The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author’s permission.

L’auteur conserve la propriété du droit d’auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-50063-2
Asymptotic Analysis of the One-Way Random Effects Models
Mahdi Alkhamisi
Ph.D, 1999
Graduate Department of Statistics
University of Toronto

Abstract

The one-way random effects model is a one-way linear model where the main effects are random variables. This model is used in animal breeding, biology, environmental experiments and other areas of statistics where the levels for the effects have been selected at random from a population of levels and the experimenter wishes to obtain some information about parameters of the distribution of those levels in the population. The main reasons for the analysis would be:

- estimating parameters of the one-way random effects model or functions of these parameters.
- testing hypotheses about the mean, and the variance components of the model or functions of these parameters.

For unbalanced data, it is often difficult to obtain the exact distribution of a statistic. In such cases we focus on some recent third order asymptotic approximations developed by Fraser and Reid (1995) to provide new approximations for the levels of significance for parameters of the model; this involves constructing third order p-values for a scalar interest parameter in the presence of nuisance parameters. For such investigations the dimension of the initial variable can be reduced by sufficiency or ancillarity. For the unbalanced one-way model ancillarity seems to be the only option. Simulations have been performed which show that the F&R method is considerably better than the standard likelihood ratio method. Furthermore, comparisons with other methods are given with data from various branches of applied statistics.
To my parents
and to Barry
Acknowledgements

I would like to thank all of the members of my extended family; in particular, my brother Hadi and my cousin Faraj Ali, for their continual support regardless of how far we live from each other.

Professor D. A. S. Fraser assigned a challenging problem to me; I thank him for his confidence, guidance, and financial support. Professor D. F. Andrews gave valuable guidance to me while I was finishing my thesis; I thank him for that. I must also thank Professors N. Reid and K. Knight for the many valuable things they have taught me on a variety of topics. I truly appreciated the conversations I have had with Professors D. Brenner, R. Neal, and R. Tibshirani while writing this dissertation. I extend my gratitude to Professors D. E. Cormack, M. W. Donnelly, and K. G. W. Moore for their support during the final stages of writing my dissertation.

For useful comments and tips on software packages, my thanks go to my friend Jeff Tupper; his \LaTeX{} advice has improved the presentation of this thesis. My thanks also go to Tom Glinos for his assistance with the departmental computer facilities. Help with administrative details was graciously provided by Andrea Carter, Laura Kerr, and Sylvia Williams.
## Contents

1 Unbalanced one-way random effects models 1

1.1 Introduction ............................................. 1

1.2 Overview of Chapters .................................. 8

2 Unbalanced Linear Models with Random effects 10

2.1 Introduction ................................................ 10

2.2 The Model .................................................. 11

2.3 Parameter Estimation ..................................... 11

2.3.1 ANOVA Method ........................................ 12

2.3.2 Henderson’s Methods ................................... 17

2.3.3 Maximum Likelihood Method ........................... 19

2.3.4 Restricted Maximum Likelihood ....................... 21

2.4 Computational Procedures ............................... 24

2.4.1 The Newton-Raphson Method ................. 24

2.4.2 Conjugate Direction Algorithms .............. 26
5.1 Introduction ................................................................. 96
5.2 Unbalanced data .............................................................. 97

6 Future Directions ............................................................... 105
6.1 Maximum Likelihood Method .............................................. 105
6.2 Restricted Maximum likelihood Method .............................. 106
6.3 Bayesian approach ......................................................... 106
6.4 Bootstrap ................................................................. 106
6.5 Regression model with random intercept ......................... 107
6.6 Mixture Model ............................................................. 107
Notations ................................................................. 108
BIBLIOGRAPHY ........................................................... 109
### List of Tables

2.1 Moisture contents for Pine Board samples ........................................... 16
2.2 ANOVA for Pine Board samples ............................................................. 17
2.3 Estimates of variance components for Pine ........................................... 23

4.1 Counts of the significance function $p(\eta)$ ........................................ 64
4.2 Data for testing accuracy of $R^*(\eta)$ vs $R(\eta)$ ................................ 66
4.3 Counts of the significance function $p(\mu)$ .......................................... 67
4.4 Data for testing accuracy of $R^*(\mu)$ vs $R(\mu)$ ................................ 67
4.5 Counts of the significance function of $p(\sigma)$ .................................... 68
4.6 Data for testing accuracy of $R^*(\sigma)$ vs $R(\sigma)$ ................................ 69
4.7 Counts of the significance function $p(\eta)$ ........................................... 77
4.8 Data for testing accuracy of $R^*(\eta)$ vs $R(\eta)$ ................................ 77
4.9 Counts of the significance function $p(\mu)$ ........................................... 78
4.10 Data for testing accuracy of $R^*(\mu)$ vs $R(\mu)$ ................................ 78
4.11 Counts of significance function of $p(\sigma)$ ........................................ 79
4.12 Data for testing accuracy of $R^*(\sigma)$ vs $R(\sigma)$ ................................ 79
Chapter 1

Unbalanced one-way random effects models

1.1 Introduction

The linear model considered in this thesis is the unbalanced one-way variance components model. This model is widely used in animal breeding, biology, clinical experiments, reliability, environmental experiments, and other areas of applied statistics. A brief description of the unbalanced one-way variance components model and of related generalizations to the one-way random effects model will be presented. Our interest then lies in developing third order theory for testing scalar interest parameters of the random effects model and in constructing highly accurate approximations to the corresponding tail probabilities.

Some basic ideas concerning variance components models can be presented easily in terms of examples. Suppose that we are interested in the distribution of the weight of newborn cows in Canada and United States. A sample of k farms is drawn at
random and the one way variance components model

\[ y_{ij} = \mu + \alpha_i + e_{ij} \]  

(1.1)
is often used to analyze the data; for this let \( y_{ij} \) designate the weight of the jth cow born on the ith farm, \( \mu \) designate the expected weight of a newborn cow, \( \alpha_i \) the effect on weight due to the ith farm, and \( e_{ij} \) be the within farm variation or random error with \( j = 1, \ldots, n_i, i = 1, \ldots, k \). The \( \alpha \)'s are then treated as random variables and in this context are called random effects. Accordingly the model (1.1) is called the unbalanced one-way random effects model or the one-way variance components model for unbalanced data. By unbalanced data we mean data with unequal number of observations in each subsample. In the present context, components of the model in (1.1) are taken to satisfy the following conditions: \( \alpha_i \sim N(0, \sigma^2_\alpha) \), \( e_{ij} \sim N(0, \sigma^2) \) where \( \alpha_i \) and \( e_{ij} \) are independent random variables. The random effects components are assumed to be from an infinite population of effects or from a very large population of effects, so large to be considered infinite. Finite populations are discussed in Tukey (1957) and Gaylor and Hartwell (1969). Rules for converting some estimation methods from an infinite population to a finite population are given in Searle and Fawcett (1970).

In the literature on variance components models, the issues of estimating variance components, deriving expressions for sampling variances of the resulting expressions, testing the hypothesis that \( \sigma^2_\alpha = 0 \), or testing the hypothesis that \( \sigma^2_\alpha / \sigma^2 \) is some constant have been discussed in some details. A general discussion of these topics is given in Crump (1951), Daniels (1939), Eisenhart (1947), Johnson (1948), Henderson (1953), Graybill(1954) and others. The procedures are often based on equating a wide variety of quadratic forms (not necessarily sums of squares) to their expected values, and can be considered as a generalization of the usual analysis of variance method. Weaknesses of these ANOVA methods are lack of available distributional
properties, possible negativity of estimates, and the absence of useful criteria on which judgments can be made to compare different applications of the ANOVA methods.

An alternative approach involves the use of the likelihood function; Hartley and Rao (1967) proposed an iterative procedure for solving the likelihood equation under the restriction of nonnegative variance components; a normality assumption was included. The behavior of the likelihood function as a real function of the variance components appears to be complex; the likelihood equation may have multiple roots or the maximum likelihood estimate may be a boundary point rather than a root. A second iterative method also based on a normality assumption is the restricted maximum likelihood method (REML) of Patterson and Thompson (1971, 1975). For the one-way random effect model this means we maximize a particular part of the likelihood function that does not include $\mu$. In some generality this is equivalent to the earlier marginal likelihood approach in Fraser (1968).

On the other hand, Rao (1971a, 1971b) proposed procedures for deriving minimum variance quadratic unbiased estimators (MINQUE) of variance components. MINQUE uses no distributional properties of random effects or the random error terms in the model other than the mean and the variance. However, estimators obtained by MINQUE have a minimum generalized variance which is equivalent to the minimum variance property when the normality assumption is included. LaMotte (1973), and Swallow and Searle (1978) also considered estimators with this property. For the one way random effects model (1.1), explicit expressions are available for these estimators (see Townsend and Searle (1971)).

A comparison of quadratic unbiased estimation (MINQUE), maximum likelihood, restricted maximum likelihood and analysis of variance estimators for variance components of the model (1.1) under weaker conditions has been given by Westfall (1987).
Westfall replaces normality assumption by the following conditions:

\( \{\alpha_i\} \) are \( iid \) with \( E|\alpha_i|^{4+\delta} < \infty \)

and

\( \{e_{ij}\} \) are \( iid \) with \( E|e_{ij}|^{4+\delta} < \infty \),

where \( iid \) designates independent and identically distributed. Westfall considered a sequence of models (1.1) with \( \sum_i n \to \infty \) and the following condition to establish the asymptotic normality of \( \sigma^2_a \) and \( \sigma^2 \).

**Assumption 1** (Westfall, 1987). Suppose the subsample sizes \( n_i \) take values from a finite set of distinct integers \( S = \{m_1, \ldots, m_k\} \). Let \( a_i(m_j) \) denote the number of \( n_i \) in \( \{n_1, \ldots, n_k\} \) which equal \( m_j \); assume \( a_i(m_j)/t = p_j + o(t^{-1/2}) \), for all \( j = 1, \ldots, k \). Further, if \( m_j = 1 \) is in \( S \), then assume \( p_j < 1 \).

Conditions of assumption 1 (Westfall, 1987) are a special case of those assumptions made by Hartley and Rao (1967) to prove the asymptotic normality of the maximum likelihood estimates for the general mixed model. Chatterjee and Das (1983) used the weighted least square method with estimated weight to develop a BAN estimator of the variance components of the model (1.1). These authors showed that these estimators are asymptotically equivalent to their corresponding maximum likelihood estimators.

Rao and Kuranchie (1988) derived several estimates for variance components of the regression model with a random intercept and unequal error variances

\[ y_{ij} = \mu + \beta x_i + \alpha_i + e_{ij}. \]

In most of the above methods a negative estimate of \( \sigma^2_a \) is replaced by zero or by a small value, say 0.001 as in Rao and Kuranchie (1988). Later we shall consider extending the domain of the \( \sigma^2_a \) to include some negative values.
Up to 1960 most researchers focused on estimating rather than testing parameters of the model (1.1) or related higher order models. Spjotvoll (1967) considered testing the hypothesis $\theta \leq \theta_0$ vs $\theta \geq \theta_0$ in the one-way random effects models, where $\theta = \sigma^2_\alpha / \sigma^2$; in particular, he considered the case of tests that give high power for alternatives distant from the null hypothesis. For the model (1.1) Donner, Wells and Eliasziw (1989) examined the degree to which two approximations to the F-distribution can provide satisfactory significance levels for testing intraclass correlation $\rho = \rho_0$ against $\rho > \rho_0$, where $\rho = \sigma^2_\alpha / (\sigma^2_\alpha + \sigma^2)$. For testing $\sigma^2_\alpha = 0$ with an unequal error variances assumption in the model (1.1) Jerayatnam and Othman (1985) used the test statistic:

$$ F_0 = \frac{\sum_{i=1}^k (\hat{y}_i - \bar{y})^2/(k - 1)}{\sum_{i=1}^k \frac{1}{k(n_i - 1)} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2}. $$

On the other hand Hill (1965) examined the estimation of variance components in the one way random effects model from a Bayesian point of view and obtained the conditional posterior distribution of $\sigma^2$, $\sigma^2_\alpha$ and $\mu$ given the ratio of variance components $\sigma^2_\alpha / \sigma^2$ for unbalanced data. Tiao and Tan (1965) adopted the same approach to analyze the balanced one way random effects model.

Our present interest lies in estimating the mean and the two variance components of the unbalanced random effects models (1.1) and constructing third order observed level of significance for interest parameters of the model or generalizations of it; this involves calculating third order p-values for functions of these parameters that are valid in the presence of nuisance parameters. Throughout this work we treat the significance function $p(\theta)$ nominally viewed as $p(\hat{\theta} \leq \hat{\theta}^0; \theta)$ as a means to provide a measure of the degree to which the observations support the null hypothesis or contradict it; it is a function of the value of the parameter being tested. In addition, it records probability to left of the data.

Third order asymptotic theory can be used directly to construct significance function of an interest parameter if the dimension of the variable is greater than or equal
to that of the parameter. For such investigations ancillarity and sufficiency are used to reduce the dimension of the variable to that of the parameter. For the unbalanced one-way random effects model ancillarity seems to be the option. For this ancillary statistic suppose that there exists a 1-1 transformation from $y$ to $(s, t)$ such that

$$f(y; \theta) \propto f(s|t; \theta)f(t);$$

then $t$ is ancillary for $\theta$ and inference for $\theta$ is based on the conditional density $f(s|t; \theta)$ for $s$. For example in the location-scale family, $t(y)$ can be the standardized residuals and the conditional density for $s$ can be obtained by a standard calculations. For a rigorous discussion see Fraser (1979) and Reid (1994).

More generally, approximate ancillarity can be used. In third order analyses the $p^*$ formula of Barndorff-Nielsen(1986) is important; the $p^*$ formula provides an approximation to the conditional density of the maximum likelihood estimate $\hat{\theta}$ given an approximate ancillary statistic $t$:

$$p^* = ce^{t(\hat{\theta}; y) - t(\hat{\theta}; y))}|f(\hat{\theta})|^{1/2} \quad (1.2)$$

Once an approximate conditional model is obtained then significance values are available by various methods. The Lugannani and Rice (1980) formula is central to these calculations and is given by

$$p(\hat{\theta} \leq \bar{\theta}; \theta) = \Phi(R) + \phi(R)\left(\frac{1}{R} - \frac{1}{Q}\right) \quad (1.3)$$

where $R$ is the signed likelihood ratio and $Q$ is a specialized maximum likelihood departure that needs to be problem specific. Suppose that $X_1, \ldots, X_n$ are independent, identically distributed random variables from a distribution function $F(X_1)$ and let $K(T)$ denote the cumulative generating function for $X_1$. Then the saddlepoint approximation of the density function of $\bar{X}$ is

$$\bar{f}_n(\bar{x}) = \left(\frac{n}{2\pi K(T_0)}\right)^{1/2} \exp\{n(K(T_0) - \bar{x}T_0)\} \quad (1.4)$$
where $T_0$ is given by $K'(T_0) = \bar{x}$. The relative error in the saddlepoint approximation is $O(n^{-1})$. The relative error can be improved to the third order by renormalization. For the scalar parameter case Daniels (1987) derived the Lugannani and Rice formula from the saddlepoint approximation. An alternative method of combining $R$ and $Q$ is given by Barndorf-Nielsen(1986) is

$$p(\hat{\theta} \leq \hat{\theta}^0; \theta) = \Phi(R + \frac{1}{R} \ln \frac{Q}{R}) = \Phi(R^*)$$

where $R^*$ is an adjusted likelihood ratio defined implicitly in (1.5).

Fraser and Reid(1995) used tangent exponential models to derive a third order likelihood test for an interest parameter. Our interest lies in using the Fraser and Reid(1990,1995) approach to obtain third order significance probabilities for scalar interest parameters in the presence of nuisance parameters for the random effects model (1.1) using the Lugannani and Rice formula or the $R^*$ formula. We will see that this requires only the observed likelihood function at the data point $\ell^0(\theta) = \ell(\theta; y^0)|_{\theta^0}$ and an appropriate gradient $\phi(\theta) = \frac{\partial \ell(\theta; y^0)}{\partial \theta}|_{\theta^0}$ of the likelihood at the data point; where $V = (V_1, \ldots, V_p)$ records tangents to a second order ancillary at the data point.

For the development it is necessary to find an answer to the question: "what does one do with a negative estimate for $\sigma^2_\alpha$?". We suggest four solutions:

a) Put zero in place of a negative value.
b) Eliminate or replace the corresponding factor from the model and then re-estimate.
c) use a Bayesian approach to analyze the model (1.1).
d) Based on the relevant information available in the sample concerning parameters of the model, (e.s the minimal sufficient statistic), the domain of the parameter space should be extended so that the likelihood equations have a solution in the extended parameter space. The domain of the extended parameter space includes negative values of the primary parameters but still nonnegative values for the relevant individual
subsample variances. The latter idea of (d) will be explored in chapter 4.

Simulation will be used to confirm the high accuracy and reliability of the present applications of recent third order asymptotic theory.

With the above ideas in mind we will present several procedures for the estimation and hypothesis testing of scalar interest parameters of the model (1.1) and for obtaining third order approximations to the significance function for a scalar parameter in the presence (or absence) of nuisance parameters.

1.2 Overview of Chapters

Chapter 2

The one-way random effects model for unbalanced data is defined in this Chapter; a mathematical expression for the model is presented and probability distributions for its components are discussed. We then discuss various methods to estimate the parameters of the model and some numerical algorithms to compute point estimates of the mean and variance components. An example to compare the methods is also discussed.

Chapter 3

For the one-way random effects model, the tangent exponential model gives highly accurate approximations for the distribution of a statistic when the exact distribution is not available. In this chapter we describe numerical methods that leads to such approximations for a scalar interest parameter of the one-way model. This includes the derivation of two first order statistics required for the construction of the third order statistic $R^*$. 
Chapter 4

The one way variance components model has three parameters of interest: the mean (say \(\mu, -\infty < \mu < \infty\)), the variance for random effects (say \(\sigma^2\)), and the variance for the experimental errors (say \(\sigma^2 > 0\)). In this chapter we discuss the construction of third order approximation for a scalar interest parameter of the one-way model when the domain of the parameter space is extended to include negative values of the primary parameters but still nonnegative values for the subsample variances. Simulations are also discussed to confirm the accuracy of third order theory.

Chapter 5

We apply the approximations obtained in Chapter 3 and Chapter 4 to unbalanced data sets. The data sets used in this chapter come from different areas of applied statistics. For each data set, details of numerical calculations, and tables are also presented. The new significance results are compared with first order approximations when no exact results are available.

Chapter 6

Generalizations and extensions to problems that arise from this thesis with suggestions and new directions for further investigations.
Chapter 2

Unbalanced Linear Models with Random effects

2.1 Introduction

Linear models are important in both theoretical and applied areas of statistics. Linear models include analysis of variance models, analysis of variance components models, linear regression models and multivariate versions of these.

In the sequel, attention is directed to the one-way random effects models with unbalanced data; by unbalanced data we mean data in which there is not the same number of observations in each subclass. Several parameters in the one-way variance components model are of interest: the mean and the two variance components corresponding to random effects and experimental errors. Our concern lies in testing the two variance components and the mean, and in emphasizing the role of the likelihood function in providing a highly accurate approximation for the observed levels of significance. Furthermore, we discuss statistical inferences such as parameter estimation
and computational procedures for the one-way random effects model.

2.2 The Model

Consider the one-way random effects model,

\[ y_{ij} = \mu + \alpha_i + e_{ij}, \]

(2.1)

where \( \{y_{ij}\} \) is a response array with \( j = 1, 2, \ldots, n_i \) and \( i = 1, 2, \ldots, k \); \( \mu \) is an unknown location parameter; and \( \alpha_i \) and \( e_{ij} \) are taken to be independent normally distributed random variables with zero means and with unknown variances \( \sigma_\alpha^2 \) and \( \sigma^2 \) respectively. We focus on the general case with an unequal number \( n_i \) of observations in the \( k \) subsamples with \( i = 1, \ldots, k \). For general notation let \( \theta' = (\lambda', \psi') \), where \( \lambda \) is a vector of nuisance parameters and \( \psi \) is a scalar interest parameter; our interest lies in finding highly accurate approximations for the significance function \( p(\psi) \), viewed nominally as \( p(\hat{\psi} \leq \psi^0; \psi) \). We shall examine the cases with \( \psi \) equal to \( \sigma_\alpha \), to \( \sigma \) and to \( \mu \).

2.3 Parameter Estimation

There are various methods available for estimating parameters from unbalanced data, we discuss these methods and their properties. A prominent question is "which of these methods is the appropriate or the practical choice?" In recent third order procedures we will find that maximum likelihood values are the needed ingredients. Accordingly we focus our discussion on these methods; they require heavy computation but the needed high speed and the high accuracy is presently available.
use these procedures to obtain highly accurate approximations for observed levels of significance for the mean and the two variance components.

2.3.1 ANOVA Method

The method of estimating parameters by equating various sums of squares to their expected values is generalized. In this generalization a class of quadratic forms in the observations is used instead of the sums of squares.

First we record some needed results on means and variances (Searle 1971). Let $A$ be a symmetric and nonnegative definite matrix; $x \sim N(\mu, V)$ where $V$ is a positive definite matrix. Then

\[ E(x'Ax) = \text{tr}(AV) + \mu' A \mu, \]
\[ \text{var}(x'Ax) = 2\text{tr}(AV)^2 + 4\mu' AV A \mu, \]
\[ \text{cov}(x'Ax, x) = 2VA \mu. \]

Let $\sigma^2$ denote the vector of variance components to be estimated and let $K$ be a vector (with the same dimension as $\sigma^2$) of linearly independent quadratic forms satisfying

\[ E(K) = B\sigma^2 \quad (2.2) \]

where $B$ is nonsingular matrix; then

\[ \hat{\sigma}^2 = B^{-1}K \quad (2.3) \]

is an unbiased estimator of $\sigma^2$ and

\[ \text{var}(\hat{\sigma}^2) = B^{-1} \text{var}(K) B^{-1} \quad (2.4) \]

It follows that the ANOVA method produces unbiased estimators whenever (2.2) and (2.3) exist. On the other hand, the ANOVA method has several disadvantages:
(a) Negative values for the estimator can occur.
(b) There is no unique or optimal way to find components for $K$ that satisfy (2.2).
(c) Distributional properties of the estimators are in general unknown.

Wang (1967) examined the distributions of variance components for the one-way classification and introduced an approximate distribution for random variables in the form $Z = a\chi^2(f_1) - b\chi^2(f_2)$ where $a$ and $b$ are known constants.

The usual ANOVA sums of squares for unbalanced data are

$$SSB = \sum_{i=1}^{k} n_i (\bar{y}_i - \bar{y}_.)^2$$

and

$$SSE = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i.)^2.$$  \hspace{1cm} (2.5)

For

$$n_0 = \left(\sum_{i=1}^{k} n_i - \sum_{i=1}^{k} n_i^2 / \sum_{i=1}^{k} n_i\right) / (k - 1)$$ \hspace{1cm} (2.6)

the expected mean squares

$$E(\text{MSB}) = n_0 \sigma^2 + \sigma^2$$ and $\text{MSE} = \sigma^2$ \hspace{1cm} (2.7)

were first available in Cochran (1939). Now applying the "equate mean squares to their expectations" principle of ANOVA estimation from balanced data gives

$$\text{MSB} = n_0 \hat{\sigma}^2 + \hat{\sigma}^2$$ and $\text{MSE} = \hat{\sigma}^2$. \hspace{1cm} (2.8)

Thus

$$\hat{\sigma}^2 = \text{MSE} = \text{SSE}/\left(\sum_{i=1}^{k} n_i - k\right),$$ \hspace{1cm} (2.9)

$$\hat{\sigma}_\alpha^2 = (k - 1)(\sum_{i=1}^{k} n_i)(\text{MSB} - \text{MSE})/((\sum_{i=1}^{k} n_i^2 - \sum_{i=1}^{k} n_i^2).$$ \hspace{1cm} (2.10)

are the ANOVA estimators. Note that in (2.10) and (2.11) if we replace $n_i$ by $n$ we get the ordinary ANOVA estimators for balanced data.
A Statistic for testing $\sigma_a^2 = 0$.

For balanced data with $n_1 = \ldots = n_k = n$ it is well known that

$$\frac{SSB}{n_0\sigma_a^2 + \sigma^2} \sim \chi^2(k - 1),$$

(2.12)

$$\frac{SSW}{\sigma^2} \sim \chi^2(nk - k),$$

(2.13)

and SSB, SSW are independent; thus

$$\frac{\sigma^2}{n_0\sigma_a^2 + \sigma^2} \frac{MSB}{MSW} \sim F(k - 1, nk - k).$$

(2.14)

The usual inference on variance components is based on these distributional properties. In contrast with unbalanced data, the MSB is not a multiple of a $\chi^2$ when $\sigma^2 > 0$, even though MSB and MSE are independent, (Searle, Casella and McCulloch (1992)). However,

$$\frac{SSB}{n_0\sigma_a^2 + \sigma^2}$$

(2.15)

has a $\chi^2(k - 1)$ distribution whenever $\sigma_a^2 = 0$. If it is known that $\sigma_a^2$ is close to zero then treating (2.15) as a $\chi^2$ may be appropriate. As another option, Wald (1940) derived a test statistic

$$F_w = \frac{\sum_{i=1}^k n_i - k}{k - 1} \frac{S_{wb}}{SSE}$$

(2.16)

where

$$w_i = \frac{n_i}{1 + n_i\sigma_a^2/\sigma^2}$$

$$S_{wb} = \sum_{i=1}^k w_i(y_i - \bar{y}_i) \frac{\sum_{i=1}^k w_i(y_i - \sum_{i=1}^k \bar{y}_i)^2}{\sum_{i=1}^k w_i}$$

on which inference about $\sigma_a$ or the intraclass correlation coefficient $\rho = \sigma_a^2/(\sigma_a^2 + \sigma^2)$ can be based and also showed that $F_w \sim F(k - 1, \sum_{i=1}^k n_i - k)$. Furthermore, $F_w$
reduces to the ratio $F = \frac{MSB}{MSW}$ when $\sigma^2 = 0$. Thus $F$ can be used for testing $\sigma^2 = 0$ although $F$ is not distributed as an $F$ random variable when $\sigma^2 > 0$. For large values of $\frac{\sigma^2}{\sigma^2}$, Spjotvoll (1967) suggested that this test is nearly optimal for alternatives which are distant from zero, when only invariant and similar test are considered. In the balanced case the test reduces to the usual $F$ statistic which Herbach (1959) has proved to be uniformly most powerful invariant and uniformly most powerful unbiased. Later, Thomas and Hultquist (1978) proposed an approximate test statistic

$$F_{TH} = \frac{S^2_0}{\sigma^2 + \frac{\sigma^2}{h}} / \frac{MSE}{\sigma^2}$$

where

$$(k - 1)S^2_0 = \sum_{i=1}^{k} \bar{y}_i^2 - (\sum_{i=1}^{k} \bar{y}_i)^2 / h,$$

$$h = \sum_{i=1}^{k} n_i / k$$

Simulations performed by Thomas and Hultquist (1978) indicate that the chi-square approximation to $(k - 1)S^2_0 / (\sigma^2 + \sigma^2/h)$ is not good when $\sigma^2/\sigma^2 < \frac{1}{4}$ and with extreme unbalance designs.

For both balanced and unbalanced data there is nothing in the preceding that prevents the occurrence of negative estimates for $\sigma^2$. Mathew, Sinha and Sutradhar (1992) cited sufficient conditions under which there is a nonnegative invariant quadratic estimator whose mean square of error is uniformly smaller than every unbiased estimator of $\sigma^2$. Due to lack of completeness of the minimal sufficient statistic, unbiased estimators of $\sigma^2$ are not unique; e.g., the ANOVA estimator and MINQUE estimator are distinct estimators of $\sigma^2$. Moreover, none of these unbiased estimators of $\sigma^2$, which are functions of the minimal sufficient statistic is nonnegative. For this they first obtained conditions under which a biased nonnegative invariant quadratic estimator is uniformly better than every unbiased invariant quadratic estimator of $\sigma^2$. 
Example 2.1 (Pine boards) Ostle and Mensing (1975) investigated data on the moisture content of pine boards. Five storage methods were examined to determine the effect on the moisture content of white pine lumber. Table 2.1 contains the data obtained from this investigation. The model (2.1) is used to analysis the data in

<table>
<thead>
<tr>
<th>Storage Conditions</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>7.3</td>
<td>5.4</td>
</tr>
<tr>
<td>8.3</td>
<td>7.4</td>
</tr>
<tr>
<td>7.6</td>
<td>7.1</td>
</tr>
<tr>
<td>8.4</td>
<td>...</td>
</tr>
<tr>
<td>8.3</td>
<td>...</td>
</tr>
<tr>
<td>Total</td>
<td>39.9</td>
</tr>
<tr>
<td>Number of observations</td>
<td>5</td>
</tr>
<tr>
<td>Number of means</td>
<td>7.98</td>
</tr>
</tbody>
</table>

Table (2.1) where $\alpha_i$ designates the effect of the $i$th level of storage conditions, $e_{ij}$ designates the effect on the moisture content of the $j$th white pine lumber sample due to the $i$th storage condition. The analysis of variance for this data is recorded in the following table.

As $F = 3.34 < F_{35}(4, 9) = 3.63$, the null hypothesis that storage conditions has no effect on the moisture content of white pine lumber produces a right tail p-value that is larger than the familiar 5% significance level. The ANOVA estimates for $\sigma^2_a$ and $\sigma^2$ are

$$\hat{\sigma}^2_a = \frac{2.6656 - .7963}{37/14} = .7073, \hat{\sigma}^2 = .7963.$$
Table 2.2: ANOVA for Pine Board samples

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Degrees of Freedom</th>
<th>Mean square</th>
<th>Expected mean square</th>
<th>F value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Storage conditions</td>
<td>4</td>
<td>2.6656</td>
<td>$\sigma^2 + 2.6429\sigma^2_a$</td>
<td>3.34</td>
</tr>
<tr>
<td>Experimental</td>
<td>9</td>
<td>.7963</td>
<td>$\sigma^2$</td>
<td></td>
</tr>
</tbody>
</table>

2.3.2 Henderson’s Methods

We present a brief summary of the three methods introduced by Henderson (1953) for estimating variance components; also a few details concerning estimators for the 2-way cross classification with $n$ observations per cell will be presented. The three methods proposed by Henderson (1953) are:

Method 1. First: compute sums of squares as in the analysis of variance method of the corresponding orthogonal data. Second: under the assumption of Model II equate these sums of squares to their expectations and then solve for the unknown variances.

Method 2. Compute the least squares estimates of fixed effects in the model. Adjust the original data by these estimates of fixed effects, and then apply Method 1 with adjusted data in place of original data.

Method 3. First: compute the mean squares by the least squares analysis of non orthogonal data (method of fitting constants, weighted squares of means, etc). Second equate these mean squares to their expectations and then solve for the unknown variances.
Method 1 is the simplest to compute and for random effects it produces estimators that are unbiased. Further, Method 1 can be used for mixed models if we consider the fixed effects as random or as non-existent; in both cases the variance component estimators are biased because their expectations under the original mixed model and the new modified model are not equal. Method 2 can not be used for mixed models with interaction between fixed and random factors (Searle 1968). Method 3 can be used for mixed model but is difficult to implement because of the computations needed to calculate the sums of squares.

Now consider the model for the 2-way classification with interaction term under Method 1.

\[ y_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ij}, \]

with \( i = 1, \ldots, a; j = 1, \ldots, b; l = 1, \ldots, n_{ij} \). All the, \( \alpha, \beta, \gamma \) and \( \epsilon \) are random variables with zero mean and variance \( \sigma_{\alpha}^2, \sigma_{\beta}^2, \sigma_{\gamma}^2 \) and \( \sigma_{\epsilon}^2 \) respectively. The effects are also assumed to be uncorrelated with each other. For example \( E(\alpha_i\gamma_{ij}) = E(\alpha_i\beta_j) = 0 \) and \( E(\alpha_i\alpha_l) = 0 \) for \( i \neq l \). Now assume that \( s \) of the \( n_{ij} \) values are different from zero. The sum of squares for first-factor, second-factor, interaction and errors are:

\[
SS_1 = \sum_{i=1}^{a} \frac{y_{i.}^2}{n_i} - \frac{y_{..}^2}{n_..},
\]

\[
SS_2 = \sum_{j=1}^{b} \frac{y_{.j}^2}{n_j} - \frac{y_{..}^2}{n_..},
\]

\[
SS_{12} = \sum_{i=1}^{a} \sum_{j=1}^{b} \frac{y_{ij}^2}{n_{ij}} - \sum_{i=1}^{a} \frac{y_{i.}^2}{n_i} - \sum_{j=1}^{b} \frac{y_{.j}^2}{n_j} + \frac{y_{..}^2}{n_..},
\]

\[
SSE = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{l=1}^{n_{ij}} y_{ijl}^2 - \sum_{i=1}^{a} \sum_{j=1}^{b} \frac{y_{ij}^2}{n_{ij}}.
\]

One can then verify the following expressions:

\[
y_{i.} = \mu n_i + \alpha_i n_e + \sum_{j=1}^{b} \beta_j n_{ij} + \sum_{j=1}^{b} \gamma_j n_{ij} + \epsilon_{i.},
\]
\[ y_{ij} = \mu n_{ij} + \sum_{i=1}^{a} \alpha_i n_{ij} + \beta_j n_{ij} + \sum_{i=1}^{a} \gamma_i n_{ij} + e_{ij}, \]
\[ y_{ijn} = \mu n_{ijn} + \alpha_i n_{ijn} + \beta_j n_{ijn} + \gamma_i n_{ijn} + e_{ijn} , \]
\[ y_{..} = \mu n_{..} + \sum_{i=1}^{a} \alpha_i n_{..} + \sum_{j=1}^{b} \beta_j n_{..} + \sum_{i=1}^{a} \sum_{j=1}^{b} n_{ij} \gamma_{ij} + e_{..} \]

Set \( N = n_{..} , k_1 = \sum_{i=1}^{a} n_{i..}^2 , k_2 = \sum_{j=1}^{b} n_{..j}^2 , k_3 = \sum_{i=1}^{a} (\sum_{j=1}^{b} n_{ij}^2)/n_{i..}, \)
\( k_4 = \sum_{j=1}^{b} (\sum_{i=1}^{a} n_{ij}^2)/n_{..j} , k_5 = \sum_{i=1}^{a} \sum_{j=1}^{b} n_{ij}^2 \cdot \) Then

\[ E(y_{ij}^2) = \mu^2 + \sigma^2_\alpha + \sigma^2_\beta + \sigma^2_\gamma + \sigma^2_\varepsilon, \]
\[ \sum_{j=1}^{b} E(y_{ij}^2/ n_{..j}) = N\mu^2 + \sigma^2_\alpha k_4 + N\sigma^2_\beta + \sigma^2_\gamma k_4 + b\sigma^2_\varepsilon, \]
\[ \sum_{i=1}^{a} E(y_{ij}^2/ n_{i..}) = N\mu^2 + N\sigma^2_\alpha + \sigma^2_\beta k_3 + \sigma^2_\gamma k_3 + a\sigma^2_\varepsilon, \]
\[ \sum_{i=1}^{a} \sum_{j=1}^{b} E(y_{ij}^2/ n_{ij}) = N(\mu^2 + \sigma^2_\alpha + \sigma^2_\beta + \sigma^2_\gamma) + s\sigma^2_\varepsilon, \]

and

\[ E(SS_1) = (N - \frac{1}{N} k_1)\sigma^2_\alpha + (k_3 - \frac{1}{N} k_2)\sigma^2_\beta + (k_3 - \frac{1}{N} k_3)\sigma^2_\gamma + (a - 1)\sigma^2_\varepsilon, \]
\[ E(SS_2) = (k_4 - \frac{1}{N} k_1)\sigma^2_\alpha + (N - \frac{1}{N} k_2)\sigma^2_\beta + (k_4 - \frac{1}{N} k_3)\sigma^2_\gamma + (b - 1)\sigma^2_\varepsilon, \]
\[ E(SS_{12}) = (\frac{1}{N} k_1 - k_4)\sigma^2_\alpha + (\frac{1}{N} k_2 - k_3)\sigma^2_\beta + (N - k_3 - k_4 + \frac{1}{N} k_3)\sigma^2_\gamma + (b - 1)(a - 1)\sigma^2_\varepsilon, \]
\[ E(SSE) = (N - s)\sigma^2_\varepsilon. \]

One can then equate the above four expected values to their observed values and obtain four equations in the four variance components. Solving the resulting four equations yields expressions for the four variance components

2.3.3 Maximum Likelihood Method

Maximization of the likelihood function was introduced as a method of estimation by Fisher(1922,1925) under the name of maximum likelihood estimation (MLE). Fisher's MLE method successfully attained a widespread popularity because of its
many positive statistical properties. During the past 40 years maximum likelihood estimation of means and variance components from unbalanced data has received less attention than other methods of estimation due to its complexities. One issue is that the standard normal equations may produce no explicit solutions for the estimators. Hartley and Rao (1967), Hartley and Vaunghn (1972), and Jennrich and Sampson (1976) introduced algorithms for maximum likelihood estimation of the mean and the variance components in the mixed analysis of variance model and proved that under certain restrictions the estimates are consistent and asymptotically normal as the size of the experiment design increases. In the present context, asymptotic normality of \( \hat{\theta} \) is also required for constructing significance functions.

The model

The existence of a solution to the following equation

\[
\frac{S^2 + \sum_{i=1}^{k} \frac{n_i(y_i - \bar{y})^2}{(\tau^2 + \gamma^2 n_i + 1)}}{N + \sum_{i=1}^{k} \frac{1}{(\tau^2 + \gamma^2 n_i + 1)^2}} = \frac{\sum_{i=1}^{k} \frac{n_i(y_i - \bar{y})^2}{(\tau^2 + \gamma^2 n_i + 1)}}{\sum_{i=1}^{k} \frac{n_i}{(\tau^2 + \gamma^2 n_i + 1)}}
\]

(2.18)

ensures the existence of the mle for the vector parameter \( \theta = (\sigma^2, \sigma^2, \mu) \) as a solution to the normal equations; see section 2.4. On the other hand, if (2.18) has no solution then the maximum likelihood solution \( \hat{\theta} \) is given by

\[
\hat{\mu} = \bar{y}, \quad \hat{\sigma}^2 = 0, \\
\sigma^2 = TSS / \sum_{i=1}^{k} n_i.
\]

For convenience and descriptive value we call such points boundary points:

**Definition 2.1** A vector parameter \( \theta = (\sigma^2, \sigma^2, \mu) \) is said to be a boundary point if \( \hat{\theta} \) is not a solution of equation (2.18).
For the balanced case where \( n_1 = \ldots = n_k = n \), the mle for \( \theta \) is given by

\[
\hat{\mu} = \bar{y},
\]

\[
\hat{\sigma}^2 = \begin{cases} 
\frac{((1 - \frac{1}{k})MSB - MSE)}{n} & \text{if } (1 - \frac{1}{k})MSB \geq MSE \\
0 & \text{otherwise}, 
\end{cases}
\]

\[
\hat{\sigma}^2 = \begin{cases} 
MSE & \text{if } (1 - \frac{1}{k})MSB \geq MSE \\
\frac{TSS}{(kn)} & \text{otherwise.}
\end{cases}
\]

2.3.4 Restricted Maximum Likelihood

Patterson and Thompson (1971) proposed transformations that can be used to partition the likelihood function under normality into two parts, one part is free of the vector \( \mu \). The maximum likelihood estimators from this part are called the restricted maximum likelihood estimators. For balanced data, the restricted maximum likelihood estimators (REML) are unbiased and identical to the ANOVA estimators; by contrast the direct maximum likelihood estimators are not unbiased. With unbalanced data both methods yield biased estimators. We will consider the general mixed model as introduced by Hartley and Rao (1967), and Corbeil and Searle (1976).

\[
y = X\mu + \sum_{i=1}^{c} U_i b_i + e
\]

where

\[
y \sim N(X\mu, H\sigma^2), \quad H = \sum_{i=1}^{c} \gamma_i U_i U_i' + I_N, \ \gamma_i = \frac{\sigma_i^2}{\sigma^2};
\]

\( U_i \) is an \( N \times m_i \) matrix of known numbers with \( \sum_{i=1}^{c} m_i + k < N \), \( \mu \) is a \( k \times 1 \) vector of unknown constants,
\( b_i \) is an \( m_i \times 1 \) vector of i.i.d \( N(0, \sigma_i^2) \),

\( e \) is an \( N \times 1 \) vector of i.i.d \( N(0, \sigma^2) \),

\( X \) is an \( N \times k \) matrix of known numbers, \( k \leq N \),

\( b_1, \ldots, b_c \) and \( e \) are mutually independent,

\( X \) and \( U_i \) are of full rank.

The key idea is to partition the following log-likelihood function into two parts one of which is free of fixed effects \( \mu \).

\[
\ell = -\frac{1}{2} N \ln(\sigma^2) - \frac{1}{2} \ln(|H|) - \frac{1}{2\sigma^2} (y - X\mu)'H^{-1}(y - X\mu).
\]

(2.20)

For this, Patterson and Thompson (1971) proposed the singular transformation \( y' [S : H^{-1} X] \) where

\[
S = I - X(X'X)^{-1}X' = \sum_{i=1}^{k} (J_{n_i} - \frac{1}{n_i} J_{n_i}),
\]

(2.21)

\( J_{n_i} \) is an \( n_i \times n_i \) matrix with every entry unity and \( \sum_{i=1}^{k} + \) represents a direct sum of matrices. Since \( SX = 0 \), \( Sy \) is distributed \( N(0, S\Sigma S'\sigma^2) \) independently of \( X'H^{-1}y \). Thus the distribution of \( Sy \) is free of fixed effects \( \mu \) and the corresponding likelihood function can be used to derive the REML estimators of the variance components involved in \( H\sigma^2 \). To avoid the singularity of \( S\Sigma S' \) arising from \( S(2.21) \), a new matrix \( M \) can be constructed from \( S \) by deleting its \( n_1 \)th, \( (n_1 + n_2) \)th, \ldots, \( (n_1 + \ldots + n_k) \)th rows. The resulting matrix \( M \) has a dimension of order \( (N - k) \times N \) and \( MX = 0 \).

Let \( z \) denote the following transformation:

\[
\begin{bmatrix}
    My \\
    X'H^{-1}y
\end{bmatrix}
\]

then

\[
z \sim N \left( \begin{bmatrix}
    0 \\
    X'H^{-1}X\mu
\end{bmatrix}, \begin{bmatrix}
    M\Sigma M'^2 & 0 \\
    0 & X'H^{-1}\Sigma\sigma^2
\end{bmatrix} \right)
\]

22
Clearly the log-likelihood function for $x$ can be expressed as the product of the log-likelihood function $\ell_1$ of $My$ and the log-likelihood function $\ell_2$ of $X'H^{-1}y$; where

$$
\ell_1 = -\frac{1}{2}(N - k) \ln(\sigma^2) - \frac{1}{2} \ln(|MM'X|) - \frac{1}{2} y'M'(MM')^{-1}My/\sigma^2, \quad (2.22)
$$

$$
\ell_2 = -\frac{k}{2} \ln(\sigma^2) - \frac{1}{2} \ln(|X'H^{-1}X|) - \frac{1}{2\sigma^2} (y - X\mu)'H^{-1}X(X'H^{-1}X)^{-1}X'H^{-1}(y - X\mu). \quad (2.23)
$$

With $\ell_1$ free of $\mu$, those values of $\sigma^2$ and $\gamma_i$'s that maximize the log-likelihood function $\ell_1$ are called the restricted maximum likelihood estimators (REML). For further details see Patterson and Thompson (1971).

**Example 2.2 (Pine Boards revisited)** The following table gives a comparison of the analysis of variance, maximum likelihood and restricted maximum likelihood estimates for the Pine Boards Data.

<table>
<thead>
<tr>
<th>Method of estimation</th>
<th>Random effects $\sigma_\alpha^2$</th>
<th>Experimenter error $\sigma^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANOVA</td>
<td>.7073</td>
<td>.7963</td>
</tr>
<tr>
<td>MLE</td>
<td>.51879</td>
<td>.77299</td>
</tr>
<tr>
<td>REML</td>
<td>.71715</td>
<td>.77735</td>
</tr>
</tbody>
</table>

Clearly the ML estimator of $\sigma_\alpha^2$ is different from those produced by the ANOVA method and REML method.
2.4 Computational Procedures

We now present numerical procedures to compute point estimates of the mean and variance components from unbalanced data. These procedures are based upon the maximum likelihood method (Fisher, 1925). Hartley and Rao (1967) and Corbeil and Searle (1976) applied the maximum likelihood method and the restricted maximum likelihood method to estimate fixed effects and variance components from a mixed model.

2.4.1 The Newton-Raphson Method

The maximum likelihood estimate can be obtained by maximizing the likelihood function \( \ell(\theta) \).

\[
\ell(\theta) = -\frac{1}{2} \sum_{i=1}^{k} \ln \sigma_i^2 - \frac{1}{2} \sum_{i=1}^{k} \frac{1}{\sigma_i^2} (\bar{y}_i - \mu)^2 - \frac{N}{2} \ln \sigma^2 - \frac{1}{2\sigma^2}S^2 \tag{2.24}
\]

where

\[
S = \sigma \chi , \\
\bar{y}_i = \mu + \sigma_i z_i , \\
\sigma_i^2 = \sigma^2 + \frac{1}{n_i} \sigma^2 \ (i = 1, \ldots, k),
\]

\( \chi^2(\sum_{i=1}^{k} n_i - k) \) is the Chi square random variable, and \( z_i \sim N(0,1) \) is the standard normal random variable.

In the present context, we compute the maximum likelihood estimate for \( \theta = (\sigma_\alpha, \sigma, \mu) \) in two steps.

**Step a.** Compute a starting value for \( \theta \).

(i) The three likelihood equations are derived by setting the gradient of the likelihood
function $\ell(\theta)$ equal to zero.

$$\hat{\mu} = \frac{\sum_{i=1}^{k} \frac{y_i}{\sigma_i^2}}{\sum_{i=1}^{k} \frac{1}{\sigma_i^2}},$$  \hspace{1cm} (2.25)

$$\frac{N}{\hat{\sigma}^2} - S^2 = \sum_{i=1}^{k} \frac{1}{n_i \hat{\sigma}_i^4} (\hat{y}_i - \hat{\mu})^2 - \sum_{i=1}^{k} \frac{1}{n_i \hat{\sigma}_i^2},$$ \hspace{1cm} (2.26)

$$\sum_{i=1}^{k} \frac{1}{\hat{\sigma}_i^2} = \sum_{i=1}^{k} \frac{1}{\sigma_i^2} (\hat{y}_i - \hat{\mu})^2.$$  \hspace{1cm} (2.27)

(ii) The above equations cannot be solved explicitly for the three parameters and so some type of numerical method must be employed. In the present context this can be dealt with by a substitution of the form $\gamma = \sigma_\alpha / \sigma$. Then partially reparametrizing equations (2.25), (2.26) and (2.27) respectively yields

$$\hat{\mu} = \frac{\sum_{i=1}^{k} \frac{n_i \hat{y}_i}{(\gamma^2 n_i + 1)}}{\sum_{i=1}^{k} \frac{n_i}{(\gamma^2 n_i + 1)}},$$ \hspace{1cm} (2.28)

$$\hat{\sigma}^2 = \frac{S^2 + \sum_{i=1}^{k} \frac{n_i (\hat{y}_i - \hat{\mu})^2}{(\gamma^2 n_i + 1)^2}}{N + \sum_{i=1}^{k} \frac{1}{(\gamma^2 n_i + 1)}},$$ \hspace{1cm} (2.29)

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{k} \frac{n_i (\hat{y}_i - \hat{\mu})^2}{(\gamma^2 n_i + 1)^2}}{\sum_{i=1}^{k} \frac{n_i}{(\gamma^2 n_i + 1)}}.$$ \hspace{1cm} (2.30)

(iii) Now combine the three expressions in (ii) into a single equation in one unknown parameter $\gamma$ of the form

$$\frac{S^2 + \sum_{i=1}^{k} \frac{n_i (\hat{y}_i - \hat{\mu})^2}{(\gamma^2 n_i + 1)^2}}{N + \sum_{i=1}^{k} \frac{1}{(\gamma^2 n_i + 1)}} = \frac{\sum_{i=1}^{k} \frac{n_i (\hat{y}_i - \hat{\mu})^2}{(\gamma^2 n_i + 1)^2}}{\sum_{i=1}^{k} \frac{n_i}{(\gamma^2 n_i + 1)}}.$$ \hspace{1cm} (2.31)

Clearly, equation (2.31) cannot be solved explicitly for $\gamma$ and an iterative procedure is required to compute $\gamma$. First solve equation (2.31) for $\gamma$ and then compute a first
approximation to the desired maximum $\hat{\theta}$ by using the current value of $\gamma$ and those equations in (ii). This can be accomplished by using software packages such as Maple, Splus, or SAS.

**Step b.** Start with a first approximation say $\theta_0$ then calculate $\theta_1, \theta_2, \ldots$ by the iterative relation

$$\theta_{k+1} = \theta_k - (H(\theta_k))^{-1}\nabla \ell(\theta_k)$$

where $H(\theta_k)$ denotes the value of the negative definite Hessian matrix of $\ell(\theta)$ at $\theta_k$ and the notation $\nabla \ell(\theta_k)$ designates the gradient of $\ell(\theta)$ at $\theta_k$. The process is terminated by the following stopping rules.

$$\|\theta_{k+1} - \theta_k\| < \delta, \quad |\ell(\theta_{k+1}) - \ell(\theta_k)| < \epsilon$$

where $\delta$ and $\epsilon$ are pre-assigned values.

### 2.4.2 Conjugate Direction Algorithms

Suppose that $F(x)$ is a positive definite quadratic function that has the form

$$F(x) = \frac{1}{2}x^T Ax + b^T x. \quad (2.32)$$

It known that finding the solution for a system of linear equations $Ax + b = 0$ is equivalent to the problem of finding the minimum for the quadratic function $F(x)(2.32)$. The general algorithm for finding the minimum of the function $F(x)$ in the x-space can be described as: Let $x_0$ be the initial point in the x-space and $h_0$ be the initial search direction. Start the search process at $x_0$ in the x-space and then proceed from there to the minimum along the search direction $h_0$. Call this point $x_1 = x_0 + t_0 h_0$, (see Fig 1).
Repeat the above step until the sequence $x_0, x_1, \ldots$ converges to the minimum point, that is, to the solution of the system of linear equations.

Of course, the search algorithm chooses the search directions and also the distance by which the search moves along these directions in such a way that $F(x)$ continually decreases. For computational reasons, the gradient of the function $F(x)$ is needed

$$\nabla F(x) = Ax + b. \quad (2.33)$$

In moving from $x$ in the search direction $h$ the search process travels through the points $x_i = x + th$ at which the quadratic function $F(x_i)$ has the form

$$F(x_i) = F(x + th) = \frac{1}{2}(x + th)^TA(x + th) + b^T(x + th)$$

$$= F(x) + th^T \nabla F(x) + \frac{1}{2} t^2 h^TAh.$$

The minimum for the above quadratic polynomial in $t$ is

$$t_m = -\frac{h^T \nabla F(x)}{h^T Ah}. \quad (2.34)$$

Thus (with $x_m = x_{t_m}$)

$$F(x_m) = F(x) - \frac{1}{2} \frac{(h^T \nabla F(x))^2}{h^T Ah}. \quad (2.35)$$

27
In fact $F(x_t)$ lies below $F(x)$ for all points in the interval $(0, 2t_m)$, therefore one needs to choose $t$ in this interval in order to reach the minimum point. Furthermore, the gradient $\nabla F(x_t)$ can be expressed as a function of $t$

$$\nabla F(x_t) = Ax_t + b = \nabla F(x) + tAh.$$  

And from

$$h^T \nabla F(x_t) = h^T \nabla F(x) + t(h^T Ah) = (t - t_m)h^T Ah;$$

it follows that the gradient $\nabla F(x)$ is orthogonal to the search direction $h$ at $x_m$.

The various methods, steepest descent, conjugate gradient, Gauss-Seidel, etc. differ only in the choice of the search direction $h_k$ and the choice of the distance $t$ in the formula $x_{k+1} = x_k + th_k$ for the new point.

**Definition 2.2** Let $D$ be an $n \times n$ symmetric matrix. The set of search direction $d_i (i = 1, 2 \ldots)$ are said to be mutually conjugate with respect to $D$ if $d_iDd_j = 0$ whenever $i \neq j$.

We shall discuss only the steepest descent method and the conjugate gradient method; see Bazaraa and Shetty (1979), Box and Draper (1987), McCormick (1983) and McIntosh (1982).

**Steepest Descent Method** Choose an initial point $x_0$ and the negative gradient $-\nabla F(x_0)$ as the initial search direction. Then determine points $x_1, x_2, \ldots$ by the iterative relation

$$x_{k+1} = x_k - \hat{\lambda}_k \nabla F|_{x_k}$$

where $\hat{\lambda}_k$ is a scalar which minimizes $F(x_k + \lambda \nabla F|_{x_k})$. The iterative process terminates if and when the difference between the values of the $x$-points is smaller than a prescribed tolerance.
Conjugate Gradient Method The method of steepest descent can be improved as follows: After the search algorithm has arrived at the minimum point \( x_k \) (coming from \( x_{k-1} \) along the search direction \( h_k = x_{k-1} + \delta h_{k-1} \)), the domain for the search algorithm extends to include the whole plane \( E_k \) spanned by the vectors \( -\nabla F(x_k) \) and \( h_{k-1} \) which pass through the point \( x_k \). (see Fig 2).

![Diagram](https://via.placeholder.com/150)

**Figure 2.** Choice of the new direction \( h_k \) conjugate to \( h_{k-1} \)
on the plane \( E_k \) spanned by \( -\nabla F \) and \( h_{k-1} \).

If we consider the curves in which this subspace \( E_k \) intersects the level surface \( F = \text{constant} \), we find that these curves are concentric ellipses, one of these ellipses touches the vector \( h_{k-1} \) at \( x_k \). The center \( m \) for that particular ellipse is the required minimum of \( F \) on \( E_k \). To reach this minimum choose at \( x_k \) a new direction \( h_k \) in the plane \( E_k \) conjugate to \( h_{k-1} \).

\[
h_k = -\nabla F(x_k) + \delta_{k-1} h_{k-1},
\]

\[
h_k^T A h_{k-1} = 0.
\]

Substituting (2.36) in (2.37) yields

\[
\delta_{k-1} = \frac{(\nabla F(x_k))^T A h_{k-1}}{h_{k-1}^T A h_{k-1}}.
\]
Thus $\delta_{k-1}$ determines the search direction $h_k$. In this search direction $h_k$ we find the minimum point which necessarily is the center of the ellipse and thus

$$x_{k+1} = x_k + s_k h_k,$$

where by (2.34)

$$s_k = -\frac{h_k \nabla F(x_k)}{h_k^T A h_k}.$$

### 2.5 Maximum Likelihood Estimates and Maximum Likelihood Solutions.

The maximum likelihood method requires that the likelihood function $l(\theta)$ be maximized over the parameter space. Therefore, the maximum likelihood solutions are the MLEs whenever they lie in the parameter space:

$$\Theta = \{ \theta \in \mathbb{R}^3 | \sigma_\alpha^2 > 0, \sigma^2 > 0, -\infty < \mu < \infty \}.$$  

For brevity denote the solutions to the maximum likelihood equations by $\bar{\mu}$, $\bar{\sigma}^2$, and $\bar{\sigma}_\alpha^2$, if any. It may happen that the likelihood equations have no solution for small samples. In Chapter 4, the domain of the parameter space (2.41) is extended so that the likelihood equations have a solution which is an interior point of the extended parameter space. The definition for the extended parameter space is given in section 4.1.

For the one way random effects model (2.1), $\sigma_\alpha^2$ is the only component of $\theta = (\sigma_\alpha^2, \sigma^2, \mu)$ that can be negative and when it is negative it is not in the parameter space. Thus, the maximum likelihood solutions for the likelihood equations are the maximum likelihood estimates if and only if $\sigma_\alpha^2$ is nonnegative. For this it suffices to show that if $\sigma_\alpha^2 < 0$ then $\sigma_\alpha^2 = 0$. To do this, we argue by contradiction that if

30
$\sigma^2 < 0$ then $\sigma^2 = 0$. Therefore, we assume that $\sigma^2 = 0$ and $\sigma^2 > 0$. This leads to two cases $\sigma^2 < \sigma^2$ and $\sigma^2 \geq \sigma^2$. For $\sigma^2 < \sigma^2$, the partial derivative of the likelihood function (2.24) with respect to $\sigma^2$ at $\sigma^2 = \sigma^2$ and $\mu = \hat{\mu}$ has the form

$$
\frac{\partial \ell}{\partial \sigma^2} = \frac{1}{2} \sum_{i=1}^{k} \frac{1}{n_i \sigma_i^2} + \frac{1}{2} \sum_{i=1}^{k} \frac{1}{n_i \sigma_i^4} (\bar{Y}_i - \hat{\mu})^2 - \frac{N}{2\sigma^2} + \frac{S^2}{2\sigma^4}
$$

$$
= -\frac{N}{\sigma^2} \{\hat{\sigma}^2 - \sigma^2\} < 0.
$$

Clearly, the likelihood function (2.24) can be increased by increasing $\sigma^2$ from $\sigma^2$. But this is a contradiction to our assumption that $\sigma^2$ is the maximum likelihood estimate. Similarly for the second case the partial derivative of the likelihood function with respect to $\sigma_i^2$ at $\sigma_i^2 = \hat{\sigma}_i^2$ and $\mu = \hat{\mu}$

$$
\frac{\partial \ell}{\partial \sigma_i^2} |_{(\hat{\sigma}_i^2, \hat{\mu})} = -\frac{1}{2\sigma_i^2} + \frac{1}{2\hat{\sigma}_i^4} (\bar{Y}_i - \hat{\mu})^2
$$

$$
= -\frac{1}{2\sigma^4} \{\hat{\sigma}_i^2 - \sigma_i^2\} < 0,
$$

shows that the likelihood function (2.24) can be increased by decreasing $\sigma_i^2$ from $\hat{\sigma}_i^2$. Again this leads to a contradiction to $\sigma_i^2$ being the maximum likelihood estimate. Therefore $\sigma_i^2$ must be zero.

For the case $\sigma^2 = 0$, the maximum likelihood estimates are given by $\sigma^2 = \text{SST}/\sum_{i=1}^{k} n_i$, and $\mu = \bar{Y}_i$, (Searle 1992).

Furthermore, we find the limiting value of $\ln \ell(\theta)$ as each of the two variance components of the model (2.1) approach either zero or infinity. This function has the form:

$$
L(\theta) = \exp(\ell(\theta)) = \left( \prod_{i=1}^{k} \frac{1}{\sigma_i^2} \right)^{N/2} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{k} \frac{1}{\sigma_i^2} (\bar{Y}_i - \mu)^2 - \frac{S^2}{2\sigma^2} \right\}
$$

Clearly,

(i) for fixed $\sigma_a^2$, $L(\theta) \to 0$ as $\sigma^2 \to 0$ or $\infty$.

(ii) for fixed $\sigma^2$, $L(\theta) \to 0$ as $\sigma_a^2 \to \infty$, and $L(\theta) < \infty$ as $\sigma_a^2 \to 0$.  

31
Similarly, finding the limiting value of the mle(2.25) of the mean is straightforward:

(i) for fixed $\sigma^2$, $\hat{\mu} \to \frac{1}{k} \sum_{i=1}^{k} \bar{y}_i$ as $\sigma^2 \to 0$, and $\hat{\mu} \to \bar{y}_i$ as $\sigma^2 \to \infty$.
(ii) for fixed $\sigma^2$, $\hat{\mu} \to \bar{y}_i$ as $\sigma^2 \to 0$, and $\hat{\mu} \to \frac{1}{k} \sum_{i=1}^{k} \bar{y}_i$ as $\sigma^2 \to \infty$.

2.6 Asymptotic Properties of the MLEs

For the general mixed model model, Hartley and Rao (1967) proved under certain restrictions that the maximum likelihood estimates are consistent and asymptotically normal as the size of the experimental design increases. Miller(1977) considered asymptotic properties of the maximum likelihood estimates for a larger class of design sequences whose size increases to infinity. Chatterjee and Das (1983) put together certain conditions to guarantee the asymptotic normality of the MLEs for the one way random effects model(2.1). Westfall (1987) also considered conditions with this property. Chatterjee and Das (1983) used the following conditions to prove that conditions of Theorem 3.1 due to Miller(1977) are met, which is enough to guarantee the asymptotic normality of the MLEs.

Assumption 2.1 For $\theta \in \Theta$, let $\{n_i\}$ be a sequence such that,

a) $\lim_{k \to \infty} \frac{1}{k} \sum_{i=1}^{k} n_i = \ell_0$ exists, $\ell_0 > 1$;

b) $\lim_{k \to \infty} \frac{1}{k} \sum_{i=1}^{k} \frac{n_i}{(\sigma^2 + n_i \sigma^2_s)} = \ell_1$ exists;

c) $\lim_{k \to \infty} \frac{1}{k} \sum_{i=1}^{k} \frac{n_i}{(\sigma^2 + n_i \sigma^2_s)} = \ell_2$ exists for $s = 0, 1, 2$.

To ensure the consistency and the asymptotic normality of the MLEs of the model(2.1), it is important to assume that the $n_i$ pattern for the subsample sizes has the form $n_1, \ldots, n_k, n_1, \ldots, n_k, \ldots$, where the number of replications of $\{n_1, \ldots, n_k\}$ tends to
infinity. Examples on asymptotic sequences of the model (2.1) will be presented in Chapter 4.
Chapter 3

Asymptotic Theory

For unbalanced data, it is often difficult to determine the exact distribution for a statistic. In such cases we focus on some recent higher order asymptotic approximations to provide accurate approximations for distribution function, density function, and integrals. Kass and Kadane (1986, 1989) applied Laplace’s method to evaluate integrals, in which the integrand has the form \( \exp\{-nh(\theta) + \ln f(\theta)\} \) and the function \( h(\theta) \) can be approximated by a quadratic. These authors obtained approximations to the posterior means and variances of positive functions of a real or vector-valued parameter \( \theta \), and to the marginal posterior densities of arbitrary parameters. On the other hand Fraser and Reid (1996) proposed recent third order procedures to determine a location model structure relative to a scalar interest parameter and then obtained a posterior for the interest parameter based on a flat or diffuse prior for only the interest parameter. For further references see, Barndorff-Nielsen and Cox (1989), Bleistein and Handelsman (1975), Efron (1993), Erdelyi (1956), Reid (1996), Tibshirani (1989), and Walker (1969).

A second method for deriving the density or the distribution function for a statistic which is asymptotically normal is the Edgeworth expansion. The Edgeworth expa-
sion is an asymptotic series that uses the cumulant generating function to provide an approximation to the density function \( f(y) \), where \( y \) is a sum or average of the values from the sample; see for example Barndorff-Nielsen and Cox (1979, 1989), Feller (1970), Fraser and Reid (1995), and Reid (1996b).

A different type of asymptotic expansion for the density of the sample mean is obtained by the saddlepoint method. The saddlepoint method approximates contour integrals in the complex plane by deforming those contours within the region of convergence of the moment generating function \( M(\theta) = E \exp(\theta^T X) \) in such away that most contributions to the values of the integral comes from a small neighborhood of the saddlepoint. Over this small neighborhood a Taylor expansion of the integrand and a term wise integration yields the saddlepoint approximation for the density function of \( \bar{x} \) (Daniels, 1954).

\[
\tilde{f}_n(\bar{x}) \sim \left( \frac{n}{2\pi K''(T_0)} \right)^{1/2} \exp\left(n(K(T_0 - T_0) - \frac{a_0}{n} + \frac{a_1}{n^2} + \ldots \right). \quad (3.1)
\]

The leading term in (3.1) is called the saddlepoint approximation. The density approximation (3.1) is generally more accurate than the corresponding normal approximation which gives an approximation to the exact density function of \( \bar{x} \) with relative error \( O(n^{-1/2}) \). The saddlepoint method was introduced to statistics by Daniels (1954); Barndorff-Nielsen and Cox (1979) extended the applications; the key is an approximate conversion of a cumulant generating function to a corresponding density function. In statistics this can be interpreted as a conversion from a likelihood function \( \ell(\theta) = a + \ln f(y; \theta) \) to a density function in the context of an exponential model. This approximation for the densities of \( y \) and \( \hat{\theta} \) are:

\[
\frac{c}{(2\pi)^{p/2}} \exp \left( \ell(\hat{\theta}; y) - \ell(\hat{\theta}; y) \right) |y|^{1/2} dy, \quad (3.2)
\]

\[
\frac{c}{(2\pi)^{p/2}} \exp \left( \ell(\hat{\theta}; y) - \ell(\hat{\theta}; y) \right) |\hat{\theta}|^{-1/2} d\hat{\theta}, \quad (3.3)
\]

where \( c \) is a constant to order \( O(n^{-3/2}) \) and \( c = 1 \) to order \( O(n^{-1}) \). The relative error
is of order $O(n^{-3/2})$.


This chapter discusses various techniques available to construct third order levels of significance for an interest parameter of the unbalanced one-way random effects model (2.1).

3.1 Dimension Reduction

In this section we review some background results on dimension reduction by means of sufficiency and ancillarity. Sufficiency and ancillarity reduction are very useful in constructing conditional and marginal distributions from an original model to achieve inference about parameters of interest. Exact conditional and marginal distributions are available for the exponential family models and the transformation family models. These exact and approximate procedures are very useful for deriving highly accurate approximations for distribution function, density function, and cumulant generating functions of a statistic. For testing a scalar interest parameter in the presence of nuisance parameters, Fraser and Reid (1995) used an ancillary statistic to reduce the dimension of an initial variable to that of the full parameter and obtained a further reduction to a scalar quantity by marginalization. For simplicity let $y$ denote a random vector of dimension $n$ from a statistical model $f(y; \theta)$ with a parameter $\theta' = (\lambda', \psi')$ of dimension $p$ where $\psi$ is a scalar interest parameter and $\lambda$ is a nuisance
parameter of dimension $p - 1$.

### 3.1.1 Sufficiency

Fisher (1922) introduced sufficiency as a method to reduce the dimension of a statistical model, sometimes to the same dimension as the parameter. It was further developed by Cox and Hinkley (1974), Fraser (1979), Kalbfleish (1975), Huzurbazar (1976) and among others. Sufficiency is defined as follows.

**Definition 3.1** Let $X$ be a random vector whose distribution depends on the parameter $\theta$. A statistic $T(x)$ is said to be sufficient for $\theta$ if the conditional distribution of $X$ given $T(x) = t$ does not depend on $\theta$.

As an example let $y_1, y_2$ denote a random sample from a normal distribution with mean $\theta$ and variance unity. Let $s = y_1 + y_2$ and $t = y_1 - y_2$ be a $1 - 1$ transformation from $(y_1, y_2)$ to $(s, t)$. The conditional distribution of $t$ given $s$ is free of $\theta$;

$$f_{T|s}(t|s) = \frac{f_{S,T}(s, t; \theta)}{f_S(s; \theta)} = f_T(t)$$

For this $s = y_1 + y_2$ is sufficient and the dimension of the variable is reduced from 2 to 1 by the marginalization in going from $y$ to $s$. Thus the model for the original variable $y$ is replaced by the marginal model for the new variable $s$ and inference about $\theta$ can be obtained from the marginal density of $s$.

To extend the preceding example to a statistical model $f(y; \theta)$; suppose that there exists a one to one transformation from $y$ to $(s, t)$ with Jacobian transformation $J$ such that

$$f(y; \theta) = f(s; \theta)f(t|s)|J|,$$  \hspace{1cm} (3.4)

then $s$ is sufficient for $\theta$ and the marginal density of $s$ in (3.4) is an appropriate basis for inference about $\theta$. In many situations sufficiency reduction of the type (3.4) is
useful in reducing the dimension of the model for the initial variable $y$ to that of the model for the sufficient statistic $s$ of dimension $p$. This sort of reduction is useful with exponential family models such as

$$ f(y; \theta) = \exp(\eta'(\theta)s(y) - k(\theta))h(y), \quad (3.5) $$

with a $p$-dimensional canonical parameter $\eta(\theta)$ and a sufficient statistic $s$ of dimension $p$.

Sometimes sufficiency reduction can be used for inference about an interest parameter $\psi$. If $\psi$ is an interest canonical parameter in (3.5) then the corresponding exponential model is sometimes expressed as

$$ f(y; \theta) = \exp(\psi s_1 + \lambda's_2 - k(\theta))h(y). \quad (3.6) $$

In (3.6) the distribution of $s_1|s_2$ depends only on the interest parameter $\psi$ and accordingly the conditional distribution of $s_1$ given $s_2$ is an appropriate distribution for inference concerning $\psi$ free of the nuisance parameter $\lambda$. A third order approximation (Fraser and Reid, 1993) to the likelihood function from this conditional distribution is given by

$$ \ell(\psi) = \ell(\hat{\theta}_\psi) + \frac{1}{2} \ln |\phi_{\lambda\lambda}(\hat{\theta}_\psi)|, \quad (3.7) $$

where $\hat{\theta}_\psi = (\psi, \hat{\lambda})$ is the constrained maximum likelihood estimate of $\theta$ and $\phi_{\lambda\lambda}$ is the observed information concerning the nuisance parameter $\lambda$ for given $\psi$.

For models with nuisance parameters various generalizations of (3.4) are available (Reid 1994);

$$ f(s; \theta) = f(s_1|s_2; \lambda)f(s_2; \psi). \quad (3.8) $$

In (3.8) $s_2$ is sufficient for $\psi$ in a general sense and the marginal density for $s_2$ can then be used for inference concerning the interest parameter $\psi$. In addition, the nuisance
parameter $\lambda$ can be eliminated by marginalization. On the other hand if interest lies in obtaining inference about the interest parameter $\lambda$, the conditional density of $s_1$ given $s_2$ would be an appropriate choice. For this case, conditioning would be the proper choice to eliminate the nuisance parameter $\psi$.

Sometimes models with nuisance parameters admit a factorization of this form;

$$f(s; \theta) = f_1(s_1; \psi, \lambda)f_2(s_2|s_1; \psi),$$

where $s_1$ is sufficient for $\lambda$ in the sense that the conditional density of $s_2$ given $s_1$ is free of $\lambda$. Inference concerning $\psi$ is based on the conditional density $f_2$. Furthermore, $f_1$ or $f$ can be used for testing $\lambda$ if plausible values of $\psi$ are available. To illustrate this point we present the following example.

**Example 3.1** For the one-way variance components model (2.1) we have

$$f(y; \theta) = f_2(s^2; \sigma^2)f_1(\bar{y}_1, \ldots, \bar{y}_k; \mu, \sigma^2_0, \sigma^2),$$

where $\text{var}(\bar{y}_i) = \sigma^2 = \sigma^2_0 + \sigma^2/n_i$, $f_2$ is the $\sigma^2 \chi^2(\sum_{i=1}^{k} n_i - k)$ density, and $\bar{y}_i \sim N(\mu, \sigma^2_i)$. Since $\bar{y}_i$ and $s^2$ are independent, inference about $\sigma^2$ is based on the conditional density $f_2$ of $s^2$ given $\bar{y}$. For $\sigma^2_0 > 0$, Wald (1940) introduced a test statistic $F_w$ (see 2.16) that can be used for inference concerning $\sigma^2_0$, and $\sigma^2_0$. If $\sigma^2_0 = 0$, $F_w$ reduces to the usual $F$ statistic and inference concerning $\sigma^2_0$ can be constructed from the $F$ statistic. Spjotvoll (1967) examined testing the hypothesis $\psi \leq \psi_0$ vs $\psi > \psi_0$ where $\psi = \sigma^2_0/\sigma^2$, and found that the most powerful invariant test against an alternative $\psi_1$ depend on $\psi_1$. Also, it maximizes the minimum power over the set of alternatives with $\psi \geq \psi_1$.

A different factorization when nuisance parameters are present is

$$f(s; \theta) = f_1(s_2; \psi, \lambda)f_2(s_1|s_2; \psi).$$

39
The conditional density of $s_1$ given $s_2$ is used for constructing inference concerning the interest parameter $\psi$. Inference about $\lambda$ is obtained from $f_1$ if needed values of $\psi$ can be constructed from the conditional density of $s_1$ given $s_2$.

**Example 3.2** Under the Restricted Maximum Likelihood method, the log-likelihood function (2.20) for the general mixed model (2.19) can be decomposed as

$$
\ell(\mu, \gamma, \sigma^2) = \ell_1(\sigma^2, \gamma)\ell_2(\sigma^2, \gamma, \mu)
$$

(3.12)

where the log-likelihood function $\ell_1$ is free of fixed effects $\mu$. Inference concerning $\sigma^2$ or $\gamma$ can be constructed from the conditional log-likelihood function $\ell_1$. The log-likelihood $\ell_2$ can be used for inference concerning $\mu$ if required values of $\sigma^2$ and $\gamma$ can be constructed from $\ell_1$.

For further consideration of various definitions of sufficiency in the presence of nuisance parameters see (Fraser, 1956; Cox and Hinkley, 1974; Basu, 1977; Barndorff-Nielsen 1978; Reid, 1994; Lindsey, 1995).

### 3.1.2 Ancillarity

Now we proceed in some opposite direction and consider statistics whose distribution is free of the parameter, ancillary statistics.

**Definition 3.1** A statistic $T(x)$ whose distribution does not depend on the parameter $\theta$ is called an ancillary statistic.

A general method for constructing an ancillary statistic does not exist and if a method exists, an ancillary statistic may not be unique. Ancillary statistics arise naturally in the transformation family model. For the transformation model exact ancillary statistics exist and are sometimes referred to as the configuration of the sample. The
transformation models are useful for measurement and production experiments. The location-scale family and the linear regression models are examples of transformation models. They also exist for experiments where the ancillary statistic is the sample size. For references see Barndorff-Nielsen (1980); Fisher (1934); Fraser (1979); McCullagh (1987); Reid (1994).

Example 3.3 Let \((X_1, \ldots, X_n)\) be a random sample from a statistical model \(\sigma^{-1} f((y-\mu)/\sigma), \theta = (\mu, \sigma), T_1 = \bar{x} \text{ and } T_2 = \sum_{i=1}^{n} (x_i - \bar{x})^2/(n - 1)\) then \((T_1, T_2)\) is a sufficient statistic given the configuration statistic (Fisher, 1934)

\[
T_3 = \left( \frac{x_1 - \hat{\mu}}{\hat{\sigma}}, \ldots, \frac{x_n - \hat{\mu}}{\hat{\sigma}} \right)
\]

which is in turn ancillary; the ancillary defines the orbit on which the observations reside. Position on an orbit is determined by the mle \((\hat{\mu}, \hat{\sigma})\). Accordingly any sample point can be represented by the pair, the mle and the configuration.

For detailed discussion see Fraser (1968).

Fisher (1934) introduced inference conditional on ancillaries as a method to reduce the size of the sample space and yet retain all the relevant information in the original sample; he claimed this leads to more relevant inference about \(\theta\). For unbalanced one-way variance components models the dimension of the initial variable is greater than that of the parameter \(\theta\) and a reduction method is needed to obtain the third order significance function of a real interest parameter \(\psi\). An initial reduction can be made by sufficiency. We then apply ancillary methods; the conditional model given the ancillary statistic is used for the asymptotic methods. A formalization of the proceeding statement is given as

\[
f(y; \theta) = f(s|t; \theta)g(t).
\]

In (3.13) \(t\) is said to be ancillary for \(\theta\) and the conditional density of \(s\) given \(t\) is used for inference concerning \(\theta\). Throughout this work we assumed if the model \(f(y; \theta)\)
satisfies (3.4) or (3.13) then the dimension of the initial problem is reduced to that of \( s \). For example, the dimensions of \( s \) and \( \theta \) are the same in both the transformation family and the exponential family models. A simple extension of (3.13) when nuisance parameters are present is given (3.8). In the location family models the likelihood function \( \ell(\psi) \) from this conditional distribution can be approximated to third order by

\[
\ell(\psi) = \ell(\hat{\theta}_\psi) - \frac{1}{2} \ln |J_{\lambda\lambda}(\hat{\theta}_\psi)|. \tag{3.14}
\]

Fraser and Reid (1993) showed that the conditional likelihood \( \ell(\psi) \) with the following reparameterization \( \phi(\psi) = -S(\psi) \), where \( S(\psi) \) is the score function calculated from the likelihood function \( \ell(\psi) \), are the ingredients for the construction of third order significance function \( p(\psi) \).

For a continuous statistical model \( f(y; \theta) \) with asymptotic properties, the current methods of constructing ancillaries for third order analysis include the tangent location model of Fraser (1964) and the cumulant method of McCullagh (1987). The cumulant method uses the cumulant of the log-likelihood derivative to obtain third order ancillaries (McCullagh, 1987) and the tangent location method constructs tangent directions for first derivative ancillaries at the data point \( y^0 \). The analysis in Fraser and Reid (1995) shows that first order derivative ancillaries can be adjusted to give second order ancillaries and no further upgrading is required for third order inference: these second order ancillaries can however be upgraded to third order ancillaries (Skovgaard, 1986).

If more than one ancillary statistic is available then a principle is required to select one of them.

**Definition 3.2** A statistic \( T(x) \) is **maximal ancillary** if every other ancillary statistic is a function of \( T(x) \).
Fraser (1979) and Schervish (1995) discussed examples when no maximal ancillary statistic exists.

For the one-way random effects model (2.1), the construction of an approximate ancillary for third order inference is based on the tangent location model method. For this we define a pivotal quantity with \( n \) components on the mle surface \( \hat{\theta} = c \) such that each component of the variable \( y \) has a location relation with \( \theta \) satisfying

\[
0 = \frac{\partial F_i(y;\theta)}{\partial \theta} d\theta + \frac{\partial F_i(y;\theta)}{\partial y_i} dy_i|(y^0, \hat{\theta}).
\]  

(3.15)

If \( p \geq 1 \) then the ancillary directions are arrays of \( p \) vectors; each vector has \( n \) components. The \((i,j)\) element of the resulting matrix \( n \times p \) is given by

\[
v_{ij} = -\left(\frac{\partial F_i(y;\theta_j)}{\partial y_i}\right)^{-1}\left(\frac{\partial F_i(y;\theta_j)}{\partial \theta_j}\right)
\]  

(3.16)

where \( i = 1, \ldots, n \) and \( j = 1, \ldots, p \).

In conclusion if an exact or an approximate ancillary exists then inference should be drawn conditionally on an ancillary statistic because it is viewed as making better use of the information available in the actual sample.

### 3.2 Conditional Inference

As pointed out earlier, inference concerning the interest parameter \( \psi \) in the presence of the nuisance parameter \( \lambda \) can be constructed from the conditional likelihood function \( \ell(\psi) \) at the observed value \( y^0 \). For this, consider the exponential model (3.6) in which the sufficient statistic \( s \) is decomposed into two components, one component, (say \( s_1 \)) of dimension one and the second component, (say \( s_2 \)) of dimension \( p - 1 \). The saddlepoint approximation to the density function of \( s \) at \((s_1, s_2)\) has the form

\[
f(s, \theta) = (2\pi)^{-p/2}\exp\{\ell(\theta) - \ell(\hat{\theta})\}|J(\hat{\theta})|^{-1/2}
\]  

(3.17)
and the same approximation to the marginal density function of $s_2$ has the form

$$f(s_2, \theta) = (2\pi)^{-\nu/2} \exp\{\ell(\theta) - \ell(\hat{\theta}_\psi)\}|J_{\lambda\lambda}(\hat{\theta}_\psi)|^{-1/2}.$$  \hspace{1cm} (3.18)

This leads to the conditional density function of $s$ given $s_2$

$$f(s|s_2, \psi) = c(2\pi)^{-1/2} \exp\{\ell(\hat{\theta}_\psi) - \ell(\hat{\theta})\}\left|\frac{|J_{\lambda\lambda}(\hat{\theta}_\psi)|}{|J(\hat{\theta})|}\right|^{-1/2}$$  \hspace{1cm} (3.19)

where $c$ is a norming constant. Skovgaard (1987) called this expression for the conditional density, the double saddlepoint approximation. Clearly, the corresponding likelihood function is given by (3.7). The distribution function can then be approximated by inverting the likelihood given by (3.7). Expressions for the conditional density and the corresponding likelihood can be obtained when the interest parameter $\psi$ is a $k$ dimensional vector and $\lambda$ is a $p-k$ dimensional nuisance parameter.

For the transformation case, consider a sample from the location model

$$f(y - \theta) = f(y_1 - \psi, y_2 - \lambda),$$  \hspace{1cm} (3.20)

where $y_2$ and $\lambda$ can be vector valued. This model permits a factorization of the form

$$f(s; \theta) = f(s_1, \psi)f(s_2|s_1; \psi, \lambda).$$  \hspace{1cm} (3.21)

If the conditional density contains no information regarding $\psi$, the marginal density can be used for inference concerning $\psi$. Now apply the Laplace method to the integral

$$\int \exp(\ell(\theta)) ds_2$$

to get an approximation to the marginal density of $\psi$

$$f(s_1, \psi) = (2\pi)^{-1/2} \exp\{\ell(\hat{\theta}_\psi)\}|J_{\lambda\lambda}(\hat{\theta}_\psi)|^{-1/2}.$$  \hspace{1cm} (3.22)

The corresponding likelihood function is given by (3.14). Third order tail probabilities can then be calculated by numerical integration. The expression for the likelihood function (3.14) is given in DiCiccio, Field and Fraser (1990).
3.3 First order approximations

Now we are in a position to discuss various ways of applying the Central Limit Theorem and the Law of Large Numbers for the purpose of obtaining asymptotic statements from the likelihood function when $n$ or some information measure is large.

Both likelihood and significance are functions of the parameter $\theta$ and provide an assessment of the data. The likelihood function measures the probability up to a multiplicative constant at the data point and the significance function measures the probability left (or right) of the data point. Three common asymptotic methods of extracting first order approximations are based on the the Law of Large Numbers and the Central Limit Theorem. The application of these methods uses the following three statistics. The standardized score

$$ s = S(\theta)j^{-1/2}, $$

the standardized maximum likelihood estimate

$$ q = (\hat{\theta} - \theta)j^{1/2} $$

and the signed likelihood ratio statistic

$$ r = \text{sgn}(\hat{\theta} - \theta)[2(\ell(\hat{\theta}) - \ell(\theta))]^{1/2} $$

where

$$ j = -\frac{\partial^2 \ln f(y; \theta)}{\partial \theta^2} \bigg|_{\theta = \hat{\theta}} $$

is the observed information and

$$ S(\theta) = \frac{\partial}{\partial \theta} \ln f(y; \theta) $$

is the score function.
Fraser (1991) examined the accuracy of $r$, $s$ and $q$ using four different models; normal, extreme value, gamma(3) and the Cauchy distribution and his results show that quite generally $r$ is superior to both $q$ and $s$. In addition, $r$ is invariant under a one to one parametric transformations. The three standardized variables lead to three approximations for the significance function: $\Phi(s)$, $\Phi(q)$ and $\Phi(r)$. For the examples, $\theta$ is considered to be a canonical parameter of an exponential model:

$$f(y; \theta) = \exp(t(y)\theta - c(\theta) + h(\theta)),$$

where $c(\theta)$ is the cumulant generating function of $f(y; 0) = \exp(h(y))$ which is a density.

In the present context the two quantities, the signed likelihood root and the maximum likelihood departure are of importance as third or lower order approximations for testing a scalar interest parameter in the unbalanced one-way variance components models.

It is a general problem to test a real interest parameter $\psi$ in the presence of a vector nuisance parameter $\lambda$. In this section we develop significance functions for testing a real parameter $\psi(\theta)$ where $\psi(\theta)$ is the canonical interest parameter; e.g., as in (3.6). Cases where the interest parameter cannot be extracted in this way will be discussed in the next section. Let

$$H_0 : \psi = \psi^0 \text{ vs } H_a : \psi \neq \psi^0.$$  

The maximum likelihood estimate for $\theta$ has an approximate $AN_p(\theta, J^\theta(\hat{\theta})$ distribution where $J^\theta$ designates the asymptotic variance of $\hat{\theta}$ and is given by

$$
\begin{pmatrix}
J^{\lambda\lambda} & J^{\lambda\psi} \\
J^{\psi\lambda} & J^{\psi\psi}
\end{pmatrix}.
$$

Using simple multivariate normal theory we find that the asymptotic distribution of
\( \hat{\psi} \) is normal with mean \( \psi \) and estimated variance \( (j^{\psi\psi}(\hat{\theta})) \) where
\[
(j^{\psi\psi}(\hat{\theta}))^{-1} = J_{\psi\psi} - J_{\psi\lambda}J_{\lambda\lambda}^{-1}J_{\lambda\psi}|\hat{\theta}_o
\]
Under the null hypothesis, the maximum likelihood departure has a well known standardized pivotal form:
\[
q = (\hat{\psi} - \psi)(j^{\psi\psi})^{-1/2}
\]  
(3.23)
Using (3.23) and the following factorization
\[
|J_{\theta\theta}| = |J_{\lambda\lambda}|(j^{\psi\psi})^{-1},
\]
(3.24)
the statistic \( q \) can be expressed as
\[
q = (\hat{\psi} - \psi)(\frac{|J_{\theta\theta}(\hat{\theta})|}{|J_{\lambda\lambda}(\hat{\theta})|})^{1/2}
\]
(3.25)
Another important statistic for the construction of third order approximations is the signed likelihood ratio statistic \( r \) defined as
\[
r = \text{sgn}(\hat{\psi} - \psi)\sqrt{D}
\]
where
\[
D = 2(\ell(\hat{\theta}) - \ell(\hat{\theta}_0))
\]
\( D \) is called the deviance and \( r \) is called the directed deviance by Lindsey(1996).

The construction of the above two standardized variables leads to two first order approximations for the significance function:
\[
p_1(\theta) = \Phi(r), \quad p_2(\theta) = \Phi(q).
\]
(3.26)
In (3.26) we have \( p_1(\theta) = p(\theta) + O(n^{-1/2}) \), with \( p(\theta) \) represents the exact probability for each case; the bound as given applies for a bound range of the standardized variable.
Third order approximations for significance functions can be expressed in the Lugannani and Rice (1980) type formula
\[ \Phi(r) + \phi(r) \left( \frac{1}{r} - \frac{1}{q} \right) \] (3.27)
and the \( r^* \) type formula
\[ \Phi \left( r + \frac{1}{r} \ln \frac{q}{r} \right). \] (3.28)

### 3.4 Third order Approximations

The one way random effects model is a member of the \((k + 1,3)\) curved exponential family because it has a \((k + 1)\) dimensional minimal sufficient statistic and a three dimensional vector parameter (Efron, 1975). The minimal sufficient statistic is
\[ \bar{y}_i = \mu + \sigma_i z_i \quad i = 1, \ldots, k, \]
\[ S = \sigma \chi, \]
where \( \chi \) with \((\sum_{i=1}^{k} n_i - k)\) degrees of freedom is the chi random variable and \( z_i \sim N(0, 1) \) is the standardized normal random variable. Clearly, the dimension of the minimal sufficient statistic \( W = (\bar{y}_1, \ldots, \bar{y}_k, S)' \) is equal to that of the vector parameter \( \theta \) whenever \( k = 2 \). For this case only, the minimal sufficient statistic is complete (Lehmann, 1986, section 4.3).

In the present context the issue of an ancillary arises if the dimension of \( W \) exceeds that of the parameter vector \( \theta \). For this we differentiate \( W \) with respect to \( \theta \) for fixed pivots \( z_i \), and \( \chi \) to obtain the tangent directions
\[ V = \frac{\partial W}{\partial \theta} \bigg|_{(\varphi^0, \varphi^0)}. \]
These direction vectors are needed for the construction of the third order approximate ancillaries. The construction of the third order tangent exponential model at the data
point $y^0$ requires only the likelihood function $\ell^0(\theta)$ and the gradient of the likelihood function at the data point $y^0$ 

\[
\varphi(\theta) = \frac{\partial \ell(\theta, y)}{\partial V} |_{y^0}
\]

taken in the ancillary directions $V$. Accordingly, we get the canonical vector parameter of the exponential model $\varphi(\theta) = (\varphi_1(\theta), \varphi_2(\theta), \varphi_3(\theta))$ where

\[
\begin{align*}
\varphi_1(\theta) &= -\sum_{i=1}^{k} \left( \frac{\tilde{y}_i^0 - \mu}{\sigma_i^2} \right), \\
\varphi_2(\theta) &= -\sum_{i=1}^{k} \left( \frac{\tilde{y}_i^0 - \mu}{\sigma_i^2} \right) \left( \frac{\tilde{y}_i^0 - \mu}{\sigma_i^2} \right) \frac{\partial \alpha}{\partial \alpha_i}, \\
\varphi_3(\theta) &= -\sum_{i=1}^{k} \left( \frac{\tilde{y}_i^0 - \mu}{\sigma_i^2} \right) \left( \frac{\tilde{y}_i^0 - \mu}{\sigma_i^2} \right) \frac{\partial \alpha}{\partial \alpha_i} - \frac{S S^0}{\sigma^2 \sigma^0}.
\end{align*}
\]

Now we are in a position to express the observed information determinants for full and constrained models recalibrated with respect to the $\varphi$ scale.

\[
\begin{align*}
|J(\theta\theta)(\hat{\theta})| &= |J_{\theta\theta}(\hat{\theta})| |J(\hat{\theta})|^{-2}, \\
|J_{(\lambda\lambda)}(\hat{\theta}_\psi)| &= |J_{\lambda\lambda}(\hat{\theta}_\psi)| |J(\hat{\theta}_\psi) J_{\lambda}(\hat{\theta}_\psi)|^{-1}
\end{align*}
\]

where $J(\theta) = (J_{\lambda}(\theta), J_{\psi}(\theta))$ is the Jacobian for parameter change, $\lambda$ is a nuisance parameter and $\hat{\theta}_\psi = (\hat{\lambda}_\psi, \psi)$ is the constrained maximum likelihood estimate with parameter of interest $\psi$ fixed.

The construction of the tangent exponential model requires the maximum likelihood estimates to be on the $\varphi$ scale. This requires a scalar linear function of $\varphi(\theta)$ say,

\[
\bar{\varphi}(\theta) = \frac{k_{\psi}(\hat{\theta}_\psi)}{|k_{\psi}(\hat{\theta}_\psi)|} \varphi(\theta)
\]

that agrees with the interest parameter $\psi$ at $\hat{\theta}_\psi$. In (3.29) if $\psi$ is the $i^{th}$ component of $\theta$ then $\frac{k_{\psi}}{|k_{\psi}|}$ is the unit vector obtained from the $i^{th}$ row in $J^{-1}(\theta)$. After further
calculations we get an important ingredient in the Lugannani and Rice formula and the $R^*$ formula; the standardized maximum likelihood departure

$$ Q = \langle \varphi(\hat{\theta}) - \varphi(\hat{\theta}_\psi) \rangle \left( \frac{|J(\theta)(\hat{\theta})|}{|J(\lambda)(\hat{\theta}_\psi)|} \right)^{1/2} \quad (3.30) $$

Another important ingredient in the analysis is the signed likelihood ratio,

$$ R = \text{sgn}(\hat{\psi} - \psi) [2\{\ell(\hat{\theta}; y) - \ell(\hat{\theta}_\psi; y)\}]^{1/2} \quad (3.31) $$

The significance function for testing an interest parameter $\psi$ is then given by the Lugannani and Rice formula (1.3) and the $R^*$ formula (1.5).

If the dimension of $W$ is equal to that of the vector parameter $\theta$, the issue of an ancillary does not arise. For this, we differentiate the likelihood function (2.24) with respect to $W$ and obtain the canonical parameters of the exponential model

$$ \varphi_1(\theta) = -\frac{1}{\sigma_1^2} (\bar{y}_{1.}^2 - \mu), $$

$$ \varphi_2(\theta) = -\frac{1}{\sigma_2^2} (\bar{y}_{2.}^2 - \mu), \text{ and} $$

$$ \varphi_3(\theta) = -\frac{S}{\sigma^2}. $$

Then after further calculations we obtain the maximum likelihood departure $Q$ and the signed likelihood ratio $R$.

### 3.5 The Tangent Exponential Model

For testing a scalar interest parameter in the presence of a vector nuisance parameter in large sample asymptotic theory, the tangent exponential model provides third order accuracy when exact ancillaries are not available or the computation of the conditional likelihood function and the conditional information matrix are cumbersome. The tangent exponential model (Fraser and Reid, 1995) gives a method of
determining a third order ancillary statistic and a way of incorporating this ancillary statistic into the model. Fraser (1988, 1990) introduced the tangent exponential model for a statistical model where dimension of the variable is the same as that of the parameter, possibly after sufficiency or ancillary reduction.

For a vector parameter with scalar interest parameter, Fraser and Reid (1993) discussed approximations for exponential models in the canonical parameterization, in which nuisances are eliminated by conditioning; DiCiccio, Field and Fraser (1990) discussed approximations in which nuisances are eliminated by marginalization; and inference for component parameters in the general case, Barndorff-Nielsen (1991) introduced approximations when ancillary reductions are possible.

Consider a continuous statistical model \( f(y; \theta) \) where dimension of the variable \( y \) is the same to that of the parameter \( \theta \). Suppose that \( \hat{\theta} \) is unique, and is of \( O(n^{-1/2}) \) about the true parameter \( \theta \). In addition assume that \( \ell(\theta; y) = \ln f(y; \theta) \) with either argument fixed is \( O(n) \) and that \( \ell_{\theta}(\theta), \ell_{\theta \theta}(\theta), \ldots \) are \( O(n) \). For further discussion on these asymptotic assumptions see DiCiccio, Field and Fraser (1990) and Fraser and Reid (1993).

Cakmak, Fraser, McDunnough, Reid and Yaun (1995) developed approximations for the observed significance function \( p(\theta) = p(\hat{\theta} \leq \theta; \theta) \) by using an exponential approximation rather than the normal approximation. For this, they expanded the log likelihood function \( \ell(\theta; y) = \sum a_{ij}(\theta - \theta_0)^i(y - y_0)^j/i!j! \) about \( (\theta_0, y_0) \) in a Taylor series expansion and then obtained a \( 4 \times 4 \) matrix of the coefficients in the form
where omitted entries are \( O(n^{-3/2}) \). The model expressed in (3.32) is exponential to order \( O(n^{-1}) \) and is exponential to order \( O(n^{-3/2}) \) in a first derivative neighborhood of the data point except for the constant \( -\frac{12c}{4n} \). The tangent exponential model is an exponential model satisfying (3.32) with \( c = 0 \). Moreover, it is uniquely determined by the first two columns of (3.32) without the first row.

Cakmak, Fraser, and Reid (1994) examined the above expansion when \( d > 1 \). The model is exponential to \( O(n^{-1}) \) and to \( O(n^{-3/2}) \) differs from an exponential model by terms of order \( O(n^{-1}) \). Expansion (3.32) and its generalization to the multivariate case lead to what can be called an inverse \( p^* \) result, (Fraser and Reid 1995). The \( p^* \) formula gives the probability density at a given point and must be recalculated at each point in the neighborhood of the data point, the tangent exponential model gives the probability in a region around the data point. To construct an exponential model yielding the first columns of (3.32) without the first row, let \( V = (v_1, \ldots, v_d) \) be an array of vectors tangent to first order derivative ancillary for parameter changes in linearly independent directions at \( \hat{\theta} \) and obtain the sample space gradient of the likelihood function in these directions as \( \varphi' = \ell'_V(\theta; y^0) \) and the corresponding score function as \( s = \ell'_\varphi(\varphi^0; y) \) at the observed data point \( y^0 \). The corresponding tangent exponential model has the form

\[
\frac{c}{(2\pi)^{d/2}} \exp \left( \ell(\theta) - \ell(\hat{\theta}) + (\varphi - \varphi^0)s \right) |\hat{I}_{\varphi\varphi}|^{1/2} d\varphi, \tag{3.33}
\]

where \( \hat{I}_{\varphi\varphi} \) is the observed information matrix in the \( \varphi \) scale. When \( d = 1 \), the model
produces \( \hat{F}(\hat{\theta}^0; \theta) \) to accuracy \( O(n^{-3/2}) \). (Fraser and Reid, 1993).

For the general model \( f_n(y_1, \ldots, y_n; \theta_1, \ldots, \theta_d) \) third order ancillaries are not unique but the distribution for their projection on the observed maximum likelihood surface is unique to \( O(n^{-3/2}) \); for this Fraser and Reid (1995) used the total derivative of the score equation at \( \hat{\theta} \) to get the differential for \( \hat{\theta} \)

\[
d\hat{\theta} = \mathbf{J}^{-1}(\theta; y)d\mathbf{y}.
\] (3.34)

The element \( d\mathbf{y} \) on the surface \( \hat{\theta} = \text{constant} \) can be decomposed into its components say

\[
d\mathbf{y} = d\mathbf{y}_p d\mathbf{y}_c
\]

where \( d\mathbf{y}_c \) is an increment on the surface \( \hat{\theta} = \text{constant} \) and \( d\mathbf{y}_p \) is an increment orthogonal to the surface \( \hat{\theta} = \text{constant} \). Since the parallel component is zero on the surface \( \hat{\theta} = \text{constant} \) it follows that

\[
d\mathbf{y} = |\ell_{\theta,y}(\hat{\theta}, y)|^{-1} |\mathbf{J}| d\hat{\theta} d\mathbf{y}_c
\] (3.35)

where \( |\ell_{\theta,y}| \) is the nominal volume of the \( d \times n \) matrix \( \ell_{\theta,y}(\hat{\theta}; y) \). Thus the general statistical model can be expressed in the form:

\[
f(y; \theta)d\mathbf{y} = \frac{c}{(2\pi)^d/2} \exp \left( \ell(\theta; y) - \ell(\hat{\theta}; y) \right) |\mathbf{J}|^{1/2} d\hat{\theta} \cdot g(y; \theta)d\mathbf{y}_c
\] (3.36)

with

\[
g(y; \hat{\theta})d\mathbf{y}_c = \frac{(2\pi)^{d/2}}{c} \exp \left( \ell(\hat{\theta}; y) \right) |\ell_{\theta,y}(\hat{\theta}; y)|^{-1} |\mathbf{J}|^{1/2} d\mathbf{y}_c.
\]

It is shown in Fraser and Reid (1995) that the marginal density \( g(y; \hat{\theta}) \) can be transformed to \( du \) where \( u \) is \( U(0,1)^{n-p} \) to third order. In other words, \( u \) is ancillary to third order, the first factor of (3.36) is the conditional density of \( \hat{\theta} \) for given \( u \), and \( \ell(\theta, y) \) is the conditional likelihood to third order.
3.5.1 Conditional model for Interest Parameter

Consider the general model \( f(y; \theta) \) as outlined earlier and assume that \( \theta = (\lambda, \psi) \) where \( \psi \) is an interest parameter of dimension \( d_\psi \) and \( \lambda \) is the nuisance parameter of dimension \( d - d_\psi \). Let \( \varphi = (\varphi_1, \varphi_2) \) be a partition of the canonical parameter \( \varphi \) in the \( \varphi \)-coordinates; \( \hat{\varphi} = \varphi(\hat{\lambda}, \hat{\psi}) \) and \( \hat{\varphi}_\psi = (\hat{\lambda}_\psi, \hat{\psi}) \) denote the full and restricted maximum likelihood estimate in the \( \varphi \)-coordinate respectively. Let \( s' = (s'_1, s'_2) \).

As pointed out in Fraser and Reid (1995), the vectors \( V \) can be chosen so that the maximum likelihood surfaces \( \hat{\theta}_\psi = \hat{\theta}_\psi^0 \) and \( \hat{\theta}_{\psi 2} = \hat{\theta}_{\psi 2}^0 \) are the same. By the same method discussed earlier, the conditional distribution for the nuisance parameter can be expressed at the observed maximum likelihood point in the form:

\[
\frac{c'}{(2\pi)^{(d-d_\varphi)/2}} \exp \left( \ell(\varphi_1, \varphi_2) - \ell(\hat{\varphi}_{1\varphi 20}, \varphi_2) \right) |\hat{\jmath}(\varphi_1, \varphi_2)(\hat{\theta}_{\psi 20})|^{-1/2} ds_1, \tag{3.37}
\]

where the conditional score is just \( s_1 \). Then the marginal density for \( s_2 \) as projected to the maximum likelihood surface \( \hat{\theta}_\psi = \hat{\theta}_\psi^0, \hat{\varphi} = \hat{\varphi}_2 \) is obtained by dividing (3.33) by (3.37):

\[
\frac{c''}{(2\pi)^{d_\varphi/2}} \exp \left( \ell(\hat{\varphi}_{1\varphi 20}, \varphi_2) - \ell(\hat{\theta}_\psi^0) + (\varphi_2 - \hat{\varphi}_2)s_2 \right) |\hat{\jmath}_{\varphi \varphi}^{-1/2}| \hat{\jmath}(\varphi_1, \varphi_2)(\hat{\theta}_{\psi 20})|^{1/2} ds_2
\]

with the nuisance information calculated for \( \varphi_1 \) with \( \psi \) fixed at \( \psi_0 \) is given by

\[
|\hat{\jmath}(\varphi_1, \varphi_2)(\hat{\theta}_{\psi 20})| = |\hat{\jmath}_{\lambda \lambda}(\hat{\theta}_\psi^0 || \varphi_{\lambda'}(\hat{\theta}_\psi^0))|^{-2} \tag{3.38}
\]

and the full information determinant is given by

\[
|\hat{\jmath}_{\varphi \varphi}| = |\hat{\jmath}(\hat{\theta}_\psi^0 || \varphi_{\varphi}(\hat{\theta}_\psi^0))|^{-2}. \tag{3.39}
\]

Furthermore, the signed profile likelihood ratio has the form:

\[
R = \text{sgn}(\hat{\varphi}_2^0 - \varphi_2)[2 \left( \ell(\hat{\theta}_\psi^0) - \ell(\hat{\theta}_{1\varphi 20}, \varphi_2) \right)]^{1/2} \tag{3.40}
\]
and the standardized nuisance -adjusted departure can be expressed in the form:

\[ Q = M \frac{\sqrt{J_{\varphi\varphi}}}{\sqrt{J_{(\varphi_1\varphi_1)}}} \]  \hspace{1cm} (3.41)

with the maximum likelihood departure \( M \) in the \( \varphi \) coordinates given as

\[ M = \varphi_2^0 - \varphi_2 = \text{sgn}(\hat{\psi} - \psi) \left( \varphi^0 - \varphi(\hat{\theta}_\varphi^0) \right) s_\varphi / |s_\varphi| \]  \hspace{1cm} (3.42)

where \( s_\varphi = \varphi^\psi(\hat{\theta}_\varphi^0) \) gives in \( \varphi \) coordinates the vector orthogonal to the curve \( \psi \) at \( \hat{\theta}^0 \)

\[ d\varphi = (d\lambda, d\psi)\varphi^\psi(\hat{\theta}_\varphi^0), \quad d\psi = d\varphi\varphi^\psi(\hat{\theta}_\varphi^0). \]

For testing a vector interest parameter, the general method is discussed in Fraser and McKay(1975), Fraser and Reid(1995) and Barndorff-Nielsen(1986).
Chapter 4

Extended Unbalanced Linear Models with Random effects

As pointed out in the introduction, it is sometimes difficult to obtain the exact tail probabilities for the one way variance components model (2.1). For this we shall extend the theory to obtain third order approximations for all subsamples, subclasses, and then use third order procedures to obtain highly accurate tail probability approximations. With the above ideas in mind we first have extended the domain of the primary parameter space to include negative values of the primary parameters but still nonnegative values for individual subsample variances. For convenience let $\eta$ denote the extended version of $\sigma_a^2$, and without loss of generality arrange the sequence of subsample variance $\{\sigma_i^2, i = 1, \ldots, k\}$ in ascending order, i.e $\sigma_1^2 \leq \sigma_2^2 \leq \ldots \leq \sigma_k^2$. 
4.1 Extended Parameter Space

In the literature for unbalanced data, the domain for the mean $\mu$, and the experimental errors $\sigma^2$ are typically given by

\[-\infty < \mu < \infty, \quad \sigma^2 > 0\]  

respectively. In the sequel, we extended the domain for $\eta$ to satisfy

\[\eta \geq -\sigma^2/n_1,\]  

redefined the subsample variances and then proceeded with the current third order analysis. To avoid notational growth, we shall denote $\eta + \sigma^2/n_i$ by $\sigma_i^2$.

4.1.1 Linear Model

For $i = 1, \ldots, k$, let

\[\hat{\sigma}_i^2 = \begin{cases} \eta + \sigma^2/n_i & \text{if } \eta > -\sigma^2/n_1, \\ 0 & \text{otherwise} \end{cases}\]  

denote the variance of the $i$-the subsample.

Accordingly we reexpress the canonical parameters for the tangent exponential model in this form:

\[\tilde{\phi}_1(\theta) = -\sum_{i=1}^{k} \frac{(\tilde{y}_i^0 - \mu_i)}{(\eta + \sigma^2/n_i)}\]  

\[\tilde{\phi}_2(\theta) = -\sum_{i=1}^{k} \frac{(\tilde{y}_i^0 - \mu_i)(\tilde{y}_i^0 - \tilde{\mu}_i)}{(\eta + \sigma^2/n_i)(\tilde{\eta}^0 + \tilde{\sigma}^{02}/n_i)}\]  

\[\tilde{\phi}_3(\theta) = -\sum_{i=1}^{k} \frac{\tilde{\sigma}_i^0(\tilde{y}_i^0 - \mu)(\tilde{y}_i^0 - \tilde{\mu})}{n_i(\eta + \sigma^2/n_i)(\tilde{\eta}^0 + \tilde{\sigma}^{02}/n_i)} - \frac{SS^0}{\sigma^2\tilde{\sigma}^0}\]  

57
4.1.2 Quartic model

For this model all subsample variances approach zero at the negative lower bound, \((-\sigma^2/n_i)\). For \(i = 1, \ldots, k\); we define the variance of the \(i\)th subsample as:

\[
\sigma_i^2 = \begin{cases} 
\eta + \sigma^2/n_i & \text{if } \eta \geq 0, \\
\eta + \sigma^2/n_i - n_i^2(n_i/n_i - 1)(\eta^4/\sigma^6) & \text{if } -\sigma^2/n_i < \eta \leq 0, \\
0 & \text{otherwise.}
\end{cases}
\] (4.7)

Now following the same method as outlined in section (3.4), we find

\[
\frac{\partial \bar{y}_i}{\partial \eta} = \frac{(\bar{y}_i^0 - \mu)}{2(\eta + \sigma^2/n_i)} \quad \text{if } \eta \geq 0, \quad \text{and}
\]

\[
= \frac{(\bar{y}_i^0 - \mu)(1 - 4n_i^2(n_i/n_i - 1)(\eta^3/\sigma^6))}{2(\eta + \sigma^2/n_i - n_i^2(n_i/n_i - 1)(\eta^4/\sigma^6))} \quad \text{if } -\sigma^2/n_i < \eta \leq 0,
\]

and compute

\[
\frac{\partial \bar{y}_i^0}{\partial \sigma} = \frac{\sigma(\bar{y}_i^0 - \mu)}{n_i(\eta + \sigma^2/n_i)} \quad \text{if } \eta \geq 0, \quad \text{and}
\]

\[
= \frac{\sigma(\bar{y}_i^0 - \mu)(1 + 3n_i^2(n_i/n_i - 1)(\eta^4/\sigma^6))}{n_i(\eta + \sigma^2/n_i - n_i^2(n_i/n_i - 1)(\eta^4/\sigma^6))} \quad \text{if } -\sigma^2/n_i < \eta \leq 0.
\]

Then derive the extended version of the canonical parameters for the tangent exponential model.

\[
\bar{\phi}_1(\theta) = -\sum_{i=1}^{k} \frac{(\bar{y}_i^0 - \mu)}{(\eta + \sigma^2/n_i)} \quad \text{if } \eta \geq 0, \quad \text{and}
\]

\[
= -\frac{(\bar{y}_i^0 - \mu)}{(\eta + \sigma^2/n_i)} - \sum_{i=2}^{k} \frac{(\bar{y}_i^0 - \mu)}{(\eta + \sigma^2/n_i - n_i^2(n_i/n_i - 1)(\eta^4/\sigma^6))} \quad \text{if } -\sigma^2/n_i < \eta \leq 0.
\]

\[
\bar{\phi}_2(\theta) = -\frac{1}{2} \sum_{i=1}^{k} \frac{(\bar{y}_i^0 - \mu)(\bar{y}_i^0 - \mu)}{(\eta + \sigma^2/n_i)(\eta^0 + \sigma^2/n_i)} \quad \text{if } \eta \geq 0, \text{ and}
\]

\[
= -\frac{1}{2} \frac{(\bar{y}_i^0 - \mu)(\bar{y}_i^0 - \mu)}{(\eta + \sigma^2/n_i)(\eta^0 + \sigma^2/n_i)} - \frac{1}{2} \sum_{i=2}^{k} \frac{(\bar{y}_i^0 - \mu)(\bar{y}_i^0 - \mu)(1 - 4n_i^2(n_i/n_i - 1)(\eta^3/\sigma^6))}{(\eta + \sigma^2/n_i - n_i^2(n_i/n_i - 1)(\eta^4/\sigma^6))} \quad \text{if } -\sigma^2/n_i < \eta \leq 0.
\]


\[ \hat{\varphi}_3(\theta) = -\sum_{i=1}^{k} \frac{\hat{\sigma}^0(\bar{y}^0_i - \mu)(\tilde{y}^0_i - \hat{\mu}^0)}{n_i(\eta + \sigma^2/n_i)(\hat{\eta}^0 + \hat{\sigma}^{02}/n_i)} - \frac{S^0S}{(\sigma^0\sigma^2)} \quad \text{if} \quad \eta \geq 0, \text{and} \]

\[ = -\frac{\sigma^0(\bar{y}^0_1 - \mu)(\tilde{y}^0_1 - \hat{\mu}^0)}{n_1(\eta + \sigma^2/n_1)(\hat{\eta}^0 + \hat{\sigma}^{02}/n_1)} \]

\[ - \sum_{i=2}^{k} \frac{(\sigma^0/n_i)(\bar{y}^0_i - \mu)(\tilde{y}^0_i - \hat{\mu}^0)(1 + 3n_i^3(n_i - n_i)(\hat{\eta}^{04}/\hat{\sigma}^{08}))}{(\eta + \sigma^2/n_i - n_i^3(n_i - 1)(\eta^4/\sigma^6))(\hat{\eta}^0 + \hat{\sigma}^{02}/n_i - n_i^3(n_i - 1)(\hat{\eta}^{04}/\hat{\sigma}^{06}))} \]

\[ - \frac{S^0S}{(\sigma^0\sigma^2)} \quad \text{if} \quad -\sigma^2/n_1 < \eta \leq 0. \]

Putting all the pieces together we get third order tail probabilities in the Lugannani and Rice type formula (1.3) and in the \( R^* \) type formula (1.5).

For the basic model, the likelihood function tends to infinity when \( \mu \) is replaced by the mean of the largest block while the corresponding variance tends to zero. We avoid this for the extended parameter space by requiring that the variances for all subsamples approach zero simultaneously; this is included in the definition of the Quartic model. The following example illustrates this point.

**Example 4.1** Let \( \mu = \bar{y}_1, n = [n_1, n_2] \) with \( n_1 > n_2 \) in the one-way random effects model (2.1) to get:

\[ \ell(\theta) = -\left(\frac{1}{2}\right)\ln(\sigma_1^2) - \left(\frac{1}{2}\right)\ln(\sigma_2^2) - \left(\frac{1}{2}\right)(\bar{y}_2 - \bar{y}_1)^2/\sigma_2^2 - (N/2)\ln(\sigma^2) - S^2/(2\sigma^2) \]

The likelihood function tends to \( \infty \) as \( \sigma_1^2 \) tends to zero. Denote the lower bound for the within-group correlation \( \rho \) by

\[ \rho = -(n_1 - 1)^{-1}, \quad (4.8) \]

then \( \sigma_1^2 \to 0 \) as \( \rho \to 0 \) (Skovgaard 1996).
4.2 Monte Carlo assessment of third order significance

As pointed out in Chapter 3, the F&R procedure produces third order tail probability $p_3(\psi)$ using the Lugannani and Rice type formula (1.3) or the Barndorff-Nielsen type formula (1.5). In repeated random sampling from the one-way random effects model (2.1) with $\psi = \psi_0$ the observed values of $p_3(\psi_0)$ should be uniform and the observed values of the $R^*$ statistics should be approximately standard normal. Note that $R^*$ is equal to $\Phi^{-1}(p_3(\psi_0))$. To assess the accuracy of the F&R third order approximation we generated small and large unbalanced data from the model (2.1) with $\mu = 0, \eta = 1 = \sigma^2$ and computed the values of $R^*, R, p_3(\psi)$, and $p_2(\psi)$ respectively. The values of the $R^*$ statistic and the $R$ statistic are plotted versus the standard normal quantiles to assess the performance of each method in the tail areas. In addition, the coverage performance of the third order method is compared to that of the first order method at the nominal 5% significance level. Thus the target type one error on each side is 2.5%. A Simulation study of 10000 trials for various designs shows that the distribution of the third order statistic $R^*$ is closer to that of the standard gaussian distribution than that of the first order statistic $R$. Moreover, the coverage performance for the F&R method is far better than that of the standard first order likelihood ratio method $R$.

4.2.1 Algorithm

There are two parts to the algorithm: the MLE step and the Third Order step.
MLE Step.

For the one-way variance components model (2.1) the likelihood function (2.24) is maximized with respect to the components of the vector parameter \( \theta = (\eta, \sigma, \mu) \) satisfying (4.7) and subject to the conditions

\[-\infty < \mu < \infty, \sigma > 0, \eta > -\sigma^2/n_i, n_i \geq n_i \text{ for } (i = 1 \ldots k).\]

Notice that the likelihood function \( \ell(\theta) \) is nonlinear in components of the vector parameter \( \theta \) and that two of these components, the variance for random effects \( \eta \) and the variance for random errors \( \sigma^2 \) are constrained. Constraints such as these are uncommon in many standard models such as the exponential families, for example. These constraints and flatness of the likelihood function (among others) seem to be the reason for failure of various numerical procedures, including Newton-Raphson, method to converge. For this reason, we reparametrized the likelihood function \( \ell(\theta) \) in terms of a new vector parameter \( (\gamma, \beta, \mu) \) with the corresponding variances satisfying (4.7) and then maximized the reparametrized likelihood function \( \ell(\gamma, \beta, \mu) \) using the following transformation:

\[\sigma^2 = \exp(\beta) \text{ and } \sigma_i^2 = \exp(\gamma)\] (4.9)

subject to the conditions:

\[-\infty < \gamma < \infty, \]
\[-\infty < \beta < \infty, \]
\[-\infty < \mu < \infty\]

to get the mle for parameters of the one-way random effects model (2.1).

The numerical algorithm for finding the maximum likelihood estimate of \( \theta \) is given below.

1. Compute the variances for all subsamples using the quartic model (4.7).
for $i = 1$ to $k$ do
  if $\eta \geq 0$
    then $\text{var}(\tilde{y}_i) = \eta + \sigma^2/n_i$
  else if $-\sigma^2/n_1 < \eta < 0$
    then $\text{var}(\tilde{y}_i) = \eta + \sigma^2/n_1 - n_i^2(n_1/n_i - 1)(\eta^4/\sigma^6)$
    else $\text{var}(\tilde{y}_i) = 0$
  endif
  endif
endfor.

allsg = seq(var($\tilde{y}_i$), $i = 1, \ldots, k$).

2. Rewrite the likelihood function in terms of $\gamma, \beta$, and $\mu$ using the previous transformation (4.9) and then maximize $\ell(\gamma, \beta, \mu)$ to get $\hat{\theta}$.

\[
\sigma^2 \leftarrow \exp(\beta) \\
\eta \leftarrow \exp(\gamma) - (1/n)(\exp(\beta)) \\
\text{lik} = \ell(\exp(\gamma) - (1/n_1) \exp(\beta), \exp(\beta), \mu) \\
\text{maxm} = \text{ms}(-\text{lik}(\gamma, \beta, \mu), \text{start} = \text{list}(\gamma = a, \beta = b, \mu = c) \\
\text{Repeat until end of the file is reached.}
\]

**Third Order Step**

Our interest lies in constructing a third order significance function $p_3(\psi)$ for testing a scalar interest parameter of the one way random effects model (2.1). This involves calculating the third order p-value for the interest parameter $\psi$ in the presence of nuisance parameters. In this step, we compute the full and the nuisance information determinants calibrated on the $\varphi$ scale, and a scalar linear function of $\varphi(\theta)$ that agrees
with the interest parameter $\psi$ at $\hat{\theta}_\psi$. This is not a difficult step to implement but it requires a reliable and an accurate procedure for computing the maximum likelihood estimate of $\theta$.

For this step we used Maple and the algorithm used is summarized in the following lines.

1. Calculate the observed Fisher information determinants for the full and the constrained models recalibrated with respect to the $\varphi$ scale.
   
   $J(\hat{\theta})$ and $J(\hat{\lambda})(\hat{\theta}_\psi)$

2. Compute the standardized likelihood ratio statistic $R(3.31)$.

3. Compute the standardized maximum likelihood departure $Q(3.30)$.

4. Report third order tail probabilities in the Lugannani and Rice form (1.3) and the Barndorff-Nielsen form (1.5).

5. Repeat until end of the data file is reached.

### 4.2.2 Design

As stated in Chapters 1 – 2, models for studies in which all factors are random are called random effects (or Model II) and those for studies in which some factors are random and some fixed are called mixed models (or Model III). In the sequel, we consider the random effects model (2.1).

We generated 10000 samples from the one-way variance components model (2.1) with $\mu = 0$, $\sigma^2 = 1 = \eta$. The subsample sizes were $n = [6, 6, 6, 5, 5, 5, 4, 4, 4]$, $n = [6, 5, 4]$, $n = [120, 120, 120, 120, 110, 110, 110, 110, 100, 100, 100, 100]$ and $n = [100, 100, 100, 100]$ respectively. The observed significance $p_0(\psi)$ using the $R^*$ statistic
and the observed significance $p_2(\psi)$ using the signed likelihood ratio statistic $R$ were recorded and tested then at the nominal 5% significance. For this, the Pearson $\chi^2$ statistic was used as a test criterion of goodness of fit.

Now we are in a position to examine several Monte Carlo simulations for testing parameters of the one-way random effects model (2.1).

**Example 4.2 (Testing $\eta$)** 10000 samples were generated from the one-way random effects model with $\mu = 0$, $\sigma^2 = 1 = \eta$, and $n = [6, 6, 6, 6, 5, 5, 5, 4, 4, 4]$. The variance of the $i$-subsample of the model (2.1) was computed in accordance with the quartic model (4.7). The significance functions $p_3(\eta)$ and $p_2(\eta)$ using the $R^*$ and the $R$ statistics were calculated and the count of the observed levels of significance in these intervals $[0, .01), [.01, .015), [.015, .025), [.025, .05), [.05, .95), [.95, .975), [.975, .985), [.985, .99), [.99, 1]$ were recorded then in the following table.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>0-.01</th>
<th>.01-.015</th>
<th>.015-.025</th>
<th>.025-.05</th>
<th>.05-.95</th>
<th>.95-.975</th>
<th>.975-.985</th>
<th>.985-.99</th>
<th>.99-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^*$</td>
<td>96</td>
<td>58</td>
<td>110</td>
<td>280</td>
<td>8942</td>
<td>249</td>
<td>99</td>
<td>41</td>
<td>125</td>
</tr>
<tr>
<td>$R$</td>
<td>258</td>
<td>138</td>
<td>200</td>
<td>460</td>
<td>8698</td>
<td>119</td>
<td>49</td>
<td>30</td>
<td>48</td>
</tr>
<tr>
<td>expected</td>
<td>100</td>
<td>50</td>
<td>100</td>
<td>250</td>
<td>9000</td>
<td>250</td>
<td>100</td>
<td>50</td>
<td>100</td>
</tr>
</tbody>
</table>

If the null hypothesis that $p_{31} = .01$, $p_{32} = .005$, $p_{33} = .01$, $p_{34} = .025$, $p_{35} = .9$, $p_{36} = .025$, $p_{37} = .01$, $p_{38} = .005$, $p_{39} = .01$ is true then the statistic

$$QF = \sum_{i=1}^{k} \frac{(x_i - np_i)^2}{np_i}$$

(4.10)

has a chi square distribution with $k - 1$ degrees of freedom. For the data in the above table the well known Pearson $\chi^2$—statistics (4.10) is used as a test criterion.
of goodness of fit. The observed frequencies for \( \Phi(R^*) \) and \( \Phi(R) \) were recorded in rows two and three respectively. The expected counts were recorded in the fourth row. At 8 degrees of freedom the overall \( \chi^2(8) \) is 14.298 for the F&R method using the \( R^* \) statistic and \( \chi^2(8) = 820.7478 \) for the first order method using the standard likelihood ratio method \( R \). In addition, the 5% critical value is \( \chi^2(8) = 15.507 \). Since 14.298 < 15.507 < 820.7478, it seems from the simulation that the F&R method leads to better approximation than with the first order method.

For brevity, the sign test for matched pairs (Fraser 1976, pp302) will be employed for testing the null hypothesis

\[ H_0: \ R^* \text{ is not a better statistic than the standard likelihood ratio statistic } R \]

against the alternative

\[ H_A: \text{ The distribution of the } R^* \text{ statistic is closer to that of the standard normal} \]

The first step is to divide each population into ten random samples of equal size; then compute

\[ d_i^* = \sqrt{QF_i} \text{ for each sample in the } R^* \text{ population}, \]

\[ d_i = \sqrt{QF_i} \text{ for each sample in the } R \text{ population}, \]

\[ s_i = \text{sgn}(d_i - d_i^*) \text{ for } (i = 1, \ldots, 10). \]

Let \( m \) designate the number of pairs \((d_i, d_i^*)\) in the matched pairs sample with \( d_i - d_i^* \neq 0 \) and let the test statistic \( D \) represent the number of pairs \((d_i, d_i^*)\) with \( d_i - d_i^* > 0 \). Clearly, the test statistic \( D \) is distributed as a Binomial random variable \( B(m, 1/2) \). For the data in Example(4.2), the observed values of \( d_i^*, d_i, \) and \( s_i \) are given in Table(4.2). Accordingly, the p-value for the observed value of the test statistic \( D \) is

\[ p(D = 10) = \frac{1}{2^{10}}. \]
A small value such as $\frac{1}{20}$ is strong evidence in favor of the alternative; the distribution of the $R^*$ statistic is closer to that of the standard normal. Then $R^*$ is a better statistic than the standard likelihood ratio statistic $R$ since the distribution of the $R^*$ statistic is closer to that of the standard normal.

The plots for values of the $R^*$ statistic and the standard likelihood ratio statistic $R$ versus the quantiles of the standard normal distribution are in good agreement with that of the standard normal distribution over the interval (-3,3), (Fig.1). At the extreme upper tail the normal quantiles increase at a lower rate. In other words, the left top panel of Fig.1 shows some deviation in the extreme upper tail, but no indication of serious violation of the normality assumption. For both methods the probability that an observation lies outside the interval (-3,3) is less than .002.

Fig.4 shows a second type of plots for the $R^*$ statistic and the $R$ statistic. In each panel of Fig.4 we find a plot, (or a scatter plot as it is sometimes called) with the ordered quantiles of the theoretical distribution on the horizontal axis and the corresponding residuals on the vertical axis. To asses how fairly the residuals deviate from zero the line $y = 0$ is superimposed on all plots of Fig.4 Clearly, the distribution of $R^*$ is symmetric about the origin with a little deviation in the upper tail but no indication of a serious violation of the normality assumption. Obviously, the distribution of the $R$ statistic is not symmetric with a serious deviation from the line $y = 0$.

In conclusion, the coverage performance and the accuracy of the F&R method is far superior to that of the first order method.

In the next example we shall consider testing the mean $\mu$ of the one way random
effects model (2.1).

**Example 4.3 (Testing $\mu$)** As outlined in the above example (4.2) 10000 samples were generated from the one-way random effects model (2.1).

Then the significance function of $\mu$ using the $R^*$ statistic and the standard likelihood statistic $R$ were then recorded. The computed values of the statistic $QF(4.10)$ for the F&R method and for the first order method are 8.073 and 128.04 respectively. At 8 degrees of freedom this translates into a better coverage for the F&R method.

By the same procedure outlined in the previous Example 4.2, the observed values of the three random samples $d_1^*, \ldots, d_{10}^*, d_1, \ldots, d_{10}, s_1, \ldots, s_{10}$ are given in Table (4.4).

Thus, the p-value for the observed value of the test statistic $D$ is

\[
p(d_i - d_i^* > 0) = p(D = 10) = \frac{1}{2^{10}}.
\]
As shown in Fig. 2 the distribution of \( R^* \) and \( R \) are both close to the standard normal distribution. But those plots of Fig. 5 and the p-value for the test statistic \( D \) show that the distribution of the third order statistic \( R^* \) is closer to that of the standard normal.

As stated earlier, our interest lies in testing all components of the vector parameter \( \theta \). For this we shall apply the F&R method to derive new approximations for observed levels of significance for the variance of the experimental errors of the one-way random effects model (2.1).

**Example 4.4 (Testing \( \sigma \))** As outlined in the preceding examples, 10000 samples were generated from the one-way random effects model (2.1) and the counts of the observed levels of significance of \( \sigma \) using the \( R^* \) statistic and the standard likelihood ratio statistic \( R \) were calculated and the counts then recorded. The observed counts

<table>
<thead>
<tr>
<th>Statistic</th>
<th>0-.01</th>
<th>.01-.015</th>
<th>.015-.025</th>
<th>.025-.05</th>
<th>.05-.95</th>
<th>.95-.975</th>
<th>.975-.985</th>
<th>.985-.99</th>
<th>.99-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>( R^* )</td>
<td>126</td>
<td>47</td>
<td>99</td>
<td>250</td>
<td>9008</td>
<td>233</td>
<td>84</td>
<td>47</td>
<td>106</td>
</tr>
<tr>
<td>( R )</td>
<td>113</td>
<td>51</td>
<td>97</td>
<td>256</td>
<td>9014</td>
<td>229</td>
<td>87</td>
<td>56</td>
<td>97</td>
</tr>
<tr>
<td>expected</td>
<td>100</td>
<td>50</td>
<td>100</td>
<td>250</td>
<td>9000</td>
<td>250</td>
<td>100</td>
<td>50</td>
<td>100</td>
</tr>
</tbody>
</table>

of \( \Phi(R^*) \) and \( \Phi(R) \) are tested at \( \alpha = 5\% \) using the chi-square test statistic (4.10). For both methods the observed values of the corresponding \( \chi^2 \)-square test statistic \( QF(4.10) \) are 11.21 and 6.63 respectively. At 8 degrees of freedom this means that the distribution for both methods fit the data. As shown in Fig. 3 the data quantiles for both statistics \( R^* \) and \( R \) agree with that of the standard normal distribution over the interval (-3,3). The data for the three random samples \((d_1, \ldots, d_{10}), (d_1, \ldots, d_{10})\)
and $s_1, \ldots, s_{10}$ with the aid of the signed test show that

$$p(D \geq 5) = \frac{638}{2^{10}}.$$

<table>
<thead>
<tr>
<th>$d_i^*$</th>
<th>2.724</th>
<th>1.892</th>
<th>3.921</th>
<th>3.090</th>
<th>3.315</th>
<th>1.914</th>
<th>3.756</th>
<th>3.141</th>
<th>2.271</th>
<th>3.728</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_i$</td>
<td>3.390</td>
<td>1.51</td>
<td>3.276</td>
<td>3.659</td>
<td>3.135</td>
<td>2.417</td>
<td>3.847</td>
<td>2.834</td>
<td>2.836</td>
<td>3.120</td>
</tr>
<tr>
<td>$s_i$</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>

**Discussion**

In Figures 1-3, the qqnorms are plots of sorted data against the corresponding quantiles of the standard normal distribution, and the qqlines are lines joining the first and the third quartile of the data and the corresponding quantiles of the standard Gaussian distribution.

As shown in Figures 1-6 and Tables 4.2,4.4,4.6 for all components of the vector parameter $\theta = (\eta, \sigma, \mu)$ the distribution of the $R^*$ statistic using the F&R method is closer to the standard normal distribution. In contrast we find a clear departure in the distribution of the standard likelihood ratio statistic $R$ when Figures 4 and 5 are examined. Furthermore the coverage performance of the F&R method is better than that of the first order method,(see Tables 4.1,4.3,4.5). Looking at the right top panel of figure 1 alone we might well conclude that these data appear to be very close to the standard normal distribution but contaminated by a small number of extreme values in both tails. This sort of departure in tails of the data occurs when the expression for the third order term $\frac{1}{R} \ln(Q/R)$ becomes unstable for $\hat{\theta}$ near $\hat{\theta}_\psi$ or $\hat{\theta} = \hat{\theta}_\psi$. To illustrate this point we shall consider this case from Example(4.2), (see left top panel of Fig. 1).
Given the mle for the full model as $\hat{\theta} = [.99982,.9828912,-.5480666]$, and the mle for the restricted model as $\hat{\theta}_n = [1, .98287105, -.5480644]$, the current third order algorithm produces these values $Q = -.0003075$, $R = -.0004$, $\Phi(R^*) = \Phi(R + (1/R)\ln(Q/R)) = \Phi(-.0004 + 191.53) = 1$, and $\Phi(R) + \phi(R)(1/R - 1/Q) = 79.911$.

Furthermore, the $n_i$ pattern for all subsample sizes in the above example are needed to ensure the asymptotic normality and consistency of all estimates under consideration (see Hartley and Rao (1967), Miller (1977), and Westfall (1987)). For this, we assumed the $n_i$ pattern has the form

$$\{6, 5, 4\}, \{6, 5, 4\}, \{6, 5, 4\}, \{6, 5, 4\}.$$
Fig. 1 qqnorm and qqline for the third order statistic $R^*$
and for the first order standard likelihood ratio
$R$ with the variance of random effects $\eta$
as the parameter of interest and
$n = [6, 6, 6, 6, 5, 5, 5, 4, 4, 4]$. 

71
qqnorm and qqline for the
Third order statistic

First order statistic

Fig. 2 qqnorm for the third order statistic $R^*$
and for the first order likelihood ratio statistic $R$ with $\mu$ as the interest parameter and

$n = [6, 6, 6, 6, 5, 5, 5, 4, 4, 4, 4]$. 

72
Fig. 3 qqnorm and qqline for the $R^*$ statistic and for the standard likelihood ratio statistic $R$ when

$\sigma$ is the interest parameter and

$n = [6, 6, 6, 6, 5, 5, 5, 5, 4, 4, 4, 4, 4]$. 

73
Fig. 4 Plots of Residuals vs Ordered Quantiles of the simulated distribution for the third order statistic $R^*$ and for the first order statistic $R$ with the variance of random effects $\eta$ as the interest parameter and

$n = [6, 6, 6, 6, 5, 5, 5, 4, 4, 4, 4]$. 

74
Residuals vs Ordered Quantiles for the Third order statistic

First order statistic

The simulated distribution for the third order statistic $R^*$ and for the first order statistic $R$ with the mean of random effects $\mu$ as the interest parameter and

$n = [6, 6, 6, 6, 5, 5, 5, 4, 4, 4, 4, 4]$. 

Fig. 5 Plots of Residuals vs Ordered Quantiles of

75
were generated from the one-way random effects model (2.1) with $n = 6, 4, 4$. Then

Example 4.5 (Testing $H_0$) As outlined in the preceding examples, 10000 samples

Next we consider testing $H'_0$ and $H'_1 = 0$, and $H'_1 = 1$. $H'_0 = 0$, and $H'_1 = 1$. $H'_1 = 0$.

\[ u = [6, 6, 6, 6, 6, 6, 5, 5, 5, 4, 4, 4, 4, 4] \]

interesting parameter and

statistic $R'$ and for the first order

the simulated distribution for the third order

\[ \text{Fig. 6: Plots of Residuals vs Ordered Quantiles of} \]

\[ \begin{align*}
\text{First Order Statistic} \\
\text{ Ordered Quantiles} &\quad 4 \quad 2 \quad 0 \quad 2 \quad 4 \\
-2.5 &\quad -1.5 &\quad -0.5 \\
\text{Residuals} &\quad -2 \quad -1 \quad 0 \quad 1 \quad 2 \quad 3
\end{align*} \]

\[ \begin{align*}
\text{Third Order Statistic} \\
\text{ Ordered Quantiles} &\quad 4 \quad 2 \quad 0 \quad 2 \quad 4 \\
-2.5 &\quad -1.5 &\quad -0.5 \\
\text{Residuals} &\quad -2 \quad -1 \quad 0 \quad 1 \quad 2 \quad 3
\end{align*} \]
the observed p-values for testing $\eta = 1$ were computed again by the standard likelihood ratio method and the present third-order approximation.

Table 4.7: Counts of the significance function $p(\eta)$

<table>
<thead>
<tr>
<th>Statistic</th>
<th>0-.01</th>
<th>.01-.015</th>
<th>.015-.025</th>
<th>.025-.05</th>
<th>.05-.95</th>
<th>.95-.975</th>
<th>.975-.985</th>
<th>.985-.99</th>
<th>.99-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^*$</td>
<td>127</td>
<td>43</td>
<td>56</td>
<td>167</td>
<td>8746</td>
<td>393</td>
<td>136</td>
<td>73</td>
<td>259</td>
</tr>
<tr>
<td>$R$</td>
<td>935</td>
<td>261</td>
<td>414</td>
<td>757</td>
<td>7509</td>
<td>66</td>
<td>26</td>
<td>10</td>
<td>22</td>
</tr>
<tr>
<td>expected</td>
<td>100</td>
<td>50</td>
<td>100</td>
<td>250</td>
<td>9000</td>
<td>250</td>
<td>100</td>
<td>50</td>
<td>100</td>
</tr>
</tbody>
</table>

In accordance with the data in the above table, the observed value of the $\chi^2$ test statistic $QF(4.10)$ at eight degrees of freedom is

420.5 for the $R^*$ statistic, and

10406.86 for the $R$ statistic.

Both methods are significant at any nominal levels of significance with $\alpha > 0$.

In the notation of Example(4.2), the data in the following Table(4.8) shows that the probability

$$p(D = 10) = \frac{1}{2^{10}}.$$

Table 4.8: Data for testing accuracy of $R^*(\eta)$ vs $R(\eta)$

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_i$</td>
<td>30.647</td>
<td>29.225</td>
<td>27.905</td>
<td>32.589</td>
<td>36.390</td>
<td>33.202</td>
<td>34.241</td>
<td>33.523</td>
<td>34.579</td>
<td>32.713</td>
</tr>
<tr>
<td>$s_i$</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>
Example 4.6 (Testing μ) As in Example (4.5), 10000 samples were generated from the one-way random effects model (2.1) with \( n = [6,5,4] \). Then the observed p-values for testing \( \mu = 0 \) were computed by the standard likelihood ratio method and the current third-order approximation using the F&R.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>0-.01</th>
<th>.01-.015</th>
<th>.015-.025</th>
<th>.025-.05</th>
<th>.05-.95</th>
<th>.95-.975</th>
<th>.975-.985</th>
<th>.985-.99</th>
<th>.99-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>( R^* )</td>
<td>133</td>
<td>46</td>
<td>95</td>
<td>255</td>
<td>8971</td>
<td>252</td>
<td>79</td>
<td>45</td>
<td>124</td>
</tr>
<tr>
<td>( R )</td>
<td>436</td>
<td>134</td>
<td>226</td>
<td>419</td>
<td>7636</td>
<td>373</td>
<td>222</td>
<td>118</td>
<td>436</td>
</tr>
<tr>
<td>expected</td>
<td>100</td>
<td>50</td>
<td>100</td>
<td>250</td>
<td>9000</td>
<td>250</td>
<td>100</td>
<td>50</td>
<td>100</td>
</tr>
</tbody>
</table>

For testing the mean parameter of the one-way random effects model, the observed value of \( QF(4.10) \) at eight degrees of freedom is

\[
22.34 \text{ for the } R^* \text{ statistic and} \\
3180.6 \text{ for the first order method.}
\]

Both methods are significant at \( \alpha = .05 \). As outlined in Example (4.2) the p-value for the test statistic \( D \) is

\[
p(D = 10) = \frac{1}{2^{10}}.
\]

Table 4.10: Data for testing accuracy of \( R^*(\mu) \) vs \( R(\mu) \)

<table>
<thead>
<tr>
<th>( d_i^* )</th>
<th>3.947</th>
<th>3.3</th>
<th>2.416</th>
<th>3.219</th>
<th>3.473</th>
<th>2.693</th>
<th>2.939</th>
<th>1.485</th>
<th>3.518</th>
<th>3.894</th>
</tr>
</thead>
<tbody>
<tr>
<td>( s_i )</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>
Example 4.7 (Testing $\sigma$) As outlined in Example (4.5), 10000 samples were generated from the one-way random effects model (2.1) with $n = [6, 5, 4]$. Then the observed levels of significance of $\eta$ were computed by the standard likelihood ratio method and the current third order method and their frequencies were recorded in the following table.

Table 4.11: Counts of significance function of $p(\sigma)$ with $n = [6, 5, 4]$

<table>
<thead>
<tr>
<th>Statistic</th>
<th>0.00-0.01</th>
<th>0.01-0.015</th>
<th>0.015-0.025</th>
<th>0.025-0.05</th>
<th>0.05-0.95</th>
<th>0.95-0.975</th>
<th>0.975-0.985</th>
<th>0.985-0.99</th>
<th>0.99-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^*$</td>
<td>88</td>
<td>51</td>
<td>106</td>
<td>265</td>
<td>8956</td>
<td>297</td>
<td>98</td>
<td>54</td>
<td>85</td>
</tr>
<tr>
<td>$R$</td>
<td>96</td>
<td>68</td>
<td>102</td>
<td>274</td>
<td>8896</td>
<td>283</td>
<td>135</td>
<td>46</td>
<td>100</td>
</tr>
<tr>
<td>expected</td>
<td>100</td>
<td>50</td>
<td>100</td>
<td>250</td>
<td>9000</td>
<td>250</td>
<td>100</td>
<td>50</td>
<td>100</td>
</tr>
</tbody>
</table>

At eight degrees of freedom the observed value of the $\chi^2$ test statistic (4.10) is 14.38111 for the F&R method and 27.112 for the first order method using $R$.

Note that the 5% critical value is $\chi^2(8) = 15.507$. Using the data in Table (4.12) and the signed test for matched pairs, the $p$-value for observed $D$ is

$p(D \geq 4) = 0.828$.

Table 4.12: Data for testing accuracy of $R^*(\sigma)$ vs $R(\sigma)$

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$s_i$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>
Fig. 7 qqnorm and qqline for the third order statistic $R^*$ and for the first order standard likelihood ratio $R$ with the variance of random effects $\eta$ as the parameter of interest and $n = [6, 5, 4]$. 
Fig. 8 Plots of Residuals vs Ordered Quantiles of the simulated distribution for the third order statistic $R^*$ and for the first order statistic $R$ with the variance of random effects $\eta$ as the interest parameter and $n = [6, 5, 4]$. 
Fig. 9 qqnorm and qqline for the third order statistic $R^*$ and for the first order standard likelihood ratio statistic $R$ with $\mu$ as the interest parameter and $n = [6, 5, 4]$. 
Fig. 10 Plots of Residuals vs Ordered Quantiles of the simulated distribution for the third order statistic $R^*$ and for the first order statistic $R$ with the mean of the model $\mu$ as the interest parameter and $n = [6, 5, 4]$. 
Fig. 11 qnorm and qline for the $R^{*}$ statistic and for the standard likelihood ratio statistic $R$ when $\sigma$ is the interest parameter and $n = [6, 5, 4]$. 
Residuals vs Ordered Quantiles for the Third order statistic

First order statistic

Fig. 12 Plots of Residuals vs Ordered Quantiles of the simulated distribution for the third order statistic $R^*$ and for the first order statistic $R$ with $\sigma$ as the interest parameter and

$n = [6, 5, 4]$.  

It seems appropriate to examine cases with larger blocks, for this we shall consider testing $\eta$ with $n = [100, 100, 100, 100, 100]$ and

$n = [120, 120, 120, 120, 110, 110, 110, 110, 100, 100, 100, 100]$.  

85
Example 4.8 (Testing $\eta$) 5000 samples were generated from the one-way random effects model (2.1) with $\eta = 1 = \sigma, \mu = 0$, and $n = [120, 120, 120, 120, 110, 110, 110, 110, 100, 100, 100, 100, 100]$. Then the observed levels of significance for the variance of random effects $\eta$ were computed using the standard likelihood ratio statistic $R$ and the current third order statistic.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>0-.01</th>
<th>.01-.015</th>
<th>.015-.025</th>
<th>.025-.05</th>
<th>.05-.95</th>
<th>.95-.975</th>
<th>.975-.985</th>
<th>.985-.99</th>
<th>.99-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^*$</td>
<td>43</td>
<td>30</td>
<td>45</td>
<td>131</td>
<td>4490</td>
<td>120</td>
<td>49</td>
<td>23</td>
<td>69</td>
</tr>
<tr>
<td>$R$</td>
<td>119</td>
<td>73</td>
<td>91</td>
<td>231</td>
<td>4349</td>
<td>67</td>
<td>28</td>
<td>9</td>
<td>33</td>
</tr>
<tr>
<td>expected</td>
<td>50</td>
<td>25</td>
<td>50</td>
<td>125</td>
<td>4500</td>
<td>125</td>
<td>50</td>
<td>25</td>
<td>50</td>
</tr>
</tbody>
</table>

The observed value of the $\chi^2$ goodness of fit test statistic $QF(4.10)$ at eight degrees of freedom is

$10.39$ for the current third order method and

$368.57$ for the first order method.

This translates into a better performance for the third order method since $10.39 < 15.507 < 368.57$. Using the signed test for matched pairs, the probability that $D = 10$ is $1/1024$, (see Table 4.14). Thus the F&R method is a better method than the first order method.

Table 4.14: data for testing accuracy of $R^*(\eta)$ vs $R(\eta)$

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$s_i$</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>
Fig. 13 qqnorm and qqline for the third order statistic $R^*$ and for the first order standard likelihood ratio statistic $R$ with the variance of random effects $\eta$ as the parameter of interest and $n = [120, 120, 120, 120, 110, 110, 110, 110, 100, 100, 100, 100, 100]$. 
Fig. 14 Plots of Residuals vs Ordered Quantiles of the simulated distribution for the third order statistic $R^*$ and for the first order statistic $R$ with the variance of random effects $\eta$ as the interest parameter and $n = [120, 120, 120, 120, 110, 110, 110, 110, 100, 100, 100, 100]$. 

For cases with five blocks, the F&R method is used to derive new approximation for the observed levels of significance of the variance of random effects $\eta$ for the model (2.1).
Example 4.9 (Testing \( \eta \)) As outlined in Example(4.8), 5000 samples were generated from the one-way random effects model(2.1) with \( \eta = 1 = \sigma, \mu = 0, \) and \( n = [100, 100, 100, 100, 100] \). Then the observed levels of significance for the variance of random effects \( \eta \) for the one-way model(2.1) were computed using the current third order method.

Table 4.15: Counts of the significance function \( p(\eta) \)

<table>
<thead>
<tr>
<th>Statistic</th>
<th>0-.01</th>
<th>.01-.015</th>
<th>.015-.025</th>
<th>.025-.05</th>
<th>.05-.95</th>
<th>.95-.975</th>
<th>.975-.985</th>
<th>.985-.99</th>
<th>.99-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>( R^* )</td>
<td>53</td>
<td>31</td>
<td>58</td>
<td>136</td>
<td>4478</td>
<td>106</td>
<td>52</td>
<td>25</td>
<td>61</td>
</tr>
<tr>
<td>( R )</td>
<td>256</td>
<td>89</td>
<td>162</td>
<td>296</td>
<td>4095</td>
<td>58</td>
<td>22</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>expected</td>
<td>50</td>
<td>25</td>
<td>50</td>
<td>125</td>
<td>4500</td>
<td>125</td>
<td>50</td>
<td>25</td>
<td>50</td>
</tr>
</tbody>
</table>

At 8 degrees of freedom the observed value of the test statistic \( QF(4.10) \) is:

- 9.364 for the third order method using the \( R^* \) statistic, and
- 1623.29 for the first order method using \( R \). With 8 degrees of freedom, the value corresponding to a 0.05 significance level is 15.507. Since the observed value of \( QF \) is less than 15.507, the data with the third order method fits the model at .05 level of significance. Again the distribution for the third order method is closer to the standard Gaussian distribution since the probability that \( D = 10 \) is 1/1024. Moreover, the observed values for \( d_1^*, d_i, \) and \( s_i \) are listed in Table(4.16).

Discussion

The F&R method is superior to the first order method for the following reasons:

- From Tables 13&15 and Figures 13-16 its clear that the coverage performance for the third order method is far better than that of the first order method.
Table 4.16: Data for testing accuracy of $R^*(\eta)$ vs $R(\eta)$

<table>
<thead>
<tr>
<th>$d_i^*$</th>
<th>1.639</th>
<th>3.113</th>
<th>2.757</th>
<th>1.791</th>
<th>2.378</th>
<th>2.953</th>
<th>1.889</th>
<th>2.539</th>
<th>1.398</th>
<th>3.185</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s_i$</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Using the signed test and the data in Tables 14 & 16 we find the distribution of the current third order statistic $R^*$ is closer to the standard Gaussian distribution.
Fig. 15 qqnorm and qqline for the third order statistic $R^*$ and for the first order standard likelihood ratio statistic $R$ with the variance of random effects $\eta$ as the parameter of interest and $n = [100, 100, 100, 100, 100]$. 
Fig. 16 Plots of Residuals vs Ordered Quantiles of the simulated distribution for the third order statistic $R^*$ and for the first order statistic $R$ with the variance of random effects $\eta$ as the interest parameter and $n = [100, 100, 100, 100, 100]$. 
4.2.3 Summary

The plots in Figures 1-6,9-16 and the data in Tables 4.1-4.6,4.11-4.16

- show that the distribution for the current third order statistic is closer to that of the standard Gaussian distribution. Moreover, the coverage performance of the third order statistic $R^*$ is far better than that of the standard first order statistic $R$.

- For testing the variance of random effects $\eta$, summary of all simulations considered in the this chapter are summerized in tables 4.17-4.18. For cases such as $n = [6, 5, 4]$, the current left tail third approximation is a good approximation when compared with that of the first order method or other methods. For the remaining cases the third order method is far superior to those of the first order method.

- As pointed out earlier, the plots in the right top panel of figures 1, 4, 6,13, 14, 15 and 16 show that the expression for the correction term $(1/R) \ln(Q/R)$ becomes unstable for $\hat{\theta} = \hat{\theta}_\psi$ or $\hat{\theta}$ near $\hat{\theta}_\psi$, (see page 69 and center of right top panel of Fig 2).

- To ensure the asymptotic normality and the consistency of all estimates under consideration, it is necessary to use asymptotic sequences of the model(2.1). This include sequences in the form

$$\{n_1, \ldots, n_k\}, \{n_1, \ldots, n_k\}, \ldots,$$

where the number of replications of $\{n_1, \ldots, n_k\}$ tends to infinity.
Table 4.17: Summary of $\%$ of $p(\eta)$ using the third order statistic

<table>
<thead>
<tr>
<th>Pattern of $n_i$</th>
<th>0-2.5</th>
<th>2.5-97.5</th>
<th>97.5-100.0</th>
<th>Both Tails</th>
<th>Replications</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Stated</strong></td>
<td>2.50</td>
<td>95.0</td>
<td>2.50</td>
<td>5.00</td>
<td>10000</td>
</tr>
<tr>
<td>[6, 6, 6, 6, 5, 5, 5, 4, 4, 4]</td>
<td>2.64</td>
<td>94.71</td>
<td>2.65</td>
<td>5.29</td>
<td>10000</td>
</tr>
<tr>
<td></td>
<td>±.312</td>
<td>±.436</td>
<td>±.312</td>
<td>±.436</td>
<td></td>
</tr>
<tr>
<td>[120, 120, 120, 120, 110, 110, 110, 100, 100, 100]</td>
<td>2.36</td>
<td>94.82</td>
<td>2.82</td>
<td>5.18</td>
<td>5000</td>
</tr>
<tr>
<td></td>
<td>±.442</td>
<td>±.616</td>
<td>±.442</td>
<td>±.616</td>
<td></td>
</tr>
<tr>
<td>[100, 100, 100, 100, 100]</td>
<td>2.84</td>
<td>94.4</td>
<td>2.76</td>
<td>5.60</td>
<td>5000</td>
</tr>
<tr>
<td></td>
<td>±.442</td>
<td>±.616</td>
<td>±.442</td>
<td>±.616</td>
<td></td>
</tr>
<tr>
<td>[6, 5, 4]</td>
<td>2.26</td>
<td>93.06</td>
<td>4.68</td>
<td>6.94</td>
<td>10000</td>
</tr>
<tr>
<td></td>
<td>±.312</td>
<td>±.436</td>
<td>±.312</td>
<td>±.436</td>
<td></td>
</tr>
</tbody>
</table>
Table 4.18: Summary of % of $p(\eta)$ using the standard likelihood ratio statistic

<table>
<thead>
<tr>
<th>Pattern of $n_i$</th>
<th>0-2.5</th>
<th>2.5-97.5</th>
<th>97.5-100.0</th>
<th>Both Tails</th>
<th>Replications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stated</td>
<td>2.50</td>
<td>95.0</td>
<td>2.50</td>
<td>5.00</td>
<td>10000</td>
</tr>
<tr>
<td>[6, 6, 6, 6, 5, 5, 5, 4, 4, 4]</td>
<td>5.96</td>
<td>92.77</td>
<td>1.27</td>
<td>7.23</td>
<td>10000</td>
</tr>
<tr>
<td></td>
<td>±.312</td>
<td>±.436</td>
<td>±.312</td>
<td>±.436</td>
<td></td>
</tr>
<tr>
<td>[120, 120, 120, 120, 110, 110, 110, 100, 100, 100, 100]</td>
<td>5.66</td>
<td>92.94</td>
<td>1.40</td>
<td>7.06</td>
<td>5000</td>
</tr>
<tr>
<td></td>
<td>±.442</td>
<td>±.616</td>
<td>±.442</td>
<td>±.616</td>
<td></td>
</tr>
<tr>
<td>[100, 100, 100, 100, 100]</td>
<td>10.14</td>
<td>88.98</td>
<td>.88</td>
<td>11.02</td>
<td>5000</td>
</tr>
<tr>
<td></td>
<td>±.442</td>
<td>±.616</td>
<td>±.442</td>
<td>±.616</td>
<td></td>
</tr>
<tr>
<td>[6, 5, 4]</td>
<td>16.1</td>
<td>83.32</td>
<td>.58</td>
<td>16.68</td>
<td>10000</td>
</tr>
<tr>
<td></td>
<td>±.312</td>
<td>±.436</td>
<td>±.312</td>
<td>±.436</td>
<td></td>
</tr>
</tbody>
</table>
Chapter 5

Applications

5.1 Introduction

The one-way variance components model (2.1) is used in studies where the levels of the effects have been selected at random and the investigator wishes to obtain information about the parameters of the distribution of these levels.

We shall now describe three different approximate methods to construct \((1 - \alpha)\%\) confidence intervals for the variance of random effects component of the model (2.1) since an exact method does not exist. These methods are the current F&R method using the \(R^*\) statistic, the first order method using the standard likelihood ratio statistic \(R\) and the Burdick & Eickman method (1986). For this four different sets of unbalanced data from various branches of applied statistics will be presented.
5.2 Unbalanced data

By unbalanced data we mean data where the number of observations in the cells are not all equal; e.g., yearly income for individuals classified by sex, education and so on. A data set is said to be balanced if the number of observations in the cells are all the same. Furthermore, we do not consider balanced data with some observations missing as unbalanced data. There are several well known procedures for estimating missing observations; after estimating missing observations, one can use the usual methods of analysis of variance for balanced data. This topic of missing observations will not be pursued here.

We now compute the 95% confidence interval for \( \sigma_a^2 \), ( \( \eta \) in the notation of Chapter 4), using the following examples.

**Example 5.1 (Cow Dairy)** In most developed countries artificial breeding of dairy cows is achieved with bull semen supplied by an artificial breeding corporation. In an experiment on artificial insemination of cows, several semen samples from a bull were tested for their ability to produce conceptions. Table (5.1) represent the percentages of conceptions to services for successive samples for six randomly sampled bulls. These measurements appear in Snedecor and Cochran (1980). It is of interest to test the variation among bulls in the population and the variation within bulls.

The one-way random effects model is used to analyze the above data:

\[
y_{ij} = \mu + \alpha_i + e_{ij} \quad j = 1, \ldots, n_i; \quad i = 1, \ldots, 5.
\]

where

- \( y_{ij} \) = percentages of conceptions obtained from the jth sample taken from the ith bull.
- \( \mu \) = designate the overall mean .
- \( \alpha_i \) = designate the effect due to the ith bull.
\( e_{ij} = \) designate the effect due to the jth sample taken from the i bull

\[
\alpha_i \sim N(0, \sigma^2_{\alpha}), \; e_{ij} \sim N(0, \sigma^2)
\]

\( \alpha_i \) and \( e_{ij} \) are mutually independent. Furthermore,

\[
var(y_{ij}) = \sigma^2_{\alpha} + \sigma^2
\]

and the covariance between pairs in the same group:

\[
cov(y_{ij}, y_{ij'}) = \sigma^2_{\alpha} \text{ for } j \neq j'
\]

Table 5.1: Percentages of conceptions to services for successive samples

<table>
<thead>
<tr>
<th>Observations</th>
<th>Bulls</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 2 3 4 5 6</td>
<td></td>
</tr>
<tr>
<td>46 70 52 47 42 35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>31 59 44 21 64 68</td>
<td></td>
<td></td>
</tr>
<tr>
<td>37 57 70 50 59</td>
<td></td>
<td></td>
</tr>
<tr>
<td>62 40 46 69 38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30 67 14 77 57</td>
<td></td>
<td></td>
</tr>
<tr>
<td>64 81 76</td>
<td></td>
<td></td>
</tr>
<tr>
<td>70 87 57</td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>206 129 394 198 470 479</td>
<td>1876</td>
</tr>
<tr>
<td>( n_i )</td>
<td>5 2 7 5 7 9</td>
<td>35</td>
</tr>
<tr>
<td>mean</td>
<td>41.20 64.50 56.29 39.60 67.14 53.22</td>
<td>321.95</td>
</tr>
</tbody>
</table>
The following table contains the 95% confidence interval for the variance of random effects of the one-way model (2.1) using three different approximation methods: The Burdick & Eickman (1986) method, the first order method using the standard likelihood ratio statistic $R$ and the current F&R method using the $R^*$ statistic.

<table>
<thead>
<tr>
<th>Method</th>
<th>95% confidence interval for $\sigma^2_a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>B &amp; E method</td>
<td>(0.744.095)</td>
</tr>
<tr>
<td>First order method</td>
<td>(0.374)</td>
</tr>
<tr>
<td>F&amp;R method</td>
<td>(0.726)</td>
</tr>
</tbody>
</table>

The two-sided $1 - 2\alpha$ confidence interval of $\sigma^2_a$ using Burdick and Eickman method is

$$\frac{S_{1U}^2 L_m}{(1 + JL_m)F(1 - \alpha, k - 1, \infty)} \leq \frac{S_{1U}^2 U_M}{(1 + JU_M)F(\alpha, k - 1, \infty)}$$

(5.1)

where

$$\bar{J} = \frac{k}{\sum_i 1/n_i} = \frac{6}{1/5 + 1/2 + 1/7 + 1/5 + 1/7 + 1/9} = 4.627,$$

$$S_{1U}^2 = \frac{\bar{J} \sum_i (\bar{y}_{i*} - \bar{y}_{**})^2}{k - 1} = \frac{4.63}{5} (41.2^2 + \ldots + 53.22^2 - (41.2 + \ldots + 53.22)^2/6) = 610.5,$$

$$m = \min(5, 2, 7, 5, 7, 9), \quad M = \max(5, 2, 7, 5, 7, 9), \quad k_2 = 29,$$

$$k - 1 = 5, \quad MSE = 248.29, \quad MSB = 664.41,$$

with

$$L_m = \frac{S_{1U}^2}{\bar{J}(MSE)F(1 - \alpha, k - 1, k_2)} - 1/m = -0.326$$
If $L_m < 0$ then the lower bound in (5.1) is defined to be zero. $F(\alpha, \nu_1, \nu_2)$ designates the $F$-value with $\nu_1$ and $\nu_2$ degrees of freedom that has $\alpha$ area to the left.

$$U_m = \frac{S_{1U}^2}{J(MSE)F(\alpha, k-1, k_2)} - 1/M = 3.19975$$
Example 5.2 (Mississippi River) The data in table 5.2 represent the nitrogen concentrations in parts per million from several sites at six of the randomly sampled influents to the Mississippi River. It is desired to estimate $\sigma^2$ and $\sigma^2$ and to test if there is a significant variation in the nitrogen concentration due to influents. These data were reported by Littell, Milliken, Stroup, Wolfinger (1996).

<table>
<thead>
<tr>
<th>Observations</th>
<th>Infuents</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 2 3 4 5 6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>21 21 20 14 7 41</td>
<td></td>
</tr>
<tr>
<td></td>
<td>27 11 19 24 15 42</td>
<td></td>
</tr>
<tr>
<td></td>
<td>29 18 20 30 18 35</td>
<td></td>
</tr>
<tr>
<td></td>
<td>17 9 11 21 4 34</td>
<td></td>
</tr>
<tr>
<td></td>
<td>19 13 14 31 28 30</td>
<td></td>
</tr>
<tr>
<td></td>
<td>12 23 27</td>
<td></td>
</tr>
<tr>
<td></td>
<td>29 2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>20</td>
<td></td>
</tr>
<tr>
<td></td>
<td>20</td>
<td></td>
</tr>
<tr>
<td></td>
<td>194 97 84 147 72 182</td>
<td></td>
</tr>
<tr>
<td>$n_i$</td>
<td>9 7 5 6 5 5 37</td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td>21.556 13.857 16.8 24.5 14.4 36.40</td>
<td></td>
</tr>
</tbody>
</table>

The model (2.1) is used to analysis the data in Table 5.2 with:

$y_{ij}$ designates the nitrogen concentration observed from the water sample taken from the $j^{th}$ site at the $i^{th}$ influent.

$\mu$ designates the overall mean.
$\alpha_i$ designates the effect due to the $i^{th}$ influent.

$e_{ij}$ designates the effect due to the $j^{th}$ site at the $i^{th}$ influent.

As outlined in Example(5.1), the 95% confidence intervals are recorded in the following table using the three different approximations.

<table>
<thead>
<tr>
<th>Method</th>
<th>95% confidence interval for $\sigma_a^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>B &amp; E method</td>
<td>(7.65, 421.4)</td>
</tr>
<tr>
<td>First order method</td>
<td>(14.23, 230.5)</td>
</tr>
<tr>
<td>F&amp;R method</td>
<td>(19.407.5)</td>
</tr>
</tbody>
</table>

**Example 5.3 (Vegetable Oil)** Five groups of bottles were drawn at random from a moving production line which was filling those bottles with vegetable oil. Table 5.3 gives the weight in ounces of vegetable oil in each bottle. Those measurements were reported by Swallow and Searle (1978). One objective of the experiment was to determine the variability in bottle weights among groups and within groups. The model used to analyze the data is

$$y_{ij} = \mu + \alpha_i + e_{ij} \quad j = 1, \ldots, n_i; \quad i = 1, \ldots, 5.$$ 

where:

$y_{ij}$ designates the weight in ounces of vegetable oil in the $j^{th}$ bottle of the $i^{th}$ group.

$\mu$ designates the over all mean.

$\alpha_i$ designates the effect of the $i^{th}$ group.

$e_{ij}$ designates the effect of the $j^{th}$ bottle within the $i^{th}$ group.

From Table(5.3), the 95% confidence interval for the variance of random effects using the same approximations as outlined in Example(5.1) are:
Table 5.3: Weights of five groups of bottles in ounces

<table>
<thead>
<tr>
<th>observations</th>
<th>Groups</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>尼</td>
<td>15.70</td>
</tr>
<tr>
<td></td>
<td>15.68</td>
</tr>
<tr>
<td></td>
<td>15.64</td>
</tr>
<tr>
<td></td>
<td>15.60</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>62.62</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>ni</td>
<td>4</td>
</tr>
<tr>
<td>mean</td>
<td>15.655</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>95% confidence interval for $\sigma^2_a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>B &amp; E method</td>
<td>(0,.0287)</td>
</tr>
<tr>
<td>First order method</td>
<td>(.00038,.01365)</td>
</tr>
<tr>
<td>F&amp;R method</td>
<td>(.00063,.0285)</td>
</tr>
</tbody>
</table>

Example 5.4 (Pine Boards revisited) Ostle and Mensing (1975) investigated data on the moisture content of pine boards. Five storage methods were examined to determine the effect on the moisture content of white pine lumber. Table(5.4) contains the data obtained from this experiment.

For the Pine Board example the 95% confidence intervals are recorded in this table.
Table 5.4: Moisture contents of pine board samples

<table>
<thead>
<tr>
<th>Observations</th>
<th>Storage Conditions</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>7.3</td>
<td>5.4</td>
</tr>
<tr>
<td>2</td>
<td>8.3</td>
<td>7.4</td>
</tr>
<tr>
<td>3</td>
<td>7.6</td>
<td>7.1</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>39.9</td>
<td>19.9</td>
</tr>
</tbody>
</table>

In the comparing the three approximations on $\sigma^2$, (or $\eta$ in the notation of Chapter 4), it is clear that the first order confidence intervals are the shortest. Furthermore, the upper limits for the B&E method are closer to upper limits of the third order confidence intervals than to those of the first order.
Chapter 6

Future Directions

As pointed out in chapters 1-2, there are various methods available for estimating and testing parameters from unbalanced data, we are now in a position to propose a number of projects which arose while working on the completion of this thesis. These ideas are based on recent third order theory developed by Fraser and Reid(1995).

6.1 Maximum Likelihood Method

Constructing third order observed levels of significance for an interest parameter of the general mixed model(2.19). As a special case, the model(2.1) with unequal error variances. This includes deriving third order p-values for functions of these parameters.
6.2 Restricted Maximum likelihood Method

The maximum likelihood method is not always the most straightforward method for estimating parameters of the mixed model (2.19) and for this Skovgaard (1996) among others recommended the REML method for estimating the variance components of the mixed model (2.19). Under normality assumption and with aid of a transformation introduced by Patterson and Thompson (1971) the likelihood function (2.19) can be decomposed into two parts, one part being free of $\mu$.

$$\ell(\tau, \sigma^2, \mu) = \ell_1(\sigma^2, \tau)\ell_2(\sigma^2, \tau, \mu).$$

Inference concerning $\sigma^2$ or $\tau$ can be constructed from $\ell_1$.

6.3 Bayesian approach

In this work the estimation of $\sigma^2$ and $\sigma^2_2$ in the general unbalanced case (2.19) is explored from a Bayesian point of view and the construction of highly accurate approximations to the posterior distribution of these parameters are considered for a specific prior distribution of these parameters (see Reid 1995b). This includes the implications of negative estimates of the variance on the posterior distribution of the variance components of the model.

6.4 Bootstrap

For unbalanced data, we propose the Bootstrap confidence interval method from a frequentist point of view and from a Bayesian point of view as alternatives to the usual approximate confidence intervals. Bootstrap confidence intervals can be constructed using the parametric Bootstrap method and the nonparametric bootstrap method.
6.5 Regression model with random intercept

The linear regression model (1.2) with unequal error variances is a generalization to the one-way model (2.1) and to the well known Behrens and Fisher problem. The variances of the model (1.2) are given by

\[ \text{var}(e_{ij}) = \sigma_i^2, \quad (j = 1, \ldots, n_i), \text{ and} \]
\[ \text{var}(\alpha_i) = \sigma_{\alpha_i}^2, \quad (i = 1, \ldots, k). \]

This includes deriving third order p-values for functions of these parameters.

6.6 Mixture Model

To understand the effects of departure from normality on inferences about location and scale parameters, the current assumption of normality with regard to the random effects component \( \alpha_i \) of the model (2.1) can be replaced by a mixture of two normals. In choosing a model to represent the distribution for \( \alpha_i \), it is important to keep the mathematical complexity to the minimum and the corresponding likelihood function should be expressible in terms of simple functions of the data.
Notations

\begin{itemize}
  \item iid Independent and Identically distributed \hfill 4
  \item $\gamma(\hat{\theta})$ designates the curvature of $\ell(\theta)$ at its maximum \hfill 6
  \item $K(X)$ designates the cumulative generating function of $X$ \hfill 7
  \item TSS Corrected total sum of squares \hfill 20
  \item $1_k$ A vector consisting of $k$ ones \hfill 22
  \item $J_n$ An $n \times n$ matrix of ones \hfill 22
  \item $N$ denotes $\sum_{i=1}^k (n_i - 1)$ \hfill 24
  \item $S^2$ denotes the sum squares within groups \hfill 24
  \item $\nabla \ell(\theta)$ denotes the gradient of $\ell(\theta)$ \hfill 26
  \item $H(\theta)$ denotes the negative hessian of $\ell(\theta)$ \hfill 26
\end{itemize}
BIBLIOGRAPHY


Patterson, H.D., and Thompson, R., “Recovery of inter-block information when block sizes are unequal,” Biometrika 58, 545-554, 1971.


Tibshirani, R. J., “Non-informative priors for one parameter of many,” Biometrika 76, 604-608.


