{r} = r^*$, and hence both are standard normal to third order. For convenience, in this section we will let V_ρ denote $V(\rho)$.

Claim 1: Consider the multivariate normal families indexed by (β, ρ):

$$\mathcal{F} = \{N_n(X\beta, V(\rho)) : \beta \in R^p \text{ and } \rho \in \Gamma_\rho\}$$

with $\Gamma_\rho = \{\rho \in R^k : V_\rho \text{ is positive definite}\}$.

Sufficient conditions for the family defined by $\mathcal{F}$ to be regular in the sense defined by Barndorff-Nielsen and Cox (1994 pg. 63), and hence $\widetilde{r} = r^*$, are given by

(i) For $\rho \in \Gamma_\rho$, $V_\rho^{-1} = \sum_{i=1}^{k} \lambda_i(\rho) Q_i$ where Q_i's are known symmetric and linearly

independent matrices. Furthermore, for each $\lambda \in R^k$, $\exists$ a W_λ such that

$W_\lambda = \sum_{i=1}^{k} \lambda_i Q_i$ and if $\lambda \in \Gamma_\lambda = \{\lambda \in R^k : W_\lambda \text{ is positive definite}\}$ then

$W_\lambda = V_\rho^{-1}$.

(ii) $\lambda(\rho): \Gamma_\rho \rightarrow \Gamma_\lambda$ is a bijective continuous function implying, in particular,

that $\lambda(\Gamma_\rho) = \Gamma_\lambda$.

(*iii*) $R(Q_i X) \subset R(X)$ for $i = 1, \ldots, k$, which is equivalent to Zyskind's condition in this model.

Proof: By (*i*)

$$l(\rho,\beta) = -\tfrac{1}{2}Y'(\sum_{i=1}^{k}Q_i\lambda_i)Y + Y'W_\lambda X\beta - \tfrac{1}{2}\beta'X'W_\lambda X\beta - \tfrac{1}{2}ln|W_\lambda|.$$

For fixed λ, condition (*iii*) $\Rightarrow$

$$R(W_\lambda X) \subset R(X) \Rightarrow \exists D_\lambda, \text{ dimension } p \times p, \text{ such that } W_\lambda X = XD_\lambda$$

$$\Rightarrow X'XD_\lambda = X'W_\lambda X \Rightarrow D_\lambda = (X'X)^{-1}X'W_\lambda X.$$

Hence, $Y'W_\lambda X\beta = Y'XD_\lambda\beta = Y'X(X'X)^{-1}X'W_\lambda X\beta.$

This implies that a $p + k$ vector of canonical statistics is

$$T = (Y'Q_1Y, \ldots, Y'Q_kY, (X'X)^{-1}X'Y)$$

and the $p + k$ vector of canonical parameters is

$$\theta = (\lambda, \gamma)$$

where $\gamma = X'W_\lambda X\beta$. For the density indexed by θ, let

$$\Phi = \{ \theta : \int e^{-\frac{1}{2}\theta \cdot T} < \infty \}.$$

It can be shown using a decomposition argument similar to the one presented in Kendall and Stewart (1987 pg. 477) that $\int e^{-\frac{1}{2}\theta \cdot T} < \infty$ if and only if W_λ is positive definite and hence $\Phi = \Gamma_\lambda \times R^p$. By the independence of the Q_i, Φ has dimension $k + p$ so that the above canonical representation is minimal. Now we will show that Φ is open.

W_λ is a continuous function into $\mathcal{K}$, the space of symmetric matrices. Note that $\Gamma_\lambda = W_\lambda^{-1}(\mathcal{D})$ where $\mathcal{D}$ is the subset of $\mathcal{K}$ containing all positive definite positive definite matrices. Since $\mathcal{D}$ is open, Γ_λ and Φ are also open (1994 pg. 62). By (*ii*) (ρ,β) one to one with (λ,γ) so that any density in $\mathcal{F}$ is also a member of a regular exponential family.

To see that $\widetilde{r} = r^*$ is such cases note that in these regular families the mean parameterization defined by $E(T) = \eta$ is one to one with θ so that the density indexed by (ρ, β) can be indexed with η. Furthermore in such cases, $\widehat{\eta} = T$ so that

$\eta(\widehat{\rho}, \widehat{\beta}) = T$. By *Lemma* 1 of Skovgaard (1996), $\eta(\widehat{\rho},\widehat{\beta}) = T$ implies that S and q are exactly equal to the sample space derivatives they approximate which gives $\widetilde{r} = r^*$. $\square$

The simplest way to obtain a λ such that condition (ii) holds is to find a parameterization of V such that $E(Y'Q_iY) = \rho_i$. This will be demonstrated with the following corollary which is applicable to most balanced variance component models.

Corollary to Claim 1: Suppose that V_ρ is an element of a commutative quadratic subspace so that V can be written in terms of its spectral decomposition

$$V_\rho = \sum_{i=1}^{k} \rho_i Q_i$$

where the Q_i's are symmetric, idempotent, pair-wise orthogonal matrices such that

$$\sum_{i=1}^{k} Q_i = 1.$$

If ρ varies freely over the set $\Gamma_\rho = \{\rho \in R^k : V(\rho) \text{ is positive definite}\}$ and if $R(Q_iX) \subset R(X) \forall i$ then $\widetilde{r} = r^*$.

Proof: In this case it can be checked that

$$V^{-1} = \sum_{i=1}^{k} \frac{1}{\rho_i} Q_i \text{ where}$$

$\rho_i = E(Y'Q_iY)$ define the eigenvalues of V_ρ.

Let $\lambda_i = \frac{1}{\rho_i}$ and $W_\lambda = \sum_{i=1}^{k} \lambda_i Q_i$. Note that the mean parameterization $\eta = (\lambda, \beta)$ so that (ii) holds (Barndorff-Nielsen and Cox 1994 pg. 62). $\square$

The next claim shows that Zyskind's condition alone give estimation orthogonality of ρ and β.

Claim 2: If condition (iii) of Claim 1 holds then for fixed ρ the constrained MLE of β is

$$\widetilde{\beta} = (X'X)^{-1}X'Y \text{ and hence } \widehat{\beta} = (X'X)^{-1}X'Y.$$

Proof: It is well known that for fixed ρ the constrained MLE of β is given by

$$\widetilde{\beta} = (X'V_\rho^{-1}X)^{-1}X'V_\rho^{-1}Y.$$

By condition (iii) we have that $R(V_\rho^{-1}X) \subset R(X)$ which implies that

$$V_\rho^{-1}X = X(X'X)^{-1}X'V_\rho^{-1}X$$

since $X(X'X)^{-1}X'$ is the orthogonal projection operator on $R(X)$. Hence

$$\widetilde{\beta} = (X'V_\rho^{-1}X)^{-1}X'V_\rho^{-1}Y = (X'V_\rho^{-1}X)^{-1}(X'V_\rho^{-1}X)(X'X)^{-1}X'Y$$

$$= (X'X)^{-1}X'Y.$$

In particular, this holds for $\rho = \widehat{\rho}$ which implies $\widehat{\beta} = \widetilde{\beta}$. $\square$

Claim 3: If Zyskind's condition holds then for fixed ρ

$$\widetilde{j}_{\beta\rho} = 0.$$

Proof. This is a direct result estimation orthogonality, since in all models

$\widetilde{\chi} = \widehat{\chi} \Leftrightarrow \widetilde{j}_{\chi\psi} = 0$ (see Barndorff-Nielsen and Cox 1994 pg. 99). $\square$

Claim 4: Suppose that

$$E(Y) = X_1\beta_1 + X_2\beta_2$$

and that interest lies in the parameter β_1 so that $\chi = (\beta_2, \rho)$. We will discuss the relationships between the first four conditions concerning the relationship between X and V_ρ and estimation and parameter orthogonality:

(1) Zyskind's in the full model: $R(V_\rho^{-1}X) \subset R(X)$, $\forall\rho$.

(2) Zyskind's in the sub-model: $R(V_\rho^{-1}(\rho)X_2) \subset R(X_2)$, $\forall\rho$.

(3) $X_1'X_2 = 0$

(4) V_ρ^{-1} orthoganality: $X_1'V_\rho^{-1}X_2 = 0$, $\forall\rho$.

(5) Estimation orthoganality of $\beta_2 : \widetilde{\beta}_2 = \widehat{\beta}_2$.

(6) Expected Information orthoganality:

$$\widetilde{\beta}_2 - \widehat{\beta}_2 = O_p(n^{-1}) \text{ and } \widetilde{\rho} - \widehat{\rho} = O_p(n^{-1}).$$

Claim 4.1: (4) $\Rightarrow$ (3).

Proof : All sensible models of interest can be parameterized so that for some ρ_o,

$V_{\rho_o}^{-1} \propto I.\square$

Claim 4.2: (2) and (3) $\Rightarrow$ (5).

Proof : For fixed β_1, $\widetilde{\beta}_2$ is the MLE for the model

$$Y_{\beta_1} = Y - X_1\beta_1 \sim N(X_2\beta_2, V(\rho)).$$

Claim 2 gives, $\widetilde{\beta}_2 = (X_2'X_2)^{-1}X_2'Y_{\beta_1}$

$$= (X_2'X_2)^{-1}X_2'(Y - X_1\beta_1)$$

$$= (X_2'X_2)^{-1}X_2'Y \text{ by (3)}.$$

In particular for $\beta_1 = \widehat{\beta}_1$ we have $\widetilde{\beta}_1 = \widehat{\beta}_1 \square$

Claim 4.3: (4) $\Rightarrow$ (5).

Proof: Clearly, (4) implies that β_1 and β_2 are parameter orthogonal in the sense defined by Cox and Reid (1987) since

$$i_{\beta_1\beta_2} = X_1'V_\rho^{-1}X_2 = 0.$$

Elementary methods show that $i_{\beta_1\rho} = 0$ so that the entire vector β_1 and (β_2, ρ) are orthogonal with respect to the observed information. This implies (Cox and Reid 1987) that $\widetilde{\beta}_2 - \widehat{\beta}_2 = O_p(n^{-1})$ and $\widetilde{\rho} - \widehat{\rho} = O_p(n^{-1})$ $\square$

Claim 4.4: (2) and (3) $\Rightarrow$ (4).

Proof: (2) $\Rightarrow \exists D_\rho$ such that $V_\rho^{-1}X_2 = X_2D_\rho$ which with (3) $\Rightarrow$
$$X_1'V_\rho^{-1}X_2 = X_1'X_2D_\rho = 0. \square$$

Claim 4.5: (1) and (4) $\Rightarrow$ (2).

Proof: As shown in the proof of claim 1, (1) $\Rightarrow \exists D_\rho$ such that $V_\rho^{-1}X = XD_\rho$ where $D_\rho = (X'X)^{-1}X'V_\rho^{-1}X$. (3) and (4) $\Rightarrow$

$$D_\rho = \begin{pmatrix} (X_1'X_1)^{-1} & 0 \\ 0 & (X_2'X_2)^{-1} \end{pmatrix} \begin{pmatrix} X_1'V_\rho^{-1}X_1 & 0 \\ 0 & X_2'V_\rho^{-1}X_2 \end{pmatrix}$$

$$= \begin{pmatrix} (X_1'X_1)^{-1}X_1'V_\rho^{-1}X_1 & 0 \\ 0 & (X_2'X_2)^{-1}X_2'V_\rho^{-1}X_2 \end{pmatrix} = \begin{pmatrix} D_{\rho 1} & 0 \\ 0 & D_{\rho 2} \end{pmatrix} \Rightarrow$$

$$V_\rho^{-1}X = XD_\rho = (X_1 D_{\rho 1} \quad X_2 D_{\rho 2}) \Rightarrow$$

$$V_\rho^{-1}X_2 = X_2(X_2'X_2)^{-1}X_2'V_\rho^{-1}X_2 \Rightarrow$$

$\exists \, D_{2\rho}$ such that $V_\rho^{-1}X_2 = X_2 D_{2\rho} \Rightarrow R(V_\rho^{-1}X_2) \subset R(X_2)$. By symmetry we also have $R(V_\rho^{-1}X_1) \subset R(X_1)$. $\square$

5. Conclusion

This thesis is concerned with applying new higher order asymptotic methods to the venerable subject of mixed linear models. Gaussian mixed linear models are among the oldest class of problems where optimal and exact frequentist methods are not available while higher order asymptotics is a relatively recent arrival to statistics with Skovgaard's seminal paper on $\hat{r}$ being published in 1996. In this way, this thesis amounts to putting old wine in new bottles and I hope I have not spilled too much.

The first paper demonstrated that Skovgaard's modified directed deviance statistic is simple to apply in Gaussian mixed linear models and is worth the trouble because of its simplicity, generality, and accuracy. The second paper showed that Patterson and Thompson's residual likelihood is a special case of Barndorff-Nielsen's approximate modified profile likelihood and discussed the appropriateness of applying the latter if interest lies with a subset of covariance parameters or in the fixed effects. The third and final paper discussed two related topics: decomposing the adjustment term used in $\hat{r}$ in a sensible manner and determining in what models Skovgaard's statistic will be accurate.

The last paper builds a foundation of, what I hope, will not only be a better understanding of the accuracy of Skovgaard's statistic but a reconciliation between likelihood inference in mixed models and other time tested sums of squares methods. The connection begins in noting that Skovgaard's modified directed likelihood is equal to Barndorff-Nielsen's third order accurate r^* statistic when the density Y is a member of a regular exponential family. The regularity of Y depends on the form the mixed model, specifically the form of the covariance matrix and its relationship with the fixed effects. These are the very same conditions discussed by Zyskind and Seely in a different context years before commercial software made likelihood based inference practical.

In practice this means that when applying higher order methods the venerable fundamentals of mixed model design, e.g., balance and the orthogonality of fixed and random effects, still matters. Contrast this with first order likelihood methods where the design does not generally affect the accuracy of either the LRT or the Wald test. I find the fact that the design has no bearing on the accuracy of the likelihood ratio test unsatisfying and the fact that $\tilde{r} = r^*$ when the data is "nice" intuitively appealing.

Of course the converse is also true: $\tilde{r}$ decreases in accuracy when the data is "messy." I also find this, perhaps perversely, appealing since a third order modified directed deviance could, in principle anyway, be constructed using a particular non-unique ancillary. The ambiguous results arising from different ancillaries is not unlike the more familiar ambiguity of different sums of squares giving different F-tests. That is, messy data gives messy third order test statistics. This takes likelihood inference in mixed linear models back to the future where messiness matters once again. This time, however, we might have a chance of quantifying messiness since the accuracy of Skovgaard's statistic is affected by the magnitude of an unspecified ancillary. By actually calculating an ancillary we might arrive at a useful definition of messiness that could serve as a guide to practitioners.

BIBLIOGRAPHY

Barndorff-Nielsen O.E. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika*, 70, 343-365.

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

Barndorff-Nielsen, O.E. (1994a). Adjusted version of profile likelihood and directed likelihood, and extended likelihood. *Journal of the Royal Statistical Society Series B*, 56, 125-140.

Barndorff-Nielsen, O.E. (1994b). Stable and invariant adjusted profile likelihood and directed likelihood for curved exponential models. Technical Report, Department of Statistics. University of Aarhus, Aarhus, Denmark.

Barndorff-Nielsen, O.E. and Cox, D.R. (1989). *Asymptotic Techniques for Use in Statistics*. Chapman and Hall, New York.

Barndorff-Nielsen, O.E. and Cox, D.R. (1994). *Inference and Asymptotics*. Chapman and Hall, New York.

Barndorff-Nielsen, O.E. and McCullagh, P. (1991). A note on the relation between modified profile likelihood and the Cox-Reid adjusted profile likelihood. *Biometrika*, 80, 321-328.

Barndorff-Nielsen, O.E., Pierce, D.A., and Peters, D. (1994). Modified directed likelihood: decomposition of the correction term. Unpublished manuscript.

Burdick, R.K. and Graybill, F.A. (1992). *Confidence Intervals on Variance Components*. Dekker, New York.

Cox, D. R. and Reid, N. (1987). Parameter orthoganality and approximate conditional inference. *Journal of the Royal Statistical Society Series B*, 49, 1-18.

Cox, D. R. and Reid, N. (1992). A note on the difference between profile and modified profile likelihood. *Biometrika*, 79, 408-411.

Cressie, N. and Lahiri, S.N. (1993). The asymptotic properties of REML estimators. *Journal of Multivariate Analysis*, 45, 217-233.

Graybill, F.A. (1983). *Matrices with Applications to Statistics*. Wadsworth Publishing, Belmont, California.

Jiang, J. (1996). REML estimation: asymptotic behavior and related topics. *Annals of Statistics*, 24, 255-286.

Johnson, R.A. and Wichern, W.W. (1983). *Applied Multivariate Statistical Analysis.* Prentice-Hall, Englewood, New Jersey.

Kackar, R.N. and Harville, D.A. (1984). Approximations for standard errors of estimators of fixed and random effects in mixed linear models. *Journal of the American Statistical Association,* 79, 853-861.

Lehmann, E.L. (1986). *Testing Statistical Hypotheses.* Wiley, New York.

Li, Y., Birkes, D., and Thomas, D.R. (1996). The residual likelihood tests for a ratio of variance components in a linear models with two variance components. *Biometrical Journal*, 38, 961-972.

Lindsey, J.K. (1996). *Parametric Statistical Inference.* Oxford University Press, Oxford.

Lyons, B. and Peters, D. (1996). A modified directed deviance statistic for mixed linear models. Technical Report 164, Department of Statistics, Oregon State University.

Lyons, B. and Peters, D. (1997a). A modified directed deviance statistic for mixed linear models. Unpublished manuscript.

Lyons, B. and Peters, D. (1997b). Modified profile likelihood for Gaussian mixed linear models. Unpublished manuscript.

Mathai, A.M. and Provost, S.B. (1992). *Quadratic Forms in Random Variables: Theory and Application.* Marcel Dekker, New York.

Mathew, T. and Sinha, S.B. (1988). Optimum tests for fixed effects and variance components in balanced models. *Journal of the American Statistical Association,* 83, 133-135.

McCullagh, P. and Tibshirani, R. (1990). A simple method for the adjustment of profile likelihoods. *Journal of the Royal Statistical Society Series B,* 52, 325-344.

McKeon, J. (1974). F approximations to the distribution of Hotelling's T_o^2. *Biometrika*, 61, 381-383.

McLean, R.A., Sanders, W.L., and Stroup, W.W. (1991). A unified approach to mixed linear models. *The American Statistician,* 45, 54-64.

Miller, J.J. (1977). Asymptotic properties of maximum likelihood estimates in the mixed model of the analysis of variance. *Annals of Statistic*, 5, 746-762.

Milliken, G.A. and Johnson, D.E. (1984). *Analysis of Messy Data. Volume 1: Designed Experiments*. Van Nostrand Rheinhold Company, New York.

Patterson, H.R. and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, 58, 545-554.

Pierce, D.A. and Peters, D. (1992). Practical use of higher order asymptotics for multiparameter exponential families. *Journal of the Royal Statistical Society Series B*, 54, 701-737.

Pillai, K.C. and Samson. P. (1959). On Hotelling's generalization of T^2. *Biometrika*, 46, 160-171.

Pothoff, R.F. and Roy, S.N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika*, 51, 313-326.

Satterthwaite, F.E. (1941). Synthesis of Variance. *Psychometrika*, 6, 309-316.

Searle, S.R., Casella, G., and McCulloch, C.E. (1992). *Variance Components*. John-Wiley and Sons, New York.

Seely, J.F. and El-Bassiouni, Y. (1983). Applying Wald's variance component test. *Annals of Statistics*, 11, 197-201.

Skovgaard, I.M. (1996). An explicit large-deviation approximation to one-parameter tests. *Journal of the Bernoulli Society*, 2, 145 -165.

Tietjen, G.L. (1974). Exact and approximate tests for unbalanced random effects designs. *Biometrics*, 30, 573-581.

Ting, N., Burdick, R.K., Graybill, F.A., Jeyaratnam, S., and Lu T.-F.C. (1990). Confidence intervals on linear combinations of variance components that are unrestricted in sign. *Journal of Statistical Computation and Simulation*, 35, 135-141.

Wolfinger, R.D. (1996). Heterogenous variance-covariance structures for repeated measures. *Journal of Agricultural, Biological, and Environmental Statistics*, 1, 205 - 230.

Wright, S.P. (1994). Adjusted F tests for repeated measures with the MIXED procedure. 328 SMC-Statistics Department, University of Tennessee.

Wright, S.P. and Wolfinger, R.D. (1997). Repeated measures analysis using mixed models: some simulation results. *Modeling Longitudinal and Spatially Correlated Data, Methods, Applications and Future Directions* (eds T.G. Gregoire, D.R. Brillinger, J.D. Diggle, E. Russek-Cohen, W.G. Warren, R.D. Wolfinger). Lecture Notes in Statistics 122. Springer, New York.

Zyskind, G. (1967). On canonical forms, non-negative covariance matrices and best and simple least square linear estimators in linear models. *Annals of Mathematical Statistics*, 38, 1092-1109.