Inference for General Random Effects Models

Colleen Hunt

School of Applied Mathematics

Master of Science by research, The University of Adelaide, October 13, 2003

Contents

List of Tables

List of Figures

Abstract

This thesis describes methods associated with general random effects models. It is divided into two parts. Part one describes a technique for investigating mean-variance relationships in random effects models. A simple one-way random effects model is proposed as a basis for deriving a score test for homogeneity of variance in one-way random effects models. An arbitrary mean-variance relationship is captured by a single parameter which allows for the possibility of detecting situations where the variance changes systematically with the mean. Part two derives an approximation to the likelihood function using a Laplace expansion to the fourth order. This approximation may be applied to general models with multiple crossed and/or nested effects. The score test of homogeneity and the approximate likelihood function are examined using simulations and simple data analyses.

Declaration

This work contains no material which has been accepted for the award of any other degree or diploma in any University or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis, when deposited in the University Library, being available for loan and photocopying.

Acknowledgements

- I thank my supervisor Associate Professor Patty Solomon for suggesting the topic of research and for providing guidance and help in the development of the research.
- I am indebted to Dr Bill Venables for his significant contribution towards this thesis and for always having time for me.
- I wish to thank Dr Brian Cullis, the staff of BiometricsSA and especially Assoc. Prof. Ari Verbyla for being there and helping me through some difficult times.
- This work was carried out in the Department of Statistics. The Department was merged to form what is now the School of Applied Mathematics. The Department of Statistics provided a positive environment for statistical research.
- This research was carried out under a Postgraduate Research Scholarship funded through a grant from the Australian Research Council.

No animals were harmed throughout the making of this thesis.

Chapter 1

Introduction

1.1 Mixed effects models

Mixed effects in linear, generalized linear and non-linear models are used widely throughout the statistical world. The range of areas where they may be used is varied and extensive. One use of random effects is to explain the error terms when we have multiple sources of variation. Classical texts such as Scheffe (1959) and Searle (1971), show the common, standard methods for describing and working with linear models with fixed and random effects, estimating coefficients, deriving expected mean squares and analysis of variance. Searle et al. (1992) have written a recent book devoted entirely to estimating variance components. It covers ANOVA estimation, matrix formulation of models, maximum likelihood estimation, prediction of random effects, all for balanced and unbalanced data.

Recent publications in the general area of variance component estimation include Aitkin (1999), Steele (1996), Cox (1998), Abrams & Sanso (1998), Breslow & Clayton (1993), Breslow & Lin (1995), Fraser et al. (1994), Lee & Nelder (1996), Lin (1997), Schall (1991) and Gilmour et al. (1985) to name a few. Most of these authors are interested in estimating the actual variance components; another use of mixed models is to estimate the effects of data that comprises from a sample of a larger population and repeated measurements are to be taken from this population. In this case we can estimate the actual random effect predicted values by using BLUPs (Best Linear Unbiased Predictions).

Robinson (1991) discusses the use of BLUP for estimating random effects. A brief history of the derivation of BLUP is given including the classical case, Bayesian, and Henderson and Goldberger's derivations. Robinson then goes on to link these with other statistical theories, such as the recovery of inter-block information, random effects models, estimation of outliers and ranking and selection, giving a number of applications.

Breslow & Clayton (1993) discuss approximate inference in generalized linear mixed models. They consider hierarchical generalized linear models with random effects which are distributed with a multivariate normal distribution, then go on to discuss the use of penalised quasi-likelihood and variance component estimation.

A major difficulty with such models is that the true likelihoods are typically intractable, and usually involve high-dimensional integrals which need to be approximated or estimated in some way. In this thesis, the major themes are firstly the application of Laplace approximations (Barndorff-Nielson & Cox, 1989) and Solomon & Cox (1992) approximations to specific models and problems, and secondly the derivation of generalizations of these approximations to models with multiple random effects. The outline of the thesis is described below.

This thesis is in two parts, Chapter two comprising part one and Chapters three to five comprising part two. The subdivision represents two distinct problems in the analysis of complex data, namely establishing if a mean-variance relationship exists, and incorporating random effects in general models. The need to approximate the likelihood in the two cases provides the connecting theme of the thesis.

1.2 Part One

This part of the thesis forms the foundation for a submitted paper (Hunt & Solomon, 1999) which is currently being revised. It is common in repeated measures data to observe a mean-variance relationship. Authors such as Liu & Pierce (1993), Commenges & Jacqmin-Gadda (1997), Jacqmin-Gadda & Commenges (1995) and Cook & Ng (1999) have considered this problem. In part one, a test to determine the possible existence of a relationship between the mean and the variance in longitudinal data is proposed. This score test statistic is based on a proposed one-way mean-variance model having a parametric representation which is indexed by a single parameter κ. The parameter κ is zero if no mean-variance relationship exists and non-zero otherwise. The likelihood function underlying the procedure must be approximated and the approximation is based here on high-order Laplace expansions. The score statistic is therefore also based on this approximation and this forms the material of Chapter two. The performance of the test is then studied via simulations and analytical results. Applications are given using data on CD4 cell counts and blood pressure.

The simulation studies showed the distribution of the test statistic to be approximately standard normal under a reasonably broad range of assumptions and thus the test is reasonable both when the parameters are assumed known and also when one or more parameters are estimated. When applied to CD4 data from the San Fransisco Men's Health Study and blood pressure data from the International Prospective Primary Prevention Study in Hypertension, mean-variance relationships were detected by the test, and transformations of the data were shown to decrease the relationships.

1.3 Part Two

Part two discusses approximations of the likelihood function with the aim of finding an expression which can represent any general model. Chapter three shows the second-order approximation as used as a basis in papers such as Shun (1997) and Lin & Breslow (1996). These authors have discussed the use of a second-order Laplace approximation and have extended this in various ways to make their approximations behave more like the true likelihood.

Chapter four establishes a general form of the approximation to the fourth-order with no adjustments. We want to show that this approximation compares well to the other approximations of this kind, but in a simpler form. Approximate likelihoods for a single variance component are extended to balanced models containing two or more variance components. Laplace expansions have been considered by other authors such as Steele (1996), Vonesh (1996), Wolfinger (1993), Lin & Breslow (1996), Shun (1997) and Shun & McCullagh (1995).

A fourth-order approximation to the true likelihood is obtained for general models with two or more random effects. This approximation is for fully crossed models, nested models and models with interactions, as long as the model has a well-defined conditional log-likelihood, which is differentiable to the fourth-order. Models with two independent crossed or nested effects are further examined in an explicit form.

The performance of these approximations is studied in Chapter five using simulations and examples. It is shown that our approximation to the likelihood for the linear model is very close to its true likelihood. Simulations are performed for a Poisson model, a model with an exponential function of nested random effects and a logistic regression model. In all cases the variance components are estimated and compared to the known simulation values, and profile likelihood plots show the shape of the approximation. The major example studied is the well-known Salamander data (McCullagh & Nelder, 1989), for which estimates for the variance components are found and compared to previous results on these data.

Chapter 2

A Test for Homogeneity in one-way variance component models

Variance components in nonlinear models and generalized linear models in particular are currently receiving considerable attention in the literature (see for instance Vonesh, 1996, Breslow & Lin, 1995, Wolfinger, 1993, Solomon & Cox, 1992, Breslow & Clayton, 1993, Lee & Nelder, 1996, Shun, 1997, Lin & Breslow, 1996 Cox, 1998 and Hodges, 1998). Such models are often handled with much attendant complexity. However simpler methods, both formal and informal, are required to give one an idea of when and why more sophisticated analysis is needed, and to provide starting values for parameter estimates in more complex models.

In this Chapter, we propose a score test for investigating whether a particular kind of relationship exists between group means and variances in repeated measures data. Such a relationship may be due to an exponential family error structure or to an arbitrary error structure, and although it may be difficult to distinguish between them, we are interested in detecting situations where the possibly arbitrary error variance changes systematically with the mean. How to handle such relationships in practice will depend on the context and purpose of the particular problem under study. Whatever the context though, it is useful to have available formal and informal techniques for assessing potentially important mean-variance relationships in models with variance components and often to check if a transformation has been effective in removing it. Solomon (1985) and Solomon & Cox (1992) suggest some informal procedures and Hodges (1998) and Cox (1998) provide overviews from rather different perspectives. The purpose of this Chapter is to provide a formal test for determining mean-variance relationships based on a general random effects model. Score tests for homogeneity in generalized linear and other models are considered by Jacqmin-Gadda & Commenges (1995), Commenges & Jacqmin-Gadda (1997) and Cook & Ng (1999), among others. Authors such as these consider, for example, testing for homogeneity in terms of differing distributions or testing for differences in variation between groups effects. Our aim is to test for a specific kind of homogeneity, that is, one defined by a simple one-way model which captures a possible mean-variance relationship.

2.1 Introduction

Suppose we have m groups or individuals, indexed by i, and r_i repeated measurements for each individual. Many forms of mean-variance relationships commonly encountered in practice may be credibly modelled by y_{ij}, $i = 1, \ldots, m$, $j = 1, \ldots, r_i$,

$$y_{ij} = \mu + \eta_i + \varepsilon_{ij} e^{\kappa \eta_i} \tag{2.1}$$

where μ is the overall mean, κ is a constant, the η_i are normally distributed with mean 0 and variance σ_η^2, independent of the errors ε, which are also assumed independent normal with mean 0 and variance σ^2. The multiplicative term $e^{\kappa\eta_i}$ allows the error variance to change systematically with increasing or decreasing η_i. When

$\kappa = 0$, we obtain the usual homoscedastic normal theory one-way variance component model.

For a situation with $\kappa \neq 0$, the usual random effects analysis of variance that ignores this relationship would not give accurate estimates for σ^2 and σ_η^2. In studies where estimation of the variance component σ_η^2 is of primary importance it is vital to detect the existence of a mean-variance relationship that can for example artificially give an impression of a large variance component between individuals where none in reality is present.

We note that the parameter κ cannot be separately estimated from the η_i, even conditional on η_i. However, for our purposes, this does not pose a problem.

Solomon & Cox (1992) propose an approximate non-linear model

$$y_{ij} = \mu + A_i + B_{ij} + \alpha_{20} A_i^2 + \alpha_{11} A_i B_{ij} + \alpha_{02} B_{ij}^2 \tag{2.2}$$

where A_i and B_{ij} are random effects, both with mean zero and standard deviations σ_A and σ_B respectively. This model is proposed as a basis for investigations of departures from the normal-theory model $y_{ij} = \mu + A_i + B_{ij}$. The parameters $(\alpha_{20}, \alpha_{11}, \alpha_{02})$ are defined to capture skewness of the random effects and heterogeneity of the within-group variation. This model may be directly compared to model (2.1). By assuming no skewness $(\alpha_{20}, \alpha_{02})$ we can rewrite the model as

$$\begin{aligned} y_{ij} &= \mu + A_i + B_{ij}(1 + \alpha_{11} A_i) \\ &\approx \mu + A_i + B_{ij} e^{\alpha_{11} A_i}. \end{aligned}$$

They define a dimensionless parameter $\rho_{11} = \alpha_{11}\sigma_A$, which measures the rate of change of the conditional standard deviation of the within-group variation with the group mean. Our κ is directly related to ρ_{11} by $\kappa = \rho_{11}/\sigma_\eta^2$, which follows from our model (2.1) and the Solomon and Cox model (2.2) (i.e. κ is approximately equal to the interaction parameter α_{11} defined in their paper). This issue is discussed further in Section 2.7.2.

By letting η_i be fixed for one individual i, the conditional mean and variance of y_{ij} (from model 2.1) become

$$\begin{aligned}\mathrm{E}(y_{ij}|\eta_i) &= \mu + \eta_i \\ \mathrm{Var}(y_{ij}|\eta_i) &= \sigma^2 \exp(2\kappa\eta_i) \\ &= \sigma^2 \exp\left\{2\kappa(\mathrm{E}(y_{ij}|\eta_i) - \mu)\right\}.\end{aligned}$$

The log likelihood for the ith group conditional on η_i for model (2.1) is given by

$$l_i(\mu, \sigma^2, \kappa; y_i|\eta_i) = -\frac{r_i}{2}\log(2\pi\sigma^2) - r_i\kappa\eta_i - \frac{\sum_{j=1}^{r_i}(y_{ij} - \mu - \eta_i)^2}{2\sigma^2\exp(2\kappa\eta_i)}; \tag{2.3}$$

throughout, this is referred to as l_i.

The likelihood for the model with random effects η_i, $i = 1, ..., n$ is obtained by taking expectations of the conditional likelihood:

$$L(\mu, \sigma^2, \kappa, \sigma_\eta^2; y) = \prod_{i-1}^{m}(2\pi\sigma_\eta^2)^{-\frac{1}{2}} \int_{-\infty}^{\infty} \exp\left(l_i - \frac{\eta_i^2}{2\sigma_\eta^2}\right) d\eta_i.$$

In section 2.2 the exact score test of null hypothesis called **mevar1** is developed. The score test statistic is simplified by using a Laplace approximation to the true likelihood. Two score test statistics are developed in section 2.4, firstly **mevar2**, a score test statistic with all parameters known, and **mevar2a**, a score test statistic which allows for estimation of parameters. We then study the performance of these tests using simulations and power calculations and explore the behaviour of observed variability in CD4 cell counts from the San Fransisco Men's Health Study and in blood pressure data.

2.2 A Score test

A detailed discussion of the score test method may be found in Cox & Hinkley (1974). Essentially, for any parameter θ, a vector of length p, the score U is defined as a vector having the ith element as the derivative of the log-likelihood function with respect to θ_i ie.

$$U = \begin{bmatrix} \frac{\partial l}{\partial \theta_1} \\ \vdots \\ \frac{\partial l}{\partial \theta_p} \end{bmatrix}$$

The variance-covariance matrix of the U_i's is the information matrix $\mathcal{I}$ with dimension $p \times p$ containing elements given by

$$\mathcal{I} = -\frac{\partial^2 l}{\partial \theta \partial \theta^T};$$

this is also known as the observed Fisher information matrix. To test the hypothesis $\theta = \theta_0$, evaluate

$$U^T \mathcal{I}^{-1} U \tag{2.4}$$

at $\theta = \theta_0$. This statistic is approximately distributed as χ^2_p if the null hypothesis is true. The score test for the null hypothesis $H_0 : \kappa = \kappa_0$ against the two-sided alternative $H_A : \kappa \neq \kappa_0$ for model 2.1 is based on this statistic.

As an example of the use of this formula consider model (2.1) as a fixed effects model. The form of the score test for fixed η_i, known μ and σ^2 and for the null hypothesis $\kappa = 0$ is given by

$$U = \left.\frac{\partial l}{\partial \kappa}\right|_{\kappa=0} = \sum_{i=1}^{m} \eta_i \left(\frac{\sum_{j=1}^{r_i} (y_{ij} - \mu - \eta_i)^2}{\sigma^2} - r_i \right)$$

$$\mathcal{I} = -\left.\frac{\partial^2 l}{\partial \kappa^2}\right|_{\kappa=0} = \sum_{i=1}^{m} \frac{2\eta_i^2 \sum_{j=1}^{r_i} (y_{ij} - \mu - \eta_i)^2}{\sigma^2}$$

The development of the test statistic **mevar1** follows.

The test for $H_0 : \kappa = 0$ is based on the statistic given in (2.4). Now let the η_i's be random effects, and assume all parameters μ, σ^2 and σ^2_η are known. There are m random effects. Then the score function is

$$U = \partial l / \partial \kappa|_{\kappa=0}$$

$$= \frac{\partial}{\partial \kappa} \sum_{i=1}^{m} \left[\log \int \exp(f(\kappa, \eta_i)) d\eta_i \right] |_{\kappa=0}$$

$$= \sum_{i=1}^{m} \left[\frac{\frac{\partial}{\partial \kappa} \left(\int \exp(f(\kappa, \eta_i)) d\eta_i \right)}{\left(\int \exp(f(\kappa, \eta_i)) d\eta_i \right)} \right] \Bigg|_{\kappa=0}.$$

For our model with conditional log likelihood l_i and random effects $\eta_i, i = 1, ..., m$, the function f is $f(\kappa, \eta_i) = l_i - \log(2\pi\sigma_\eta^2)/2 - \eta_i^2/2\sigma_\eta^2$, note that the full log-likelihood can be written as a summation since the η_i are independent.

Now consider just one term in the summation.

Assuming sufficient regularity to interchange the order of integration and differentiation, we can write

$$\frac{\partial}{\partial \kappa} \left[\int \exp(f(\kappa, \eta_i)) d\eta_i \right] |_{\kappa=0} = \int \left(\frac{\partial}{\partial \kappa} \exp(f(\kappa, \eta_i))|_{\kappa=0} \right) d\eta_i.$$

Using $l_i = -r_i \log(2\pi\sigma^2)/2 - r_i\kappa\eta_i - \sum_{j=1}^{r_i}(y_{ij} - \mu - \eta_i)^2/(2\sigma^2 \exp(2\kappa\eta_i))$ we get

$$(2\pi\sigma^2)^{-\frac{r_i}{2}} (2\pi\sigma_\eta^2)^{-\frac{1}{2}} \int \left(-r_i\eta_i + \frac{\eta_i \sum_{j=1}^{r_i}(y_{ij} - \mu - \eta_i)^2}{\sigma^2} \right)$$

$$\exp\left\{ \frac{-\sum_{j=1}^{r_i}(y_{ij} - \mu - \eta_i)^2}{2\sigma^2} - \frac{\eta_i^2}{2\sigma_\eta^2} \right\} d\eta_i$$

$$= \frac{(2\pi\sigma^2)^{(1-r_i)/2}}{\sqrt{\sigma^2 + r_i\sigma_\eta^2}} \exp\left(\frac{\sum_{j=1}^{r_i}(y_{ij} - \mu)^2}{-2\sigma^2} + \frac{(y_{i.} - r_i\mu)^2\sigma_\eta^2}{2\sigma^2(\sigma^2 + r_i\sigma_\eta^2)} \right)$$

$$\times \frac{\sigma_\eta^2(y_{i.} - r_i\mu)}{\sigma^2(\sigma^2 + r_i\sigma_\eta^2)^3} \times \left[(\sum_{j=1}^{r_i}(y_{ij} - \mu)^2 - r_i\sigma^2 - 2\sigma^2)(\sigma^2 + r_i\sigma_\eta^2)^2 \right.$$

$$\left. -\sigma_\eta^2(2\sigma^2 + r_i\sigma_\eta^2)(y_{i.} - r_i\mu)^2 + 3\sigma^2\sigma_\eta^2 r_i(\sigma^2 + r_i\sigma_\eta^2) \right].$$

The expression $[\int \exp(f(\kappa, \eta_i)) d\eta_i] |_{\kappa=0}$ reduces to the true likelihood (considering just one term η_i from the total summation) for a standard linear model, i.e.

$$\frac{(2\pi\sigma^2)^{(1-r_i)/2}}{\sqrt{\sigma^2 + r_i\sigma_\eta^2}} \exp\left(\frac{\sum_{j=1}^{r_i}(y_{ij} - \mu)^2}{-2\sigma^2} + \frac{(y_{i.} - r_i\mu)^2\sigma_\eta^2}{2\sigma^2(\sigma^2 + r_i\sigma_\eta^2)} \right),$$

giving the full expression for the score, U, as

$$\sum_{i=1}^{m} \frac{\sigma_\eta^2(y_{i.} - r_i\mu)}{\sigma^2(\sigma^2 + r_i\sigma_\eta^2)^3} \left[(\sum_{j=1}^{r_i}(y_{ij} - \mu)^2 - r_i\sigma^2 - 2\sigma^2)(\sigma^2 + r_i\sigma_\eta^2)^2 \right.$$

$$\left. -\sigma_\eta^2(2\sigma^2 + r_i\sigma_\eta^2)(y_{i.} - r_i\mu)^2 + 3\sigma^2\sigma_\eta^2 r_i(\sigma^2 + r_i\sigma_\eta^2) \right]. \quad (2.5)$$

Now the information is written

$$\begin{aligned} \mathcal{I} &= -\frac{\partial^2 l}{\partial \kappa^2} \\ &= -\frac{d^2}{d\kappa^2}\sum_{i=1}^{m}\left[\log \int \exp(f(\kappa, \eta_i))d\eta_i\right]\Big|_{\kappa=0} \\ &= \sum_{i=1}^{m}\left[-\frac{\frac{d^2}{d\kappa^2}(\int \exp(f(\kappa,\eta_i))d\eta_i)}{(\int \exp(f(\kappa,\eta_i))d\eta_i)}\right]\Big|_{\kappa=0} \\ &\quad + \sum_{i=1}^{m}\left[\frac{\frac{d}{d\kappa}(\int \exp(f(\kappa,\eta_i))d\eta_i)}{(\int \exp(f(\kappa,\eta_i))d\eta_i)}\right]^2\Big|_{\kappa=0}, \end{aligned} \quad (2.6)$$

and considering one term only from the total summation gives

$$\frac{d^2}{d\kappa^2}\left[\int \exp(f(\kappa,\eta_i))d\eta_i\right]|_{\kappa=0}$$

$$= \int \left(\frac{\partial^2}{\partial^2\kappa}\exp(f(\kappa,\eta_i))|_{\kappa=0}\right)d\eta_i$$

$$= \text{Const}\int \left[\left(-\frac{2\eta_i^2\sum_{j=1}^{r_i}(y_{ij}-\mu-\eta_i)^2}{\sigma^2}\right) + \left((-r_i\eta_i + \frac{\eta_i\sum_{j=1}^{r_i}(y_{ij}-\mu-\eta_i)^2}{\sigma^2})\right)^2\right]$$

$$\times \exp\left\{\frac{-\sum_{j=1}^{r_i}(y_{ij}-\mu-\eta_i)^2}{2\sigma^2} - \frac{\eta_i^2}{2\sigma_\eta^2}\right\} d\eta_i$$

$$= \frac{(2\pi\sigma^2)^{(1-r_i)/2}}{\sqrt{\sigma^2 + r_i\sigma_\eta^2}}\exp\left(\frac{\sum_{j=1}^{r_i}(y_{ij}-\mu)^2}{-2\sigma^2} + \frac{(y_{i.}-r_i\mu)^2\sigma_\eta^2}{2\sigma^2(\sigma^2+r_i\sigma_\eta^2)}\right) \times$$

$$-\frac{2\sigma_\eta^2}{\sigma^2(\sigma^2+r_i\sigma_\eta^2)^4}\left[\sum_{j=1}^{r_i}(y_{ij}-\mu)^2(\sigma^2+r_i\sigma_\eta^2)^2\left((\sigma^2+r_i\sigma_\eta^2)\sigma^2+(y_{i.}-r_i\mu)^2\sigma_\eta^2\right)\right.$$

$$\left.+\,3r_i\sigma^4\sigma_\eta^2(\sigma^2+r_i\sigma_\eta^2)^2-6(\sigma^2+r_i\sigma_\eta^2)\sigma_\eta^2\sigma^4(y_{i.}-r_i\mu)^2-\sigma_\eta^4(y_{i.}-r_i\mu)^4(2\sigma^2+r_i\sigma_\eta^2)\right]$$

$$+\frac{\sigma_\eta^2}{\sigma^2(\sigma^2+r_i\sigma_\eta^2)^5}N_i^2+\frac{6r_i\sigma_\eta^4}{(\sigma^2+r_i\sigma_\eta^2)^4}N_i+2\frac{\sigma_\eta^2(r_i\sigma_\eta^2-2\sigma^2)}{(\sigma^2+r_i\sigma_\eta^2)^2}(y_{..}-r_i\mu)M_i+M_i^2$$

$$+15\frac{r_i^2\sigma^2\sigma_\eta^6}{(\sigma^2+r_i\sigma_\eta^2)^3}+\frac{\sigma_\eta^4(r_i\sigma_\eta^2-2\sigma^2)^2}{(\sigma^2+r_i\sigma_\eta^2)^4}\left(2(y_{i.}-r_i\mu)^2+(y_{..}-r_i\mu)^2\right) \tag{2.7}$$

where

$$\begin{aligned}
N_i &= (\sigma^2+r_i\sigma_\eta^2)^2\Big(\sum_{j=1}^{r_i}(y_{ij}-\mu)^2-r_i\sigma^2\Big)-\sigma_\eta^2(y_{i.}-r_i\mu)^2(4\sigma^2+r_i\sigma_\eta^2)\\
M_i &= \frac{\sigma_\eta^2(y_{i.}-r_i\mu)}{\sigma^2(\sigma^2+r_i\sigma_\eta^2)^3}\left((\sigma^2+r_i\sigma_\eta^2)^2\Big(\sum_{j=1}^{r_i}(y_{ij}-\mu)^2-r_i\sigma^2\Big)-\sigma_\eta^2(y_{i.}-r_i\mu)^2(r_i\sigma_\eta^2+2\sigma^2)\right)
\end{aligned}$$

This leads to the full expression (2.6) for the information as

$$-\sum_{i=1}^{m}\frac{2\sigma_\eta^2}{\sigma^2(\sigma^2+r_i\sigma_\eta^2)^4}\left[\sum_{j=1}^{r_i}(y_{ij}-\mu)^2(\sigma^2+r_i\sigma_\eta^2)^2\left((\sigma^2+r_i\sigma_\eta^2)\sigma^2+(y_{i.}-r_i\mu)^2\sigma_\eta^2\right)\right.$$

$$\left.+\,3r_i\sigma^4\sigma_\eta^2(\sigma^2+r_i\sigma_\eta^2)^2-6(\sigma^2+r_i\sigma_\eta^2)\sigma_\eta^2\sigma^4(y_{i.}-r_i\mu)^2-\sigma_\eta^4(y_{i.}-r_i\mu)^4(2\sigma^2+r_i\sigma_\eta^2)\right]$$

$$+\frac{\sigma_\eta^2}{\sigma^2(\sigma^2+r_i\sigma_\eta^2)^5}\sum_{i=1}^{m}N_i^2+\frac{6r_i\sigma_\eta^4}{(\sigma^2+r_i\sigma_\eta^2)^4}\sum_{i=1}^{m}N_i+15\frac{mr_i^2\sigma^2\sigma_\eta^6}{(\sigma^2+r_i\sigma_\eta^2)^3}$$

$$+\frac{2\sigma_\eta^4(r_i\sigma_\eta^2-2\sigma^2)^2\sum_{i=1}^{m}(y_{i.}-r_i\mu)^2}{(\sigma^2+r_i\sigma_\eta^2)^4} \tag{2.8}$$

The expression for the score test statistic under the null hypothesis $\kappa = 0$ is written as the score function, (2.5) divided by the square root of (2.8). This version of the test will be referred to as **mevar1**.

Although **mevar1** is an exact expression for the score test statistic for testing the null hypothesis $\kappa = 0$, it is very cumbersome, therefore we seek to find some simpler expression that performs well. It is also only obtained relatively simply under the null hypothesis $\kappa = 0$. It would be an advantage to have a more general analytical form which allows for non-zero hypotheses. Another alternative could be, for example, numerical integration. We have chosen to explore an approximate score test statistic by approximating the true likelihood as detailed in the following Section.

2.3 An approximate likelihood function

We base our approximation on a Laplace expansion of the true likelihood (see Barndorff-Nielson & Cox (1989) for a discussion of Laplace expansions). The application of Laplaces' method has had much discussion, see for instance Barndorff-Nielson & Cox (1989), Liu & Pierce (1993), Breslow & Lin (1995).

Essentially, the approximation is derived from a Taylor series expansion. Previous work by Solomon & Cox (1992), Breslow & Lin (1995), Shun & McCullagh (1995) have shown that including higher-order terms in the expansion may substantially improve the approximation to the true likelihood. More detailed discussion of Laplace approximations and previous work in this area is given later in the Thesis, see particularly Chapter 3.

We consider including up to 4 terms. The fourth-order Taylor expansion for $\left(l_i - \frac{\eta_i^2}{2\sigma_\eta^2}\right)$ about its (local) maximum, denoted by $\tilde{\eta}_i$, evaluated at $\mu, \sigma^2, \sigma_\eta^2$ and κ, is given by

$$\begin{aligned} l_i - \frac{\eta_i^2}{2\sigma_\eta^2} &= \tilde{l}_i - \frac{\tilde{\eta}_i^2}{2\sigma_\eta^2} + \left(\tilde{l}_i^{(1)} - \frac{\tilde{\eta}_i}{\sigma_\eta^2}\right)(\eta_i - \tilde{\eta}_i) + \frac{1}{2}\left(\tilde{l}_i^{(2)} - \frac{1}{\sigma_\eta^2}\right)(\eta_i - \tilde{\eta}_i)^2 \\ &\quad + \frac{1}{6}\tilde{l}_i^{(3)}(\eta_i - \tilde{\eta}_i)^3 + \frac{1}{24}\tilde{l}_i^{(4)}(\eta_i - \tilde{\eta}_i)^4 \end{aligned}$$

where $\tilde{l}_i = l_i(\tilde{\eta}_i)$, $\tilde{l}_i^{(k)} = \frac{\partial^k l_i}{\partial \eta_i^k}|_{\tilde{\eta}_i}$. By definition, $\tilde{\eta}_i/\sigma_\eta^2 = \tilde{l}_i^{(1)}$, therefore the approximation to the full likelihood becomes

$$\prod_{i=1}^{m}(2\pi\sigma_\eta^2)^{-\frac{1}{2}} \int_{-\infty}^{\infty} \exp\left\{-\frac{\tilde{\eta}_i^2}{2\sigma_\eta^2} + \tilde{l}_i + \frac{1}{2}(\tilde{l}_i^{(2)} - \frac{1}{\sigma_\eta^2})(\eta_i - \tilde{\eta}_i)^2 \right.$$

$$\left. + \frac{1}{6}\tilde{l}_i^{(3)}(\eta_i - \tilde{\eta}_i)^3 + \frac{1}{24}\tilde{l}_i^{(4)}(\eta_i - \tilde{\eta}_i)^4 \right\} d\eta_i$$

This is simplified by the expansion of the exponential term

$$\begin{aligned} L \quad &\approx \quad \prod_{i=1}^{m} \frac{\exp(\tilde{l}_i - \frac{\tilde{\eta}_i^2}{2\sigma_\eta^2})}{\sqrt{2\pi\sigma_\eta^2}} \int_{-\infty}^{\infty} \exp\left\{\frac{1}{2}(\tilde{l}_i^{(2)} - \frac{1}{\sigma_\eta^2})(\eta_i - \tilde{\eta}_i)^2\right\} \\ &\quad \times \left[1 + \frac{1}{6}\tilde{l}_i^{(3)}(\eta_i - \tilde{\eta}_i)^3 + \frac{1}{24}\tilde{l}_i^{(4)}(\eta_i - \tilde{\eta}_i)^4\right] d\eta_i \end{aligned}$$

Now make a substitution $y(\eta_i) = \sqrt{\frac{1}{\sigma_\eta^2} - \tilde{l}_i^{(2)}}(\eta_i - \tilde{\eta}_i)$ and the equation becomes

$$\begin{aligned} L \quad &= \quad \prod_{i=1}^{m} \frac{\exp(\tilde{l}_i - \frac{\tilde{\eta}_i^2}{2\sigma_\eta^2})}{\sqrt{2\pi\sigma_\eta^2}} \frac{1}{\sqrt{\frac{1}{\sigma_\eta^2} - \tilde{l}_i^{(2)}}} \int_{-\infty}^{\infty} \exp\{-\frac{1}{2}y(\eta_i)^2\} \\ &\quad \times \left[1 + \frac{\tilde{l}_i^{(3)} y(\eta_i)^3}{6(\frac{1}{\sigma_\eta^2} - \tilde{l}_i^{(2)})^{3/2}} + \frac{\tilde{l}_i^{(4)} y(\eta_i)^4}{24(\frac{1}{\sigma_\eta^2} - \tilde{l}_i^{(2)})^2}\right] dy(\eta_i) \\ &= \quad \prod_{i=1}^{m} \frac{\exp(\tilde{l}_i - \frac{\tilde{\eta}_i^2}{2\sigma_\eta^2})}{\sqrt{2\pi\sigma_\eta^2}} \frac{\sigma_\eta}{\sqrt{1 - \sigma_\eta^2\tilde{l}_i^{(2)}}} \left[\sqrt{2\pi} + \frac{\sqrt{2\pi}\sigma_\eta^4\tilde{l}_i^{(4)}}{8(1 - \sigma_\eta^2\tilde{l}_i^{(2)})^2)}\right] \\ &= \quad \prod_{i=1}^{m} \frac{\exp(\tilde{l}_i - \frac{\tilde{\eta}_i^2}{2\sigma_\eta^2})}{\sqrt{1 - \sigma_\eta^2\tilde{l}_i^{(2)}}} \left[1 + \frac{\sigma_\eta^4\tilde{l}_i^{(4)}}{8(1 - \sigma_\eta^2\tilde{l}_i^{(2)})^2}\right] \\ &\approx \quad \prod_{i=1}^{m} \frac{1}{\sqrt{1 - \sigma_\eta^2\tilde{l}_i^{(2)}}} \exp\left\{\tilde{l}_i - \frac{\tilde{\eta}_i^2}{2\sigma_\eta^2} + \frac{\sigma_\eta^4\tilde{l}_i^{(4)}}{8(1 - \sigma_\eta^2\tilde{l}_i^{(2)})^2}\right\} \end{aligned}$$

This leads to an approximation of the log-likelihood

$$l \approx \sum_{i=1}^{m}\left[-\frac{1}{2}\log(1 - \sigma_\eta^2\tilde{l}_i^{(2)}) + \tilde{l}_i - \frac{\tilde{\eta}_i^2}{2\sigma_\eta^2}\right] + \frac{1}{8}\sum_{i=1}^{m}\left[\frac{\sigma_\eta^4\tilde{l}_i^{(4)}}{(1 - \sigma_\eta^2\tilde{l}_i^{(2)})^2}\right] \tag{2.9}$$

For the model with conditional log likelihood given by equation 2.3:

$$\tilde{l}_i^{(2)} = \frac{1}{\sigma^2\exp(2\kappa\tilde{\eta}_i)}\left[-r_i - 4\kappa\sum_{j=1}^{r_i}(y_{ij}-\mu-\tilde{\eta}_i) - 2\kappa^2\sum_{j=1}^{r_i}(y_{ij}-\mu-\tilde{\eta}_i)^2\right]$$

$$\tilde{l}_i^{(4)} = \frac{8\kappa^2}{\sigma^2\exp(2\kappa\tilde{\eta}_i)}\left[-3r_i - 4\kappa\sum_{j=1}^{r_i}(y_{ij}-\mu-\tilde{\eta}_i) - \kappa^2\sum_{j=1}^{r_i}(y_{ij}-\mu-\tilde{\eta}_i)^2\right];$$

substituting these into equation (2.9) then the approximation to the true log likelihood for our heterogeneity model is

$$\sum_{i=1}^{m}\left[\tilde{l}_i - \frac{\tilde{\eta}_i^2}{2\sigma_\eta^2} - \frac{1}{2}\log(\sigma^2\exp(2\kappa\tilde{\eta}_i) + \sigma_\eta^2(r_i + 4\kappa S_i + 2\kappa^2 Q_i)) + \frac{1}{2}\log(\sigma^2) + \kappa\tilde{\eta}_i\right]$$

$$+\sum_{i=1}^{m}\frac{\sigma^2\exp(2\kappa\tilde{\eta}_i)\sigma_\eta^4\kappa^2\left(-3r_i - 4\kappa S_i - \kappa^2 Q_i\right)}{\left(\sigma^2\exp(2\kappa\tilde{\eta}_i) + \sigma_\eta^2(r_i + 4\kappa S_i + 2\kappa^2 Q_i)\right)^2}, \tag{2.10}$$

where l_i is given by equation (2.3), $S_i = \sum_{j=1}^{r_i}(y_{ij}-\mu-\tilde{\eta}_i)$ and $Q_i = \sum_{j=1}^{r_i}(y_{ij}-\mu-\tilde{\eta}_i)^2$.

Note that the expression includes the parameter κ, therefore we can more readily explore behaviour under non-zero hypotheses for κ.

Expansion around zero

One can alternatively expand the Taylor series about $\eta = 0$ (Solomon & Cox (1992), rather than the m.l.e.. Then the approximation to the true likelihood takes the form

$$\prod_{i=1}^{m} L_i\{1-\sigma_\eta^2 l_{i0}^{(2)}\}^{-\frac{1}{2}}\exp\left\{\frac{\sigma_\eta^2 l_{i0}^{(1)2}}{2(1-\sigma_\eta^2 l_{i0}^{(2)})} + \frac{1}{2}(l_{i0}^{(3)}l_{i0}^{(1)} + \frac{1}{4}l_{i0}^{(4)})\sigma_\eta^4\right\}\{1+o(\sigma_\eta^4)\}. \tag{2.11}$$

Here, the higher order terms are included in the exponent. In this case, $l_{i0}^{(\omega)}$ is the ωth derivative of the conditional log-likelihood with respect to η_i, evaluated at $\eta_i = 0$ (i.e. $l_{i0}^{(\omega)} = \left.\frac{\partial^\omega l_i}{\partial\eta_i^\omega}\right|_{\eta_i=0}$). For the conditional log likelihood as given by 2.3, the Solomon-Cox approximation to the true log likelihood for the one-way random effects model is

$$\sum_{i=1}^{m}\Big[-\frac{r_i}{2}\log(2\pi\sigma^2)-\frac{Q_{i0}}{2\sigma^2}-\frac{1}{2}\log(\sigma^2+\sigma_\eta^2(r_i+4\kappa S_{i0}+2\kappa^2 Q_{i0}))+\frac{1}{2}\log(\sigma^2)$$

$$+\frac{\sigma_\eta^2(S_{i0}-\kappa Q_{i0}-r_i\kappa\sigma^2)^2}{2\sigma^2(\sigma^2+\sigma_\eta^2(r_i+4\kappa S_{i0}+2\kappa^2 Q_{i0}))}-\frac{\kappa^2\sigma_\eta^4}{\sigma^2}(3r_i+4\kappa S_{i0}+\kappa^2 Q_{i0})$$

$$+\frac{\sigma_\eta^4}{\sigma^4}(3\kappa r_i+6\kappa^2 S_{i0}+2\kappa^3 Q_{i0})(S_{i0}-\kappa Q_{i0}-r_i\kappa\sigma^2)\Big]$$

where $S_{i0}=\sum_{j=1}^{r_i}(y_{ij}-\mu)$ and $Q_{i0}=\sum_{j=1}^{r_i}(y_{ij}-\mu)^2$.

Breslow & Lin (1995) compare the performances of Laplace, Solomon-Cox and penalized quasi-likelihood approximations. For situations where the between-group variance component σ_η^2 is small they find there is not much to choose between the methods.

2.4 Approximate score tests

2.4.1 All parameters assumed known

The score test statistic **mevar2** is developed by assuming μ, σ^2, and σ_η^2 are known, and we treat the estimates $\tilde{\eta}_i$ as "metaparameters" which are by definition functions of μ, σ^2, and σ_η^2 and thus are also assumed to be known. From equation (2.10),

$$\frac{\partial l}{\partial \kappa} = \sum_{i=1}^{m}\Bigg[\frac{\partial \tilde{l}_i}{\partial \kappa}+\tilde{\eta}_i-\frac{\sigma^2\tilde{\eta}_i e^{2\kappa\tilde{\eta}_i}+\sigma_\eta^2(2S_i+2\kappa Q_i)}{\sigma^2 e^{2\kappa\tilde{\eta}_i}+\sigma_\eta^2(r_i+4\kappa S_i+4\kappa^2 Q_i)}$$

$$-\frac{2\sigma^2\sigma_\eta^4 e^{2\kappa\tilde{\eta}_i}\kappa(3r_i+6\kappa S_i+2\kappa^2 Q_i+\tilde{\eta}_i\kappa(3r_i+4\kappa S_i+\kappa^2 Q_i))}{\left(\sigma^2 e^{2\kappa\tilde{\eta}_i}+\sigma_\eta^2(r_i+4\kappa S_i+4\kappa^2 Q_i)\right)^2}$$

$$-\frac{4\sigma^2\sigma_\eta^4 e^{2\kappa\tilde{\eta}_i}\kappa^2(3r_i+4\kappa S_i+\kappa^2 Q_i)(\sigma^2\tilde{\eta}_i e^{2\kappa\tilde{\eta}_i}+\sigma_\eta^2(2S_i+2\kappa Q_i))}{\left(\sigma^2 e^{2\kappa\tilde{\eta}_i}+\sigma_\eta^2(r_i+4\kappa S_i+4\kappa^2 Q_i)\right)^3}\Bigg]$$

$$
\begin{aligned}
\frac{\partial^2 l}{\partial \kappa^2} &= \sum_{i=1}^{m}\left[\frac{\partial^2 \tilde{l}_i}{\partial \kappa^2} - \frac{2\sigma^2\tilde{\eta}_i^2 e^{2\kappa\tilde{\eta}_i} + 2\sigma_\eta^2 Q_i}{\sigma^2 e^{2\kappa\tilde{\eta}_i} + \sigma_\eta^2(r_i + 4\kappa S_i + 4\kappa^2 Q_i)}\right. \\
&\quad + \frac{(2\sigma^2\tilde{\eta}_i e^{2\kappa\tilde{\eta}_i} + \sigma_\eta^2(4S_i + 4\kappa Q_i))^2}{2\left(\sigma^2 e^{2\kappa\tilde{\eta}_i} + \sigma_\eta^2(r_i + 4\kappa S_i + 4\kappa^2 Q_i)\right)^2} \\
&\quad \left. - \frac{2\sigma^2\sigma_\eta^4 e^{2\kappa\tilde{\eta}_i}(3r_i + 12\kappa S_i + 6\kappa^2 Q_i + 6\tilde{\eta}_i r_i\kappa + 12\tilde{\eta}_i\kappa^2 S_i) + \tilde{\eta}_i\kappa(..)}{\left(\sigma^2 e^{2\kappa\tilde{\eta}_i} + \sigma_\eta^2(r_i + 4\kappa S_i + 4\kappa^2 Q_i)\right)^2}\right] \\
&\quad + \text{terms containing multiples of } \kappa.
\end{aligned}
$$

We first consider the simplest null hypothesis, $\kappa = 0$. Substitute $\kappa = 0$ into the above, and the score and information become:

$$
\frac{\partial l}{\partial \kappa}|_{\kappa=0} = \quad U_0 = \quad \sum_{i=1}^{m}\left[-r_i\tilde{\eta}_i + \frac{\tilde{\eta}_i\sum_j Q_i}{\sigma^2} + \tilde{\eta}_i - \frac{\tilde{\eta}_i\sigma^2 + 2\sigma_\eta^2 S_i}{\sigma^2 + r_i\sigma_\eta^2}\right]
$$

$$
-\frac{\partial^2 l}{\partial \kappa^2}|_{\kappa=0} = \quad \mathcal{I}_{11} = \quad \sum_{i=1}^{m}\left[\frac{2\tilde{\eta}_i^2 Q_i}{\sigma^2} + \frac{2\tilde{\eta}_i^2\sigma^2 + 2\sigma_\eta^2 Q_i}{\sigma^2 + r_i\sigma_\eta^2} + \frac{6r_i\sigma^2\sigma_\eta^4 - (\tilde{\eta}_i\sigma^2 + 2\sigma_\eta^2 S_i)^2}{(\sigma^2 + r_i\sigma_\eta^2)^2}\right]
$$

The score test statistic **mevar2** for testing the null hypothesis $H_0 : \kappa = 0$ is now given by

$$
\frac{\sum_{i=1}^{m}\left[-r_i\tilde{\eta}_i + \frac{\tilde{\eta}_i Q_i}{\sigma^2} + \tilde{\eta}_i - \frac{\tilde{\eta}_i\sigma^2 + 2\sigma_\eta^2 S_i}{\sigma^2 + r_i\sigma_\eta^2}\right]}{\sqrt{\sum_{i=1}^{m}\left[\frac{2\tilde{\eta}_i^2 Q_i}{\sigma^2} + \frac{2\tilde{\eta}_i^2\sigma^2 + 2\sigma_\eta^2 Q_i}{\sigma^2 + r_i\sigma_\eta^2} + \frac{6r_i\sigma^2\sigma_\eta^4 - (\tilde{\eta}_i\sigma^2 + 2\sigma_\eta^2 S_i)^2}{(\sigma^2 + r_i\sigma_\eta^2)^2}\right]}}. \tag{2.12}
$$

Under H_0, this should be approximately normally distributed with zero mean and standard deviation one.

When expanding about zero rather than the maximum functions $\tilde{\eta}_i$, $i = 1, ..., m$, U_0 and $\mathcal{I}_{11}$ reduce to

$$
U_0 = \sum_{i=1}^{m}\left[-\frac{\sigma_\eta^2\sigma^2 S_{i0}(2 + r_i) + S_{i0}Q_{i0}\sigma_\eta^2}{\sigma^2(\sigma^2 + r_i\sigma_\eta^2)} - \frac{2\sigma_\eta^4(S_{i0})^3}{\sigma^2(\sigma^2 + r_i\sigma_\eta^2)^2} + \frac{3\sigma_\eta^4 r_i S_{i0}}{\sigma^4}\right]
$$

$$\mathcal{I}_{11} = \sum_{i=1}^{m} \left[\frac{2\sigma^2\sigma_\eta^2 Q_{i0} - \sigma_\eta^2(Q_{i0} + r_i\sigma^2)^2}{\sigma^2(\sigma^2 + r_i\sigma_\eta^2)} - \frac{2\sigma_\eta^4(S_{i0})^2(3Q_{i0} + 4\sigma^2 + 4r_i\sigma^2)}{\sigma^2(\sigma^2 + r_i\sigma_\eta^2)^2} \right.$$

$$\left. + \frac{6r_i\sigma_\eta^4}{\sigma^2} - \frac{8\sigma_\eta^6(S_{i0})^4}{\sigma^2(\sigma^2 + r_i\sigma_\eta^2)^4} - \frac{\sigma_\eta^4}{\sigma^4}(12(S_{i0})^2 - 6r_i(Q_{i0} + r_i\sigma^2)) \right].$$

The advantage of this version is that the η_i do not need to be estimated. Unfortunately though, this form of the score test is not as tractable as that based on the Laplace expansion about the $\tilde{\eta}_i$. This is due to the information not being positive-definite for some combinations of variance components.

2.4.2 Using the estimation of parameters $(\sigma^2, \sigma_\eta^2, \mu)$

The score test statistic **mevar2a** follows. In practice, one is likely to estimate the mean and variance parameters using the data, and this additional variability should be accounted for in the test statistic thus improving its accuracy. When the remaining parameters are estimated, the null and alternative hypotheses are:

$$H_0: \quad \kappa = 0, \text{ with } \mu, \sigma^2 \text{ and } \sigma_\eta^2 \text{ unspecified.}$$

$$H_A: \quad \kappa \neq 0, \text{ with } \mu, \sigma^2 \text{ and } \sigma_\eta^2 \text{ unspecified.}$$

Unlike previous forms of the score test, the estimated residual or measurement error variance $\tilde{\sigma}^2$ may be very large in the case where $\kappa \neq 0$.

The test statistic becomes a χ_1^2 score test statistic with score given by

$$U = \frac{\partial l}{\partial \kappa}$$

and information obtained by defining a (4×4) matrix of second derivatives of the log-likelihood with respect to the four parameters and partitioned as

$$\mathbf{I} = \begin{bmatrix} \partial^2 l/\partial\kappa^2 & \mathbf{I}_{12} \\ \mathbf{I}_{12}^T & \mathbf{I}_{22} \end{bmatrix}$$

where $\mathbf{I}_{22}$ is the 3×3 matrix containing all the second derivatives with respect to all combinations of the three parameters σ^2, σ_η^2 and μ. Both the score and information are evaluated at $\kappa = 0$ and the maximum likelihood estimates of $\sigma^2 = \tilde{\sigma}^2$, $\sigma_\eta^2 = \tilde{\sigma_\eta^2}$, $\mu = \tilde{\mu}$.

The score test statistic for the null hypothesis $\kappa = 0$ becomes

$$\left.\frac{\partial l}{\partial \kappa}\right|_{\kappa=0} \cdot \left(\frac{|\mathbf{I}_{22}|}{|\mathbf{I}|}\right)^{\frac{1}{2}} \sim N(0,1). \tag{2.13}$$

For our heterogeneity model, l is given by the log likelihood approximation (2.10), and the terms from the 1×3 vector $\mathbf{I}_{12}$:

$$\left.\frac{\partial^2 l}{\partial \kappa \partial \sigma^2}\right|_{\kappa=0} = \sum_{i=1}^{m} \left[-\frac{\tilde{\eta}_i Q_i}{\sigma^4} - \frac{\tilde{\eta}_i}{\sigma^2 + r_i \sigma_\eta^2} + \frac{\sigma^2 \tilde{\eta}_i + 2S_i \sigma_\eta^2}{(\sigma^2 + r_i \sigma_\eta^2)^2}\right]$$

$$\left.\frac{\partial^2 l}{\partial \kappa \partial \sigma_\eta^2}\right|_{\kappa=0} = \sum_{i=1}^{m} \left[-\frac{2S_i}{\sigma^2 + r_i \sigma_\eta^2} + \frac{r_i(\tilde{\eta}_i \sigma^2 + 2\sigma_\eta^2 S_i)}{(\sigma^2 + r_i \sigma_\eta^2)^2}\right]$$

$$\left.\frac{\partial^2 l}{\partial \kappa \partial \mu}\right|_{\kappa=0} = \sum_{i=1}^{m} \left[\frac{-2\tilde{\eta}_i S_i}{\sigma^2} + \frac{2\sigma_\eta^2 r_i}{\sigma^2 + r_i \sigma_\eta^2}\right]$$

and from the matrix $\mathbf{I}_{22}$:

$$\left.\frac{\partial^2 l}{\partial \sigma^2 \partial \sigma^2}\right|_{\kappa=0} = \sum_{i=1}^{m} \left[\frac{r_i - 1}{2\sigma^4} + \frac{1}{2(\sigma^2 + r_i \sigma_\eta^2)^2} - \frac{Q_i}{\sigma^6}\right]$$

$$\left.\frac{\partial^2 l}{\partial \sigma_\eta^2 \partial \sigma_\eta^2}\right|_{\kappa=0} = \sum_{i=1}^{m} \left[-\frac{\tilde{\eta}_i^2}{\sigma_\eta^6} + \frac{r_i^2}{2(\sigma^2 + r_i \sigma_\eta^2)^2}\right]$$

$$\left.\frac{\partial^2 l}{\partial \sigma^2 \partial \sigma_\eta^2}\right|_{\kappa=0} = \sum_{i=1}^{m} \left[\frac{r_i}{2(\sigma^2 + r_i \sigma_\eta^2)^2}\right]$$

$$\left.\frac{\partial^2 l}{\partial \mu^2}\right|_{\kappa=0} = \sum_{i=1}^{m} \left[\frac{-r_i}{\sigma^2}\right]$$

$$\left.\frac{\partial^2 l}{\partial \sigma^2 \partial \mu}\right|_{\kappa=0} = \sum_{i=1}^{m} \left[-\frac{(y_{i\cdot} - r_i \mu)}{\sigma^4}\right]$$

$$\left.\frac{\partial^2 l}{\partial \sigma_\eta^2 \partial \mu}\right|_{\kappa=0} = 0$$

Substituting these into equation (2.13) gives the score test statistic **mevar2a**. The performance of this version of the test statistic is examined in the next Section.

mevar2 and **mevar2a** contain terms $\tilde{\eta}_i$ which we have referred to as "meta parameters". In a general model these will need further investigation, however here we are considering a score test, where under the null hypothesis $\kappa = 0$ we have a linear model. Therefore in this case we are able to estimate $\tilde{\eta}_i$:

$$\begin{aligned}
\tilde{\eta}_i &= \tilde{l}_i^{(1)} \tilde{\sigma_\eta^2} \\
&= \frac{\sum_{j=1}^{r_i}(y_{ij} - \tilde{\mu})\tilde{\sigma_\eta^2}}{\tilde{\sigma}^2 + r_i \tilde{\sigma_\eta^2}} \\
&= \frac{r_i \tilde{\sigma_\eta^2}(\bar{y}_{i.} - \tilde{\mu})}{\tilde{\sigma}^2 + r_i \tilde{\sigma_\eta^2}}.
\end{aligned}$$

From here we can notice that when $r_i\tilde{\sigma_\eta^2} >> \tilde{\sigma}^2$ the estimates $\tilde{\eta}_i$ are essentially $(\bar{y}_{i.} - \tilde{\mu})$. When $\tilde{\sigma}^2 >> r_i\tilde{\sigma_\eta^2}$ then the $\tilde{\eta}_i$ are shrunken versions of $(\bar{y}_{i.} - \tilde{\mu})$. This feature is picked up further in next section.

2.5 Simulations

Data were simulated from model (2.1) on page 7 under the null hypothesis with $\kappa = 0$ and various variance component combinations ranging from $0.01 \leq \sigma^2, \sigma_\eta^2 \leq 1$, $\mu = 1$, $m = 20$ individuals and $r = 20$ replications.

The simulated data were used to calculate **mevar2**, equation (2.12), **mevar2a**, equation 2.13, using estimated variance components $\tilde{\sigma}^2 = \sum_{ij}(y_{ij} - \bar{y}_{i.})^2/m(r-1)$ and $\tilde{\sigma}_\eta^2 = \sum_{i=1}^{m}(\bar{y}_{i.} - \bar{y}_{..})^2/(m-1) - \tilde{\sigma}^2/r$, i.e. the estimates from a standard normal linear model and **mevar1**, Section 2.2. This procedure was repeated 100 times to give the three score test statistics for 100 simulated data sets. Table 2.1 shows a typical sample of the sample means and variances from these simulated data. The number of rejected points (5% level) from these 100 values when compared to the standard normal distribution were noted and the results given in figure 2.2.

This selected sample is typical of the overall results obtained from all possible combinations of parameters. Generally, the distributions of the statistic under $H_0 : \kappa = 0$ are observed to be approximately standard normal, although there is a tendency in the approximation versions for the standard deviation to be slightly underestimated. It also appears that the test works better when $\sigma_\eta^2 \geq \sigma^2$, i.e. the between variance component large relative to the random error. The test seems to behave differently when σ^2 is very much larger than σ_η^2. This is because the error term, $\sigma^2 \exp(2\kappa\eta_i)$ is very much larger then the mean, $\mu + \eta_i$ for all values of η_i, so it is hard to detect a true mean/variance relationship.

Tables 2.4, 2.5 and 2.6 show the results of the null hypothesis test $H_0 : \kappa = 0$ when κ departs from 0. The number of rejected values appears unstable when $\sigma^2 > \sigma_\eta^2$ (as noted for $\kappa = 0$ simulations above). The rejection is quite strong for other variance component combinations. Notice however, particularly in figure 2.6, that the test becomes unstable as the magnitude of the variance components increases. This instability is dependent on the combination of κ and σ_η^2.

To investigate what happens as κ is changed keeping σ_η^2 and σ^2 fixed, data were simulated using $\sigma^2 = \sigma_\eta^2 = 1$ and the results are shown in table 2.3. For this set of parameter values, the score test is definitely detecting when κ departs from 0. Also, as κ departs further from zero the mean becomes larger (so a P value would be more significant), but when κ is around ± 0.4 away from 0, the standard deviation starts to increase enormously and the significance level may start to drop, the exact test appears to be more sensitive in these areas.

Figures 2.1 and 2.2 show typical representations when using the same simulated data to find the test statistics **mevar2a**, **mevar2** and **mevar1**. Figure 2.1

shows data simulated with $\kappa = 0$ with **mevar2a** as the solid line. It shows that **mevar2** and **mevar1** are very similar under the null hypothesis. The distributions are roughly symmetric around 0 for all versions of the test statistic. The plots are very similar between versions of the score test. This is typical for all other plots with different parameters.

Figure 2.2 shows the distributions of the score test statistics for simulated data with $\kappa = 1$. There is now a non-zero mean and slight skewness indicating a distribution which is not normal. There is also present a change in standard deviation between versions. This is very apparent for $\kappa = 1$, but as κ gets closer to zero the standard deviations all approach 1 and the differences between the versions of the statistic become less obvious.

It may be reasonable to assume that the test **mevar2a** may be different to the others since its form is composed of a larger information matrix, therefore let's compare the differences in **mevar2** and **mevar1**. In figure 2.3, both **mevar1** and **mevar2** were found for the same set of simulated data with parameters $m = 30, r = 10, \mu = 5, \kappa = 0, \sigma_\eta^2 = 2, \sigma^2 = 1$. After 100 repetitions the mean and standard deviation for the exact statistic were 0.009,1.114 and for the approximate statistic were 0.010,1.073. These results together with the plots show that both distributions are very similar.

In table 2.3 **mevar2** appears to be more stable than the other two. This distribution is represented graphically by figures 2.4 and 2.5. Figure 2.4 shows that the distribution of **mevar2** is roughly normal for simulated data with $\kappa \neq 0$. As κ increases the distribution indicates an increase in standard deviation. Figure 2.5 shows that on the same scale, the means are increasing and the standard deviations are staying roughly equal for $\kappa = 0.2$, however for $\kappa = 0.6$ the distribution starts to skew and become non-normal with a large standard deviation.

The combination of $\sigma_\eta^2 = 1$ and $\sigma^2 = 1$ seems to give rather odd results in figure

2.6 and table 2.3, this was examined further. Figure 2.6 shows the distributions for $\kappa = 0.2$ and $\kappa = 0.6$ with variance components $\sigma_\eta^2 = \sigma^2 = 1$. These plots show how the distributions change as κ becomes larger. **mevar1** appears to have the greatest change in standard deviation, showing "skewness" for this combination of large κ and large variance components. It also appears that **mevar2a** is behaving better than the others, which was not seen by first examination of the table.

2.6 Power of the test

The power of the test is defined as $P(H_A|H_0 \text{ false })$, the probability of rejecting the null hypothesis given that H_0 is false, i.e. given that $\kappa = 0$. Given $\kappa \neq 0$ and the score test statistic S with distribution $N(\mu_A, \sigma_A)$, for a 5% rejection region

$$\text{Power} = 1 - P(-1.96 < S < 1.96).$$

The power was calculated using sample distributions similar to those given in table 2.3. For each value of κ, 100 repetitions of simulated data with $m = 30, r = 10, \mu = 0$ and $\sigma^2 = 1$ were used with two different values for σ_η^2. Figure 2.7 shows that for simulated data with $\sigma_\eta^2 = 1$, power is 1 around $\kappa = 0.2$ and for simulated data with $\sigma_\eta^2 = 0.5$ power is 1 around $\kappa = 0.3$. For larger variance σ_η^2, the score test will reject the null hypothesis for smaller values of κ.

The drop after about $\kappa = 0.5$ is due to the increase in standard deviation around this value. The exact test, shown by the dashed line, appears different from the two approximations, it rises more sharply and then levels off roughly, but appears not to decrease again like the approximate tests.

We now wish to consider what will happen to the score test statistic when σ_η^2 becomes very small. All three expressions contain terms involving the data and approximations of the fitted effects from a linear model, and the $\tilde{\eta}_i$ terms. Firstly note that as σ_η^2 approaches zero $\tilde{\eta}_i = \tilde{l}_i^{(1)} \sigma_\eta^2$ also approaches zero. Also, as $\sigma_\eta^2 \to 0$, η_i

will tend towards its expected value of zero, therefore the model $y_{ij} = \mu + \eta_i + \varepsilon_{ij} e^{\kappa \eta_i}$ will become $y_{ij} = \mu + \varepsilon_{ij}$, a model which is independent of κ, therefore estimates of σ^2 and μ will be unaffected by changes in κ.

This implies that the score test statistic for the null hypothesis $\kappa = 0$ for data with very small σ_η^2 will remain constant independent of the value of κ and therefore the power for very small σ_η^2 will always be small. κ would need to be very large to show any significance at all against a very small σ_η^2, a κ this large would not realistically fit the model proposed here.

Table 2.1: Means and standard deviations of the score statistic under the null hypothesis $H_0 : \kappa = 0$ for simulated data with $\kappa = 0$, with variance components ranging from 0.001 to 5.

	mevar2		mevar2a		mevar1	
$\frac{\sigma_\eta^2}{\sigma^2}$	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s
0.1	-0.065	0.728	-0.038	0.629	-0.097	1.077
0.01	0.034	0.471	0.093	0.462	0.158	3.332
10	-0.133	1.022	-0.008	0.962	0.118	1.176
0.1	0.013	0.710	0.074	0.705	0.107	1.131
100	0.095	0.970	0.158	1.031	-0.064	1.063
10	-0.110	0.890	-0.096	1.069	0.046	1.058
1	0.074	0.952	0.037	0.969	-0.086	0.971
0.5	-0.121	0.955	-0.005	0.817	0.039	0.985
2	0.108	0.925	-0.081	0.947	0.107	0.989
0.5	-0.039	0.826	0.043	0.918	-0.118	1.045
1	-0.038	1.014	0.099	0.983	0.037	1.009
5	-0.070	1.057	-0.144	0.887	-0.045	1.038
2.5	0.076	0.905	0.118	1.028	0.017	0.937

σ^2	Test	$\sigma_\eta^2 = 0.01$	0.05	0.1	0.25	0.5	1
1	1	11	11	6	5	5	5
	2	0	1	1	1	3	2
	2a	0	1	1	1	3	2
0.5	1	13	6	7	6	4	6
	2	0	1	1	5	3	5
	2a	0	1	1	5	3	6
0.25	1	9	8	2	3	4	4
	2	0	2	1	2	4	4
	2a	0	2	1	2	4	4
0.1	1	6	4	3	3	5	6
	2	0	3	3	3	5	6
	2a	0	3	3	3	5	6
0.01	1	3	5	5	4	3	6
	2	3	5	5	4	3	6
	2a	3	5	5	4	3	6

Table 2.2: Number of rejected points from 100 repetitions of the score test statistic for simulated data with $\kappa = 0$

Table 2.3: Means and standard deviations of the score statistic under the null hypothesis $H_0 : \kappa = 0$ for simulated data with variance components $\sigma_\eta^2 = \sigma^2 = 1$ and various κ.

	mevar2		mevar2a		mevar1	
κ	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s
-1	-5.35	12.56	-8.31	6.47	-12.98	14.07
-0.5	-6.93	1.06	-8.82	2.03	-11.86	7.49
-0.4	-6.32	0.99	-7.72	1.71	-9.66	6.80
-0.3	-5.02	1.05	-5.66	1.47	-6.19	1.53
-0.2	-3.68	1.09	-3.89	1.11	-4.09	0.93
-0.1	-1.99	0.95	-1.99	1.01	-2.17	0.82
-0.05	-0.96	0.90	-1.04	0.933	-0.99	0.98
0	-0.19	1.03	0.14	1.00	-0.04	0.93
0.05	1.09	0.81	0.96	0.99	1.01	1.05
0.1	1.78	0.98	1.99	0.98	2.13	0.89
0.2	3.69	0.94	3.81	1.23	3.94	1.04
0.3	5.18	0.78	6.06	1.47	6.05	1.30
0.4	6.11	1.07	7.31	1.46	19.95	13.67
0.5	6.93	0.99	8.70	2.14	11.42	4.65
1	6.15	3.94	8.97	7.29	18.75	25.37

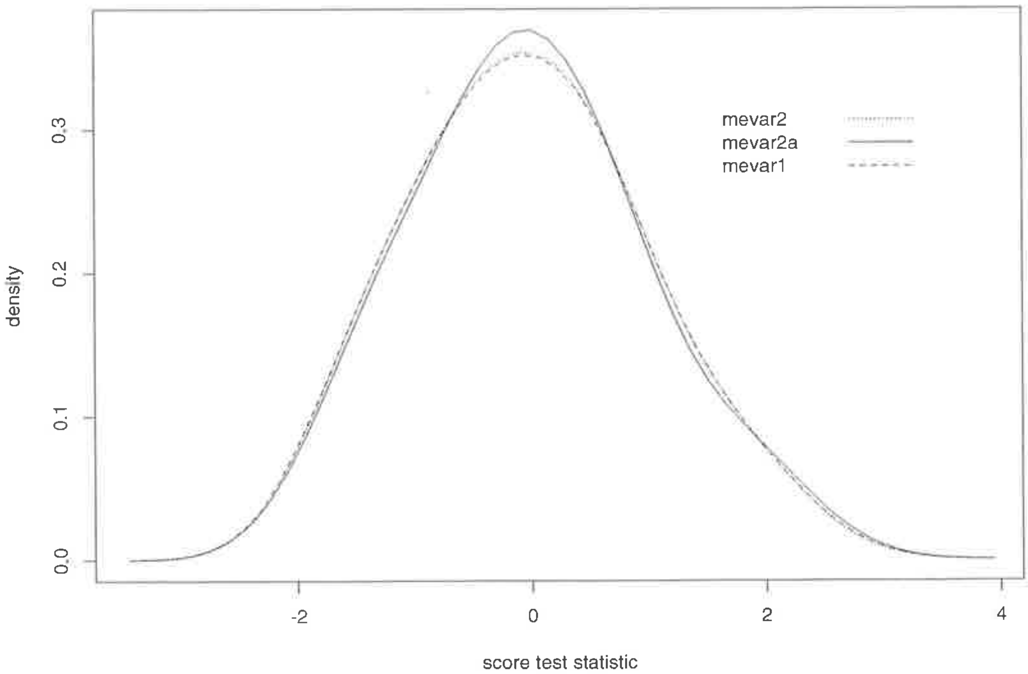


Figure 2.1: Distributions of the score test statistic for simulated data with $\kappa = 0$, $\sigma_\eta^2 = 0.5, \sigma^2 = 0.1$, 100 repetitions

Table 2.4: Means and standard deviations of the score statistic under the null hypothesis $H_0 : \kappa = 0$ for simulated data with $\kappa = 0.5$, with variance components ranging from 0.01 to 0.5

		mevar2		mevar2a		mevar1	
σ_η^2	σ^2	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s
0.01	0.01	1.134	0.994	0.984	0.966	1.053	1.042
0.01	0.1	0.598	0.610	0.476	0.717	0.758	1.215
0.1	0.01	3.524	1.200	3.579	1.064	3.422	0.931
0.1	0.1	3.203	0.902	3.132	1.276	3.270	1.234
0.25	0.5	4.637	1.149	4.277	1.335	5.254	1.464
0.5	0.25	7.446	1.695	7.563	1.662	6.924	1.149
0.5	0.5	6.716	1.250	6.776	1.624	7.412	2.244
1	0.5	10.134	2.532	10.227	1.963	9.494	2.036
1	1	9.578	2.185	8.882	1.845	11.028	4.287

σ^2	Test	$\sigma_\eta^2 = 0.01$	0.05	0.1	0.25	0.5	1
1	1	25	39	67	90	94	93
	2	0	7	36	88	100	100
	2a	0	3	32	93	99	100
0.5	1	13	59	59	97	99	88
	2	0	16	62	99	100	100
	2a	0	17	60	96	100	100
0.25	1	11	62	83	99	100	100
	2	0	30	70	100	100	100
	2a	0	30	71	99	100	100
0.1	1	41	59	89	100	100	100
	2	2	48	83	99	100	100
	2a	3	52	87	99	100	100
0.01	1	17	66	90	100	100	100
	2	13	67	94	100	100	100
	2a	13	71	92	99	100	100

Table 2.5: Number of rejected points from 100 repetitions of the score test statistic for simulated data with $\kappa = 0.5$

σ^2	Test	σ_η^2 = 0.01	0.05	0.1	0.25	0.5	1
1	1	26	73	83	89	88	90
	2	0	28	63	96	97	85
	2a	0	21	73	96	94	89
0.5	1	30	79	90	93	92	92
	2	0	49	90	100	100	97
	2a	1	58	94	99	100	93
0.25	1	51	88	95	93	94	94
	2	1	85	99	100	100	100
	2a	2	80	99	100	100	100
0.1	1	59	98	95	99	99	99
	2	10	94	99	100	100	100
	2a	14	93	99	100	100	100
0.01	1	55	99	100	100	100	100
	2	49	100	100	100	100	100
	2a	48	99	100	100	100	100

Table 2.6: Number of rejected points from 100 repetitions of the score test statistic for simulated data with $\kappa = 1$

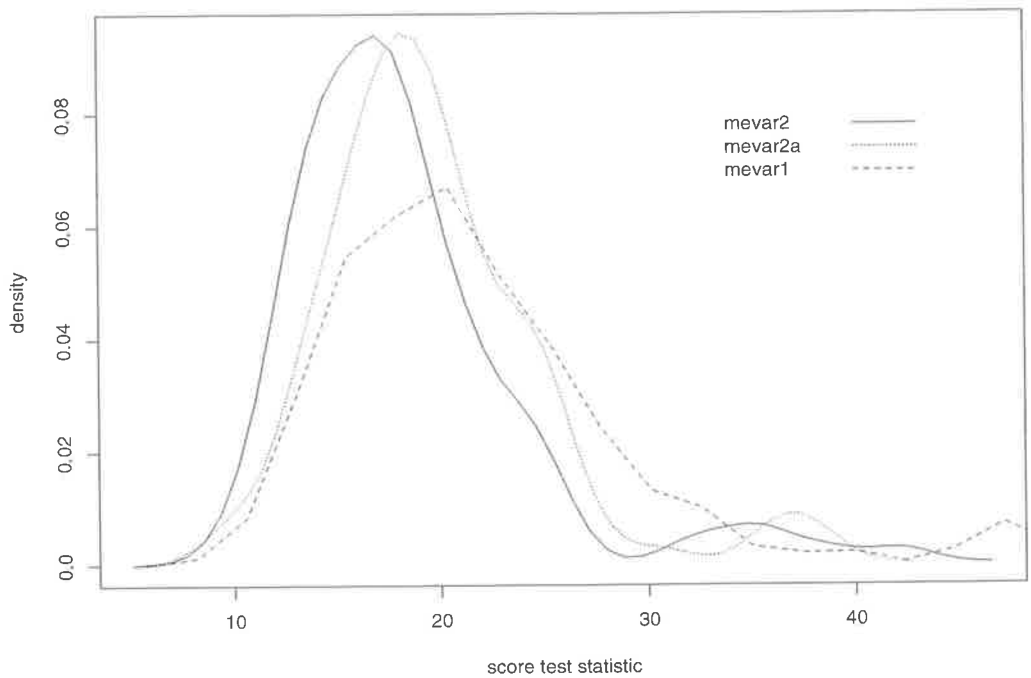


Figure 2.2: Distributions of the score test statistic for simulated data with $\kappa = 1$, $\sigma_\eta^2 = 0.5, \sigma^2 = 0.1$, 100 repetitions

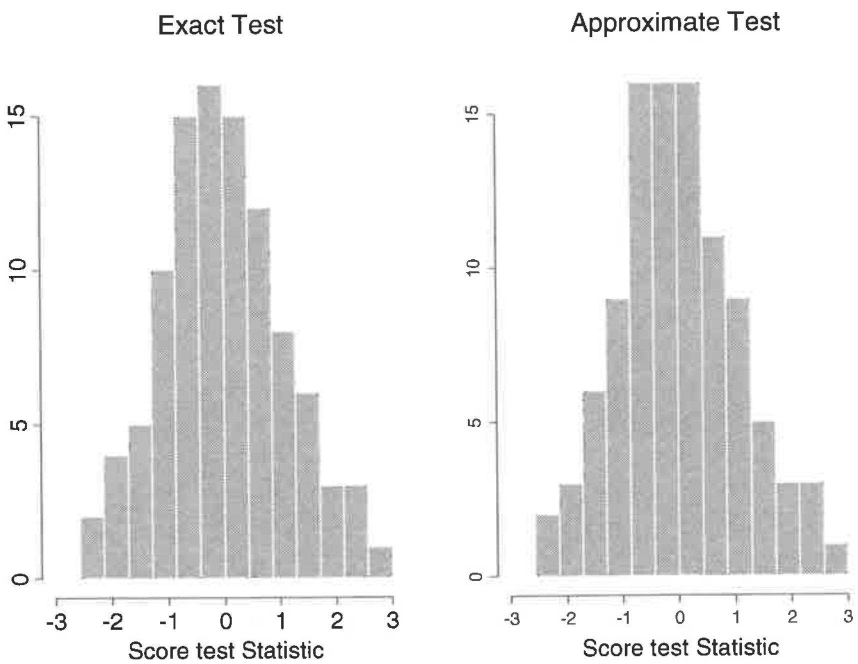


Figure 2.3: Comparisons of **mevar1** and **mevar2** distributions data simulated with $\kappa = 0, \sigma_\eta^2 = 2, \sigma^2 = 1$.

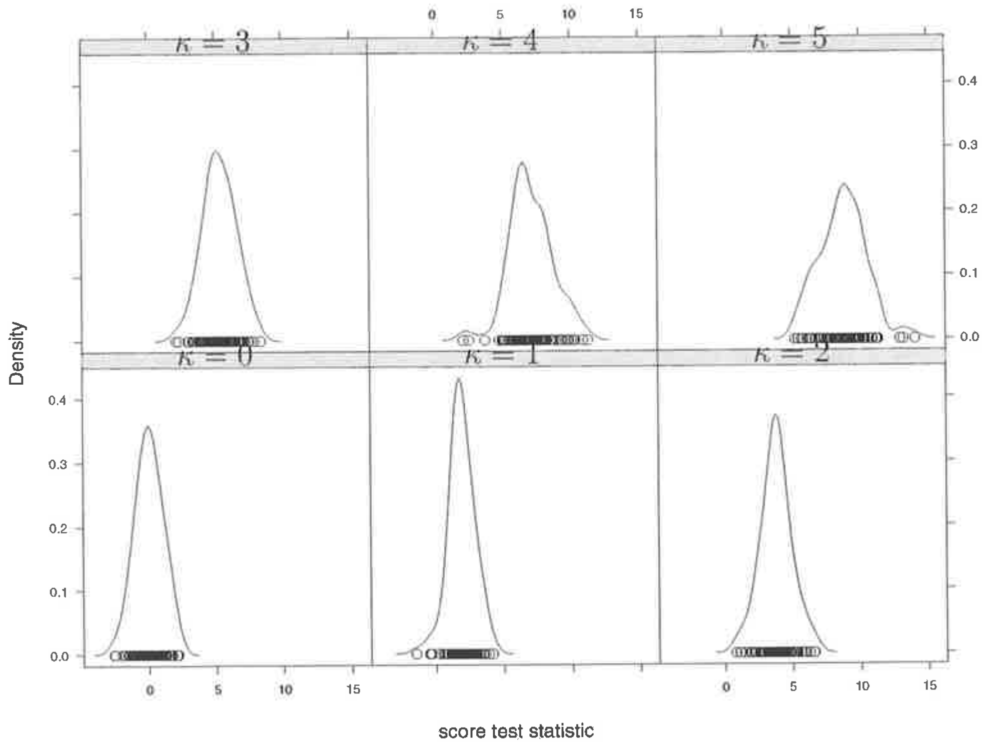


Figure 2.4: Distributions of the score test statistic **mevar2** $\sigma_\eta^2 = 0.5, \sigma^2 = 0.1, \mu = 10$ and κ ranging from 0 to 0.5.

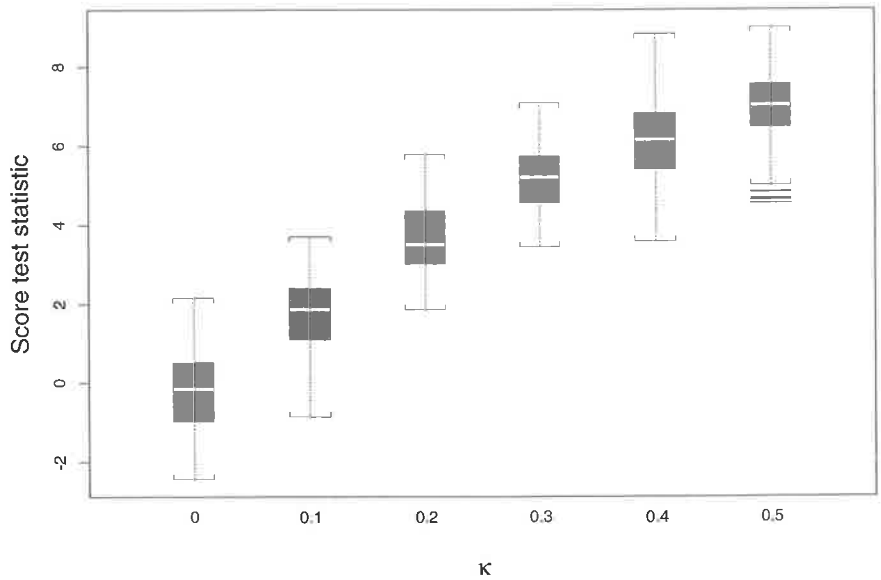


Figure 2.5: Distributions of the score test statistic **mevar2** $\sigma_\eta^2 = \sigma^2 = 1$ and $\mu = 10$.

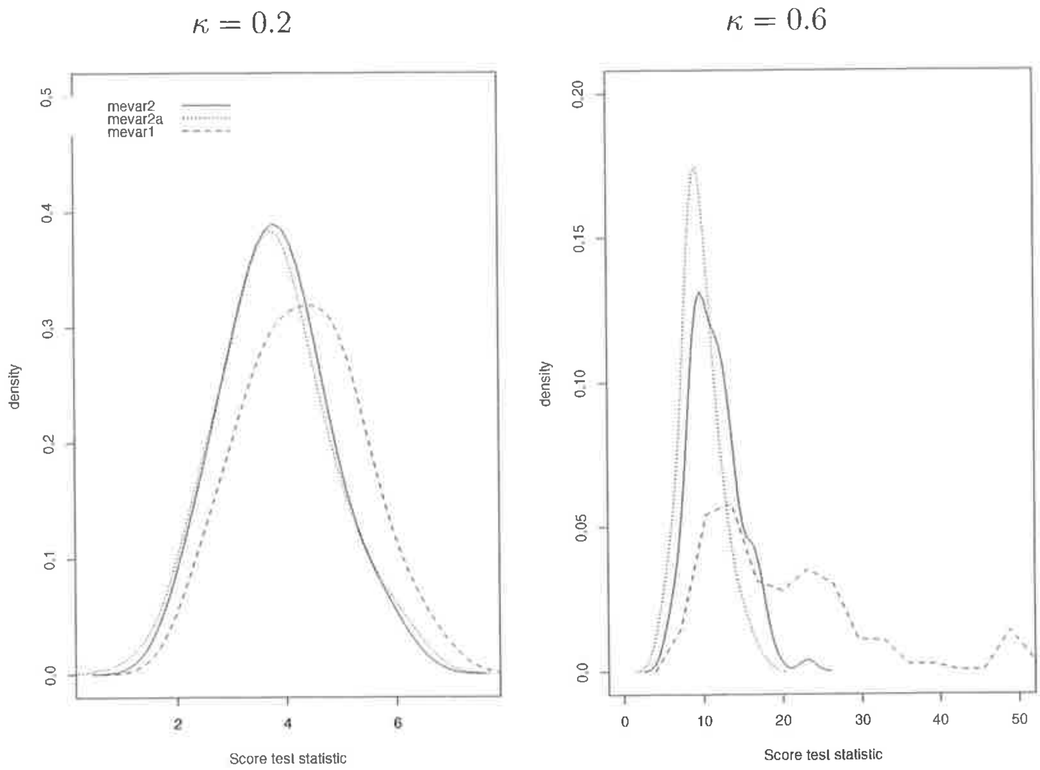
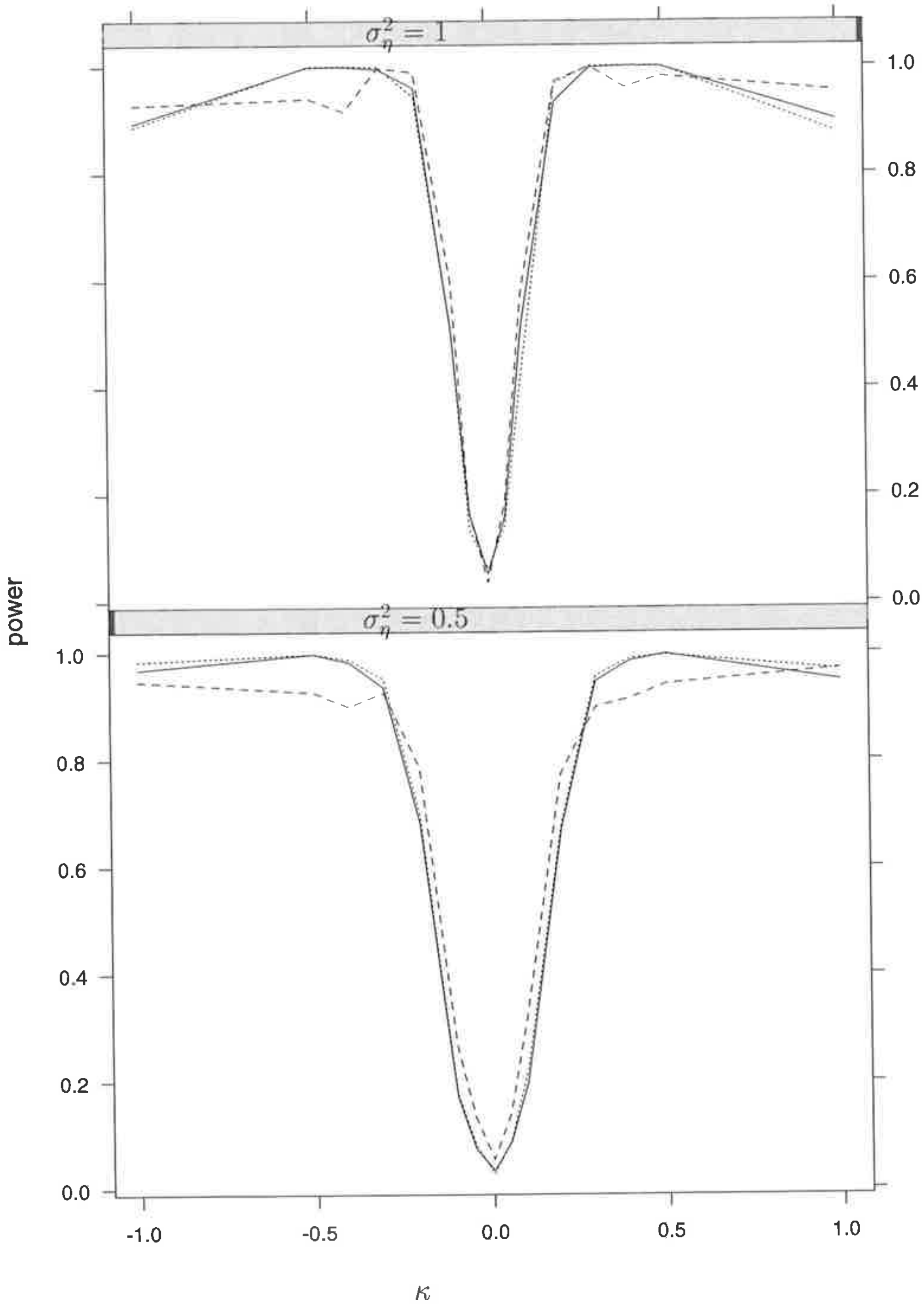


Figure 2.6: Distributions of the score test statistic with simulated data with $\kappa = 0.2, 0.6, \sigma_\eta^2 = \sigma^2 = 1$.

Figure 2.7: Power of the score test



2.7 Examples

2.7.1 CD4 Count Data

Our data come from the San Fransisco Men's Health Study (SFMHS) as described by Winkelstein et al. (1987). Recent papers which have considered these data include Bacchetti & Moss (1989), Brookmeyer & Gail (1988), Boscardin et al. (1998), DeGruttola et al. (1991), Lange et al. (1992), Pawitan & Self (1993) and Satten & Longini (1994). Bacchetti & Moss (1989) and DeGruttola et al. (1991), for example, use the SFMHS data to model times to infection using a parametric linear growth curve model with random effects. Satten & Longini (1994) studied the probability density function of the time to HIV infection.

The data consist of 427 HIV^+ male patients in San Fransisco who where scheduled to visit a clinic every 6 months, however some patients visited over irregular periods of time. Each patient was enrolled in the programme sometime between June 1984 and January 1985 and were followed up to September 1992, some patients were HIV^- some were already HIV^+. The data we are considering here is only concerned with the measurements taken after the subject was declared HIV^+. At each visit a blood sample was taken and from this a CD4 cell count was determined. The virus that causes AIDS is accompanied by a decline in the CD4 cell count (measured in counts per microlitre) which are white blood cells associated with the immune system. The first HIV^+ observations appeared between May 1984 and May 1990. The dataset contains the five baseline covariates for each patient: patient ID number, month and year of the first HIV^+ observation, age of the subject at entry, and the number of visits, and with each visit is associated a triplet consisting of time, CD4 count and treatment indicator, and finally an indicator of AIDS status (-10 if the patient developed full AIDS).

Early analysis on the entire data set consisted of plotting the distribution of age, starting CD4 cell count and month of entry, these are shown by the kernel density estimates in figure 2.8. The starting CD4 cell counts are those counts taken at the time the patients were first diagnosed as HIV^+. The mean starting CD4 count is $687.28mm^3$ with a standard deviation of $292.66mm^3$. The ages of patients range from 25 to 54 years with an average of 38.61 years and standard deviation 8.4 years. The month of entry ranges from 5 (May 1984) to 77 (May 1990), this is an indication of when each patient was diagnosed as HIV^+. The month of entry histogram shows that 93 percent of patients were declared HIV^+ in the first 24 months, with 78 percent in the first 12 months.

The data were split into 2 cohorts. The "AIDS" cohort contains those patients who contracted full AIDS (indicated by the -10 as mentioned above) and then were removed from the study, and "NoAIDS" contains all patients who never contracted full AIDS throughout the duration of the trial.

Figure 2.9 shows plots of the age distributions and the distributions of CD4 cell count at entry for the AIDS and NoAIDS data separately. The mean starting CD4 cell count for the AIDS patients is lower than the starting CD4 cell count for the NoAIDS patients (545.91 versus 772.54), as expected since the patients who developed AIDS are sicker with poorer immune function and therefore lower CD4. The NoAIDS cases are more variable (standard deviation is 239.51 for AIDS, 289.13 for NoAIDS). The age distributions do not differ very much, the mean age for AIDS is 35.91, versus 34.27 for NoAIDS.

Figure 2.10 shows the trajectories for the four combinations of AIDS and treatment factors. It shows that the patients' CD4 counts decrease more rapidly when no treatment is given, in particular, the plot representing Untreated AIDS decreases at a very rapid rate. We can also see from these plots that the starting CD4 cell count for the "AIDS" patients were overall lower than the "NoAIDS" group. The entire data set contains unequal numbers of observations for each patient either because

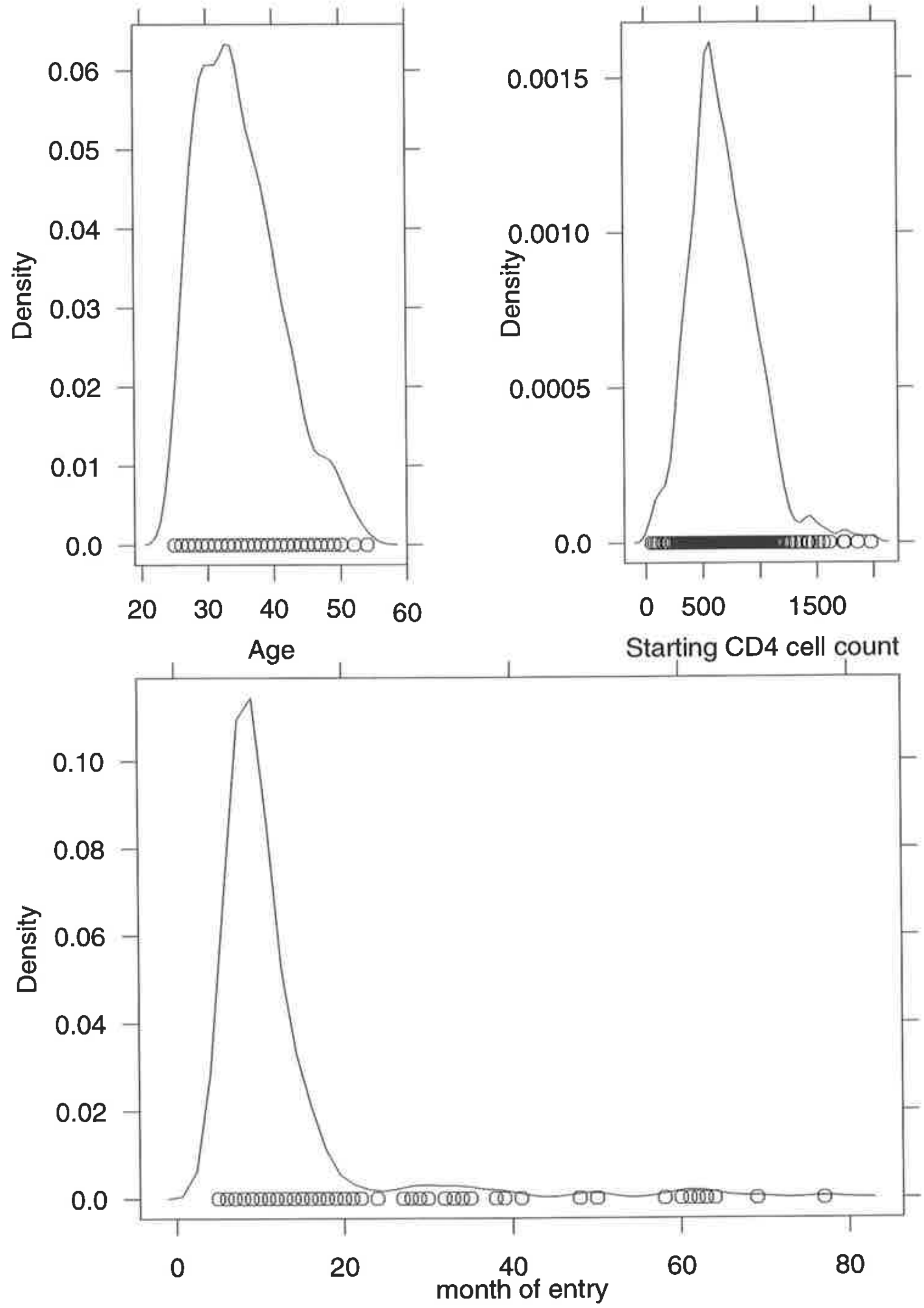


Figure 2.8: Kernel density estimates of whole CD4 data

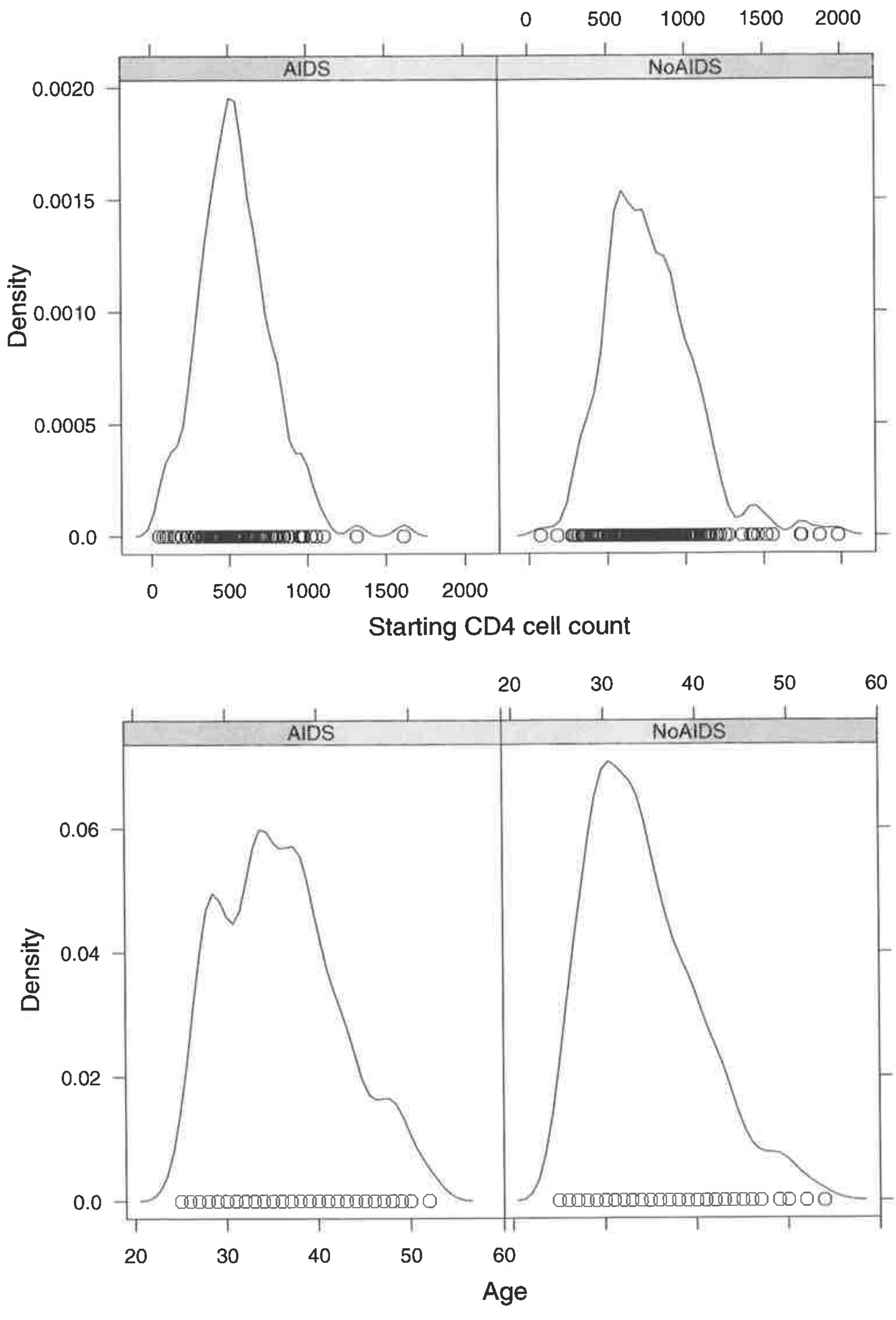


Figure 2.9: Kernel density estimates of CD4 data showing AIDS and NoAIDS separately

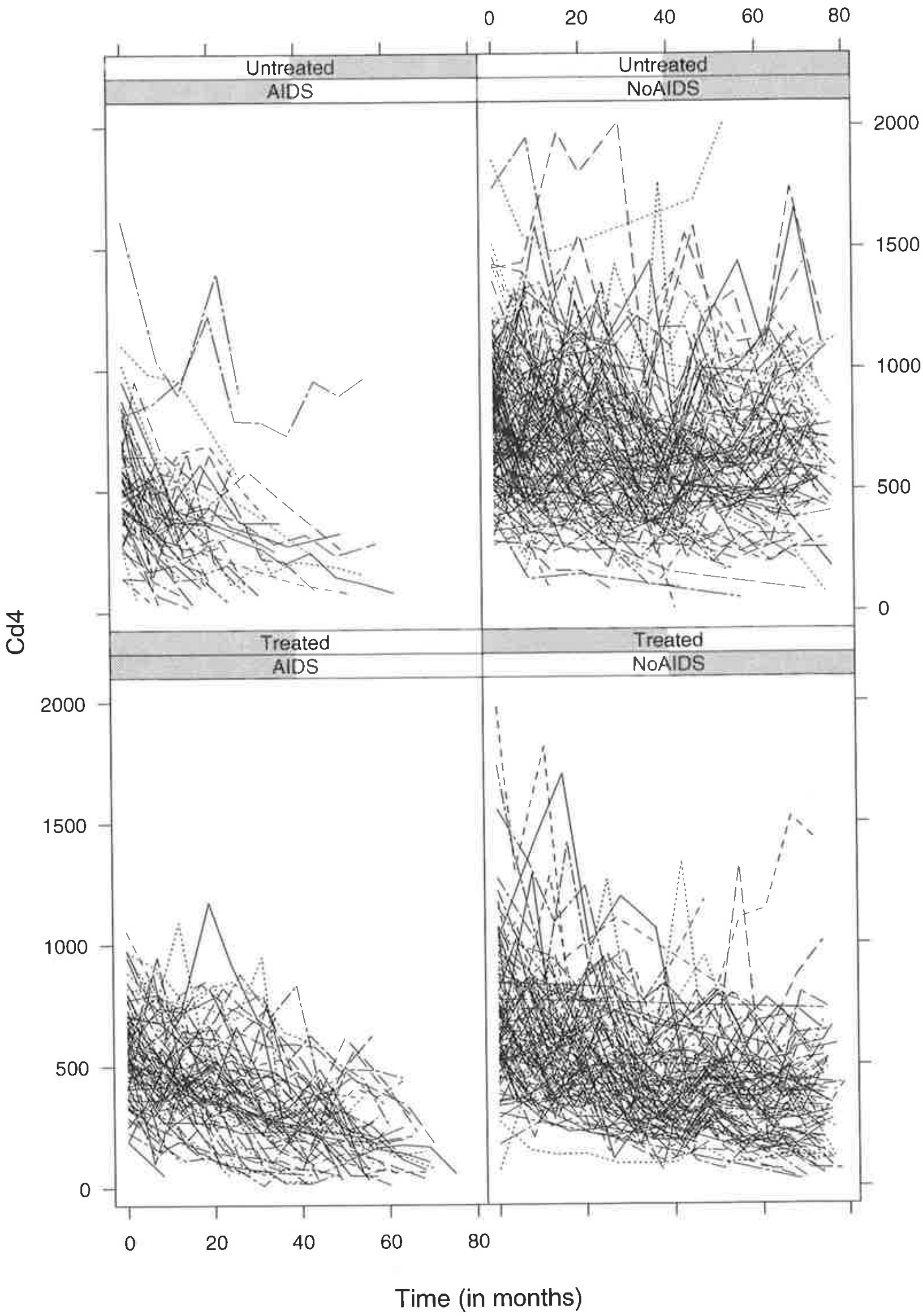


Figure 2.10: Individual trajectories of CD4 cell counts over time t in months

the patient developed full AIDS, or because the patient failed to appear regularly. To enable a more complete study, two balanced cohorts were chosen.

The selected Cohorts

From the two groups (AIDS and NoAIDS above) cohorts of 38 AIDS patients and 141 non-AIDS patients were analysed. Each cohort was selected as those patients who attended the clinic 9 or more times regularly throughout the course of the trial, the data from the first 9 visits were used as balanced data sets for the analyses. The analyses conducted included investigating transformations of the CD4 counts and simple regression analysis. The procedure for selecting the cohorts is outlined below.

Firstly we took the number of visits for each subject (column 5), and selected those who had 9 or more observations. The intervals between visits by each patient was examined. A good cohort would consist of those patients whose data was measured at regular intervals. Although the aim of the trial was to have regular six monthly visits, there were patients who were quite irregular, a histogram of intervals between visits for the total data set showed a large peak at 6 months. It was reasonable to assume there were enough 6 monthly intervals to create a sizable cohort. The intervals were firstly examined for all patients from the AIDS group who had nine or more visits. It could be seen that there were about 50 patients who had 9 or more visits. After a patient was diagnosed AIDS (or perhaps died) labels of '-10' are used repeatedly to indicate a visit occurred but no reasonable measurement could be determined, the CD4 data from the AIDS cohort was then examined to ensure that all the data was count data. It was found that only 38 of these patients had 9 or more CD4 cell count measurements. These 38 patients were then analysed and shown to be a good cohort for the AIDS patients.

A similar procedure was carried out for the group of patients who didn't develop

AIDS. In this case there was no need to look for '-10' labels.

Transformations

Let the CD4 counts be represented by Y. The transformations considered were

$$Y^{\lambda}, \lambda = -1, -\frac{1}{2}, 0, \frac{1}{2}, 1. \tag{2.14}$$

For each subject, the mean and standard deviation of their (9 or more) readings was calculated. Figures 2.11 and 2.12 show these means versus standard deviations of CD4 cell counts for each of these transformations for the AIDS and NoAIDS groups respectively. It was found that Y, Y^{-1} and $Y^{-\frac{1}{2}}$ all had increasing mean-variance relationships, $\log(Y)$ appeared to have a slight decreasing relationship, but mostly random, and $\sqrt{Y}$ had a random looking relationship.

Because of the assumptions for standard normal theory linear models having a random mean variance relationship, $\log(Y)$ and $\sqrt{Y}$ were chosen as the 'best' transformations and used in further investigations. Figure 2.13 shows individual CD4 cell count trajectories for the two cohorts over all transformations. The raw CD4 cell counts declined over time as the mens' immune systems were increasingly compromised. Overall, the counts are lower and less variable for the AIDS cohort. The relationship between the individual means and standard deviations appears to be positive and linear for both cohorts. The square root transformation stabilises the variance in the AIDS cohort, but log transformation is better for the more variable NoAIDS cohort.

Although CD4 cell counts are count data, the numbers are large and known to be well approximated by normal or log normal distributions.

Fitting a 2-Stage Random Effects Model

Assume the relationship between CD4count and time is linear for each patient, and that the linear regression parameters vary amongst patients. Here time is characterised by visit number (in this case $1 \ldots 9$). To fit the best straight line to each trajectory we consider the following model

$$Y_{ij} = \mu + P_i + b_i x_{ij} + V_{ij} \tag{2.15}$$

Where μ is the group mean, $P_i \sim \mathrm{N}(0, \sigma_p^2)$ is the patient effect, $b_i \sim \mathrm{N}(b, \sigma_b^2)$ is the trend, x_{ij} is the jth visit number for patient i and $V_{ij} \sim \mathrm{N}(0, \sigma_V^2)$ is the error term.

For each individual i, fit

$$Y_{ij} = \mu_i + b x_{ij} + V_{ij} \tag{2.16}$$

where $\mu_i = \mu + P_i$. Under a fixed efects model, $\mathrm{E}(\hat{\mu}_i|\mu_i) = \mu_i$, $\mathrm{E}(\hat{b}_i|b_i) = b_i$, $var(\hat{\mu}_i) = \sigma_V^2 c_i^\mu$, $var(\hat{b}_i) = \sigma_V^2 c_i^b$, $c_i^\mu = \left(\frac{1}{r} + \frac{\bar{x}^2}{\sum(x-\bar{x})^2}\right)$, and $c_i^b = \frac{1}{\sum(x-\bar{x})^2}$.

Now assume random effects $\mu_i \sim \mathrm{N}(\mu, \sigma_p^2)$ $b_i \sim \mathrm{N}(b, \sigma_b^2)$, and let $\mathcal{E}$ denote expectation under random effects. Then we can show

$$\begin{aligned} \mathcal{E}\mathrm{E}\left[\frac{1}{(m-1)}\sum_{i=1}^{m}(\hat{\mu}_i - \bar{\mu}.)^2\right] &= \sigma_p^2 + \sigma_V^2 \bar{c}^\mu_. \\ \mathcal{E}\mathrm{E}\left[\frac{1}{(m-1)}\sum_{i=1}^{m}(\hat{b}_i - \bar{b}.)^2\right] &= \sigma_b^2 + \sigma_V^2 \bar{c}^b_. \end{aligned} \tag{2.17}$$

Here σ_v^2 is the pooled mean square error from individual regressions $\frac{rss}{n-p'}$ (Weisburg, 1985) The results from this theory are presented in table 2.7

The intercept in the above model, μ_i, represents the initial CD4 count for the ith individual. In table 2.7, it can easily be seen that the mean initial CD4 count is lower for the AIDS group than the non-AIDS group. The difference between each group may be accounted for by assuming the patients that contracted full AIDS were brought into the observation program at a later stage of their disease than those who didn't contract full AIDS.

The slope, b_i, is the rate of change in CD4 count over each time interval. In this case the time interval is 6 monthly. Therefore, in the raw data AIDS case, a mean slope of -40.83 indicates an average drop in CD4 count of 40.83 cubic centimeters

Table 2.7: Results of regression on raw data, log and square root transformations

	$\hat{\mu}$	$\hat{b}$	Pooled σ_v^2	σ_p^2	σ_b^2
AIDS					
Raw Data	648.5497	-40.83333	14687.0	30377.8	429.58
Sqrt Data	25.6378	-1.056128	8.09219	14.00732	0.3546412
Log Data	6.536823	-0.1195582	0.1076965	0.110335	0.007317455
Without AIDS					
Raw Data	762.9612	-29.55331	96644.5	269616.1	1691.69
Sqrt Data	27.1866	-0.612743	35.82079	84.92848	0.6462528
Log Data	6.574461	-0.0535784	0.2739325	0.477713	0.005485728

over 6 months. Table 2.7 shows that on average the CD4 counts for the non-AIDS group decrease at a rate of 29cm^3 over 6 months, which is a less rapid decrease over time than the AIDS group. For the logged data set, the AIDS group has an average percentage drop of 12% over each 6 monthly period.

The variance of the initial CD4 counts, σ_p^2, is in general a lot smaller for the AIDS group then the non-AIDS group. Making transformations of the data does not alter these differences dramatically.

For the raw data the slope variance, σ_b^2, is smaller in the AIDS case. However, taking transformations alters the slope variance in such a way that for the log transformation, the AIDS case has a larger slope variance than the non-AIDS case.

Score test results

Table 2.8 gives the estimated score test statistics under the null hypothesis $\kappa = 0$ for the two HIV$^+$ cohorts on the original and transformed scales. All the P-values are small, demonstrating the presence of systematic relationships between the individual means and the variances of the CD4 cell counts even after transformation. Note though, that the raw and square root counts for the AIDS cohort are not statistically significant at the 1% level. As a comparison, table 2.9 shows similar statistics when the parameters are assumed known and equal to the estimates from a standard one-way analysis of variance using maximum likelihood estimation.

Table 2.8: Score test **mevar2a** results for the CD4 cell count data

	Score test statistic **mevar2a**	P-value
AIDS data		
Raw Data	2.0563	0.040
Sqrt Data	-2.0901	0.037
Log Data	-7.5574	0.000
NoAIDS data		
Raw Data	10.4504	0.000
Sqrt Data	3.3185	0.000
Log Data	-6.5767	0.000

Table 2.9: Score test results for the CD4 cell count data versions **mevar2** and **mevar1**

	σ_η^2	σ^2	**mevar2**	**mevar1**
AIDS data				
Raw Data	19103.85	29942.85	2.0311	3.296
Sqrt Data	11.491	18.835	-2.0328	-3.542
Log Data	0.14142	0.2655	-6.596	-12.268
NoAIDS data				
Raw Data	50736.94	35240.48	10.533	11.944
Sqrt Data	19.9456	13.4876	3.287	3.756
Log Data	0.1456	0.1042	-6.342	-7.440

Square root transformation (NoAIDS-Cohort) reduces the mean-variance relationship and variability observed in the data, but there is still substantial variability unaccounted for in our modelling process. The one-way random effects model for the CD4 cell counts on which the variance component estimates are based ignores serial correlation, trend and possibly other effects. Nonetheless, the score test appears to be responsive to real effects in the data, and behaves in trend as we expect from the exploratory analyses.

The results from the two-stage regression procedure above indicated strong variation between individuals decline in CD4 cell count over time. We can use the variance components obtained from this method (table 2.7) to get score test statistics from our all parameters known methods **mevar1** and **mevar2**. These statistics are shown in table 2.10. They show a similar behaviour as those given previously in table 2.9. In particular notice that the mean-variance relationship for the square root transformation (NoAIDS-Cohort) is now not significant for the approximate score test **mevar2**. Allowing for trend in our analyses has not made a substantial difference in our conclusions of the behaviour of these data.

2.7.2 Blood Pressure Data

Blood pressure data from the International Prospective Primary Prevention Study in Hypertension (IPPPSH) were analysed by Solomon (1985) and Solomon & Cox (1992). Repeated quarterly measurements were made on 25 men over a four year period ($r = 16$). The results are shown in Table 2.11 and confirm Solomon's findings that log transformation stabilises the variance, and substantially reduces the relationship between the individual means and standard deviations.

Solomon and Cox estimated their parameter ρ_{11} to be 0.1254 and 0.0411 for the diastolic blood pressure on the original and log transformed scales, with corresponding estimates 0.1669 and 0.0479 for systolic blood pressure. Estimating κ by

Table 2.10: Score test results for the CD4 cell count data versions **mevar2** and **mevar1** using variance component estimates from the two-stage regression.

	mevar2	**mevar1**
AIDS data		
Raw Data	3.0244	6.176
Sqrt Data	-3.4364	-9.298
Log Data	-10.6457	-28.914
NoAIDS data		
Raw Data	5.8413	8.207
Sqrt Data	1.8027	2.731
Log Data	-3.5111	-5.692

Table 2.11: **mevar2a** Score test statistic for the blood pressure data

BP Data	Score test statistic **mevar2a**	P-value
Diastolic		
raw data	2.9264	0.003
sqrt	1.9170	0.055
log	0.9412	0.347
Systolic		
raw data	3.7178	0.000
sqrt	2.4122	0.016
log	1.0859	0.278

ρ_{11}/σ_A, gives 0.0248 and 0.7568 on the original and log scales for diastolic pressure, and 0.0121 and 0.5308 respectively for systolic blood pressure. Here, the behaviour of ρ_{11} should be consistent with the performance of the score test statistic under $H_0 : \kappa = 0$, and the results presented in Table 2.11 confirm that this is so. H_0 is retained for both log diastolic and systolic blood pressure data, but rejected in the direction of $\kappa > 0$ on the original scale.

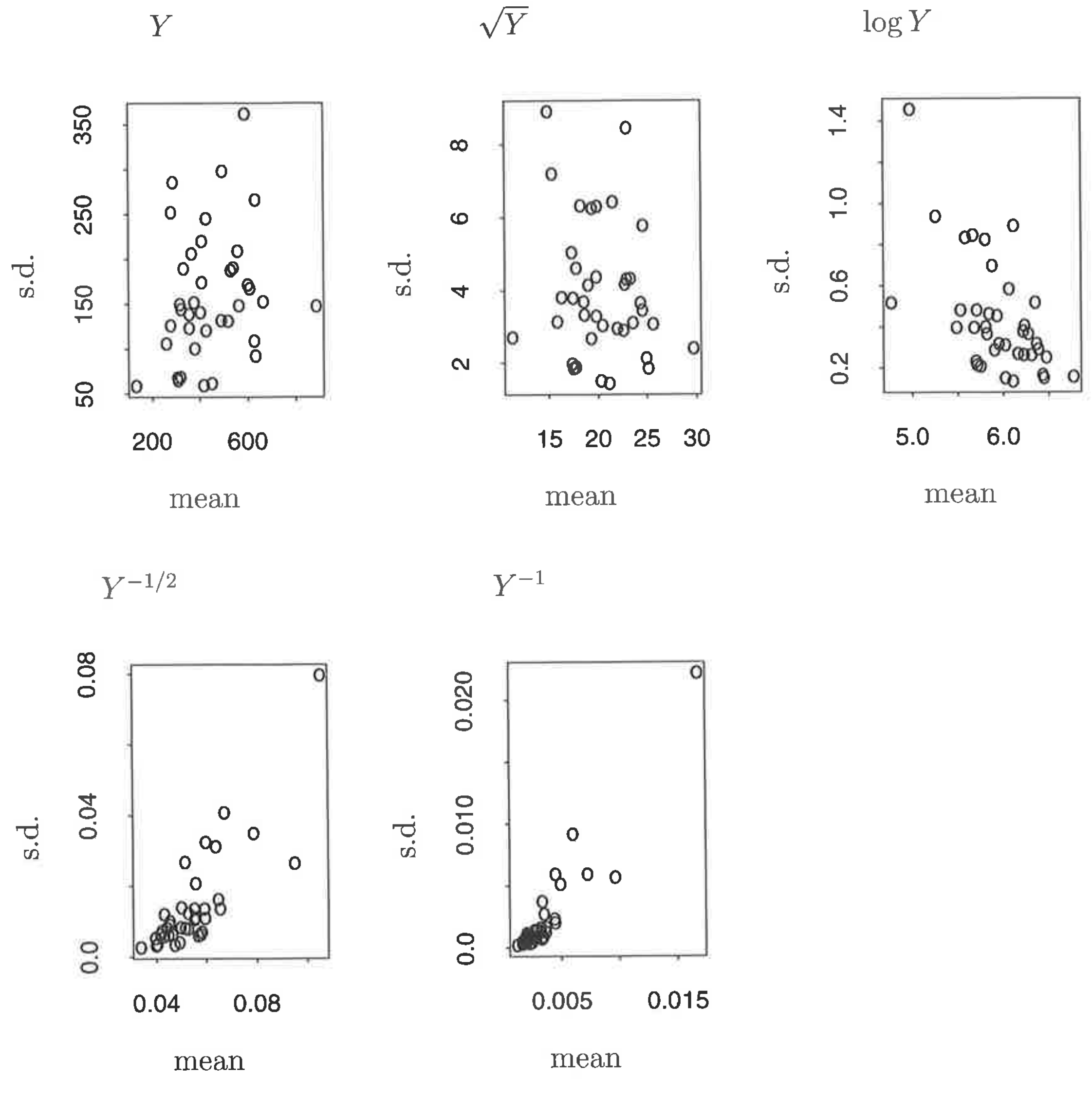


Figure 2.11: Means versus standard deviations for transformation of the AIDS data

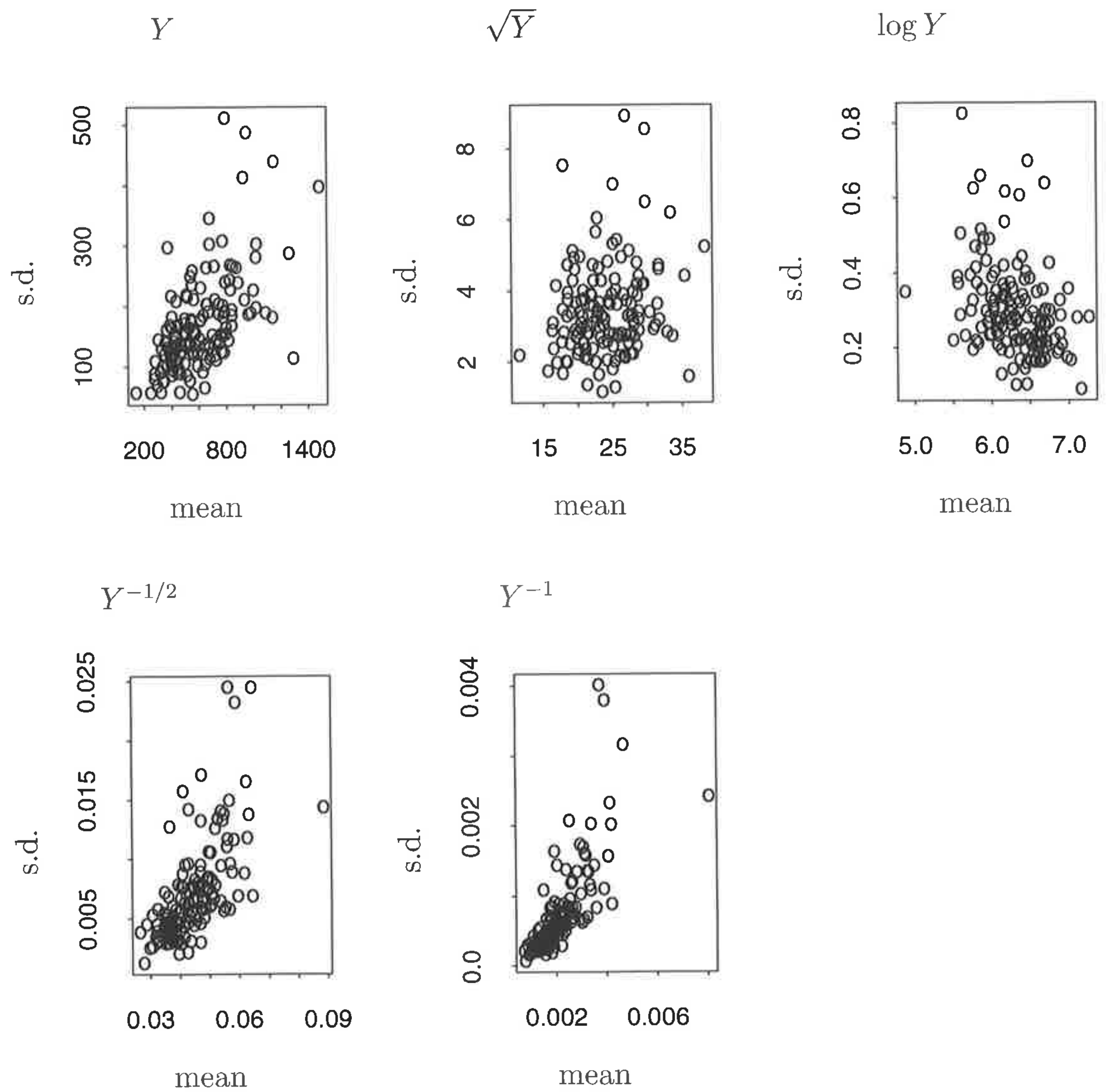
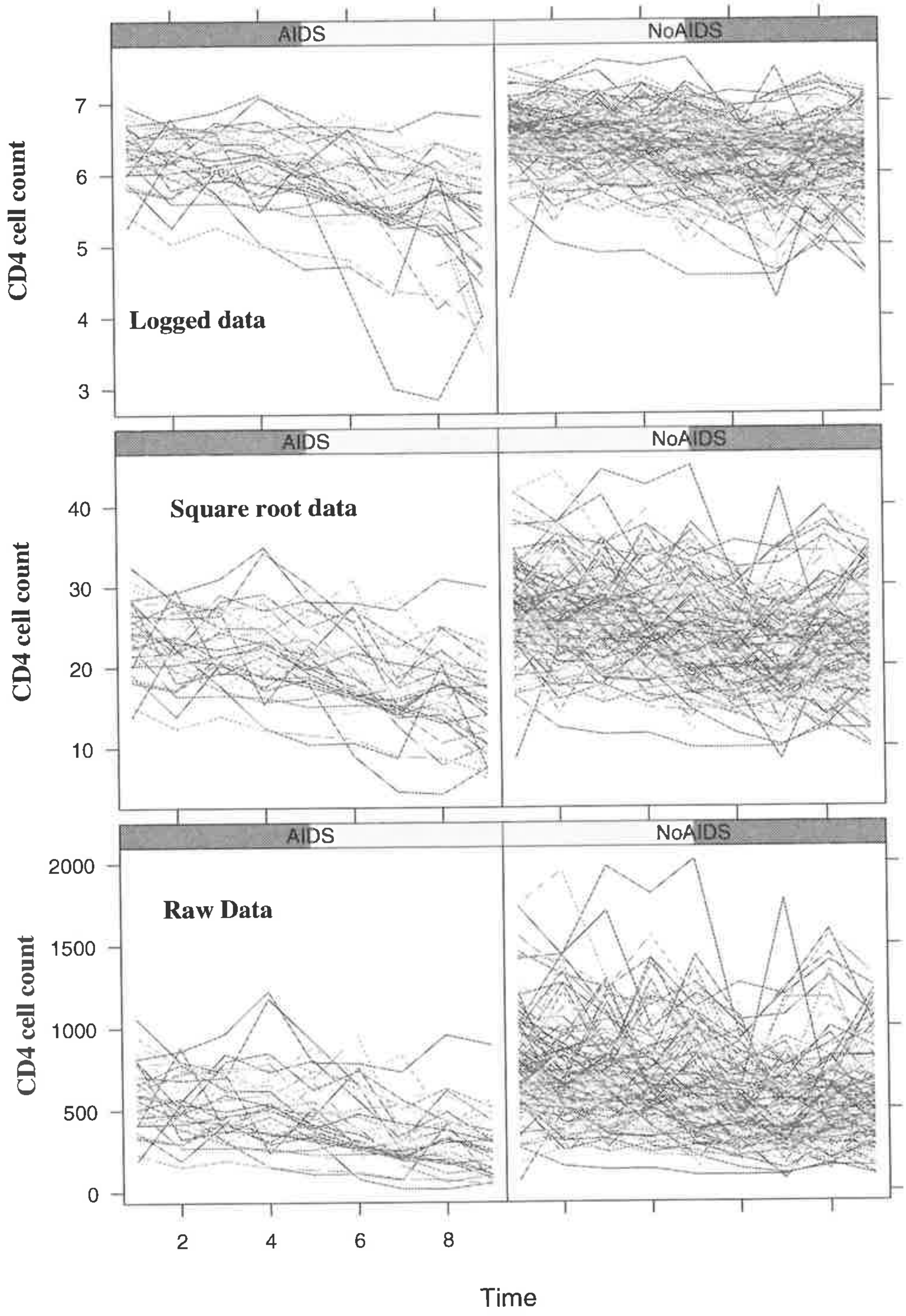


Figure 2.12: Means versus standard deviations for transformation of the NoAIDS data

Figure 2.13: Individual trajectories of the CD4 data for AIDS and NoAIDS cohorts



Chapter 3

Laplace approximations

The following chapters deal with likelihood approximations for general models with multiple random effects. Works in the field of likelihood approximation functions for generalized linear models include Fraser et al. (1994), Wolfinger (1993), Lee & Nelder (1996), Shun (1997), Shun & McCullagh (1995), Lin & Breslow (1996), Breslow & Lin (1995) and Lin (1997) to name a few. Classic papers on the use of likelihoods in random effects models include Gilmour et al. (1985), Lindstrom & Bates (1988) and Schall (1991), while other authors such as Abrams & Sanso (1998) use a likelihood approximation as a basis for a meta-analysis.

When considering a general approximation most authors firstly consider taking a second-order Laplace approximation to the true likelihood for generalized linear mixed models and add in some type of correction term. For example Shun & McCullagh (1995) add a correction term inside the exponential term of the second-order Laplace approximation to the likelihood. Shun (1997) uses this modified Laplace approach to analyse data from the Salamander mating experiment (McCullagh & Nelder, 1989). Lin & Breslow (1996) extend their results from Breslow & Lin (1995) (see previous chapter) to allow for GLMMs with multiple random effects. Their likelihood approximation uses a correction factor derived from Penalised Quasi Likelihood estimates (Breslow & Clayton, 1993).

The form of likelihood approximation presented in the following chapters is based on a multivariate fourth-order Taylor series expansion. We will consider both the second-order and fourth-order expansions and compare these to decide if taking the expansion to fourth-order offers any advantage for the problems studied.

The rest of the current chapter leads discussions on background into Laplace approximations including previous work in this area. Chapter 5 is a complete derivation of the approximation to the likelihood function starting with the most general case. This general case is shown to the fourth-order and is applicable to all models with terms that are crossed, nested or mixtures of these. To explore further the form of this approximation, a complete algebraic expression is found for a fully crossed model with two independent terms followed by a brief discussion of the form for a model with an interaction term. Next we consider a model with two nested terms, the form for the likelihood here is shown for the case of the linear model and a nice form is found for the fourth-order approximation to the likelihood for a general model. Chapter 4 concludes with discussions on expanding around the true values of the random terms (that is expansion around zero). Finally, Chapter 5 shows the behaviour of the approximation for simulated data, the linear form, a crossed Poisson model, a nested exponential type model and finally a logit example using data from the Salamander mating experiment.

3.1 Example: 2-way table

Take a model with two independent random effects $\mathbf{a} = (a_1, a_2, ..., a_m)$ and $\mathbf{b} = (b_1, b_2, ..., b_r)$, distributed normal with zero mean and standard deviations σ_A and σ_B respectively. An example is the linear model:

$$y_{ijk} = \mu + a_i + b_j + \varepsilon_{ijk}, i = 1, ..., m, j = 1, ..., r, k = 1, ..., n,$$

where ε is the error term which is normally distributed with mean zero and standard deviation σ^2. The *conditional log-likelihood* is defined as the log-likelihood obtained by fixing the random effects a_i and b_j, therefore becoming a function of the fixed effect μ and these two fixed terms. For the linear model mentioned above this would be written $l_{ij}^c(y_{ijk}, \mu, \sigma^2 | a_i, b_j) = -\frac{n}{2}\log 2\pi\sigma^2 - \frac{\sum_{k=1}^{n}(y_{ijk}-\mu-a_i-b_j)^2}{2\sigma^2}$, separately for each i and j.

The expression for the true likelihood for any model with conditional log likelihood l_{ij}^c is found by taking expectations over the distributions of all the random effects ($m + r$ integrals):

$$L = (2\pi)^{-(m+r)/2}\sigma_A^{-m}\sigma_B^{-r}\int\cdots\int \exp\left\{\sum_{i=1}^{m}\sum_{j=1}^{r} l_{ij}^c - \sum_{i=1}^{m}\frac{a_i^2}{2\sigma_A^2} - \sum_{j=1}^{r}\frac{b_j^2}{2\sigma_B^2}\right\} da_1 ... da_m db_1 ... db_r.$$

or in matrix form

$$L = (2\pi)^{-(m+r)/2}\sigma_A^{-m}\sigma_B^{-r}\int\cdots\int \exp\left\{\sum_{i=1}^{m}\sum_{j=1}^{r} l_{ij}^c - \frac{\mathbf{a}^T\mathbf{a}}{2\sigma_A^2} - \frac{\mathbf{b}^T\mathbf{b}}{2\sigma_B^2}\right\} d\mathbf{a}d\mathbf{b} \qquad (3.1)$$

Define ϕ as a function of $(y_{ijk}, \mu, \sigma^2, \mathbf{a}, \mathbf{b}, \sigma_A^2, \sigma_B^2)$ then (3.1) becomes

$$L = (2\pi)^{-(m+r)/2}\sigma_A^{-m}\sigma_B^{-r}\int\cdots\int \exp(\phi) d\mathbf{a}d\mathbf{b}. \qquad (3.2)$$

The resulting likelihood is a function of the variance components σ_A^2, σ_B^2 and the fixed effects (for now assume there is only one fixed effect μ, in reality there could be many such as for a regression component $X\tau$).

As an example of the types of general models applicable here, consider for example models whose conditional expectation is a linear function of the two random effects, $E(y_{ij}|a_i, b_j) = \mu_{ij} = f(\mu + a_i + b_j)$, for example, suppose $y_{ij} \sim \text{Bin}(1, p_{ij})$ and we have a logit link:

$$\text{logit}(p_{ij}) = \mu + a_i + b_j.$$

Then we have conditional log likelihood for this written

$$l_{ij}^c(\mathbf{y}|a_i, b_j) = y_{ij}\log(p_{ij}) + (1 - y_{ij})\log(1 - p_{ij})$$

$$= \quad y_{ij}(\mu + a_i + b_j) - \log(1 + \exp(\mu + a_i + b_j)), \qquad (3.3)$$

and therefore the likelihood is

$$(2\pi)^{-(m+r)/2}\sigma_A^{-m}\sigma_B^{-r} \int \cdots \int \exp\left[\sum_{i=1}^{m}\sum_{j=1}^{r}\left\{y_{ij}(\mu + a_i + b_j) - \right.\right.$$

$$\left.\left.\log(1 + \exp(\mu + a_i + b_j)) - \frac{a_i^2}{2r\sigma_A^2} - \frac{b_j^2}{2m\sigma_B^2}\right\}\right] da_1...da_m db_1...db_r. \qquad (3.4)$$

Although these integrals may be approximated numerically without too much fuss, our purpose is to find an explicit algebraic expression for any model with independent crossed random effects and conditional log likelihood represented by l_{ij}^c. High order integrals such as (3.4) are complex, so we propose an approximation based on the Laplace expansion for this expression.

The Laplace approximation is based on a Taylor series expansion for the exponent ϕ, which in this case is given by $\sum_{i=1}^{m}\sum_{j=1}^{r}(l_{ij}^c - \frac{a_i^2}{2r\sigma_A^2} - \frac{b_j^2}{2m\sigma_B^2})$, about estimators denoted $\tilde{\mathbf{a}}$ and $\tilde{\mathbf{b}}$, where $(\tilde{a}_1, \ldots \tilde{a}_m, \tilde{b}_1 \ldots \tilde{b}_r)$ is the point where the exponent is a (local) maximum. These estimates will behave like functions of $(\mu, \sigma^2, \sigma_A^2, \sigma_B^2)$ that maximize the function ϕ

$$\begin{aligned} \phi \quad &\approx \quad \phi(\tilde{\mathbf{a}}, \tilde{\mathbf{b}}) + \frac{\partial \phi}{\partial \mathbf{a}}(\tilde{\mathbf{a}}, \tilde{\mathbf{b}})(\mathbf{a} - \tilde{\mathbf{a}}) + \frac{\partial \phi}{\partial \mathbf{b}}(\tilde{\mathbf{a}}, \tilde{\mathbf{b}})(\mathbf{b} - \tilde{\mathbf{b}}) + \frac{\partial^2 \phi}{\partial \mathbf{a}\partial \mathbf{b}}(\tilde{\mathbf{a}}, \tilde{\mathbf{b}})(\mathbf{a} - \tilde{\mathbf{a}})(\mathbf{b} - \tilde{\mathbf{b}}) \\ &\quad + \frac{1}{2}\frac{\partial^2 \phi}{\partial \mathbf{a}^2}(\tilde{\mathbf{a}}, \tilde{\mathbf{b}})(\mathbf{a} - \tilde{\mathbf{a}})^2 + \frac{\partial^2 \phi}{\partial \mathbf{b}^2}(\tilde{\mathbf{a}}, \tilde{\mathbf{b}})(\mathbf{b} - \tilde{\mathbf{b}})^2 \end{aligned}$$

$$= \quad \phi(\tilde{\mathbf{a}}, \tilde{\mathbf{b}}) + \frac{1}{2}\begin{bmatrix} (\mathbf{a} - \tilde{\mathbf{a}}) \\ (\mathbf{b} - \tilde{\mathbf{b}}) \end{bmatrix}^T \begin{pmatrix} \frac{\partial^2 \phi}{\partial \mathbf{a}^2}(\tilde{\mathbf{a}}, \tilde{\mathbf{b}}) & \frac{\partial^2 \phi}{\partial \mathbf{a}\partial \mathbf{b}}(\tilde{\mathbf{a}}, \tilde{\mathbf{b}}) \\ \frac{\partial^2 \phi}{\partial \mathbf{a}\partial \mathbf{b}}(\tilde{\mathbf{a}}, \tilde{\mathbf{b}}) & \frac{\partial^2 \phi}{\partial \mathbf{b}^2}(\tilde{\mathbf{a}}, \tilde{\mathbf{b}}) \end{pmatrix} \begin{bmatrix} (\mathbf{a} - \tilde{\mathbf{a}}) \\ (\mathbf{b} - \tilde{\mathbf{b}}) \end{bmatrix}. \qquad (3.5)$$

Define a vector $u = (\mathbf{a}, \mathbf{b})$ of length $m + r$ and an $(m + r \times m + r)$ matrix $\mathcal{D}$ of negative second-order derivatives of ϕ, sometimes referred to as the "Hessian" matrix for ϕ, making the Taylor series expansion for ϕ

$$\phi(\tilde{\mathbf{a}}, \tilde{\mathbf{b}}) - \frac{1}{2}(u - \tilde{u})^T \mathcal{D}(u - \tilde{u}).$$

Then the integral (3.2) is approximated by

$$\frac{\exp \sum_{i=1}^{m}\sum_{j=1}^{r}\left(\tilde{l}_{ij}^{c} - \frac{\tilde{a}_i^2}{2r\sigma_A^2} - \frac{\tilde{b}_j^2}{2m\sigma_B^2}\right)}{(2\pi)^{(m+r)/2}\sigma_A^m\sigma_B^r} \int \exp\left\{-\frac{1}{2}(u-\tilde{u})^T\mathcal{D}(u-\tilde{u})\right\} du.$$

Since the matrix $\mathcal{D}$ is symmetric and positive definite, the integral here is in the form of a multivariate normal distribution, and therefore may be approximated as $|\mathcal{D}|^{-\frac{1}{2}}(2\pi)^{\frac{m+r}{2}}$ and the approximation to the log likelihood is

$$-\frac{m}{2}\log(\sigma_A^2) - \frac{r}{2}\log(\sigma_B^2) + \sum_{i=1}^{m}\sum_{j=1}^{r}\left(\tilde{l}_{ij}^{c} - \frac{\tilde{a}_i^2}{2r\sigma_A^2} - \frac{\tilde{b}_j^2}{2m\sigma_B^2}\right) - \frac{1}{2}\log(|\mathcal{D}|), \qquad (3.6)$$

which is a common representation of a second-order Laplace approximation to the true log likelihood.

3.1.1 Other representations

Shun & McCullagh (1995) and Shun (1997) use this representation in their modified Laplace approximation. They refer to the third and fourth-order terms as correction terms, ϵ_0 below, and put them inside the exponent of the likelihood to obtain an approximation for the log likelihood of the form

$$-\frac{m}{2}\log(\sigma_A^2) - \frac{r}{2}\log(\sigma_B^2) - \frac{1}{2}\log(|\mathcal{D}|) + \sum_{i=1}^{m}\sum_{j=1}^{r}\left(\tilde{l}_{ij}^{c} - \frac{\tilde{a}_i^2}{2r\sigma_A^2} - \frac{\tilde{b}_j^2}{2m\sigma_B^2}\right) + \epsilon_0. \qquad (3.7)$$

$\mathcal{D}$ may be written as a function of the information matrix, $\mathcal{I}$, (matrix of negative second derivatives):

$$\begin{aligned}\mathcal{D} &= \begin{pmatrix} I_m\frac{1}{\sigma_A^2} - \frac{\partial^2\tilde{l}^c}{\partial\mathbf{a}^2} & -\frac{\partial^2\tilde{l}^c}{\partial\mathbf{a}\partial\mathbf{b}} \\ -\left(\frac{\partial^2\tilde{l}^c}{\partial\mathbf{a}\partial\mathbf{b}}\right)^T & I_r\frac{1}{\sigma_B^2} - \frac{\partial^2\tilde{l}^c}{\partial\mathbf{b}^2}\end{pmatrix} \\ &= \Sigma^{-1} + \mathcal{I}\end{aligned}$$

where Σ is the $(m+r)\times(m+r)$ variance matrix for $(\mathbf{a},\mathbf{b})$.

To show that (3.6) is the same as the expression used by Lin & Breslow (1996) we need to express it slightly differently. By rearranging (3.6) slightly the log likelihood approximation is equal to

$$\begin{aligned}
& \sum_{i=1}^{m}\sum_{j=1}^{r}\left(\tilde{l}_{ij}^{c} - \frac{\tilde{a}_i^2}{2r\sigma_A^2} - \frac{\tilde{b}_j^2}{2m\sigma_B^2}\right) - \frac{1}{2}\log(\prod_i \sigma_A^2 \prod_j \sigma_B^2 |I + \mathcal{I}\Sigma||\Sigma^{-1}|) \\
= & \sum_{i=1}^{m}\sum_{j=1}^{r}\left(\tilde{l}_{ij}^{c} - \frac{\tilde{a}_i^2}{2r\sigma_A^2} - \frac{\tilde{b}_j^2}{2m\sigma_B^2}\right) - \frac{1}{2}\log(|I + \mathcal{I}\Sigma|) \qquad (3.8)
\end{aligned}$$

which is written in the form used by Lin & Breslow (1996). They refer to (3.8) as the first order Laplace approximation for generalized linear mixed models. They extend this to create an approximation using penalised quasi likelihood estimates; they concentrate on correcting for bias due to these estimates.

Chapter 4

Approximate likelihood functions

4.1 An approximation to the likelihood function for general random effects models

We wish to find an expression for the approximate likelihood function which allows for a model containing p sets of variables. Given p sets of independent normal variables, $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_p$ with variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_p^2$ respectively, each $\mathbf{b}_i$ is a vector containing r_i elements. Let $\sum_{i=1}^{p} r_i = N$. The likelihood is given by finding the expected value of the conditional log likelihood l_c, where the N variables have multivariate normal distribution with mean zero and variance matrix Σ:

$$\frac{1}{(2\pi)^{\frac{N}{2}}|\Sigma|^{\frac{1}{2}}} \int\int \ldots \int \exp\left\{\sum l_c - \frac{1}{2}u^T\Sigma^{-1}u\right\} du \tag{4.1}$$

where u is a vector of length N, $u = (\mathbf{b}_1^T, \mathbf{b}_2^T, \ldots \mathbf{b}_p^T)$. The conditional log likelihood l_c is a single number dependent on one particular value from each of the p variables.

Firstly define a matrix $\mathcal{D}$ of order $(N \times N)$. Take the conditional log likelihood as a function of the p sets of variables and find the second derivative w.r.t. either one or two of these and sum over the remaining $(p-2)$ sets of variables to create a matrix for each combination of sets of variables:

$$\begin{aligned}
\tilde{l}_{11}(i,j) &= \sum_{r_2,\ldots,r_p} \frac{\partial^2 l_c}{\partial \mathbf{b}_{1i}\mathbf{b}_{1j}} \text{ (diagonal matrix)} \\
\tilde{l}_{12}(i,j) &= \sum_{r_3,\ldots,r_p} \frac{\partial^2 l_c}{\partial \mathbf{b}_{1i}\mathbf{b}_{2j}} \\
\tilde{l}_{13}(i,j) &= \sum_{r_2,r_4,\ldots,r_p} \frac{\partial^2 l_c}{\partial \mathbf{b}_{1i}\mathbf{b}_{3j}} \\
&\text{etc...}
\end{aligned}$$

Then $\tilde{l}_{kl}$ is an $(r_k \times r_l)$ matrix corresponding to the sets $\mathbf{b}_k$ and $\mathbf{b}_l$. The matrix $\mathcal{D}$, as defined before, is partitioned as

$$\mathcal{D} = \begin{pmatrix} \frac{1}{\sigma_1^2} I_{r_1} - \tilde{l}_{11} & -\tilde{l}_{12} & -\tilde{l}_{13} & \cdots \\ -\tilde{l}_{12}^T & \frac{1}{\sigma_2^2} I_{r_2} - \tilde{l}_{22} & -\tilde{l}_{23} & \cdots \\ -\tilde{l}_{13}^T & -\tilde{l}_{23}^T & \frac{1}{\sigma_3^2} I_{r_3} - \tilde{l}_{33} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \tag{4.2}$$

where I_{r_1} is the identity matrix of size r_1. Let $J = (j_1, j_2, ..., j_p)$ be a sequence of nonnegative integers with *norm* $|J| = j_1 + j_2 + \ldots + j_p$ and $J! = j_1!j_2!\ldots j_p!$. A general fourth derivative is given by $|J| = 4$ and the differential operator

$$d_J = \frac{\partial^4}{\partial \mathbf{b}_1^{j_1} \partial \mathbf{b}_2^{j_2} \ldots \partial \mathbf{b}_p^{j_p}}.$$

A fourth-order Taylor expansion gives the expression for the approximate likelihood as

$$\frac{\exp\left\{\sum \tilde{l}_c - \frac{1}{2}\tilde{u}^T \Sigma^2 \tilde{u}\right\}}{|\Sigma \mathcal{D}|^{1/2}} \int \frac{|\mathcal{D}|^{1/2}}{(2\pi)^{N/2}} \exp\left\{-\frac{1}{2}(u-\tilde{u})^T \mathcal{D}(u-\tilde{u})\right\}$$

$$\times \left[1 + \sum_{i=1}^{r_1} \sum_{j=r_1+1}^{r_1+r_2} \cdots \sum_{p=N-r_p}^{N} \left\{ \sum_{|J|=4} \frac{1}{J!} d_J \tilde{l}_c \times (u-\tilde{u})_i^{j_1} (u-\tilde{u})_j^{j_2} ... (u-\tilde{u})_p^{j_p} \right\}\right] du$$

The first term of the integral is in the form of a multivariate normal distribution, $u - \tilde{u} \sim N_N(0, \mathcal{D}^{-1})$, since the matrix $\mathcal{D}$ is symmetric and positive definite.

Moment generating function

The moment generating function for a multivariate normal distribution is written

$$M_{(u-\tilde{u})}(t) = \exp\{\frac{1}{2}t^T\mathcal{D}^{-1}t\},$$

so to find the fourth-order terms of the likelihood approximation we need to use this expression to find

$$E\left((u-\tilde{u})_1^{k_1}(u-\tilde{u})_2^{k_2}....)\right) = \frac{\partial^{k_1+k_2+...}}{\partial t_1^{k_1}\partial t_2^{k_2}...}M_{(u-\tilde{u})}(t)\Big|_{t=0}.$$

In our fourth-order likelihood approximation we have 5 different types of expected values to find; define these by

$$\begin{aligned}
J(4) &= (4,0,\ldots,0), j_i = 4 \text{ for some } i \\
J(3,1) &= (3,1,0,\ldots,0), j_i = 3, j_j = 1 \text{ for some } i \neq j \\
J(2,2) &= (2,2,0,\ldots,0), j_i = 2, j_j = 2 \text{ for some } i \neq j \\
J(2,1,1) &= (2,1,1,0\ldots,0), j_i = 2, j_j = 1, j_k = 1 \text{ for some } i \neq j \neq k \\
J(1,1,1,1) &= (1,1,1,1,0\ldots,0), j_i = 1, j_j = 1, j_k = 1, j_l = 1 \text{ for some } i \neq j \neq k \neq l.
\end{aligned}$$

Now we find expressions for each of these in turn.

Consider the following expectations (where $\mathcal{D}_{ij}^{-1}$ is the ijth element of $\mathcal{D}^{-1}$):

$$E(u-\tilde{u})_1^4 = \frac{\partial^4}{\partial t_1^4}\exp\{\frac{1}{2}t^T\mathcal{D}^{-1}t\}\Big|_{t=0}$$

$$= \left(3(\mathcal{D}_{11}^{-1})^2 + 6\mathcal{D}_{11}^{-1}(t_1\mathcal{D}_{11}^{-1} + \sum_{j\neq 1}t_j\mathcal{D}_{1j}^{-1})^2 + (t_1\mathcal{D}_{11}^{-1} + \sum_{j\neq 1}t_j\mathcal{D}_{1j}^{-1})^4\right)\exp\{\frac{1}{2}t^T\mathcal{D}^{-1}t\}\Big|_{t=0}$$

$$= 3(\mathcal{D}_{11}^{-1})^2$$

and

$$E(u-\tilde{u})_1^2(u-\tilde{u})_2^2 = \frac{\partial^4}{\partial t_1^2\partial t_2^2}\exp\{\frac{1}{2}t^T\mathcal{D}^{-1}t\}\Big|_{t=0}$$

$$= \left(\mathcal{D}_{11}^{-1}\mathcal{D}_{22}^{-1} + \mathcal{D}_{11}^{-1}(t_2\mathcal{D}_{22}^{-1} + \sum_{j\neq 2} t_j\mathcal{D}_{2j}^{-1})^2 \right.$$

$$+2(\mathcal{D}_{12}^{-1})^2 + 2\mathcal{D}_{12}^{-1}(t_1\mathcal{D}_{11}^{-1} + \sum_{j\neq 1} t_j\mathcal{D}_{1j}^{-1})(t_2\mathcal{D}_{22}^{-1} + \sum_{j\neq 2} t_j\mathcal{D}_{2j}^{-1})$$

$$\left. + \text{ terms involving } t \right) \exp\{\frac{1}{2}t^T\mathcal{D}^{-1}t\}\Big|_{t=0}$$

$$= \mathcal{D}_{11}^{-1}\mathcal{D}_{22}^{-1} + 2(\mathcal{D}_{12}^{-1})^2$$

Similarly,

$$E\left((u-\tilde{u})_1^3(u-\tilde{u})_2\right) = 3\mathcal{D}_{11}^{-1}\mathcal{D}_{12}^{-1}$$

$$E\left((u-\tilde{u})_1^2(u-\tilde{u})_2(u-\tilde{u})_3\right) = \mathcal{D}_{11}^{-1}\mathcal{D}_{23}^{-1} + 2\mathcal{D}_{12}^{-1}\mathcal{D}_{13}^{-1}$$

$$E\left((u-\tilde{u})_1(u-\tilde{u})_2(u-\tilde{u})_3(u-\tilde{u})_4\right) = \mathcal{D}_{12}^{-1}\mathcal{D}_{34}^{-1} + \mathcal{D}_{13}^{-1}\mathcal{D}_{24}^{-1} + \mathcal{D}_{14}^{-1}\mathcal{D}_{23}^{-1}$$

Therefore the Laplace approximation to the true log likelihood for a model with multiple crossed main effects is

$$\sum_N \tilde{l}_c - \frac{1}{2}\tilde{u}^T\Sigma^{-1}\tilde{u} - \frac{1}{2}\log(|\mathcal{D}\Sigma|) + \sum_N \left\{ \frac{1}{4!}\sum_{|J(4)|} d_{J(4)}\tilde{l}_c\left(3\mathcal{D}_{ii}^{-1}\right)^2 \right.$$

$$+\frac{1}{3!1!}\sum_{|J(3,1)|} d_{J(3,1)}\tilde{l}_c 3\mathcal{D}_{ii}^{-1}\mathcal{D}_{ij}^{-1} + \frac{1}{2!2!}\sum_{|J(2,2)|} D_{J(2,2)}\tilde{l}_c\left(\mathcal{D}_{ii}^{-1}\mathcal{D}_{jj}^{-1} + 2(\mathcal{D}_{ij}^{-1})^2\right)$$

$$+\frac{1}{2!1!1!}\sum_{|J(2,1,1)|} d_{J(2,1,1)}\tilde{l}_c\left(\mathcal{D}_{ii}^{-1}\mathcal{D}_{jk}^{-1} + 2\mathcal{D}_{ij}^{-1}\mathcal{D}_{ik}^{-1}\right)$$

$$\left. + \sum_{|J(1,1,1,1)|} d_{J(1,1,1,1)}\tilde{l}_c\left(\mathcal{D}_{ij}^{-1}\mathcal{D}_{kl}^{-1} + \mathcal{D}_{ik}^{-1}\mathcal{D}_{jl}^{-1} + \mathcal{D}_{il}^{-1}\mathcal{D}_{jk}^{-1}\right) \right\} \tag{4.3}$$

When directly compared to (3.7) we can see that this is in exactly the same form and the correction term here is simply the terms resulting from the fourth-order expansion.

4.1.1 Models with nested random effects

The nested case may be represented in a similar way as equation 4.3 for crossed models. In the case of two random effects, the matrix $\mathcal{D}$ (partitioned) is given by the $(r_1 + r_1 r_2) \times (r_1 + r_1 r_2)$ matrix

$$\mathcal{D} = \left(\begin{array}{cccc|cccc} A_1 & 0 & \dots & 0 & \tilde{l}_{12}(1j) & 0_{r_2} & \dots & 0_{r_2} \\ 0 & A_2 & \dots & 0 & 0_{r_2} & \tilde{l}_{12}(2j) & \dots & 0_{r_2} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & A_{r_1} & 0_{r_2} & 0_{r_2} & \dots & \tilde{l}_{12}(r_1 j) \\ \hline \tilde{l}^T_{12}(1j) & 0_{r_2} & \dots & 0_{r_2} & B_{11} & 0 & \dots & 0 \\ 0_{r_2} & \tilde{l}^T_{12}(2j) & \dots & 0_{r_2} & 0 & B_{12} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0_{r_2} & 0_{r_2} & \dots & \tilde{l}^T_{12}(r_1 j) & 0 & 0 & \dots & B_{r_2 r_2} \end{array}\right)$$

where 0_{r_2} is a 0 vector of length r_2, and

$$\begin{aligned} A_i &= \frac{1}{r_2 \sigma_1^2} - \sum_{j=1}^{r_2} \tilde{l}_{11}(ij) \\ B_{ij} &= \frac{1}{\sigma_2^2} - \tilde{l}_{22}(ij) \\ \tilde{l}_{12}(ij) &= (\tilde{l}_{12}(i1), \tilde{l}_{12}(i2), \dots, \tilde{l}_{12}(ir_2)). \end{aligned}$$

For many random effects this matrix has dimensions $(r_1 + r_1 r_2 + r_1 r_2 r_3 + \dots + \prod_{i=1}^m r_i) \times (r_1 + r_1 r_2 + r_1 r_2 r_3 + \dots + \prod_{i=1}^m r_i)$ which is very big! The next section shows an explicit form of the second-order approximation for nested models.

Additive models

For models where the effects always appear as a summation, as in the previous logit example, each combination of second and fourth-order derivatives are the same, i.e. $l_{11} = l_{12} = l_{22}$ and $l_{1111} = l_{1112} = l_{1122} = ...$ Proof is given by the chain rule as follows:
The conditional log likelihood is a function of a summation

$$l = f\left(g(x)\right), \text{ where } g(x) = \sum_i x_i.$$

Now,

$$\begin{aligned}
g(x) &= x_1 + x_2 + x_3 + ... \\
\frac{\partial g(x)}{\partial x_i} &= 1 \text{ for all } i \\
\frac{\partial^2 g(x)}{\partial x_i \partial x_j} &= 0 \text{ for all } i, j \text{ combinations}
\end{aligned}$$

The derivatives of the function $f(g(x))$ are

$$\begin{aligned}
\frac{\partial f\left(g(x)\right)}{\partial x_i} &= g'(x).f'\left(g(x)\right) = f'\left(g(x)\right) \\
\frac{\partial^2 f\left(g(x)\right)}{\partial x_i \partial x_j} &= g''(x).f'\left(g(x)\right) + \left(g'(x)\right)^2.f''\left(g(x)\right) = f''\left(g(x)\right) \\
\text{similarly} & \\
\frac{\partial^4 f\left(g(x)\right)}{\partial x_i \partial x_j \partial x_k \partial x_l} &= f''''\left(g(x)\right).
\end{aligned}$$

This simpilifies the expression for the approximate log likelihood by grouping all the fourth-order derivatives. This proof will be especially useful in Chapter 5, where we find the approximation to the likelihood for some examples of additive models.

4.2 Approximations for models with two random effects: explicit expressions

4.2.1 Two-way crossed with no interaction

The fourth-order Taylor series expansion for a function $f(x, y)$ around two variables (a, b) is

$$\sum_{p=0}^{4}\sum_{d=0}^{p}\frac{1}{d!(p-d)!}\frac{\partial^p}{\partial a^d \partial b^{p-d}}f(a,b)(x-a)^d(y-b)^{p-d}.$$

The function we wish to expand is ϕ, as defined earlier. "$f(a, b)$" is $\phi(\mathbf{a}, \mathbf{b})$:

$$\phi(\mathbf{a},\mathbf{b}) = \sum_{i=1}^{m}\sum_{j=1}^{r} l_c - \sum_{i=1}^{m}\frac{a_i^2}{2\sigma_A^2} - \sum_{j=1}^{r}\frac{b_j^2}{2\sigma_B^2}.$$

We wish to expand around $(\tilde{\mathbf{a}}, \tilde{\mathbf{b}})$, hence our approximation to the true likelihood is found by expanding this as below:

$$(2\pi)^{-(m+r)/2}\sigma_A^{-m}\sigma_B^{-r}\int\cdots\int \exp\sum_{i=1}^{m}\sum_{j=1}^{r}\left[\tilde{l}_c - \frac{\tilde{a}_i^2}{2r\sigma_A^2} - \frac{\tilde{b}_j^2}{2m\sigma_B^2}\right.$$

$$+\frac{1}{2}(\tilde{l}_{11}(ij) - \frac{1}{r\sigma_A^2})(a_i-\tilde{a}_i)^2 + \frac{1}{2}(\tilde{l}_{22}(ij) - \frac{1}{m\sigma_B^2})(b_j-\tilde{b}_j)^2$$

$$+\tilde{l}_{12}(ij)(a_i-\tilde{a}_i)(b_j-\tilde{b}_j) + \frac{1}{6}\tilde{l}_{111}(a_i-\tilde{a}_i)^3 + \frac{1}{6}\tilde{l}_{222}(b_i-\tilde{b}_j)^3$$

$$+\frac{1}{2}\tilde{l}_{112}(a_i-\tilde{a}_i)^2(b_j-\tilde{b}_j) + \frac{1}{2}\tilde{l}_{122}(a_i-\tilde{a}_i)(b_j-\tilde{b}_j)^2 + \frac{1}{24}\tilde{l}_{1111}(a_i-\tilde{a}_i)^4$$

$$+\frac{1}{24}\tilde{l}_{2222}(b_i-\tilde{b}_j)^4 + \frac{1}{6}\tilde{l}_{1112}(a_i-\tilde{a}_i)^3(b_j-\tilde{b}_j)$$

$$\left. + \frac{1}{6}\tilde{l}_{1222}(a_i-\tilde{a}_i)(b_j-\tilde{b}_j)^3 + \frac{1}{4}\tilde{l}_{1122}(a_i-\tilde{a}_i)^2(b_j-\tilde{b}_j)^2\right] da_1 \ldots da_m db_1 \ldots db_r.$$

In the notation above, the number of digits in the subscripts for each $\tilde{l}$ representing the order of differentiation of the conditional log likelihood $\tilde{l}_{ij}$, that is, 2 digits such as $\tilde{l}_{11}(ij)$ or $\tilde{l}_{12}(ij)$ represents a second derivative, 3 digits such as $\tilde{l}_{111}$ represents a third derivative etc. The actual numbers themselves represent the variables the differentiation is with respect to, 1 refers to an a_i and 2 refers to a b_j. For example, $\tilde{l}_{11}(ij) = d^2 l_{ij}/da_i^2|_{a_i=\tilde{a}_i}$, $\tilde{l}_{12}(ij) = d^2 l_{ij}/da_1 ... da_m db_1 ... db_r|_{a_i=\tilde{a}_i, b_j=\tilde{b}_j}$ and similarly for higher orders.

The method is to complete the square and make substitutions for $a_i - \tilde{a}_i$ and $b_j - \tilde{b}_j$ so that the integral is in the form $\int\int \exp\{-\sum_{i=1}^m y_i^2/2 - \sum_{j=1}^r x_j^2/2\} dydx$. First rewrite the above expression as

$$\text{constant} \int \cdots \int \exp\left\{-\frac{1}{2}\left(\sum_{i=1}^{m} \frac{A_i}{\sigma_A^2}(a_i - \tilde{a}_i)^2 + \sum_{j=1}^{r} \frac{B_j}{\sigma_B^2}(b_j - \tilde{b}_j)^2 \right.\right.$$

$$\left.\left. - 2\sum_{i=1}^{m}\sum_{j=1}^{r} \tilde{l}_{12}(ij)(a_i - \tilde{a}_i)(b_j - \tilde{b}_j)\right)\right\}$$

$$\times (1 + \text{third and fourth derivatives}) da_1 ... da_m db_1 ... db_r$$

where $A_i = 1 - \sum_{j=1}^r \tilde{l}_{11}(ij)\sigma_A^2$ and $B_j = 1 - \sum_{i=1}^m \tilde{l}_{22}(ij)\sigma_B^2$. A second-order approximation to the true likelihood is defined as the first term of this integral, i.e. the exponential term. A second-order approximation is

$$L \approx \frac{\exp \sum_{i=1}^m \sum_{j=1}^r \left(\tilde{l}_c - \frac{\tilde{a}_i^2}{2r\sigma_A^2} - \frac{\tilde{b}_j^2}{2m\sigma_B^2}\right)}{(2\pi)^{(m+r)/2}\sigma_A^m \sigma_B^r} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}\left\{\sum_{j=1}^{r} \frac{B_j}{\sigma_B^2}(b_j - \tilde{b}_j)^2 \right.\right.$$

$$\left.\left. + \sum_{i=1}^{m} \frac{A_i}{\sigma_A^2}(a_i - \tilde{a}_i)^2 - 2\sum_{i=1}^{m}\sum_{j=1}^{r} \tilde{l}_{12}(ij)(a_i - \tilde{a}_i)(b_j - \tilde{b}_j)\right\}\right] da_1 da_2 \ldots db_r \quad (4.4)$$

Since all the terms a_i will have similar integrals we can make a general substitution.

Let $y_i = \frac{\sqrt{A_i}}{\sigma_A}(a_i - \tilde{a}_i) - \frac{\sigma_A \sum_{j=1}^r \tilde{l}_{12}(ij)(b_j - \tilde{b}_j)}{\sqrt{A_i}}$ and $da_i = \frac{\sigma_A}{\sqrt{A_i}} dy_i$, and we get

$$\frac{\exp \sum_{i=1}^m \sum_{j=1}^r \left(\tilde{l}_c - \frac{\tilde{a}_i^2}{2r\sigma_A^2} - \frac{\tilde{b}_j^2}{2m\sigma_B^2}\right)}{(2\pi)^{(m+r)/2}\sigma_A^m \sigma_B^r} \left[\prod_i \frac{\sigma_A}{\sqrt{A_i}}\right] \int \cdots \int \exp\left[-\frac{1}{2}\left\{\sum_{j=1}^r \frac{B_j}{\sigma_B^2}(b_j - \tilde{b}_j)^2\right.\right.$$

$$\left.\left. + \sum_{i=1}^m y_i^2 - \sum_{i=1}^m \frac{\sigma_A^2(\sum_{j=1}^r \tilde{l}_{12}(ij)(b_j - \tilde{b}_j))^2}{A_i}\right\}\right] dy_1 \ldots dy_m db_1 \ldots db_r$$

$$= \frac{\exp \sum_{i=1}^m \sum_{j=1}^r \left(\tilde{l}_c - \frac{\tilde{a}_i^2}{2r\sigma_A^2} - \frac{\tilde{b}_j^2}{2m\sigma_B^2}\right)}{(2\pi)^{(m+r)/2}\sigma_B^r \prod_i \sqrt{A_i}} \times \int \cdots \int \exp\left[-\frac{1}{2}\left\{\sum_{i=1}^m y_i^2\right.\right.$$

$$+ \sum_{j=1}^r (b_j - \tilde{b}_j)^2 \left(\frac{B_j}{\sigma_B^2} - \sum_{i=1}^m \frac{\sigma_A^2 \tilde{l}_{12}(ij)^2}{A_i}\right)$$

$$\left.\left. -2\sum_{i=1}^m \sum_{k \neq j} \frac{\sigma_A^2 \tilde{l}_{12}(ij)(b_j - \tilde{b}_j)\tilde{l}_{12}(ik)(b_k - \tilde{b}_k)}{A_i}\right\}\right] dy_1 \ldots dy_m db_1 \ldots db_r.$$

Now create an $(r \times r)$ symmetric matrix $C = \text{diag}(B_j) - \sigma_A^2 \sigma_B^2 \tilde{l}_{12}(ij)^T \text{diag}(A_i^{-1}) \tilde{l}_{12}$, where

$$\text{diag}(B_j) = \begin{pmatrix} B_1 & 0 & \ldots \\ 0 & B_2 & \ldots \\ \vdots & \vdots & \ddots \end{pmatrix}.$$

The above integrals can now be written in the form we require,

$$\int \cdots \int \exp\left\{-\sum_{i=1}^m \frac{y_i^2}{2} - \frac{1}{2\sigma_B^2}(b - \tilde{b})^T C (b - \tilde{b})\right\} dy_1 \ldots dy_m db_1 \ldots db_r$$

$$= \frac{\sigma_B^r}{|C|^{1/2}} \int \cdots \int \exp\left\{-\sum_{i=1}^m \frac{y_i^2}{2} - \frac{1}{2}X^T X\right\} dy_1 \ldots dy_m dx_1 \ldots dx_r$$

where X is a vector of length r with elements x_j and there exists an $(r \times r)$ matrix P such that $X = P(b - \tilde{b})/\sigma_B$ and $P^T P = C$, also $|P| = |C|^{1/2}$.

The second-order terms of the Laplace approximation can now be written

$$\frac{\exp \sum_{i=1}^{m} \sum_{j=1}^{r} \left(\tilde{l}_c - \frac{\tilde{a}_i^2}{2r\sigma_A^2} - \frac{\tilde{b}_j^2}{2m\sigma_B^2}\right)}{(2\pi)^{(m+r)/2} \prod_i \sqrt{A_i} |C|^{\frac{1}{2}}} \int \cdots \int \exp\left\{-\frac{\sum_{i=1}^{m} y_i^2}{2} - \frac{1}{2} X^T X\right\} dy_1 ... dx_r$$

$$= \frac{\exp \sum_{i=1}^{m} \sum_{j=1}^{r} \left(\tilde{l}_c - \frac{\tilde{a}_i^2}{2r\sigma_A^2} - \frac{\tilde{b}_j^2}{2m\sigma_B^2}\right)}{\prod_i \sqrt{A_i} |C|^{\frac{1}{2}}} \tag{4.5}$$

The second-order terms of the approximation to the true log likelihood may also be written

$$\sum_{i=1}^{m} \sum_{j=1}^{r} \left(\tilde{l}_c - \frac{\tilde{a}_i^2}{2r\sigma_A^2} - \frac{\tilde{b}_j^2}{2m\sigma_B^2}\right) - \frac{1}{2} \log\left(|\mathrm{diag}(A_i)|\right.$$

$$\left. \times \left|\mathrm{diag}(B_j) - \sigma_A^2 \sigma_B^2 \tilde{l}_{12}(ij)^T \mathrm{diag}(A_i^{-1}) \tilde{l}_{12}(ij)\right|\right). \tag{4.6}$$

If we create a partitioned matrix D,

$$\mathrm{D} = \begin{pmatrix} \mathrm{diag}(A_i) & 0 \\ 0 & C \end{pmatrix}$$

then 4.6 becomes

$$\sum_{i=1}^{m} \sum_{j=1}^{r} \left(\tilde{l}_c - \frac{\tilde{a}_i^2}{2r\sigma_A^2} - \frac{\tilde{b}_j^2}{2m\sigma_B^2}\right) - \frac{1}{2} \log(|D|). \tag{4.7}$$

$|D|$ is equal to $|\mathcal{D}|$ in section 3.1, and thus compares directly to past forms of the Laplace approximation, especially Lin & Breslow (1996).

Adding in the higher order terms

Now return to the fourth-order expansion. We need to express the third and fourth-order terms in terms of y_i and some vector X representing the substitution for b_j. The full substitutions are

$$a_i - \tilde{a}_i = \frac{\sigma_A}{\sqrt{A_i}} y_i + \frac{\sigma_A^2}{A_i} \sum_{j=1}^{r} \tilde{l}_{12}(ij)(b_j - \tilde{b}_j)$$

$$\text{and } b_j - \tilde{b}_j = \sigma_B (P^{-1} X)_j.$$

Note here that the transformation for $b_j - \tilde{b}_j$ can be written $\sigma_B \sum_{k=1}^{r} P_{jk}^{-1} x_k$ and also

$$P^T P = C \quad \rightarrow \quad \sum_{j=1}^{r} C_{jj} = \sum_{i,j} P_{ij}^2$$

and

$$P^{-1}(P^{-1})^T = C^{-1} \quad \rightarrow \quad \sum_{j=1}^{r} C_{jj}^{-1} = \sum_{i,j} (P_{ij}^{-1})^2$$

Remembering that the first and third moments are zero for a standard normal, the third and fourth-order terms in the series become

$$\sum_{i=1}^{m}\sum_{j=1}^{r}\Bigg[\frac{1}{24}\tilde{l}_{1111}\left(\frac{y_i^4\sigma_A^4}{A_i^2} + \frac{\sigma_A^8}{A_i^4}(\sum_{j=1}^{r}\tilde{l}_{12}(ij)\sigma_B(P^{-1}X)_j)^4 + \frac{6\sigma_A^6}{A_i^3}y_i^2(\sum_{j=1}^{r}\tilde{l}_{12}(ij)\sigma_B(P^{-1}X)_j)^2\right)$$

$$+\frac{1}{6}\tilde{l}_{1112}\left(\frac{3\sigma_A^4\sigma_B^2}{A_i^2}y_i^2(\sum_{j=1}^{r}\tilde{l}_{12}(ij)(P^{-1}X)_j)(P^{-1}X)_j + \frac{\sigma_A^6\sigma_B^4}{A_i^3}(\sum_{j=1}^{r}\tilde{l}_{12}(ij)(P^{-1}X)_j)^3(P^{-1}X)_j\right)$$

$$+\frac{1}{4}\tilde{l}_{1122}\left(\frac{\sigma_A^2\sigma_B^2 y_i^2}{A_i}((P^{-1}X)_j)^2 + \frac{\sigma_A^4\sigma_B^4}{A_i^2}(\sum_{j=1}^{r}\tilde{l}_{12}(ij)(P^{-1}X)_j)^2((P^{-1}X)_j)^2\right)$$

$$+\quad \frac{1}{6}\tilde{l}_{1222}\frac{\sigma_A^2\sigma_B^4}{A_i}\sum_{j=1}^{r}\tilde{l}_{12}(ij)(P^{-1}X)_j((P^{-1}X)_j)^3 + \frac{1}{24}\tilde{l}_{2222}\sigma_B^4((P^{-1}X)_j)^4\Bigg]$$

The fourth-order approximation to the true log likelihood for a two-way crossed model with no interaction is found by multiplying this by $\exp\left\{-\frac{1}{2}\sum_{i=1}^{m} y_i^2 - \frac{1}{2}X^TX\right\}$ then integrating w.r.t. all the y_i then the x_j and putting this together with (4.5) to get

$$l \approx \sum_{j=1}^{m}\sum_{i=1}^{r}\left[\tilde{l}_c - \frac{\tilde{a}_i^2}{2r\sigma_A^2} - \frac{\tilde{b}_j^2}{2m\sigma_B^2}\right] - \frac{1}{2}\sum_{i=1}^{m}\log(A_i) - \frac{1}{2}\log(|C|)$$

$$+\frac{1}{8}\sum_{j=1}^{m}\sum_{i=1}^{r}\Bigg[\tilde{l}_{1111}\frac{\sigma_A^4}{A_i^2}\left(1 + \frac{\sigma_A^2\sigma_B^2}{A_i}\sum_{j=1}^{r}\sum_{k=1}^{r}\left[\tilde{l}_{12}(ij)\tilde{l}_{12}(ik)C_{kj}^{-1}\right]\right)^2$$

$$+2\tilde{l}_{1122}\frac{\sigma_A^2\sigma_B^2}{A_i}\left((1 + \frac{\sigma_A^2\sigma_B^2}{A_i}\sum_{j=1}^{r}\sum_{k=1}^{r}\left[\tilde{l}_{12}(ij)\tilde{l}_{12}(ik)C_{kj}^{-1}\right])C_{jj}^{-1} + 2\frac{\sigma_A^2\sigma_B^2}{A_i}(\sum_{k=1}^{r}\tilde{l}_{12}(ik)C_{kj}^{-1})^2\right)$$

$$+4\tilde{l}_{1112}\frac{\sigma_A^4\sigma_B^2}{A_i^2}\left(1+\frac{\sigma_A^2\sigma_B^2}{A_i}\sum_{j=1}^{r}\sum_{k=1}^{r}\left[\tilde{l}_{12}(ij)\tilde{l}_{12}(ik)C_{kj}^{-1}\right]\right)\sum_{k=1}^{r}\tilde{l}_{12}(ik)C_{kj}^{-1}$$

$$+4\tilde{l}_{2221}\sigma_B^2C_{jj}^{-1}\frac{\sigma_A^2\sigma_B^2}{A_i}\sum_{k=1}^{r}C_{jk}^{-1}\tilde{l}_{12}^T(ki)+\tilde{l}_{2222}\sigma_B^4(C_{jj}^{-1})^2\Bigg] \qquad (4.8)$$

where C_{jj}^{-1} are the diagonal elements of the matrix C^{-1}, the inverse of C. This may be written in semi matrix form as follows:

Define a vector Λ of length 2 with elements referring to a particular i, j combination:

$$\Lambda=\begin{pmatrix}\frac{\sigma_A^2}{A_i}(1+\frac{\sigma_A^2\sigma_B^2}{A_i}\sum_{j=1}^{r}\tilde{l}_{12}(ij)^2C_{jj}^{-1})\\ \sigma_B^2C_{jj}^{-1}\end{pmatrix}.$$

Then

$$l\approx\sum_{j=1}^{m}\sum_{i=1}^{r}\left[\tilde{l}_c-\frac{\tilde{a}_i^2}{2r\sigma_A^2}-\frac{\tilde{b}_j^2}{2m\sigma_B^2}\right]-\frac{1}{2}\sum_{i=1}^{m}\log(A_i)-\frac{1}{2}\log(|C|)$$

$$+\frac{1}{8}\sum_{j=1}^{m}\sum_{i=1}^{r}\left[\Lambda^T\begin{pmatrix}\tilde{l}_{1111} & \tilde{l}_{1122}\\ \tilde{l}_{1122} & \tilde{l}_{2222}\end{pmatrix}\Lambda+4\frac{\sigma_A^2\sigma_B^2}{A_i}\tilde{l}_{12}(ij)C_{jj}^{-1}\begin{pmatrix}\tilde{l}_{1112} & \tilde{l}_{1222}\end{pmatrix}\Lambda\right] \qquad (4.9)$$

Investigations of expression (4.3) for two random effects ($p = 2$)

As an example of the indepth form of expression (4.3) we now compare (4.8) above to our previous general expression. We will now take our approximation (4.3) and rewrite it in its explicit form; this is made easier by the fact that there are no terms with $J(2, 1, 1)$ or $J(1, 1, 1, 1)$ and some nice symmetries appear. Firstly rewrite matrix $\mathcal{D}$ in terms of A and B,

$$A_i = 1-\sigma_A^2\sum_{j=1}^{r}\tilde{l}_{11}(ij)$$

$$B_j = 1-\sigma_B^2\sum_{i=1}^{m}\tilde{l}_{22}(ij)$$

$$\mathcal{D} = \begin{pmatrix} \frac{1}{\sigma_A^2}\text{diag}(A_i) & -\tilde{l}_{12} \\ -\tilde{l}_{12}^T & \frac{1}{\sigma_B^2}\text{diag}(B_j) \end{pmatrix}$$

An expression for the inverse of $\mathcal{D}$ ($= \mathcal{D}^{-1}$) may be partitioned as

$$\begin{aligned} \mathcal{D}^{-1} &= \begin{pmatrix} \mathcal{D}_{11}^{-1} & \mathcal{D}_{12}^{-1} \\ \mathcal{D}_{21}^{-1} & \mathcal{D}_{22}^{-1} \end{pmatrix} \\ &= \begin{pmatrix} \sigma_A^2\text{diag}(A_i^{-1})\left(I + \sigma_A^2\sigma_B^2\tilde{l}_{12}C^{-1}\tilde{l}_{12}^T\text{diag}(A_i^{-1})\right) & \sigma_A^2\sigma_B^2\text{diag}(A_i^{-1})\tilde{l}_{12}C^{-1} \\ \sigma_A^2\sigma_B^2C^{-1}\tilde{l}_{12}^T\text{diag}(A_i^{-1}) & \sigma_B^2C^{-1} \end{pmatrix} \end{aligned}$$

Consider the term $\tilde{l}_{1111}$ in equation 4.8:

$$\tilde{l}_{1111} = d_{J(4)}\tilde{l}_c$$

Remember here that $J = (j_1, j_2, \ldots, j_{r+m})$, so for the term $\tilde{l}_{1111}$ $J(4)$ refers to some $j_i = 4$, for one i in $[1:m]$:

$$\begin{aligned} \tilde{l}_{1111}(\mathcal{D}_{11}^{-1}(ii))^2 &= \frac{\sigma_A^4}{A_i^2}\left\{1 + \sigma_A^2\sigma_B^2\sum_{j=1}^{r}\left[\sum_{k=1}^{r}\tilde{l}_{12}(ik)C_{kj}^{-1}\tilde{l}_{12}^T(ji)A_i^{-1}\right]\right\}^2 \\ &= \frac{\sigma_A^4}{A_i^2}\left\{1 + \frac{\sigma_A^2\sigma_B^2}{A_i}\sum_{j=1}^{r}\sum_{k=1}^{r}\left[\tilde{l}_{12}(ij)\tilde{l}_{12}(ik)C_{kj}^{-1}\right]\right\}^2. \end{aligned}$$

Similarly:

$$\tilde{l}_{2222}(\mathcal{D}_{22}^{-1}(jj))^2 = \sigma_B^4(C_{jj}^{-1})^2$$

$$\tilde{l}_{2221}\mathcal{D}_{22}^{-1}(jj)\mathcal{D}_{21}^{-1}(ji) = \sigma_B^2C_{jj}^{-1}\frac{\sigma_A^2\sigma_B^2}{A_i}\sum_{k=1}^{r}C_{jk}^{-1}\tilde{l}_{12}^T(ki)$$

$$\tilde{l}_{1112}\mathcal{D}_{11}^{-1}(ii)\mathcal{D}_{12}^{-1}(ij) = \frac{\sigma_A^2}{A_i}\left\{1 + \frac{\sigma_A^2\sigma_B^2}{A_i}\sum_{j=1}^{r}\sum_{k=1}^{r}\left[\tilde{l}_{12}(ij)\tilde{l}_{12}(ik)C_{kj}^{-1}\right]\right\}\frac{\sigma_A^2\sigma_B^2}{A_i}\sum_{k=1}^{r}\tilde{l}_{12}(ik)C_{kj}^{-1}$$

$$\tilde{l}_{1122}\left(\mathcal{D}_{11}^{-1}(ii)\mathcal{D}_{22}^{-1}(jj) + 2\mathcal{D}_{12}^{-1}(ij)\mathcal{D}_{21}^{-1}(ji)\right)$$

$$= \frac{\sigma_A^2}{A_i}\left\{1 + \frac{\sigma_A^2\sigma_B^2}{A_i}\sum_{j=1}^{r}\sum_{k=1}^{r}\left[\tilde{l}_{12}(ij)\tilde{l}_{12}(ik)C_{kj}^{-1}\right]\right\}\sigma_B^2 C_{jj}^{-1} + 2\frac{\sigma_A^4\sigma_B^4}{A_i^2}\left\{\sum_{k=1}^{r}\tilde{l}_{12}(ik)C_{kj}^{-1}\right\}^2$$

All these expression can be directly compared to (4.8), so for models with two crossed terms, these expression are equivalent.

4.2.2 Two way crossed with interaction

Models in this category, such as $y_{ijk} = \mu + a_i + b_j + c_{ij} + \varepsilon_{ijk}$ have two random effects a_i and b_j, as before, but now include a random interaction term c_{ij} which is distributed with mean zero and standard deviation σ_C. The likelihood for these type of models is of the form

$$\int\cdots\int\left[\int\cdots\int \exp\left\{\sum_{i=1}^{m}\sum_{j=1}^{r}\left(\tilde{l}_c - \frac{a_i^2}{2r\sigma_A^2} - \frac{b_j^2}{2m\sigma_B^2} - \frac{c_{ij}^2}{2\sigma_C^2}\right)\right\} dc_{11}...dc_{mr}\right] da_1...da_m db_1...db_r$$

and we need to integrate out the three random effects by making an expansion around three terms etc.

Consider the second-order terms. Firstly integrate out the c_{ij} variables to get an approximate integral

$$\prod_{ij}\left[\sqrt{\frac{\sigma_C^2}{1-\tilde{l}_{33}\sigma_C^2}}\exp\left(\frac{\sigma_C^2(\tilde{l}_{13}(a_i-\tilde{a}_i)+\tilde{l}_{23}(b_j-\tilde{b}_j))^2}{2(1-\tilde{l}_{33}\sigma_C^2)}\right)\times\right.$$

$$\exp\left[-\frac{1}{2}\left\{(a_i-\tilde{a}_i)^2\left(\frac{1-\tilde{l}_{11}(ij)r\sigma_A^2}{r\sigma_A^2}\right) + (b_j-\tilde{b}_j)^2\left(\frac{1-\tilde{l}_{22}(ij)m\sigma_B^2}{m\sigma_B^2}\right)\right.\right.$$

$$\left.\left.\left. - 2\tilde{l}_{12}(ij)(a_i-\tilde{a}_i)(b_j-\tilde{b}_j)\right\}\right]\right]$$

Now define

$$A1_i = \sum_{j=1}^{r}\left(\frac{1-\tilde{l}_{11}(ij)r\sigma_A^2}{r\sigma_A^2} - \frac{\tilde{l}_{13}(ij)^2\sigma_C^2}{1-\tilde{l}_{33}(ij)\sigma_C^2}\right)$$

and

$$B1_j = \sum_{i=1}^{m}\left(\frac{1-\tilde{l}_{22}(ij)m\sigma_B^2}{m\sigma_B^2} - \frac{\tilde{l}_{23}(ij)^2\sigma_C^2}{1-\tilde{l}_{33}(ij)\sigma_C^2}\right)$$

and the approximation becomes

$$\int\cdots\int \exp -\frac{1}{2}\left\{\sum_{i=1}^{m}(a_i-\tilde{a}_i)^2 A1_i + \sum_{j=1}^{r}(b_j-\tilde{b}_j)^2 B1_j - 2\sum_{i=1}^{m}\sum_{j=1}^{r}(a_i-\tilde{a}_i)(b_j-\tilde{b}_j)\right.$$

$$\left.\times\ (\tilde{l}_{12}(ij)+\frac{\tilde{l}_{13}(ij)\tilde{l}_{23}(ij)\sigma_C^2}{1-\tilde{l}_{33}(ij)\sigma_C^2})\right\} da_1 \ldots da_m db_1 \ldots db_r. \tag{4.10}$$

This is of the same form as (4.4), so the procedure from here follows exactly the same method as before. Remembering that each second derivative of the conditional log likelihood has $(m \times r)$ terms corresponding to each ij combination, we can define an $(m \times r)$ matrix $L2$ with elements $L2_{ij}$ given by

$$L2_{ij} = \tilde{l}_{12}(ij) + \frac{\tilde{l}_{13}(ij)\tilde{l}_{23}(ij)\sigma_C^2}{1-\tilde{l}_{33}(ij)\sigma_C^2}.$$

Now define a new matrix $C1$ as

$$C1 = \text{diag}(B1_j) - \sigma_A^2\sigma_B^2 L2^T \text{diag}(A1_i^{-1}) L2$$

and the second-order Laplace approximation is

$$\sum_{i=1}^{m}\sum_{j=1}^{r}\left(\tilde{l}_c - \frac{\tilde{a}_i^2}{2r\sigma_A^2} - \frac{\tilde{b}_j^2}{2m\sigma_B^2} - \frac{\tilde{c}_{ij}^2}{2\sigma_C^2}\right) - \frac{1}{2}\sum_{i=1}^{m}\log(A1_i) - \frac{1}{2}\log(|C1|). \tag{4.11}$$

The fourth-order terms may be found explicitly by following the procedure outlined previously, or using the general formula with $\mathcal{D}$ defined using a mixture of nested and crossed terms.

4.2.3 Models with two nested random effects

The standard normal theory linear two-way nested model, with b_{ij} nested in a_i, may be written $y_{ijk} = \mu + a_i + b_{ij} + \varepsilon_{ijk}$ with $i = 1, \ldots, m$, $j = 1, \ldots, r_i$ and $k = 1, \ldots, n$. The a_i and b_{ij} are random effects with $a_i \sim \text{N}(0, \sigma_A^2)$ independent of $b_{ij} \sim \text{N}(0, \sigma_B^2)$, $\varepsilon \sim \text{N}(0, \sigma^2)$ is the random error term.

The true likelihood in the case of a nested model is

$$
\begin{aligned}
L &= \prod_{i=1}^{m}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\sigma_A^2}}e^{-\frac{a_i^2}{2\sigma_A^2}}\prod_{j=1}^{r_i}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\sigma_B^2}}e^{-\frac{b_{ij}^2}{2\sigma_B^2}}e^{l_c}db_{ij}da_i \\
&= \prod_{i=1}^{m}\int_{-\infty}^{\infty}\left[\prod_{j=1}^{r_i}\int_{-\infty}^{\infty}(2\pi\sigma_A^2)^{-\frac{r_i}{2}}(2\pi\sigma_B^2)^{-\frac{1}{2}}e^{l_c-\frac{b_{ij}^2}{2\sigma_B^2}-\frac{a_i^2}{2r_i\sigma_A^2}}db_{ij}\right]da_i
\end{aligned}
$$

where l_{ij} is the conditional log likelihood.

The true log likelihood for a normal theory linear model (balanced) is

$$
-\frac{mrn}{2}\log(2\pi)-\frac{mr(n-1)}{2}\log(\sigma^2)-\frac{m}{2}\log(nr\sigma_A^2+n\sigma_B^2+\sigma^2)
$$

$$
-\frac{m(r-1)}{2}\log(n\sigma_B^2+\sigma^2)-\frac{1}{2\sigma^2}\left\{\sum_{i,j,k}(y_{ijk}-\mu)^2-\frac{n^2\sigma_B^2\sum_{i,j}(\bar{y}_{ij.}-\mu)^2}{n\sigma_B^2+\sigma^2}\right.
$$

$$
\left.-\frac{n^2\sigma_A^2\sum_{i=1}^{m}\left(\sum_{j=1}^{r}(\bar{y}_{ij.}-\mu)\right)^2}{(n\sigma_B^2+\sigma^2)(nr\sigma_A^2+n\sigma_B^2+\sigma^2)}\right\}
$$

as given by Searle et al. (1992).

Random effects may enter non-linearly, which causes the integral representing the true likelihood to be intractable. The integral may be approximated with a Laplace approximation using a Taylor series expansion for the exponent. Similarly to the crossed case, the expansion is

$$
\begin{aligned}
&\tilde{l}_c-\frac{\tilde{b}_{ij}^2}{2\sigma_B^2}-\frac{\tilde{a}_i^2}{2r_i\sigma_A^2}+(\tilde{l}_1-\frac{\tilde{a}_i^2}{r_i\sigma_A^2})(a_i-\tilde{a}_i)+(\tilde{l}_2-\frac{\tilde{b}_{ij}}{\sigma_B^2})(b_{ij}-\tilde{b}_{ij}) \\
&+\frac{1}{2}(\tilde{l}_{11}(ij)-\frac{1}{r_i\sigma_A^2})(a_i-\tilde{a}_i)^2+\frac{1}{2}(\tilde{l}_{22}(ij)-\frac{1}{\sigma_B^2})(b_{ij}-\tilde{b}_{ij})^2+\tilde{l}_{12}(ij)(a_i-\tilde{a}_i)(b_{ij}-\tilde{b}_{ij}) \\
&+\frac{1}{24}\tilde{l}_{1111}(a_i-\tilde{a}_i)^4+\frac{1}{24}\tilde{l}_{2222}(b_{ij}-\tilde{b}_{ij})^4+\frac{1}{4}\tilde{l}_{1122}(a_i-\tilde{a}_i)^2(b_{ij}-\tilde{b}_{ij})^2 \\
&+\frac{1}{6}\tilde{l}_{1112}(a_i-\tilde{a}_i)^3(b_{ij}-\tilde{b}_{ij})+\frac{1}{6}\tilde{l}_{1222}(a_i-\tilde{a}_i)(b_{ij}-\tilde{b}_{ij})^3
\end{aligned}
$$

As in the crossed case, $\tilde{l}_c=l_c(\tilde{a}_i,\tilde{b}_{ij})$ i.e. the conditional log likelihood (fixed a_i and b_{ij}) substituting in the maximizing estimates $\tilde{a}_i$ and $\tilde{b}_{ij}$. Similarly for all conditional log likelihood derivatives.

Firstly integrate w.r.t. b_{ij}:

$$\int \exp\left\{-\frac{1}{2}\left[(b_{ij}-\tilde{b}_{ij})^2(\frac{1}{\sigma_B^2}-\tilde{l}_{22}(ij)) \quad - \quad 2(b_{ij}-\tilde{b}_{ij})\tilde{l}_{12}(ij)(a_i-\tilde{a}_i) \right.\right.$$
$$\left.\left. + \quad (a_i-\tilde{a}_i)^2(\frac{1}{r_i\sigma_A^2}-\tilde{l}_{11}(ij))\right]\right\} db_{ij}$$

$$= \int \exp\left\{-\frac{1}{2}\left[\left(\frac{\sqrt{1-\tilde{l}_{22}(ij)\sigma_B^2}}{\sigma_B}(b_{ij}-\tilde{b}_{ij}) - \frac{\tilde{l}_{12}(ij)\sigma_B(a_i-\tilde{a}_i)}{\sqrt{1-\tilde{l}_{22}(ij)\sigma_B^2}}\right)^2\right.\right.$$
$$\left.\left. - \frac{\tilde{l}_{12}^2(ij)\sigma_B^2(a_i-\tilde{a}_i)^2}{1-\tilde{l}_{22}(ij)\sigma_B^2} + (a_i-\tilde{a}_i)^2(\frac{1}{r_i\sigma_A^2}-\tilde{l}_{11}(ij))\right]\right\} db_{ij}$$

Now integration yields

$$\frac{\sqrt{2\pi}\sigma_B}{\sqrt{1-\tilde{l}_{22}(ij)\sigma_B^2}} \exp\left\{-\frac{1}{2}(a_i-\tilde{a}_i)^2\left[\frac{1-\tilde{l}_{11}(ij)r_i\sigma_A^2}{r_i\sigma_A^2} - \frac{\tilde{l}_{12}^2\sigma_B^2}{1-\tilde{l}_{22}(ij)\sigma_B^2}\right]\right\}$$

Now multiply by $(2\pi\sigma_B^2)^{-\frac{1}{2}}$ and take the product over j to get

$$\prod_{j=1}^{r_i}\frac{1}{\sqrt{1-\tilde{l}_{22}(ij)\sigma_B^2}} \exp\sum_{j=1}^{r_i} -\frac{1}{2}(a_i-\tilde{a}_i)^2\left[\frac{1-\tilde{l}_{11}(ij)r_i\sigma_A^2}{r_i\sigma_A^2} - \frac{\tilde{l}_{12}^2\sigma_B^2}{1-\tilde{l}_{22}(ij)\sigma_B^2}\right]$$

Integrate over a_i

$$\prod_{i=1}^{m}\left\{\prod_{j=1}^{r_i}\frac{1}{\sqrt{1-\tilde{l}_{22}(ij)\sigma_B^2}}\left[\sum_{j=1}^{r_i}\frac{(1-\tilde{l}_{11}(ij)r_i\sigma_A^2)(1-\tilde{l}_{22}(ij)\sigma_B^2) - \tilde{l}_{12}^2\sigma_B^2\sigma_A^2 r_i}{r_i(1-\tilde{l}_{22}(ij)\sigma_B^2)}\right]^{-\frac{1}{2}}\right\}$$

So the second-order approximation to the true log likelihood becomes

$$\sum_{i,j}\left(\tilde{l}_c - \frac{\tilde{b}_{ij}^2}{2\sigma_B^2} - \frac{\tilde{a}_i^2}{2r_i\sigma_A^2} - \frac{1}{2}\log\mathcal{B}\right) - \sum_{i=1}^{m}\frac{1}{2}\log(\mathcal{C}) \tag{4.12}$$

where

$$\mathcal{B} = 1-\tilde{l}_{22}(ij)\sigma_B^2$$
$$\mathcal{C} = \sum_{j=1}^{r}\frac{(1-\tilde{l}_{11}(ij)r_i\sigma_A^2)\mathcal{B} - \tilde{l}_{12}^2\sigma_B^2\sigma_A^2 r_i}{\mathcal{B}r_i}$$

Now, making the substitutions

$$\begin{aligned} b_{ij} - \tilde{b}_{ij} &= \frac{\sigma_B}{\sqrt{\mathcal{B}}} Y_{ij} + \frac{\tilde{l}_{12}\sigma_B^2}{\mathcal{B}}(a_i - \tilde{a}_i) \\ a_i - \tilde{a}_i &= \mathcal{C}^{-\frac{1}{2}} X_i \end{aligned}$$

into the Taylor expansion for the higher-order terms, we get an expression for the fourth-order Laplace approximation to the true log likelihood:

$$\begin{aligned} &\sum_{i,j} \left(\tilde{l}_c - \frac{\tilde{b}_{ij}^2}{2\sigma_B^2} - \frac{\tilde{a}_i^2}{2r_i\sigma_A^2} - \frac{1}{2}\log\mathcal{B} \right) + \sum_{i=1}^{m} -\frac{1}{2}\log(\sigma_A^2\mathcal{C}) \\ &+ \sum_{i,j} \frac{1}{8\mathcal{B}^4\mathcal{C}^2} \left[\tilde{l}_{1111}\mathcal{B}^4 + 4\tilde{l}_{1112}\tilde{l}_{12}\sigma_B^2\mathcal{B}^3 + 2\tilde{l}_{1122}\sigma_B^2\mathcal{B}^2(\mathcal{C}\mathcal{B} + 3\tilde{l}_{12}^2\sigma_B^2 \right. \\ &+ \left. 4\tilde{l}_{1222}\tilde{l}_{12}\sigma_B^4\mathcal{B}(\mathcal{B}\mathcal{C} + \tilde{l}_{12}^2\sigma_B^2) + \tilde{l}_{2222}(\mathcal{C}\mathcal{B}\sigma_B^2 + \tilde{l}_{12}^2\sigma_B^4)^2 \right] \end{aligned} \tag{4.13}$$

4.2.4 Expansion around the true values

Solomon & Cox (1992) find an approximation to the true likelihood by taking the Taylor series expansion around zero, i.e. the true values, using a fourth-order approximation. For a crossed model with two independent random effects, the second-order approximation is

$$\sum_{i=1}^{m}\sum_{j=1}^{r}\tilde{l}_c - \frac{1}{2}\log(|C|) - \frac{1}{2}\sum_{i=1}^{m}\log(A_i) + \frac{\sigma_A^2}{2}J_{1\times r}\tilde{l}_1^T\mathrm{diag}(A_i^{-1})\tilde{l}_1 J_{r\times 1} + \frac{\sigma_B^2}{2}K^T C^{-1} K \tag{4.14}$$

where C is the $(m \times m)$ symmetric matrix

$$C = \mathrm{diag}(B_j) - \sigma_A^2\sigma_B^2\tilde{l}_{12}^T\mathrm{diag}(A_i^{-1})\tilde{l}_{12},$$

K is a $(r \times 1)$ vector:

$$K = \sigma_A^2\tilde{l}_{12}^T\mathrm{diag}(A_i^{-1})\tilde{l}_1 + \tilde{l}_2$$

and $\tilde{l}_c$ is evaluated at $a_i = 0, b_j = 0$.

The second-order approximation to the log likelihood for a nested model with two random effects is

$$\sum_{i=1}^{m}\sum_{j=1}^{r_i}\left(\tilde{l}_c - \frac{1}{2}\log\mathcal{B}\right) - \sum_{i=1}^{m}\frac{1}{2}\log(\mathcal{C})$$

$$+\frac{1}{2}\sum_{i=1}^{m}\left\{\frac{\sigma_A^2(\sum_{j=1}^{r_i}(\tilde{l}_1\mathcal{B}+\sigma_B^2\tilde{l}_2\tilde{l}_{12})/\mathcal{B})^2}{\mathcal{C}} + \sum_{j=1}^{r_i}\frac{\tilde{l}_2^2\sigma_B^2}{\mathcal{B}}\right\}.$$

where

$$\mathcal{B} = 1 - \tilde{l}_{22}\sigma_B^2$$

$$\mathcal{C} = \sum_{j=1}^{r_i}\frac{(1-\tilde{l}_{11}r_i\sigma_A^2)\mathcal{B} - \tilde{l}_{12}^2\sigma_B^2\sigma_A^2 r_i}{\mathcal{B}r_i}$$

also defined before.

Lin & Breslow (1996) generalize these forms and discuss Solomon-Cox approximations for GLMMs with multiple components of dispersion, and use it as a "stepping stone" to derive their bias correction procedure.

These approximations have the advantage that there is no need to calculate the maximizers $\tilde{a}_i$ and $\tilde{b}_j$. One problem is that to expand beyond the second derivative is not as straight forward as the Laplace approximation since the expected values for odd powers do not cancel and thus the expression becomes very complicated.

4.3 Using approximations to the log-likelihood in estimation of effects in mixed models

Firstly take the Laplace approximations. To estimate the components of some (known) model we take the conditional log-likelihood for that model, choose starting values, estimate the local maximisers (for example $\tilde{a}_i = \tilde{l}_1(i,j)\sigma_A^2$ and $\tilde{b}_j = \tilde{l}_2(i,j)\sigma_B^2$) as functions of the starting values, substitute these into the approximate log-likelihood function, maximise this to obtain new estimates of the model components, use these to find new local maximisers and so on until the iterative process converges. This process is similar to a Newton-Raphson type algorithm.

From the expansion around the true values, as mentioned previously, to estimate components of a model we avoid the internal process of finding the local maximum at each level and use the values from the previous iteration, similar to an EM algorithm approach.

Lindstrom & Bates (1988) compare the Newton-Raphson and EM algorithms. They conclude that although the EM will converge, it will take many more iterations than the Newton-Raphson. We can make a comparison here with the two methods of likelihood approximations and conclude that the Laplace may be preferable when estimating components from a model. For this reason the simulation Chapter which follows considers the Laplace approximation only.

Chapter 5

Simulations of the likelihood function

To illustrate properties of the likelihood approximation, I will use it to approximate all the fixed effects and variance components for four general models. The examples I will use are

1. linear model – this will enable comparison with known results,
2. Poisson,
3. a logistic regression simulation including an example from the Salamander mating experiment,
4. and a normal nested model with exponential mean.

Suppose, for example, our model contains one fixed effect and two random effects. The parameters we wish to estimate are $(\mu, \sigma_A^2, \sigma_B^2)$. The estimation procedure may be considered as two stage. The steps are as follows:

- Firstly fix initial values for $(\mu, \sigma_A^2, \sigma_B^2)$

1 Find $\tilde{a}_i, \tilde{b}_j$ as functions of $(\tilde{\mu}, \tilde{\sigma}_A^2, \tilde{\sigma}_B^2)$

2 Substitute these into expression for the Laplace approximation and maximize this to find estimates $(\tilde{\mu}, \tilde{\sigma}_A^2, \tilde{\sigma}_B^2)$

- Repeat stages 1 and 2 until the approximate log likelihood is a maximum.

5.1 Two-way linear crossed model

To investigate the behaviour of the approximations to the likelihood, firstly consider the normal theory linear model. The approximation for this model is simplified because the third and fourth-order derivatives of the conditional log likelihoods are zero, and so we will only need to consider up to two derivatives. A linear crossed model with 2 random effects and no interaction is

$$y_{ijk} = \mu + a_i + b_j + \varepsilon_{ijk}, \text{with } i = 1, .., m \quad j = 1, ..., r \quad k = 1, ..., n$$

where $a \sim N(0, \sigma_A^2)$, $b \sim N(0, \sigma_B^2)$ are the random effects and $\varepsilon \sim N(0, \sigma^2)$ is the random error term. We simulated data using this model with $r = m = 20$, $n = 2$ and found the Laplace approximation to the log likelihood using the conditional log likelihood

$$l_c(y_{ijk}|a_i, b_j) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{\sum_k (y_{ijk} - \mu - a_i - b_j)^2}{2\sigma^2}.$$

The estimated values of $\mu, \sigma_A, \sigma_B, \sigma$ were compared with the results obtained from a random effects analysis of variance (with maximum likelihood method). Table 5.1 shows estimates from two different sets of simulated data. They compare reasonably well. The estimates of the fixed effects are identical, but there is a tendency for the variance components σ_A^2 and σ_B^2 to be slightly underestimated by the approximations.

Table 5.1: Estimates from a linear model with two random effects

	μ	σ_A^2	σ_B^2	σ^2
Approximate likelihood	1.07471	0.93751	0.58527	1.037188
RAOV	1.07471	0.96906	0.59796	1.037165
Approximate likelihood	0.42698	0.09894	0.14636	0.41794
RAOV	0.42698	0.10141	0.15142	0.41792

5.2 A Poisson model

Consider Poisson simulated data with two independent crossed random effects, $a_i \sim N(0, \sigma_A^2)$ and $b_j \sim N(0, \sigma_B^2)$, and $\mathrm{E}(y_{ij}|a_i, b_j) = \exp(\mu + a_i + b_j)$.

The distribution is Poisson with conditional log likelihood given by

$$l_c(y_{ij}|a_i, b_j) = y_{ij}(\mu + a_i + b_j) + \exp(\mu + a_i + b_j) - \log(y_{ij}!).$$

From here:

$$\begin{aligned}
\tilde{l}_1 &= y_{ij} + \exp(\mu + \tilde{a}_i + \tilde{b}_j) \\
\tilde{l}_{11} &= \exp(\mu + \tilde{a}_i + \tilde{b}_j) \\
&\quad \text{and} \\
\tilde{l}_{1111} &= \exp(\mu + \tilde{a}_i + \tilde{b}_j)
\end{aligned}$$

The expression for a Laplace approximation to the true log likelihood has the form (4.8) with

$$\begin{aligned}
A_i &= 1 - \sum_j \exp(\mu + \tilde{a}_i + \tilde{b}_j)\sigma_A^2 \\
B_j &= 1 - \sum_i \exp(\mu + \tilde{a}_i + \tilde{b}_j)\sigma_B^2
\end{aligned}$$

Table 5.2: Estimates of the variance components for simulated Poisson data with various starting values

Simulated values			2nd order				4th order			
μ	σ_A	σ_B	μ	σ_A	σ_B	Log-lik	μ	σ_A	σ_B	Log-lik
1	1	1	1.711	0.962	0.706	-9227.734	1.711	0.962	0.706	-9227.671
1	1	1	1.220	0.936	0.999	-6327.444	1.221	0.936	0.999	-6327.349
2	0.707	1	2.093	0.656	1.009	-18390.44	2.093	0.656	1.009	-18390.40
3	0.316	0.316	2.932	0.305	0.321	-17436.00	2.932	0.305	0.321	-17435.98

Table 5.2 shows estimated variance components from the second and fourth-order likelihood approximations for four simulated datasets with $m = r = 20$ and values for the fixed and random terms as given in the table. Most estimated values appear close to the true (simulated) values. There is little to choose between the second and fourth-order approximations in these simulations. μ is over-estimated by both methods. Note though that the log likelihood is slightly higher for the fourth-order approximations in each case.

The Laplace approximation to log likelihood for simulated Poisson data with values $\sigma_A^2 = \sigma_B^2 = 0.1$, $m = r = 20$ is shown graphically by figures 5.1 and 5.2. The estimated values are $\tilde{\sigma_A}^2 = 0.074$ and $\tilde{\sigma_B}^2 = 0.106$. The plots show a well defined curve and definite maximising near the true values.

5.3 A nested model

Consider normal data y_{ijk} with two nested random effects, a_i and b_{ij} and mean $\mathrm{E}(y_{ijk}|a_i, b_{ij}) = \exp(\mu + a_i + b_{ij})$ and variance σ^2. The conditional log likelihood is written

$$l_c = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_k (y_{ijk} - \exp(\mu + a_i + b_{ij}))^2 \quad (5.1)$$

The derivatives of the log likelihood are

$$
\begin{aligned}
l_1 &= l_2 = \frac{1}{\sigma^2}\sum_k \exp(\mu + a_i + b_{ij})\,(y_{ijk} - \exp(\mu + a_i + b_{ij})) \\
l_{11} &= l_{12} = l_{22} = l_1 - \frac{n}{\sigma^2}\exp(\mu + a_i + b_{ij})^2 \\
l_{1111} &= l_{11} - \frac{6n}{\sigma^2}\exp(\mu + a_i + b_{ij})^2 \\
l_{1112} &= l_{1122} = l_{1222} = l_{2222}
\end{aligned}
$$

Data were simulated over a range of variance components using $r = m = 20, n = 5, \sigma^2 = 0.1$ and $\mu = 5$. For each set of variance components, the same simulated set of data was used to find both the second-order approximation and the fourth-order approximation to the log likelihood by substituting these into equations (4.12) and (4.13) respectively. The results are shown in table 5.3. The estimated values appear to be close to the simulated values and, for this model at least, a second order approximation is enough.

Table 5.3: Variance component estimates for the exponential nested model for both a second-order Laplace approximation and a fourth-order approximation.

		second-order		fourth-order	
σ_A	σ_B	$\tilde{\sigma}_A$	$\tilde{\sigma}_B$	$\tilde{\sigma}_A$	$\tilde{\sigma}_B$
1	1	0.876	1.031	0.876	1.031
1	0.71	1.036	0.747	1.036	0.747
0.71	1	0.662	1.026	0.662	1.026
0.71	0.71	0.744	0.756	0.744	0.756
0.5	0.71	0.479	0.719	0.479	0.719
0.5	0.5	0.485	0.538	0.485	0.538
0.71	0.5	0.726	0.498	0.726	0.498

In figure 5.3, data were simulated with $\sigma_A^2 = 1, \sigma_B^2 = 1, \sigma^2 = 0.1, \mu = 1, m = 20, r = 20, n = 5$. The approximation to the log likelihood function for these data was plotted against its parameters σ_A^2 and σ_B^2. The function appears to be maximizing near the true values of the parameters, although the log likelihood functions are rather flat for values greater than the maximum.

5.4 The Salamander mating data

Shun (1997) used a Laplace approximation with data from the salamander mating experiment, where ten male and ten female salamanders from two populations (rough butt (RB) and white side (WS)) were mated. The full set of data is described in McCullagh and Nelder (1989 page 440). We wish to use these data to find estimates for the variance components, ignoring the populations and assuming 20 independent males and females (random).

This binary y_{ij} is a measure of success or failure (0=failure, 1=success) when mating female $f_i, i = 1, \ldots, 20$ with male $m_j, j = 1, \ldots, 20$. The model is

$$\text{logit}(P(y_{ij} = 1|f_i, m_j)) = X\alpha + f_i + m_j, \tag{5.2}$$

where $f_i \sim N(0, \sigma_f^2)$ and $m_j \sim N(0, \sigma_m^2)$ are the random effects, and α is the vector of fixed effects, $\alpha = (\mu, WS_f, WS_m, WS_f \times WS_m)$. The conditional log likelihood is

$$l_c(y_{ij}|f_i, m_j) = y_{ij}(X\alpha + f_i + m_j) - \log(1 + \exp(X\alpha + f_i + m_j)). \tag{5.3}$$

The derivatives are

$$\begin{aligned} l_{11} &= l_{12} = l_{22} = -K + K^2 \\ l_{1111} &= -K + 7K^2 - 12K^3 + 6K^4 \\ &= l_{1112} = l_{1122} = l_{1222} = l_{2222} \end{aligned}$$

where

$$K = \frac{\exp(X\alpha + f_i + m_j)}{1 + \exp(X\alpha + f_i + m_j)}.$$

The estimated values for both the fixed and random effects are shown in table 5.4. The estimated standard errors for the random effects were calculated numerically by fitting a quadratic regression through the approximate log-likelihood function.

Table 5.4: Estimates of parameters for the Salamander mating experiment

	second-order	fourth-order
μ	1.335	1.330
WS_f	-2.940	-2.935
WS_m	-0.422	-0.436
$WS_f \times WS_m$	3.181	3.196
$\sigma_f \pm SE$	1.255±.63	1.259±.67
$\sigma_m \pm SE$	0.269±.79	0.344±.75

The second-order estimates are comparable to the estimates obtained from past authors, see Shun (1997) Uncorrected Summer '86 results. Contour plots of the two approximate likelihoods are shown in figure 5.4. From this we can see that the two are very close, although again the likelihood is better defined for the fourth-order approximation.

Table 5.5 was produced using simulated data with 20 females ($m = 20$), 20 males ($r = 20$) and $\mu = 1$. In most cases the estimates are greater than the simulated values. There is only a slight difference between the fourth-order and second-order estimates in about the third decimal place.

Table 5.5: Variance component estimates for the logit model for both a second-order Laplace approximation and a fourth-order approximation.

		fourth-order		second-order	
σ_A	σ_B	$\tilde{\sigma}_A^2$	$\tilde{\sigma}_B^2$	$\tilde{\sigma}_A^2$	$\tilde{\sigma}_B^2$
$m = r = 20, \mu = 1$					
1	1	1.230	1.250	1.240	1.258
1	0.1	1.552	0.222	1.584	0.224
0.1	1	0.218	1.144	0.217	1.153
0.5	0.5	0.675	1.039	0.672	1.040
0.2	0.5	0.441	0.569	0.440	0.569
0.1	0.5	0.171	0.589	0.169	0.585
0.1	0.2	0.244	0.301	0.244	0.301
$m = r = 20, \mu = 2$					
0.5	0.5	1.104	0.697	1.157	0.724
0.2	0.5	0.349	0.765	0.361	0.793
$m = 40, r = 10, \mu = 2$					
0.2	0.5	0.517	0.771	0.537	0.794
0.2	0.2	0.665	0.395	0.697	0.408
0.1	0.1	0.333	0.209	0.340	0.212

Figure 5.1: Profile plots for Poisson simulated data with $\sigma_A^2 = 0.1$ and $\sigma_B^2 = 0.1$

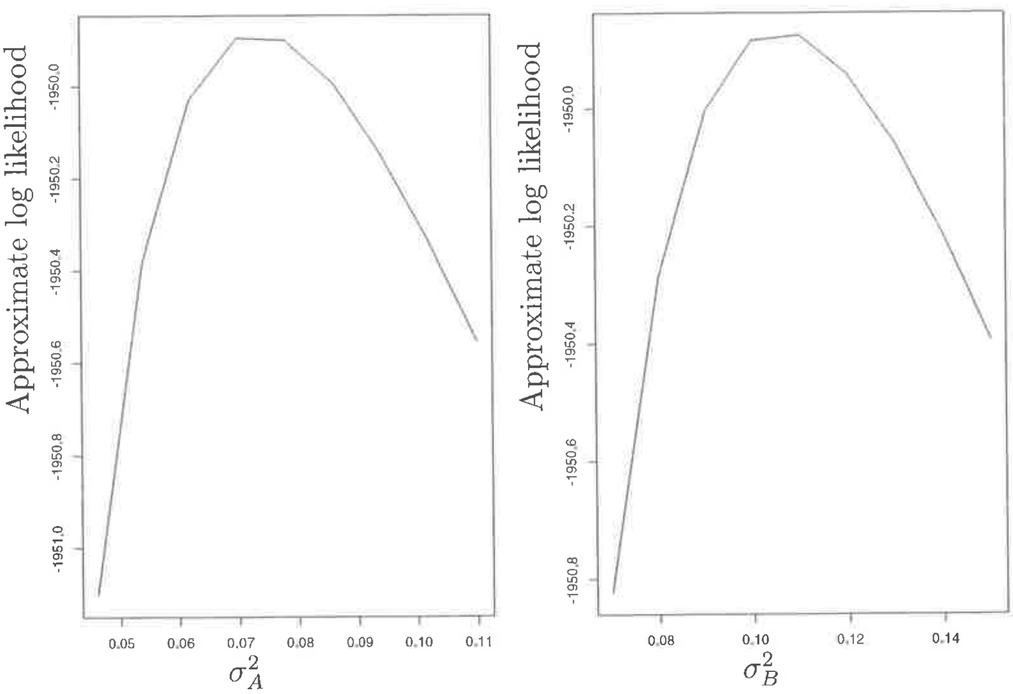


Figure 5.2: Approximate log likelihood for Poisson simulation model with $\sigma_A^2 = 0.1$ and $\sigma_B^2 = 0.1$.

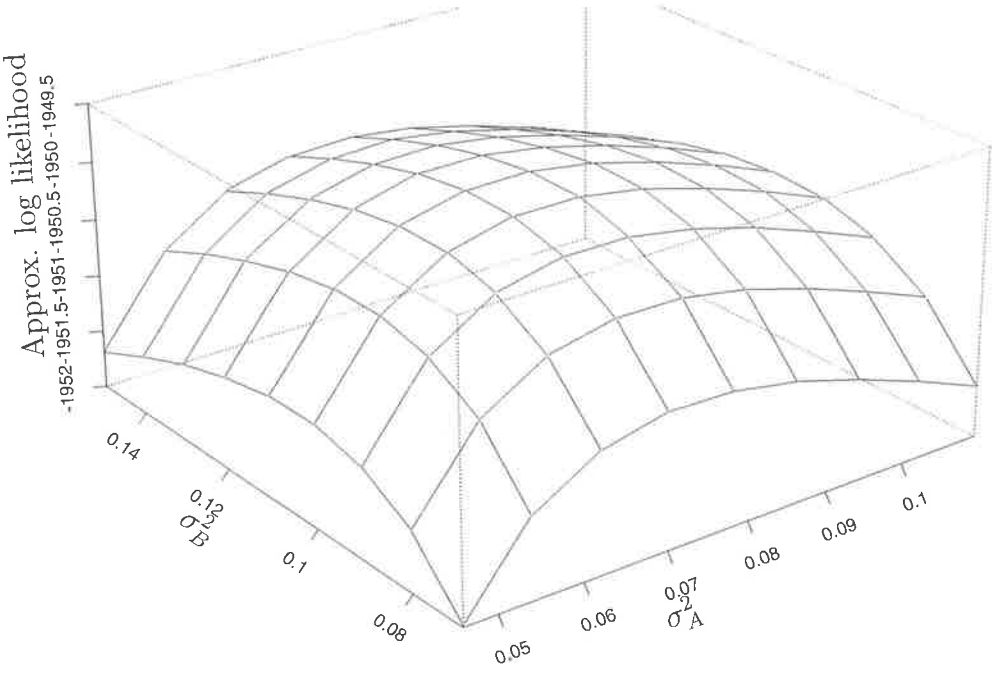


Figure 5.3: Profiles and 3D perspective plots showing the approximation to the log likelihood for data simulated from the nested model with exponential mean and $\sigma_A^2 = 1, \sigma_B^2 = 1$.

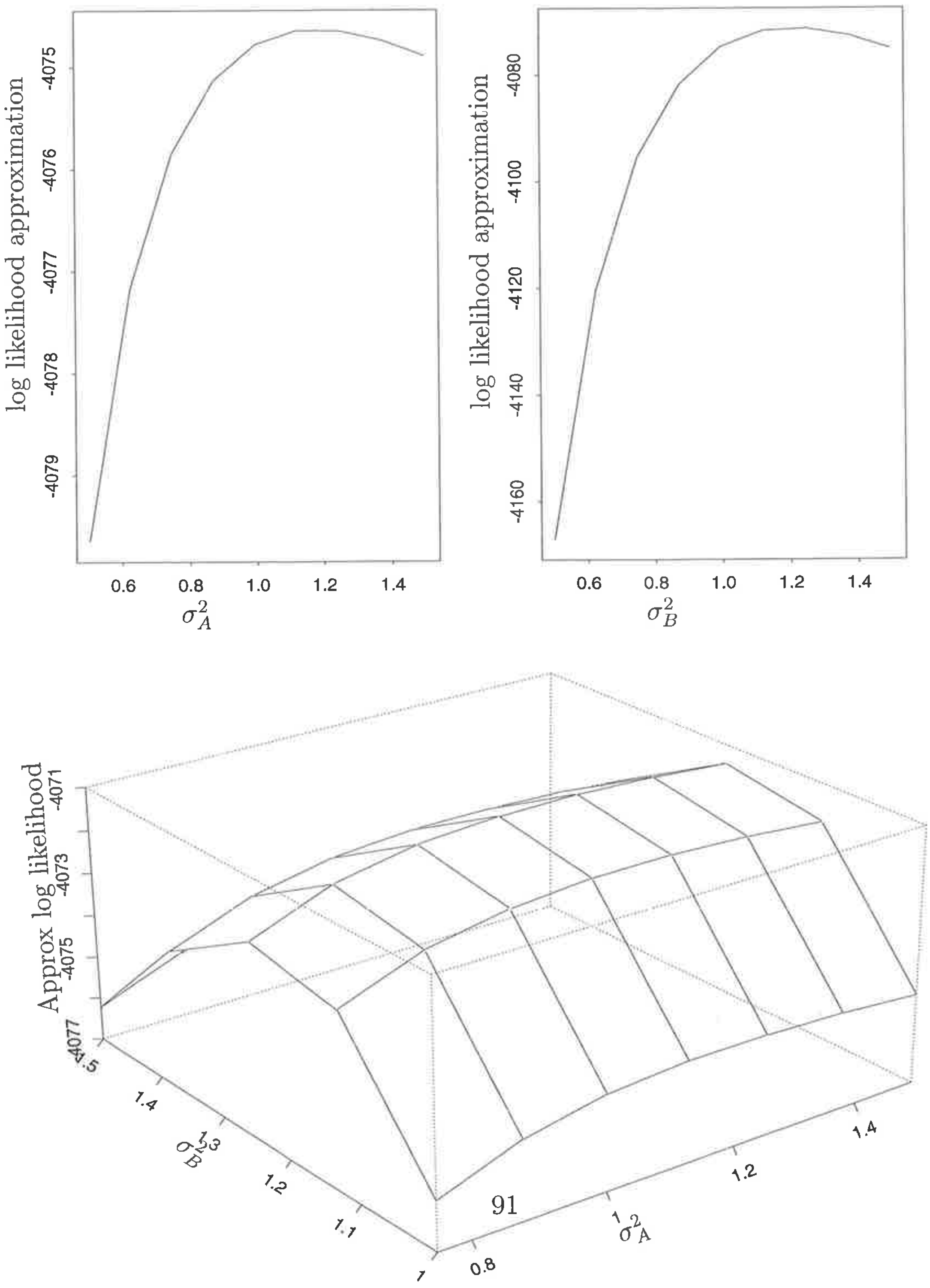
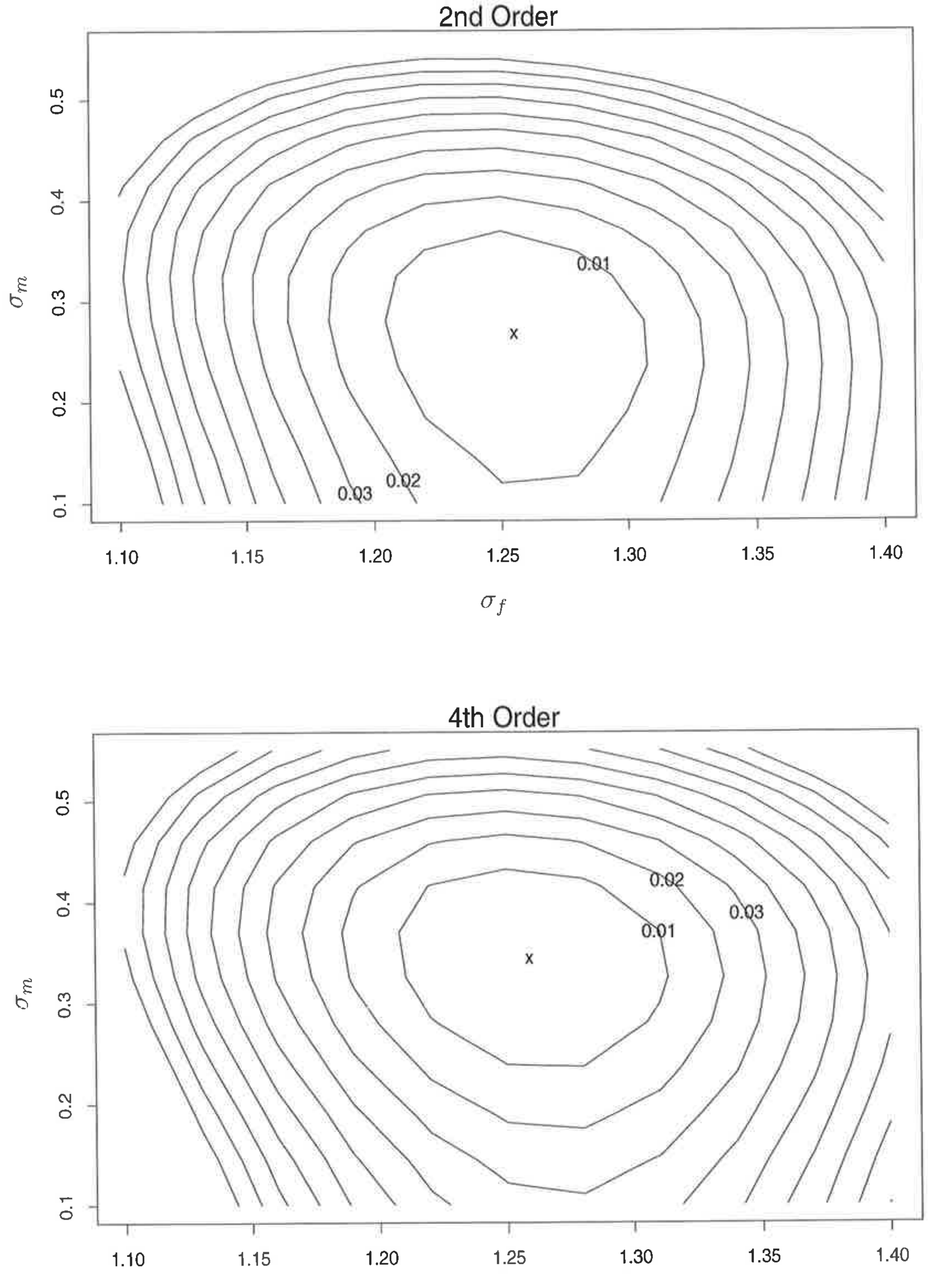


Figure 5.4: Contours of approximate log-likelihood functions for Salamander data.



Appendix A

Splus functions

A.1 Score test statistics

```
mevar1<-function(data)
{
#y = data in matrix form, mu = grand mean
#s2 = error variance, s2a = eta variance
#mm = number of individuals, r = number of reps
#this functions calculates my derived score test based on exact
#likelihood
        size <- dim(data)
        mm <- size[1]
        r <- size[2]
        s2 <- sum((data - apply(data, 1, mean))^2)/(mm * (r - 1))
        s2a <- sum((apply(data, 1, mean) - mean(data))^2)/(mm - 1) -
                s2/r
        mu <- mean(data)
        A <- s2 + r * s2a
```

```
    sumy <- apply(data, 1, sum) - r * mu
    sumy2 <- apply((data - mu)^2, 1, sum)
    bot1 <- A^2 * sumy2 * (A * s2 + sumy^2 * s2a) + 3 * r * s2^2
            * s2a * A^2 - 6 * s2a * sumy^2 * A * s2^2 - sumy^
            4 * s2a^2 * (r * s2a + 2 * s2)
    N2 <- (s2a * (A^2 * (sumy2 - r * s2) - s2a * sumy^2 * (4 *
            s2 + r * s2a))^2)/(s2 * A^5)
    NP <- (r * s2a^2 * (A^2 * (sumy2 - r * s2) - s2a * sumy^2 *
            (4 * s2 + r * s2a)))/A^4
    O <- (s2a * sumy * (-2 * s2 + r * s2a))/A^2
    M <- (s2a * sumy * (A^2 * (sumy2 - r * s2) - s2a * sumy^2 *
            (r * s2a + 2 * s2)))/(s2 * A^3)
    P2 <- (r^2 * s2 * s2a^3)/A^3
    totop <- sum(O + M)
    bot2 <- sum(N2) + 6 * sum(NP) + 15 * mm * P2 + 2 * sum(O^2)
    tobot <- (2 * s2a * sum(bot1))/(A^4 * s2) - bot2
    total <- (totop/sqrt(abs(tobot)))
    return(total)
}
```

```
mevar2<-function(data) {
#y = data in matrix form, mu = grand mean
#s2 = error variance, s2a = eta variance
#mm = number of individuals, n = number of reps
#this functions calculates the score test with all parameters known
        size <- dim(data)
        mm <- size[1]
        n <- size[2]
        eta <- sumy <- sumy2 <- top <- numeric(mm)
        bot1 <- bot2 <- bot3 <- numeric(mm)
        s2 <- sum((data - apply(data, 1, mean))^2)/(mm * (n - 1))
        s2a <- sum((apply(data, 1, mean) - mean(data))^2)/(mm - 1) -
                s2/n
        mu <- mean(data)
        eta <- ((apply(data, 1, sum) - n * mu) * s2a)/(s2 + n * s2a)
        sumy <- apply((data - mu - eta), 1, sum)
        sumy2 <- apply((data - mu - eta)^2, 1, sum)
#top (numerator) of score test
        top <- eta * (1 - n + sumy2/s2) * (s2 + n * s2a) - (eta *
                s2 + 2 * s2a * sumy)
#bottom (denominator) of score test
        bot1 <- (2 * eta^2 * sumy2 * (s2 + n * s2a)^2)/s2
        bot2 <- (2 * eta^2 * s2 + 2 * s2a * sumy2) * (s2 + n * s2a)
        bot3 <-  - (eta * s2 + 2 * s2a * sumy)^2 + 6 * n * s2 * s2a^2
        totop <- sum(top)
        tobot <- sum(bot1 + bot2 + bot3)
        return(totop/sqrt(abs(tobot)))  }
```

```
mevar2a<-function(data)
{
#y = data in matrix form, mu = grand mean
#s2 = error variance, s2a = eta variance
#mm = number of individuals, n = number of reps
#this functions calculates  score test with 4x4
#information matrix.
        size <- dim(data)
        mm <- size[1]
        n <- size[2]
        Info <- array(dim = c(4, 4))
        s2 <- sum((data - apply(data, 1, mean))^2)/(mm * (n - 1))
        s2a <- sum((apply(data, 1, mean) - mean(data))^2)/(mm - 1) -
                s2/n
        mu <- mean(data)
        eta <- ((apply(data, 1, sum) - n * mu) * s2a)/(s2 + n * s2a)
        sumy <- apply((data - mu - eta), 1, sum)
        sumy2 <- apply((data - mu - eta)^2, 1, sum)
        A <- s2 + n * s2a
#top (numerator) of score test
        top <- eta * (1 - n + sumy2/s2) * A - (eta * s2 + 2 * s2a *
                sumy)
#bottom (denominator) of score test
        bot1 <- (2 * eta^2 * sumy2 * A^2)/s2
        bot2 <- (2 * eta^2 * s2 + 2 * s2a * sumy2) * A
        bot3 <-  - (eta * s2 + 2 * s2a * sumy)^2 + 6 * n * s2 * s2a^2
        I22 <- ((1 - n)/(2 * s2^2) + sumy2/s2^3) * A^2 - 1/2
```

```
        I12 <- (A^2 * eta * sumy2)/s2^2 + eta * A - (s2 * eta + 2 *
                sumy * s2a)
        I33 <- (eta^2 * A^2)/s2a^3 - n^2/2
        I13 <- 2 * sumy * A - n * (eta * s2 + 2 * s2a * sumy)
        I23 <- ( - mm * n)/2
        Info[1, 1] <- sum(bot1 + bot2 + bot3)
        Info[2, 2] <- sum(I22)
        Info[1, 2] <- Info[2, 1] <- sum(I12)
        Info[3, 3] <- sum(I33)
        Info[1, 3] <- Info[3, 1] <- sum(I13)
        Info[2, 3] <- Info[3, 2] <- I23
        Info[4, 4] <- (n * mm * A^2)/s2
        Info[3, 4] <- Info[4, 3] <- 0
        Info[2, 4] <- Info[4, 2] <- (A^2 * sum(sumy))/(s2^2)
        Info[1, 4] <- Info[4, 1] <- (2 * A^2 * sum(eta * sumy))/s2 -
                2 * A * n * mm * s2a
        tobot <- det(Info[2:4, 2:4])
        C <- det(Info)
        total <- sum(top) * sqrt(abs(tobot/C))  #distrib. N(0,1)
        return(total)
}
```

A.2 Approximate likelihood functions

Second order Laplace approximation to the true log likelihood using the Salamander data.

```
function(alpha,sgma)
{
        on.exit(if(((started <- exists(".newmin", frame = 1)) && (v <
                .newmin)) || !started) {
                assign(".newmin", v, frame = 1)
                cat(v, ":", format(sigma), "\n")
        }
        )
        sigma <- c(alpha,sgma)
        sigma[5] <- sigma[5]^2
        sigma[6] <- sigma[6]^2
        m <- length(levels(Sal$Female))
        r <- length(levels(Sal$Male))
        if(!exists("Est", frame = 1))
                Est <- rep(0, m + r)
        tmp <- nlminb(start = Est, objective = Logitmax, gradient
                 = GLogitmax, hessian = T, sigma = sigma)
        assign("Est", tmp$parameters, frame = 1)
        v <- tmp$objective + log(det(tmp$hessian))/2 + (m * log(
                sigma[5]))/2 + (r * log(sigma[6]))/2
        v
}
```

Fourth order Laplace approximation to the true log likelihood using the Salamander data.

```
function(alpha, sgmas)
{
        on.exit(if(((started <- exists(".newmin", frame = 1)) && (v <
                .newmin)) || !started) {
                assign(".newmin", v, frame = 1)
                cat(v, ":", format(sigma), "\n")
        }
        )       #on.exit(cat("Likelihood =", v, "\n"))
        sigma <- c(alpha, sgmas)
        sigma[5] <- sigma[5]^2
        sigma[6] <- sigma[6]^2
        m <- length(levels(Sal$Female))
        r <- length(levels(Sal$Male))
        n <- m + r
        if(!exists("Est", frame = 1))
                Est <- rep(0, n)
        tmp <- nlminb(start = Est, objective = Logitmax, gradient
                 = GLogitmax, hessian = T, sigma = sigma)
        assign("Est", tmp$parameters, frame = 1)
        a <- Est[1:20]
        b <- Est[21:40]
        eta <- as.vector(Sal$X %*% sigma[1:4] + a[Sal$Female] + b[
                Sal$Male])
        K <- exp(eta)/(1 + exp(eta))
        l1111 <-  - K + 7 * K^2 - 12 * K^3 + 6 * K^4
```

```
    # Now calculate the approximation to the log-likelihood
    i <- cbind(as.numeric(Sal$Female), as.numeric(Sal$Male))
    L1111 <- array(0, c(m, r))
    L1111[i] <- l1111
    Dinv <- ginv(tmp$hessian)
    tot1 <- diag(Dinv)^2 * c(tapply(l1111, Sal$Female, sum),
            tapply(l1111, Sal$Male, sum))
    Dm <- rep(diag(Dinv)[1:m], r)
    Dr <- rep(diag(Dinv)[(m + 1):n], rep(m, r))
    D3 <- as.vector(Dinv[1:m, (m + 1):n])
    D2 <- 4 * Dm * D3 + 4 * Dr * D3 + 2 * (Dm * Dr + 2 * D3^2)
    tot2 <- sum(tot1) + sum(as.vector(L1111) * D2)
    v <- tmp$objective + log(det(tmp$hessian))/2 + (m * log(
            sigma[5]))/2 + (r * log(sigma[6]))/2 - tot2/8
    v
}
```

```
Logitmax<-function(Est, sigma = ST) {
#  on.exit(cat("Logitmax =", v, "\n"))
        a <- Est[1:20]
        b <- Est[21:40]
        eta <- Sal$X %*% sigma[1:4] + a[Sal$Female] + b[Sal$Male]
        s2a <- sigma[5]
        s2b <- sigma[6]
        v <- sum(a^2)/(2 * s2a) + sum(b^2)/(2 * s2b) - sum(Sal$Y *
                eta - log(1 + exp(eta)))
        v  }

GLogitmax<-function(Est, sigma = ST) {
        a <- Est[1:20]
        b <- Est[21:40]
        eta <- as.vector(Sal$X %*% sigma[1:4] + a[Sal$Female] + b[
                Sal$Male])
        s2a <- sigma[5]
        s2b <- sigma[6]
        p <- 1 - 1/(1 + exp(eta))
        H <- crossprod(p * (1 - p) * Sal$Xab, Sal$Xab)
        diag(H) <- diag(H) + rep(1/sigma[5:6], c(length(a), length(
                b)))
        list(gradient = c(a/s2a, b/s2b) - crossprod(Sal$Xab, Sal$Y -
                p), hessian = H[!lower.tri(H)])  }
```

Bibliography

Abrams, K. & Sanso, B. (1998). Approximate Bayesian inference for random effects meta-analysis. *Statistics in Medicine* **17**, 201–218.

Aitkin, M. (1999). A general maximum likelihood analysis of variance components in generalized linear models. *Biometrics* **55**, 117–128.

Bacchetti, P. & Moss, A. R. (1989). Incubation period of AIDS in San Francisco. *Nature* **338**, 251–253.

Barndorff-Nielson, O. E. & Cox, D. R. (1989). *Asymptotic techniques for use in statistics.* Chapman and Hall.

Boscardin, W. J., Taylor, J. M. G., & Law, N. (1998). Longitudinal models for AIDS marker data. *Statistical Methods in Medical Research* **7**, 13–27.

Breslow, N. E. & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* **88**, 9–25.

Breslow, N. E. & Lin, X. (1995). Bias correction in generalised linear mixed models with a single component of dispersion. *Biometrika* **82**, 81–91.

Brookmeyer, R. & Gail, M. (1988). A method for obtaining short-term projections and lower bounds on the size of the AIDS epidemic. *Journal of the American Statistical Association* **82**, 301–308.

Commenges, D. & Jacqmin-Gadda, H. (1997). Generalized score test of homogeneity based on correlated random effects models. *Journal of the Royal Statistical Society* **B 59**, 157–171.

Cook, R. J. & Ng, E. T. M. (1999). Adjusted score tests of homogeneity for Poisson processes. *Journal of the American Statistical Association* **94**, 308–319.

Cox, D. R. (1998). Components of variance: a miscellany. *Statistical Methods in Medical Research* **7**, 3–12.

Cox, D. R. & Hinkley, D. V. (1974). *Theoretical statistics.* Chapman and Hall.

DeGruttola, V., Lange, N., & Dafni, U. (1991). Modeling the progression of HIV infection. *Journal of the American Statistical Association* **86**, 569–577.

Fraser, D. A. S., Monette, G., Ng, K. W., & Wong, A. (1994). Higher order approximations with generalised linear models. *Multivariate analysis and its applications - IMS lecture notes* **24**, 253–262.

Gilmour, A. R., Anderson, R. D., & Rae, A. L. (1985). The analysis of binomial data by a generalised linear mixed model. *Biometrika* **72**, 593–599.

Hodges, J. S. (1998). Some algebra and geometry for hierarchical models, applied to diagnostics (with discussion). *Journal of the Royal Statistical Society* **B 60**, 497–536.

Hunt, C. H. & Solomon, P. J. (1999). On investigating mean-variance relationships in general component of variance models. *Australian and New Zealand Journal of Statistics. Under revision* .

Jacqmin-Gadda, H. & Commenges, D. (1995). Tests of homogeneity for generalized linear models. *Journal of the American Statistical Association* **90**, 1237–46.

Lange, N., Carlin, B. P., & Gelfand, A. E. (1992). Hierarchical Bayes models for the progression of HIV infection using longitudinal CD4 T-cell numbers. *Journal of the American Statistical Association* **87**, 615- 626.

Lee, Y. & Nelder, J. A. (1996). Hierarchical generalized linear models. *Journal of the Royal Statistical Society* **B 58**, 619–678.

Lin, X. (1997). Variance component testing in generalised linear models with random effects. *Biometrika* **84**, 309–326.

LIN, X. & BRESLOW, N. E. (1996). Bias correction in generalized linear mixed models with multiple components of dispersion. *Journal of the American Statistical Association* **91**, 1007–16.

LINDSTROM, M. J. & BATES, D. M. (1988). Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. *Journal of the American Statistical Association* **83**, 1014–22.

LIU, Q. & PIERCE, D. A. (1993). Heterogeneity in Mantel-Haenszel-type models. *Biometrika* **80**, 543–556.

MCCULLAGH, P. & NELDER, J. A. (1989). *Generalized linear models.* Chapman and Hall, 2nd edition.

PAWITAN, Y. & SELF, S. (1993). Modelling disease marker processesion AIDS. *Journal of the American Statistical Association* **88**, 719–726.

ROBINSON, G. K. (1991). That BLUP is a good thing: The estimation of random effects. *Statistical Science* **6**, 15–51.

SATTEN, G. A. & LONGINI, I. M. (1994). Estimation of incidence of HIV infection using cross-sectional marker surveys. *Biometrics* **50**, 675–688.

SCHALL, R. (1991). Estimation in generalised linear models with random effects. *Biometrika* **78**, 719–27.

SCHEFFE, H. (1959). *The analysis of variance.* Wiley.

SEARLE, S. R. (1971). *Linear models.* Wiley.

SEARLE, S. R., CASELLA, G., & MCCULLOCH, C. E. (1992). *Variance Components.* Wiley.

SHUN, Z. (1997). Another look at the salamander mating data: A modified Laplace approximation approach. *Journal of the American Statistical Association* **92**, 341–349.

SHUN, Z. & MCCULLAGH, P. (1995). Laplace approximation of high dimensional integrals. *Journal of the Royal Statistical Society* **B 57**, 749–760.

Solomon, P. J. (1985). Transformations for components of variance and covariance. *Biometrika* **72**, 233–9.

Solomon, P. J. & Cox, D. R. (1992). Nonlinear component of variance models. *Biometrika* **79**, 1–11.

Steele, B. M. (1996). A modified EM algorithm for estimation in generalized mixed models. *Biometrics* **52**, 1295–1310.

Vonesh, E. F. (1996). A note on the use of Laplace's approximation for nonlinear mixed-effects models. *Biometrika* **83**, 447–52.

Weisburg, S. (1985). *Applied Linear Regression.* Wiley, 2nd edition.

Winkelstein, W., Lyman, D. M., Padian, N., Grant, R., Samuel, M., Wiley, J. A., Anderson, R. E., Lang, W., Riggs, J., & Levy, J. A. (1987). Sexual practices and risk of infection by the human immunodeficiency virus. *Journal of the American Medical Association* **257**, 321–25.

Wolfinger, R. (1993). Laplace's approximation for nonlinear mixed models. *Biometrika* **80**, 791–5.