IMPLEMENTATION OF THE BAYESIAN PARADIGM FOR HIGHLY PARAMETERISED LINEAR MODELS

T.D. Lee

Thesis submitted to the University of Nottingham
for the degree of Doctor of Philosophy

September 1986

PAGE NUMBERING AS ORIGINAL

Implementation of the Bayesian paradigm for highly parameterised linear models

Chapter 1) Introduction and review of Bayes Linear Models

Chapter 2) Multivariate Normal Results and generalisations to other distributions

Chapter 3) Characterisation of the Likelihood Evaluation Problem

Chapter 4) Reference Prior Distributions

Chapter 5) Bayesian Analysis of a multi-stratum experimental design

Chapter 6) A Bayesian analysis of inter and intra-laboratory variation

Chapter 7 Classical Variance Component Analysis vs Bayesian Procedures

Chapter 8 Conclusion

Papers:

The following papers contain material presented in Chapters 5 and 6 respectively.

A.M. Skene, J.E.H. Shaw and T.D. Lee, "Bayesian modelling and sensitivity analysis" *The Statistician* 35, pp 281-288 (1986).

T.D. Lee, "Assessment of intra and inter laboratory variances - A Bayesian alternative to BS5497" Presented to the Institute of Statisticians conference *Applied Bayesian Statistics* Cambridge July 1986. To appear in *The Statistician*.

Table of Figures

List of Tables

Implementation of the Bayesian paradigm

for highly parameterised linear models

Abstract

This thesis re-examines the Bayes hierarchical linear model and the associated issue of variance component estimation in the light of new numerical procedures, and demonstrates that the Bayes linear model is indeed a practical proposition. Technical issues considered include the development of analytical procedures essential for efficient evaluation of the likelihood function, and a partial characterisation of the difficulty of likelihood evaluation. A general non-informative prior distribution for the hierarchical linear model is developed. Extensions to spherically symmetric error distributions are shown to be practicable and useful. The numerical technique enables the sensitivity of the results to the prior structure, error structure and model structure to be investigated. An extended example is considered which illustrates these analytical and numerical techniques in a 15 dimensional problem. A second example provides a critical examination of a British Standards Institute paper, and develops further techniques for handling alternative spherically symmetric error distributions. Recent work on variance component estimation is viewed from the Bayesian perspective, and areas for further work are identified.

Acknowledgements

My thanks are due to Dr. A.M. Skene for his invaluable advice, support and encouragement during the time of this thesis, and to Joy and Pamela for putting up with us. I am also grateful for the inspiration provided by Prof. A.F.M. Smith.

Thanks are also due to Dr. J.C. Naylor and J.E.H. Shaw, whose suggestions have been highly valued, and whose work in the development of the BAYES4 and GR packages has made this research possible.

Finally, I would like to thank the SERC for providing financial support for the duration of my research.

Chapter 1 Introduction and review of Bayes Linear Models

1.1 Introduction

Over the last 20 years the accumulated literature on alternatives to the classical linear model has grown steadily. Important developments in the literature over this period have included the introduction of Bayesian ideas, particularly the hierarchical model which has, loosely speaking, similar properties to the classical random effects model. However, issues of tractability have tended to dominate and there has been little demonstration that the hierarchical linear model is of practical use. Associated with this, there has been little movement from the standard assumptions (such as normality at all stages in the model). This may be because the relaxation of these assumptions introduces even more numerical difficulties. Recently, considerable progress has been made with numerical integration in the Bayesian context to the extent that multi-parameter models can be handled effectively within the Bayesian framework. These numerical techniques enable the integration of a posterior distribution with upwards of 15 to 20 dimensions allowing the production of marginal posterior densities on individual parameters or pairs of parameters. It is argued that a lot more information can be obtained from these posterior densities than from a few point estimates. The posterior densities are also good starting points for the calculation of predictive densities. This thesis re-examines the Bayes hierarchical linear model and the associated issue of variance component estimation in the light of these new numerical procedures, and demonstrates that the Bayes linear model is indeed a practical proposition.

Calculating a marginal density via numerical integration requires repeated evaluation of the likelihood and prior densities at points in some sense covering the parameter space, or at least covering that subset of parameter space which contains "almost all" of the non-zero probability. There are two important issues here. First is the choice of points in parameter space at which to evaluate the likelihood and prior, and second is the ease with which the likelihood and prior can be evaluated at any specified point. The progress with numerical integration has provided techniques for choosing the evaluation points. For the hierarchical linear model, the evaluations themselves are essentially the evaluations of a quadratic form involving the inverse of a dispersion matrix. In numerical terms, evaluating the inverse dominates, and thus different classes of hierarchical linear model are more or less easy to handle numerically. This thesis shows that completely balanced factorial models require minimum numerical effort, followed by unbalanced nested models

and finally other unbalanced models, and introduces a new class of slightly unbalanced models which can be handled efficiently. Thus, substantially freed from the previous restrictions imposed by tractability, numerical integration allows much more freedom in the choice of both error distribution and prior. Given the ease with which these perturbations can be made to the original model, it is argued that a sensitivity analysis should be a routine feature of Bayesian linear model analysis, so that the robustness or otherwise of the results can be reported.

1.2 Structure of the thesis

Chapter 2 reviews the standard prior to posterior analyses adopted for the hierarchical linear model. It happens that for the hierarchical linear model certain parameters can be integrated out analytically. The notion of selective margins is introduced, whereby the dimensionality of the numerical integration can be reduced by first carrying out some form of analytical integration. It is also shown that certain margins can be viewed as weighted mixtures, where the mixing distribution is the posterior density arising from a simpler model. Such mixture densities can be handled elegantly by the BAYES4 numerical integration package using "Special Function Analysis" (see Section 1.4). Attention is then focused on alternative error distributions, and it is shown that the use of alternative spherically symmetric distributions has almost no impact on the numerical complexity. It is further observed that only a slight relaxation of the usual assumptions associated with the error term in the linear model allows the substitution of scale mixtures of multivariate normals for the joint error term. This prompts the advocacy of more general error distributions for Bayes linear models especially as it is shown that such a change has little effect on the analytic results or on the numerical complexity.

Chapter 3 characterises the problem of evaluating the likelihood. Completely balanced factorial models are shown to have dispersion matrices that are easy to invert and hence such models are algebraically and numerically easy to handle, as naive expressions can be simplified prior to coding. Unbalanced factorial models pose more difficulties. A new class of "slightly unbalanced factorial models" is introduced for which the likelihood can still be evaluated without requiring numerical inversion of a large dispersion matrix. Grossly unbalanced nested factorial models are shown to be tractable, but grossly unbalanced crossed designs without interaction terms are shown to be difficult. Algebraic results that simplify

the inversion of the dispersion matrix are produced for these cases.

Chapter 4 discusses the problem of deriving a reference prior distribution for the general hierarchical linear model. The Jeffreys' prior is derived under the assumption of normality, and is shown to be unchanged when the error distribution is changed to a general multivariate t distribution. Finally Bernardo reference priors are considered and are shown to be the same as the Jeffreys' priors.

Chapter 5 demonstrates the feasibility of the numerical approach by providing a full worked analysis of a multi-stratum experimental design. The maximal model is a 15 parameter model comprising 3 variance components and 12 fixed effects. The marginal posterior density for the eigenvalues of the dispersion matrix is produced analytically under the assumption of normality, together with a series of distributions conditional on the eigenvalues. A sub-model with only 6 fixed effects is also considered. Two methods of integration are used, one based on the 3 dimensional likelihood for the variance components, and the other based on the full 15 dimensional likelihood. The analysis is repeated using multivariate t errors, and both a non-informative prior and an informative prior are considered. The results of these analyses are presented, and a comparison is made with the results of the original analysis. Finally the performance of the integration routines is discussed. The full 15 dimensional integral is used to show the feasibility of Bayesian analysis of high dimensional linear models, even with t distribution errors. The development of this example took place at a time when the possibilities of using BAYES4 beyond 10 dimensions were just being realised. The example demonstrated that numerical techniques were applicable to higher dimensional problems than had previously been considered. It is believed that this example is probably the first Bayesian analysis to involve numerical integration over as many as 15 dimensions.

The one way analysis of variance model has already received considerable attention from the Bayesian perspective by Box and Tiao (1973) and others. However, one routine use of the one way random effects model is in the estimation of inter and intra laboratory variation which is the subject of a British Standards Institute paper number 5497. Chapter 6 provides a critical analysis of the Standard, in the spirit of Box and Tiao, but incorporating the ideas of this thesis with respect to error distributions. The analysis generalises the model of the Standard by allowing t distributions instead of normal distributions for either the error distribution or the distribution of the laboratory means. The numerical techniques used in this chapter contrast with the brute force numerical integration of Chapter 5. Since

the t distributions can be considered as scale mixtures of normal distributions, the mixing parameter can enter the likelihood as a nuisance parameter. This parameterisation enables an algebraic simplification of the likelihood yielding a low dimensional integral. It is concluded that a Bayesian analysis is efficient and informative in this context, and that the procedures advocated by the BSI analysis are suspect unless it is known for certain that both distributions are normal.

Chapter 7 presents a survey of classical variance component analysis. The traditional ANOVA estimates are considered together with the more recent Minimum Variance Unbiased Estimators, Minimum Norm Quadratic Estimation (MiNQE) and Maximum Likelihood Estimates (MLE). The inter-relationships between the estimators are shown, and comments on the estimators are made from a Bayesian perspective.

Chapter 8 contains a concluding discussion and identifies directions for further work.

1.3 Bayesian theory for linear models

The book by Box and Tiao provides much of the theory for Bayesian linear model analysis and deals specifically with a variety of random effects and mixed models. Zellner (1971) provides an introduction to linear model theory, and gives a Bayesian analysis of time series and econometric models, including simultaneous equation models. More recently a good survey of the work is supplied by Broemeling (1985), and an efficient statement of the algebra is given by Berger (1985). The hierarchical model is described by Lindley and Smith (1972) who advocated the use of a multistage structure to describe the relationships between the model parameters. The authors indicated that exchangeable prior distributions for (subsets of) the location parameters may sometimes be appropriate. Smith (1973) examines the Bayesian model in more detail and lists some general properties of the resulting Bayes estimators.

Throughout the Seventies, a number of authors have suggested applications of this model to areas including growth curve models, non-linear regression, time series and econometric models. Fearn (1975) examines the generalised growth models of Potthoff and Roy (1964) which have also been studied from a Bayesian

viewpoint by Geisser (1970). Such models are shown to be within the scope of the general Bayesian linear models of Lindley and Smith. Fearn generalises the results and produces approximations for the posterior distribution of the means. Sweeting (1982) also extends the results of Smith.

The book by Zellner appears to be the only one devoted to Bayesian analysis of parametric time series models. The mid-Seventies saw a lot of interest in econometric data. Bayesian studies of time series models have been carried out by amongst others Newbold (1973), and Harrison and Stevens (1976). An important reference on structural change in linear models is Poirier (1976) who reviews statistical and econometric literature and presents new ideas on modelling structural change using spline functions. Other recent structural change papers include Smith (1975), Holbert and Broemeling (1977), Abraham and Wei (1979), Chin Chay and Broemeling (1980), Tsurami (1980) and Salazar, Broemeling and Chi (1981). Bayesian methods for adaptive fitting for the Kalman filter are given by Hawkes (1973) and Alspach (1974).

However, throughout this period, the tractability of the posterior density was always a problem, and many early papers advocate the use of modes as approximations to the posterior means. This leads to a discussion of the relative merits of joint modes and marginal modes (see for example O'Hagan (1976)).

Later in the decade, a lot of interest focused on robustness. Papers by Box (1980), Chen and Box (1979), Bailey and Box (1980), and an early paper by Rubin (1977) consider the robustness of results to outliers from the Bayesian perspective.

Throughout this period there were still few examples of practical data analysis.

In the years between 1979 and the start of this thesis in 1983, this work essentially languished. Over this period however, analytical approximations and numerical procedures were developed as authors have tackled the problem of obtaining margins rather than modes (see Skene (1983), Smith et al. (1985)). Naylor and Smith (1982) identified two aspects to numerical Bayesian analysis. First is the problem of the numerical integration, and second is the difficulty of reconstructing a marginal posterior density from a sparse set of evaluation points. These problems are now tackled by two computer packages BAYES4 and GR which have been developed under a SERC research project at the University of Nottingham during

1983-1986. These developments allow useful summaries to be produced for a much broader class of models than previously possible.

At the commencement of work on this thesis, the full potential of multidimensional integration using, for example, BAYES4 was only just being appreciated. It was anticipated that numerical procedures of this general type would become available for routine Bayesian analysis involving models with several parameters, but the power of such techniques for highly parameterised linear models had not been investigated. Similarly, GR was of unknown utility. The Bayesian statistics community had had little opportunity to express the results of an analysis via a large number of interesting marginal densities and the most useful form of such plots was still an open question.

1.4 The BAYES4 numerical integration package

The BAYES4 package enables the production of summaries from high dimensional posterior densities using numerical integration. BAYES4 is built upon four important concepts, i) transformations of the parameter space to make the integral "easier", ii) initial estimates of the first and second moments of the transformed parameters, iii) calculation of many integrals in parallel, and iv) iterative re-estimation of the first and second moments of the parameters. Naylor and Smith (1982) describe the basic iterative philosophy used by BAYES4, and further details and illustrations of its use can be found in Shaw (1986).

BAYES4 works by evaluating the likelihood and prior distributions at selected points in parameter space. If the moments of the posterior distribution were known in advance, then the points at which the likelihood was to be evaluated could be chosen in some optimal way so as to cover the densest part of the multi-dimensional posterior distribution, and thus achieve maximum accuracy for any statistics computed from the set of points chosen. Having chosen a set of points using initial estimates of the moments, it is then possible to estimate the moments of the posterior distribution by evaluating the appropriate integrals numerically. These moments can be used to choose a *better set of evaluation points*, which in turn lead to more reliable estimates for the moments. Thus by iteratively upgrading the estimates for the moments, it is possible to home in on the region of high posterior probability in parameter space, and thus efficiently compute every other integral of

interest.

BAYES4 uses three distinct numerical integration techniques to tackle problems with different numbers of dimensions. The first of these is repeated Gauss-Hermite integration over a cartesian product grid. This works by approximating the posterior density by a low order polynomial times a multivariate normal distribution. With careful choice of the evaluation points and the weights attached to them, functions of this form can be integrated exactly using only a sparse grid of points provided that the first two moments of the normal kernel are known. For example a n^k grid will yield exact results for the product of an order $2n-1$ polynomial and an k dimensional normal. If first and second moments are evaluated, then these are calculated exactly for the functions that are the product of a k dimensional normal and an order $2n-3$ polynomial. Thus even a 4^k grid is very powerful.

It is thus highly advantageous to use transformations of parameter space that yield posterior surface of the form of a normal multiplied by a low order polynomial. Considerable thought should be given to a suitable (if arbitrary) transformation eg. log, logit, probit. This theme is further discussed in Chapters 5, 6 and 8. Even with integration rules of this type, the number of evaluations becomes prohibitive as the number of dimensions increases. On the current generation of computer hardware, integration with up to 6 dimensions can be undertaken using Gauss-Hermite integration.

For higher dimensions a class of spherical integration rules are available, and these are highly recommended for regular problems with 4 to 8 parameters (see Stroud (1971)). For problems with even more dimensions, Monte-Carlo integration techniques are available which have been demonstrated in as many as 23 dimensions (see Skene and Wakefield (1986)). These techniques are discussed in Shaw (1985a,1985b) and use a form of importance sampling (see Hammersley and Handscomb (1964)) based on transforming a configuration of points in the unit k-dimensional hypercube to a corresponding configuration of points in $\mathbb{R}^k$. In any problem, the integration techniques may be mixed with (say) some Gauss-Hermite dimensions and some Monte-Carlo dimensions. Variants of Monte-Carlo integration again prove to be very accurate and efficient when the joint posterior density is close to a multivariate normal in form.

For each of the integration techniques, BAYES4 iterates to yield a stable set of first and second moments from a user supplied set. It also enables the calculation of

univariate or bivariate marginal distributions. Finally BAYES4 has a feature called *Special Function Analysis* that enables the evaluation of functions of the form $\int f(\theta)p(\theta|y)d\theta$, where $f(\theta)$ is an arbitrary function of the parameters, and $p(\theta|y)$ is the posterior density for the parameters.

To use BAYES4, the user is required to provide code for evaluating the likelihood and prior, and an initial estimate of the parameters. The likelihood should be coded for maximum efficiency. Data translation can often help here, as this often yields simpler expressions for the likelihood. Similarly great care must be taken with linear models to sort out the inverse of the dispersion matrix. In ideal cases the quadratic form can be reduced to a simple expression involving a few sums of squares. Experience shows that the choice of initial estimates for the parameters in linear model problems seldom poses difficulties. Any estimates that are even vaguely sensible will usually enable rapid convergence to the correct values. A final point is that the code for the likelihood can in fact be code for several alternative likelihoods controlled by a selection mechanism at run time. An appendix to this thesis contains the code which was used for some of the analyses performed using BAYES4. This illustrates the work which must be done to use BAYES4 and gives some practical tips.

1.5 GR - a graphical presentation and manipulation package

In addition to the first and second moments, and possibly special function analysis, BAYES4 is usually used to produce a small number of spot heights from one or more univariate or bivariate posterior densities. It is required to reconstruct the univariate or bivariate posterior densities from these spot heights. Typically these spot heights will be unevenly spaced (as the roots of a Gauss-Hermite polynomial) or in the bivariate cases as a sheared grid of unevenly spaced points. Often the points will not even be on the required scale as a transformation of parameter space was made to enable BAYES4 to work efficiently. Thus the reconstruction of the posterior density on the natural scale is difficult.

This task is achieved by GR which will take a sparse set of points and reconstruct the posterior density by interpolation, with extrapolation if required, using cubic splines applied to the log ordinate. GR includes a large range of univariate and bivariate transformations to enable the posterior density to be

produced on the natural scale or any other scale. GR also enables univariate densities to be calculated from bivariates by marginalisation or conditioning. As output GR will produce bivariate contour plots and/or univariate probability densities. For further details see the GR User Guide, Shaw (1986). All the marginal densities and contour plots in this thesis were produced using GR.

Chapter 2
Multivariate Normal results and generalisations to other distributions

2.1 Introduction

Any Bayesian analysis involving numerical integration requires repeated evaluations of the posterior density. Linear models may have many parameters (counting both fixed effects and variance components) giving rise to a posterior density with many dimensions. Numerical evaluation of posterior distributions from hierarchical linear models must be carried out efficiently to avoid a herculean amount of work.

There are two different aspects of efficient evaluation. First it may be possible to invert the dispersion matrix analytically, rather than tackle it numerically, or failing that, it is usually possible to obtain the inverse of the dispersion matrix from the inverse of a matrix of smaller size, which has to be inverted numerically. Secondly it may be possible to take the high dimensional posterior distribution and analytically integrate out parameters that are not of direct interest to get a marginal posterior density with fewer dimensions.

With linear models it is common to propose some linear structure to describe the fixed effects, and to assume the errors between the data and the model come from a multivariate normal distribution. Symbolically, the data y has expectation $X\alpha$ where α is a vector of fixed effects and X is a design matrix. Then writing V for the dispersion matrix gives:

$$y \sim N(X\alpha, V)$$

For most of the models considered in this thesis, V is of the form $V(\sigma^2)$ where σ^2 is a vector of dispersion parameters commonly referred to as variance components. This model gives rise to a likelihood:

$$l(y|\alpha,\sigma^2) = (2\pi)^{-\frac{1}{2}n}|V|^{-\frac{1}{2}}\exp-\tfrac{1}{2}(y-X\alpha)'V^{-1}(y-X\alpha)$$

and hence from Bayes' theorem

$$p(\alpha,\sigma^2|y) \propto |V|^{-\frac{1}{2}}\exp-\tfrac{1}{2}(y-X\alpha)'V^{-1}(y-X\alpha)\, p(\alpha,\sigma^2) \qquad (2.1{:}1)$$

where $p(\alpha,\sigma^2)$ is the prior distribution on the fixed effects and variance components. The location parameters and scale parameters are frequently assumed to

be *a priori* independent, hence $p(\underline{a},\sigma^2) = p(\underline{a})p(\sigma^2)$. With the assumption of normality, the evaluation of the joint posterior density for $\underline{a}$ and σ^2 at any point is seen to be a matter of calculating $\underline{V}^{-1}$ and $|\underline{V}|$. However, the joint posterior for $\underline{a}$ and σ^2 may have many dimensions, and it may be difficult to produce margins numerically from the joint density. Fortunately, this is seldom necessary for two reasons. First the margins on the fixed effects are typically similar to each other subject to a difference in location (and possibly scale), so it is not necessary to consider them all at once. Second, the primary interest may be in σ^2, or in functions that can be derived from the marginal posterior density for σ^2.

Thus lower dimensional summaries of equation 2.1:1 are desirable. In general the distribution of a fixed effect conditional on the variance components is easy to produce. Also, analytic integration to remove the fixed effects is possible to yield a marginal posterior density for σ^2. Since these basic integrations can be done it leads to the notion of selective margins whereby the dimensionality of the posterior numerical integration can be reduced by first performing analytical integrations. These analytical integrations depend upon the prior specification, particularly the factorisation of the prior into a term for the fixed effects and a term for the variance components. In the analyses in this chapter, non-informative priors are used for parameters that are to be analytically integrated out of the posterior density.

The first part of this chapter (Section 2.2) lists a series of posterior densities from Normal models, after integrating out different parameters from $\underline{a}$ and σ^2. Specifically the following distributions are produced:

i) The marginal posterior density for the variance components σ^2.

ii) The joint posterior density for the variance components σ^2 and a single fixed effect (eg σ^2, a_1).

iii) The marginal posterior density for a single fixed effect (eg. a_1) or a pair of fixed effects (a_1, a_2).

It is shown that evaluation of these distributions requires $\underline{V}^{-1}$ and $|\underline{V}|$ as well as $(\underline{X}'\underline{V}^{-1}\underline{X})^{-1}$, $|\underline{X}'\underline{V}^{-1}\underline{X}|$. Efficient methods of evaluating these matrices and determinants are developed in Chapter 3.

The later part of this chapter extends the discussion to error distributions other than the normal distribution, as there may be little direct evidence for the assumption of normality, except by appealing to the Central Limit theorem. When considering alternatives to normality it is important to ensure that the distribution used retains the properties that enable analytic simplification of the posterior density, and also allow efficient numerical implementation of the densities produced. It is seen that the evaluation of spherical error distributions is dominated by the same matrices and determinants as for the normal case.

The discussion of alternative error distributions is preceded by a brief review of the properties of the multivariate normal distribution (Section 2.3) that enable analytic simplifications and numerical tractability. The first alternative error distribution considered (Section 2.4) is the multivariate *t* distribution (see Johnson and Kotz (1970)). This distribution has the same parameters and first and second moments as the multivariate Normal distribution, thus allowing direct comparison of models with different error structures. The multivariate *t* distribution is shown to keep all the properties of the multivariate normal that are useful for analytic and numerical progress except for one - namely the multivariate *t* distribution does not have independent errors. Thus there is no computational penalty associated with the use of the general multivariate *t* distribution rather than the multivariate Normal distribution.

A general discussion then follows in which some of the useful properties of the multivariate normal are taken as axioms, and the space of distributions obeying these axioms is explored. Specifically it is required that marginals and conditionals should have the same functional form as the original density. It is shown that the assumption of a spherical density, but without independence between the errors, is equivalent to using the set of scale mixtures of multivariate normal distributions.

2.2 Marginal Posterior Densities assuming a Normal Error Structure

2.2.1 The Marginal Posterior Density for the Variance Components

Sometimes the variance components themselves, or functions derived from them, are of primary interest. In other cases, a density is required that can be

formed as a mixture of distributions where the mixing distribution is the posterior density for the variance components. In these cases, the fixed effects can be integrated out from equation 2.1:1 yielding a marginal posterior density on the σ^2, which has considerably fewer dimensions than the joint posterior density for $\underset{\sim}{a}$ and σ^2.

Take the joint posterior density from equation 2.1:1, and assume a non-informative (uniform) prior for $\underset{\sim}{a}$. Integrate across the fixed effects to get a marginal posterior for the variance components:

$$p(\sigma^2 \mid y) \propto |V|^{-\frac{1}{2}} \int_{a} \exp -\tfrac{1}{2}(y - Xa)' V^{-1}(y - Xa)\, da\; p(\sigma^2)$$

$$\propto p(\sigma^2)\, |V|^{-\frac{1}{2}} \exp -\tfrac{1}{2}\, y'V^{-1}y \int_{a} \exp -\tfrac{1}{2}\{a'X'V^{-1}Xa - 2y'V^{-1}Xa\}\, da$$

The integration is an exercise in completing the square and yields

$$p(\sigma^2|y) \propto p(\sigma^2)\, |V|^{-\frac{1}{2}} |X'V^{-1}X|^{-\frac{1}{2}} \exp -\tfrac{1}{2} y'V^{-1}y \exp \tfrac{1}{2} y'V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}y \quad (2.2.1:1)$$

Efficient use of this requires $X'V^{-1}X$ to be invertible analytically, and $|X'V^{-1}X|$ to be calculable.

2.2.2 Joint Posterior Density for a single fixed effect and the Variance Components

Suppose that instead of wanting a marginal posterior density for the variance components σ^2, the joint posterior density for a fixed effect and the variance components is required. This posterior density may be of specific interest, or it may be used to determine the posterior correlation between one of the fixed effects and the variance components.

This posterior density can be produced simply as it factorises into the posterior density for the variance components, multiplied by the density of the fixed effect conditional on the variance components. The latter density is itself a univariate normal distribution. Thus, writing a_1 for the fixed effect:

$$p(a_1, \sigma^2|y) = p(a_1|\sigma^2, y)\, p(\sigma^2|y)$$

It is interesting to note that an alternative form for $p(a_1,\sigma^2|y)$ can be produced as follows: Partition X into X_1 and X_2, and a into a_1 and a_2 where $a_1 = (a_1)$ and a_2 is a vector of the other fixed effects. Then integrate out over the nuisance parameters a_2:

$$y \sim N(Xa,V) \quad \longrightarrow \quad y - X_1a_1 \sim N(X_2a_2,V)$$

hence integrating out a_2 yields:

$$p(a_1,\sigma^2|y) \propto |V|^{-\frac{1}{2}} \exp -\tfrac{1}{2}(y-X_1a_1)'V^{-1}(y-X_1a_1)$$
$$\times |X_2'V^{-1}X_2|^{-\frac{1}{2}} \exp \tfrac{1}{2}(y-X_1a_1)'X_2(X_2'V^{-1}X_2)^{-1}X_2'(y-X_1a_1) \qquad (2.2.2:1)$$

Equation 2.2.2:1 can be obtained by substituting $y-X_1a_1$ for y and X_2 for X in equation 2.2.1:1. Note that this formulation involves an unbalanced matrix $X_2'V^{-1}X_2$, which makes equation 2.2.2:1 difficult to evaluate (see section 3.2.3 in Chapter 3).

2.2.3 Marginal Posterior Densities for the fixed effects

Finally consider the marginal posterior distribution for the fixed effects. It is not usually possible to integrate out σ^2 analytically to leave a marginal posterior for $a|y$. Instead a marginal distribution for a single fixed effects a_1 or a pair of fixed effects a_1,a_2 can be obtained by numerically integrating σ^2 out of the joint density for the fixed effect(s) and the variance components (as produced in section 2.2.2).

$$p(a_1|y) = \int_{\sigma^2} p(a_1|\sigma^2,y)\,p(\sigma^2|y)\,d\sigma^2 \qquad (2.2.3:1)$$

$$\text{and } p(a_1,a_2|y) = \int_{\sigma^2} p(a_1,a_2|\sigma^2,y)\,p(\sigma^2|y)\,d\sigma^2 \qquad (2.2.3:2)$$

where $p(\sigma^2|y)$ is the posterior density for σ^2 as given by equation 2.2.1:1. Note that $p(\sigma^2|y)$ can be dependent on the choice of prior distribution for a,σ^2.

Conditional on $\sigma^2$and y, the fixed effects are normally distributed, and hence $p(a_1|y)$ and $p(a_1,a_2|y)$ from equations 2.2.3:1 and 2.2.3:2 are just mixtures of normal distributions with the mixing distribution being the marginal posterior density for the variance components.

In general, it is not possible to perform this integration analytically, but it is easy to do it numerically using Special Function Analysis in BAYES4. A special case

occurs for the single variance fixed effects model, where it is possible to produce $p(\boldsymbol{a}|\boldsymbol{y})$ analytically:

Special Case

Consider the linear model $\boldsymbol{y} \sim N(\boldsymbol{X}\boldsymbol{a}, \sigma^2\boldsymbol{I})$ where $\boldsymbol{X}$ is an $n \times k$ design matrix, $\boldsymbol{a}$ is an unknown vector of k fixed effects and the residual variance σ^2 is unknown. A particular case of this model is the single variance fixed effects model given by $y_{ij} = a_i + \epsilon_{ij}$, where $i = 1,2...p$, and $j = 1,2...q$. The ϵ_{ij} are independent with variance σ^2.

For this linear model it is possible to analytically integrate out σ^2 from the marginal for $\boldsymbol{a},\sigma^2$. The usual assumption of a non-informative prior for σ^2 leads to a marginal posterior for σ^2 of $\nu s^2 \chi_\nu^{-2}$ with $\nu = n-k$, and $s^2 = (\boldsymbol{y} - (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y})^2/\nu$. The distribution for $\boldsymbol{a}|\sigma^2$ is multivariate normal, leading to a heavy tailed marginal for $\boldsymbol{a}$. Let $\hat{\boldsymbol{a}}$ denote $(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$, then:

$$p(\boldsymbol{a}|\boldsymbol{y}) \propto \left[1 + \frac{(\boldsymbol{a}-\hat{\boldsymbol{a}})'\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{a}-\hat{\boldsymbol{a}})}{\nu s^2}\right]^{-\frac{1}{2}(\nu+k)} \qquad (2.2.3{:}3)$$

which is the multivariate t distribution discovered independently by Cornish (1954) and Dunnet & Sobel (1954). A derivation can be found in Box & Tiao (1973) section 2.7. Note that even if $\boldsymbol{X}'\boldsymbol{X}$ is a diagonal matrix, the margins for the fixed effects are not independent of each other as equation 2.2.3:3 does not factorise. Marginals from the multivariate t distribution are derived in section 2.4.1, thus allowing $p(a_i|\boldsymbol{y})$ to be evaluated.

2.3 Properties of the multivariate normal distribution.

A standard exposition of normal linear models considers a vector of data $\boldsymbol{y}$ with expectation $\boldsymbol{X}\boldsymbol{a}$ and errors that are independent identically distributed $\sim N(0,\sigma^2)$. The assumption of normality allows the vector of independent errors to be viewed as a vector from a multivariate normal distribution. Difficulties arise as soon as alternatives to normality are considered. For example, it may be considered appropriate to use a distribution with heavier tails than a normal distribution. However, if the errors are taken to be independent, then it is no longer possible to view the errors as being a vector from a heavy tailed spherical multivariate

distribution. Conversely, if a heavy tailed spherical multivariate distribution is assumed as an alternative to the multivariate normal distribution, then the independence property of the errors is lost.

These properties are discused in the following sections. First it is necessary to review those properties of the multivariate normal distribution which enable analytic progress to be made with the posterior density, and allow numerical tractability. This review permits the relative merits of other multivariate distributions to be assessed. Sections 2.4, 2.4.1, 2.4.2 and 2.4.3 derive the equivalent properties for the multivariate t distribution.

Consider $\boldsymbol{y}$, an n-vector from a multivariate Normal distribution with mean $\boldsymbol{X}\boldsymbol{a}$ and dispersion matrix $\boldsymbol{V}$, where $\boldsymbol{X}$ is a design matrix and $\boldsymbol{a}$ is a vector of fixed effects. Without loss of generality, it is assumed that $\boldsymbol{V} = \sigma^2\boldsymbol{I}_n$. This simplifies the resulting discussion, as $\boldsymbol{y}'\boldsymbol{V}^{-1}\boldsymbol{y}$ becomes $\boldsymbol{y}'\boldsymbol{y}/\sigma^2$, but the results hold for a general $\boldsymbol{V}$ matrix. Then the multivariate normal has the following useful properties:

i) It is spherically symmetric (or elliptically symmetric for general dispersion matrices).

ii) There is independence between $\{y_{m+1}\ldots y_n\}$ and $\{y_1, y_2 \ldots y_m\}$ conditional on the σ^2. Thus the probability density factorises and hence:

$$P(y_{m+1}\ldots y_n | y_1\ldots y_m, \sigma^2) = P(y_{m+1}\ldots y_n | \sigma^2)$$

iii) The conditional distribution for $y_{m+1}\ldots y_n | y_1\ldots y_m$ is multivariate normal with dispersion matrix $\boldsymbol{V} = \sigma^2\boldsymbol{I}_{n-m}$.

iv) The marginal distribution for $y_1\ldots y_m$ is multivariate normal with dispersion matrix $\boldsymbol{V} = \sigma^2\boldsymbol{I}_m$.

v) The distribution of the fixed effects $\boldsymbol{a}$ conditional on the data and the variances is multivariate Normal with variance $(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}$ and mean $(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y}$.

The objective is to generalise the class of possible error distributions, yet to retain as many of the properties i) - v) as are possible. This idea is certainly not new, eg. West (1984) considers extensions to simple normality by allowing scale mixtures of normal distributions in order to accommodate potential outliers without giving them too much influence on the final results. Dickey and Chen (1983) consider the

whole class of spherically symmetric distributions. Property i) spherical symmetry implies that $r^2 = \sum_{i=1}^{n} y_i^2$ is a minimal sufficient statistic for $\boldsymbol{y}$. Kelker (1970) examined the class of spherically symmetric distributions, and showed that the only spherically symmetric distribution that had the independence property was the multivariate normal. Properties i) and ii) thus imply a multivariate normal distribution. Conversely to extend the class of error distributions either property i) symmetry, or property ii) independence, must be dropped.

Generalised versions of properties iii) and iv) are that the marginals and conditionals have the same functional form as the joint distribution, and with the same variance structure. Similarly property v) requires that the fixed effects conditional on the data have the same functional form as the joint distribution and with mean $(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y}$ and dispersion matrix $(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}$.

2.4 Properties of the general multivariate *t* distribution

The multivariate *t* distribution is an obvious choice for an alternative error distribution to the multivariate normal distribution. It is now examined to discover how many of the properties i) - v) it possesses. Let $t_\nu(\boldsymbol{X}\boldsymbol{a}, \boldsymbol{V})$ denote a general multivariate *t* distribution with ν degrees of freedom, mean $\boldsymbol{X}\boldsymbol{a}$ and dispersion matrix $\boldsymbol{V}$. Note that there is not independence in the sense of property ii). The probability of a data vector $\boldsymbol{y} \sim t_\nu(\boldsymbol{X}\boldsymbol{a}, \boldsymbol{V})$ is usually written as:

$$p(\boldsymbol{y}|\boldsymbol{a},\boldsymbol{V}) = \frac{\Gamma(\frac{1}{2}(\nu+n))}{\Gamma(\frac{1}{2}\nu)} (\pi\nu)^{-\frac{1}{2}n} |\boldsymbol{V}|^{-\frac{1}{2}} \left[1 + \frac{(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{a})'\boldsymbol{V}^{-1}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{a})}{\nu}\right]^{-\frac{\nu+n}{2}} \qquad (2.4:1)$$

Unfortunately the second moments of this are $\frac{\nu}{\nu-2}\boldsymbol{V}$ rather than $\boldsymbol{V}$, and this is inconvenient when comparing models with different error distributions. A simple reparameterisation overcomes this difficulty, hence whenever the general multivariate *t* distribution is mentioned, the distribution given in equation 2.4:2 will be meant.

$$p(\boldsymbol{y}|\boldsymbol{a},\boldsymbol{V}) = \frac{\Gamma(\frac{1}{2}(\nu+n))}{\Gamma(\frac{1}{2}\nu)} \pi^{-\frac{1}{2}n} (\nu-2)^{-\frac{1}{2}n} |\boldsymbol{V}|^{-\frac{1}{2}} \left[1 + \frac{(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{a})'\boldsymbol{V}^{-1}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{a})}{\nu-2}\right]^{-\frac{\nu+n}{2}} \qquad (2.4:2)$$

This distribution has second moments of $\boldsymbol{V}$.

Now since this only depends on $\underset{\sim}{y}$ through the quadratic form $(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{a})'\underset{\sim}{V}^{-1}(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{a})$, it follows that this is an elliptically symmetric distribution, hence property i) from Section 2.3 holds. Also since it is a function of $(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{a})'\underset{\sim}{V}^{-1}(\underset{\sim}{y}-\underset{\sim}{X}\underset{\sim}{a})$, it is computationally no harder to evaluate than the multivariate Normal distribution.

2.4.1 Marginal Densities from a general multivariate t distribution

Consider the marginal distribution formed by integrating out y_n from the multivariate t distribution of $\underset{\sim}{y}$ ($\equiv y_n$ say). This yields the distribution for $\underset{\sim}{y}_{n-1}$. Without loss of generality, and to simplify the algebra assume that the mean of the t distribution is 0. As before, without loss of generality take $\underset{\sim}{V} = \sigma^2 I_n$.

$$p(y_n|\underset{\sim}{V}) = \frac{\Gamma(\frac{1}{2}(\nu+n))}{\Gamma(\frac{1}{2}\nu)}\pi^{-\frac{1}{2}n}(\nu-2)^{-\frac{1}{2}n}|\underset{\sim}{V}_n|^{-\frac{1}{2}}\left(1+\frac{\underset{\sim}{y}_n'\underset{\sim}{V}_n^{-1}\underset{\sim}{y}_n}{\nu-2}\right)^{-\frac{\nu+n}{2}}$$

Now integrate out y_n to produce a marginal distribution for $y_1,\ldots,y_{n-1}$

$$p(y_1\ldots y_{n-1}|\underset{\sim}{V}) = \int_{y_n} p(y_n|\underset{\sim}{V})\,dy_n$$

$$p(y_1\ldots y_{n-1}|\underset{\sim}{V}) = \int_{y_n}\frac{\Gamma(\frac{1}{2}(\nu+n))}{\Gamma(\frac{1}{2}\nu)}\pi^{-\frac{1}{2}n}(\nu-2)^{-\frac{1}{2}n}|\underset{\sim}{V}_n|^{-\frac{1}{2}} \times\left(1+\frac{\underset{\sim}{y}_{n-1}'\underset{\sim}{V}_{n-1}^{-1}\underset{\sim}{y}_{n-1}}{\nu-2}+\frac{1}{\sigma^2}\frac{y_n^2}{\nu-2}\right)^{-\frac{\nu+n}{2}}dy_n$$

Now let $Q = 1+\frac{\underset{\sim}{y}_{n-1}'\underset{\sim}{V}_{n-1}^{-1}\underset{\sim}{y}_{n-1}}{\nu-2}$, and substitute $\sigma^2(\nu-2)Q\tan^2\theta$ for y_n^2.

Thus $dy_n = (\sigma^2)^{\frac{1}{2}}Q^{\frac{1}{2}}(\nu-2)^{\frac{1}{2}}\sec^2\theta$

$$p(y_1\ldots y_{n-1}|\underset{\sim}{V}) = \frac{\Gamma(\frac{1}{2}(\nu+n))}{\Gamma(\frac{1}{2}\nu)}\pi^{-\frac{1}{2}n}(\nu-2)^{-\frac{1}{2}n}|\underset{\sim}{V}_{n-1}|^{-\frac{1}{2}}(\sigma^2)^{-\frac{1}{2}}Q^{-\frac{1}{2}(\nu+n)} \times\int_{\theta=0}^{\frac{1}{2}\pi}\cos^{\nu+n-2}\theta\,(\sigma^2)^{\frac{1}{2}}Q^{\frac{1}{2}}(\nu-2)^{\frac{1}{2}}\,d\theta$$

$$= \frac{\Gamma(\frac{1}{2}(\nu+n))}{\Gamma(\frac{1}{2}\nu)}\pi^{-\frac{1}{2}n}(\nu-2)^{-\frac{1}{2}(n-1)}|\underset{\sim}{V}_{n-1}|^{-\frac{1}{2}}Q^{-\frac{1}{2}(\nu+n-1)}\int_{\theta=0}^{\frac{1}{2}\pi}\cos^{\nu+n-2}\theta\,d\theta$$

$$= \frac{\Gamma(\frac{1}{2}(\nu+n))}{\Gamma(\frac{1}{2}\nu)}\pi^{-\frac{1}{2}n}(\nu-2)^{-\frac{1}{2}(n-1)}|\underset{\sim}{V}_{n-1}|^{-\frac{1}{2}}Q^{-\frac{1}{2}(\nu+n-1)}\frac{\Gamma(\frac{1}{2}(\nu+n)-\frac{1}{2})}{\Gamma(\frac{1}{2}(\nu+n))}\pi^{\frac{1}{2}}$$

$$p(y_{n-1}|V) = \frac{\Gamma(\frac{1}{2}(\nu+n-1))}{\Gamma(\frac{1}{2}\nu)} \pi^{-\frac{1}{2}(n-1)} (\nu-2)^{-\frac{1}{2}(n-1)}$$
$$\times |V_{n-1}|^{-\frac{1}{2}} \left[1 + \frac{y'_{n-1} V_{n-1}^{-1} y_{n-1}}{\nu-2} \right]^{-\frac{\nu+n-1}{2}}$$

A marginal distribution is thus also a general multivariate t distribution with ν degrees of freedom, and with a dispersion matrix V_{n-1}, ie that formed by deleting rows and columns out of V. If y_n is not independent of y_{n-1} then by completing the square the same argument holds. Property iv) thus holds.

2.4.2 Conditional Densities from a general multivariate t distribution

Now consider the conditional distributions. For simplicity, again take $V = \sigma^2 I_n$ and let $r^2 = \sum_{i=1}^{n} y_i^2$, $r_1^2 = \sum_{i=1}^{m} y_i^2$ and $r_2^2 = \sum_{m+1}^{n} y_i^2$. Then:

$$p(y_{m+1} \cdots y_n | y_1 \cdots y_m, \sigma^2) = \frac{\Gamma(\frac{1}{2}(\nu+n))(\pi\sigma^2)^{-\frac{1}{2}n}(\nu-2)^{-\frac{1}{2}n}(1 + r^2/(\nu-2)\sigma^2)^{-\frac{1}{2}(\nu+n)}}{\Gamma(\frac{1}{2}(\nu+m))(\pi\sigma^2)^{-\frac{1}{2}m}(\nu-2)^{-\frac{1}{2}m}(1 + r_1^2/(\nu-2)\sigma^2)^{-\frac{1}{2}(\nu+m)}}$$

let $Q = 1 + r_1^2/(\nu-2)\sigma^2$, and $\nu^\dagger = \nu+m$, then by rearranging:

$$p(y_{m+1} \cdots y_n | y_1 \cdots y_m, \sigma^2)$$
$$= \frac{\Gamma(\frac{1}{2}(\nu+n))}{\Gamma(\frac{1}{2}(\nu+m))} (\pi\sigma^2)^{-\frac{1}{2}(n-m)} (\nu-2)^{-\frac{1}{2}(n-m)} \frac{(Q + r_2^2/(\nu-2)\sigma^2)^{-\frac{1}{2}(\nu+n)}}{Q^{-\frac{1}{2}(\nu+m)}}$$
$$= \frac{\Gamma(\frac{1}{2}(\nu+n))}{\Gamma(\frac{1}{2}(\nu+m))} (\pi\sigma^2 Q)^{-\frac{1}{2}(n-m)} (\nu-2)^{-\frac{1}{2}(n-m)} \left[1 + \frac{r_2^2/Q\sigma^2}{\nu-2} \right]^{-\frac{\nu^\dagger+n-m}{2}}$$

This is a general multivariate t distribution with $\nu^\dagger = \nu+m$ degrees of freedom on $n-m$ observations, with a variance of $\sigma^2 Q(\nu-2)/(\nu^\dagger-2)$ rather than σ^2. Thus property iii) almost holds. The factor of $Q(\nu-2)/(\nu^\dagger-2)$ leaves the variance unchanged if the observed y_i values have variance σ^2. If however the observed y_i have a greater variance than σ^2, then this factor increases the variance for the remaining y_i, and conversely the factor shrinks the variance for the remaining y_i if the observed y_i have a lower variance than σ^2. More interesting, the conditional distribution is *more nearly normal* than the joint distribution, due to the increase in the number of degrees of freedom.

The increase in variance if the observed $y'y \gg m\sigma^2$ is to be expected since a conditional slice through the tail of a heavy tailed density will have greater variance then a conditional slice through the mean of a heavy tailed density.

2.4.3 Distribution of the fixed effects conditional on the data

In this section an analogous result to property v) is produced for multivariate t distribution. Start with an n-vector of data y, and a p-vector of fixed effects a. Apply Bayes theorem:

$$p(a|\sigma^2,y) = p(y|\sigma^2,a)p(a)$$

where $p(a)$ is the prior distribution for a. Assuming a uniform prior on a this gives:

$$p(a|\sigma^2,y) \propto \left(1 + \frac{(y-Xa)'V^{-1}(y-Xa)}{\nu-2}\right)^{-\frac{\nu+n}{2}}$$

$$\propto \left(1 + \frac{(a-B)'A^{-1}(a-B) + C}{\nu-2}\right)^{-\frac{\nu+n}{2}}$$

where $A^{-1} = X'V^{-1}X$, $B = (X'V^{-1}X)^{-1}X'V^{-1}y$,
and $C = y'V^{-1}y - y'V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}y$.

$$p(a|\sigma^2,y) \propto \left(1 + \frac{C}{\nu-2}\right)^{-\frac{\nu+n}{2}} \left(1 + \frac{(a-B)'A^{-1}(a-B)/(1+\frac{C}{\nu-2})}{\nu-2}\right)^{-\frac{\nu+n}{2}}$$

$$\propto \left(1 + \frac{C}{\nu-2}\right)^{-\frac{\nu+n}{2}} \left(1 + \frac{\frac{\nu^\dagger-2}{\nu-2}(a-B)'A^{-1}(a-B)/(1+\frac{C}{\nu-2})}{\nu^\dagger-2}\right)^{-\frac{\nu^\dagger+p}{2}}$$

where $\nu^\dagger = \nu + n - p$. This is a multivariate t distribution with mean B and dispersion matrix $A(1+\frac{C}{\nu-2})\frac{\nu-2}{\nu^\dagger-2}$, and $\nu^\dagger$ degrees of freedom. Hence substituting for A, B and C:

$$y \sim t_\nu(Xa,V) \longrightarrow a \sim t_{\nu^\dagger}(B, A(\nu-2+y'V^{-1}y-y'V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}y)/\nu^\dagger-2)$$

Thus property v) from Section 2.3 is partially kept. The mean of this distribution is the same as that from a Normal distribution, but the dispersion matrix has been scaled. However, the evaluation of this density is still dominated by the same manipulations of V and $X'V^{-1}X$ as in the multivariate normal case. Thus it is

computationally no more difficult than the multivariate normal.

Note that usually there will be far more data values than fixed effects so that $n \gg p$. This gives $\nu^{\dagger} \gg \nu$. Even if there are very few degrees of freedom for the original t distribution on the data, the conditional distribution for the fixed effects will usually have many degrees of freedom. As a consequence the conditional distribution for the fixed effects will often thus look *roughly normal* even when the error distribution is distinctly heavy tailed.

2.5 Multivariate distributions with independence

Now consider the class of distributions that are produced if property i) spherical symmetry is dropped and property ii) independence is kept. These properties can be accomplished simply by setting the multivariate distribution equal to the product of n identical independently distributed univariate distributions. Thus:

$$p(Y_1=y_1 \ldots Y_n=y_n) = \prod_{i=1}^{n} p(Y_i=y_i)$$

However if the tails of the univariate distribution are heavier than the tails for a normal distribution, then this probability density function will have *star-shaped* non-convex contours. Also in general the sufficient statistics for the observations will be the entire y vector.

2.6 Multivariate distributions with spherical symmetry

Finally consider the class of distributions that are produced if property i) spherical symmetry is kept, and property ii) independence is dropped. Any density function of the form $f(y'y)$ will satisfy this, for example the general multivariate t distribution or the multivariate normal distribution. Any conditional distribution from a probability density function of the form $f(y'y)$, will also depend on y only through $y'y$, but in general the conditional distribution will have a different density function.

Consider the conditional distribution for $(y_{m+1}\cdots y_n|y_1\cdots y_m)$. Let $r^2 = \sum_{i=1}^{n} y_i^2$, $r_1^2 = \sum_{i=1}^{m} y_i^2$ and $r_2^2 = \sum_{m+1}^{n} y_i^2$. Then the conditional distribution has the form:

$$p(y_{m+1}\cdots y_n|y_1\cdots y_m) = f(r^2)/f(r_1^2) = f(r_1^2 + r_2^2)/f(r_1^2) \propto f(r_1^2 + r_2^2)$$

This will have a different form to $f(\ )$ unless the density $f(\ )$ is of the form $f(a_n + b_n r^2)$. If the density function is of this form, then the conditional distribution for $(y_{m+1}\cdots y_n|y_1\cdots y_m)$ will be of the same form but with a_n replaced by $a_{n-m} = a_n + b_n r_1^2$, and $b_{n-m} = b_n$. Now suppose $f(x) \propto x^{-c_n}$, then without loss of generality take $a_n = a_{n-m} = 1$. The density function is then:

$$p(y_1\cdots y_n) \propto (1+b_n r^2)^{-c_n} \qquad (2.6:1)$$

The conditional distribution will be of the form:

$$p(y_{m+1}\cdots y_n|y_1\cdots y_m) \propto (1+b_n r^2)^{-c_n} / (1+b_m r_1^2)^{-c_m}$$

$$\propto (1+b_n r_1^2 + b_n r_2^2)^{-c_n}$$

$$\propto (1+b_{n-m} r_2^2)^{-c_{n-m}}$$

where $b_{n-m} = \dfrac{b_n}{1+b_n r_1^2}$, and $c_{n-m} = c_n$

The change from b_n to b_{n-m} shows that though the distributional form has been preserved, the variance of the conditional distribution changes, depending on the observed y_i values. Thus property iii) is only partially kept. Similarly marginal distributions will be of the same form, but with (potentially) different variances.

Now reparametrise equation 2.6:1. Assuming $c_n > \frac{1}{2}n$; take $\nu = 2c_n - n$, and $\sigma^2 = 1/\nu b_n$. Then equation 2.6:1 represents a general multivariate t distribution with ν degrees of freedom, and a dispersion matrix $V = \sigma^2 I_n$. Clearly by letting $\nu \longrightarrow \infty$, equation 2.6:1 tends to a multivariate normal distribution.

The significance of the assumption that $f(x) = x^{-c_n}$ can now be seen. This assumption leads to $f(\ )$ representing a general multivariate t distribution. Generally (assuming sufficient regularity) $1/f(x)$ can be expressed an even polynomial, and hence correspond to a mixture of general multivariate t distributions.

The class of scale mixtures of general multivariate t distributions is the same as the class of scale mixtures of multivariate Normal distributions. Clearly since the

multivariate Normal distribution is a limiting form of the general multivariate t distribution, the class of scale mixtures of multivariate Normal distributions is a subset of the class of scale mixtures of general multivariate t distributions. Conversely any general multivariate t distribution can be viewed as a scale mixture of multivariate Normal distributions, so a scale mixture of t distributions must also be a scale mixture of Normal distributions.

Dickey and Chen (1983) consider other tractable spherical distributions by considering other forms for the radial density $f(\ )$.

Thus spherical symmetry combined with conditionals of the same form leads to scale mixtures of multivariate Normal distributions. In the subsequent analyses only multivariate Normal distributions, and general multivariate t distributions are considered rather than arbitrary scale mixtures. Allowing multivariate Normal distributions to be replaced by general multivariate t distributions allows heavier tailed distributions to be considered with only the loss of the independence property of the multivariate Normal, and with only one extra parameter, the number of degrees of freedom ν - which determines the mixing weights.

Since independence does not apply, it is intriguing to note the effect of conditioning on the first m out of n observations. What then can be deduced about $y_{m+1} \ldots y_n$ given $y_1 \ldots y_m$. Suppose that the m observed values all lie far from zero, so that $r_1^2 \gg m\sigma^2$. Given independence the variance of the remaining y_i must equal σ^2. Without independence, one might reasonably expect the variance of the remaining y_i to be increased, so that $y_{m+1} \ldots y_n$ agree better with the observed data $y_1 \ldots y_m$. If $r_1^2 \gg m\sigma^2$, then b_{n-m} is reduced compared with b_n, corresponding to an increase in variance.

Chapter 3
Characterisation of the Likelihood Evaluation Problem

3.1 Introduction

Chapter 2 presented the standard algebraic methods for integrating out location parameters to yield marginal distributions on just the variance components. Within the class of models with normally distributed errors, and also in the wider class of spherically symmetric error distributions, eg. general multivariate t distributions, it was shown that efficient evaluation of the likelihood function required efficient evaluation V^{-1} and $|V|$, and also $(X'V^{-1}X)^{-1}$ and $|X'V^{-1}X|$, where V is the dispersion matrix, and X is the design matrix. This chapter explores efficient methods of evaluating these quantities, and demonstrates the techniques on a range of common models.

The first section of the chapter defines the notation used for 10 common variance component models that are used as examples throughout the chapter. Using the results of Searle and Henderson (1979), it is shown in section 3.2 that the dispersion matrices from balanced factorial dispersion models are easy to invert, and hence such models are numerically easy to handle. A simple characterisation of the fixed effects structure also indicates whether there is a general analytic form for the inverse of the $X'V^{-1}X$ matrix which occurs in the marginal posterior density for the variance components. A recursive method is also introduced for the implicitly unbalanced $(X_2'V^{-1}X_2)$ matrix that occurs in section 2.2.2 of Chapter 2. Unbalanced factorial models are considered in section 3.3. A new classification of slightly unbalanced models is introduced for models that would be balanced except that they have an odd number of observations in one cell. An analytic technique is developed in section 3.4 for finding the inverse and determinant of the dispersion matrix from slightly unbalanced factorial models, enabling efficient numerical evaluation. A variant of this technique in used in section 3.5 to give a recursive method of handling unbalanced nested models. This produces results for the same models as considered by La Motte (1972). Illustrations of all these unbalanced inversion techniques are given. Section 3.6 considers grossly unbalanced crossed designs without interaction terms. For these models an analytic inverse is not found, but techniques are developed which substantially reduce the numerical difficulty. Finally section 3.7 examines the computational implications of each of these methods.

3.1.1 Notation

Throughout this chapter the symbols I_n, J_n, 1_n, 0_n, C_n, $1_{a,b}$, $0_{a,b}$, $\otimes$, $diag(,,)$ and D[,,] are used extensively. I_n denotes an $n \times n$ identity matrix, J_n is an $n \times n$ matrix of ones, and 1_n (0_n) is a column of n ones (zeroes). Similarly $1_{a,b}$ ($0_{a,b}$) is used for an $a \times b$ array of ones (zeroes). Thus $J_n = 1_{n,n} = 1_n 1_n'$. C_n is used for a vector of n elements the first of which is one and the remaining elements are zero. $\otimes$ denotes Kronecker products. $diag(,,)$ is used for a diagonal matrix with prescribed elements. D[,,] denotes a block diagonal matrix with prescribed blocks. Note that D[,,] will in general not be a square matrix. For example $D[1_2, 1_2, 1_2] = I_3 \otimes 1_2$ is a matrix of 6 rows and 3 columns.

3.1.2 The Models considered in detail

This section lists the models considered in detail. They are denoted by the letter M followed by a digit and a letter. The digit represents the number of variance components in the model, and the letter serves to distinguish between different models with the same number of variance components.

All of the models listed assume multivariate Normal errors but the same techniques for manipulating V (and the matrices derived from V) apply to all spherical error distribution functions. The models detailed here are not the set of models for which the subsequent theory applies, but rather they form a set of examples from a general class of models.

M1) One Way Fixed Effects Model

$$y \sim N(Xa, V) \qquad V = \sigma_\varepsilon^2 I \qquad \text{i.e. } y_{ij} = a_i + \varepsilon_{ij}$$

where a is a vector of fixed effects.

M2a) 1 Way Random Effects Model

$$y \sim N(1\mu, V) \qquad V \equiv V(\sigma_a^2, \sigma_\varepsilon^2) \qquad \text{i.e. } y_{ij} = \mu + a_i + \varepsilon_{ij}$$

$$E[y_{ij}] = \mu + a_i \qquad var(y_{ij}) = \sigma_a^2 + \sigma_\varepsilon^2 \qquad cov(y_{ij}, y_{ik}) = \sigma_a^2$$

M2b) Bayesian hierarchical model; 1 factor, levels exchangeable

$$\left.\begin{array}{l} y|\theta_i \sim N(\theta_i, \sigma_\varepsilon^2) \\ \theta_i|\mu \sim N(\mu, \sigma_a^2) \\ \mu \sim \textit{uniform} \end{array}\right\} \begin{array}{l}\text{integrate} \\ \text{out } \theta_i\end{array} \longrightarrow y \sim N(1\mu, V) \qquad \text{as in model M2a}$$

M2c) Randomised Block Design

$$y \sim N(X\mu, V) \qquad V \equiv V(\sigma_b^2, \sigma_\varepsilon^2) \qquad \text{i.e. } y_{ijk} = \mu_i + \beta_j + \varepsilon_{ijk}$$

where β_j is random, and μ is a vector of fixed effects.
Several cases for replication equal to 1, greater than 1, or unbalanced

M2d) Arbitrary treatment structure - two level nested block structure

$$y \sim N(X\mu, V) \qquad V \equiv V(\sigma_b^2, \sigma_\varepsilon^2)$$

M3a) Classical 2 level nested random effects model

$$y = N(1\mu, V) \qquad V \equiv V(\sigma_a^2, \sigma_b^2, \sigma_\varepsilon^2) \qquad \text{i.e. } y_{ijk} = \mu + a_i + \beta_{ij} + \varepsilon_{ijk}$$

groups/subgroups/observations.

M3b) Bayesian 2 level nested hierarchical model

$$y_{ijk}|\mu_{ij} \sim N(\mu_{ij}, \sigma_\varepsilon^2)$$

$$\mu_{ij}|\alpha_i \sim N(\alpha_i, \sigma_b^2)$$

$$\alpha_i|\mu \sim N(\mu, \sigma_a^2)$$

$$\mu \sim \textit{uniform}$$

Integrating out parameters μ_{ij} and α_i yields model M3a

M3c) General treatment structure on a 2 factor nested blocking structure

$$y \sim N(X\mu, V) \qquad V \equiv V(\sigma_a^2, \sigma_b^2, \sigma_\varepsilon^2)$$

e.g. $y_{ijk} = a_j + b_k + \alpha_i + \beta_{ij} + \varepsilon_{ijk}$

or with interaction:

$$y_{ijk} = a_j + b_k + c_{jk} + \alpha_i + \beta_{ij} + \varepsilon_{ijk}$$

which can be reparametrised as:

$$y_{ijk} = \mu_{jk} + \alpha_i + \beta_{ij} + \varepsilon_{ijk}$$

M3d) Two random effects crossed model

$$y = N(1\mu, V) \qquad V \equiv V(\sigma_a^2, \sigma_b^2, \sigma_\varepsilon^2) \qquad \text{i.e. } y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk}$$

M3e) Bayesian hierarchical model 2 crossed factors

$$y_{ijk} \sim N(\mu + \alpha_i + \beta_j, \sigma_\varepsilon^2)$$

$$\mu \sim \textit{uniform}$$

$$\alpha_i \sim N(0, \sigma_a^2)$$

$$\beta_j \sim N(0, \sigma_b^2)$$

Integrating out parameters α_i and β_j yields model M3d.

3.1.3 Matrix Results

In the algebra, a large number of matrix manipulations are performed on unbalanced dispersion matrices. The following matrix identities are frequently used to obtain one inverse matrix in terms of another matrix inverse. Usually the latter inverse is a smaller or simpler matrix than the original, or is already known.

$$(A + HWH')^{-1} = A^{-1} - A^{-1}H(H'A^{-1}H + W^{-1})^{-1}H'A^{-1} \qquad (3.1.3:1)$$

Usually A will be a diagonal matrix, and hence evaluating A^{-1} is trivial. Also H is often a design matrix with only a single 1 in each row. This ensures that $T^{-1} = H'A^{-1}H$ is also a diagonal matrix. Even if A is not diagonal the resulting numerical inverse is of smaller size provided that H has more rows than columns. (This should always be the case if the model is sensible and a full rank parametrisation is used). If A and T are diagonal, then using equation 3.1.3:1 on $(H'A^{-1}H + W^{-1})^{-1} = (T^{-1} + W^{-1})^{-1}$ gives:

$$(A + HWH')^{-1} = A^{-1} - A^{-1}HTH'A^{-1} + A^{-1}HT(T + W)^{-1}TH'A^{-1} \qquad (3.1.3:2)$$

This result requires only one non-trivial inverse namely $(T + W)^{-1}$, whereas equation 3.1.3:1 requires two non-trivial inverses. Sometimes the W matrix has yet more structure and equation 3.1.3:1 or equation 3.1.3:2 can be used on $(T + W)^{-1}$ (recursively).

3.2 Completely Balanced Factorial Models

Historically, balanced factorial designs have been popular since they lead to orthogonal columns. For the Bayesian, balance is useful because it simplifies the algebra of the model, and also reduces the computational burden of evaluating the likelihood. Using Kronecker product notation (denoted by $\otimes$), it is possible to write the dispersion matrix V as the weighted sum of Kronecker products of identity matrices and square block matrices with common block sizes - see equation 3.2:1 below. Without loss of generality, all the examples presented in this chapter assume a multivariate Normal error distribution since all spherical error distributions lead to the same manipulations of the dispersion matrix.

In general a completely balanced factorial experiment with s-1 random (or blocking) factors, factor levels $r_1, r_2, \ldots, r_{s-1}$ and replication r_s has a dispersion

matrix:

$$V = \sum_{i_1 i_2 \ldots i_s} K_{i_1 i_2 \ldots i_s} J_{r_1}^{i_1} \otimes J_{r_2}^{i_2} \otimes \ldots \otimes J_{r_s}^{i_s} \qquad (3.2:1)$$

where the summation is over all the binary permutations of zero and one for $i_1 i_2 \ldots i_s$ and $J_p^0 = I_p$. This is the form of dispersion matrix considered by Searle and Henderson (1979). Note that no assumptions have been made about the fixed effect structure.

Multiplying two matrices of the form given in equation 3.2:1 together will give a third of the same form, and since the identity matrix is also of this form, it follows that V^{-1} must also be of this form. Finding V^{-1} can thus be viewed as being equivalent to solving 2^s simultaneous equations.

Frequently the design matrix X can be expressed as a Kronecker product of column vectors of ones and identity matrices. This is useful as $X'V^{-1}X$ is then of the same form as V, so evaluating $(X'V^{-1}X)^{-1}$ and $|X'V^{-1}X|$ is essentially the same problem as evaluating V^{-1} and $|V|$.

Using the structure of equation 3.2:1 for the dispersion matrix, the eigen-values and eigen-vectors of V can be produced analytically. The eigen-values of V are linear combinations of the K_i, with the multiplying constants being dependent on the number of levels for each factor $r_1, \ldots, r_{s-1}$ and on the number of replications r_s. This yields an analytical solution for V^{-1} and $|V|$, hence the joint likelihood can be evaluated efficiently. An extended description of finding the eigen-values and eigen-vectors of such patterned dispersion matrices can be found in Searle and Henderson (1979). Note that in a nested model with s-1 levels all the $K_{1,i_2,\ldots,i_s}$ coefficients are equal to zero. This can be used effectively to halve the computational task of evaluating V^{-1} using the analytic results.

3.2.1 Joint Posterior Density for the fixed effects and variance components

The joint posterior density for the fixed effects and variance components is a function of V^{-1}. For completeness, in this section, the form of V and V^{-1} is presented for all the models listed. Other models can be approached by the Searle and

Henderson method.

Example 1 Model M1

Consider the balanced one way fixed effects model M1, with p fixed effects $(a_1, a_2, \ldots, a_p)$, and q observations from each effect. Then $y_{ij} = a_i + \epsilon_{ij}$ for $i=1,2,\ldots,p$, $j=1,2,\ldots,q$ where ϵ_{ij} are independent identically distributed $\sim N(0, \sigma_\epsilon^2)$, that is:

$$y \sim N(Xa, V) \quad \text{with} \quad V = \sigma_\epsilon^2 I_p \otimes I_q \quad \text{and} \quad X = I_p \otimes 1_q.$$

Calculation of the inverse and determinant of V for this model is trivial, but it is presented below in the form from equation 3.2:1 with $r_1 = p$, $r_2 = q$

Coefficients of V	
$K_{00} = \sigma_\epsilon^2$	$K_{01} = 0$
$K_{10} = 0$	$K_{11} = 0$

Coefficients of V^{-1}	
$K_{00} = \frac{1}{\sigma_\epsilon^2}$	$K_{01} = 0$
$K_{10} = 0$	$K_{11} = 0$

giving $V^{-1} = \frac{1}{\sigma_\epsilon^2} I_p \otimes I_q$

Example 2 Model M2a

Consider the 1 way random effects model M2a. $y_{ij} = \mu + a_i + \epsilon_{ij}$. Again assume Normality so that $a_i \sim N(0, \sigma_a^2)$, and ϵ_{ij} are iid with $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$. This can be written as:

$$y \sim N(X\mu, V) \quad \text{with } V = \sigma_a^2 I_p \otimes J_q + \sigma_\epsilon^2 I_p \otimes I_q \quad \text{and } X = 1_p \otimes 1_q$$

hence in equation 3.2:1 $r_1 = p$, $r_2 = q$

Coefficients of V	
$K_{00} = \sigma_\epsilon^2$	$K_{01} = \sigma_a^2$
$K_{10} = 0$	$K_{11} = 0$

Coefficients of V^{-1}	
$K_{00} = \frac{1}{\sigma_\epsilon^2}$	$K_{01} = \frac{-\sigma_a^2}{\sigma_\epsilon^2(\sigma_\epsilon^2 + q\sigma_a^2)}$
$K_{10} = 0$	$K_{11} = 0$

giving $V^{-1} = \frac{1}{\sigma_\varepsilon^2} I_p \otimes I_q - \frac{\sigma_a^2}{\sigma_\varepsilon^2(\sigma_\varepsilon^2 + q\sigma_a^2)} I_p \otimes J_q$

The same result holds for Model M2b.

Example 3 Model M2c

The randomised block design

$$y \sim N(X\mu, V)$$

or $y_{ijk} = \mu_i + \beta_j + \varepsilon_{ijk}$

This has $X = I_p \otimes 1_q \otimes 1_r$ and $V = \sigma_b^2 J_p \otimes I_q \otimes J_r + \sigma_\varepsilon^2 I_p \otimes I_q \otimes I_r$.
hence in equation 3.2:1 $r_1 = p, r_2 = q, r_3 = r$

Coefficients of V			
$K_{000} = \sigma_\varepsilon^2$	$K_{001} = 0$	$K_{010} = 0$	$K_{011} = 0$
$K_{100} = 0$	$K_{101} = \sigma_b^2$	$K_{110} = 0$	$K_{111} = 0$

Coefficients of V^{-1}			
$K_{000} = \frac{1}{\sigma_\varepsilon^2}$	$K_{001} = 0$	$K_{010} = 0$	$K_{011} = 0$
$K_{100} = 0$	$K_{101} = \frac{-\sigma_b^2}{\sigma_\varepsilon^2(\sigma_\varepsilon^2 + r\sigma_b^2)}$	$K_{110} = 0$	$K_{111} = 0$

giving $V^{-1} = \frac{1}{\sigma_\varepsilon^2} I_p \otimes I_q \otimes J_r - \frac{\sigma_b^2}{\sigma_\varepsilon^2(\sigma_\varepsilon^2 + r\sigma_b^2)} J_p \otimes I_q \otimes J_r$

Example 4 Model M3a

The classical 2 level nested random effects model

$$y_{ijk} = \mu + a_i + \beta_{ij} + \varepsilon_{ijk}$$

or $y \sim N(X\mu, V)$

$$X = 1_p \otimes 1_q \otimes 1_r$$

and $V = \sigma_a^2 I_p \otimes J_q \otimes J_r + \sigma_b^2 I_p \otimes I_q \otimes J_r + \sigma_\varepsilon^2 I_p \otimes I_q \otimes I_r$

hence in equation 3.2:1 $r_1 = p, r_2 = q, r_3 = r$

Coefficients of V			
$K_{000} = \sigma_\varepsilon^2$	$K_{001} = \sigma_b^2$	$K_{010} = 0$	$K_{011} = \sigma_a^2$
$K_{100} = 0$	$K_{101} = 0$	$K_{110} = 0$	$K_{111} = 0$

Coefficients of V^{-1}			
$K_{000} = \frac{1}{\sigma_\varepsilon^2}$	$K_{001} = \frac{-\sigma_b^2}{\sigma_\varepsilon^2(\sigma_\varepsilon^2 + r\sigma_b^2)}$	$K_{010} = 0$	$K_{011} = \frac{-\sigma_a^2}{(\sigma_\varepsilon^2 + r\sigma_b^2)(\sigma_\varepsilon^2 + r\sigma_b^2 + qr\sigma_a^2)}$
$K_{100} = 0$	$K_{101} = 0$	$K_{110} = 0$	$K_{111} = 0$

giving

$$V^{-1} = \frac{-\sigma_a^2}{(\sigma_\varepsilon^2 + r\sigma_b^2)(\sigma_\varepsilon^2 + r\sigma_b^2 + qr\sigma_a^2)} I_p \otimes J_q \otimes J_r + \frac{-\sigma_b^2}{\sigma_\varepsilon^2(\sigma_\varepsilon^2 + r\sigma_b^2)} I_p \otimes I_q \otimes J_r + \frac{1}{\sigma_\varepsilon^2} I_p \otimes I_q \otimes I_r$$

Example 5 Model M3c

General treatment structure on a 2 factor nested blocking structure

e.g. $y_{ijk} = a_j + b_k + \alpha_i + \beta_{ij} + \varepsilon_{ijk}$

or $y_{ijk} = \mu_{jk} + \alpha_i + \beta_{ij} + \varepsilon_{ijk}$

$$X = 1_p \otimes I_q \otimes 1_r \,|\, 1_p \otimes 1_q \otimes I_r \quad \text{or } X = 1_p \otimes I_q \otimes I_r$$

$$V = \sigma_a^2 I_p \otimes J_q \otimes J_r + \sigma_b^2 I_p \otimes I_q \otimes J_r + \sigma_\varepsilon^2 I_p \otimes I_q \otimes I_r$$

clearly V^{-1} is the same as in M3a

Example 6 Model M3d

The two random effects crossed model

$$y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk}$$

or $y \sim N(X\mu, V)$

$$X = 1_p \otimes 1_q \otimes 1_r$$

and $V = \sigma_a^2 I_p \otimes J_q \otimes J_r + \sigma_b^2 J_p \otimes I_q \otimes J_r + \sigma_\varepsilon^2 I_p \otimes I_q \otimes I_r$

hence in equation 3.2:1 $r_1 = p, r_2 = q, r_3 = r$

Coefficients of V			
$K_{000} = \sigma_\varepsilon^2$	$K_{001} = 0$	$K_{010} = 0$	$K_{011} = \sigma_a^2$
$K_{100} = 0$	$K_{101} = \sigma_b^2$	$K_{110} = 0$	$K_{111} = 0$

Coefficients of V^{-1}			
$K_{000} = \frac{1}{\sigma_\varepsilon^2}$	$K_{001} = 0$	$K_{010} = 0$	$K_{011} = \frac{-\sigma_a^2}{\sigma_\varepsilon^2(\sigma_\varepsilon^2+q r\sigma_a^2)}$
$K_{100} = 0$	$K_{101} = \frac{-\sigma_b^2}{\sigma_\varepsilon^2(\sigma_\varepsilon^2+p r\sigma_b^2)}$		
$K_{110} = 0$	$K_{111} = \frac{-\sigma_a^2}{p\sigma_\varepsilon^2(\sigma_\varepsilon^2+q r\sigma_a^2)} - \frac{\sigma_a^2}{p(\sigma_\varepsilon^2+p r\sigma_b^2)(\sigma_\varepsilon^2+q r\sigma_a^2+p r\sigma_b^2)}$		

3.2.2 Marginal Posterior Density for the Variance Components

Recall equation 2.2.1:1 from Chapter 2:

$$p(\sigma^2|y) \propto |V|^{-\frac{1}{2}}|X'V^{-1}X|^{-\frac{1}{2}} \exp -\tfrac{1}{2}y'V^{-1}y \exp \tfrac{1}{2}y'V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}y\, p(\sigma^2)$$

This requires the use of $|X'V^{-1}X|$ and $(X'V^{-1}X)^{-1}$.

The design matrix X can often be expressed as $H_{r_1} \otimes H_{r_2} \otimes \ldots \otimes H_{r_s}$ where H_i represents either I_i or 1_i. Design matrices of this form always occur for factorial models unless there are crossed fixed effects without interaction terms - see Model M3c(i). If there are interaction terms, then after reparametrising, there can be a single fixed effect term for each cell and hence a X matrix of the above form - see Model M3c(ii).

Given a convenient X matrix, $|X'V^{-1}X|$ and $(X'V^{-1}X)^{-1}$ can be obtained easily since $X'V^{-1}X$ has the *same form* as V. Thus such X matrices lead to efficient computation of the marginal posterior density.

$$X'V^{-1}X = (H_{r_1} \otimes H_{r_2} \otimes \ldots \otimes H_{r_s})' \left(\sum \bar{K}_{i_1 i_2 \ldots i_s} J_{r_1}^{i_1} \otimes J_{r_2}^{i_2} \otimes \ldots \otimes J_{r_s}^{i_s} \right) (H_{r_1} \otimes H_{r_2} \otimes \ldots \otimes H_{r_s})$$

$$= \sum K_{i_1 i_2 \ldots i_s} H'_{r_1} J_{r_1}^{i_1} H_{r_1} \otimes H'_{r_2} J_{r_2}^{i_2} H_{r_2} \otimes \ldots \otimes H'_{r_s} J_{r_s}^{i_s} H_{r_s}$$

which is another matrix of the same form as equation 3.2:1, but with fewer levels because some of the H_i correspond to 1_i. Thus this matrix can be inverted in the same manner as V. Often the matrix $X'V^{-1}X$ is either a 1×1 matrix or a diagonal matrix. In either case the evaluation of $(X'V^{-1}X)^{-1}$ and $|X'V^{-1}X|$ is trivial.

Example 1 Model M1

The balanced one way fixed effects model M1

$$y \sim N(Xa, V) \quad \text{with} \quad V = \sigma_\varepsilon^2 I_p \otimes I_q \quad \text{and} \quad X = I_p \otimes 1_q.$$

hence $X'V^{-1}X = q\sigma_\varepsilon^2 I_p$

which is trivially of the form given by equation 3.2:1

$$(X'V^{-1}X)^{-1} = \frac{1}{q\sigma_\varepsilon^2} I_p$$

Example 2 Model M2a The 1 way random effects model.

$$y \sim N(X\mu, V) \quad \text{with} \quad V = \sigma_a^2 I_p \otimes J_q + \sigma_\varepsilon^2 I_p \otimes I_q \quad \text{and} \quad X = 1_p \otimes 1_q$$

hence $X'V^{-1}X = pq^2\sigma_a^2 + pq\sigma_\varepsilon^2$

$$(X'V^{-1}X)^{-1} = \frac{1}{pq(q\sigma_a^2 + \sigma_\varepsilon^2)}$$

Example 3 Model M2c The randomised block design

$$y \sim N(X\mu, V) \quad \text{or} \quad y_{ijk} = \mu_i + \beta_j + \varepsilon_{ijk}$$

$$V = \sigma_b^2 J_p \otimes I_q \otimes J_r + \sigma_\varepsilon^2 I_p \otimes I_q \otimes I_r \quad \text{and} \quad X = I_p \otimes 1_q \otimes 1_r$$

hence $X'V^{-1}X = qr^2\sigma_b^2 J_p + qr\sigma_\varepsilon^2 I_p$

$$(X'V^{-1}X)^{-1} = \frac{-\sigma_b^2}{q\sigma_\varepsilon^2(\sigma_\varepsilon^2 + pr\sigma_b^2)} J_p - \frac{1}{qr\sigma_\varepsilon^2} I_p$$

Example 4 Model M3a The classical 2 level nested random effects model

$$\boldsymbol{y} \sim N(\boldsymbol{X}\mu, \boldsymbol{V}) \quad \text{or} \quad y_{ijk} = \mu + a_{i-} + \beta_{ij} + \varepsilon_{ij}$$

$$\boldsymbol{V} = \sigma_a^2 \boldsymbol{I}_p \otimes \boldsymbol{J}_q \otimes \boldsymbol{J}_r + \sigma_b^2 \boldsymbol{I}_p \otimes \boldsymbol{I}_q \otimes \boldsymbol{J}_r + \sigma_\varepsilon^2 \boldsymbol{I}_p \otimes \boldsymbol{I}_q \otimes \boldsymbol{I}_r \quad \text{and} \quad \boldsymbol{X} = 1_p \otimes 1_q \otimes 1_r$$

hence $\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X} = pq^2r^2\sigma_a^2 + pqr^2\sigma_b^2 + pqr\sigma_\varepsilon^2$

$$(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1} = \frac{1}{pqr(qr\sigma_a^2 + r\sigma_b^2 + \sigma_\varepsilon^2)}$$

Example 5 Model M3c General treatment structure on a 2 factor nested blocking structure

e.g. $y_{ijk} = \mu_{jk} + \alpha_i + \beta_{ij} + \varepsilon_{ijk}$

$$\boldsymbol{V} = \sigma_a^2 \boldsymbol{I}_p \otimes \boldsymbol{J}_q \otimes \boldsymbol{J}_r + \sigma_b^2 \boldsymbol{I}_p \otimes \boldsymbol{I}_q \otimes \boldsymbol{J}_r + \sigma_\varepsilon^2 \boldsymbol{I}_p \otimes \boldsymbol{I}_q \otimes \boldsymbol{I}_r \quad \text{and} \quad \boldsymbol{X} = 1_p \otimes \boldsymbol{I}_q \otimes \boldsymbol{I}_r$$

hence $\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X} = p\{\sigma_a^2 \boldsymbol{J}_q \otimes \boldsymbol{J}_r + \sigma_b^2 \boldsymbol{I}_q \otimes \boldsymbol{J}_r + \sigma_\varepsilon^2 \boldsymbol{I}_q \otimes \boldsymbol{I}_r\}$

giving

$$(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1} = \frac{1}{p}\left\{\frac{-\sigma_a^2}{(\sigma_\varepsilon^2 + r\sigma_b^2)(\sigma_\varepsilon^2 + r\sigma_b^2 + qr\sigma_a^2)}\boldsymbol{J}_q \otimes \boldsymbol{J}_r + \frac{-\sigma_b^2}{\sigma_\varepsilon^2(\sigma_\varepsilon^2 + r\sigma_b^2)}\boldsymbol{I}_q \otimes \boldsymbol{J}_r + \frac{1}{\sigma_\varepsilon^2}\boldsymbol{I}_q \otimes \boldsymbol{I}_r\right\}$$

Example 6 Model M3d The two random effects crossed model

$$\boldsymbol{y} \sim N(\boldsymbol{X}\mu, \boldsymbol{V}) \quad \text{or} \quad y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk}$$

$$\boldsymbol{V} = \sigma_a^2 \boldsymbol{I}_p \otimes \boldsymbol{J}_q \otimes \boldsymbol{J}_r + \sigma_b^2 \boldsymbol{J}_p \otimes \boldsymbol{I}_q \otimes \boldsymbol{J}_r + \sigma_\varepsilon^2 \boldsymbol{I}_p \otimes \boldsymbol{I}_q \otimes \boldsymbol{I}_r \quad \text{and} \quad \boldsymbol{X} = 1_p \otimes 1_q \otimes 1_r$$

hence $\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X} = pq^2r^2\sigma_a^2 + p^2qr^2\sigma_b^2 + pqr\sigma_\varepsilon^2$

$$(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1} = \frac{1}{pqr(qr\sigma_a^2 + pr\sigma_b^2 + \sigma_\varepsilon^2)}$$

A model that does not conveniently fit into this form is the 3 factor crossed model. If there are no interaction terms, or only first order interaction terms, then $\boldsymbol{X}$ cannot be expressed as the Kronecker product of $\boldsymbol{H}$ matrices, and the matrix $\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}$ is difficult to handle. If there are first and second order interaction terms, then after a reparametrisation to get a single fixed effect in each cell, an $\boldsymbol{X}$ matrix of the desired form is obtained. This issue is returned to in Chapter 8.

3.2.3 Joint Posterior Density for a fixed effect and the Variance Components

As shown in section 2.2.2 of Chapter 2, the direct way of evaluating these posteriors is to multiply the posterior density for the variance components by the density for the fixed effects conditional on the variance components. This introduces no new numerical problems. It is interesting to consider the alternative formulation developed in the same section of Chapter 2. This involves the inverse of the potentially unbalanced matrix $X_2'V^{-1}X_2$ formed by the deletion of a row and a column from $X'V^{-1}X$. This unbalanced matrix makes the evaluation difficult. A method is developed here for evaluating $(X_2'V^{-1}X_2)^{-1}$ and $|X_2'V^{-1}X_2|$. This method leads naturally to a method of handling more general unbalanced dispersion matrices as developed in sections 3.4 (slightly unbalanced factorial models) and 3.5 (grossly unbalanced nested factorial models).

The method for handling $X_2'V^{-1}X_2$ is illustrated by considering Model M3c(ii). This has a row, column and interaction treatment structure on a 2 factor nested blocking structure, and is the simplest model considered for which $X'V^{-1}X$ is non-diagonal, and hence the simplest model for which $X_2'V^{-1}X_2$ is unbalanced. Recall:

$$V^{-1} = \beta I_p \otimes J_q \otimes J_r + \gamma I_p \otimes I_q \otimes J_r + \delta I_p \otimes I_q \otimes I_r$$

hence $X'V^{-1}X = p\delta I_q \otimes I_r + p\gamma I_q \otimes J_r + p\beta J_q \otimes J_r$

thus $X_2'V^{-1}X_2 = p\delta I_{qr-1} + p\gamma D[J_{r-1}, J_r, \ldots, J_r] + p\beta J_{qr-1}$ (3.2.3:1)

In general, it can be seen that $X_2'V^{-1}X_2$ will be the sum of a diagonal matrix and a matrix of blocks of differing sizes (caused by the deletion of a row and a column). The second term can be expressed as $H_t V_t H_t'$ where V_t is the balanced matrix formed by collapsing the blocks of differing sizes to a single element, and $H_t = D[1_{k_1}, 1_{k_2}, \ldots]$ indicates the sizes of the blocks. Thus:

$$X_2'V^{-1}X_2 = \Lambda_t + H_t V_t H_t' \qquad (3.2.3:2)$$

where Λ_t is a constant times an identity matrix. The form of this decomposition is crucial to the success of the method here, and the related methods for slightly unbalanced models, and grossly unbalanced nested models. Note that the balanced matrix V_t has one fewer level (in the sense of Searle and Henderson) than $X_2'V^{-1}X_2$.

Equation 3.2.3:2 can be inverted analytically using equation 3.1.3:2 yielding:

$$(X_2'V^{-1}X_2)^{-1} = \Lambda_t^{-1} - \Lambda_t^{-1}H_tT_tH_t'\Lambda_t^{-1} + \Lambda_t^{-1}H_tT_t(T_t + V_t)^{-1}T_tH_t'\Lambda_t^{-1} \qquad (3.2.3:3)$$

where T_t is the diagonal matrix $(H_t'\Lambda_t^{-1}H_t)^{-1}$. Note that $T_t + V_t$ is also the sum of a diagonal matrix and a balanced matrix of blocks. Thus it can be expressed as $\Lambda_{t-1} + H_{t-1}V_{t-1}H_{t-1}'$, where Λ_{t-1} is a diagonal matrix and V_{t-1} is a balanced matrix. The procedure recurses trivially.

The calculation of $|X_2V^{-1}X_2'|$ proceeds as follows. Clearly $X_2'V^{-1}X_2$ has a large number of eigenvalues of σ_ε^2 corresponding to the replications. All the other eigenvalues correspond to eigenvectors of the form $v_t = Hv_{t-1}$. These remaining eigenvectors are the eigenvectors of the matrix $H'(X_2V^{-1}X_2')H$ divided by the replications. This form of determinant is calculated (recursively) in Section 3.4.1

3.3 Unbalanced Factorial Models

In this section, and the following ones, the effect of unbalance is explored. Two different types of unbalance should be distinguished. The first type is where a design was originally balanced, but one (or several) observations have been lost/are missing/were discarded as outliers or a few additional observations are available, so that the number of replications varies in one (or a few) cells. The second type of unbalance occurs when no attempt at balance has ever been made, for example in a two level nested model where the number of subgroups varies, and the number of observations in each subgroup also varies. For clarity, the former designs are referred to as being *slightly unbalanced*, whereas the latter designs are *grossly unbalanced.*

A new class of slightly unbalanced models is introduced in section 3.4. It is shown that all models that have only a *single odd cell* (too few or too many observations in just one cell) can be tackled analytically. The inverse and determinant of the dispersion matrix from such models are produced analytically. This enables efficient numerical evaluation of these models.

Grossly unbalanced nested models are tackled in Section 3.5. An algebraic technique is developed that calculates the inverse and determinant of the dispersion matrix from an unbalanced nested model in terms of another dispersion matrix with one fewer level. The technique can be applied recursively to yield an analytic form for the inverse and determinant of the dispersion matrix. The theory from this

section contains as special cases some of the slightly unbalanced models which were considered in Section 3.4 - namely the slightly unbalanced nested models.

Grossly unbalanced crossed models are considered in Section 3.6. For these models complete analytic solutions are not available for the inverse or the determinant of the dispersion matrix. However, algebraic techniques are presented that reduce the numerical difficulties of evaluating the likelihood, by substantially reducing the size of the matrix that must be inverted numerically.

3.4 Slightly Unbalanced Factorial Models

This section develops a new class of unbalanced models, namely those models that would be balanced except that one cell has an odd number of observations. These models are called "slightly unbalanced models". Section 3.4.1 develops procedures for evaluating the inverse and determinant of such dispersion matrices. These are used in section 3.4.2 for the slightly unbalanced one way random effects model. Section 3.4.3 produces the inverse and determinant of the dispersion matrix for a slightly unbalanced 2 level nested random effects model, and section 3.4.4 does the same for the slightly unbalanced 2 random effects crossed model. These worked examples demonstrate the technique for handling slightly unbalanced models, though the procedure is directly applicable to any slightly unbalanced model.

The basic technique used in the derivation of the inverse is to strip the residual variance from the dispersion matrix leaving a block structured matrix. This is collapsed to a balanced matrix by taking a single entry from each block. The balanced matrix can be inverted analytically, and from it the inverse of the original dispersion matrix can be derived.

Similarly the determinant is found by extracting all the eigenvalues corresponding to the residual variance, then finding a simpler matrix whose determinant is equal to the product of the remaining eigenvalues.

3.4.1 Slightly Unbalanced Factorial Models with one odd cell

Consider a completely balanced factorial model with s-1 levels (in the sense of equation 3.2:1) with dispersion matrix V_s, where V_s is a $n_s \times n_s$ matrix. Now suppose that there are m observation(s) missing from the first cell. Denote the resulting dispersion matrix by $V_s^{\uparrow}$, and the data by y_s. Let r be the number of observations in each of the n_{s-1} cells except the first which has $r-m > 0$ observations. Then the $n_{s-1}r-m \times n_{s-1}r-m$ dispersion matrix is:

$$V_s^{\uparrow} = \sigma_\varepsilon^2 I + H V_{s-1} H'$$

where $H = D[1_{r-m}, 1_r, 1_r, \ldots, 1_r]$ and V_{s-1} is the $n_{s-1} \times n_{s-1}$ balanced matrix formed from taking one element from each block of $V_s^{\uparrow} - \sigma_\varepsilon^2 I$.

Calculation of $V_{s-1}^{\uparrow -1}$

Using equation 3.1.3:1 on $V_s^{\uparrow}$:

$$V_s^{\uparrow -1} = \frac{1}{\sigma_\varepsilon^2} I - \frac{1}{\sigma_\varepsilon^2} H(H'H + \sigma_\varepsilon^2 V_{s-1}^{-1})^{-1} H'$$

$$= \frac{1}{\sigma_\varepsilon^2} I - \frac{1}{\sigma_\varepsilon^2} H(\sigma_\varepsilon^2 V_{s-1}^{-1} + diag(r-m, r, r, \ldots, r))^{-1} H' \qquad (3.4.1:1)$$

But $\sigma_\varepsilon^2 V_{s-1}^{-1} + diag(r-m, r, r, \ldots, r) = S - diag(m, 0, \ldots, 0) = S - \frac{m}{n_{s-1}} TT'$

where S is the balanced matrix given by: $S = \sigma_\varepsilon^2 V_{s-1}^{-1} + r I_{n_{s-1}}$ (3.4.1:2)

and T is an $n_{s-1} \times n_{s-1}$ matrix whose first row is ones and all other entries are zeroes. Thus $T = 1'_{n_{s-1}} \otimes C_{n_{s-1}}$ where $C_{n_{s-1}}$ is a vector of length n_{s-1} with the first element as one, and the other elements as zero. Using equation 3.1.3:1

$$(\sigma_\varepsilon^2 V_{s-1}^{-1} + diag(r-m, r, r, \ldots, r))^{-1} = S^{-1} - S^{-1} T \left(T' S^{-1} T - \frac{n_{s-1}}{m} I_{n_{s-1}} \right)^{-1} T' S^{-1}$$

$$= S^{-1} - S^{-1} T (s_{11} J_{n_{s-1}} - \frac{n_{s-1}}{m} I_{n_{s-1}})^{-1} T' S^{-1} \qquad (3.4.1:3)$$

where s_{11} is the top left element of S^{-1}.

But $(s_{11} J_{n_{s-1}} - \frac{n_{s-1}}{m} I_{n_{s-1}})^{-1} = -\frac{m}{n_{s-1}} I_{n_{s-1}} - (\frac{m}{n_{s-1}})^2 \frac{s_{11}}{1 - m s_{11}} J_{n_{s-1}}$

thus $T(s_{11}J_{n_{s-1}} + \frac{n_{s-1}}{m}I_{n_{s-1}})^{-1}T' = \{-m - m^2\frac{s_{11}}{1-ms_{11}}\}diag(1,0\ldots0) = \frac{-m}{1-ms_{11}}\frac{TT'}{n_{s-1}}$

substituting this into equation 3.4.1:3 gives:

$$(\sigma_\varepsilon^2 V_{s-1}^{-1} + diag(r-m,r,r,\ldots,r))^{-1} = S^{-1} + \frac{m}{1-ms_{11}}S^{-1}\frac{TT'}{n_{s-1}}S^{-1}$$

$$= S^{-1} + \frac{m}{1-ms_{11}}\frac{(1'_{n_{s-1}} \otimes S_1)(1'_{n_{s-1}} \otimes S_1)'}{n_{s-1}}$$

where S_1 is the column vector formed by taking the first column of S^{-1}. Substituting for $(\sigma_\varepsilon^2 V_{s-1}^{-1} + diag(r-m,r,r,\ldots,r))^{-1}$ in equation 3.4.1:1 gives:

$$V_s^{\uparrow -1} = \frac{1}{\sigma_\varepsilon^2}I - \frac{1}{\sigma_\varepsilon^2}H(S^{-1} + \frac{m}{1-ms_{11}}\frac{(1'_{n_{s-1}} \otimes S_1)(1'_{n_{s-1}} \otimes S_1)'}{n_{s-1}})H'$$

$$= \frac{1}{\sigma_\varepsilon^2}I - \frac{1}{\sigma_\varepsilon^2}HS^{-1}H' - \frac{1}{\sigma_\varepsilon^2}\frac{m}{1-ms_{11}}H\frac{(1'_{n_{s-1}} \otimes S_1)(1'_{n_{s-1}} \otimes S_1)'}{n_{s-1}}H' \qquad (3.4.1:4)$$

All the terms in equation 3.4.1:4 are easy to evaluate as S is balanced. Thus $V_s^{\uparrow -1}$ has been produced analytically and can be evaluated without too much numerical work.

$$y'V^{-1}y = y_s'V_s^{\uparrow -1}y_s$$

$$= \frac{1}{\sigma_\varepsilon^2}y_s'y_s - \frac{1}{\sigma_\varepsilon^2}y_{s-1}'S^{-1}y_{s-1} - \frac{1}{\sigma_\varepsilon^2}\frac{m}{1-ms_{11}}y_{s-1}'\frac{(1'_{n_{s-1}} \otimes S_1)(1'_{n_{s-1}} \otimes S_1)'}{n_{s-1}}y_{s-1}$$

where $y_{s-1} = H'y_s$ is the vector formed from the sum of the y observations within each cell.

$$y_s'V_s^{\uparrow -1}y_s = \frac{1}{\sigma_\varepsilon^2}y_s'y_s - \frac{1}{\sigma_\varepsilon^2}y_{s-1}'S^{-1}y_{s-1} - \frac{1}{\sigma_\varepsilon^2}\frac{m}{1-ms_{11}}(S_1'y_{s-1})^2 \qquad (3.4.1:5)$$

As S^{-1} is known, $S_1'y_{s-1}$ and $y_{s-1}'S^{-1}y_{s-1}$ can be produced without effort - giving the sufficient statistics for the y.

Note - Calculation of S^{-1}

Note that though S was defined in equation 3.4.1:2 only S^{-1} is ever used. The definition of S involves V_{s-1}^{-1} so the calculation of S^{-1} requires the inversion of two (balanced) matrices. This is unnecessary and also assumes the existence of V_{s-1}^{-1}. Alternatively equation 3.1.3:1 can be applied to equation 3.4.1:2 yielding:

$$S^{-1} = \frac{1}{r} I_{n_{s-1}} - \frac{\sigma_\varepsilon^2}{r}(rV_{s-1} + \sigma_\varepsilon^2 I_{n_{s-1}})^{-1} \qquad (3.4.1:6)$$

This is a convenient representation as it only involves one inverse of a matrix that is balanced, and $rV_{s-1} + \sigma_\varepsilon^2 I_{n_{s-1}}$ is always non-singular.

Calculation of $|V_s^\uparrow|$

The likelihood evaluation also needs $|V_s^\uparrow|$. Clearly $V_s^\uparrow$ has $n_{s-1}(r-1)-m$ eigenvalues of σ_ε^2. The remaining n_{s-1} eigenvalues correspond to eigenvectors of the form $v_s = Hv_{s-1} = (k_1 1'_{r-m}, k_2 1'_r, k_3 1'_r, \ldots, k_{n_{s-1}} 1'_r)$, where $v_{s-1} = (k_1, k_2, k_3, \ldots, k_{n_{s-1}})$. Any eigenvector v_s of $V_s^\uparrow$ corresponds to an eigenvector v_{s-1} of $V_{s-1} + \sigma_\varepsilon^2 diag(\frac{1}{r-m}, \frac{1}{r}, \frac{1}{r}, \ldots, \frac{1}{r})$ with an eigenvalue scaled by $r-m$ or r. Thus:

$$|V_s^\uparrow| = (\sigma_\varepsilon^2)^{n_{s-1}(r-1)-m} \left| V_{s-1} + \sigma_\varepsilon^2 diag(\frac{1}{r-m}, \frac{1}{r}, \frac{1}{r}, \ldots, \frac{1}{r}) \right| (r-m) r^{n_{s-1}-1} \qquad (3.4.1:7)$$

There are two ways of evaluating the determinant on the right hand side of equation 3.4.1:7. The first method is applicable whenever the model is a nested model, and calculates the determinant recursively. The second method is applicable to all models, and works by comparing the eigenvalues and eigenvectors of the matrix on the right of equation 3.4.1:7 with those of the balanced matrix with no missing observations. The second method is more efficient, but is harder to formalise. Both techniques are presented.

Method 1

Note $V_{s-1} + \frac{\sigma_\varepsilon^2}{r} I_{n_{s-1}}$ is the balanced matrix that appeared in equation 3.4.1:2. Since the inverse of this has already been calculated, its determinant can be obtained trivially. The difference between this determinant, and the determinant on the right of equation 3.4.1:7 is $\sigma_\varepsilon^2(\frac{1}{r} - \frac{1}{r-m})|V_{s-1}^\uparrow|$, where $V_{s-1}^\uparrow$ is the matrix obtained by deleting the first row and column of $V_{s-1} + \frac{\sigma_\varepsilon^2}{r} I$. But $V_{s-1}^\uparrow$ is a matrix of exactly the same form as $V_s^\uparrow$ except that it has one fewer levels. Thus using this procedure recursively, one obtains $|V_s^\uparrow|$.

$$|V_s^{\dagger}| = (\sigma_\varepsilon^2)^{n_{s-1}(r-1)-m}\left\{\left|V_{s-1} + \frac{\sigma_\varepsilon^2}{r} I_{\tilde{n}_{s-1}}\right| + \frac{m\sigma_\varepsilon^2}{r(r-m)}\left|V_{s-1}^{\dagger}\right|\right\}(r-m)r^{n_{s-1}-1} \quad (3.4.1{:}8)$$

Method 2

An alternative method is now presented that is applicable to all models. For illustration consider the slightly unbalanced 2 level nested random effects model. This has:

$$|V_2^{\dagger}| = (\sigma_\varepsilon^2)^{pq(r-1)-m}\left|V_1 + \sigma_\varepsilon^2 diag(\frac{1}{r-m}, \frac{1}{r}, \frac{1}{r}, \ldots, \frac{1}{r})\right|(r-m)r^{pq-1}$$

But for the $\frac{1}{r-m}$ term, the matrix on the right of this would have $p(q-1)$ eigenvalues of $\frac{\sigma_\varepsilon^2}{r}$ and p of $q\sigma_a^2 + \frac{\sigma_\varepsilon^2}{r}$. The $\frac{1}{r-m}$ term destroys one of each of these eigenvalues, and the two new eigenvalues must have eigenvectors of the form $(1, g1'_{q-1}, 0'_{(p-1)q})'$. Multiplying out yields two equations in λ (the eigenvalue) and g.

$$\sigma_a^2(q-1)g = \lambda - \sigma_a^2 - \sigma_b^2 - \frac{\sigma_\varepsilon^2}{r-m}$$

$$\{\sigma_a^2(q-1) + \sigma_b^2 + \frac{\sigma_\varepsilon^2}{r} - \lambda\}g = -\sigma_a^2$$

Eliminating g yields a quadratic in λ.

$$\lambda^2 - \lambda(\sigma_a^2 + \sigma_b^2 + \frac{\sigma_\varepsilon^2}{r-m} + (q-1)\sigma_a^2 + \sigma_b^2 + \frac{\sigma_\varepsilon^2}{r}) + ((\sigma_a^2 + \sigma_b^2 + \frac{\sigma_\varepsilon^2}{r-m})(\sigma_a^2(q-1) + \sigma_b^2 + \frac{\sigma_\varepsilon^2}{r}) - (q-1)(\sigma_a^2)^2)$$

There is no need to solve this since the two λ values are of no interest in themselves - only their product is required and this must be:

$$(\sigma_a^2 + \sigma_b^2 + \frac{\sigma_\varepsilon^2}{r-m})(\sigma_a^2(q-1) + \sigma_b^2 + \frac{\sigma_\varepsilon^2}{r}) - (q-1)(\sigma_a^2)^2$$

$$= \frac{1}{r(r-m)}\{(\sigma_\varepsilon^2 + (r-m)\sigma_b^2)(\sigma_\varepsilon^2 + r\sigma_b^2 + qr\sigma_a^2) - m\sigma_a^2\sigma_\varepsilon^2\}$$

Comparison with section 3.4.3 shows that the same results are obtained.

3.4.2 The slightly unbalanced 1 Way Random Effects Model

This is an extension to model M2a - the completely balanced one way random effects model, and illustrates the technique for manipulating slightly unbalanced

models that was developed in Section 3.4.1.

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \qquad i = 1,\ldots,p \qquad j = 1,\ldots,n_i \qquad n_i = \begin{cases} q-m & i = 1 \\ q & i \neq 1 \end{cases}$$

$$V_2^{\dagger} = \sigma_\varepsilon^2 I_{pq-m} + \sigma_a^2 D[J_{q-m}, J_q, \ldots, J_q] = \sigma_\varepsilon^2 I_{pq-m} + HV_1H'$$

where $H = D[1_{q-m}, 1_q, \ldots, 1_q]$, and $V_1 = \sigma_a^2 I_p$.

Then from equation 3.4.1:2 $S = \frac{\sigma_\varepsilon^2}{\sigma_a^2} I_p + qI_p$, hence $S^{-1} = \frac{\sigma_a^2}{\sigma_\varepsilon^2 + q\sigma_a^2} I_p$. This gives $s_{11} = \frac{\sigma_a^2}{\sigma_\varepsilon^2 + q\sigma_a^2}$, and S_1 as a column vector of s_{11} and p-1 zeroes. Finally from equation 3.4.1:4

$$V_2^{\dagger -1} = \frac{1}{\sigma_\varepsilon^2} I - \frac{1}{\sigma_\varepsilon^2} s_{11} D[J_{q-m}, J_q, \ldots, J_q] - \frac{1}{\sigma_\varepsilon^2} \frac{m}{1 - ms_{11}} s_{11}^2 D[J_{q-m}, 0_{q,q}, \ldots, 0_{q,q}]$$

where $0_{q,q}$ is a $q \times q$ block of zeroes.

$$V_2^{\dagger -1} = \frac{1}{\sigma_\varepsilon^2} I - \frac{1}{\sigma_\varepsilon^2} D[\frac{s_{11}}{1 - ms_{11}} J_{q-m}, s_{11} J_q, \ldots, s_{11} J_q]$$

$$= \frac{1}{\sigma_\varepsilon^2} I - \frac{\sigma_a^2}{\sigma_\varepsilon^2} D[\frac{1}{\sigma_\varepsilon^2 + (q-m)\sigma_a^2} J_{q-m}, \frac{1}{\sigma_\varepsilon^2 + q\sigma_a^2} J_q, \ldots, \frac{1}{\sigma_\varepsilon^2 + q\sigma_a^2} J_q]$$

The determinant can also be calculated easily from equation 3.4.1:8

$$|V_2^{\dagger}| = (\sigma_\varepsilon^2)^{p(q-1)-m} \left\{ \left| V_1 + \frac{\sigma_\varepsilon^2}{q} I_p \right| + \frac{m\sigma_\varepsilon^2}{q(q-m)} |V_1^{\dagger}| \right\} (q-m) q^{p-1}$$

where $V_1^{\dagger}$ is obtained by deleting the first row and column from V_1.

But $V_1 = \sigma_a^2 I_p$, hence $V_1 + \frac{\sigma_\varepsilon^2}{q} I_p = \frac{\sigma_\varepsilon^2 + q\sigma_a^2}{q} I_p$, and $V_1^{\dagger} = \frac{\sigma_\varepsilon^2 + q\sigma_a^2}{q} I_{p-1}$

Substituting:

$$|V_2^{\dagger}| = (\sigma_\varepsilon^2)^{p(q-1)-m} \left\{ (\frac{\sigma_\varepsilon^2 + q\sigma_a^2}{q})^p + \frac{m\sigma_\varepsilon^2}{q(q-m)} (\frac{\sigma_\varepsilon^2 + q\sigma_a^2}{q})^{p-1} \right\} (q-m) q^{p-1}$$

$$= (\sigma_\varepsilon^2)^{p(q-1)-m} (\sigma_\varepsilon^2 + q\sigma_a^2)^{p-1} (\sigma_\varepsilon^2 + (q-m)\sigma_a^2)$$

Finally as $X'V_2^{\dagger -1}X$ is a 1×1 matrix, the calculation of $(X'V_2^{\dagger -1}X)^{-1}$ and $|X'V_2^{\dagger -1}X|$ present no difficulty.

3.4.3 The slightly unbalanced 2 level nested Random Effects Model with one odd cell

This is an extension to model M3a - the completely balanced two level nested random effects model.

$$y_{ijk} = \mu + \alpha_i + \beta_{ij} + \varepsilon_{ijk} \quad i = 1...p \quad j = 1...q \quad k = 1...n_{ij} \quad n_{ij} = \begin{cases} r-m & i,j = 1 \\ r & i,j \neq 1 \end{cases}$$

$$V_3^{\dagger} = \sigma_\varepsilon^2 I_{pqr-m} + \sigma_b^2 D[J_{r-m}, J_r, \ldots, J_r] + \sigma_a^2 D[J_{qr-m}, J_{qr}, \ldots, J_{qr}]$$

$$= \sigma_\varepsilon^2 I_{pqr-m} + HV_2H'$$

where $H = D[1_{r-m}, 1_r, \ldots, 1_r]$, and $V_2 = \sigma_b^2 I_p \otimes I_q + \sigma_a^2 I_p \otimes J_q$.

Using equation 3.4.1:6

$$S^{-1} = \frac{1}{r} I_{pq} - \frac{\sigma_\varepsilon^2}{r}(rV_2 + \sigma_\varepsilon^2 I_{pq})^{-1}$$

$$= \frac{\sigma_b^2}{\sigma_\varepsilon^2 + r\sigma_b^2} I_p \otimes I_q + \frac{\sigma_a^2 \sigma_\varepsilon^2}{(\sigma_\varepsilon^2 + r\sigma_b^2)(\sigma_\varepsilon^2 + r\sigma_b^2 + qr\sigma_a^2)} I_p \otimes J_q$$

$$= AI_p \otimes I_q + BI_p \otimes J_q$$

for appropriate choices of A and B.

thus $s_{11} = A + B$, and $S_1 = A \begin{pmatrix} 1 \\ 0_{pq-1} \end{pmatrix} + B \begin{pmatrix} 1_p \\ 0_{p(q-1)} \end{pmatrix}$.

Then substituting into equation 3.4.1:4

$$V_3^{\dagger -1} = \frac{1}{\sigma_\varepsilon^2} I_{pqr-m} - \frac{1}{\sigma_\varepsilon^2} AD[J_{r-m}, J_r \ldots J_r] - \frac{1}{\sigma_\varepsilon^2} BD[J_{qr-m}, J_{qr} \ldots J_{qr}] \qquad (3.4.3:1)$$

$$- \frac{1}{\sigma_\varepsilon^2} \frac{m}{1 - ms_{11}} D\left[\begin{pmatrix} (A+B)^2 J_{r-m} & (A+B)B\, 1_{(q-1)r,r-m} \\ (A+B)B\, 1_{r-m,(q-1)r} & B^2 J_{(q-1)r} \end{pmatrix}, 0_{qr,qr} \ldots 0_{qr,qr} \right]$$

To obtain $|V_3^{\dagger}|$ use equation 3.4.1:8

$$|V_3^{\dagger}| = (\sigma_\varepsilon^2)^{pq(r-1)-m} \left\{ \left| V_2 + \frac{\sigma_\varepsilon^2}{r} I \right| + \frac{m\sigma_\varepsilon^2}{r(r-m)} \left| V_2^{\dagger} \right| \right\} (r-m) r^{pq-1}$$

where $V_2^{\dagger}$ is the matrix $V_2 + \frac{\sigma_\varepsilon^2}{r} I$ with the first row and column deleted.

$$V_2 + \frac{\sigma_\varepsilon^2}{r} I = \frac{\sigma_\varepsilon^2 + r\sigma_b^2}{r} I_p \otimes I_q + \sigma_a^2 I_p \otimes J_q$$

$$|V_2 + \frac{\sigma_\varepsilon^2}{r} I| = (\frac{\sigma_\varepsilon^2 + r\sigma_b^2}{r})^{p(q-1)} (\frac{\sigma_\varepsilon^2 + r\sigma_b^2 + qr\sigma_a^2}{r})^p$$

$$= (\sigma_\varepsilon^2 + r\sigma_b^2)^{p(q-1)} (\sigma_\varepsilon^2 + r\sigma_b^2 + qr\sigma_a^2)^p / r^{pq}$$

Using equation 3.4.1:8 on $V_2^\dagger$

$$|V_2^\dagger| = (\frac{\sigma_\varepsilon^2 + r\sigma_b^2}{r})^{p(q-1)-1} \left\{ \left| \frac{\sigma_\varepsilon^2 + r\sigma_b^2 + qr\sigma_a^2}{qr} I_p \right| + \frac{\sigma_\varepsilon^2 + r\sigma_b^2}{q(q-1)r} \left| \frac{\sigma_\varepsilon^2 + r\sigma_b^2 + qr\sigma_a^2}{qr} I_{p-1} \right| \right\} (q-1) q^{p-1}$$

$$= (\sigma_\varepsilon^2 + r\sigma_b^2)^{p(q-1)-1} (\sigma_\varepsilon^2 + r\sigma_b^2 + qr\sigma_a^2)^{p-1} (\sigma_\varepsilon^2 + r\sigma_b^2 + (q-1)r\sigma_a^2) / r^{pq-1}$$

Substituting:

$$|V_3^\dagger| = (\sigma_\varepsilon^2)^{pq(r-1)-m} \left\{ \frac{(\sigma_\varepsilon^2 + r\sigma_b^2)^{p(q-1)} (\sigma_\varepsilon^2 + r\sigma_b^2 + qr\sigma_a^2)^p}{pq} + \frac{m\sigma_\varepsilon^2}{r(r-m)} |V_2^\dagger| \right\} (r-m) r^{pq-1}$$

$$= (\sigma_\varepsilon^2)^{pq(r-1)-m} (\sigma_\varepsilon^2 + r\sigma_b^2)^{p(q-1)-1} (\sigma_\varepsilon^2 + r\sigma_b^2 + qr\sigma_a^2)^{p-1}$$

$$\{(\sigma_\varepsilon^2 + (r-m)\sigma_b^2)(\sigma_\varepsilon^2 + r\sigma_b^2 + qr\sigma_a^2) - m\sigma_a^2\sigma_\varepsilon^2\} \qquad (3.4.3:2)$$

Finally as X is a vector of ones, $X' V_3^{\dagger\,-1} X$ is a 1×1 matrix so $(X' V_3^{\dagger\,-1} X)^{-1}$ and $|X' V_3^{\dagger\,-1} X|$ are trivial.

3.4.4 Slightly unbalanced two random effects crossed Models with one odd cell

This is an extension to model M3d - the completely balanced two random effects crossed model.

$$y_{ijk} = \alpha_i + \beta_j + \varepsilon_{ijk} \quad i = 1 \ldots p \quad j = 1 \ldots q \quad k = 1 \ldots n_{ij} \quad n_{ij} = \begin{cases} r-m & i,j = 1 \\ r & i,j \neq 1 \end{cases}$$

$$V_3^\dagger = \sigma_\varepsilon^2 I_{pqr-m} + H V_2 H'$$

where $H = D[1_{r-m}, 1_r, 1_r]$, and $V_2 = \sigma_a^2 I_p \otimes J_q + \sigma_b^2 J_p \otimes I_q$.

then from equation 3.4.1:6

$$S^{-1} = \frac{1}{r}I_{pq} - \frac{\sigma_\varepsilon^2}{r}(rV_2 + \sigma_\varepsilon^2 I_{pq})^{-1} -$$

$$= \frac{1}{r}I_{pq} - \frac{\sigma_\varepsilon^2}{r}(\sigma_\varepsilon^2 I_{pq} + r\sigma_a^2 I_p \otimes J_q + r\sigma_b^2 J_p \otimes I_q)^{-1}$$

The techniques for a balanced matrix give:

$$(\sigma_\varepsilon^2 I_{pq} + r\sigma_a^2 I_p \otimes J_q + r\sigma_b^2 J_p \otimes I_q)^{-1}$$

$$= \frac{1}{\sigma_\varepsilon^2} I_p \otimes I_q - \frac{r\sigma_a^2}{\sigma_\varepsilon^2(\sigma_\varepsilon^2 + qr\sigma_a^2)} I_p \otimes J_q - \frac{r\sigma_b^2}{\sigma_\varepsilon^2(\sigma_\varepsilon^2 + pr\sigma_b^2)} J_p \otimes I_q$$

$$+ \frac{r\sigma_a^2 r\sigma_b^2 \{(\sigma_\varepsilon^2 + qr\sigma_a^2) + (\sigma_\varepsilon^2 + pr\sigma_b^2)\}}{\sigma_\varepsilon^2(\sigma_\varepsilon^2 + qr\sigma_a^2)(\sigma_\varepsilon^2 + pr\sigma_b^2)(\sigma_\varepsilon^2 + qr\sigma_a^2 + pr\sigma_b^2)} J_p \otimes J_q$$

Hence

$$S^{-1} = \frac{\sigma_a^2}{\sigma_\varepsilon^2(\sigma_\varepsilon^2 + qr\sigma_a^2)} I_p \otimes J_q + \frac{\sigma_b^2}{\sigma_\varepsilon^2(\sigma_\varepsilon^2 + pr\sigma_b^2)} J_p \otimes I_q$$

$$- \frac{r\sigma_a^2 \sigma_b^2 \{(\sigma_\varepsilon^2 + qr\sigma_a^2) + (\sigma_\varepsilon^2 + pr\sigma_b^2)\}}{\sigma_\varepsilon^2(\sigma_\varepsilon^2 + qr\sigma_a^2)(\sigma_\varepsilon^2 + pr\sigma_b^2)(\sigma_\varepsilon^2 + qr\sigma_a^2 + pr\sigma_b^2)} J_p \otimes J_q$$

Substituting this form of S^{-1} into 3.4.1:4 gives $V_3^{\dagger -1}$.

Determinant of $V_3^\dagger$

The determinant can be evaluated directly using the second technique from section 3.4.1. First use equation 3.4.1:7

$$|V_3^\dagger| = (\sigma_\varepsilon^2)^{pq(r-1)-m} \left| V_2 + \sigma_\varepsilon^2 diag(\frac{1}{r-m}, \frac{1}{r}, \frac{1}{r}, \ldots, \frac{1}{r}) \right| (r-m) r^{pq-1}$$

Thus the determinant of $\sigma_a^2 I_p \otimes J_q + \sigma_b^2 J_p \otimes I_q + \sigma_\varepsilon^2 diag(\frac{1}{r-m}, \frac{1}{r}, \frac{1}{r}, \ldots, \frac{1}{r})$ is needed. Pretending the r-m was an r, there would then be $(p-1)(q-1)$ eigenvalues of σ_ε^2/r, $p-1$ eigenvalues of $q\sigma_a^2 + \sigma_\varepsilon^2/r$, $q-1$ eigenvalues of $p\sigma_b^2 + \sigma_\varepsilon^2/r$ and one eigenvalue of $q\sigma_a^2 + p\sigma_b^2 + \sigma_\varepsilon^2/r$. Since there is an r-m and not an r it follows that there must be $(p-1)(q-1)-1$ eigenvalues of σ_ε^2/r, $p-2$ eigenvalues of $q\sigma_a^2 + \sigma_\varepsilon^2/r$ and $q-2$ eigenvalues of $p\sigma_b^2 + \sigma_\varepsilon^2/r$. The missing 4 eigenvalues must have eigenvectors of the form:

$$C_p \otimes C_q + x C_p \otimes 1_q + y 1_p \otimes C_q + z 1_p \otimes 1_q$$

where C_p denotes a vector of p elements, of which the first element is a one, and the remaining $p-1$ elements are zero. Multiplying out yields four different equations.

$$\sigma_a^2(1+qx+y+qz) + \sigma_b^2(1+x+py+pz) + \frac{\sigma_\epsilon^2}{r-m}(1+x+y+z) = \lambda(1+x+y+z)$$

$$\sigma_a^2(1+qx+y+qz) + \sigma_b^2(x+pz) + \frac{\sigma_\epsilon^2}{r}(x+z) = \lambda(x+z)$$

$$\sigma_a^2(y+qz) + \sigma_b^2(1+x+py+pz) + \frac{\sigma_\epsilon^2}{r}(y+z) = \lambda(y+z)$$

$$\sigma_a^2(y+qz) + \sigma_b^2(x+pz) + \frac{\sigma_\epsilon^2}{r}z = \lambda z$$

Subtracting the second and fourth equations, and the third and fourth yields:

$$x = \sigma_a^2/L_1 \quad \text{where } L_1 = \lambda - q\sigma_a^2 - \sigma_\epsilon^2/r$$

$$y = \sigma_b^2/L_2 \quad \text{where } L_2 = \lambda - p\sigma_b^2 - \sigma_\epsilon^2/r$$

Then the fourth gives:

$$z = (\sigma_a^2 y + \sigma_b^2 x)/L_3 \quad \text{where } L_3 = \lambda - q\sigma_a^2 - p\sigma_b^2 - \sigma_\epsilon^2/r$$

Finally (first plus fourth) minus (second plus third) gives:

$$\lambda = \frac{\sigma_\epsilon^2}{r-m}(1+x+y+z) - \frac{\sigma_\epsilon^2}{r}(x+y+z) = \frac{\sigma_\epsilon^2}{r-m} + \frac{m\sigma_\epsilon^2}{r(r-m)}(x+y+z)$$

substituting for x, y and z and rearranging:

$$(\lambda - \frac{\sigma_\epsilon^2}{r-m})L_1L_2L_3 = \frac{m\sigma_\epsilon^2}{r(r-m)}\{\sigma_b^2L_1(L_3+\sigma_a^2) + \sigma_a^2L_2(L_3+\sigma_b^2)\}$$

This is a quartic in λ, but only the constant term is needed, as only the product of the λ's is required. Thus the awkward determinant is:

$$\left| V_z + \sigma_\epsilon^2 \, diag(\frac{1}{r-m}, \frac{1}{r}, \frac{1}{r}, \ldots, \frac{1}{r}) \right| =$$

$$(\frac{\sigma_\epsilon^2}{r})^{(p-1)(q-1)-1}(q\sigma_a^2 + \frac{\sigma_\epsilon^2}{r})^{p-2}(p\sigma_b^2 + \frac{\sigma_\epsilon^2}{r})^{q-2}$$

$$\{\frac{\sigma_\epsilon^2}{r-m}(q\sigma_a^2 + \frac{\sigma_\epsilon^2}{r})(p\sigma_b^2 + \frac{\sigma_\epsilon^2}{r})(q\sigma_a^2 + p\sigma_b^2 + \frac{\sigma_\epsilon^2}{r})$$

$$- \frac{m\sigma_\epsilon^2}{r(r-m)}(\sigma_b^2(q\sigma_a^2 + \frac{\sigma_\epsilon^2}{r})((q-1)\sigma_a^2 + p\sigma_b^2 + \frac{\sigma_\epsilon^2}{r}) + \sigma_a^2(p\sigma_b^2 + \frac{\sigma_\epsilon^2}{r})(q\sigma_a^2 + (p-1)\sigma_b^2 + \frac{\sigma_\epsilon^2}{r})\}$$

3.5 Grossly Unbalanced Factorial Nested Models

In this section nested models that are grossly unbalanced are examined using the same techniques as for slightly unbalanced models. Though superficially complex, these designs may be practically sensible as they can allow more information to be gained about the higher levels of the design. By contrast a balanced nested model usually supplies detailed information about the residual variance, and the low level variances, and only scant information on the higher levels. Anderson (1973) describes the benefits of unbalanced nested designs, or staggered designs as they are also known.

A general method for handling unbalanced nested models is developed in Section 3.5.1. Both the inverse and determinant of the dispersion matrix from a s-1 level model are derived in terms of the dispersion matrix from a s-2 level model. Thus the inverse and determinant can be calculated recursively. This section parallels the work of La Motte (1972), who used a similar recursive technique to tackle any unbalanced nested model. The results are equivalent. It is also observed that this technique enables unequal residual variances to be used without difficulty for all factorial nested models.

These results are illustrated by Section 3.5.2 which details the 1 way random effects model. A special case of this is the slightly unbalanced one way random effects model, and it is shown that the results agree with those in Section 3.4.2. Section 3.5.3 which details the 2 level nested random effects model. A special case of this is the slightly unbalanced 2 level nested random effects model of section 3.4.3.

3.5.1 Grossly unbalanced s-1 level nested random effects model

The algebra in this section is a generalisation of the algebra for the balanced nested random effects models such as models M2a, M3a, M3c. Analytic expressions are derived for the inverse and determinant of the dispersion matrix V_s from any nested random effects model, allowing arbitrary replication at any level. Define V_s recursively.

$$V_{i+1} = \sigma^2_{i+1} I + H_i V_i H_i' \quad \text{where } V_1 = \sigma^2_1 I$$

then $V = V_s = \sigma_s^2 I + H_{s-1} V_{s-1} H'_{s-1}$.

Note that the stretching matrix H_{s-1} contains all the unbalance at the bottom level of the model by containing the number of replications within each of the observational cells. Similarly the matrix H_{s-2} contains all the information about the number of cells at the next to bottom level of the model.

For later convenience define $y_s = y$ and $T_s = I$, then use equation 3.1.3:2 on V_s

$$V_s^{-1} = (\sigma_s^2 T_s + H_{s-1} V_{s-1} H'_{s-1})^{-1} \qquad (3.5.1:1)$$

$$= \frac{1}{\sigma_s^2} T_s^{-1} - \frac{1}{\sigma_s^2} T_s^{-1} H_{s-1} (H'_{s-1} \frac{1}{\sigma_s^2} T_s^{-1} H_{s-1})^{-1} H'_{s-1} \frac{1}{\sigma_s^2} T_s^{-1}$$

$$+ \frac{1}{\sigma_s^2} T_s^{-1} H_{s-1} (H'_{s-1} \frac{1}{\sigma_s^2} T_s^{-1} H_{s-1})^{-1}$$

$$((H'_{s-1} \frac{1}{\sigma_s^2} T_s^{-1} H_{s-1})^{-1} + V_{s-1})^{-1}$$

$$(H'_{s-1} \frac{1}{\sigma_s^2} T_s^{-1} H_{s-1})^{-1} H'_{s-1} \frac{1}{\sigma_s^2} T_s^{-1}$$

Now use the recursive definition for V_{s-1} on a term from above:

$$(H'_{s-1} \frac{1}{\sigma_s^2} T_s^{-1} H_{s-1})^{-1} + V_{s-1} = (H'_{s-1} \frac{1}{\sigma_s^2} T_s^{-1} H_{s-1})^{-1} + \sigma_{s-1}^2 I + H_{s-2} V_{s-2} H'_{s-2}$$

Define $T_{s-1} = \frac{1}{\sigma_{s-1}^2} (H'_{s-1} \frac{1}{\sigma_s^2} T_s^{-1} H_{s-1})^{-1} + I$ (3.5.1:2)

thus $(H'_{s-1} \frac{1}{\sigma_s^2} T_s^{-1} H_{s-1})^{-1} + V_{s-1} = \sigma_{s-1}^2 T_{s-1} + H_{s-2} V_{s-2} H'_{s-2}$ (3.5.1:3)

Note that since T_s is diagonal, so is T_{s-1}, and hence recursively all the T matrices will be diagonal. Substituting for equation 3.5.1:3 gives:

$$V_s^{-1} = \frac{1}{\sigma_s^2} T_s^{-1} - \frac{1}{\sigma_s^2} T_s^{-1} H_{s-1} (H'_{s-1} T_s^{-1} H_{s-1})^{-1} H'_{s-1} T_s^{-1}$$

$$+ T_s^{-1} H_{s-1} (H'_{s-1} T_s^{-1} H_{s-1})^{-1}$$

$$(\sigma_{s-1}^2 T_{s-1} + H_{s-2} V_{s-2} H'_{s-2})^{-1}$$

$$(H'_{s-1} T_s^{-1} H_{s-1})^{-1} H'_{s-1} T_s^{-1} \qquad (3.5.1:4)$$

Recall that T_s is diagonal and so also is $H'_{s-1} T_s H_{s-1}$. Consequently the inverses of these two matrices are trivial to calculate. Thus the only non-trivial inverse in equation 3.5.1:4 is the inverse of $\sigma_{s-1}^2 T_{s-1} + H_{s-2} V_{s-2} H'_{s-2}$ which is of exactly the

same form as equation 3.5.1:1. Thus the process can be applied recursively.

It is required to calculate $y'V^{-1}y = y_s'V_s^{-1}y_s$ efficiently

let $y_{s-1} = (H_{s-1}'T_s^{-1}H_{s-1})^{-1}H_{s-1}'T_s^{-1}y_s$

then $y_s'V^{-1}y_s = \frac{1}{\sigma_s^2}y_s'T_s^{-1}y_s$

$$- \frac{1}{\sigma_s^2}y_s'T_s^{-1}H_{s-1}(H_{s-1}'T_s^{-1}H_{s-1})^{-1}H_{s-1}'T_s^{-1}y_s$$

$$+ y_{s-1}V_{s-1}^{-1}y_{s-1}$$

This clearly recurses. All the T matrices are diagonal. Note also that if the original observations had unequal residual variances, then this would make T_s not equal to I, but would not complicate the algebra in any way. Thus unequal residual variances can be handled without difficulty.

Calculation of $|V_s|$

The calculation of the determinant of V_s proceeds in a similar recursive fashion, along the lines of the recursive method in section 3.4.1.

$$|\sigma_s^{-2}V_s| = |I + H_{s-1}\frac{1}{\sigma_s^2}V_{s-1}H_{s-1}'|$$

This clearly has many eigenvalues of one, the others all corresponding to eigenvectors of the form $v_s = H_{s-1}v_{s-1}$. Suppose the eigenvalue associated with v_s is λ then:

$$(I + H_{s-1}\frac{1}{\sigma_s^2}V_{s-1}H_{s-1}')v_s = \lambda v_s$$

$$(I + H_{s-1}\frac{1}{\sigma_s^2}V_{s-1}H_{s-1}')H_{s-1}v_{s-1} = \lambda H_{s-1}v_{s-1}$$

$$H_{s-1}(I + \frac{1}{\sigma_s^2}V_{s-1}H_{s-1}'H_{s-1})v_{s-1} = H_{s-1}\lambda v_{s-1}$$

Thus the non-unity eigenvalues of $I+H_{s-1}\frac{1}{\sigma_s^2}V_{s-1}H_{s-1}'$ are the same as the eigenvalues of $I+\frac{1}{\sigma_s^2}V_{s-1}H_{s-1}'H_{s-1}$.

hence $|\sigma_s^{-2}V_s| = |I + \frac{1}{\sigma_s^2}V_{s-1}H_{s-1}'H_{s-1}|$ (3.5.1:5)

Note that the determinant on the right hand side is a matrix with one fewer level than V_s. Now repeat the procedure. First replace V_{s-1} by $\sigma_{s-1}^2 I + H_{s-2}V_{s-2}H_{s-2}'$

$$|\sigma_s^{-2}V_s| = |I + \frac{\sigma_{s-1}^2}{\sigma_s^2}H'_{s-1}H_{s-1} + \frac{1}{\sigma_s^2}H_{s-2}V_{s-2}H'_{s-2}H'_{s-1}H_{s-1}|$$

$$|\sigma_s^{-2}V_s| = |I + \frac{1}{\sigma_{s-1}^2}H_{s-2}V_{s-2}H'_{s-2}\tilde{T}_{s-1}^{-1}|\,|T_{s-1}\frac{\sigma_{s-1}^2}{\sigma_s^2}H'_{s-1}H_{s-1}| \qquad (3.5.1:6)$$

where $T_{s-1} = I + \frac{\sigma_s^2}{\sigma_{s-1}^2}(H'_{s-1}H_{s-1})^{-1}$ as defined in equation 3.5.1:3.

The T matrix is diagonal, and so is the matrix in the second determinant of equation 3.5.1:6 - hence it is trivial to evaluate. The first determinant can be treated in the same manner as the original V_s determinant. This procedure can be repeated recursively. Thus:

$$|\sigma_s^{-2}V_s| = |T_{s-1}\frac{\sigma_{s-1}^2}{\sigma_s^2}H'_{s-1}H_{s-1}||I + \frac{1}{\sigma_{s-1}^2}V_{s-2}H'_{s-2}T_{s-1}^{-1}H_{s-2}| \qquad (3.5.1:7)$$

$$= |T_{s-1}\frac{\sigma_{s-1}^2}{\sigma_s^2}H'_{s-1}H_{s-1}||I + \frac{\sigma_{s-2}^2}{\sigma_{s-1}^2}H'_{s-2}T_{s-1}^{-1}H_{s-2} + \frac{1}{\sigma_{s-1}^2}H_{s-3}V_{s-3}H'_{s-3}H'_{s-2}T_{s-1}^{-1}H_{s-2}|$$

$$= |T_{s-1}\frac{\sigma_{s-1}^2}{\sigma_s^2}H'_{s-1}H_{s-1}||I + \frac{1}{\sigma_{s-1}^2}H'_{s-3}V_{s-3}H'_{s-3}T_{s-2}^{-1}||T_{s-2}\frac{\sigma_{s-2}^2}{\sigma_{s-1}^2}H'_{s-2}T_{s-1}^{-1}H_{s-2}|$$

$$= |T_{s-1}\frac{\sigma_{s-1}^2}{\sigma_s^2}H'_{s-1}H_{s-1}||T_{s-2}\frac{\sigma_{s-2}^2}{\sigma_{s-1}^2}H'_{s-2}T_{s-1}^{-1}H_{s-2}||I + \frac{1}{\sigma_{s-1}^2}H'_{s-3}V_{s-3}H'_{s-3}T_{s-2}^{-1}|$$

with $T_{s-2} = I + \frac{\sigma_{s-1}^2}{\sigma_{s-2}^2}(H'_{s-2}T_{s-1}^{-1}H_{s-2})^{-1}$

3.5.2 Grossly unbalanced 1 way random effects model

The inverse and determinant of any arbitrarily unbalanced nested model have been constructed recursively in section 3.5.1. By way of illustration, consider the unbalanced equivalent of model M2a - the one way random effects model:

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \qquad \alpha_i \sim N(0,\sigma_a^2) \quad i = 1,2,\ldots,p$$
$$\varepsilon_{ij} \sim N(0,\sigma_\varepsilon^2) \quad j = 1,2,\ldots,q_i \quad n = \sum_i q_i$$

Alternatively $y \sim N(\mu,V_2)$ where the dispersion matrix for this model is denoted by V_2, the subscript 2 serving as a reminder that there are two variance components. Then

$$V_2 = \sigma_\varepsilon^2 I_n + \sigma_a^2 D[1_{q_1}, 1_{q_2}, \ldots, 1_{q_p}] D[1_{q_1}, 1_{q_2}, \ldots, 1_{q_p}]'$$

$$V_2 = \sigma_2^2 I_n + H_1 V_1 H_1'$$

where $\sigma_2^2 = \sigma_\varepsilon^2$, $H_1 = D[1_{q_1}, 1_{q_2}, \ldots, 1_{q_p}]$, $V_1 = \sigma_1^2 I_p$ and $\sigma_1^2 = \sigma_a^2$ (3.5.2:1)

Note that the stretching matrix H_1 contains all of the information about the unbalance in the design. Also note that $H_1'H_1$ is $diag(q_1, \ldots, q_p)$. Now let $T_2 = I$:

$$\sigma_1^2 T_1 = (H_1' \frac{1}{\sigma_2^2} T_2^{-1} H_1)^{-1} + \sigma_1^2 I$$

$$= \sigma_\varepsilon^2 diag(\frac{1}{q_1}, \ldots, \frac{1}{q_p}) + \sigma_a^2 I$$

$$T_1 = \frac{1}{\sigma_a^2} diag(\frac{\sigma_\varepsilon^2 + q_1\sigma_a^2}{q_1}, \ldots, \frac{\sigma_\varepsilon^2 + q_p\sigma_a^2}{q_p})$$

$$T_1^{-1} = \sigma_a^2 diag(\frac{q_1}{\sigma_\varepsilon^2 + q_1\sigma_a^2}, \ldots, \frac{q_p}{\sigma_\varepsilon^2 + q_p\sigma_a^2})$$

then substituting terms into equation 3.5.1:4 gives:

$$V_2^{-1} = \frac{1}{\sigma_\varepsilon^2} I - \frac{1}{\sigma_\varepsilon^2} H_1 (H_1'H_1)^{-1} H_1' + H_1 (H_1'H_1)^{-1} (\sigma_1^2 T_1)^{-1} (H_1'H_1)^{-1} H_1'$$

$$= \frac{1}{\sigma_\varepsilon^2} \{I - D[\frac{1}{q_1} J_{q_1}, \ldots, \frac{1}{q_p} J_{q_p}] + \sigma_\varepsilon^2 D[\frac{1}{q_1} \frac{1}{\sigma_\varepsilon^2 + q_1\sigma_a^2} J_{q_1}, \ldots, \frac{1}{q_p} \frac{1}{\sigma_\varepsilon^2 + q_p\sigma_a^2} J_{q_p}]\}$$

$$= \frac{1}{\sigma_\varepsilon^2} \{I - D[\frac{\sigma_a^2}{\sigma_\varepsilon^2 + q_1\sigma_a^2} J_{q_1}, \ldots, \frac{\sigma_a^2}{\sigma_\varepsilon^2 + q_p\sigma_a^2} J_{q_p}]\} \quad (3.5.2:2)$$

Thus V_2^{-1} has been obtained explicitly. The determinant of V_2 is equally straight forward using equation 3.5.1:5

$$|\sigma_2^{-2} V_2| = |I + \frac{1}{\sigma_2^2} V_1 H_1' H_1| = |I + \frac{\sigma_a^2}{\sigma_\varepsilon^2} diag(q_1, \ldots, q_p)| = \prod_{i=1}^{p} (1 + q_i \frac{\sigma_a^2}{\sigma_\varepsilon^2})$$

$$|V_2| = (\sigma_\varepsilon^2)^{n-p} \prod_{i=1}^{p} (\sigma_\varepsilon^2 + q_i\sigma_a^2)$$

Special Case - the slightly unbalanced 1 way random effects model

Note that if $q_1 = q-m$ and $q_i = q$ for $i = 2,\ldots,p$, then V_2 is the same as for the slightly unbalanced model from Section 3.4.2. Substituting for $q_1,\ldots,q_p$ in equation 3.5.2:2 gives:

$$V_2^{-1} = \frac{1}{\sigma_\varepsilon^2}(I - D[\frac{\sigma_a^2}{\sigma_\varepsilon^2+(q-m)\sigma_a^2}J_{q-m}, \frac{\sigma_a^2}{\sigma_\varepsilon^2+q\sigma_a^2}J_q, \ldots, \frac{\sigma_a^2}{\sigma_\varepsilon^2+q\sigma_a^2}J_q])$$

$$|V_2| = (\sigma_\varepsilon^2)^{p(q-1)-m}(\sigma_\varepsilon^2 + q\sigma_a^2)^{p-1}(\sigma_\varepsilon^2 + (q-m)\sigma_a^2)$$

The inverse and determinant are of course the same as those obtained in Section 3.4.2, but this section is far more flexible and allows far more unbalance in the design.

3.5.3 Grossly unbalanced 2 level nested random effects model

As a second illustration of the technique, consider the unbalanced equivalent of model M3a - the two level nested random effects model. Allow an arbitrary number of observations ($n_{ij} > 0$) within each cell, and an arbitrary number of subgroups within each group.

$$y_{ijk} = \mu + \alpha_i + \beta_{ij} + \varepsilon_{ijk} \qquad \begin{aligned} \alpha_i &\sim N(0,\sigma_a^2) \quad i = 1,2,\ldots,p \\ \beta_{ij} &\sim N(0,\sigma_b^2) \quad j = 1,2,\ldots,q_i \\ \varepsilon_{ijk} &\sim N(0,\sigma_\varepsilon^2) \quad k = 1,2,\ldots,n_{ij} \quad n = \Sigma n_{ij} \end{aligned}$$

then $y \sim N(\mu, V_3)$ where

$$V_3 = \sigma_\varepsilon^2 I_n + \sigma_b^2 D[1_{n_{11}},1_{n_{12}},\ldots,1_{n_{pq_p}}]D[1_{n_{11}},1_{n_{12}},\ldots,1_{n_{pq_p}}]'$$

$$+ \sigma_a^2 D[1_{n_{1.}},1_{n_{2.}},\ldots,1_{n_{p.}}]D[1_{n_{1.}},1_{n_{2.}},\ldots,1_{n_{p.}}]'$$

$$V_3 = \sigma_3^2 I_n + H_2 V_2 H_2'$$

where $\sigma_3^2 = \sigma_\varepsilon^2$, $H_2 = D[1_{n_{11}},1_{n_{12}},\ldots,1_{n_{pq_p}}]$,

and $V_2 = \sigma_b^2 I_{q.} + \sigma_a^2 D[1_{q_1},\ldots,1_{q_p}]$

$= \sigma_2^2 I_{q.} + H_1 V_1 H_1'$

where $V_1 = \sigma_1^2 I_p$, $H_1 = D[1_{q_1},\ldots,1_{q_p}]$ and $\sigma_1^2 = \sigma_a^2$

Note that $H_2'H_2$ is $diag(n_{11}, n_{12},\ldots, n_{pq_p})$. Let $T_3 = I$, then from equation 3.5.1:2

$$\sigma_2^2 T_2 = (H_2' \frac{1}{\sigma_3^2} T_3 H_2)^{-1} + \sigma_2^2 I$$

$$= \sigma_\epsilon^2 diag(\frac{1}{n_{11}},\ldots,\frac{1}{n_{pq_p}}) + \sigma_b^2 I$$

$$= diag(\frac{\sigma_\epsilon^2 + n_{11}\sigma_b^2}{n_{11}},\ldots,\frac{\sigma_\epsilon^2 + n_{pq_p}\sigma_b^2}{n_{pq_p}})$$

Then from equation 3.5.1:4

$$V_3^{-1} = \frac{1}{\sigma_\epsilon^2} I - \frac{1}{\sigma_\epsilon^2} H_2 (H_2'H_2)^{-1} H_2' + H_2 (H_2'H_2)^{-1} (\sigma_2^2 T_2 + H_1 V_1 H_1')^{-1} (H_2'H_2)^{-1} H_2' \qquad (3.5.3{:}1)$$

Now use the same argument again, or alternatively use equation 3.1.3:1

$$(\sigma_2^2 T_2 + H_1 V_1 H_1')^{-1} = \frac{1}{\sigma_2^2} T_2^{-1} - \frac{1}{\sigma_2^2} T_2^{-1} H_1 (H_1' \frac{1}{\sigma_2^2} T_2^{-1} H_1 + V_1^{-1})^{-1} T_2^{-1} \frac{1}{\sigma_2^2}$$

hence substituting this into equation 3.5.3:1

$$V_3^{-1} = \frac{1}{\sigma_\epsilon^2} I - \frac{1}{\sigma_\epsilon^2} H_2 (H_2'H_2)^{-1} H_2'$$

$$+ H_2 (H_2'H_2)^{-1} \frac{1}{\sigma_2^2} T_2^{-1} (H_2'H_2)^{-1} H_2'$$

$$- H_2 (H_2'H_2)^{-1} \frac{1}{\sigma_2^2} T_2^{-1} H_1 (H_1' \frac{1}{\sigma_2^2} T_2^{-1} H_1 + V_1^{-1})^{-1} \frac{1}{\sigma_2^2} T_2^{-1} (H_2'H_2)^{-1} H_2'$$

The only inverse which is not already known to be diagonal (and hence trivial) is $(H_1' \frac{1}{\sigma_2^2} T_2^{-1} H_1 + V_1^{-1})^{-1}$ which can be inverted as follows:

$$H_1' \frac{1}{\sigma_2^2} T_2^{-1} H_1 + V_1^{-1} = diag(\sum_j \frac{n_{1j}}{\sigma_\epsilon^2 + n_{1j}\sigma_b^2},\ldots,\sum_j \frac{n_{pj}}{\sigma_\epsilon^2 + n_{pj}\sigma_b^2}) + \frac{1}{\sigma_1^2} I_p$$

$$(H_1' \frac{1}{\sigma_2^2} T_2^{-1} H_1 + V_1^{-1})^{-1} = \sigma_a^2 diag(1 + \sigma_a^2 \sum_j \frac{n_{ij}}{\sigma_\epsilon^2 + n_{1j}\sigma_b^2},\ldots, 1 + \sigma_a^2 \sum_j \frac{n_{pj}}{\sigma_\epsilon^2 + n_{pj}\sigma_b^2})^{-1}$$

Substituting for this gives:

$$V_3^{-1} = \frac{1}{\sigma_\varepsilon^2} I - \frac{1}{\sigma_\varepsilon^2} H_2 diag(\frac{1}{n_{11}}, \ldots, \frac{1}{n_{pq_p}}) H_2'$$

$$+ \frac{1}{\sigma_\varepsilon^2} H_2 diag(\frac{1}{n_{11}} \frac{1}{\sigma_\varepsilon^2 + n_{11}\sigma_b^2}, \ldots, \frac{1}{n_{pq_p}} \frac{1}{\sigma_\varepsilon^2 + n_{pq_p}\sigma_b^2}) H_2'$$

$$- H_2 diag(\frac{1}{\sigma_\varepsilon^2 + n_{11}\sigma_b^2}, \ldots, \frac{1}{\sigma_\varepsilon^2 + n_{pq_p}\sigma_b^2})$$

$$H_1 \sigma_a^2 diag(1 + \sigma_a^2 \sum_j \frac{n_{1j}}{\sigma_\varepsilon^2 + n_{1j}\sigma_b^2}, \ldots, 1 + \sigma_a^2 \sum_j \frac{n_{pj}}{\sigma_\varepsilon^2 + n_{pj}\sigma_b^2})^{-1} H_1'$$

$$diag(\frac{1}{\sigma_\varepsilon^2 + n_{11}\sigma_b^2}, \ldots, \frac{1}{\sigma_\varepsilon^2 + n_{pq_p}\sigma_b^2}) H_2'$$

Simplifying this yields

$$V_3^{-1} = \frac{1}{\sigma_\varepsilon^2} I - \frac{\sigma_b^2}{\sigma_\varepsilon^2} D[\frac{1}{\sigma_\varepsilon^2 + n_{11}\sigma_b^2} J_{n_{11}}, \ldots, \frac{1}{\sigma_\varepsilon^2 + n_{pq_p}\sigma_b^2} J_{n_{pq_p}}]$$

$$- D[\frac{1}{\sigma_\varepsilon^2 + n_{11}\sigma_b^2} 1_{n_{11}}, \ldots, \frac{1}{\sigma_\varepsilon^2 + n_{pq_p}\sigma_b^2} 1_{n_{pq_p}}]$$

$$\sigma_a^2 D[(1 + \sigma_a^2 \sum_j \frac{n_{1j}}{\sigma_\varepsilon^2 + n_{1j}\sigma_b^2})^{-1} J_{q_1}, \ldots, (1 + \sigma_a^2 \sum_j \frac{n_{pj}}{\sigma_\varepsilon^2 + n_{pj}\sigma_b^2})^{-1} J_{q_p}]$$

$$D[\frac{1}{\sigma_\varepsilon^2 + n_{11}\sigma_b^2} J_{n_{11}}, \ldots, \frac{1}{\sigma_\varepsilon^2 + n_{pq_p}\sigma_b^2} J_{n_{pq_p}}]' \qquad (3.5.3:2)$$

The determinant of V_3 is obtained from a direct application of equation 3.5.1:7

$$|\sigma_3^{-2} V_3| = |T_2 \frac{\sigma_2^2}{\sigma_3^2} H_2' H_2| |I + \frac{1}{\sigma_2^2} V_1 H_1' T_2^{-1} H_1|$$

where $T_2 = \frac{1}{\sigma_b^2} diag(\frac{\sigma_\varepsilon^2 + n_{11}\sigma_b^2}{n_{11}}, \ldots, \frac{\sigma_\varepsilon^2 + n_{pq_p}\sigma_b^2}{n_{pq_p}})$

hence $T_2 \frac{\sigma_2^2}{\sigma_3^2} H_2' H_2 = \frac{1}{\sigma_\varepsilon^2} diag(\sigma_\varepsilon^2 + n_{11}\sigma_b^2, \ldots, \sigma_\varepsilon^2 + n_{pq_p}\sigma_b^2)$

$$|T_2 \frac{\sigma_2^2}{\sigma_3^2} H_2' H_2| = (\sigma_\varepsilon^2)^{-q} \cdot \prod_{i,j=1}^{p,q_i} (\sigma_\varepsilon^2 + n_{ij}\sigma_b^2)$$

$$I + \frac{1}{\sigma_2^2} V_1 H_1' T_2^{-1} H_1 = diag(1 + \sum_j \frac{n_{1j}\sigma_a^2}{\sigma_\epsilon^2 + n_{1j}\sigma_b^2}, \ldots, 1 + \sum_j \frac{n_{pj}\sigma_a^2}{\sigma_\epsilon^2 + n_{pj}\sigma_b^2})$$

hence

$$|V_3| = (\sigma_3^2)^{n-q} \cdot \prod_{i,j=1}^{p,q_i} (\sigma_3^2 + n_{ij}\sigma_2^2) \prod_{i=1}^{p} (1 + \sum_j \frac{n_{ij}\sigma_1^2}{\sigma_3^2 + n_{ij}\sigma_2^2}) \qquad (3.5.3:3)$$

Special Case - the slightly unbalanced two level nested random effects model

By way of example suppose that $n_{11} = r-m$ and all the other $n_{ij} = r$. Suppose there are always q subgroups within each group. Then V_3 is the $V_3^{\dagger}$ from the slightly unbalanced model in Section 3.4.3 Substituting into equation 3.5.3:2 yields the inverse $V_3^{\dagger -1}$

$$V_3^{\dagger -1} = \frac{1}{\sigma_\epsilon^2} I - \frac{\sigma_b^2}{\sigma_\epsilon^2} D[\frac{1}{\sigma_\epsilon^2 + (r-m)\sigma_b^2} J_{r-m}, \frac{1}{\sigma_\epsilon^2 + r\sigma_b^2} J_r, \ldots, \frac{1}{\sigma_\epsilon^2 + r\sigma_b^2} J_r]$$

$$- D[\frac{1}{\sigma_\epsilon^2 + (r-m)\sigma_b^2} 1_{r-m}, \frac{1}{\sigma_\epsilon^2 + r\sigma_b^2} 1_r, \ldots, \frac{1}{\sigma_\epsilon^2 + r\sigma_b^2} 1_r)$$

$$\sigma_a^2 D[\frac{(\sigma_\epsilon^2 + r\sigma_b^2)(\sigma_\epsilon^2 + (r-m)\sigma_b^2)}{(\sigma_\epsilon^2 + (r-m)\sigma_b^2)(\sigma_\epsilon^2 + r\sigma_b^2 + qr\sigma_a^2) - m\sigma_\epsilon^2\sigma_a^2} J_q, \frac{\sigma_\epsilon^2 + r\sigma_b^2}{\sigma_\epsilon^2 + r\sigma_b^2 + qr\sigma_a^2} J_q \cdots \frac{\sigma_\epsilon^2 + r\sigma_b^2}{\sigma_\epsilon^2 + r\sigma_b^2 + qr\sigma_a^2} J_q)$$

$$D[\frac{1}{\sigma_\epsilon^2 + (r-m)\sigma_b^2} 1_{r-m}, \frac{1}{\sigma_\epsilon^2 + r\sigma_b^2} 1_r, \ldots, \frac{1}{\sigma_\epsilon^2 + r\sigma_b^2} 1_r]$$

$$V_3^{\dagger -1} = \frac{1}{\sigma_\epsilon^2} I - \frac{1}{\sigma_\epsilon^2} A D[J_{r-m}, J_r \ldots J_r] - \frac{\sigma_b^2}{\sigma_\epsilon^2} \frac{m\sigma_b^2}{(\sigma_\epsilon^2 + r\sigma_b^2)(\sigma_\epsilon^2 + (r-m)\sigma_b^2)} D[J_{r-m}, 0_{r,r} \ldots 0_{r,r}]$$

$$- \frac{1}{\sigma_\epsilon^2} D\left[\begin{pmatrix} P J_{r-m} & Q 1_{(q-1)r, r-m} \\ Q 1_{r-m, (q-1)r} & R J_{(q-1)r} \end{pmatrix}, BJ_{qr}, \ldots, BJ_{qr}\right]$$

where $A = \dfrac{\sigma_b^2}{\sigma_\epsilon^2 + r\sigma_b^2}$ and $B = \dfrac{\sigma_a^2\sigma_\epsilon^2}{(\sigma_\epsilon^2 + r\sigma_b^2)(\sigma_\epsilon^2 + r\sigma_b^2 + qr\sigma_a^2)}$

where $P = \dfrac{\sigma_\epsilon^2\sigma_a^2(\sigma_\epsilon^2 + r\sigma_b^2)}{[(\sigma_\epsilon^2 + (r-m)\sigma_b^2)(\sigma_\epsilon^2 + r\sigma_b^2 + qr\sigma_a^2) - m\sigma_\epsilon^2 siga)(\sigma_\epsilon^2 + (r-m)\sigma_b^2)}$

$$Q = \frac{\sigma_\epsilon^2\sigma_a^2}{(\sigma_\epsilon^2 + (r-m)\sigma_b^2)(\sigma_\epsilon^2 + r\sigma_b^2 + qr\sigma_a^2) - m\sigma_\epsilon^2\sigma_a^2}$$

$$\text{and } R = \frac{\sigma_\varepsilon^2\sigma_a^2(\sigma_\varepsilon^2+(r-m)\sigma_b^2)}{\{(\sigma_\varepsilon^2+(r-m)\sigma_b^2)(\sigma_\varepsilon^2+r\sigma_b^2+qr\sigma_a^2)-m\sigma_\varepsilon^2\sigma_a^2)(\sigma_\varepsilon^2+r\sigma_b^2)}$$

Now evaluate $Q-B$ and $A+B$ and compare:

$$(Q-B)/\sigma_a^2\sigma_\varepsilon^2 =$$

$$\frac{m(\sigma_b^2(\sigma_\varepsilon^2+r\sigma_b^2+qr\sigma_a^2)+\sigma_\varepsilon^2\sigma_a^2)}{(\sigma_\varepsilon^2+r\sigma_b^2)(\sigma_\varepsilon^2+r\sigma_b^2+qr\sigma_a^2)\{(\sigma_\varepsilon^2+r\sigma_b^2)(\sigma_\varepsilon^2+r\sigma_b^2+qr\sigma_a^2)-m(\sigma_b^2(\sigma_\varepsilon^2+r\sigma_b^2+qr\sigma_a^2)+\sigma_\varepsilon^2\sigma_a^2)\}}$$

$$A+B = \frac{\sigma_b^2(\sigma_\varepsilon^2+r\sigma_b^2+qr\sigma_a^2)+\sigma_\varepsilon^2\sigma_a^2)}{(\sigma_\varepsilon^2+r\sigma_b^2)(\sigma_\varepsilon^2+r\sigma_b^2+qr\sigma_a^2)}$$

$$\text{thus } Q-B = \frac{m(A+B)B}{1-m(A+B)}$$

Rearranging

$$R-B = \frac{mB^2}{1-m(A+B)} \text{ and } P-B+\frac{m(\sigma_b^2)^2}{(\sigma_\varepsilon^2+r\sigma_b^2)(\sigma_\varepsilon^2+r\sigma_b^2+qr\sigma_a^2)} = \frac{m(A+B)^2}{1-m(A+B)}$$

Substituting for all these leaves:

$$V_3^{\uparrow -1} = \frac{1}{\sigma_\varepsilon^2}I - \frac{1}{\sigma_\varepsilon^2}A\mathrm{D}[J_{r-m},J_r,\ldots,J_r] - \frac{1}{\sigma_\varepsilon^2}B\mathrm{D}[J_{qr-m},J_{qr},\ldots,J_{qr}]$$

$$- \frac{1}{\sigma_\varepsilon^2}\frac{m}{1-m(A+B)}\mathrm{D}\left[\begin{pmatrix}(A+B)^2J_{r-m} & (A+B)B\,1_{(q-1)r,r-m}\\ (A+B)B\,1_{r-m,(q-1)r} & B^2J_{(q-1)r}\end{pmatrix},0_{qr,qr}\cdots 0_{qr,qr}\right]$$

This is of course identical to equation 3.4.3:1

$$|V_3| = (\sigma_\varepsilon^2)^{pqr-m-pq}(\sigma_\varepsilon^2+r\sigma_b^2)^{pq-1}(\sigma_\varepsilon^2+(r-m)\sigma_b^2)$$

$$\left(\frac{\sigma_\varepsilon^2+r\sigma_b^2+qr\sigma_a^2}{\sigma_\varepsilon^2+r\sigma_b^2}\right)^{p-1}\frac{\{(\sigma_\varepsilon^2+(r-m)\sigma_b^2)(\sigma_\varepsilon^2+r\sigma_b^2+qr\sigma_a^2)-m\sigma_a^2\sigma_\varepsilon^2\}}{(\sigma_\varepsilon^2+r\sigma_b^2)(\sigma_\varepsilon^2+(r-m)\sigma_b^2)}$$

$$|V_3| = (\sigma_\varepsilon^2)^{pq(r-1)-m}(\sigma_\varepsilon^2+r\sigma_b^2)^{p(q-1)-1}(\sigma_\varepsilon^2+r\sigma_b^2+qr\sigma_a^2)^{p-1}$$

$$\{(\sigma_\varepsilon^2+(r-m)\sigma_b^2)(\sigma_\varepsilon^2+r\sigma_b^2+qr\sigma_a^2)-m\sigma_a^2\sigma_\varepsilon^2\}$$

This is of course identical to equation 3.4.3:2

3.6 Grossly unbalanced Crossed Designs without Interaction

For these cases an analytic solution to the inverse of the dispersion matrix has not been found. A method is presented which reduces the computational burden substantially by replacing the numerical inversion of V by that of a series of smaller matrices. The technique is however much worse than all the other techniques discussed - see section 3.7 for a discussion. Consider model M3d - a two random effects crossed model without interaction. Thus the observation from a cell depends on a *row effect* and a *column effect*, and an error term, where both the row and column effects are random. Thus:

$$y_{ijk} = \alpha_i + \beta_j + \epsilon_{ij} \qquad \begin{aligned} \alpha_i &\sim N(0,\sigma_a^2) \quad i = 1,2,\ldots,p \\ \beta_j &\sim N(0,\sigma_b^2) \quad j = 1,2,\ldots,q \\ \epsilon_{ij} &\sim N(0,\sigma_\epsilon^2) \quad k = 1,2,\ldots,n_{ij} \end{aligned}$$

Alternatively:

$$y \sim N(X_1 a_1 + X_2 a_2, V_3)$$

where $X_1 = D[1_{n_{1.}}, 1_{n_{2.}}, \ldots, 1_{n_{p.}}]$

$$\text{and } X_2 = \begin{pmatrix} D[1_{n_{11}}, 1_{n_{12}}, \ldots, 1_{n_{1q}}] \\ D[1_{n_{21}}, 1_{n_{22}}, \ldots, 1_{n_{2q}}] \\ \\ D[1_{n_{p1}}, 1_{n_{p2}}, \ldots, 1_{n_{pq}}] \end{pmatrix}$$

and $V_3 = \sigma_\epsilon^2 I_n + \sigma_a^2 X_1 X_1' + \sigma_b^2 X_2 X_2'$

The subscript 3 on the V_3 serving as a reminder that this is a 3 variance component model. Now let V_2 be $\sigma_\epsilon^2 I_n + \sigma_a^2 X_1 X_1'$ and use equation 3.1.3:1 first on V_2 to get V_2^{-1} explicitly, then on V_3:

$$V_2^{-1} = \frac{1}{\sigma_\epsilon^2} I_n - \frac{1}{\sigma_\epsilon^2} X_1 (X_1' X_1 + \frac{\sigma_\epsilon^2}{\sigma_a^2} I_p)^{-1} X_1'$$

$$V_3^{-1} = V_2^{-1} - V_2^{-1} X_2 (X_2' V_2^{-1} X_2 + \frac{1}{\sigma_b^2} I_q)^{-1} X_2' V_2^{-1}$$

Thus the inverse of V_3 can be calculated numerically from the inverse of a q by q matrix. This is a much smaller task than inverting V_3

In general consider an s level crossed linear model with replications $r_1, r_2, \ldots, r_s$, and random effects $a_1, a_2, \ldots, a_s$. This can be written as:

$$y_{i_1 i_2 \ldots i_s j} = (a_1)_{i_1} + (a_2)_{i_2} + (a_s)_{i_s} + \varepsilon_{i_1 i_2 \ldots i_s j}$$

$$i_1 = 1,2,\ldots r_1 \quad i_2 = 1,2,\ldots r_2 \quad i_s = 1,2,\ldots r_s \quad j = 1,2,\ldots n_{i_1,i_2,\ldots,i_s}$$

$$y \sim N(X_1 a_1 + X_2 a_2 + \ldots + X_s a_s, V_s)$$

$$\text{where } V_s = \sigma_\varepsilon^2 I_n + \sigma_1^2 X_1 X_1' + \sigma_2^2 X_2 X_2' + \ldots + \sigma_s^2 X_s X_s'$$

$$= V_{s-1} + \sigma_s^2 X_s X_s'$$

$$\text{where } V_{s-1} = \sigma_\varepsilon^2 I_n + \sigma_1^2 X_1 X_1' + \sigma_2^2 X_2 X_2' + \sigma_s^2 X_{s-1} X_{s-1}'$$

$$\text{then } V_s^{-1} = V_{s-1}^{-1} - V_{s-1}^{-1} X_s (X_s' V_{s-1}^{-1} X_s + \frac{1}{\sigma_r^2} I_{r_s}) X_s' V_{s-1}^{-1}$$

to evaluate this requires a r_s by r_s matrix to be inverted numerically as well as V_{s-1}. Recurse and note that V_2 is analytically invertible. Thus V_s can be inverted in terms of a series of inverses of sizes r_s, r_{s-1}, ..., r_3 and r_2. Note that since there is not a matrix of size r_1 to be inverted, then the data should be arranged so that r_1 is greater than the other r_i.

3.7 Computational Considerations

The purpose of this chapter was to examine ways of efficiently evaluating the likelihood function to enable Bayesian analysis, based on numerical integration, to be feasible. This section provides an order of magnitude guide to the effort required to evaluate the likelihoods for the models discussed.

In Section 3.2, and thereafter, frequent use is made of balanced dispersion matrices. Following the methods of Searle and Henderson, the eigenvalues for a dispersion matrix from a s-1 level model can be obtained in about $s\,2^{s-1}$ multiplications, thus giving $|V_s|$ in $(s+2)2^{s-1}$ multiplications. V_s^{-1} takes $2s\,2^{s-1} + 2^s = (s+1)2^s$ multiplications and $y_s' V_s^{-1} y_s$ takes $(s+2)\,2^s$ multiplications. Both $|V_s|$ and $y_s' V^{-1} y_s$ can be evaluated in $(s+3)\,2^s$ multiplications. Thus a 2 level model (nested or crossed) takes about 48 multiplications, and a 3 level takes 112 multiplications, regardless of the number of observations. Since nested models have many zero coefficients in

equation 3.2:1, an s-1 level nested model only takes as many operations as a general s-2 level model.

Section 3.2.2 for marginal σ^2 uses two balanced matrices $\boldsymbol{V}$ and $\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}$, the latter having fewer levels. Thus the number of multiplications for the former dominates the number of multiplications for the latter. Similarly for Section 3.2.3 for marginal α_1,σ^2. Though the matrix $\boldsymbol{X}_2'\boldsymbol{V}^{-1}\boldsymbol{X}_2$ is unbalanced, the recursive method for manipulating it is efficient, and the number of multiplications handling $\boldsymbol{V}$ should dominate as $\boldsymbol{V}$ has more levels.

A slightly unbalanced model from Section 3.4.1 with s-1 levels requires an s-2 level balanced matrix to be inverted in the calculation of $\boldsymbol{S}^{-1}$. Method 1 for the determinant requires a series of balanced matrices from s-2 levels to 0 levels. Method 2 requires the determinant of an s-2 matrix and a similar amount of work for the *extra* λ values. Method 2 is thus more efficient and the number of multiplications is approximately $(3s+4)2^{s-1}$. This compares with $(s+1)2^{s-1}$ for a balanced nested s-1 level model and $(s+2)2^s$ for any balanced s-1 level model. Thus the slight unbalance increases the computational load for a model by about a half.

The grossly unbalanced nested models in Section 3.5 require the inversion of a set of diagonal matrices (equation 3.5.1:2) and a the determinants of a series of diagonal matrices (equation 3.5.1:7). These are dominated by the number of cells as this is the number of operations for the biggest inverse and determinant.

The unbalanced crossed designs without interactions from Section 3.6 require a series of inverses of sizes $r_2,\ldots,r_s$, and thus the number of operations is proportional to $r_2^3 + \ldots + r_s^3$.

Chapter 4
Non-Informative Prior Distributions

4.1 Introduction

One of the strengths of Bayesian analysis is the way in which it allows the data to modify your beliefs prior to the experiment, through the likelihood, yielding a posterior distribution encapsulating both the prior beliefs and the data. This posterior distribution can subsequently be used as the prior distribution to a later experiment and so on. There remains the question of choosing an initial prior distribution to the first experiment, when the experimenter is in a state of considerable ignorance about the true or likely values of the parameter(s). Hopefully the information supplied by the data in this experiment will vastly outweigh the information supplied by the initial prior distribution, and then the particular initial prior distribution that is chosen will matter little. Nevertheless, it is wise to use a *non-informative prior distribution*, so that the contribution to the posterior distribution made by initial prior distribution is small.

Two alternative philosophies for selecting non-informative prior distributions for linear models are explored. In sections 4.2, 4.3 and 4.4 the ideas of Jeffreys (1961) are presented, and the standard Jeffreys' priors are produced. Section 4.2 outlines the principle of Jeffreys' priors, and section 4.3 deals with Jeffreys' priors for Normal models. Section 4.4 is concerned with Jeffreys' priors for models with general multivariate t errors. The latter priors are shown to be the same as the priors from the equivalent Normal models. In section 4.5 the information theory approach taken by Bernardo (1979) is given. Despite the different theoretical justifications, both philosophies lead to the same prior distributions for the models considered.

4.2 Jeffreys' Priors

In any model, there is always a certain arbitrariness about the choice of parameters. Suppose the observations (y_i) are known to be from a Normal distribution with unknown mean θ and unit variance, thus $y_i \sim N(\theta,1)$. Suppose however that the quantity of interest was not θ itself but some function $\phi(\theta)$, for example $1/\theta$ or $\sqrt{\theta}$. In the model specification, this presents no difficulty, but difficulties arise with the choice of non-informative prior distribution, as a uniform prior on θ will

not be uniform on $\phi(\theta)$.

The information about θ in the data is expressed through the likelihood function. In general, the location, scale and shape of the posterior distribution will depend on the data y. If however for some choice of θ, or a 1-1 mapping of $\theta \longrightarrow \phi(\theta)$, the shape of the posterior distribution remains invariant, then the posterior density is termed *data translated*. The shape of the posterior density for $\phi(\theta)$ is determined *a priori*. The experimenter can thus express prior ignorance about θ by saying that no value of $\phi(\theta)$ is preferable to another. That is by taking a prior that assumes $\phi(\theta)$ to be locally uniform, the resulting posterior distribution is seen to be a function of the data alone.

For a normal distribution with an unknown mean, and a known variance, this corresponds to a uniform prior on the mean. A normal distribution with an known mean, but unknown variance σ^2, gives rise to a non-informative prior of $1/\sigma^2$.

It is not generally possible to find a transformation that produces the data translation property, and hence it is not possible to produce exact non-informative priors. The metric $\phi(\theta)$ for which a locally uniform prior is approximately non-informative is:

$$\frac{d\phi}{d\theta} \propto \{I(\theta)\}^{\frac{1}{2}}$$

where $I(\theta)$ is the Fisher information defined by:

$$I(\theta) = -\underset{y|\theta}{E}\left[\frac{d^2 \log p(y|\theta)}{d^2\theta}\right] = \underset{y|\theta}{E}\left[\frac{d \log p(y|\theta)}{d\theta}\right]^2$$

This result was first given by Jeffreys (1961).

The argument generalises to multi-dimensional problems yielding $p(\theta) \propto |I_n(\theta)|^{\frac{1}{2}}$ for a vector of n parameters. However care should be taken before mechanistically applying this result, especially if there are different types of parameter in θ. Sometimes it is known *a priori* that certain sets of parameters are independent of each other. For example, location parameters θ and scale parameters σ^2 are frequently assumed to be independent *a priori*. In such cases $p(\theta|\sigma^2) = p(\theta)$ and hence $p(\theta,\sigma^2) = p(\theta)p(\sigma^2)$. Then the non-informative prior distributions $p(\theta)$ and $p(\sigma^2)$ should be calculated yielding:

$$p(\theta,\sigma^2) \propto |I_n(\theta)|^{\frac{1}{2}} |I_n(\sigma^2)|^{\frac{1}{2}}$$

4.3 Jeffreys' Priors for Normal Models

Consider a set of observations y from a multivariate normal distribution with mean 0 and dispersion $V(\sigma^2)$ where σ^2 are unknown. Decompose V into its eigenvalues λ_i and eigenspaces S_i giving $V = \sum_{i=1}^{s} \lambda_i S_i$. The S_i matrices are the eigen-projections of the λ_i and have ranks m_i. Hence $|V| = \prod_{i=1}^{s} \lambda_i^{m_i}$. The likelihood for this model is then:

$$l(\lambda \mid y) \propto \prod_{i=1}^{s} \lambda_i^{-\frac{1}{2}m_i} \exp -\tfrac{1}{2} \sum_{i=1}^{s} \lambda_i^{-1} y' S_i y$$

$$L(\lambda \mid y) = \log l(\lambda \mid y) = constant - \tfrac{1}{2} \sum_{i=1}^{s} m_i \log \lambda_i - \tfrac{1}{2} \sum_{i=1}^{s} \lambda_i^{-1} y' S_i y$$

$$\frac{\partial L}{\partial \lambda_i} = -\tfrac{1}{2} \frac{m_i}{\lambda_i} + \tfrac{1}{2} \frac{y' S_i y}{\lambda_i^2}$$

$$\frac{\partial^2 L}{\partial^2 \lambda_i} = \tfrac{1}{2} \frac{m_i}{\lambda_i^2} - \frac{y' S_i y}{\lambda_i^3}$$

$$\frac{\partial^2 L}{\partial \lambda_i \partial \lambda_j} = 0 \quad \text{for } i \neq j$$

To calculate $i(\theta)$ the expectation of the second derivative of L is needed. This requires the expectation of the quadratic form $y'S_i y$. For this case, standard results give:

$$E[y' S_i y] = tr S_i V = m_i \lambda_i$$

$$\text{hence } E\left[\left|\frac{\partial^2 L}{\partial^2 \lambda}\right|\right] = \prod_{i=1}^{s} \left(\tfrac{1}{2} \frac{m_i}{\lambda_i^2} - \frac{y' S_i y}{\lambda_i^3} \right) \propto \prod_{i=1}^{s} \lambda_i^{-2}$$

$$\text{then } |i(\theta)|^{\frac{1}{2}} \propto \prod_{i=1}^{s} \lambda_i^{-1}$$

hence the Jeffreys' prior is proportional to the reciprocal of the product of the eigenvalues.

4.3.1 Jeffreys' Priors for Normal Models with fixed effects and variance components

This section shows that the Jeffrey's prior for likelihoods with both variance components and fixed effects, is the same as the Jeffreys' prior for a model with just

the variance components, provided that it is assumed *a priori* that the location parameters and the scale parameters are independent. Thus the prior $p(\lambda,a) = p(\lambda)p(a)$.

$$l(a|y,V) \propto |V|^{-\frac{1}{2}}\exp-\tfrac{1}{2}(y-Xa)'V^{-1}(y-Xa)$$

$$L = \log l(a|y,V) = constant - \tfrac{1}{2}\log|V| - \tfrac{1}{2}(y-Xa)'V^{-1}(y-Xa)$$

$$\frac{\partial L}{\partial a} = X'V^{-1}(y-Xa)$$

$$\frac{\partial^2 L}{\partial^2 a} = -X'V^{-1}X = constant$$

thus $l(\theta) = constant \longrightarrow p(a) = constant$
Hence $p(\sigma^2,a) = p(\sigma^2)$.

If the factorisation of the prior is not assumed, then the Fisher information $i(\lambda,a)$ must be obtained. This requires the expectation of $\partial^2 L / \partial^2(\lambda,a)$, that is:

$$\mathbf{E}\begin{pmatrix} \frac{\partial^2 L}{\partial^2 \lambda} & \frac{\partial^2 L}{\partial \lambda\, \partial a} \\ \frac{\partial^2 L}{\partial \lambda\, \partial a} & \frac{\partial^2 L}{\partial^2 a} \end{pmatrix}$$

Since $\mathbf{E}\frac{\partial^2 L}{\partial \lambda\, \partial a} \neq 0$, the resulting prior is no longer the reciprocal of the product of the eigenvalues multiplied by a constant. Thus the choice of prior is, strictly speaking, dependent on the initial assumptions.

4.4 Jeffreys' Priors for models with general multivariate *t* errors

The Jeffreys' prior for a general multivariate *t* likelihood with a single variance component is produced in Section 4.4.2. In this restricted case the Jeffreys' prior is the same as the Jeffreys' prior for a multivariate normal likelihood. This result is extended in Section 4.4.3 to allow for many variance components within a diagonal dispersion matrix V. In Section 4.4.4, the result is shown to hold even if V is not diagonal, thus establishing the result for all dispersion matrices V.

4.4.1 Basic Likelihood theory

Let $y \sim t_\nu(Xa, V)$. Thus:

$$p(y|a,\sigma^2) = \frac{\Gamma(\frac{1}{2}(\nu+n))}{\Gamma(\frac{1}{2}\nu)} \pi^{-\frac{1}{2}n} (\nu-2)^{-\frac{1}{2}n} |V|^{-\frac{1}{2}} \left[1 + \frac{(y-Xa)'V^{-1}(y-Xa)}{\nu-2}\right]^{-\frac{\nu+n}{2}}$$

where n is the number of observations in y. Bayes theorem gives:

$$l(\sigma^2,a|y) \propto |V|^{-\frac{1}{2}} \left[1 + \frac{(y-Xa)'V^{-1}(y-Xa)}{\nu-2}\right]^{-\frac{\nu+n}{2}} p(\sigma^2,a) \qquad (4.4.1:1)$$

4.4.2 Jeffreys' Prior for single variance problem

From equation 4.4.1:1, the log likelihood L is:

$$L = constant - \tfrac{1}{2}\log|V| - \tfrac{1}{2}(\nu+n)\log\left[1 + \frac{(y-Xa)'V^{-1}(y-Xa)}{\nu-2}\right]$$

In this section, it is assumed that $V = \sigma^2 I_n$, and that $a = 0$. Then:

$$L = constant - \tfrac{1}{2}n\log\sigma^2 - \tfrac{1}{2}(\nu+n)\log\left[1+\frac{y'y}{(\nu-2)\sigma^2}\right] \qquad (4.4.2:1)$$

Hence:

$$\frac{dL}{d\sigma^2} = -\tfrac{1}{2}\frac{n}{\sigma^2} + \tfrac{1}{2}(\nu+n)\frac{1}{\nu-2+y'y/\sigma^2}\frac{y'y}{(\sigma^2)^2}$$

$$\frac{d^2L}{d^2\sigma^2} = \tfrac{1}{2}n(\sigma^2)^{-2} - \tfrac{1}{2}(\nu+n)(\sigma^2)^{-2}\left\{2\frac{y'y}{(\nu-2)\sigma^2+y'y} - \left[\frac{y'y}{(\nu-2)\sigma^2+y'y}\right]^2\right\}$$

$$i(\sigma^2) = E\left[\frac{d^2L}{d^2\sigma^2}\right] = \int_y \frac{d^2L}{d^2\sigma^2}\, p(y|\sigma^2)\, dy$$

$$i(\sigma^2) = \tfrac{1}{2}\frac{\Gamma(\frac{1}{2}(\nu+n))}{\Gamma(\frac{1}{2}\nu)}(\pi\sigma^2)^{-\frac{1}{2}n}(\nu-2)^{-\frac{1}{2}n}(\sigma^2)^{-2}$$

$$\times \int_{y_1}\cdots\int_{y_n}\left[n-(\nu+n)\left\{2\frac{y'y}{(\nu-2)\sigma^2+y'y}-\left[\frac{y'y}{(\nu-2)\sigma^2+y'y}\right]^2\right\}\right]$$

$$\times\left[1+\frac{y'y}{(\nu-2)\sigma^2}\right]^{-\frac{\nu+n}{2}} dy_n\ldots dy_1$$

Note that the distribution given by equation 4.4.2:1 is spherically symmetric, and that both the likelihood and the second derivative of the log likelihood depend on y only through $y'y$. Change variables to the multi-dimensional polar co-ordinate system specified by $y_1 = r\cos\theta_1$, $y_2 = r\sin\theta_1\cos\theta_2$, ... $y_{n-1} = r\sin\theta_1\ldots\sin\theta_{n-2}\cos\theta_{n-1}$ and $y_n = r\sin\theta_1\ldots\sin\theta_{n-2}\sin\theta_{n-1}$. Thus replace $\{y_1,y_2,\ldots,y_n\}$ with $\{r,\theta_1,\theta_2,\ldots,\theta_{n-1}\}$ where $0 \leqslant r < \infty$, $0 \leqslant \theta_1,\ldots,\theta_{n-2} < \pi$ and $0 \leqslant \theta_{n-1} < 2\pi$. The Jacobian of this transformation is developed in the lemma in Section 4.4.5 and has the value $r^{n-1}\prod_{i=1}^{n-1}\sin^{n-i-1}\theta_i$.

$$i(\sigma^2) = \tfrac{1}{2}\frac{\Gamma(\frac{1}{2}(\nu+n))}{\Gamma(\frac{1}{2}\nu)}(\pi\sigma^2)^{-\frac{1}{2}n}(\nu-2)^{-\frac{1}{2}n}(\sigma^2)^{-2}$$

$$\times\int_{r=0}^{\infty}\int_{\theta_1}\cdots\int_{\theta_{n-1}}\left[n-(\nu+n)\left\{2\frac{r^2}{(\nu-2)\sigma^2+r^2}-\left[\frac{r^2}{(\nu-2)\sigma^2+r^2}\right]^2\right\}\right]$$

$$\times\left[1+\frac{r^2}{(\nu-2)\sigma^2}\right]^{-\frac{\nu+n}{2}} r^{n-1}\prod_{i=1}^{n-1}\sin^{n-i-1}\theta_i\, d\theta_{n-1}\ldots d\theta_1\, dr$$

Note that σ^2 is independent of all of the θ_i integrals. Hence

$$i(\sigma^2) \propto (\sigma^2)^{-2}(\sigma^2)^{-\frac{1}{2}n}$$

$$\times\int_{r=0}^{\infty}\left[n-(\nu+n)\left\{2\frac{r^2}{(\nu-2)\sigma^2+r^2}-\left[\frac{r^2}{(\nu-2)\sigma^2+r^2}\right]^2\right\}\right]\left[1+\frac{r^2}{(\nu-2)\sigma^2}\right]^{-\frac{\nu+n}{2}} r^{n-1}\,dr$$

Now substitute r^2 for $(\nu-2)\sigma^2\tan^2\phi$, hence $dr = \sqrt{(\nu-2)\sigma^2}\sec^2\phi$, and simplify:

$$i(\sigma^2) \propto (\sigma^2)^{-2}(\sigma^2)^{-\frac{1}{2}n}$$

$$\times\int_0^{\frac{1}{2}\pi}\{n-(\nu+n)(2\sin^2\phi-\sin^4\phi)\}\cos^{\nu+n}\phi\,(\nu-2)^{\frac{1}{2}n-\frac{1}{2}}(\sigma^2)^{\frac{1}{2}n-\frac{1}{2}}\tan^{n-1}\phi\,(\nu-2)^{\frac{1}{2}}(\sigma^2)^{\frac{1}{2}}\sec^2\phi\,d\phi$$

$$i(\sigma^2) \propto (\sigma^2)^{-2}\int_0^{\frac{1}{2}\pi}\{n-(\nu+n)(2\sin^2\phi-\sin^4\phi)\}\cos^{\nu+n}\phi\tan^{n-1}\phi\sec^2\phi\,d\phi$$

Hence $i(\sigma^2) \propto (\sigma^2)^{-2}$. Thus the Jeffreys' prior is proportional to $1/\sigma^2$ as in the Normal case.

4.4.3 Jeffreys' Prior for a problem with several variance components

Generalise the results of section 4.4.2 to allow for several different eigenvalues in the dispersion matrix V. First consider a V matrix with two distinct eigenvalues. Suppose that V had n_1 entries of σ_1^2 and $n_2 = n-n_1$ entries of σ_2^2. Then from equation 4.4.1:1 the log likelihood L is:

$$L = constant - \tfrac{1}{2}n_1 \log\sigma_1^2 - \tfrac{1}{2}n_2 \log\sigma_2^2 - \tfrac{1}{2}(\nu+n)\log\left[1 + \frac{Y_1^2/\sigma_1^2 + Y_2^2/\sigma_2^2}{\nu-2}\right] \quad (4.4.3:1)$$

where $Y_1^2 = \sum_{i=1}^{n_1} y_i^2$ and $Y_2^2 = \sum_{n_1+1}^{n_2} y_i^2$

differentiating equation 4.4.3:1 gives:

$$\frac{\partial L}{\partial\sigma_1^2} = -\tfrac{1}{2}\frac{n_1}{\sigma_1^2} + \tfrac{1}{2}(\nu+n)\frac{1}{\nu-2+Y_1^2/\sigma_1^2+Y_2^2/\sigma_2^2}\frac{Y_1^2}{(\sigma_1^2)^2}$$

$$\frac{\partial^2 L}{\partial^2\sigma_1^2} = \tfrac{1}{2}\frac{n_1}{(\sigma_1^2)^2} - \tfrac{1}{2}\frac{\nu+n}{(\sigma_1^2)^2}\left\{2\frac{Y_1^2/\sigma_1^2}{\nu-2+Y_1^2/\sigma_1^2+Y_2^2/\sigma_2^2} - \left[\frac{Y_1^2/\sigma_1^2}{\nu-2+Y_1^2/\sigma_1^2+Y_2^2/\sigma_2^2}\right]^2\right\}$$

and

$$\frac{\partial^2 L}{\partial\sigma_1^2\partial\sigma_2^2} = \tfrac{1}{2}\frac{\nu+n}{\sigma_1^2\sigma_2^2}\frac{Y_1^2/\sigma_1^2\, Y_2^2/\sigma_2^2}{(\nu-2+Y_1^2/\sigma_1^2+Y_2^2/\sigma_2^2)^2}$$

$$i(\sigma^2) = E\left|\left[\frac{\partial^2 L}{\partial^2\sigma^2}\right]\right|$$

consider first $E\left[\frac{\partial^2 L}{\partial^2\sigma_1^2}\right]$

$$E\left[\frac{\partial^2 L}{\partial^2\sigma_1^2}\right] = \tfrac{1}{2}(\sigma_1^2)^{-2}\frac{\Gamma(\frac{1}{2}(\nu+n))}{\Gamma(\frac{1}{2}\nu)}(\pi\nu)^{-\frac{1}{2}n}(\sigma_1^2)^{-\frac{1}{2}n_1}(\sigma_2^2)^{-\frac{1}{2}n_2}$$
$$\times \int_y\left[n_1 - (\nu+n)\left\{2\frac{Y_1^2/\sigma_1^2}{\nu-2+Y_1^2/\sigma_1^2+Y_2^2/\sigma_2^2} - \left[\frac{Y_1^2/\sigma_1^2}{\nu-2+Y_1^2/\sigma_1^2+Y_2^2/\sigma_2^2}\right]^2\right\}\right]$$
$$\times\left[1+\frac{Y_1^2}{(\nu-2)\sigma_1^2}+\frac{Y_2^2}{(\nu-2)\sigma_2^2}\right]^{-\frac{n+1}{2}} dy$$

Now change variables as in section 4.4.2 to map $\{y_1, y_2 \ldots y_{n_1}\}$ to $\{r_1, \theta_{1,1}, \theta_{1,2} \ldots \theta_{1,n_1-1}\}$ and to map $\{y_{n_1+1} \ldots y_n\}$ to $\{r_2, \theta_{2,1}, \theta_{2,2} \ldots \theta_{2,n_2-1}\}$. Then all the θ integrations drop out leaving:

$$E\left[\frac{\partial^2 L}{\partial^2 \sigma_1^2}\right] \propto (\sigma_1^2)^{-2}\,(\sigma_1^2)^{-\frac{1}{2}n_1}\,(\sigma_2^2)^{-\frac{1}{2}n_2} -$$

$$\times \int_{r_1}\int_{r_2}\left[n_1-(\nu+n)\left\{2\frac{r_1^2/\sigma_1^2}{\nu-2+r_1^2/\sigma_1^2+r_2^2/\sigma_2^2}-\left(\frac{r_1^2/\sigma_1^2}{\nu-2+r_1^2/\sigma_1^2+r_2^2/\sigma_2^2}\right)^2\right\}\right]$$

$$\times \left(\nu-2+\frac{r_1^2}{\sigma_1^2}+\frac{r_2^2}{\sigma_2^2}\right)^{-\frac{\nu+n}{2}} r_1^{n_1-1}\,r_2^{n_2-1}\,dr_2\,dr_1$$

Substitute $(\nu-2+r_1^2/\sigma_1^2)\sigma_2^2\tan^2\phi_2$ for r_2^2, integrate out over ϕ_2. Then substitute $(\nu-2)\sigma_1^2\tan^2\phi_1$ for r_1^2 and integrate out over ϕ_1 to get:

$$E\left[\frac{\partial^2 L}{\partial^2 \sigma_1^2}\right] \propto (\sigma_1^2)^{-2}$$

In an identical manner the other terms can be calculated yielding:

$$i(\sigma^2) = \left|E\left[\frac{\partial^2 L}{\partial^2 \sigma^2}\right]\right| = \begin{vmatrix} K_{11}(\sigma_1^2)^{-2} & K_{12}(\sigma_1^2)^{-1}(\sigma_2^2)^{-1} \\ K_{21}(\sigma_1^2)^{-1}(\sigma_2^2)^{-1} & K_{22}(\sigma_2^2)^{-2} \end{vmatrix} \quad \text{some } K_{ij}$$

$$i(\sigma^2) \propto (\sigma_1^2)^{-2}(\sigma_2^2)^{-2} \qquad (4.4.3{:}2)$$

Thus the Jeffreys' prior is proportional to $1/(\sigma_1^2\sigma_2^2)$ as in the Normal case.

Generalisation

Now consider V to be a diagonal matrix with m distinct eigenvalues. The terms in equation 4.4.3:2 demonstrate both the diagonal and off-diagonal entries in the expectation of the second derivative log likelihood. Thus by exchanging subscripts we can obtain the equivalent version of equation 4.4.3:2 without further work.

$$i(\sigma^2) = \left|E\left[\frac{\partial^2 L}{\partial^2 \sigma^2}\right]\right| \propto \begin{vmatrix} K_{11}(\sigma_1^2)^{-2} & K_{12}(\sigma_1^2)^{-1}(\sigma_2^2)^{-1} & .. & K_{1m}(\sigma_1^2)^{-1}(\sigma_m^2)^{-1} \\ K_{21}(\sigma_1^2)^{-1}(\sigma_2^2)^{-1} & K_{22}(\sigma_2^2)^{-2} & .. & K_{2m}(\sigma_2^2)^{-1}(\sigma_m^2)^{-1} \\ .. & .. & .. & .. \\ K_{m1}(\sigma_1^2)^{-1}(\sigma_m^2)^{-1} & K_{m2}(\sigma_2^2)^{-1}(\sigma_m^2)^{-1} & .. & K_{mm}(\sigma_m^2)^{-2} \end{vmatrix}$$

$$i(\sigma^2) \propto \left(\prod_{i=1}^{m}\sigma_i^2\right)^{-2}$$

Hence the Jeffreys' prior is the reciprocal of the product of the eigenvalues, as is the case for the Normal distribution.

4.4.4 Jeffreys' Prior for any general multivariate *t* distribution

In this section, the Jeffreys' prior for any multivariate *t* distribution is produced. Recall:

$$i(\sigma^2) = \left| \int_y \frac{\partial^2 L(y, V(\sigma^2))}{\partial^2 \sigma^2} p(y|V(\sigma^2))\, dy \right| = \left| \int_y \frac{\partial^2 L(y'V^{-1}y))}{\partial^2 \sigma^2} p(y'V^{-1}y)\, dy \right| \quad (4.4.4:1)$$

Since V is a symmetric real, it can be diagonalised. Thus let $V = M'\Lambda M$ where M is a rotation matrix (hence $M^{-1} = M'$), and Λ is a diagonal matrix of eigenvalues. Define $z = My$, $y = M^{-1}z$, $y' = z'M$. Change the integral in equation 4.4.4:1 from y space to z space.

$$i(\sigma^2) = \left| \int_z \frac{\partial^2 L(y'V^{-1}y)}{\partial^2 \sigma^2} p(y'V^{-1}y) |M^{-1}|\, dz \right|$$

Now $i(\sigma^2)$ depends on y only through the quadratic form $y'V^{-1}y$, but $y'V^{-1}y = (z'M)(M'\Lambda^{-1}M)(M^{-1}z) = z'\Lambda^{-1}z$. Also $|M| = 1$ as M is a rotation matrix. Thus

$$i(\sigma^2) = \left| \int_z \frac{\partial^2 L(z'\Lambda^{-1}z)}{\partial^2 \sigma^2} p(z'\Lambda^{-1}z)\, dz \right| = \left| \int_z \frac{\partial^2 L(z,\Lambda)}{\partial^2 \sigma^2} p(z|\Lambda)\, dz \right|$$

$$i(\sigma^2) = \left| \int_y \frac{\partial^2 L(y,\Lambda)}{\partial^2 \sigma^2} p(y|\Lambda)\, dy \right| \quad (4.4.4:2)$$

Comparing equations 4.4.4:1 and 4.4.4:2 it is seen that we can replace V by Λ without changing the results. Consequently section 4 shows that the Jeffreys' prior for a general multivariate *t* distribution is the same as the Jeffreys' prior for a multivariate Normal distribution.

4.4.5 Lemma - the Jacobian of the transformation to polar coordinates

This lemma determines the Jacobian necessary for transforming the y of n observations into an n dimensional polar co-ordinate system. Thus we transform $y = \{y_1, y_2, \ldots, y_n\}$ into $\{r, \theta_1, \ldots, \theta_{n-1}\}$.

where $y_1 = r\cos\theta_1$

$y_2 = r\sin\theta_1\cos\theta_2$

$y_3 = r\sin\theta_1\sin\theta_2\cos\theta_3$

$$y_{n-1} = r\sin\theta_1\sin\theta_2\sin\theta_3\ldots\sin\theta_{n-2}\cos\theta_{n-1}$$

$$y_n = r\sin\theta_1\sin\theta_2\sin\theta_3\ldots\sin\theta_{n-2}\sin\theta_{n-1}$$

where $0 \leqslant r < \infty$, $0 \leqslant \theta_1,\ldots,\theta_{n-2} < \pi$ $0 \leqslant \theta_{n-1} < 2\pi$ for $i = 1\ldots n-1$

Then

$$\frac{\partial y_i}{\partial r} = y_i/r \text{ and } \frac{\partial y_i}{\partial \theta_j} = \begin{cases} 0 & i < j \\ -y_i\tan\theta_i & i = j \\ y_i\cot\theta_i & i > j \end{cases}$$

Thus
$$\left|\frac{\partial(y)}{\partial(r,\theta)}\right| = \frac{1}{r}\prod_{i=1}^{n} y_i \prod_{i=1}^{n-1}\cot\theta_i \begin{vmatrix} 1 & -\tan^2\theta_1 & 0 & 0 & .. & 0 \\ 1 & 1 & -\tan^2\theta_2 & 0 & .. & 0 \\ 1 & 1 & 1 & -\tan^2\theta_3 & .. & 0 \\ .. & .. & .. & .. & .. & .. \\ 1 & 1 & 1 & 1 & .. & -\tan^2\theta_{n-1} \\ 1 & 1 & 1 & 1 & .. & 1 \end{vmatrix}$$

$$\left|\frac{\partial(y)}{\partial(r,\theta)}\right| = \frac{1}{r}\prod_{i=1}^{n} y_i \prod_{i=1}^{n-1}\cot\theta_i \prod_{i=1}^{n-1}(1+\tan^2\theta_i)$$

$$= \frac{1}{r}\prod_{i=1}^{n} y_i \prod_{i=1}^{n-1}\frac{1}{\cos\theta_i\sin\theta_i}$$

$$\left|\frac{\partial(y)}{\partial(r,\theta)}\right| = r^{n-1}\prod_{i=1}^{n-1}\sin^{n-i-1}\theta_i$$

4.5 Bernardo Priors

In contrast to the Jeffreys' approach based on invariance, Bernardo (1979) introduces a new philosophy for priors based on the expected information to be gained from an experiment. Bernardo shows that for simple regular cases with asymptotic normality, these new priors correspond to the Jeffreys' priors. Thus Bernardo priors for t distributions are the same as the Jeffreys' priors.

It is interesting to note that although this material has been around for a few years, little work has been done on large multiparameter models. The problem of whether to use a prior for the joint or conditional distribution has also received little attention and requires further work. Similarly the question of what is a nuisance

parameter is important as different priors are selected depending on whether the scale or location and scale parameters are of interest (see section 4.3.1 where a related problem arises with the Jeffrey's prior).

The Bernardo philosophy is based on the expected information to be gained from an experiment $\varepsilon = \{X, \Theta, p(x|\theta)\}$ which is one observation of the random variable $x \in X$, where x is distributed according to $p(x|\theta)$ for some $\theta \in \Theta$. $\varepsilon(n)$ denotes n replications of the experiment. Take $p(\theta)$ to be a prior density for θ. Without loss of generality take $p(\theta) > 0 \; \forall \theta \in \Theta$. Then define the *expected information* about θ provided by ε when the prior is $p(\theta)$ as:

$$I^{\theta}\{\varepsilon, p(\theta)\} = \int p(x) \int p(\theta|x) \log \frac{p(\theta|x)}{p(\theta)} d\theta \, dx$$

where $p(x) = \int p(x|\theta) p(\theta) \, d\theta$ and $p(\theta|x) = p(x|\theta) p(\theta) / p(x)$

Let $I^{\theta}\{\varepsilon(n), p(\theta)\}$ denote the information to be gained from n independent replications of ε. By performing ∞ replications of ε, one would get to know θ exactly. Thus $I^{\theta}\{\varepsilon(\infty), p(\theta)\}$ is the amount of missing information about θ when the prior is $p(\theta)$. It is sensible to define a non-informative prior (what Bernardo calls *vague initial knowledge*) as the density $\pi(\theta)$ which maximises the missing information over the admissible class of priors.

If Θ is a continuous space, then $I^{\theta}\{\varepsilon(\infty), p(\theta)\}$ will usually be ∞ as an infinite amount of information is required to determine a real number. In these cases define the non-informative prior as the limit as $n \longrightarrow \infty$ of the priors which maximise $I^{\theta}\{\varepsilon(n), p(\theta)\}$.

Often a non-informative prior can be obtained more rapidly than using the limiting process above. Suppose y is the data obtained from $\varepsilon(k)$, then under sufficient regularity:

$$I^{\theta}\{\varepsilon, p(\theta)\} = \int_{\theta} p(\theta) \log \left\{ \frac{\exp - \int p(y|\theta) H\{p(\theta|y)\} \, dy}{p(\theta)} \right\} d\theta \qquad (4.5:1)$$

or alternatively

$$I^{\theta}\{\varepsilon, p(\theta)\} = \int_{\theta} p(\theta) \log \left\{ \frac{\exp \int p(y|\theta) \log p(\theta|y) \, dy}{p(\theta)} \right\} d\theta \qquad (4.5:2)$$

where $H\{.\}$ is the entropy:

$$H\{p(\theta)\} = - \int_{\theta} p(\theta) \log p(\theta) \, d\theta$$

A simple exercise in calculus of variations shows that given the constraint $\int a(\theta)d\theta < \infty$, integrals of the form $\int p(\theta)\log\{a(\theta)/b(\theta)\}d\theta$ are maximised when $a(\theta) \propto b(\theta)$. Maximising equation 4.5:1 or 4.5:2 would appear to deliver the non-informative prior $p(\theta)$. However, this is somewhat misleading as in both equations 4.5:1 and 4.5:2 the numerator is a function of the denominator, as $p(\theta|y)$ depends on $p(\theta)$. However it will still be true that:

$$p(\theta) \propto \exp(-\int p(y|\theta) H\{p(\theta|y)\}\, dy)$$

or $$p(\theta) \propto \exp(\int p(y|\theta) \log p(\theta|y)\, dy)\,.$$

Note that these definitions are cyclic, in the sense that $p(\theta)$ is needed to evaluate the right hand integrals as it as a term in the posterior $p(\theta|y)$. However, under sufficient regularity conditions, asymptotically $p(\theta|y)$ is independent of $p(\theta)$ hence:

$$\pi_n(\theta) \propto \exp\left(-\int p(y|\theta) H(p^{\dagger}(\theta|y))\, dy\right) \qquad (4.5:3)$$

or equivalently

$$\pi_n(\theta) \propto \exp\left(\int p(y|\theta) \log p^{\dagger}(\theta|y)\, dy\right)$$

for large n, with $p^{\dagger}(\theta|y)$ as the asymptotic posterior density for θ (which does not depend on the prior).

4.5.1 Bernardo Prior for a single variance problem

It is interesting to note how some priors can be derived directly from the definitions. For example, the single variance model can be tackled as follows. Take n observations from a Normal distribution with known mean μ and unknown variance σ^2.

$$p(\sigma^2|y) \propto p(y|\sigma^2)\, p(\sigma^2) \propto p(\sigma^2)\, (\sigma^2)^{-\frac{1}{2}n} \exp -\tfrac{1}{2}\sum_{i=1}^{n} (y_i - \mu)^2/\sigma^2$$

$$\propto p(\sigma^2)\, (\sigma^2)^{-\frac{1}{2}n} \exp -\tfrac{1}{2}ns^2/\sigma^2$$

where $s^2 = \sum_{i=1}^{n} (y_i - \mu)^2/n$

Recall that $s^2 \sim \frac{\sigma^2}{n}\chi^2_n$, and that a χ^2_n distribution has mean n and variance $2n$. Asymptotically as $n \longrightarrow \infty$, the distribution of s^2 tends to $N(\sigma^2, \frac{2(\sigma^2)^2}{n})$. But s^2 is $\hat{\sigma^2} \mid y$, hence the asymptotic density of $\hat{\sigma^2} \mid y$ is Normal.

It is well known that the entropy of a Normal distribution with variance v is $\frac{1}{2}\log 2\pi e v$, hence the entropy of the asymptotic density of $\hat{\sigma^2} \mid y$ is $\frac{1}{2}\log\{2\pi e \frac{2(\sigma^2)^2}{n}\}$ which does not depend on the data.

$$H(p^{\dagger}(\hat{\sigma}^2|y)) = \log\sigma^2 + \tfrac{1}{2}\log\{4\pi e/n\}$$

Thus from equation 4.5:3

$$\pi_n(\sigma^2) \propto \exp\left(-\log\sigma^2 - \tfrac{1}{2}\log\{4\pi e/n\}\right)$$

$$\pi_n(\sigma^2) \propto 1/\sigma^2$$

Thus the Bernardo prior is seen to be the same as the Jeffrey's prior from Section 4.3.

Chapter 5
Bayesian Analysis of a multi-stratum experimental design

5.1 Introduction

In his 1983 thesis, Knuiman (1983) describes an agricultural field experiment in which 72 apple trees are subject to two treatments, namely irrigation and thinning. The trees are arranged in a nested block structure of six blocks, three plots per block and four trees per plot. Three irrigation regimes are considered, assigned at random to plots. Each of the four trees in each plot is allocated a different thinning regime. Thus, each irrigation and thinning combination is represented exactly once in each block. The experiment was repeated over several years. This analysis is concerned with the weight (Kg) of apples produced per tree in the 1975 season.

The Knuiman example is used in this chapter to illustrate three different areas.

i) It demonstrates the possibilities of Bayesian analysis undertaken by high dimensional numerical integration such as the BAYES4 computer package. The main features of BAYES4 were discussed in section 1.4 of Chapter 1. Numerical integration is carried out in 3, 9 and 15 dimensions. As indicated in Chapter 2, many of the margins of interest can either be evaluated numerically from the full joint distribution, or after some analytic integration has been performed. This example thus allows several marginal distributions to be computed by different routes, so providing a useful check on the accuracy of the suite of integration routines central to BAYES4.

ii) An extended sensitivity analysis is performed by changing the prior distribution, the error distribution, and the fixed effect structure. The effects of each of these changes can be seen by looking at the resulting marginal distributions for the model parameters. This example provides a powerful argument that such sensitivity studies should be part of any routine Bayesian analysis.

iii) The analyses performed call upon the algebra of the preceding chapters for the analytical integrations and matrix manipulations.

5.1.1 Details of the Agricultural Field Trial

Data is available for the apple yields from two years, namely 1975 and 1977. The irrigation and thinning treatments were applied to the trees over a long period of time prior to the experiment. The three irrigation regimes are:

W1 No irrigation
W2 Three or four irrigations at monthly intervals
W3 Soil maintained at field capacity by weekly irrigation

The four thinning policies are:
T1 All fruit removed in the first seven years (1965 - 1971)
T2 Two chemical thinning sprays every two years
T3 Normal commercial thinning, ie. one spray every two years
T4 Minimal thinning - a few fruit removed to prevent limb breakage

5.1.2 The data

Two years of data are available. The analyses can be performed using the data from either year. Alternatively, the two sets of data may be combined by using the posterior density from one year's data as the prior distribution for the other year's data. The analyses presented in this chapter are based principally upon the 1975 data (Table 5.1.2:1). In the later sections of the chapter a sensitivity analysis is performed using (amongst other things) an informative prior rather than a reference prior. The 1977 data (Table 5.1.2:2) are used to get values describing a plausible informative prior.

Apple yields (Kg) for 1975							
Irrigation	Thinning	Block 1	2	3	4	5	6
W1	T1	139	332	342	398	193	342
	T2	233	299	428	406	244	351
	T3	241	265	422	394	243	332
	T4	268	333	412	351	297	446
W2	T1	401	402	241	329	162	440
	T2	359	283	265	380	313	163
	T3	296	487	296	407	303	388
	T4	488	353	347	502	431	406
W3	T1	560	164	312	363	379	508
	T2	400	360	401	469	498	455
	T3	528	419	447	513	464	364
	T4	586	135	426	519	488	483

Table 5.1.2:1

Apple yields (Kg) for 1977							
Irrigation	Thinning	Block 1	2	3	4	5	6
W1	T1	50	181	208	270	97	107
	T2	153	208	85	253	154	216
	T3	156	194	270	261	154	296
	T4	213	179	326	328	249	349
W2	T1	357	256	166	120	64	299
	T2	329	150	65	234	176	12
	T3	328	342	347	309	329	312
	T4	400	204	293	300	367	359
W3	T1	442	28	268	158	295	451
	T2	379	200	231	302	97	434
	T3	498	310	350	407	520	346
	T4	452	306	342	237	531	454

Table 5.1.2:2

5.1.3 The Models

Write y_{ijk} for the yield of tree k in plot j, block i, and assume normally distributed errors. The maximal model may be expressed as:

$$y \sim N(X\tau, V)$$

where τ is a vector of 12 treatment means, X gives the allocation of treatments to experimental units and the dispersion matrix V is a 72×72 block diagonal matrix with non-zero entries:

$$var(y_{ijk}) = \sigma_b^2 + \sigma_a^2 + \sigma^2$$

$$cov(y_{ijk}, y_{ijl}) = \sigma_b^2 + \sigma_a^2 \qquad k \neq l$$

$$cov(y_{ijk}, y_{ilm}) = \sigma_b^2 \qquad j \neq l$$

This model thus involves fifteen parameters: twelve treatment means and three variance components. Let $\lambda = (\lambda_1, \lambda_2, \lambda_3)'$ be the distinct eigenvalues of V, ie. $\lambda_1 = \sigma^2$, $\lambda_2 = \sigma^2 + 4\sigma_a^2$ and $\lambda_3 = \sigma^2 + 4\sigma_a^2 + 12\sigma_b^2$. V is positive definite if and only if $\lambda_i > 0 \forall i$. Limits for integration over the parameters of the block structure are thus simplified if the likelihood is expressed as a function of τ and λ.

A standard analysis of variance for these data gives an F statistic for interaction between treatments of approximately 1·0. Therefore two models are considered one assuming main-effects only (and hence with only six treatment parameters), the other allowing interaction.

5.1.4 The use of the BAYES4 numerical integration package

As outlined in Chapter 1, the BAYES4 package enables numerical integration to be performed on high dimensional likelihoods, to yield moments for all the parameters, and marginal distributions for selected parameters or pairs of parameters. BAYES4 also allows the calculation of user defined integrals over parameter space using *special function analysis*. Thus if the parameters in the likelihood were a set of variances σ^2, special function analysis would enable the calculation of integrals of the form $\int_{\sigma^2} f(\sigma^2)\, p(\sigma^2|y)\, d\sigma^2$ for an arbitrary function $f(\sigma^2)$.

Consider the 15 dimensional maximal model. This comprises 3 variance components and 12 fixed effects. As shown in Chapter 2, the 12 fixed effects can be analytically integrated out to yield a 3 dimensional marginal posterior distribution on the variance components, or equivalently a 3 dimensional posterior distribution on the eigenvalues. However, it is not possible to integrate out analytically the 3 variance components. These results lead to two possible ways of performing the analyses.

i) Use the full 15 dimensional joint posterior density and let BAYES4 integrate out the 13 or 14 parameters that are not of interest. This technique would allow the calculation of univariate or bivariate marginal distributions for either variance components or fixed effects or combinations with equal ease, together with all the first and second moments of the joint density.

ii) An alternative and much faster technique is to use the 3 dimensional marginal posterior density for the eigenvalues in conjunction with either a 1 dimensional distribution for a fixed effect or a 2 dimensional distribution for a pair of fixed effects conditional on the eigenvalues, the latter being viewed as *special functions*. Assuming a normal error structure, the algebraic form of the 3 dimensional marginal density for the eigenvalues was produced in Chapter 2 equation 2.2.1:1. For the maximal model this density has a simple analytic form (see Section 5.2.2). Assuming t errors the corresponding density is produced in Section 5.5.1.

Methods i) and ii) have led to the development of two computer programs APPLES and MAPPLES. The APPLES program tackles the maximal model via the 3 dimensional marginal distribution for the eigenvalues as used in method ii). The MAPPLES program directly implements the full 15 dimensional likelihood as in method i). An option allows for the main-effects only model, and then maps the 6 fixed effects into the 12 cell effects. With MAPPLES univariate or bivariate distributions can be obtained using a Gauss-Hermite grid over one or two dimensions, and Monte-Carlo integration over the other 14, 13, 8 or 7 dimensions as appropriate.

Recall that BAYES4 requires initial estimates of the first and second moments of the parameters. These estimates are then iteratively updated until they have stabilised. For a 15 parameter model, the provision of good estimates is important, because otherwise a highly improbable region of parameter space may be examined. Frequently BAYES4 can recover from a poor set of initial estimates, but this may take many iterations and waste computer time. In this particular example, the

moments for the three eigenvalue components, can be obtained using the moments calculated by the 3 dimensional APPLES program. Second moments for the 12 fixed effects can also be obtained from the APPLES program again using *special function analysis*.

Clearly the computational load is substantially heavier using the method i) rather than with method ii). However, the efficiency of the spherical Monte-Carlo integration rules within BAYES4 makes the numerical integration method practical even with 15 dimensions. During the development of this example, marginal densities were frequently computed using the Monte-Carlo technique on the joint 15 dimensional density $l(\tau,\lambda|y)$ and compared with results obtained via the alternative route which used analytic integration. The exercise proved to be a very useful check on the accuracy of the high dimensional numerical integration procedures, and was in itself a good reason for implementing the full 15 dimensional problem.

A final point concerns the parametrisation used with the APPLES and MAPPLES programs. The BAYES4 package works best on likelihoods that are well approximated by a low degree polynomial times a normal distribution. Transformations of the parameter space that yield a likelihood surface that is nearer to this form, help BAYES4 to converge and stabilise. Marginal distributions for variance components (or more precisely, for eigenvalues of dispersion matrices) typically have a marked right skew, and are also constrained to be positive. Incorporating this type of constraint into BAYES4 may destroy stability since successive iterations may drop points either side of the constraint. A way of avoiding the positivity constraint and making the likelihood more nearly Normal, is to reparametrise in terms of the logs of the eigenvalues rather than the eigenvalues. This is done in both the APPLES and the MAPPLES programs, and good results are obtained (see Sections 5.7 and 5.9). A consequence of this is that BAYES4 does not produce marginal distributions for the eigenvalues, but rather marginal distributions for the log eigenvalues are produced. The graphical presentation and manipulation program GR is used to transform these marginals back to marginals on the eigenvalues and hence the variance components, and also to calculate first and second moments on this scale rather than on the log scale.

5.1.5 Use of Special Function Analysis

Note that the effective posterior densities considered in method ii) from section 5.1.4, and implemented in the APPLES program, are 4 or 5 dimensional. In practice it is better to implement these distributions as a 3 dimensional likelihood with 1 or 2 dimensions (for the fixed effect, or pair of fixed effects) implemented using special function analysis. There are three reasons for this choice.

i) By coding the problem this way, it is possible to iterate on the 3 eigenvalue dimensions until they have stabilised, whilst leaving the special function analysis switched off. This enables very rapid and accurate calculation of the marginal distribution on the eigenvalues.

ii) Given the analytic form for the marginal distribution on the eigenvalues, the only purpose of extending the dimensionality of the problem by incorporating 1 or 2 fixed effects is to enable calculation of the marginal distribution(s) for the fixed effect(s). Depending on the error structure, the univariate and bivariate marginals for the τ_{jk}'s unconditional on the eigenvalues can be obtained as a weighted mixtures of either normal distributions (see Section 5.3) or t distributions (see Section 5.5.2). These marginals may be computed more efficiently using special function analysis than using the standard BAYES4 integration routines, as is demonstrated in Section 5.1.6 below.

iii) Conditional on the eigenvalues, the univariate distribution for a fixed effect is symmetric, and the bivariate distribution of a pair of fixed effects is rotationally symmetric of order two. These observations enable the number of points needed in τ space to be halved - thus substantially reducing the execution time.

5.1.6 Gauss-Hermite integration v Special Function Analysis

The standard method of producing univariate or bivariate marginal distributions using BAYES4 is to make BAYES4 integrate over those dimensions using a Gauss-Hermite grid of points. BAYES4 can then produce a lattice of spot heights over the requested Gauss-Hermite dimensions. However, the Gauss-Hermite integration rule scatters points fairly widely, and some of the points will lie outside regions of appreciable probability for well behaved marginals. Thus from the perspective of statistical analysis (though not from the perspective of numerical

integration) the evaluation of the likelihood at some of these points is a waste of time and effort. Conversely, if the range of statistically plausible values were known in advance, then all the points could be constrained to be within this region. The extent of this problem is illustrated below.

Consider a one dimensional numerical integration performed using a Gauss-Hermite integration rule. The points at which the likelihood is evaluated are determined by the current estimates for the mean and variance, and also by the number of points n used by the rule. The number of standard deviations of these points from the mean is shown in Table 5.1.6:1 below for values of n from 2 to 12. (Note that since the Gauss-Hermite rule is symmetric, only the positive half of the values need be shown.)

Points at which a Gauss-Hermite integration rule is evaluated											
n	number of standard deviations from mean at which to evaluate the likelihood										
2	1·000										
3	0	2·680									
4	0·742		2·334								
5	0	1·356		2·857							
6	0·617		1·889		3·324						
7	0	1·154		2·367		3·750					
8	0·539		1·637		2·802		4·145				
9	0	1·023		2·077		3·205		4·513			
10	0·485		1·466		2·484		3·582		4·850		
11	0	0·929		1·876		2·865		3·936		5·188	
12	0·444		1·340		2·260		3·224		4·272		5·501

Table 5.1.6:1

In practice it is found that points further than 3 or 3·5 standard deviations from the mean are useless for statistical inference in the Knuiman example, as the probability of being so far in the tails of the likelihood is very low. To get accurate precise numerically calculated marginals requires there to be many effective points at which the marginal density has been evaluated. (In this context effective points are those within 3 or 2·5 standard deviations of the mean). As can be seen in Table 5.1.6:2 below, simply increasing n is not an efficient way of increasing the number of effective points once n exceeds 6.

Number of effective points using a Gauss-Hermite integration rule											
total number of points n	2	3	4	5	6	7	8	9	10	11	12
number of points $< 3{\cdot}5\sigma$	2	3	4	5	6	5	6	7	6	7	8
number of points $< 3\sigma$	2	3	4	5	4	5	6	5	6	7	6

Table 5.1.6:2

Thus when using a 12 point Gauss-Hermite grid, only 50% of the points calculated fall within $\pm 3\sigma$ of the mean. From the statistical point of view this is most inefficient. (It must be remembered that as a method of integrating the posterior density the Gauss-Hermite rule is efficient. Evaluating points that are far into the tails of the density allows the integration method to recover more quickly from poor initial estimates for the first and second moments of the density.)

5.1.7 Implementation of the main-effects only model

The joint posterior density for the main-effects only model has 9 dimensions. Once again, interest will usually be in the marginal distribution for a particular τ_{jk}, or a bivariate pair of τ_{jk}'s. The direct solution for these marginals would require 8 or 9 dimensions to be integrated out analytically and this is not feasible. The marginal density for the eigenvalues can be written down algebraically, but this does not have a simple analytic form. However the main-effects only model can be viewed as a submodel of the maximal model and analysed using the Monte-Carlo method that was feasible on the 15 dimensional problem. Since there are only 9 dimensions, the computational load is substantially lower than with the maximal model.

5.1.8 Notation

A number of symbols are heavily used throughout this chapter. I_4 and J_4 refer to a 4×4 identity matrix, and a 4×4 block of ones. 1_6 denotes a column of 6 ones, and $\otimes$ is used for Kronecker products. Finally σ^2 refers to the residual variance, and $\boldsymbol{\sigma}^2$ to the vector of variance components $(\sigma^2, \sigma_a^2, \sigma_b^2)$. Using this notation, the

dispersion matrix for the Knuiman problem can be expressed:

$$V(\sigma^2) = \sigma_b^2 I_6 \otimes J_3 \otimes J_4 + \sigma_a^2 I_6 \otimes I_3 \otimes J_4 + \sigma^2 I_6 \otimes I_3 \otimes I_4 \qquad (5.1.8{:}1)$$

5.2 The marginal posterior density for the eigenvalues

In terms of its eigenvalues λ_1, λ_2 and λ_3, the dispersion matrix given in equation 5.1.8:1 can be expressed as:

$$V = \tfrac{1}{12}(\lambda_3 - \lambda_2) I_6 \otimes J_3 \otimes J_4 + \tfrac{1}{4}(\lambda_2 - \lambda_1) I_6 \otimes I_3 \otimes J_4 + \lambda_1 I_6 \otimes I_3 \otimes I_4 \qquad (5.2{:}1)$$

Thus the maximal model can be expressed as:

$$y \sim N(X\tau, \tfrac{1}{12}(\lambda_3 - \lambda_2) I_6 \otimes J_3 \otimes J_4 + \tfrac{1}{4}(\lambda_2 - \lambda_1) I_6 \otimes I_3 \otimes J_4 + \lambda_1 I_6 \otimes I_3 \otimes I_4)$$

where $\tau = (\tau_{11}, \tau_{12}, \ldots, \tau_{34})$ and $X = 1_6 \otimes I_3 \otimes I_4$.

Now apply Bayes' theorem:

$$p(\lambda,\tau|y) \propto p(y|\lambda,\tau)\, p(\lambda,\tau) \sim N(X\tau, V)\, p(\lambda,\tau)$$

where $p(\lambda,\tau)$ is the prior distribution for λ and τ. Throughout the following discussion, the prior distribution on λ and τ is assumed to factorise into a λ part and a τ part. Thus $p(\lambda,\tau) = p(\lambda)p(\tau)$. Two alternative forms for $p(\lambda)$ are considered, and an improper non-informative prior is taken for τ. Hence $p(\tau) = 1$.

Integrating out over τ gives the marginal posterior density for the eigenvalues λ, or equivalently for the variance components σ^2. This posterior density was derived in equation 2.2.1:1 of chapter 2 and is given below in equation 5.2:2. This is the three dimensional likelihood coded in the LOGLIK subroutine in the APPLES program (see Appendix).

$$p(\lambda|y) = \int_{\tau} l(\tau,\lambda|y)\, d\tau$$

$$\propto p(\lambda)|V|^{-\frac{1}{2}}|X'V^{-1}X|^{-\frac{1}{2}} \exp -\tfrac{1}{2}y'V^{-1}y \exp \tfrac{1}{2}y'V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}y \qquad (5.2{:}2)$$

5.2.1 Efficient evaluation of the marginal density for the eigenvalues

To evaluate equation 5.2:2 efficiently, algebraic expressions are required for $|V|$, $y'V^{-1}y$, $|X'V^{-1}X|$, and $y'V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}y$. These four quantities are evaluated in this section. Start with equation 5.2:1

$$V = \tfrac{1}{12}(\lambda_3-\lambda_2)I_6\otimes J_3\otimes J_4 + \tfrac{1}{4}(\lambda_2-\lambda_1)I_6\otimes I_3\otimes J_4 + \lambda_1 I_6\otimes I_3\otimes I_4$$

As V has been written in terms of its eigenvalues, its inverse and determinant can be written down directly. An alternative method † is provided by Searle and Henderson (1979). Using either method on the dispersion matrix V gives:

$$V^{-1} = \tfrac{1}{12}(\lambda_3^{-1}-\lambda_2^{-1})I_6\otimes J_3\otimes J_4 + \tfrac{1}{4}(\lambda_2^{-1}-\lambda_1^{-1})I_6\otimes I_3\otimes J_4 + \lambda_1^{-1}I_6\otimes I_3\otimes I_4$$

$$\text{and } |V| = \lambda_1^{6.3.(4-1)}\lambda_2^{6.(3-1)}\lambda_3^{6} = \lambda_1^{54}\lambda_2^{12}\lambda_3^{6} \qquad (5.2.1{:}1)$$

Now let $\beta = \frac{1}{12}(\lambda_3^{-1}-\lambda_2^{-1})$, $\gamma = \frac{1}{4}(\lambda_2^{-1}-\lambda_1^{-1})$, and $\delta = \lambda_1^{-1}$, then:

$$V^{-1} = \beta I_6\otimes J_3\otimes J_4 + \gamma I_6\otimes I_3\otimes J_4 + \delta I_6\otimes I_3\otimes I_4$$

$$\text{and } y'V^{-1}y = \beta\sum_{i} y_{i..}^2 + \gamma\sum_{i,j} y_{ij.}^2 + \delta\sum_{i,j,k} y_{ijk}^2 \qquad (5.2.1{:}2)$$

Recall that $X = 1_6\otimes I_3\otimes I_4$ which gives that:

$$V^{-1}X = 1_6\otimes(\beta J_3\otimes J_4 + \gamma I_3\otimes J_4 + \delta I_3\otimes I_4)$$

$$X'V^{-1}X = 6(\beta J_3\otimes J_4 + \gamma I_3\otimes J_4 + \delta I_3\otimes I_4)$$

Then using either of the matrix inversion techniques

$$(X'V^{-1}X)^{-1} = \tfrac{1}{6}\left\{\tfrac{1}{12}(\lambda_3-\lambda_2)J_3\otimes J_4 + \tfrac{1}{4}(\lambda_2-\lambda_1)I_3\otimes J_4 + \lambda_1 I_3\otimes I_4\right\}$$

$$|X'V^{-1}X| = 6\delta^{3.(4-1)}(\delta+4\gamma)^{(3-1)}(\delta+4\gamma+12\beta) = 6\lambda_1^{-3(4-1)}\lambda_2^{-(3-1)}\lambda_3^{-1} \qquad (5.2.1{:}3)$$

---

† Searle and Henderson produce the inverse and determinant of any balanced dispersion matrix. Applying their technique to the matrix A where A is given by:

$$A = pI_6\otimes J_3\otimes J_4 + qI_6\otimes I_3\otimes J_4 + rI_6\otimes I_3\otimes I_4$$

yields the inverse matrix A^{-1}:

$$A^{-1} = \tfrac{1}{12}\left\{\frac{1}{12p+4q+r} - \frac{1}{4q+r}\right\}I_6\otimes J_3\otimes J_4 + \tfrac{1}{4}\left\{\frac{1}{4q+r} - \frac{1}{r}\right\}I_6\otimes I_3\otimes J_4 + \frac{1}{r}I_6\otimes I_3\otimes I_4$$

$$|A| = r^{6.3.(4-1)}(r+4q)^{6.(3-1)}(r+4q+12p)^{6}$$

Multiplying out yields:

$$(X'V^{-1}X)^{-1}X'V^{-1} = \tfrac{1}{6}1'_6 \otimes I_3 \otimes I_4$$

$$V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1} = \tfrac{1}{6}J_6 \otimes (\beta J_3 \otimes J_4 + \gamma I_3 \otimes J_4 + \delta I_3 \otimes I_4)$$

$$y'V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}y = \tfrac{1}{6}\beta y^2_{...} + \tfrac{1}{6}\gamma\sum_j y^2_{.j.} + \tfrac{1}{6}\delta\sum_{j,k} y^2_{.jk} \qquad (5.2.1{:}4)$$

Combining equations 5.2.1:1 and 5.2.1:3 gives

$$|V||X'V^{-1}X| = 6\lambda_1^{(6-1).3.(4-1)}\lambda_2^{(6-1).(3-1)}\lambda_3^{6-1} \qquad (5.2.1{:}5)$$

$$\log|V||X'V^{-1}X| = constant + 45\log\lambda_1 + 10\log\lambda_2 + 5\log\lambda_3$$

recall equation 5.2:2

$$l(\lambda|y) \propto |V|^{-\frac{1}{2}}|X'V^{-1}X|^{-\frac{1}{2}}\exp-\tfrac{1}{2}y'V^{-1}y\exp\tfrac{1}{2}y'V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}y$$

Substituting from equations 5.2.1:2, 5.2.1:4 and 5.2.1:5

$$\log l(\lambda|y) = constant - \tfrac{1}{2}\{45\log\lambda_1 + 10\log\lambda_2 + 5\log\lambda_3\} \qquad (5.2.1{:}6)$$

$$-\tfrac{1}{2}(\beta\sum_i y^2_{i..} - \tfrac{1}{6}\beta y^2_{...} + \gamma\sum_{i,j} y^2_{ij.} - \tfrac{1}{6}\gamma\sum_j y^2_{.j.} + \delta\sum_{i,j,k} y^2_{ijk} - \tfrac{1}{6}\delta\sum_{j,k} y^2_{.jk})$$

5.2.2 Data Translation to simplify the log likelihood

Since the location parameters τ have been integrated out of equation 5.2.1:6, this likelihood is invariant under translations $y \longrightarrow y - Xa$, as these translations merely send $\tau \longrightarrow \tau + a$. The log likelihood is substantially simplified by translating by an amount $a = (\bar{y}_{.11}, \ldots, \bar{y}_{.34})'$ ($= \hat{\tau}$ say). Writing y^* for $y - X\hat{\tau}$, this yields:

$$\log l(\lambda|y^*) = constant - \tfrac{1}{2}\{45\log\lambda_1 + 10\log\lambda_2 + 5\log\lambda_3\}$$

$$- \tfrac{1}{2}\{\beta\sum_i y^{*2}_{i..} + \gamma\sum_{i,j} y^{*2}_{ij.} + \delta\sum_{i,j,k} y^{*2}_{ijk}\} \qquad (5.2.2{:}1)$$

where $\beta = \frac{1}{12}(\lambda_3^{-1} - \lambda_2^{-1})$, $\gamma = \frac{1}{4}(\lambda_2^{-1} - \lambda_1^{-1})$ and $\delta = \lambda_1^{-1}$.

An alternative formulation is:

$$\log l(\lambda|y^*) = constant - \tfrac{1}{2}(45\log\lambda_1 + 10\log\lambda_2 + 5\log\lambda_3)$$
$$- \tfrac{1}{2}(\lambda_3^{-1}Y_3^2 + \lambda_2^{-1}Y_2^2 + \lambda_1^{-1}Y_1^2) \qquad (5.2.2:2)$$

where $Y_3^2 = \frac{1}{12}\sum_i y_{i..}^{*2}$, $Y_2^2 = \frac{1}{4}(\sum_{i,j} y_{ij.}^{*2} - \frac{1}{3}\sum_i y_{i..}^{*2})$ and $Y_1^2 = \sum_{i,j,k} y_{ijk}^{*2} - \frac{1}{4}\sum_{i,j} y_{ij.}^{*2}$

Equation 5.2.2:2 provides the most efficient way of evaluating the likelihood for the eigenvalues, as the maximum amount of work is done once only (in the calculation of Y_1^2, Y_2^2 and Y_3^2), and the minimum amount of work is done at each point in λ space. This form of the likelihood is implemented in the APPLES program.

Note that after the data translation:

$$y^{*\prime}V^{-1}y^* = \lambda_3^{-1}Y_3^2 + \lambda_2^{-1}Y_2^2 + \lambda_1^{-1}Y_1^2$$
$$\text{and } y^{*\prime}V^{-1}X(X'V^{-1}X)'X'V^{-1}y^* = 0 \qquad (5.2.2:3)$$

5.3 Distributions conditional on the Variance Components

Special function analysis allows integrals of the form $\int f(\sigma^2)p(\sigma^2|y)\,d\sigma^2$ to be calculated. Integrals of this form allow the calculation of marginal densities and moments for the fixed effects given suitable choices for the function $f(\sigma^2)$. For example, the marginal density for τ_{11} is obtained when $f(\sigma^2)$ is $p(\tau_{11}|\sigma^2,y)$. These marginal distributions are developed in this section.

First, conditional on the variances, any fixed effect (location parameter) τ_{jk} is normally distributed:

$$\tau_{jk}|\sigma^2,y \sim N(\bar{y}_{.jk}, \tfrac{1}{6}(\sigma^2 + \sigma_a^2 + \sigma_b^2))$$

Similarly $\tau_{jk},\tau_{lm}|\sigma^2,y$ is distributed as a bivariate normal with common variance $\frac{1}{6}(\sigma^2 + \sigma_a^2 + \sigma_b^2)$ and a covariance of $\frac{1}{6}(\sigma_a^2 + \sigma_b^2)$ if $j = l$ or a covariance of $\frac{1}{6}\sigma_b^2$ if $j \neq l$.

The second moments for the τ are simply:

$$var(\tau_{jk}|\sigma^2) = \tfrac{1}{6}(\sigma^2 + \sigma_a^2 + \sigma_b^2)$$
$$corr(\tau_{jk},\tau_{lm}|\sigma^2) = \frac{\sigma_a^2 + \sigma_b^2}{\sigma^2 + \sigma_a^2 + \sigma_b^2} \qquad j = l$$

$$corr(\tau_{jk},\tau_{lm}|\sigma^2) = \frac{\sigma_b^2}{\sigma^2+\sigma_a^2+\sigma_b^2} \qquad j \neq l$$

Thus the following marginal densities and moments are calculated:

$$p(\tau_{jk}|y) = \int_{\sigma^2} N(\bar{y}_{.jk}, \tfrac{1}{6}(\sigma^2+\sigma_a^2+\sigma_b^2))\, p(\sigma^2|y)\, d\sigma^2 \qquad (5.3:1)$$

$$p(\tau_{jk},\tau_{lm}|y) = \int_{\sigma^2} N\left(\begin{bmatrix}\bar{y}_{.jk}\\ \bar{y}_{.lm}\end{bmatrix}, \Sigma\right) p(\sigma^2|y)\, d\sigma^2 \qquad (5.3:2)$$

where Σ has diagonal entries of $\frac{1}{6}(\sigma^2+\sigma_a^2+\sigma_b^2)$ and off-diagonal entries of $\frac{1}{6}(\sigma_a^2+\sigma_b^2)$ or $\frac{1}{6}\sigma_b^2$ for $i = l$ and $i \neq l$ respectively.

$$var(\tau_{jk}) \quad v = \int_{\sigma^2} \tfrac{1}{6}(\sigma^2+\sigma_a^2+\sigma_b^2)\, p(\sigma^2|y)\, d\sigma^2 \qquad (5.3:3)$$

$$corr(\tau_{jk},\tau_{jm}) \quad c = \int_{\sigma^2} \frac{\sigma_a^2+\sigma_b^2}{\sigma^2+\sigma_a^2+\sigma_b^2}\, p(\sigma^2|y)\, d\sigma^2 \qquad k \neq m \qquad (5.3:4)$$

$$corr(\tau_{jk},\tau_{lm}) \quad d = \int_{\sigma^2} \frac{\sigma_b^2}{\sigma^2+\sigma_a^2+\sigma_b^2}\, p(\sigma^2|y)\, d\sigma^2 \qquad j \neq l \qquad (5.3:5)$$

Equations 5.3:1 and 5.3:2 are coded in the APPLES program using the special function analysis routines of the BAYES4 integration package. A single *special function* may be used to give a single spot height for either $p(\tau_{jk}|y)$ or $p(\tau_{jk},\tau_{lm}|y)$ using equation 5.3:1 or 5.3:2. A lattice of spot height thus calculated can give these two marginal distributions. Similarly the variance of and correlations between the τ can be computed as integrals over σ^2 using equations 5.3:3, 5.3:4 and 5.3:5.

5.4 Joint Posterior Density for the variance components and a fixed effect

The previous formulation produced summaries of λ or σ^2 using the marginal posterior density, and summaries of τ using *special function analysis*. Preceding sections have produced univariate and bivariate distributions for both λ (or σ^2) and τ. However, the previous formulation does not allow the calculation of the joint posterior density of σ^2 and a fixed effect. If it is required to investigate σ^2 and a fixed effect jointly, then a joint likelihood for σ^2 and the fixed effect must be coded in the program. This is achieved most easily by writing this posterior density as a product and recalling that, conditional on the variance components, τ_{11} is normally

distributed with variance $\frac{1}{6}(\sigma^2+\sigma_a^2+\sigma_b^2)$.

Substitute from equation 5.2.2:2 for $p(\lambda|y)$ to yield:

$$p(\lambda,\tau_1\,|\,y) \propto (\lambda_1^{45}\lambda_2^{10}\lambda_3^{5})^{-\frac{1}{2}}(\sigma^2+\sigma_a^2+\sigma_b^2)^{-\frac{1}{2}}$$
$$\times \exp-\tfrac{1}{2}\{\lambda_3^{-1}Y_3^2+\lambda_2^{-1}Y_2^2+\lambda_1^{-1}Y_1^2+6\tau_{11}^2/(\sigma^2+\sigma_a^2+\sigma_b^2)\} \qquad (5.4:1)$$

This joint distribution was straight forward to obtain because of the simple form of the distribution of $\tau_{11}|\lambda,y$.

5.5 Alternatives to a Normal error structure

It is argued in Chapter 2 that the general multivariate t distribution is an alternative to the multivariate normal error distribution. In this example, a distinctly heavy-tailed alternative to normality was sought and thus the degrees of freedom, υ, was chosen equal to 5. The parameterisation of equation 2.4:2 is used so that $\boldsymbol{V}$ has the same interpretation under both error distributions. It is noted that adopting this distribution implies that the y_{ijk}'s are no longer assumed to be independent given τ. (Recall the discussion in Chapter 2 section 2.4). For a further discussion of alternatives to normality see section 6.4 in Chapter 6.

5.5.1 Marginal Posterior Density for the Variance Components with t errors

Equation 2.4:2 gives the joint posterior distribution for τ and σ^2. To obtain a marginal posterior density for σ^2, the τ must be integrated out.

now $(y-X\tau)'V^{-1}(y-X\tau) = \tau'X'V^{-1}X\tau - 2\tau'X'V^{-1}y + y'V^{-1}y$

Denote $X'V^{-1}X$ by A^2, and $X'V^{-1}y$ by AB, where B is a vector. Also let $\tau^\dagger = \tau A$, then:

$$(y-X\tau)'V^{-1}(y-X\tau) = \tau^{\dagger\prime}\tau^\dagger - 2\tau^{\dagger\prime}B + y'V^{-1}y$$
$$= (\tau^\dagger - B)^2 + C$$

where $C = y'V^{-1}y - B'B$

$$= y'V^{-1}y - y'V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}y$$

Proceed with the integration to remove τ - recall equation 2.4:2

$$p(\sigma^2 \mid y) = \int_{\tau} p(\sigma^2,\tau \mid y)\, p(\sigma^2)\, d\tau$$

$$\propto p(\sigma^2)\,|V|^{-\frac{1}{2}} \int_{\tau} \left[1 + \frac{(\tau^{\dagger} - B)^2 + C}{\nu-2}\right]^{-\frac{\nu+n}{2}} d\tau$$

change from τ space to $\tau^{\dagger}$ and remember that $d\tau = |A|^{-1}\, d\tau^{\dagger}$.

$$\propto p(\sigma^2)\,|V|^{-\frac{1}{2}} |X'V^{-1}X|^{-\frac{1}{2}} \int_{\tau^{\dagger}} \left[1 + \frac{(\tau^{\dagger} - B)^2 + C}{\nu-2}\right]^{-\frac{\nu+n}{2}} d\tau^{\dagger}$$

$$\propto p(\sigma^2)\,|V|^{-\frac{1}{2}} |X'V^{-1}X|^{-\frac{1}{2}} \left[1 + \frac{C}{\nu-2}\right]^{-\frac{\nu+n}{2}}$$

$$\times \int_{\tau^{\dagger}} \left[1 + \frac{(\tau^{\dagger} - B)^2/(1 + \frac{C}{\nu-2})}{\nu-2}\right]^{-\frac{\nu+n}{2}} d\tau^{\dagger}$$

The integrand on the right looks like a multivariate t distribution with dispersion matrix $(1 + \frac{C}{\nu-2})I$ and mean B. Hence:

$$p(\sigma^2 \mid y) \propto p(\sigma^2)\,|V|^{-\frac{1}{2}} |X'V^{-1}X|^{-\frac{1}{2}} (1 + \frac{C}{\nu-2})^{-\frac{1}{2}(\nu+n-12)}$$

$$\propto p(\sigma^2)\,|V|^{-\frac{1}{2}} |X'V^{-1}X|^{-\frac{1}{2}} \left[1 + \frac{y'V^{-1}y - y'V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}y}{\nu-2}\right]^{-\frac{\nu+n-12}{2}}$$

Using the data translation of Section 5.2.2 and substituting from equations 5.2.1:5 and 5.2.2:3 yields:

$$p(\sigma^2 \mid y) \propto p(\sigma^2)(\lambda_1^{45}\lambda_2^{10}\lambda_3^{5})^{-\frac{1}{2}} \left[1 + \frac{\lambda_3^{-1}Y_3^2 + \lambda_2^{-1}Y_2^2 + \lambda_1^{-1}Y_1^2}{\nu-2}\right]^{-\frac{\nu+n-12}{2}} \qquad (5.5.1{:}1)$$

Thus equation 5.5.1:1 gives the marginal density for σ^2 as required. As $\nu \longrightarrow \infty$, then the $(1 + C/\nu\text{-}2)^{-\frac{1}{2}(\nu+n-12)}$ term tends to $\exp\text{-}\frac{1}{2}C$, and the marginal posterior density converges to:

$$p(\sigma^2 \mid y) \propto p(\sigma^2)(\lambda_1^{45}\lambda_2^{10}\lambda_3^{5})^{-\frac{1}{2}} \exp(-\tfrac{1}{2}C)$$

$$\propto p(\sigma^2)(\lambda_1^{45}\lambda_2^{10}\lambda_3^{5})^{-\frac{1}{2}} \exp\text{-}\tfrac{1}{2}(\lambda_3^{-1}Y_3^2 + \lambda_2^{-1}Y_2^2 + \lambda_1^{-1}Y_1^2)$$

As expected, this is exactly the same as the posterior obtained from a Normal distribution in equation 5.2.2:2.

5.5.2 Marginal Posterior Density for a fixed effect and the variance components with general multivariate t errors

In order to calculate a marginal distribution on a single fixed effect (for example τ_{11}), it is necessary to derive an expression for $p(\tau_{11}|\sigma^2,y)$. As derived in Chapter 2, the distribution the the fixed effects τ conditional on the variance components σ^2 and the data y is multivariate t with:

Mean	$(X'V^{-1}X)^{-1}X'V^{-1}y$
Dispersion matrix	$(X'V^{-1}X)^{-1}(\nu-2+y'V^{-1}y-y'V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}y)/(\nu+n-12-2)$
Degrees of freedom	$\nu+n-12$

Now integrate out 11 of the τ_{jk} to leave $p(\tau_{11}|\sigma^2,y)$. As shown by in section 2.4.3 of Chapter 2, this yields a t distribution, with the same number of degrees of freedom, and with a mean (and variance) formed by deleting 11 rows (and columns) from the 12 dimensional values. Also doing the data translation of Section 5.2.2, $y \longrightarrow y^* = y - X\hat{\tau}$ takes the mean to zero, and simplifies the variance to $\frac{1}{6}(\sigma^2+\sigma_a^2+\sigma_b^2)\ (\nu-2+y^{*\prime}V^{-1}y^*)/(\nu+n-12-2)$. Hence:

$$p(\tau_{11}^*|\sigma^2,y^*) \propto \left[1+\frac{6\tau_{11}^{*2}(\nu+n-12-2)/(\sigma^2+\sigma_a^2+\sigma_b^2)(\nu-2+y^{*\prime}V^{-1}y^*)}{\nu+n-12-2}\right]^{-\frac{\nu+n-12+1}{2}}$$

$$\propto \left[1+\frac{6\tau_{11}^{*2}/(\sigma^2+\sigma_a^2+\sigma_b^2)}{\nu-2+y^{*\prime}V^{-1}y^*}\right]^{-\frac{\nu+n-11}{2}} \qquad (5.5.2{:}1)$$

where $\tau_{11}^* = \tau_{11} - \hat{\tau}_{11}$.

5.5.3 Marginal Posterior Density for $\tau_{jk}, \tau_{lm}, \sigma^2$ with general multivariate t errors

Instead of integrating out 11 of the τ_{jk} as done in the calculation of equation 5.5.2:1, only 10 of the τ_{jk} are integrated out. This leaves the joint posterior for two fixed effects conditional on the eigenvalues. Denote the fixed effects by τ_{jk} and τ_{lm}. Let v be the variance of the τ_{jk}^* and c be the correlation between τ_{jk}^* and τ_{lm}^*. Thus v is $\frac{1}{6}(\sigma^2+\sigma_a^2+\sigma_b^2)$, and c is $\frac{1}{6}(\sigma_a^2+\sigma_b^2)$ if $l=l$ or c is $\frac{1}{6}\sigma_b^2$ if $l \neq l$:

$$p(\tau^*_{jk},\tau^*_{lm}|\sigma^2,y^*) \propto \left[1+\frac{(v\tau^{*2}_{jk}+v\tau^{*2}_{lm}-2c\tau^*_{jk}\tau^*_{lm})/(v^2-c^2)}{\nu-2+y^{*\prime}V^{-1}y^*}\right]^{-\frac{\nu+n-10}{2}} \qquad (5.5.3:1)$$

Equation 5.5.3:1 is again implemented in the special function analysis code of the APPLES program.

5.6 The analyses performed, and their sensitivity to the assumptions

One of the strengths of the Bayesian philosophy is the ease with which different assumptions can be accommodated into the model. Full analyses can then be performed under a set of differing assumptions, enabling the effects of the assumptions to be seen. For the sake of illustration, assume that in this example there are three summaries of interest:

i) The marginal posterior density $p(\tau_{jk}|y)$, as a basis for inference about the mean yield, τ_{jk}, of a particular combination of irrigation j and thinning treatment k. Also the bivariate marginal $p(\tau_{jk},\tau_{lm}|y)$ may be of interest whenever τ_{jk} and τ_{lm} are not independent.

ii) The predictive distribution

$$p(z|y) = \int p(z|\tau_{jk},\lambda)\,p(\tau_{jk},\lambda|y)\;d\tau_{jk}\,d\lambda$$

for a future observation z from the combination of irrigation i and thinning treatment j.

iii) The joint posterior density $p(\lambda|y)$, which may be useful in the analysis of some future experiment involving identical experimental material but possibly different treatments.

The sensitivity of each of these summaries is investigated with respect to the choice of linear structure (maximal vs main effects) for the treatment effects, the choice of error distribution and the choice of prior distribution on λ.

The predictive distribution in ii) can easily be implemented in BAYES4 using special function analysis, as it is an integral across parameter space.

5.6.1 Prior Distributions

The accepted non-informative reference prior for λ under both the normal and the multivariate t formulation is $\prod\lambda_i^{-1}$ (see Chapter 4). As an example of a possible alternative consider the informative prior:

$$l(\lambda) \propto \prod_{i=1}^{3} \lambda_i^{-\frac{1}{2}(\nu_i+2)} e^{-\frac{1}{2}\nu_i m_i/\lambda_i} \qquad (5.6.1:1)$$

where $\nu_1 = 43$, $m_1 = 6933$, $\nu_2 = 8$, $m_2 = 22016$, $\nu_3 = 3$ and $m_3 = 27696$. The functional form is a product of inverse χ^2 distributions, and the values are based on the data given by Knuiman for the year 1977 (see Table 5.1.2:2).

Using the Monte-Carlo integration technique, there is no need for analytic integration, and a proper prior could be assigned to τ. Alternatively τ and λ need not be assumed independent.

5.7 Results

It should be remembered that there is far more information provided by the data for λ_1 than for λ_2, and more information for λ_2 than for λ_3. This is because λ_1 has 45 degrees of freedom, λ_2 has 10 degrees of freedom and λ_3 has only 5 degrees of freedom. The presentation of the results is split into three sections, the first dealing with the fixed effects, the second with the eigenvalues and the final section dealing with the variance components. All the bivariate plots show contours at 1%, 5%, 10%, 30%, 50%, 70% and 90% of the height of the mode.

5.7.1 The fixed effects

The mean values for the estimates of the fixed effects clearly do not depend on the particular choice of error distribution or prior. (For the maximal model they are simply $\bar{y}_{.jk}$). The fixed effect means for the maximal and main effects only models are given in Tables 5.7.1:1 and 5.7.1:2.

Cell means for the 12 fixed effects			
291·0	326·8	316·2	351·2
329·2	293·8	362·8	421·2
381·0	430·5	455·9	439·5

Table 5.7.1:1

Cell means for the main effects model			
288·4	305·1	333·0	358·7
318·8	335·5	363·4	389·1
393·8	410·5	438·4	464·1

Table 5.7.1:2

The global mean of the data values is 366·6. The decomposition of the 12 fixed effect into row, column and interaction effects in given in Table 5.7.1:3

Classical Parameterisation for the fixed effects				
Row Effects	−45·3	−14·9	60·1	
Column Effects	−32·9	−16·2	11·7	37·4
Interaction Effects	2·6	21·7	−16·8	−7·5
	10·4	−41·7	−0·6	32·1
	−12·8	20·0	17·5	−24·6

Table 5.7.1:3

Figure 5.7.1:1 displays $l(\tau_{12}|y)$ for various choices of model, error distribution and prior distribution. Most noticeable is the dependence of the location of the margin on the inclusion or exclusion of interaction terms in the model. The magnitude of this displacement varies, of course, with the choice of τ_{jk}. The

maximum displacement observed is 41·7 with τ_{22}, the minimum 0·6 with τ_{23}. For τ_{12} the displacement is 21·7. The decision to display the margin for τ_{12} rather than for any other treatment parameter reflects a desire to demonstrate the effect of choice of linear structure without selecting an extreme instance. It can be seen that the displacement for τ_{11} is only 2·6. Clearly, if τ_{11} is the *only* location parameter of interest then sensitivity to choice of linear structure would not be an issue. As might be expected, the choice of error distribution has little effect. The informative prior for σ^2 does have the effect of tightening the margin slightly when the main-effects model is adopted.

Figure 5.7.1:2 shows the distribution of a future observation from cell (1,1) for three choices of error distribution and prior distribution. The effect of choice of linear structure is not displayed but is identical to that demonstrated in figure 5.7.1:1. Here one might anticipate rather more sensitivity to the choice of normal or *t* errors. However, the practical consequences are slight unless one is interested in probabilities obtained from the extreme tails of this distribution.

Figures 5.7.1:3 and 5.7.1:4 shows two bivariate plots under the three sets of conditions considered. The two plots are for τ_{11} against τ_{12} and τ_{11} against τ_{21}. All bivariate τ plots that use elements of τ with the same irrigation regime are, subject to a translation, identical. Similarly, all τ plots with different irrigation regimes are, subject to a translation, identical. Thus the τ_{11} against τ_{12} and τ_{11} against τ_{21} plots are representative of all bivariate τ plots. It is noted that the change of error distribution makes almost no effect, and the informative prior merely tightens the distributions a little. The results are summarised in table 5.7.1:4 below:

Moments of the Fixed Effects				
Error Distribution	Prior Distribution	*variance* τ_{jk}	*correlation* (τ_{jk},τ_{jm})	(τ_{jk},τ_{lm})
Normal	Jeffrey s'	1830	0·490	−0·027
Normal	Inverse χ^2	1520	0·354	−0·036
t, ν=5	Jeffrey s'	1824	0·490	−0·027

Table 5.7.1:4

5.7.2 The eigenvalues

The first and second moments for the eigenvalues are tabulated below. Following these are 3 bivariate contour plots showing (λ_1, λ_2), (λ_1, λ_3) and (λ_2, λ_3) and three univariate marginal distributions for the three combinations of prior and error distribution. Tables 5.7.2:1, 5.7.2:2 and 5.7.2:3 below give the posterior moments of $l(\lambda|y)$ for the three combinations with the maximal model. These have been calculated by GR rather than BAYES4, as BAYES4 calculates moments for the parameters, which in this case are the log eigenvalues. Table 5.7.2:4 gives these moments for $\log l(\lambda|y)$ with Normal errors and a Jeffrey's prior. These have come directly from BAYES4. Figures 5.7.2:1, 5.7.2:2 and 5.7.2:3 give the bivariate plots. The ranges of these plots are: λ_1 0 to 25000, λ_2 0 to 150000 and λ_3 0 to 130000.

Moments of $l(\lambda|y)$

	Normal errors		**Jeffrey's prior**		
			correlation		
	mean	*variance*	λ_3	λ_2	λ_1
λ_3	26500	8·85e8		0·000	0·000
λ_2	28900	2·71e8	0·000		0·000
λ_1	5230	1·33e6	0·000	0·000	

Table 5.7.2:1

It is to be expected that the correlations in Table 5.7.2:1 are all zero since equation 5.2.2:2 for the likelihood on the λ factorises, and the uniform prior on the log λ also factorises. Thus under these assumptions the λ_i are mutually independent.

Normal errors χ^{-2} prior					
			correlation		
	mean	*variance*	λ_3	λ_2	λ_1
λ_3	16300	6·59*e*7		0·000	0·000
λ_2	20400	4·62*e*7	0·000		0·000
λ_1	5810	7·68*e*5	0·000	0·000	

Table 5.7.2:2

Again, it is to be expected that the correlations in Table 5.7.2:2 are all zero since the likelihood on the λ factorises, and the informative prior given in equation 5.6:1 also factorises. Thus under these assumptions the λ_i are mutually independent.

t errors, ν=5 Jeffrey s' prior					
			correlation		
	mean	*variance*	λ_3	λ_2	λ_1
λ_3	44500	4·95*e*8		0·272	0·371
λ_2	48400	2·03*e*9	0·272		0·627
λ_1	8730	3·51*e*7	0·371	0·627	

Table 5.7.2:3

These correlations in Table 5.7.2:3 are not surprising as the spread of plausible values for the variances σ^2 and σ_a^2 are increased dramatically with the *t* errors. As the eigenvalues are linear combinations of σ^2, σ_a^2 and σ_b^2, a large spread in the estimation of the variances will induce correlation between the eigenvalues.

Moments of the log $l(\lambda|y)$

Normal errors Jeffrey's prior					
			correlation		
	mean	*variance*	log λ_3	log λ_2	log λ_1
log λ_3	9·89	0·698		0·000	0·000
log λ_2	10·15	0·470	0·000		0·000
log λ_1	8·54	0·213	0·000	0·000	

Table 5.7.2:4

The moments in Table 5.7.2:4 are the moments which BAYES4 produces, that is the moments for the parameters. Because of the log transform, they are of little use in this particular example. However, these moments feature in a discussion of the stability of the first and second moments, deferred to section 5.9 (see Table 5.9.1:1), but are displayed here for completeness.

Now consider the effect of the informative inverse Chi-squared prior for λ. As this prior is based on as much information as is contained in the likelihood through the data, it predictably tightens the joint density considerably. This is seen by the reduction in the variances for the λ. The variance for λ_1 reduces by 42%, the variance for λ_2 by 83%, and the variance of λ_3 by 92%. From figures 5.7.2:2 and 5.7.2:3, it can be seen that the reason for the variance of λ_1 not shrinking as much as the variances for λ_2 and λ_3 is due to the differing locations for the mode of λ_1 in the prior and the likelihood.

The sensitivity of the correlation structure to changes in the error distribution is dramatic. The means of λ_1, λ_2 and λ_3 increase by 50%, 67% and 68%, yet figures 5.7.2:1, 5.7.2:2 and 5.7.2:3 show that all the modes shrink. At the same time the variances increase by factors of 26·4, 7·5 and 5·6, indicating considerably increased uncertainty about the values for λ. This is because outlying data values can be explained either by having high variances and hence high λ, or alternatively, outlying data values can be explained by the heavy tails implicit in the t distribution. The large correlations are to be expected because the variance of $\lambda_1 = \sigma^2$ has increased so much, and this term appears as a constituent of λ_2 and λ_3.

It should be noted that all three analyses give a lower mean value for λ_2 than for λ_3. Classically this would give rise to a negative estimate for σ_b^2, and suggests that there is a significant probability that σ_b^2 is zero, or that negative correlations exist between observations. However with only 5 degrees of freedom for λ_3 it was not possible to investigate whether σ_b^2 was really zero.

5.7.3 The Variance Components

The σ^2 and σ_a^2 variance components can be reconstructed from the eigenvalues by taking the bivariate plot of λ_1 against λ_2 and shearing it. From this bivariate plot, the two univariate distributions can be obtained together with the first and second moments. These moments are given below in Table 5.7.3:1 for the three combinations of error distribution and prior considered. Figure 5.7.3:1 gives the corresponding bivariate and univariate plots.

As is to be expected from the eigenvalues, the effect of the informative inverse chi-squared prior is to tighten the distribution for the σ_a^2 and σ^2. The mean for σ^2 rises because the prior has a substantially larger mode than the likelihood, and the mean for σ_a^2 falls because the right tail of the distribution collapses. Both second moments shrink in line with the tighter posterior distribution. Curiously the correlation becomes $-0{\cdot}128$.

The multivariate t distribution with 5 degrees of freedom has a dramatic effect on the correlation, the mean and the variance of the posterior. From being essentially uncorrelated, the correlation between σ_a^2 and σ^2 becomes $0{\cdot}536$, and the means increase by 83% and 67%. The variances also increase by factors of $6{\cdot}3$ and 26.

In all three cases the modal values change relatively little compared with the means.

Moments of Variance Components						
Error Distribution	Prior Distribution	mean		variance		correlation
		σ_a^2	σ^2	σ_a^2	σ^2	
Normal	Jeffrey s'	5390	5230	1·72e7	1·34e6	−0·070
Normal	Inverse Chi-Squared	3640	5810	2·93e6	7·68e5	−0·128
t, ν=5	Jeffrey s'	9900	8730	1·08e8	3·57e7	0·536

Table 5.7.3:1

5.8 Comparison of results with those of Knuiman

Knuiman provides an analysis of the both the 1975 and 1977 data sets assuming Normal errors and a non-informative prior distribution. These can be compared with the Bayesian results obtained with Normal errors and a non-informative prior.

The estimates for the 12 fixed effects are identical. This is to be expected as the mean estimates correspond to the maximum likelihood (modal) estimates.

The estimates for the eigenvalues differ substantially. Knuiman estimates λ_1, λ_2 and λ_3 as 4997, 23138 and 15997 which compare with the Bayesian estimates of 5230, 28900 and 26500. The Bayesian estimates are significantly larger since they represent mean rather rather modal estimates and the univariate marginal distributions on the eigenvalues have a marked right skew (see Figure 5.7.2:8).

Knuiman was principally concerned with approximate updating of modal estimates from a sequence of identical experiments. No attempt is made here to judge the validity of these approximations. Rather, it is wished to emphasise the worth to the experimenter of a marginal distribution rather than a point estimate.

5.9 Performance of the integration routines

This section looks critically at the results obtained from running APPLES and MAPPLES using BAYES4 with a view to learning about the behaviour of the integration techniques rather than learning about growing apples in Australia.

5.9.1 Performance of the Gauss-Hermite integration rule in the APPLES program

In this section the reliability of the moments of the log eigenvalues is considered for different sizes of Gauss-Hermite grid, along with the cpu times for those grids. The results presented below are for Normal errors, and a Jeffrey's prior distribution. Since these conditions give independence between the λ, only the means and variances for the log λ need be presented. For each size of grid, sufficient iterations were performed for the moments to stabilise, and the results are given in Table 5.9.1:1 below. Identical behaviour was observed when Normal errors and an informative prior were used. Also t distribution errors stabilised in the same manner, but the correlations between the λ were non zero (see Table 5.7.2:3). To cater for this a linear transformation was introduced to orthogonalise the log λ.

Moments of log eigenvalues for different grid sizes					
size of grid	cpu seconds	log *eigenvalues* λ_1	λ_2	λ_3	
2×2×2	0·09	8·52404	10·0833	9·74967	*mean*
		0·21470	0·45351	0·65054	*variance*
3×3×3	0·09	8·53897	10·1523	9·89267	*mean*
		0·21200	0·45871	0·66592	*variance*
4×4×4	0·16	8·53887	10·1503	9·88434	*mean*
		0·21160	0·45481	0·65440	*variance*
5×5×5	0·22	8·53880	10·1494	9·88350	*mean*
		0·21320	0·47114	0·70238	*variance*
6×6×6	0·28	8·53899	10·1530	9·89572	*mean*
		0·21312	0·46809	0·68830	*variance*
7×7×7	0·44	8·53898	10·1520	9·88915	*mean*
		0·21315	0·46955	0·69747	*variance*
8×8×8	0·57	8·53898	10·1524	9·89318	*mean*
		0·21318	0·47035	0·69873	*variance*

Table 5.9.1:1

Note that because the roots or zeroes of an n^{th} order Gauss-Hermite polynomial interleave the roots of an $n+1^{th}$ order Gauss-Hermite polynomial, changing from a grid of n points to a grid with $n+1$ points is a good test for convergence of the moments. Such a change in grid guarantees that substantially different points are used in the two grids. If this has only a slight effect on the moments, then the moments are robust.

Note also 6 significant figures are quoted for the means, and 5 for the variances. It is not intended to suggest that the values have been obtained that precisely, but quoting all the figures does enable the degree of convergence to be seen. It can be seen that a precision of 3 figures for the mean of λ_1 and 4 figures for the mean of λ_2 can be obtained from only a 3×3×3 grid, but that it takes a 7×7×7 grid to achieve the same precision for λ_3. This is largely because the skewness of λ_3 is greater than the other two skewnesses, and λ_3 also requires a larger grid to

stabilise. This is to be expected as λ_3 has far fewer degrees of freedom than λ_2 or λ_1.

5.9.2 Execution times for the APPLES program

Ultimately the applicability or non-applicability of numerical methods depend on the amount of computer time used. There are several important quantities for the APPLES program.

i) The number of iterations required to gain stability for the moments of the log eigenvalues.

ii) The time taken to perform a single iteration of the 3 dimensional likelihood for the log λ.

iii) Once the moments for the eigenvalues have converged, a single further iteration is required with *special function analysis* switched on. The *special function analysis* code calculates the three conditional distributions for the fixed effects conditional on the eigenvalues, and the time that this code takes is important.

Clearly the times in ii) and iii) are functions of the number of points used for the integration rules, and the number of iterations in i) depends inversely on the number of points in ii) and also on how good the initial estimates of the moments were. In practice it is found that although iii) is only done once, the time taken in iii) substantially dominates the other times. This is because the *special function analysis* code evaluates two 2 dimensional grids of points and a 1 dimensional vector of points at every point in the 3 dimensional log eigenvalue grid of points.

It is thus sensible to iterate on a fairly large grid for ii) until stability is reached, then do iii). Working on a final grid of $8\times8\times8$ points takes less than 0·6 seconds per iteration on a Vax 11/785 computer. Convergence is fairly rapid and even very poor estimates for the moments converge in 6-10 iterations (most of which are on a smaller grid than $8\times8\times8$). Obtaining an 8×8 array of spot heights for the bivariate marginal distributions for the fixed effects and 8 points on the univariate distribution takes about 4·6 seconds (of which 0·6 is the time taken for the likelihood on the variance components).

5.9.3 The Monte-Carlo integration rule in the MAPPLES program

Using the Monte-Carlo rule there is no distinction between fixed effects and log eigenvalues - there are simply 15 dimensions with a complicated dispersion matrix. It must be remembered that each iteration of BAYES4 updates the first and second moments of the parameters, and it is usual to continue iterating until these moments have stabilised. With MAPPLES there are 15 means, 15 variances and 105 correlations, thus giving 135 values that have to stabilise. For a Monte-Carlo procedure to estimate 135 values in a stable fashion clearly requires a large number of points, and also an efficient importance sampling algorithm for placing those points. Clearly stability can always be gained by taking enough points, but this may take a prohibitive amount of time.

Once all 135 moments have stabilised, the integration must be broken up into 1 or 2 dimensions for a Gauss-Hermite grid, and 14 or 13 dimensions for Monte-Carlo integration. Marginal distributions can be obtained for the 1 or 2 dimensions tackled via Gauss-Hermite.

BAYES4 provides two different measures of stability when using Monte-Carlo integration. First, internal to BAYES4, several Monte-Carlo integrations are performed and the results merged, rather than doing a single integration. This technique achieves greater stability as problems with ridges in the likelihood are reduced, and also provides an *Estimated Error* based on the difference between the internal estimates. The second measure compares the moments from this iteration with the moments from the previous iteration. Changes in the means are reported as the percentage change in standard deviations, changes in variance are reported as percentage fractional changes, and changes in correlation are reported as absolute changes. These 135 changes are combined to give a single *Normalised Error*.

5.9.4 Execution times for the MAPPLES program

A summary of the execution times and stabilities for given numbers of points are listed below. These are all for the 15 dimensional Monte-Carlo integration. It has been found that when calculating marginal distributions for some parameters, the number of points needed for the Monte-Carlo integration may be reduced by a large factor (say 10) without affecting the marginal (though poor estimates will be

made of the other moments).

Performance of the Monte-Carlo Integration rule			
Number of points	cpu time (seconds)	Estimated Error %	Normalised Error %
5000	35·3	0·95	0·10
20000	138·0	0·55	0·15

Table 5.9.4:1

Running on 5000 points, the following level of stability is observed for the moments of the 15 parameters:

i) The 12 fixed effect means change on average by 0·02 on each iteration, or about $\frac{1}{20}$% of a standard deviation. Similarly the means of the log eigenvalues change by about 0·0002 or (again) $\frac{1}{20}$% of a standard deviation.

ii) The standard deviations of the fixed effects change by about 0·012 on each iteration or 0·3 % of a standard deviation, and the standard deviations of the log eigenvalues also vary by 0·3 %.

iii) Finally the correlations typically change by about 0·05 %.

It can be seen that from the point of view of Monte-Carlo integration within BAYES4, there is no difference between the 12 fixed effects and the 3 log eigenvalues. They are simply 15 parameters, and are all estimated with the same level of precision.

Finally compare the two measures of stability given in Table 5.9.4:1. It can be seen that changing from 5000 points to 20000 effectively halves the estimated error. This is the within iteration estimate of error, and it is to be expected that quadrupling the number of points will halve it. Rather surprisingly the normalised error (between iteration error) is higher using 20000 points than with 5000. This is probably an illusion caused by doing the iterations on 20000 points before the iterations on 5000 points. It is to be expected that given more iterations at 20000 points, its normalised error would become less than that for 5000 points.

Figure 5.7.1:1 Marginal τ_{12} for both interaction and no interaction model under three sets of assumptions

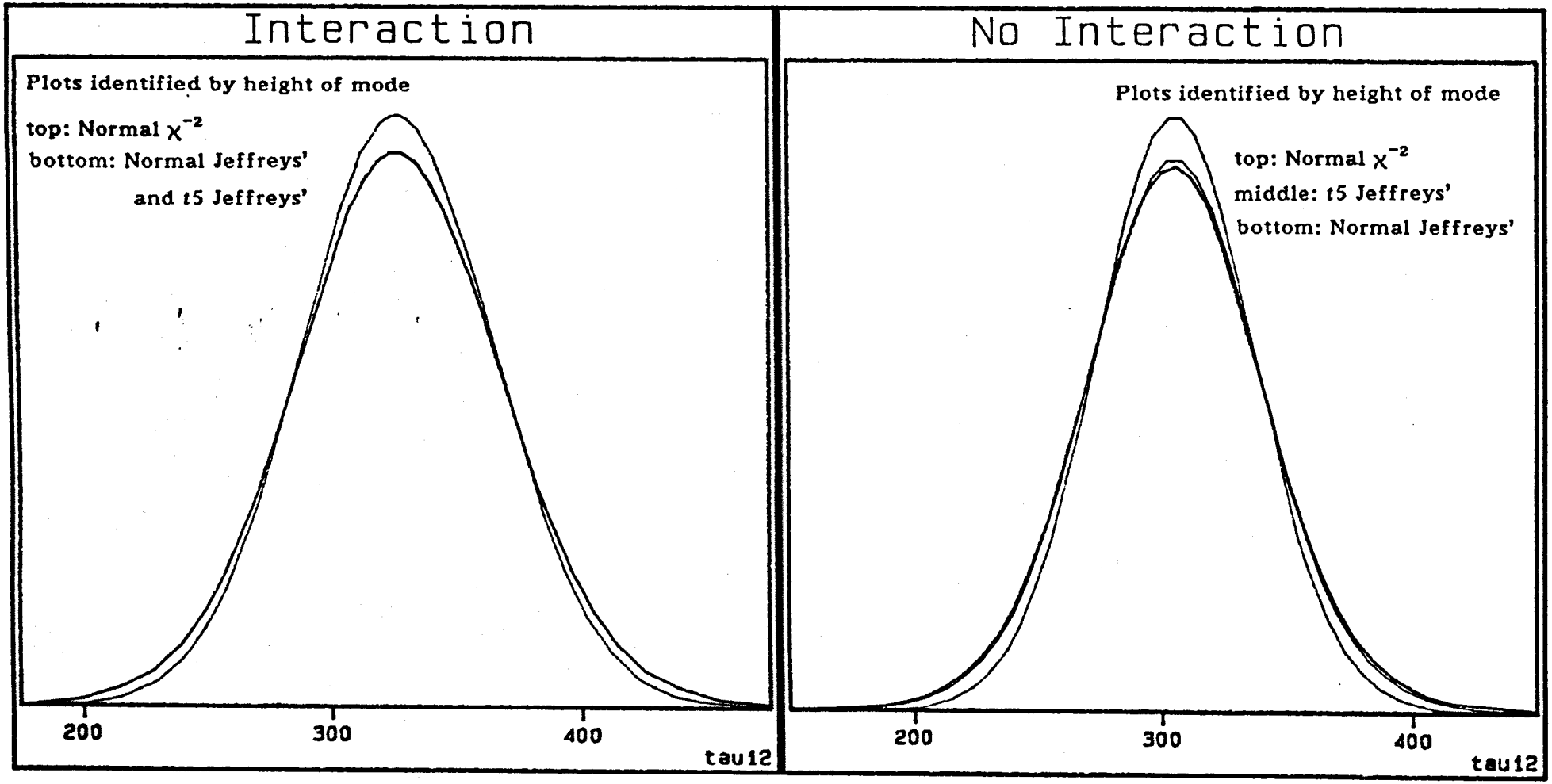
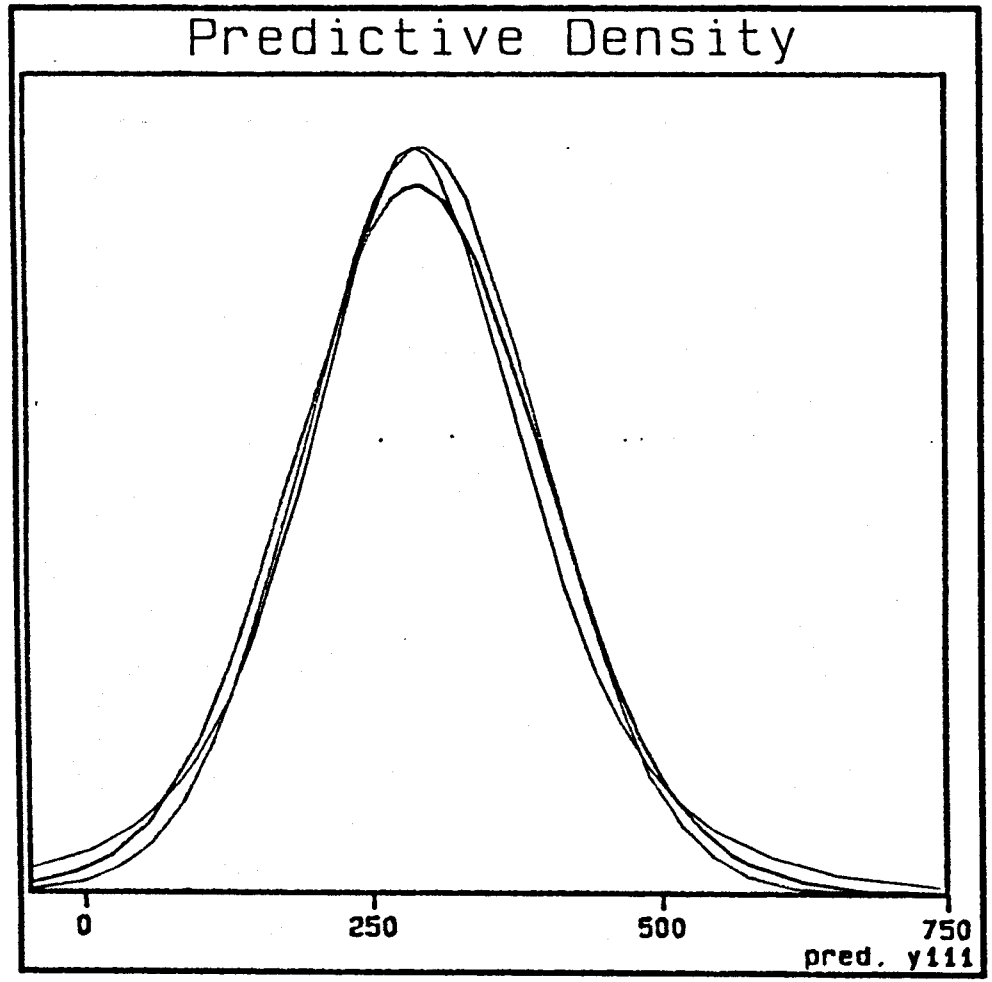


Figure 5.7.1:2 Marginal Density for a future observation from cell 1,1 under three sets of assumptions



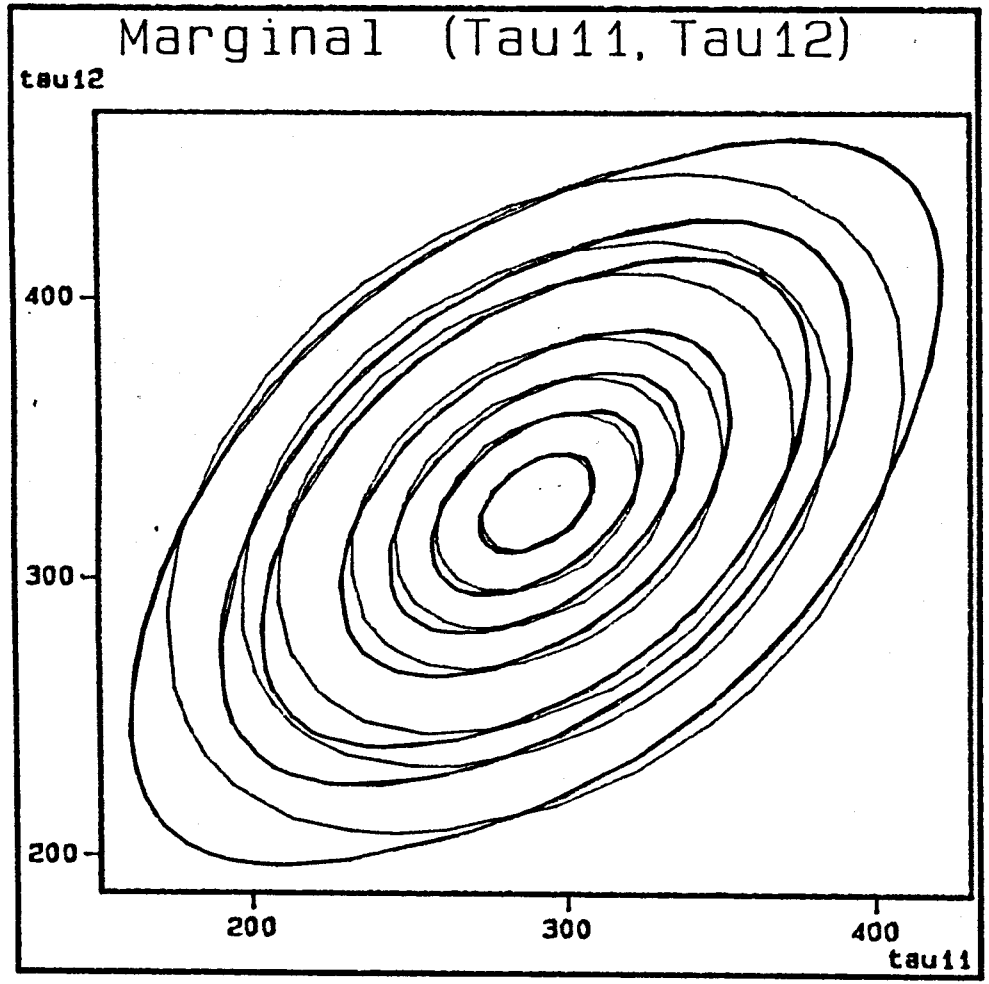
Plots identified by height of mode

top right: Normal χ^{-2}

top left: $t5$ Jeffreys'

bottom: Normal Jeffreys'

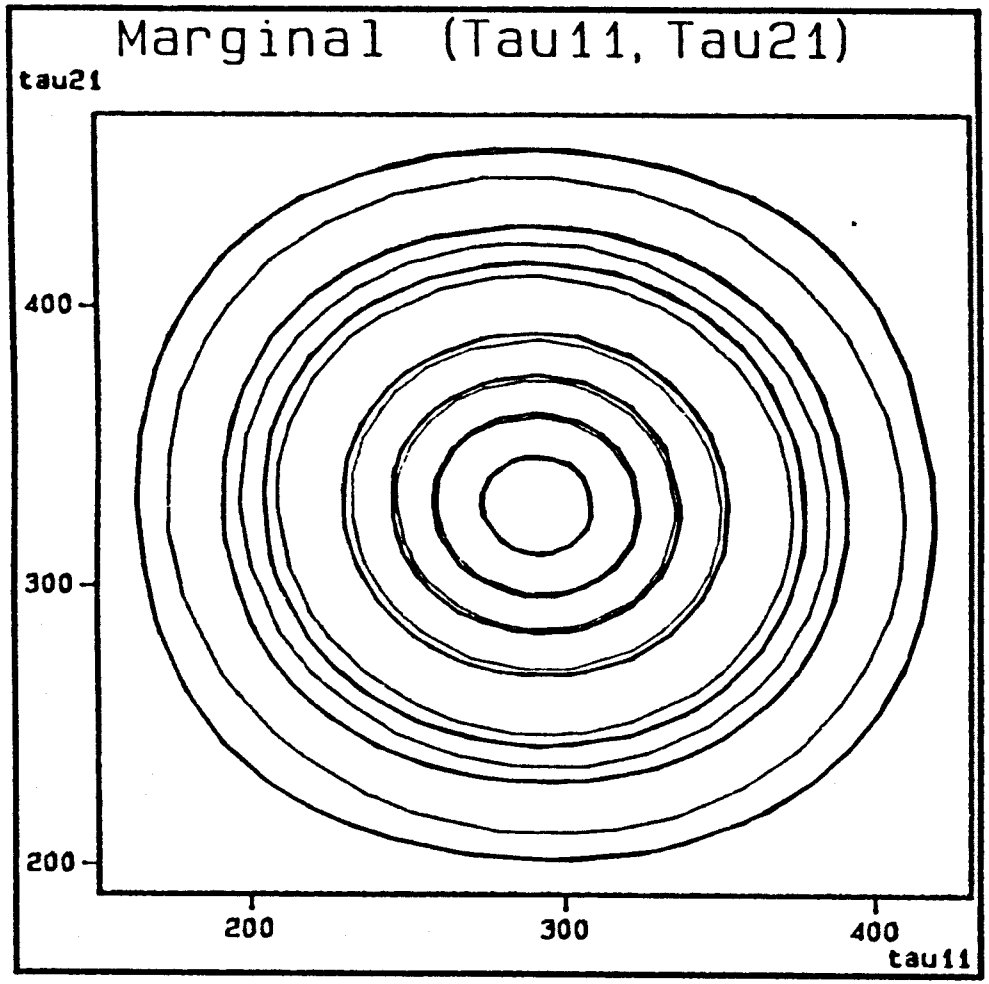
Figure 5.7.1:3 Marginal $\tau_{11} \times \tau_{12}$ under three sets of assumptions



least elliptical: Normal χ^{-2}

most elliptical: Normal Jeffreys' and $t5$ Jeffreys'

Figure 5.7.1:4 Marginal $\tau_{11} \times \tau_{21}$ under three sets of assumptions

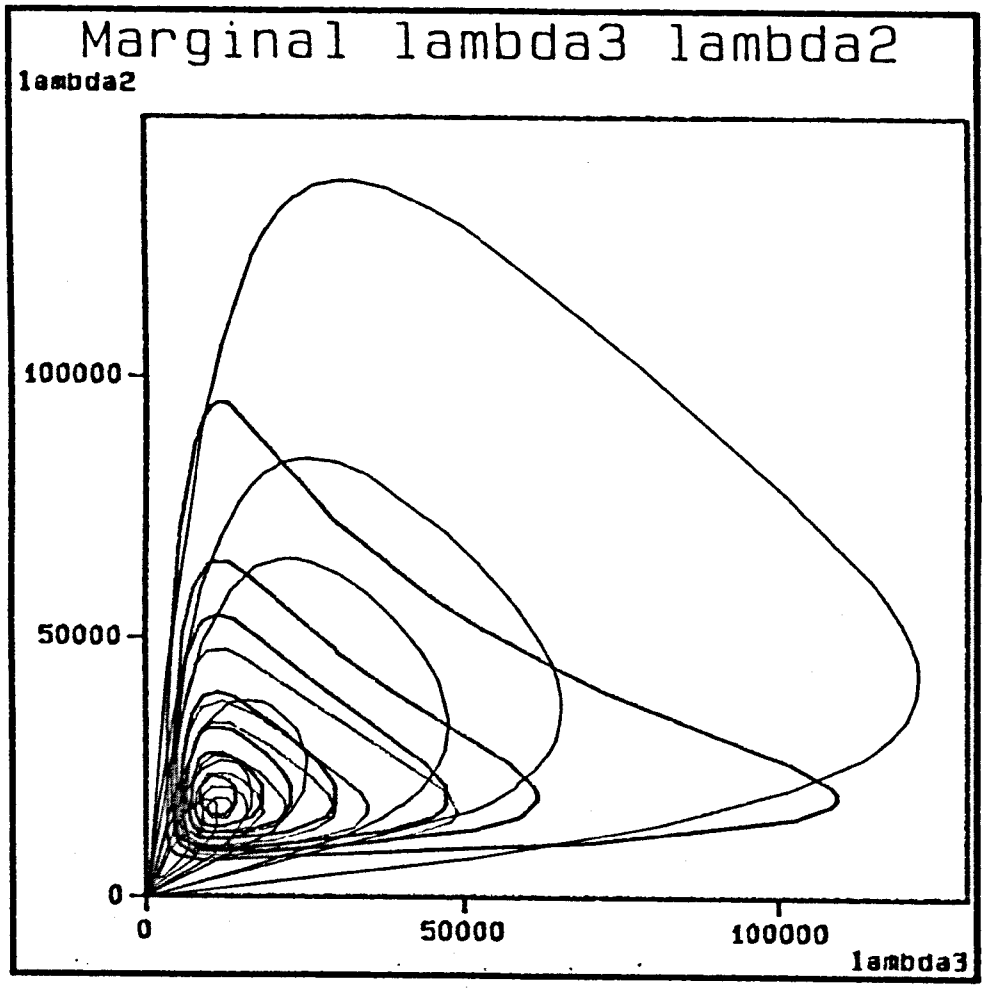


plots identified by variance

least: Normal χ^{-2}

greatest: Normal Jeffreys'
and t5 Jeffreys'

Figure 5.7.2:1 Marginal $\lambda_3 \times \lambda_2$ under three sets of assumptions



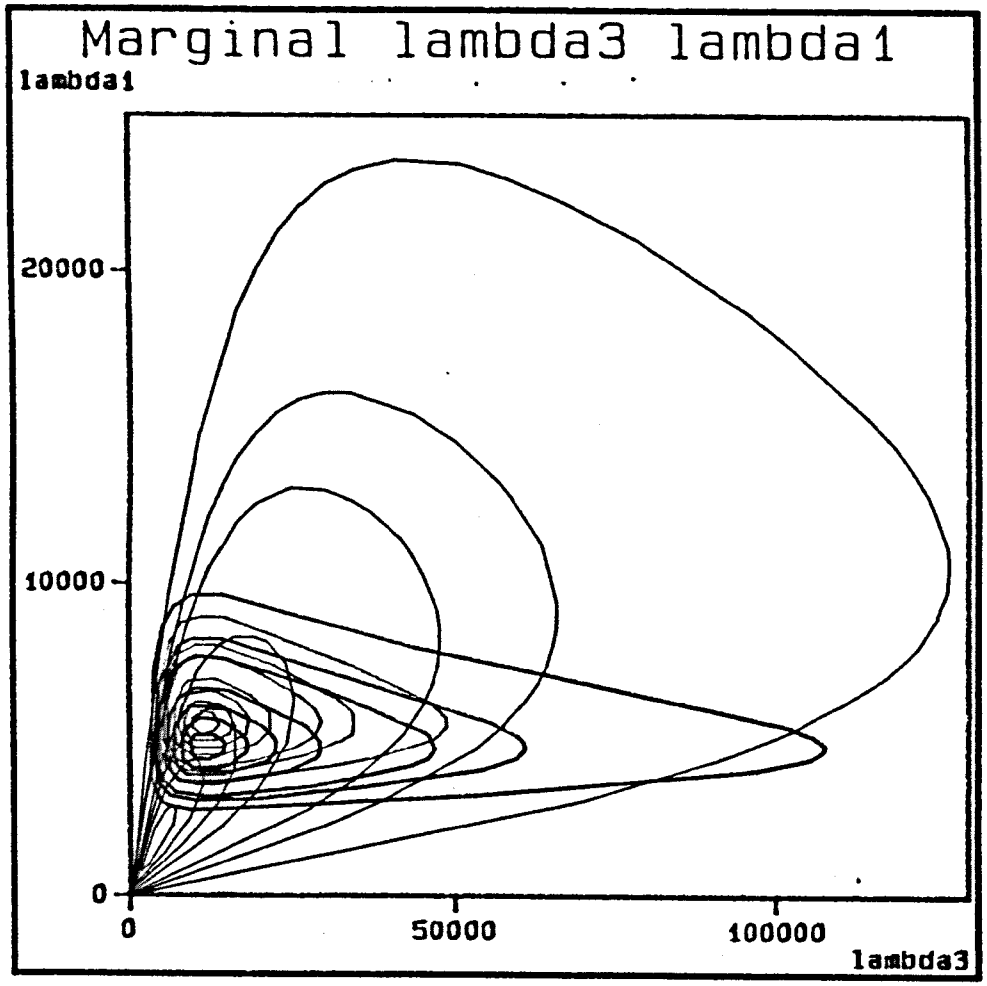
plots identified by variance

least: Normal χ^{-2}

middle: Normal Jeffreys'

greatest: *t*5 Jeffreys'

Figure 5.7.2:2 Marginal $\lambda_3 \times \lambda_1$ under three sets of assumptions



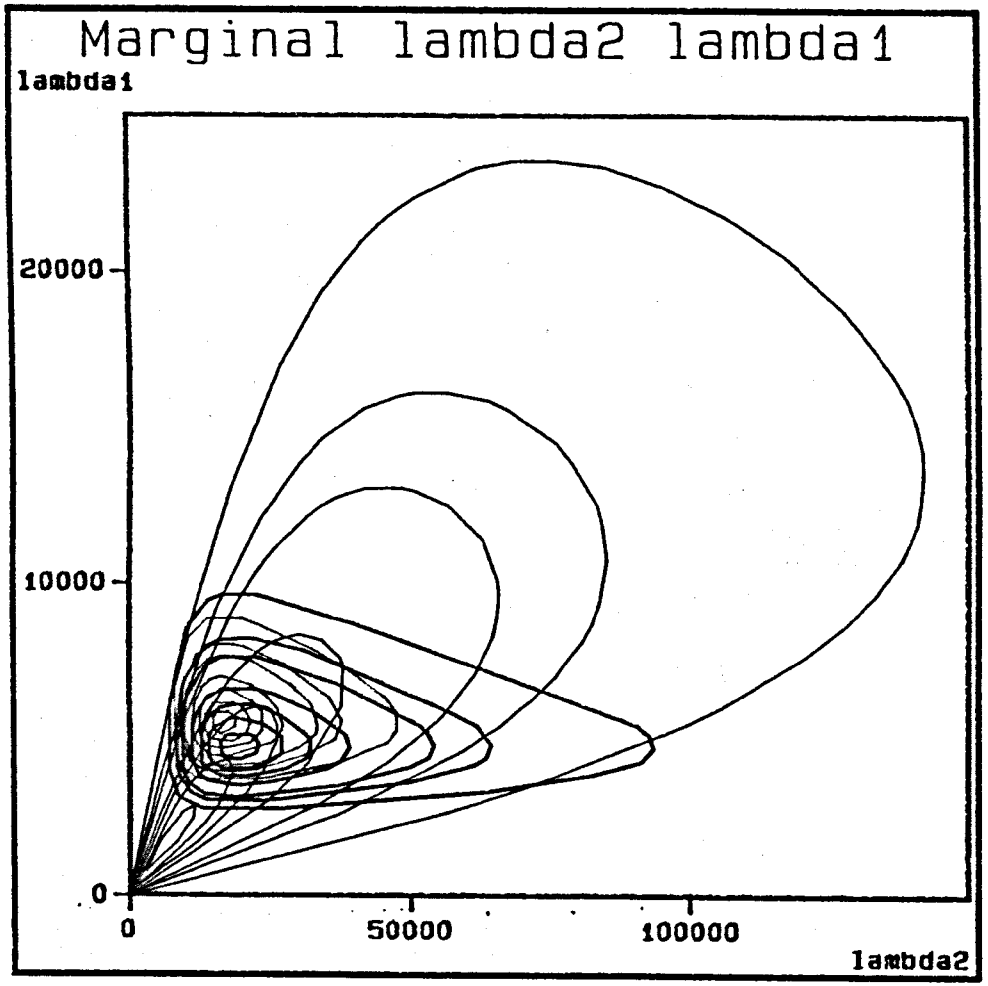
plots identified by variance

least: Normal χ^{-2}

middle: Normal Jeffreys'

greatest: t5 Jeffreys'

Figure 5.7.2:3 Marginal $\lambda_2 \times \lambda_1$ under three sets of assumptions



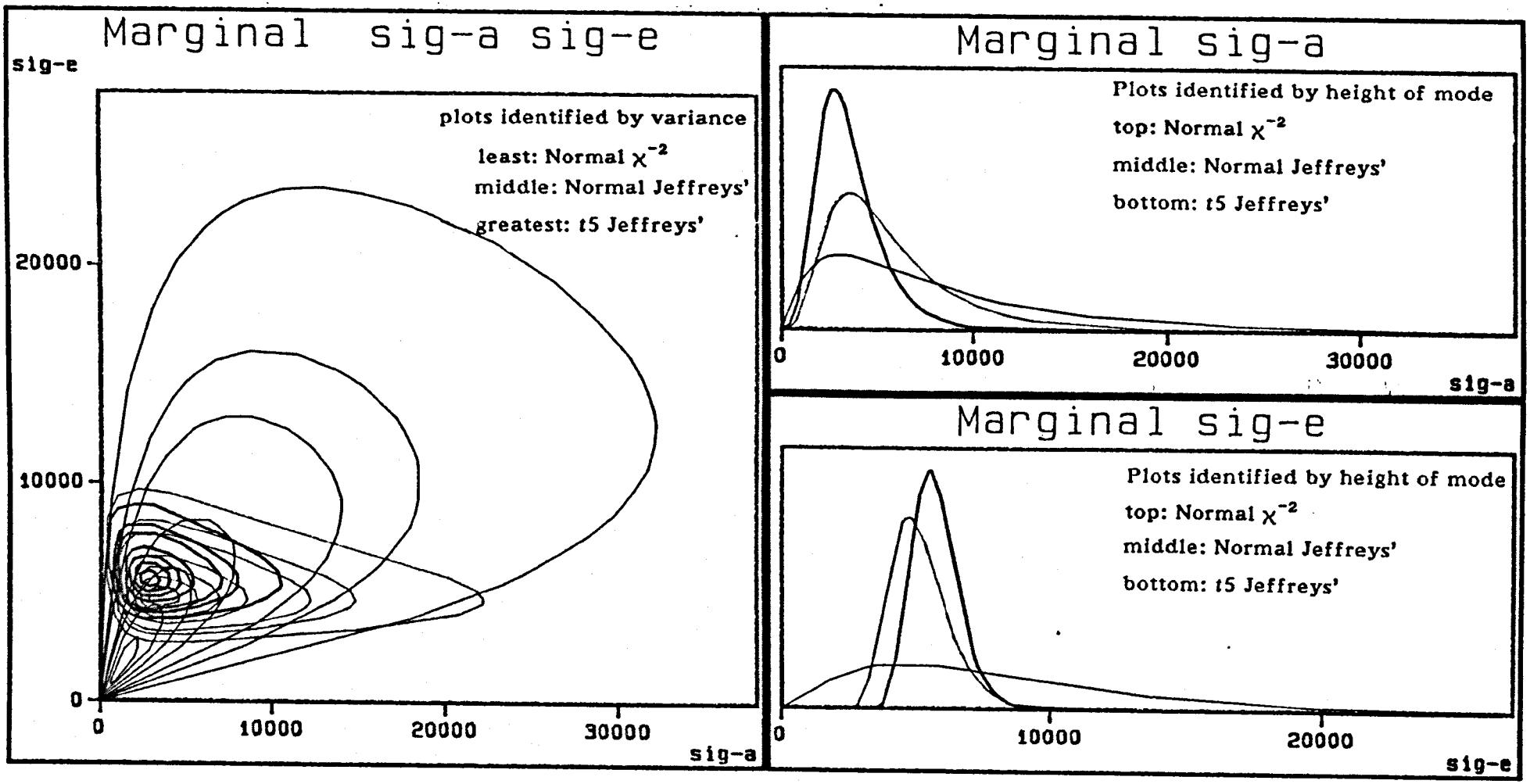
plots identified by variance

least: Normal χ^{-2}

middle: Normal Jeffreys'

greatest: $t5$ Jeffreys'

Figure 5.7.3:1 Marginal $\sigma_a^2 \times \sigma^2$, σ_a^2, σ^2 under three sets of assumptions



Chapter 6
A Bayesian analysis of inter and intra laboratory variation

6.1 Introduction

The *British Standards Institute* has defined a standard for expressing the precision of a test method in the document BS5497. This standard addresses itself to determining the precision of a standard test method and isolates two different sources of variation called *repeatability* r, and *reproducibility* R. Repeatability measures the variation between observations made in as far as possible identical conditions (same laboratory, same operator and machine, same day), whereas reproducibility R is the variation between observations from different points in time and space. By way of example, in a nested design, repeatability is based on the residual variance, and reproducibility is based on the sum of the variances from all the levels.

The standard lays down a statistical method for producing point estimates of r and R from an initial calibration experiment, but does not consider the variances of these estimates. The precise definitions for r and R are closely related to the *critical differences* at the 95% probability level for two single results obtained under the conditions of repeatability and reproducibility. For example, in an experiment involving a number of laboratories each providing replicate measurements of the same test material, point estimates are produced for the residual error σ^2 and the inter laboratory variance σ_L^2. Assuming normality, the critical difference at the 95% probability level for the difference of two observations made under identical conditions, would be $1.96\sqrt{2\sigma^2}$. The standard defines r to be $2\sqrt{2}\sigma$ which is 2.83σ. Similarly, the critical difference at the 95% probability level for the difference of two observations made at different laboratories, would be $1.96\sqrt{2\sigma^2+2\sigma_L^2}$. The standard defines R as $2.83\sqrt{\sigma^2+\sigma_L^2}$.

This chapter provides an alternative statistical analysis using Bayes theorem to produce the marginal distributions for r and R conditional on the data, and also the predictive densities for the difference of two observations under the conditions of repeatability and reproducibility. Marginal distributions on variance components tend to have considerably heavier right tails than normal distributions with the same first two moments, and thus critical differences based on point estimates will tend to be biased towards zero. It is argued that the predictive distribution for the difference of two future observations under repeatability or reproducibility may be

a more appropriate summary.

6.2 A Bayesian approach

The simplest calibration experiment considered by the standard involves J laboratories each taking K measurements of a common test material. More complicated structures can be collapsed to this as only the inter and intra laboratory variances are considered. (This will often imply more structure than the simple exchangeability that is assumed below, but the model considered can be modified to handle this.) The data can be specified as y_{jk} where j is the number of the laboratory and k distinguishes replicate observations of the sample made at that laboratory. Observations are assumed to be independent of each other and have mean μ_j and variance σ^2. All the laboratories are taken to have the same precision. The laboratory means, in the language of the standard, are considered to be random effects, and are taken to have mean μ and variance σ_L^2.

Assuming normal distributions at both the error and laboratory level this gives:

$y_{jk} \sim N(\mu_j,\sigma^2)$ independent of all the other y

$\mu_j \sim N(\mu,\sigma_L^2)$ independent of all the other μ_l

Thus $y|\mu,\sigma^2$ is multivariate normal and $\mu|\mu,\sigma_L^2$ is also multivariate normal. Integrating out the laboratory means, the two multivariate normal distributions yield:

$$y|\mu,\sigma^2,\sigma_L^2 \sim N(\mu,V(\sigma^2,\sigma_L^2))$$

where V is a block diagonal matrix with $\sigma^2+\sigma_L^2$ on the diagonal and σ_L^2 in the off diagonal elements of the blocks.

This standard model is discussed in some detail by Box and Tiao (1973). The accepted non-informative prior $p(\mu,\sigma^2,\sigma_L^2) = p(\sigma^2,\sigma_L^2) = \frac{1}{\sigma^2(\sigma^2+K\sigma_L^2)}$. Note also that r and R do not depend on the global mean μ, which can be integrated out analytically leaving the two dimensional marginal posterior density:

$$p(\sigma^2,\sigma_L^2|y) \propto (\sigma^2)^{-\frac{1}{2}J(K-1)}(\sigma^2+K\sigma_L^2)^{-\frac{1}{2}J}\exp-\tfrac{1}{2}\frac{Y_1^2}{\sigma^2}\exp-\tfrac{1}{2}\frac{Y_2^2}{\sigma^2+K\sigma_L^2} \qquad \sigma^2,\sigma_L^2 > 0 \quad (6.2{:}1)$$

where $Y_1^2 = \sum_{j=1}^{J}\sum_{k=1}^{K}(y_{jk}-\bar{y}_{j.})^2$ and $Y_2^2 = \sum_{j=1}^{J}(\bar{y}_{j.}-\bar{y}_{..})^2$.

Conditional on the within laboratory variance σ^2, the predictive density for the difference between two observations from the same laboratory is $N(0, 2\sigma^2)$. Similarly, conditional on the σ^2 and σ_L^2, the predictive density for the difference between two observations from the different laboratories is $N(0, 2\sigma^2+2\sigma_L^2)$. Denote these two differences by d_1 and d_2.

Thus $d_1|\sigma^2 \sim N(0,2\sigma^2)$ (6.2:2)

and $d_2|\sigma^2,\sigma_L^2 \sim N(0,2\sigma^2+2\sigma_L^2)$ (6.2:3)

Predictive densities are simply:

$$p(d_1|y) \propto \int_{\sigma^2,\sigma_L^2} p(d_1|\sigma^2)\,p(\sigma^2,\sigma_L^2|y)\,d\sigma^2\,d\sigma_L^2 \qquad (6.2:4)$$

$$p(d_2|y) \propto \int_{\sigma^2,\sigma_L^2} p(d_2|\sigma^2,\sigma_L^2)\,p(\sigma^2,\sigma_L^2|y)\,d\sigma^2\,d\sigma_L^2 \qquad (6.2:5)$$

From equations 6.2:2 and 6.2:3 it can be seen that the integrals in equations 6.2:4 and 6.2:5 depend on the variance components only through σ^2, $\sigma^2+K\sigma_L^2$ and $\sigma^2+\sigma_L^2$. Since the joint posterior density for σ^2,σ_L^2 (given in equation 6.2:1) factorises into a term involving σ^2 and a term involving $\sigma^2+K\sigma_L^2$, the form of the posterior density for σ^2 is known. This simplifies the integration in equation 6.2:4 to a 1 dimensional integral which can be done analytically, yielding a t distribution with mean 0 and variance $2Y_1^2/(\nu-2)$ with $\nu = J(K-1)-1$ degrees of freedom. Neither the marginal posterior for $\sigma^2+\sigma_L^2$, nor the integration in equation 6.2:5 can be calculated analytically. Numerical integration is straight forward.

It is also of interest to see the marginal densities for r and R conditional on the data. These can be obtained as simple transformations of univariate margins calculated equation 6.2:1.

6.3 Two examples

Two examples are presented based on the first two numerical examples from BS 5497. The first example, called *Table* 1 in BS 5497, presents data from 7 laboratories with 2 observations from each laboratory. The standard analysis of variance is given in Table 6.3:1 below:

Analysis of Variance from BS 5497 Table 1			
Sum of Squares	d.o.f.	Mean Square	Expected Mean Square
$Y_1^2 = 0{\cdot}290$	$J(K-1) = 7$	0·04143	σ^2
$Y_2^2 = 0{\cdot}984$	$J-1 = 6$	0·1641	$\sigma^2+2\sigma_L^2$

Table 6.3:1

Thus $\hat{\sigma^2} = 0{\cdot}04143$ and $\hat{\sigma_L^2} = 0{\cdot}06132$. From these estimates of the variances, values of 0·58 and 0·91 are calculated for r and R. It can be seen from Figure 6.3:1 that the BS value for R lies just below the modal value for the marginal density for R. Examination of the predictive density for the difference of two observations from different laboratories, it can be seen that the range $\pm 0{\cdot}91$ covers about 93 % of the density, rather than the claimed 95%. Similarly the range $\pm 0{\cdot}58$ covers about 92½% of the predictive density for the difference of two observations from the same laboratory.

Though not correct, the coverage probabilities do not look too badly wrong. However, in some cases a manufacturer whose product was within specification might experience difficulty if 7 % of his product was rejected, when he had only expected 5 % to be rejected. Similarly, too many laboratories that are in reality up to precision will be rejected.

The second example, called *Table* 2 in BS 5497, presents data from 9 laboratories with 3 observations from each laboratory. The analysis of variance is given in Table 6.3:2 below:

Analysis of Variance from BS 5497 Table 2			
Sum of Squares	d.o.f.	Mean Square	Expected Mean Square
$Y_1^2 = 44{\cdot}8062$	$J(K-1) = 18$	2·4892	σ^2
$Y_2^2 = 445{\cdot}037$	$J-1 = 8$	55·630	$\sigma^2+3\sigma_L^2$

Table 6.3:2

Thus $\hat{\sigma^2} = 2{\cdot}4892$ and $\hat{\sigma_L^2} = 17{\cdot}727$. Using these estimates of the variances, values of 4·46 and 12·72 are calculated for r and R. It can be seen from Figure 6.3:2 that the BS value for R lies just above the modal value for the marginal density for R. From the predictive density for the difference of two observations from different laboratories, it can be seen that the range $\pm 12{\cdot}72$ covers about 93% of the density, rather than the claimed 95%. Similarly the range $\pm 4{\cdot}46$ covers about 93½% of the predictive density for the difference of two observations from the same laboratory.

An interesting property of the non-informative prior used in this analysis is that it actually tightens the predictive densities when compared with a uniform prior on all the parameters. This is because the non-informative prior causes the estimates for σ^2 and $\sigma^2+K\sigma_L^2$ to shrink towards zero, and lower estimates yield tighter predictive densities. Had uniform priors been used, greater discrepancy between nominal and actual coverage probabilities would have been observed.

6.4 Extending the model to include t distributions.

In Section 5.5 of Chapter 5, a multivariate normal error distribution was replaced by a multivariate t distribution and the effect observed. Again, in this section, the effects of changing to t distributions are investigated.

In reality there is often little reason for the assumption of normal distributions at the laboratory mean and residual error levels. BS 5497 acknowledges this possibility but argues that the final values for r and R will be fairly stable to changes in these distributions, provided that the distributions remain unimodal.

In classical terms, a test to check the normality of the residuals will not have much power unless there is plenty of replicate data. This is unlikely to occur for two reasons:

i) All the measurements in a laboratory must be done by the same operator as this is one of the conditions of repeatability. It is hard to imagine that the operator will make many independent measurements on the same test material.

ii) If there is sufficient money available to enable many replicate measurements at all the laboratories, then the model is usually made more complex by introducing more variables eg. several different test materials may be used.

It is thus difficult to test the assumption of normality for the residual errors, and even harder to believe it without first considering possible alternatives. Similarly it is easy to justify the assumption that the laboratory means are exchangeable from some distribution, but the form of that distribution is not clear.

Two exploratory analyses were made to test the sensitivity of the assumptions of normality at the two levels. First, the normal distribution for the laboratory means was replaced by a t distribution with a small number of degrees of freedom, and the analysis repeated. Second, the normal distribution for the residuals was replaced by a t distribution, again with a small number of degrees of freedom. Specifically the choice of a t distribution is appropriate when it is believed that there may be more observations in the tails of the distribution than would be expected with a normal distribution.

6.4.1 Laboratory Means distributed as a t distribution

Consider first the case of a multivariate t distribution on the laboratory means, and a normal distribution on the residuals:

$$y_{jk} \sim N(\mu_j, \sigma^2) \qquad \text{independent of all the other } y$$

$$\mu \sim t_\nu(\mu, \sigma_L^2 I)$$

Note that it is no longer possible to analytically integrate out the μ from these distributions to produce $y|\mu,\sigma^2,\sigma_L^2$ in the manner of Section 6.2. Instead $y|\mu,\mu,\sigma^2,\sigma_L^2$ is produced. Only μ can be integrated out of the joint density leaving a marginal on μ,σ^2,σ_L^2. This is a high $(J+2)$ dimensional likelihood. While it *can* be integrated numerically (for reasonable values of J), an alternative approach is possible.

It is well known that a t_ν distribution with variance σ^2 can be expressed as a scale mixture of normal densities with the mixing density given by $\sigma^2(\nu-2)\chi_\nu^{-2}$. Similarly a multivariate t distribution can be represented as a scale mixture of a multivariate normal densities. Thus the model can be expressed:

$$y_{jk} \sim N(\mu_j,\sigma^2) \qquad \text{independent of all the other } y$$

$$\mu \sim N(\mu,\lambda I) \qquad \text{with } \lambda \sim \sigma_L^2(\nu-2)\chi_\nu^{-2}$$

and collapsing the normals:

$$p(\mu,\sigma^2,\sigma_L^2|y) \propto \int_\lambda N(\mu,V(\sigma^2,\lambda))\,p(\lambda)\,d\lambda\,p(\mu,\sigma^2,\sigma_L^2)$$

There are two distinct ways of viewing this. First it can be considered as a 3 dimensional posterior density with each posterior evaluation requiring a one dimensional integral. Second it can be viewed as a 4 dimensional posterior with λ acting as a nuisance parameter. Both methods are equivalent, but the second is more convenient for the numerical integration package. Regarding the mixing parameter as an extra dimension in the posterior is advocated by Berger (1985) as a method of replacing the J dimensions of the original integral by a single dimension.

μ can be integrated out analytically from the 4 dimensional posterior for $\mu,\sigma^2,\sigma_L^2,\lambda$ yielding:

$$p(\sigma^2,\sigma_L^2,\lambda|y) \propto (\sigma^2)^{-\frac{1}{2}J(K-1)}(\sigma^2+K\lambda)^{-\frac{1}{2}J}\exp-\tfrac{1}{2}\frac{Y_1^2}{\sigma^2}\exp-\tfrac{1}{2}\frac{Y_2^2}{\sigma^2+K\lambda}$$
$$\left[\frac{\lambda}{(\nu-2)\sigma_L^2}\right]^{-\frac{1}{2}(\nu+2)}\exp-\tfrac{1}{2}\frac{(\nu-2)\sigma_L^2}{\lambda}\,p(\sigma^2,\sigma_L^2)$$

As this is only a three parameter problem it poses no difficulty to BAYES4. Although BAYES4 could be used on the 2 dimensional posterior density for σ^2,σ_L^2 with the integration over λ carried out within the likelihood evaluation, it is substantially more efficient to use BAYES4 on the 3 dimensional posterior density for two reasons:

i) BAYES4 will choose better placed ordinates for λ than any straight forward integration technique.

ii) BAYES4 can use the correlations between λ and σ^2 & σ_L^2.

6.4.2 Some technical points

As mentioned in chapters 1 and 5, efficient evaluation of the numerical integrals requires estimates of the first and second moments of the parameters. It is clear that λ and σ_L^2 will be highly correlated. The correlation would pose no problem to the BAYES4 integration package provided that a reasonable estimate of it was available. Unfortunately no such estimate is available. However, *a priori* one would expect less correlation between λ/σ_L^2 and σ_L^2 than between λ and σ_L^2. Consequently, it is better to parameterise the posterior in terms of $\sigma^2, \sigma_L^2, \lambda/\sigma_L^2$.

Again, transformations to improve the sphericity of the posterior are necessary, and thus the final parameterisation was in terms of $\log\sigma^2$, $\log\sigma_L^2$ and $\log(\lambda/\sigma_L^2)$. Hence the integration was over the space of positive variance components rather than the space of positive eigenvalues. Initial estimates for the mean and variance of $\log(\lambda/\sigma_L^2)$, and its correlation with $\log\sigma^2$ and $\log\sigma_L^2$ were needed. Rough estimates calculated from the unbiased estimates of σ^2 and σ_L^2 together with assumption of zero off diagonal elements proved satisfactory, and BAYES4 converged rapidly. A high negative correlation between $\log(\lambda/\sigma_L^2)$ and $\log\sigma_L^2$ was established, ($-0{\cdot}5$ and $-0{\cdot}68$ in the two examples), but this was numerically much less than the correlation between $\log\lambda$ and $\log\sigma_L^2$ which was of the order of $0{\cdot}95$.

Correlations as high as these will give acute problems in the evaluation of such triple integrals unless considerable care is taken. BAYES4 can handle this without difficulty, but as the example illustrates, even linear models can lead to densities where the naive parameterisation is not appropriate.

6.4.3 Predictive densities for differences in observations

Finally it is required to produce the predictive densities for two observations with laboratory means t distributed under the conditions of repeatability and reproducibility. It is clear that d_1, the difference between two observations from the same laboratory, is distributed as $N(0,2\sigma^2)$ as in equation 6.2:2. The predictive density for d_2, the difference between observations from different laboratories is more complicated. This is composed of two parts, a $t_\nu(0,2\sigma_L^2)$ from the difference in laboratory means, and $N(0,2\sigma^2)$ from the difference of two residuals. Writing the univariate t as a scale mixture of univariate normal distributions, and collapsing yields:

$$d_1|\sigma^2 \sim N(0,2\sigma^2)$$

$$\text{and } d_2|\sigma^2,\sigma_L^2 \sim \int_{\lambda^\dagger} N(0,2\lambda^\dagger+2\sigma^2)\,p(\lambda^\dagger)\,d\lambda^\dagger \quad \text{with } \lambda^\dagger \sim (\nu-2)\sigma_L^2\chi_\nu^{-2}$$

Thus an extra nuisance parameter $\lambda^\dagger$ is introduced in the evaluation of $p(d_2)$. A marginal density for d_2 is thus effectively the result of a 5 dimensional integral $(d_2,\lambda^\dagger,\sigma^2,\sigma_L^2,\lambda)$.

6.4.4 Residuals distributed as a t distribution

Now consider the case of a t distribution for the residuals and a normal distribution on the laboratory means:

$$y \sim t_\nu(\mu,\sigma^2) \qquad (6.4.4:1)$$

$$\mu \sim N(\mu,\sigma_L^2 I)$$

As in section 6.4.1, it is no longer possible to analytically integrate out the μ from these distributions to produce $y|\mu,\sigma^2,\sigma_L^2$. However, the t distribution can be written as a scale mixture of normal distributions.

$$y \sim N(\mu,\lambda) \qquad \text{with } \lambda \sim \sigma^2(\nu-2)\chi_\nu^{-2}$$

$$\mu \sim N(\mu,\sigma_L^2 I)$$

Then collapsing the normals:

$$l(y|\mu,\sigma^2,\sigma_L^2) \propto \int_{\lambda} N(\mu,V(\sigma_L^2,\lambda))\,p(\lambda)\,d\lambda \qquad \text{with } \lambda \sim (\nu-2)\sigma^2\chi_\nu^{-2}$$

Direct application of Bayes' theorem then yields a 3 dimensional likelihood for μ,σ^2,σ_L^2 based on an integral over λ. μ can be integrated out analytically, and λ can be viewed as a nuisance parameter in the 3 dimensional likelihood for $\sigma^2,\sigma_L^2,\lambda$. The same technical points apply as described in section 6.4.2.

6.4.5 Other Technical Issues

The formulation of the model in equation 6.4.4:1 specifies that the residuals across all the laboratories are jointly distributed as a multivariate *t* distribution. This induces correlation between residuals, even between residuals from different laboratories. It could be argued that the residuals in different laboratories should be independent of each other, but equally it could be argued that each laboratory should have its own precision, rather than all the laboratories having a common residual error σ^2. If correlation between residuals from different laboratories is to be avoided, then a separate multivariate *t* distribution should be applied to each laboratory.

All *J* multivariate *t* distributions can be written as a product of a Inverse Chi squared distribution and a multivariate normal distribution. The *J*+1 normal distributions can be collapsed into a single normal distribution (thereby loosing the *J* laboratory means μ_j), and the global mean μ can be integrated out. However the likelihood is still specified in terms of 2 variance components and *J* nuisance parameters. At first sight this yields a *J*+2 dimensional likelihood. Fortunately it is not as bad as that, as the likelihood can be considered as being 2 dimensional but involving *J* one dimensional integrals. This technique for handling high dimensional likelihoods that are the product of many low dimensional integrals is discussed in Skene and Wakefield (1986).

6.4.6 Predictive densities for differences in observations

Following the procedure of section 6.4.3, it is clear that d_1, the difference between two observations from the same laboratory, is distributed as $t_\nu(0,2\sigma^2)$ as in

equation 6.2:2. The predictive density for d_2, the difference between observations from different laboratories is composed of two parts, a $N(0,2\sigma_L^2)$ from the difference in laboratory means, and $t_\nu(0,2\sigma^2)$ from the difference of two residuals. Writing the univariate t as a scale mixture of univariate normal distributions, and collapsing yields:

Thus $d_1|\sigma^2 \sim t_\nu(0,2\sigma^2)$

and $d_2|\sigma^2,\sigma_L^2 \sim \int_{\lambda^\dagger} N(0,2\sigma_L^2+2\lambda^\dagger)\,p(\lambda^\dagger)\,d\lambda^\dagger$

with $\lambda^\dagger \sim (\nu-2)\sigma^2\chi_\nu^{-2}$. Again d_2 is effectively the result of a 5 dimensional problem.

6.5 Examples revisited

The two alternative models using t distributions were applied to the same two data sets from BS 5497. As can be seen these mildly different assumptions have a profound effect. The spread of the estimates for the inter laboratory variance σ_L^2 increases by a factor of 4, and the spread of the estimates for the within laboratory variance σ^2 increases by a factor of about 3. These yield larger values for R and r as appropriate.

It can also be seen from Figures 6.5:1 and 6.5:2 that there is a substantial amount of independence between the two stages of the model, in the sense that the t distribution on the laboratory means has little effect on the residuals, and vice versa. Consequently, the marginals for r are little changed under the assumption of a t distribution on the laboratory means, and hence the predictive density for the difference of two observations from the same laboratory is hardly changed. They are of course substantially changed with a t distribution on the residuals.

R and the predictive density for the difference of two observations from different laboratories both change under both sets of assumptions. In both examples, $\sigma_L^2 > \sigma^2$, so the distributions are more sensitive to changes in the laboratory error structure, than the residual error structure.

The coverage probabilities are summarised in Table 6.5:1 below:

Coverage Probabilities				
	Table 1		Table 2	
	d_1	d_2	d_1	d_2
normal	92·5%	92·9%	93·3%	92·9%
t on laboratories	92·8%	72·7%	93·3%	80·2%
t on residuals	77·5%	87·3%	83·2%	91·7%

Table 6.5:1

This table indicates that the procedure given in the Standard has severe limitations as the results depend critically on the assumption of normality at both stages in the model.

6.6 Discussion

A natural extension of the work is to consider t distributions for both the laboratory means and the residuals. More interestingly:

The choice of the number of degrees of freedom ν has been somewhat arbitrary. Several possibilities suggest themselves. Simplest of all, the analysis could be repeated with several different values for ν (say 4,6,8 and 10) and the sensitivity to the choice investigated. Alternatively prior probabilities could be assigned to these values, and posterior probabilities produced. Along these lines, ν could be viewed as a index to distributions ranging from Cauchy (ν=2) to normal (ν=∞). Viewing ν as a continuous parameter, it could then be considered as another nuisance parameter in the model, giving a 4 dimensional likelihood for $\sigma^2,\sigma_L^2,\lambda,\nu$. Naylor (1982) chapter 5 views ν in this manner in a discussion about elaborated models. The extra nuisance parameter ν would pose no numerical problems for integration using the BAYES4 package.

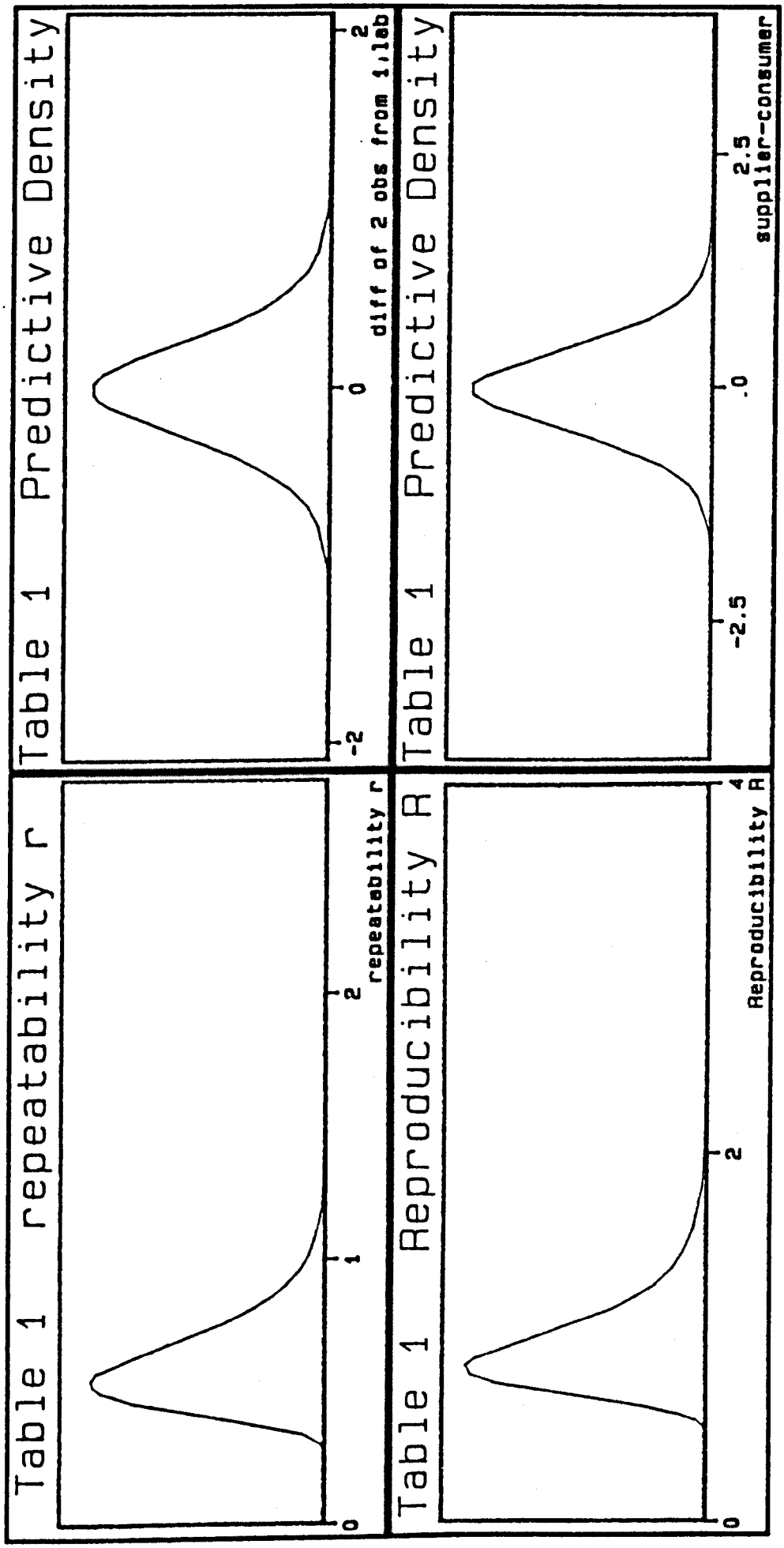


Figure 6.3:1 repeatability r, Reproducibility *R* and predictive densities for Table 1 of BS 5497 assuming normality

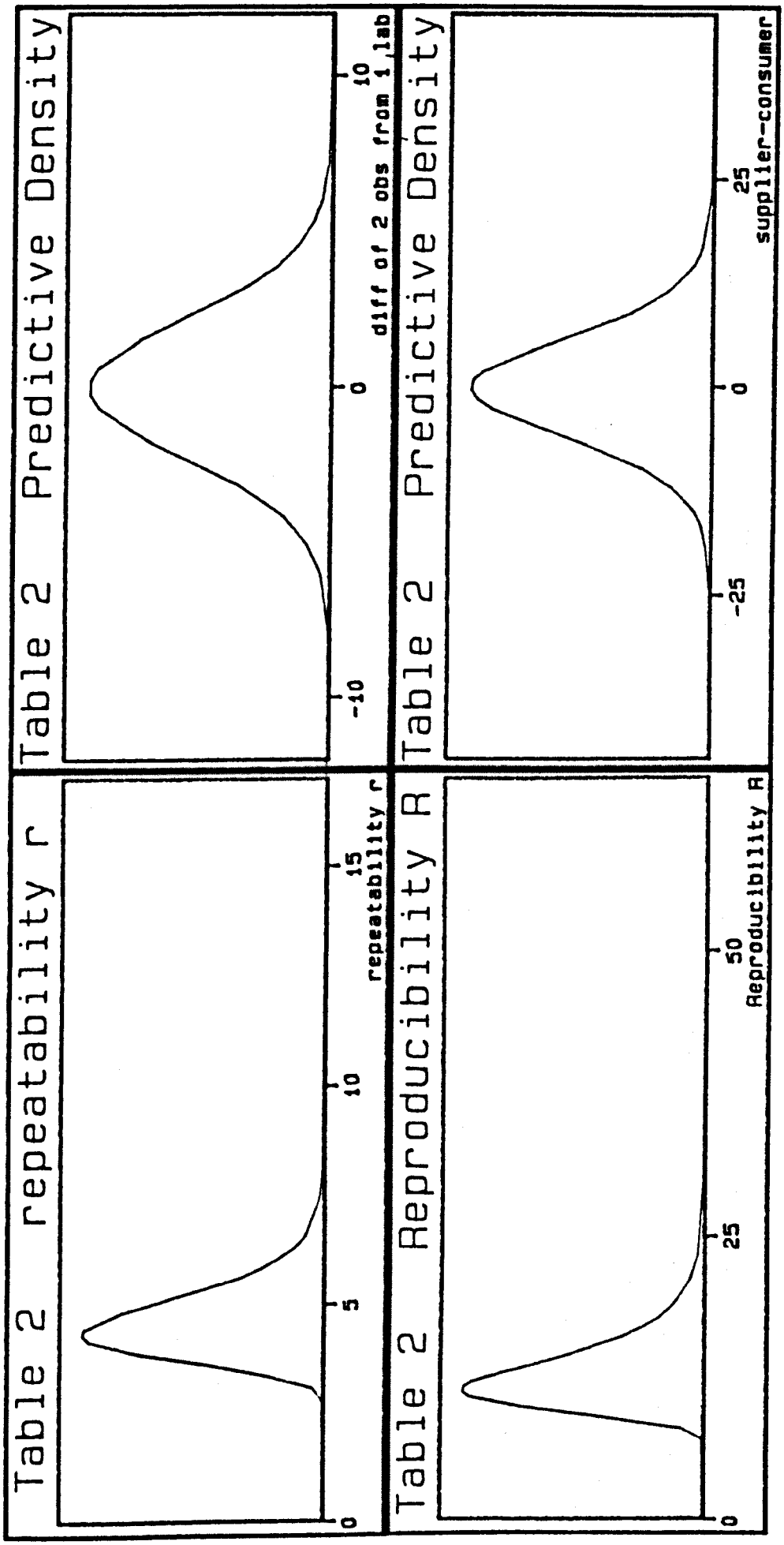


Figure 6.3:2 repeatability r, Reproducibility *R* and predictive densities for Table 2 of BS 5497 assuming normality

Figure 6.5:1 repeatability r, Reproducibility R and predictive densities for Table 1 of BS 5497 under three sets of assumptions

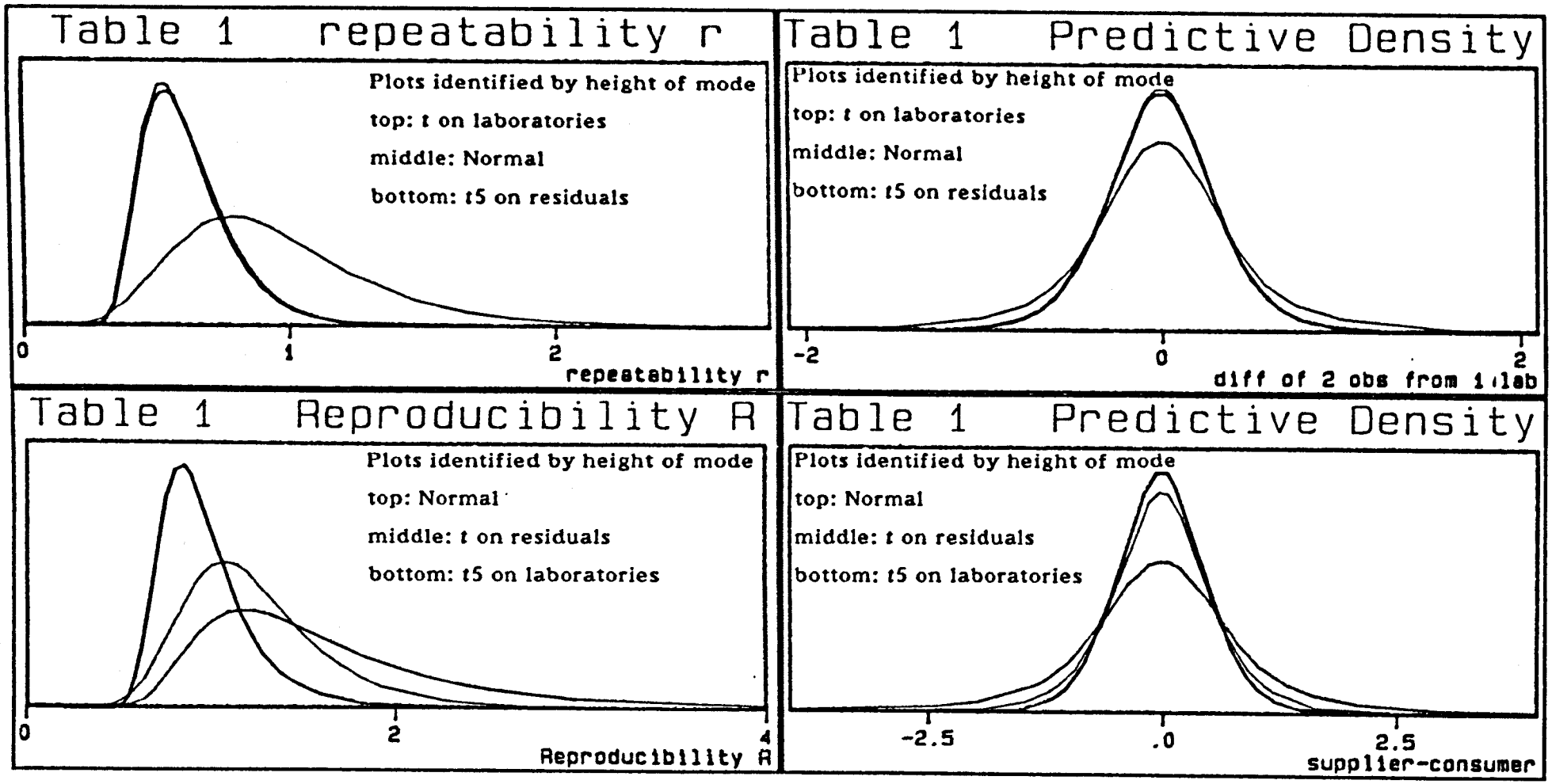
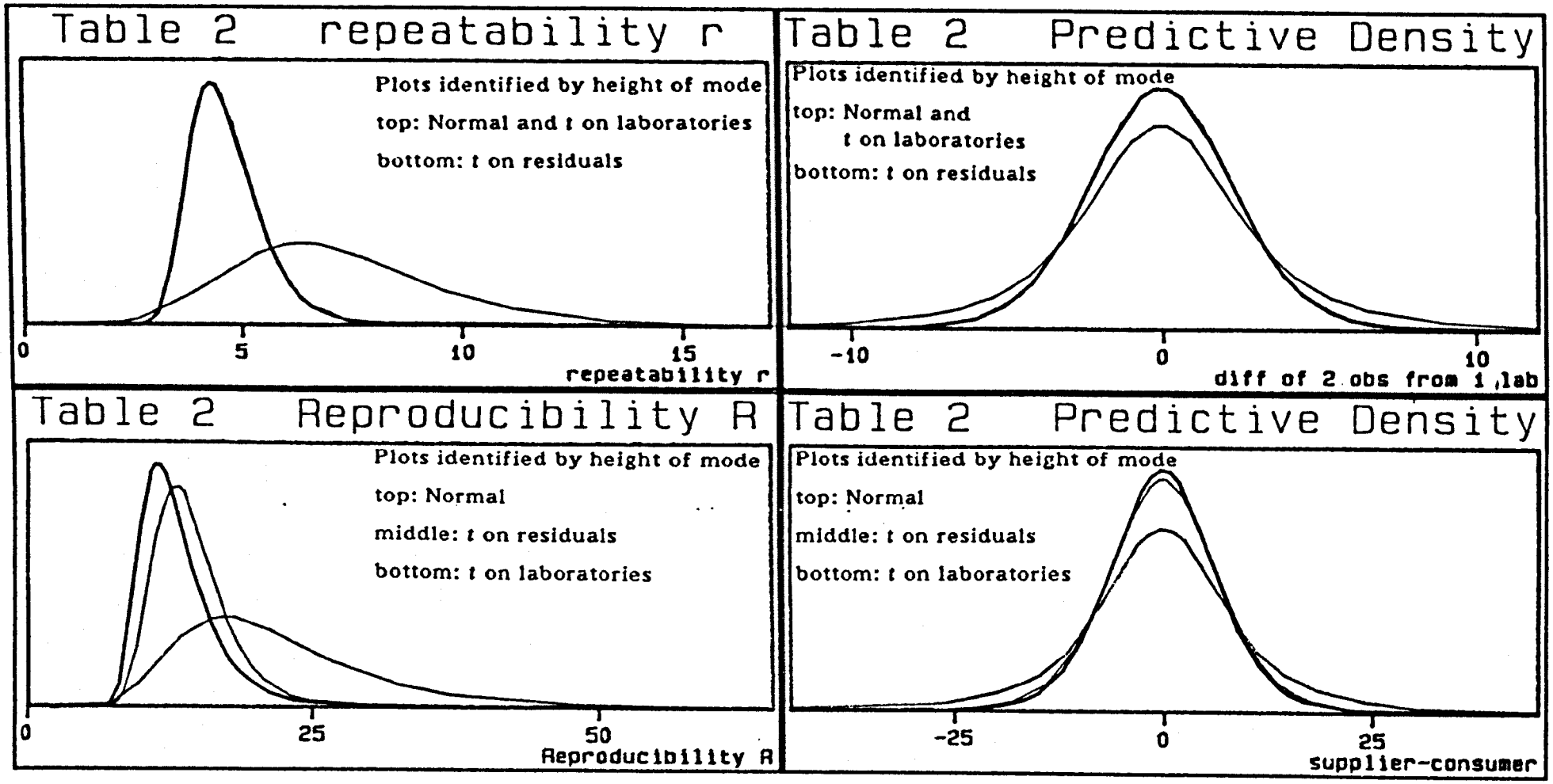


Figure 6.5:2 repeatability r, Reproducibility R and predictive densities for Table 2 of BS 5497 under three sets of assumptions



Chapter 7
Classical Variance Component Analysis vs Bayesian Procedures

7.1 Introduction

The preceding chapters have developed, amongst other ideas, algebraic and numerical techniques that allow Bayesian estimation of variance components via marginal posterior densities in a range of practical situations. Such machinery has only recently become available with the advent of powerful computers and good numerical algorithms. Variance component estimation has been discussed for much longer than this, and in the last 20 years many different point estimators of variance components have been suggested in addition to the traditional ANOVA estimates.

This chapter summarises this recent work and concludes with a discussion of estimates comparing classical methods with the previously described numerical Bayes procedures. The two main classes of new point estimators are those based on minimum variance or minimum norm (see sections 7.3, 7.4 and 7.5), and those based on maximising some sort of likelihood (see section 7.6). An extended survey can be found in either Rao (1979) or Rao and Kleffe (1980).

Throughout the chapter the mixed linear model is considered:

$$y = X\tau + \varepsilon \qquad E(\varepsilon) = 0$$

$$D(\varepsilon) = \theta_1 V_1 + \ldots + \theta_p V_p = V_\theta \qquad (7.1{:}1)$$

τ is an unknown vector, and X is a known design matrix. The main interest is in the variance components θ.

7.2 ANOVA methods

A traditional method of estimating variance components is to equate the observed and expected mean squares in an ANOVA table and solve the resulting equations for the estimators. These estimators are usually called the ANOVA estimators and for balanced designs they are unbiased, are easy to calculate and have minimum variance amongst unbiased quadratic estimators. ANOVA estimates yield translation invariant, quadratic, unbiased estimators. Under normality they are minimum variance amongst all unbiased estimators (see for example Searle

(1971a,1971b)). Henderson (1953) extended the techniques to unbalanced data, but these estimators have few desirable properties. For balanced data they correspond to the MML estimators (see later) unless non-negativity constraints come into play.

7.3 Minimum Variance Unbiased Estimators

A Minimum Variance Unbiased Estimator is sought for θ (or a linear combination of the θ_i, ie. $f'\theta$), placing no restrictions upon the class of estimators. A Locally Minimum Variance Unbiased Estimator LMVUE of $f'\theta$, γ say, can be derived at any chosen point (τ_0, θ_0), where (τ_0, θ_0) is a *prior estimate* of (τ, θ). If the estimator is not a function of τ_0 or θ_0 then the estimator is Uniformly Minimum Variance Unbiased Estimator UMVUE, but these only occur in simple cases. Rao and Kleffe (1980) suggest the use of LMVUE at a point (τ_0, θ_0) which is based on previous considerations, or alternatively they advocate iterative use of LMVUE to produce an IMVUE (Iterated Minimum Variance Unbiased Estimator). The IMVUE is not necessarily unbiased, and is discussed further in Section 7.5.

An estimator $\hat{\gamma}_*$ is defined to be the LMVUE of γ at (τ_0,θ_0) iff $E(\hat{\gamma}_*) = \gamma$ and $V(\hat{\gamma}_*|\tau_0, \theta_0) \leqslant V(\hat{\gamma}|\tau_0, \theta_0)$ for all $\hat{\gamma}$ that are unbiased. The LMVUE for $f'\theta$ given an *a priori* estimate (τ_0,θ_0) for (τ,θ) is:

$$\hat{\gamma} = \sum_{i=1}^{p} \lambda_i (y - X\tau_0)' A_{i\theta_0} (y - X\tau_0) \qquad (7.3:1)$$

where λ is any solution of $K_{\theta_0}\lambda = f$,

$$K_\theta = (tr\, A_{i\theta} V_j)$$

$$A_{i\theta} = V_\theta^{-1}(V_i - P_\theta V_i P_\theta')V_\theta^{-1}$$

and $$P_\theta = X(X'V_\theta^{-1}X)^{-1}X'V_\theta^{-1}$$

The result in equation 7.3:1 can be established by showing that $cov(g(y),\hat{\gamma}|\tau_0,\theta_0) = 0$ for all $g(y)$ such that $E[g(y)|\tau,\theta] = 0$ for all τ,θ and using a theorem due to Rao (1973) p317 on minimum variance estimation (see for example Rao (1979)).

LMVUEs were obtained by La Motte (1973) in the class of quadratic functions under normality constraints, and by Rao (1971a,1971b) in the class of all functions under normality constraints. Rao denoted the LMVUE by MiVQUE (Minimum Variance Quadratic Unbiased Estimates). Papers by Kleffe and Pincus (1974a,1974b) and Kleffe (1977a,1977b) extended the theory to include quadratic forms in $(y - X\beta)$ and they proved that under normality MiVQUE is LMVUE in the whole class of unbiased estimators.

7.3.1 LMVIUE

Locally Minimum Variance Invariant Unbiased Estimators (LMVIUE) can be developed in a similar way. It is clearly a desirable property that the variance estimators should be invariant to translations of the data. The LMVIUE class of estimators is restricted to those $g(y)$ such that $E[g(y)|\tau,\theta] = f'\theta$ and $g(y+X\tau)=g(y)$ for all τ. The LMVIU estimator is similar to the LMVUE but there is a different set of linear equations to solve.

Let P be the orthogonal projection onto the column space of X, ie.

$$P = X(X'X)^{-}X \qquad M = I - P$$

$$H_{UI}(\theta) = (tr[(MV_\theta M)^{+} V_i (MV_\theta M)^{+} V_j]) = (tr[V_\theta^{-1}(I-P_\theta)V_i(I-P_\theta')V_\theta^{-1}V_j]) \quad (7.3.1:1)$$

$$h_I(y,\theta) = [y'(MV_\theta M)^{+} V_1 (MV_\theta M)^{+} y, \ \dots\ , y'(MV_\theta M)^{+} V_p (MV_\theta M)^{+} y)'$$

$$= [y'V_\theta^{-1}(I-P_\theta)V_1(I-P_\theta)'V_\theta^{-1}y, \dots, y'V_\theta^{-1}(I-P_\theta)V_p(I-P_\theta)'V_\theta^{-1}y]' \quad (7.3.1:2)$$

where $(\ )^{+}$ denotes the Moore Penrose inverse (see Rao and Mitra (1972)). The LMVIUE of $f'\theta$ at θ_0 is $\hat{\gamma} = \lambda' h_I(y, \theta_0)$ where λ is any solution to $[H_{UI}(\theta_0)]\lambda = f$. (see for example Rao (1979)).

7.4 Minimum Norm Quadratic Estimation (MiNQE)

In section 7.3 on LMVUEs, no restriction except unbiasedness was placed on the estimating functions, yet the estimating functions always turned out to be a quadratic. It is thus intuitive to explore the field of quadratic estimators more fully. In this section only quadratic estimators are considered, and the assumptions of

unbiasedness and/or normality are dropped. This leads to the MiNQE family of estimators as proposed by Rao in a series of papers (1970,1971a,1971b,1972,1973).

Suppose that $\epsilon = U_1\phi_1 + \ldots + U_{p-1}\phi_{p-1} + \phi_p$ where U_i are $(n \times n_i)$ matrices, and ϕ_i are independent observable variables with mean zero and variance σ_i^2. Then $V_i = U_iU_i'$, and *natural estimates* $\hat{\theta}_i$ of θ_i are $\phi_i'\phi_i/n_i$ yielding an estimator of γ of the form $\hat{\gamma}_* = f_1\hat{\theta}_1 + \ldots + f_p\hat{\theta}_p$. For later convenience suppose that α is a prior estimate of θ, define $\eta' = (\alpha_1^{\frac{1}{2}}\phi_1', \ldots, \alpha_p^{\frac{1}{2}}\phi_p')'$, and N such that $\hat{\gamma}_* = \eta'N\eta$. Note that $\eta'N\eta$ does not depend on α_i.

In the general model, the error structure is less defined. Let $V_\alpha = \alpha_1V_1 + \ldots + \alpha_pV_p$, and let $\eta = V_\alpha^{\frac{1}{2}}\epsilon$. Then a *natural estimator* is:

$$\hat{\gamma}_* = \eta'(\Sigma\lambda_i V_\alpha^{\frac{1}{2}}V_iV_\alpha^{\frac{1}{2}})\eta = \eta'N\eta \qquad (7.4:1)$$

where λ is chosen to make $E(\eta'N\eta) = f'\theta$, which implies that $[H_J(\alpha)]\lambda = f$ where $H_J(\alpha) = (tr V_\alpha^{-1}V_iV_\alpha^{-1}V_j)$. It should be stressed that this *natural estimator* cannot be calculated as the fixed effects τ are unknown, hence ϵ (and thus η) are unobservable.

Now consider a general quadratic estimator. Assume (τ_0, K) as a prior mean and dispersion matrix for τ, and let $\nu = K^{-\frac{1}{2}}(\tau - \tau_0)$.

$$\hat{\gamma} = (y - X\tau_0)'A(y - X\tau_0) = \left(\eta', \nu'\right)\begin{pmatrix} V_\alpha^{\frac{1}{2}}AV_\alpha^{\frac{1}{2}} & V_\alpha^{\frac{1}{2}}AXK^{\frac{1}{2}} \\ K^{\frac{1}{2}}X'AV_\alpha^{\frac{1}{2}} & K^{\frac{1}{2}}X'AXK^{\frac{1}{2}} \end{pmatrix}\begin{pmatrix}\eta \\ \nu\end{pmatrix} \qquad (7.4:2)$$

The difference between the two estimators given in equations 7.4:1 and 7.4:2 is:

$$\left(\eta', \nu'\right)\begin{pmatrix} V_\alpha^{\frac{1}{2}}AV_\alpha^{\frac{1}{2}} - N & V_\alpha^{\frac{1}{2}}AXK^{\frac{1}{2}} \\ K^{\frac{1}{2}}X'AV_\alpha^{\frac{1}{2}} & K^{\frac{1}{2}}X'AXK^{\frac{1}{2}} \end{pmatrix}\begin{pmatrix}\eta \\ \nu\end{pmatrix}$$

The MiNQE is $y'Ay$ where A is chosen to minimise the norm $\| \ \|$ of the matrix above, that is A is chosen to minimise:

$$\left\| \begin{matrix} V_\alpha^{\frac{1}{2}}AV_\alpha^{\frac{1}{2}} - N & V_\alpha^{\frac{1}{2}}AXK^{\frac{1}{2}} \\ K^{\frac{1}{2}}X'AV_\alpha^{\frac{1}{2}} & K^{\frac{1}{2}}X'AXK^{\frac{1}{2}} \end{matrix} \right\| = \left\| \begin{matrix} D_{11} & D_{21} \\ D_{21} & D_{22} \end{matrix} \right\| \quad \text{(say)}$$

for some suitably chosen norm, for example a Euclidean Norm

$$tr\, D_{11}D_{11} + 2tr\, D_{12}D_{21} + tr\, D_{22}D_{22}. \qquad (7.4:3)$$

Different types of MiNQEs can be obtained by imposing restrictions on the set of matrices A.

i) MiNQE(U) - Unbiased

ii) MiNQE(I) - Invariant with respect to translation of y by τ

iii) MiNQE(NND) - Non Negative Definite
or any combination.

The properties of the estimators depend on which of the restrictions is applied, and on the norm used. Note that τ_0 does not appear in the final MiNQE. It should however be remembered that MiNQEs do require a prior estimate for θ and K. Many papers do not consider the prior dispersion matrix for τ and simply set K to an identity matrix. If a prior estimate is not available for θ then MiNQEs can be iterated to produce IMiNQE, starting with all the α_i equal to 1.

7.4.1 MINQE(U)

An estimator $y'Ay$ can be shown to be unbiased for $\gamma = f'\theta$ if $X'AX = 0$ and $tr\, AV_i = f_i$ for $i=1,2,\ldots,p$. See Rao (1970,1971a,1971b). The square of the Euclidean norm in equation 7.4:3 becomes:

$$\|V_\alpha^{\frac{1}{2}} A V_\alpha^{\frac{1}{2}} - N\|^2 + 2\|V_\alpha^{\frac{1}{2}} TAX\|^2 \quad \text{where } T = V_\alpha + XKX' \qquad (7.4.1{:}1)$$

which reduces to:

$$tr\, AV_\alpha AV_\alpha + 2tr\, AV_\alpha AXX' = tr\, ATAT$$

To obtain the MiNQE(U) the trace is minimised by:

$$\hat{\gamma} = \sum_i \lambda_i y'A_i y, \quad A_i = T^{-1}(V_i - P_T V_i P_T')T^{-1} \qquad (7.4.1{:}2)$$

where λ is any solution to $[H_U(\alpha)]\lambda = f$ where $H_U(\alpha)$ is the matrix $(tr\, A_i V_j)$

Focke and Dewess (1972) consider an alternative to the Euclidean Norm giving different weights to the two terms in equation 7.4.1:1, yielding the r-MiNQE(U) which is the same as equation 7.4.1:2 but with T replaced by $(V_\alpha + r^2 XX')$.

7.4.2 MINQE(U,I)

Consider the class of Unbiased Invariant quadratic estimators, that is estimators of the form $y'Ay$ where A satisfies $AX = 0$ and $tr\,AV_i = f_i$ $(i=1,2,...p)$. Define:

$$T = V_\alpha + XKX' > 0, \quad V_\alpha = \alpha_1 V_1 + \ldots + \alpha_p V_p \quad \text{where } \alpha \text{ is an } a\ priori \text{ value for } \theta.$$

$$P_T = X(X'T^{-1}X)^{-}X'T^{-1}, \quad M_T = (I - P_T)$$

Under these conditions the square of the Euclidean Norm given in equation 7.4:3 simplifies considerably (see for example Rao (1979)) and becomes:

$$\| V_\alpha^{\frac{1}{2}} A V_\alpha^{\frac{1}{2}} - N \|^2 = tr(V_\alpha A V_\alpha A) - 2trA V_\alpha^{\frac{1}{2}} N V_\alpha^{\frac{1}{2}} + trNN.$$

Note that due to the choice of N, the second term is independent of A, Also the third term does not involve A, so only the first term must be minimised. This yields the MINQE(U,I) of $f'\theta$ as:

$$\hat{\gamma} = \sum_i \lambda_i y'A_i y, \quad \text{for } A_i = T^{-1}M_T V_i M_T' T^{-1} \qquad (7.4.2{:}1)$$

where λ is any solution to $[H_{UI}(\alpha)]\lambda = f$ with $H_{UI}(\alpha)$ as the matrix $(tr\,A_i V_j)$. The solution to equation 7.4.2:1 can be written in the form $f'\hat{\theta}$ where $\hat{\theta}$ is a solution to:

$$[H_{UI}(\alpha)]\theta = h_I(y, \alpha).$$

where the i^{th} element of $h_I(y, \alpha)$ is defined to be:

$$(h_I(y, \alpha))_i = y'A_i y = y'T^{-1}M_T V_i M_T' T^{-1} y$$

Note that unlike the MINQE(U), the MINQE(U,I) does not depend on the prior estimate K of the dispersion matrix for the τ.

7.5 Iterated MVIUE and Iterated MINQE(U,I)

Two different classical approaches both finding considerable favour have been discussed. Both techniques rely on *prior information*, and this frequently poses difficulties. Two solutions are discussed in the literature. First take for α a vector of ones, and second iterate starting from any initial estimate until converge occurs. The iterated MINQE(U,I) known as the IMINQE(U,I) satisfies

$$[H_{UI}(\theta)]\theta = h_I(y, \theta)$$

where $h_I(y, \theta)$ is the same as in equation 7.3.1:2. The IMiNQE(U,I) is the same as the IMVIUE and the MML estimator (see section 7.6.1), and may be biased. The equivalence between both iterated estimators and the MML estimator is interesting since the iterated estimators do not explicitly require normality but the MML estimator does. This equivalence is probably to be expected due to the close liaison between normality and quadratic functions.

7.6 Maximum Likelihood Estimates

Seek an MLE for θ under the assumption that:

$$y \sim N(X\tau, V_\theta) \tag{7.6:1}$$

$$\text{log likelihood} \quad l(\tau, \theta, y) = -\tfrac{1}{2}\log|V_\theta| - \tfrac{1}{2}(y - X\tau)'V_\theta^{-1}(y - X\tau) \tag{7.6:2}$$

Partial differentiation with respect to τ and θ_i produces

$$\frac{dl}{d\tau} \quad X'V_\theta^{-1}X\tau = X'V_\theta^{-1}y \tag{7.6:3}$$

$$\frac{dl}{d\theta_i} \quad tr\, V_\theta^{-1}V_i = (y - X\tau)'V_\theta^{-1}V_iV_\theta^{-1}(y - X\tau) \qquad i=1,\ldots,p \tag{7.6:4}$$

re-arranging equation 7.6:3 yields:

$$X\tau = P_\theta y \qquad P_\theta = X(X'V_\theta^{-1}X)^{-1}X'V_\theta^{-1}$$

substitute this into equation 7.6:4 to yield:

$$[H_I(\theta)]\theta = h_I(y, \theta) \tag{7.6:5}$$

where the $H_I(\theta)$ matrix $= (tr\, V_\theta^{-1}V_iV_\theta^{-1}V_j)$ and the i^{th} element of $h_I(y,\theta) = y'(I - P_\theta)'V_\theta^{-1}V_iV_\theta^{-1}(I - P_\theta)y$. Note that the same equations can be obtained from an Iterated MiNQE(I).

The MLE provides an estimator of $(\hat{\tau},\hat{\theta})$. An iterative method must be employed to obtain a solution to equation 7.6:5. If a solution is found then it is the MLE. There are several problems however. If the supremum of equation 7.6:2 occurs at a boundary rather than in the interior of the permissible space, then there may be no solution to equation 7.6:5.

Under the normality assumptions in equation 7.6:1 the MLE of τ is of course unbiased. However, the solution to equation 7.6:5 is biased as it does not take into account the loss of degrees of freedom in estimating τ.

$$ie.\ E[h_I(y, \theta)] \neq [H_I(\theta)]\theta$$

The MLE is asymptotically unbiased for large samples, however the amount of bias may be large if p is large and n is of moderate size. For this reason, Patterson and Thompson (1975) proposed the Marginal Maximum Likelihood estimator.

7.6.1 Marginal Maximum Likelihood estimation MML

Instead of solving equation 7.6:5, the following equation is solved to yield a MML estimator for $\hat{\theta}$.

$$h_I(y, \theta) = E[h_I(y, \theta)] = [H_{UI}(\theta)]\theta$$

That is, MML estimators can be obtained by maximising the likelihood of θ based on error contrasts. Error contrasts are any $u'y$ such that $E(u'y) = 0$ and $u'X = 0$ where u does not depend on θ or τ. The maximum number of linearly independent error contrasts is $n - p^{\dagger}$, where $p^{\dagger}$ is the number of linearly independent columns of the X matrix. Define $X^{\dagger}$ to be any $p^{\dagger}$ linearly independent columns of X. A particular set of $n - p^{\dagger}$ linearly independent error contrasts is Ty, a $(n - p^{\dagger})$ by n matrix, where T is any $n - p^{\dagger}$ linearly independent columns from $I - X(X'X)^{-}X'$.

$$l_1(\theta, Ty) = -\tfrac{1}{2}\log|T'V_\theta T| - \tfrac{1}{2}y'T(T'V_\theta T)^{-1}T'y \qquad (7.6.1{:}1)$$

MML maximises l_1 rather than l in equation 7.6:2. Differentiating with respect to θ_i gives:

$$\frac{dl_1}{d\theta_i}\ tr\,(T(T'V_\theta T)^{-1}T'V_i) = y'T(T'V_\theta T)^{-1}T'V_iT(T'V_\theta T)^{-1}T'y \quad i=1,2,\ldots,p \quad (7.6.1{:}2)$$

using the identity $T(T'V_\theta T)^{-1}T' = V_\theta^{-1}(I - P_\theta)$ due to Rao, equation 7.6.1:1 becomes:

$$tr\,(V_\theta^{-1}(I - P_\theta)V_i) = y'V_\theta^{-1}(I - P_\theta)V_i(I - P_\theta')V_\theta^{-1}y \qquad i=1,2,\ldots,p \qquad (7.6.1{:}3)$$

Note that this expression is independent of T which had been chosen arbitrarily. It can be written in the form $[H_{UI}(\theta)]\theta = h_I(y,\theta)$ which establishes that the MML estimates are equivalent to the IMiNQE(U,I) and IMiVIUE estimators.

If the supremum of equation 7.6.1:1 occurs at a boundary rather than in the interior of the permissible (positive) space, then there may be no solution to equation 7.6.1:2. As with the ML estimates, the MML estimator is invariant with respect to translations of the data by $X\tau$ for all τ. The bias in the MML estimates may not be as large as in the ML estimates, particularly when p is large compared with n.

As noted in section 7.2, the MML estimates correspond to the ANOVA estimates if the data are balanced, and the non-negativity constraints are not required. Note also that in the review paper by Harville (1977) these estimators were called REML (REstricted Maximum Likelihood).

7.7 Equivalence of MML and Bayes Marginal modes

The relationship between MML estimators and Bayesian estimators can be seen by considering the marginal posterior density function for θ. Assuming the prior on θ,τ factorises into a term in θ and a term in τ, the posterior density on θ can be written as the prior on θ multiplied by the Marginal Likelihood. Thus the MML estimate is seen to be the joint posterior mode for θ assuming a uniform prior on θ (see Harville (1974)).

This equivalence suggests that the MML estimate (and thus the IMVIUE and IMINQE(U,I) as well) have to be viewed with caution compared with the information obtainable from a posterior margin. In particular:

i) Uniform Priors: In Bayesian terminology, the MML takes a uniform prior on θ as a "non-informative" prior. This expands the estimates considerably compared with the accepted reference prior.

ii) Joint vs Marginal Modes: The MML estimate for θ corresponds to the Joint mode for the θ. This is somewhat inconsistent. The τ have been regarded as nuisance parameters in the estimation of θ and have been integrated out, yet all the θ have been considered at once. It would be more consistent to consider integrating out all but one of the θ_i to enable estimation of the θ_i individually, that is use marginal modes for θ_i rather than the joint mode for θ.

iii) Modes vs Means: Marginals on variance components tend to have marked positive skew, and hence modes provide unrealistically low estimates compared with the means. Note that this shrinks the estimates and i) explodes them, so these two factors may to some extent cancel each other out.

iv) IMVIUE and IMiNQE(U,I) are "independent" of the error distribution as this was not specified for their derivation. By implication, it is to be expected that these estimators are equally applicable across a range of error distributions, yet it is observed from Chapters 5 and 6 that changes in the error structure radically affect the variances.

v) Standard errors of estimates from the MML algorithm are based on local curvature at the mode. This is a poor procedure for distributions that may be very skewed.

vi) As discussed later, there are similarities in the amount of numerical effort required to produce Bayes or MML estimators. Convergence with the MML algorithm is best for surfaces that are approximately quadratic, hence transformations of parameter space may be useful (eg. work with the square root of the ratios of the variances to the residual error). Quadratic convergence typically happens in six to ten iterations, thus requiring rather fewer function evaluations than efficient Gauss-Hermite integration, but the difference is not marked unless the dimension of θ is massive.

7.8 Computational Methods

The ML and MML estimators require the inverse of the dispersion matrix V to be calculated once per iteration, and the Bayesian approach requires V^{-1} at each point of a multi-dimensional lattice. As written, the MiNQEs also require the inverse of a matrix of size $n \times n$ where n is the number of observations. Clearly for large data sets this becomes computationally very time consuming. Algebraic tricks that reduce the computational load are thus highly desired.

Various authors have discussed ways of reducing the numerical effort involved in evaluating MiNQE, MIVQUE, and MML estimators. Much of this work offered a basis for the work of Chapter 3. For example, if the data are balanced, then V can be inverted analytically using the results of Searle and Henderson (1979). Wansbeek (1982) provides an analytical way of calculating more general V^{-1} using the inverse

of a much smaller matrix than V, thereby considerably reducing the computational burden. Giesbrecht and Burrows (1979) consider efficient ways of computing MiNQE(U,I) and MML estimates for hierarchical classifications. Kaplan (1983) provides a method of evaluating MiNQEs efficiently, and in their consideration of the calculation of MML estimates, Corbeil and Searle (1976) show how to reduce the n by n inversion to the inversion of a q by q matrix where q is the number of random levels in the structure.

7.9 Further work

There is scope for much further work characterising the way in which *prior information* can is used in the classical estimates. It would be intriguing to know whether these estimates handle prior information consistently as viewed from the Bayesian perspective. Also consider the case of strong prior information. Do the one-step methods such as MiNQE(U,I) have utility in providing estimates for the prior mode, such as may be used as a starting point for BAYES4? The type of prior information used by the classical estimates is also interesting as point estimates are provided for θ, but higher moments are not considered.

There is only a small amount of literature on confidence intervals producing approximate intervals for specific variances or ratios of variances in a few specific models, for example: Artega, Jeyartnam and Graybill (1982), Bross (1950), Bulmer (1957), Burdick and Sielken (1978), Graybill (1976), Graybill and Wang (1979), Wang and Graybill (1981), Green (1954), Howe (1974), Jeyartnam and Graybill (1980), Khuri (1981), Moriguti (1954), Tukey (1951) and Williams (1962). Some of these approximate methods can now be compared with the highest posterior density intervals that can be produced using BAYES4.

Chapter 8 Conclusion

8.1 Applicability of the numerical Bayesian approach

The algebra of the early chapters, and the extended examples in Chapters 5 and 6 have demonstrated that the numerical Bayesian approach employing hierarchical linear models offers a practical means of data analysis, demonstrating something of the potential of the Bayes paradigm in highly parameterised linear models. These chapters have shown that the hierarchical linear model is practical and can be used efficiently given the appropriate computer software such as that developed at the University of Nottingham. In turn, BAYES4 has been shown to work well for the hierarchical linear model using the normal times polynomial approximation to the posterior density, given a restriction to spherical error distributions. It has been shown that for hierarchical linear models, the log eigenvalue transformation is sufficient to produce a posterior density that has nearly spherical contours. The success of the log eigenvalue transformation can be seen in section 5.9.4 of chapter 5. Not only is the marginal posterior density for the log eigenvalues spherical, but also the joint posterior distribution for the variance components and the fixed effects is spherical. This leads to BAYES4 evaluating the fixed effects and the log eigenvalues with the same precision, and all the correlations have the same precision. As far as the integration routines are concerned there are simply a set of parameters with a spherical joint posterior density. Thus the joint posterior density fits well with BAYES4.

The viability of the numerical Bayesian approach to hierarchical linear models is seen to be a function of three things. First there is the need for algebraic manipulation of likelihood function to enable its speedy evaluation. Second, there is the requirement for efficient numerical integration and surface reconstruction routines. Finally there is also the choice of which margins to produce analytically, and which to do numerically. In many cases it is worth doing as much work analytically as is possible since this reduces the dimensionality of the numerical integration. However, this is not always the case, as sometimes an analytical integration yields a lower dimensional posterior which is much harder to evaluate. In some situations, the increased difficulty in evaluating the posterior, more than counterbalances the saving caused by having a lower dimensional numerical integral. Both cases can be seen by considering the Knuiman example from Chapter 5.

i) The maximal model has 15 dimensions comprising 12 fixed effects and 3 variance components in a balanced factorial design. Accordingly the posterior density is easy to evaluate. The 12 fixed effects can be integrated out analytically yielding a 3 dimensional posterior that is also easy to evaluate. If margins are required for the fixed effects, then these can be calculated by Special Function Analysis. The production of univariate and bivariate margins using the analytical method thus requires a grid of points to be calculated where each point relies upon a 3 dimensional numerical integral. By contrast, for the numerical method each point in the margin requires a 13 or 14 dimensional numerical integral. It is clear that the analytic technique is more efficient.

ii) By contrast consider the main effects only model. The posterior density has 6 fixed effects and 3 variance components, and is easy to evaluate. It is possible to integrate out the 6 fixed effects leaving only a three dimensional integral, but the posterior density thus produced requires the numerical inversion of a 6×6 matrix at each evaluation point. Thus using the analytic route, margins for the fixed effects could be produced using Special Function Analysis on a grid of points, where each point is based on a numerical integral of a 3 dimensional integral involving a difficult posterior. This must be compared with the numerical technique which requires an 8 or 9 dimension numerical integral of an easy function. The efficiency of the Monte-Carlo routines may make the latter technique more efficient than the analytic approach.

With balanced factorial models, it is worth considering the use of eigenvalues as parameters rather than variance components. This naturally raises the question: *Should the numerical integral be calculated over the space of all positive definite dispersion matrices, or over the space of strictly positive variance components?* The variance components are the more natural parameterisation, but experience has shown that eigenvalues are frequently more convenient. There are two main reasons, for this:

i) One or more variance, but not usually the residual variance, may be close to zero. This may lead to convergence difficulties if the variance components are parameters. These problems will not arise with the eigenvalue parameterisation as none of the eigenvalues are close to zero.

ii) The posterior correlation structure is often much simpler on the eigenvalues than on the variance components.

It should be noted that working in the space of strictly positive eigenvalues does not conflict with ones prior opinions. This can be seen by considering the one way analysis of variance. The same dispersion matrix is compatible with two models, one in which all the observations are independent, and one in which all observations within a group are equally correlated. These models lead to the same likelihood, but with different parameters, and the natural parameter spaces for the two models different: namely the space of positive variance components and the space of positive eigenvalues. With BAYES4 it is convenient to work in terms of the eigenvalues since the log eigenvalues give a posterior with approximately spherical contours. Note that this implies a slightly wider class of models than is usually considered where the variance components are positive. However, provided that the data are suggesting that the variance components are positive, it is possible within BAYES4 to work with log variance components as parameters. In Chapter 6 this was done, as it is consistent with the model of the Standard.

8.2 Comments on the use of BAYES4

The philosophy underlying BAYES4 can be found in Naylor and Smith (1982) and a general strategy for its use can be found in the BAYES4 User Guide (Naylor and Shaw 1985). In addition to these techniques, the following algorithms are useful.

In low (≤ 6) dimensional problems, BAYES4 can use Gauss-Hermite integration rules, and from any Gauss-Hermite dimension(s) BAYES4 can calculate univariate (or bivariate) margins. However, in practice it is usually necessary to apply a linear transformation to the parameters to make the integration easier. This implies that univariate and bivariate densities can only be calculated from the first two dimensions. If other margins are required, then the order of the parameters must be shuffled, and this can be done within BAYES4.

In higher dimensional problems, using spherical or Monte-Carlo rules, it is no longer possible to produce margins, though the methods give good convergence and perform the calculation of the first and second moments of all the parameters. This leads to the following algorithm for production of margins from a high dimensional density.

i) Use the spherical integration or the Monte-Carlo integration as appropriate on all the parameters.

ii) Repeat i) until first and second moments have converged. In the case of Monte-Carlo integration, now increase the number of evaluation points to see if the estimates remain stable.

iii) To produce margins, reorder the parameters if required, and set the first one or two dimensions to be done by Gauss-Hermite integration. The number of points chosen for the Gauss-Hermite integration becomes the number of points from which the posterior density will be reconstructed using GR. Use a spherical or Monte-Carlo integration on the remaining dimensions. If a Monte-Carlo integration is being used, then the number of Monte-Carlo evaluation points can typically be reduced by a factor of 5 or 10 without affecting the accuracy of the margin of interest. (Such a reduction will of course lead to very poor re-estimates of the moments of the Monte-Carlo parameters, but these estimates should be discarded). Perform one final iteration with this configuration to produce the margins.

This is another illustration of a general principle that applies to BAYES4. Iterate to gain convergence on something that is as simple as possible, then extend the problem to evaluate the margins/predictive densities of real interest. This minimises the computer time necessary. Exactly the same principle applies to Special Function Analysis - converge first, then switch on the extra analysis.

The production of predictive densities using Special Function Analysis has frequently been discussed in this thesis. The production of such densities is not an automatic procedure with BAYES4 and the user is required to write a substantial amount of code (often considerably more code than was needed for the likelihood). This is because Special Function Analysis was designed to evaluate the integral of a prescribed function $f(\theta)$ across parameter space θ, rather than evaluate a density that has been written as a distribution conditional on the posterior. A particular choice of $f(\theta)$ enables the calculation of a function such as a moment of θ or a single point corresponding to a single evaluation from a predictive density. Thus to produce a predictive density, an array of different functions $f_i(\theta)$ must be evaluated, with each $f_i(\theta)$ corresponding to a different evaluation point. Bivariate predictive densities require a lattice of evaluation points. GR can then be used to reconstruct the surface from the set of evaluation points in the same manner as for densities produced directly from BAYES4. More work could be done with BAYES4 to simplify this procedure.

Finally, consider the coding aspects of the main effects only model from Chapter 5. A particularly simple and effective method of analysing this model is the method taken by the MAPPLES program. Originally MAPPLES was coded to handle the 15 dimensional full model. Having done this a separate program could have been written to analyse the sub-model, but instead a single extra subroutine was added to MAPPLES that would take the 9 parameters of the sub-model and map them back into the 15 parameters of the full model. This vector of 15 parameters is then passed to the full 15 dimensional likelihood as if it were the full model being analysed. This saves the user from a considerable amount of work as only the one new subroutine must be written. This method of coding generalises to all cases where submodels of a maximal model are considered.

As noted before, and as illustrated in the Appendix, the code for the likelihood can in fact be code for several alternative likelihoods controlled by a selection mechanism at run time.

8.3 Alternative error distributions and heavy tails

The algebra of Chapter 2 (distributions) and Chapter 4 (priors) combines with the worked examples of Chapters 5 and 6 to show that the multivariate t error distribution is a viable and useful alternative to normality in the analysis of hierarchical linear models. The method of handling the t distribution in Chapter 6 using scale mixtures together with the early algebra illustrates that any scale mixture of normals can be used as an alternative error distribution without incurring much of an increase in numerical difficulty. Although the theory of this has been discussed by West (1984), Berger (1985) and others, this thesis provides a first practical demonstration of the value of such an analysis.

This raises the question of routine sensitivity analysis, and it is clear that some of the margins produced are robust with respect to the choice of error distribution (eg the fixed effects in Chapter 5), and others are not.

It is argued that care should still be taken, as heavy tailed distributions plus the notion of exchangeability are not a panacea for all situations. An analysis of a scatter plot may lead to a belief in a heavy tailed distribution such as a t. Equally, however, it may suggest that one or two laboratories are different from the others,

or that there are some wildly discordant data. In the first case, it may be better to use a mixture model, and in the second case it may be better to discard the outliers and use an unbalanced design respectively, otherwise a "model" may be developed that is consistent with the data, but in no way models the underlying process of interest.

8.4 Conclusion

Considerable progress has been made in the field of Bayesian estimation of variance components for a range of linear models. The algebra allows the possibility of some unbalance, but a full characterisation of the difficulty of a model has yet to be achieved. This difficulty is shown to be a function of both the structure of the dispersion matrix V and the design matrix X. A level of generality in the algebra for unbalanced cases has been established, but as yet these have not been utilised. The algebra exists for others to make use of, and also serves to prove that a lot of headway can be made. This raises the question of whether a general computer package could be written that used these balanced and unbalanced results to automatically write a likelihood for use in BAYES4.

Using BAYES4, the numerical Bayesian approach has been demonstrated to be practical for a range of highly parameterised linear models. This methodology yields marginal posterior distributions on the parameter(s) of interest, rather than a few point estimates. It has been shown that a routine sensitivity analysis can and should be carried out, producing marginal distributions under a range of possible assumptions. It is argued that no matter what approximations are used, there is no substitute for the real thing.

References

B. Abraham and W.W.S. Wei, "Inferences in a switching time series.", *American Statistical Association Proceedings of the Business and Econometrics Statistics Section*, pp.354-358 (1979).

D.L. Alspach, "A parallel filtering algorithm for linear systems with unknown time varying noise statistics.", *IEEE Transactions for Automatic Control* **AC-19**, pp.555-556 (1974).

R.L. Anderson, *Design and Estimators for Variance Components.*, University of Kentucky Technical Report 44 (1973).

C. Artega, S. Jeyartnam, and F.A. Graybill, "Confidence Intervals for proportions of total Variance in two-way cross Components of Variance model.", *Commun. Statist. Theor. Meth.* 4, pp.1643-1658 (1982).

BS5497, *Precision of test methods. Part 1 Guide for the determination of repeatability and reproducibility for a standard test method*, British Standards Institute (1979).

S.P. Bailey and G.E.P Box, "Modelling the nature and frequency of outliers", *Technical Report 2085* (1980).

J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*, Springer-Verlag (1985).

J.M. Bernardo, "Reference Posterior Distributions for Bayesian Analysis", *Journal of the Royal Statistical Society, B* 2, pp.113-147 (1979).

G.E.P. Box and G.C. Tiao, *Bayesian Inference in Statistical Analysis*, Addison Wesley Publishing Company (1973).

G.E.P. Box, "Sampling and Bayes' inference in scientific modelling and robustness (with discussion)", *Journal of the Royal Statistical Society, A* 143, pp.383-430 (1980).

L.D. Broemeling, *Bayesian Analysis of Linear Models*, Marcel Dekker Inc. (1985).

I. Bross, "Fiducial Intervals for Variance Components.", *Biometrics* 6, pp.136-144 (1950).

M.G. Bulmer, "Approximate Confidence Limits for Components of Variance.", *Biometrika* 44, pp.159-167 (1957).

R.K. Burdick and R.L. Sielken, "Exact Confidence Intervals for Linear Combinations of Variance Components in Nested Classifications.", *Journal American Statistical Association* 73, pp.632-635 (1978).

J.H. Chin Chay and L.D. Broemeling, "Some Bayesian inferences for a changing linear model.", *Technometrics* 22(1), pp.71-78 (1980).

G.G. Chen and G.E.P. Box, "Further study of robustification via a Bayesian approach", *Technical Report 1998* (1979).

R.R. Corbeil and S.R. Searle, "REML estimation of Variance Components in the mixed linear model.", *Technometrics* 18, pp.31-38 (1976).

E.A. Cornish, "The Multivariate t distribution associated with a set of normal deviates", *Australian Journal of Physics* 7, pp.531-542 (1954).

J.M. Dickey and C.H. Chen, "Direct Subjective Probability Modelling using ellipsoidal distributions", *Bayesian Statistics 2*, pp.157-182, Valencia University Press (1983).

C.W. Dunnet and M. Sobel, "A Bivariate generalisation of Student's t distribution with tables for certain special cases", *Biometrika* 41, pp.153-169 (1954).

T. Fearn, "A Bayesian approach to growth curves", *Biometrics* 62, pp.89-100 (1975).

J. Focke and G. Dewess, "Uber die schatzmethode MINQUE von C.R. Rao and ihre Verallgemeinerung.", *Math. Operationforsch Statist.* 3, pp.129-143 (1972).

S. Geisser, "Bayesian Analysis of growth curves", *Sankyha A* 32, pp.53-64 (1970).

F.G. Giesbrecht and P.M. Burrows, "Estimating Variance Components in hierarchical structures using MINQUE and Restricted Maximum Likelihood.", *Commun. Statist. Theor. Meth.* A7(9), pp.891-904 (1979).

F.A. Graybill, *The Theory and Application of the Linear Model.*, Duxbury Press,, North Scituate, Mass. (1976).

F.A. Graybill and Wang C.M., "Confidence Intervals for proportions of total variability in two factor nested Variance Components problems.", *Journal American Statistical Association* 74, pp.368-374 (1979).

J.R. Green, "A Confidence Interval for Variance Components.", *Annals of Mathematical statistics* 25, pp.671-686 (1954).

J.M. Hammersley and D.C. Handscomb, *Monte Carlo Methods*, Menthuen, London (1964).

P.J. Harrison and C.F. Stevens, "Bayesian Forecasting", *Journal of the Royal Statistical Society B* 38, pp.205-247 (1976).

D.A. Harville, "Bayesian Inference for Variance Components using only error contrasts.", *Biometrika* 61, pp.383-385 (1974).

D.A. Harville, "Maximum Likelihood approaches for Variance Components estimation and to related problems.", *Journal American Statistical Association* 72, pp.320-340 (1977).

R.M. Hawkes, *unpublished masters thesis*, 1973.

C.R. Henderson, "Estimation of Variance and Covariance Components.", *Biometrics* 9, pp.226-252 (1953).

D. Holbert and L.D. Broemeling, "Bayesian inference related to shifting sequences and 2 phase regression.", *Commumications in Statistics* A6(3), pp.265-275 (1977).

W.G. Howe, "Approximate confidence limits for the mean of X+Y where X and Y are two tabled independent random variables.", *Journal American Statistical Association* 69, pp.789-794 (1974).

H. Jeffreys, *Theory of Probability 3rd Edition*, Clarendon Press, Oxford (1961).

S. Jeyartnam and F.A. Graybill, "Confidence intervals for Variance Components in 3

factor cross classification.", *Technometrics* 22, pp.375-380 (1980).

N.L. Johnson and S. Kotz, *Distributions in Statistics: Continuous Multivariate Distributions*, John Wiley, New York (1970).

J.S. Kaplan, "A method of calculating MINQUE estimators of variance components", *Journal of the American Statistics Association* 78, pp.476-477 (1983).

D. Kelker, "Distribution Theory of Spherical Distributions and a Location-scale parameter generalisation", *Sankhya A* 32, pp.419-430 (1970).

A.I. Khuri, "Simultaneous confidence intervals for functions of Variance Components in random models.", *Journal American Statistical Association* 76, pp.878-885 (1981).

J. Kleffe and R. Pincus, "Bayes and best quadratic unbiased estimators for the parameters of the Covariance matrix in a normal linear model.", *Math. Operationsforsch. Statist.* 5, pp.47-67 (1974).

J. Kleffe and R. Pincus, "Bayes and best quadratic unbiased estimators for Variance Components and heteroscedastic Variances in linear models.", *Math. Operationsforsch. Statist.* 5, pp.147-159 (1974).

J. Kleffe, "Invariant methods for estimating Variance Components in mixed linear models.", *Math. Operationsforsch. Statist.* 8, pp.233-250 (1977).

J. Kleffe, "A note on infinity-MINQUE in Variance-covariance Components models.", *Math. Operationsforsch. Statist.* 8, pp.337-343 (1977).

M.W. Knuiman, "Combining Information in Linear and Log-Linear Models", *PhD Thesis.* (1983).

D.V. Lindley and A.F.M. Smith, "Bayes Estimates for the Linear Model", *Journal of the Royal Statistical Society, B* 34, pp.1-31 (1972).

S. Moriguti, "Confidence Limits for Variance Components. Union of Japanese Scientists and Engineers,", *Reports of Statistical Application Research* 3, pp.7-19 (1954).

L.R. La Motte, "Notes on the Covariance Matrix of a Random Nested ANOVA model", *Annals of Mathematical Statistics* 43, pp.659-662 (1972).

L.R. La Motte, "Quadratic estimation of Variance Components.", *Biometrics* 29, pp.311-330 (1973).

J. C. Naylor and A. F. M. Smith, "Applications of a method for the efficient computation of posterior distributions", *Applied Statistics* 31, pp.214-225 (1982).

J. C. Naylor, "Some Numerical Aspects of Bayesian Inference", *PhD Thesis. University of Nottingham* (1982).

J. C. Naylor and J. E. H. Shaw, *Bayes Four User Guide*, Nottingham Statistics Group, Department of Mathematics, University of Nottingham, Nottingham, U.K (1985).

P. Newbold, "Bayesian estimation of Box Jenkins transfer function noise models", *Journal of the Royal Statistical Society B* 35(2), pp.323-336 (1973).

A.O'Hagan, "On Joint and Marginal Modes", *Biometrika* 63(2), pp.329-333 (1976).

H.D. Patterson and R. Thompson, "Maximum Likelihood Estimation of Components of Variance.", *Proceedings of 8th International Biometrics Conference*, pp.197-207 (1975).

D.J. Poirier, *The Economics of Structural Change*, North-Holland Publishing Company, Amsterdam (1976).

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

C.R. Rao, "Estimation of heteroscedastic Variances in linear models.", *Journal American Statistical Association* 65, pp.161-172 (1970).

C.R. Rao, "Estimation of Variance and Covariance Components.", *Journal Multivariate Analysis* 1, pp.257-275 (1971).

C.R. Rao, "Minimum Variance Quadratic Unbiased estimation for Variance Components.", *Journal Multivariate Analysis* 1, pp.445-456. (1971).

C.R. Rao and S.K. Mitra, *Generalised Inverse of Matrices and its applications*, John Wiley, New York (1972).

C.R. Rao, "Estimation of Variance and Covariance Components in linear models.", *Journal American Statistical Association* 67, pp.112-115 (1972).

C.R. Rao, *Linear Statistical Inference and its applications. Second Edition.*, John Wiley, New York (1973).

C.R. Rao, "MINQUE Theory and it relation to ML and MML estimation of variance components", *Sankhya B* 41(3,4), pp.138-153 (1979).

C.R. Rao and J. Kleffe, *Chapter 1 of Handbook of Statistics 1, Analysis of Variance*, North Holland Publishing Company (1980).

H. Rubin, "Robust Bayesian Estimation", *Statistical Decision Theory and Related Topics, II*, Academic Press (1977).

D. Salazar, L.D. Broemeling, and A. Chi, "Parameter changes in a regression model with autocorrelated errors.", *Communications in Statistics Part A Theory and Methods* A10(17), pp.1751-1758 (1981).

S.R. Searle, "Topics in variance component estimation", *Biometrics* 27, pp.1-76 (1971).

S.R. Searle, *Linear Models*, John Wiley, New York (1971).

S.R. Searle and H.V. Henderson, "Dispersion matrices for Variance Components models.", *Journal American Statistical Association* 74, pp.465-470 (1979).

J. E. H. Shaw, "GR - User Guide", Research report, Nottingham University Statistics Group, Department of Mathematics, University of Nottingham, Nottingham, U.K (1985).

J. E. H. Shaw, "A quasirandom approach to integration in Bayesian statistics", Research report, Nottingham University Statistics Group, Department of Mathematics, University of Nottingham, Nottingham, U.K (1985).

J. E. H. Shaw, "A wide family of univariate distributions of use in Monte-Carlo

studies", Research report, Nottingham University Statistics Group, Department of Mathematics, University of Nottingham, Nottingham, U.K (1985).

A.M. Skene, "Computing marginal distributions for the dispersion parameters of analysis of variance models", *Practical Bayesian Statistics*, Longman (1983).

A.M. Skene and J.C. Wakefield, "Models for Multi-centre Binary Response Studies", *Statistician* to appear (1986).

A.F.M. Smith, "A general Bayesian Linear Model", *Journal of the Royal Statistical Society B* 35, pp.67-75 (1973).

A.F.M. Smith, "A Bayesian approach to inference about a change point in a sequence of random variables", *Biometrika* 62, pp.407-416 (1975).

A. F. M. Smith, A. M. Skene, J. E. H. Shaw, J. C. Naylor, and M. Dransfield, "The implementation of the Bayesian paradigm", *Comm. Statist. Theory* 14, pp.1079-1102 (1985).

A.H. Stroud, *Approximate Calculation of Multiple Integrals*, Prentice-Hall Inc., Englewood Cliffs, New Jersey. (1971).

T.J. Sweeting, "A Bayesian Analysis of some Pharmacological Data using a Random Coefficients Regression Model", *Applied Statistics* 31(3), pp.205-213 (1982).

H. Tsurami, "A Bayesian estimation of structural shifts in gradual switching regression with an application to the US gasoline market.", *Bayesian Analysis in Econometrics and Statistics*, North-Holland Publishing Company (1980).

J.W. Tukey, "Components in Regression.", *Biometrics* 7, pp.33-69 (1951).

C.M. Wang and F.A. Graybill, "Confidence Intervals on a ratio of Variances in the two factor nested Components of Variance model.", *Commun. Statist. Theor. Meth.* A10, pp.1357-1368 (1981).

T. Wansbeek, "Another approach to inverting a Covariance matrix when the data are unbalanced.", *Commun. Statist. Simula. Computa.* 11(5), pp.583-588 (1982).

M. West, "Outlier Models and Prior Distributions in Bayesian Linear Regression", *Journal of the Royal Statistical Society B* 46 (3), pp.431-439 (1984).

J.S. Williams, "A Confidence Interval for Variance Components.", *Biometrika* 49, pp.278-281 (1962).

A. Zellner, *An Introduction to Bayesian Inference in Econometrics*, John Wiley, New York (1971).

Appendix

This appendix contains the APPLES and MAPPLES program that were used for the analyses in Chapter 5. A brief functional description of these programs is included to help the reader to write similar programs that use the BAYES4 package. Simple programs for BAYES4 need only supply the following subroutines PROBLD, LOGLIK, PRIOR. The PROBLD routine must define the number of parameters and give them names, LOGLIK must evaluate the log likelihood at a point in parameter space, and PRIOR must evaluate the prior distribution at a point in parameter space.

The APPLES and MAPPLES programs both have large PROBLD subroutines. There are several reasons for this. First the programs read in the raw apple yields as data and calculate from these the sufficient statistics that enable efficient likelihood evaluation. Second both programs ask the user whether a normal or a t error structure is required, and the variable *nu* is set to -1 or the number of degrees of freedom as appropriate. This variable is passed through a COMMON block to the LOGLIK subroutine where the appropriate likelihood is evaluated. Thirdly the PROBLD routines ask the user to choose a particular prior. The code for the selected prior is stored in the *nprior* variable and is passed to PRIOR through a COMMON block. Finally the PROBLD subroutines calculate the cell means and do the necessary data translation to simplify the resulting likelihood.

Both APPLES and MAPPLES use special function analysis to produce densities. MAPPLES produces a predictive density for a future observation from cell (1,1) whilst APPLES produces a univariate distribution for *tau*11, a bivariate distribution for *tau*11,*tau*12, a bivariate distribution for *tau*11,*tau*21, and a predictive density for a future observation from cell (1,1). It is easier to see what is happening by reference to MAPPLES. Three subroutines are necessary for special function analysis, namely BXINIT, BXFUN and BXOUT. BXINIT performs any necessary initialisation for the special function analysis, and must indicate the number of special functions that are being used. Typically, all that is done here is to define evaluation points for the special functions themselves. The BXFUN subroutine is responsible for evaluating all of the special functions, conditional on the current parameter vector. The BXFUN subroutine in MAPPLES merely evaluates a normal (or t) distribution at a series of ordinates as defined by BXINIT. The values returned by BXFUN are weighted according to the posterior density and summed to produce the vector of special functions. This vector is passed to BXOUT which is responsible

for its output. The BXOUT subroutine in MAPPLES simply writes this vector to an output file in the format used by GR, so that the predictive density can be reconstructed using GR. In APPLES, the subroutines are more complicated, as there are more densities to consider. BXFUN calls a subroutine CALBIV twice to evaluate the two bivariate densities. Similarly BXOUT is more complicated and calls DRWBIV twice for outputing the bivariate densities. BXOUT sends the densities both to output files for later use with GR, and also to the terminal.

Another subroutine used in MAPPLES is BTFTRN. This is used to map the parameter vector as used by BAYES4 into the parameter vector wanted by LOGLIK. This may be used to implement a transformation of parameter space, but in MAPPLES it is used to map the 6 fixed effects from the main effects only model to the 12 cell means used in the maximal model. A question in PROBLD asks which model is required for a particular run.

The APPLES program

```
c     program APPLES
c
c     Program to analyse one year's data from the Knuiman apples data set
c
c     t distributions are used that have the same variances as the Normal
c     distribution
c
c     This program deals with the full model (12 fixed effects and 3
c     variance components).  All 12 fixed effects have been integrated
c     out, and the data vector y is translated to take each of the plot
c     means to zero, as this considerably simplifies the algebra (and
c     the FORTRAN!)
c
c     Two different error distributions are supported:
c     1) Multivariate Normal errors
c     2) Multivariate t errors
c
c     This program produces:
c     1) The posterior distribution for the three variance components,
c        based on y data translated to have to give a zero total in each
c        of the 12 plots.
c     2) The marginal distribution for a fixed effect (they are all identical)
```

```
c       and for both of the different pairs of fixed effects
c      eg. Tau11 v Tau12 or Tau11 v Tau21
c      3) A predictive density for any observation (eg y111).  All other
c        predictive densities are identical
c      4) The variance of and correlations between the Tau are calculated
c
c      Three different forms of the Prior distribution are supported
c      1)  Uniform Prior on the log eigenvalues.
c          This is the Jeffreys prior, ie minimum information.
c      2)  Independent inverse Chi-squared priors on the eigenvalues Eps.
c          The ICs are chosen to have the appropriate number of degrees
c          of freedom, and have a mode matching the estimates made for
c          the eigenvalues from the other years data.
c      3)  Log Normal Prior on the eigenvalues
c
       print *,'APPLES (3 dimensional)'
       Call BAYLD
       Call BAYES
       Call BAYEND
       stop
       end
c
c
       subroutine PROBLD(vnam,ndim)
c
c      subroutine to read the data file, determine the type of problem
c      calculate sufficient statistics (various sums of y), and perform
c      other housekeeping to enable fast execution of Log_Lik
c
       implicit none
       common /problm/Y12,Y22,Y32
       common /problm2/normal, nu2,nun10,nun11,nun12
       common /probpr/nprior,priorm,priorv
       common /probDr1/Tau0,kpin
       common /probDr2/ksout,type,type2
       common /probsf2/npoint
c
       real y(72), yijk2,yijd2,yidd2, Y12,Y22,Y32, Tau0(12)
       real priorm(3),priorv(3)
       integer ndata, nu,nprior,kpin,ksout
```

```
      logical normal
c
      real nu2,nun10,nun11,nun12
      integer i,j,ndim, npoint
      real sum,ymean,row(3),column(4),intr(3,4),t
      character*8 vnam(3)
      character*4 ty
      character*3 ty2
      character*7 type,type2
      character*40 string
c
c     Get IO channel numbers
      call bfgpio(kpin,ksout)
      read (kpin,*) string
c
      write (ksout,1) string
    1 format(//' <APPLES in Ch5>  Analysis of.',a40/)
c
c     6 blocks of 3 plots       of 4 trees
c     6 blocks of 3 irrigations of 4 thinnings
c
      ndata = 6 * 3 * 4
c
c     set number of dimensions for integration routine
      ndim = 3
c
c     set names for each dimension
      vnam(1) = 'llambd1'
      vnam(2) = 'llambd2'
      vnam(3) = 'llambd3'
c
c     get all the raw data y values
      read (kpin,*) (y(i),i=1,ndata)
c
      write (*,2)
    2 format(/' Choose between a multivariate Normal error distribution'
     +       /' and a multivariate t error distribution.'
     +       /' Type 0 for a Normal, or the number of degrees of'
     +        ' freedom for a t')
      read (*,*) nu
```

```
      if (nu.lt.1) then
        write (*,3)
        write (ksout,3)
    3   format(' Errors have a Normal distribution')
        normal = .TRUE.
        ty  = 'Nor/'
        ty2 = 'Nor'
      else
        write (*,4) nu
        write (ksout,4) nu
    4   format(' Errors have a t distribution with',i4,' d.o.f.')
        normal = .FALSE.
        ty  = 't05/'
        ty2 = 't05'
      endif
      nu2 = nu - 2
      nun10 = nu + ndata - 10
      nun11 = nu + ndata - 11
      nun12 = nu + ndata - 12
c
    5 print *,'Please indicate the type of Prior required'
      print *,'Type 1 for a Jeffreys prior (uniform on log Eigenvalues)'
      print *,'     2 for a Inverse Chi-squared prior on the ',
     !            'eigenvalues'
      print *,' or 3 for a log-normal prior on the Eigenvalues'
      read (*,*) nprior
c
      if (nprior.eq.1) then
        write (ksout,*) ' Prior is Jeffrey"s prior'
        type  = ty // 'Jef'
        type2 = ty2 // 'Jef'
      else if (nprior.eq.2) then
        write (ksout,*) ' Prior is Inverse Chi on the Eigenvalues'
        type  = ty // 'IC '
        type2 = ty2 // 'IC'
      else if (nprior.eq.3) then
        write (ksout,*)' Prior is log-normal on the Eigenvalues'
        write (ksout,*)' Prior Means and variances:'
        print *,'Type the prior mean, and variance for each ',
     +   'Eigenvalue in turn'
```

```
      do 6 i=1,ndim
        read (*,*) priorm(i), priorv(i)
    6   write (ksout,*) priorm(i), priorv(i)
      type  = ty  // 'Nor'
      type2 = ty2 // 'Nor'
    else
      print *,'Reply not understood - Please retype'
      go to 5
    endif
c
    print *,'Type the number of points required on the'
    print *,'Special Function Analysis grids for the fixed effects'
    print *,'(an even value is suggested)'
    read (*,*) npoint
c
c   calculate yd11, yd12, ... yd34, then subtract yd11/6 from yi11 etc
c   this maps Tau ---> Tau + Tau0, where Tau0 is (yd11/6, ... yd34/6)
c
    do 12 i=1,12
      sum = 0.0
      do 10 j=1,6
   10   sum = sum + y((j-1)*12+i)
      Tau0(i) = sum / 6.0
      do 11 j=1,6
   11   y((j-1)*12+i) = y((j-1)*12+i) - Tau0(i)
   12 continue
c
    write (ksout,13) (Tau0(i),i=1,12)
   13 format(/' These are the MLE for the 12 fixed effects',
     !    3(/4f7.1,' '))
c
c   Produce a Classical table of interaction effects
c
    do 14 i=1,3
   14   row(i) = ( Tau0((i-1)*4+1)+Tau0((i-1)*4+2)
     !        + Tau0((i-1)*4+3)+Tau0(i*4) ) / 4.0
c
    ymean = (row(1) + row(2) + row(3)) / 3.0
    write (ksout,*) ' The Mean of the data values is ', ymean
c
```

```
      do 15 j=1,4
   15   column(j) = (Tau0(j)+Tau0(j+4)+Tau0(j+8) ) / 3.0 - ymean
      do 16 i=1,3
        do 16 j=1,4
   16     intr(i,j) = Tau0((i-1)*4+j) - row(i) - column(j)
c
      write (ksout,17)
   17 format(/' Classical Parameterisation:'
     !     /'              Interaction effects      Column effects')
      write (ksout,18)
     !    intr(1,1),intr(1,2),intr(1,3),intr(1,4),row(1)-ymean
      write (ksout,18)
     !    intr(2,1),intr(2,2),intr(2,3),intr(2,4),row(2)-ymean
      write (ksout,18)
     !    intr(3,1),intr(3,2),intr(3,3),intr(3,4),row(3)-ymean
      write (ksout,*) '            ------------------------|-------'
      write (ksout,19) column(1),column(2),column(3),column(4),ymean
   18 format('           ',4f6.1,' | ',f6.1)
   19 format(' Row effects',4f6.1,' I ',f6.1,' = Global Mean')
c
c    Produce a GLIM table of interaction effects
c
      t = Tau0((1-1)*4+1)
      do 20 j=2,4
   20   column(j) = Tau0((1-1)*4+j) - t
      do 21 i=2,3
   21   row(i) = Tau0((i-1)*4+1) - t
      do 22 i=2,3
        do 22 j=3,4
   22   intr(i,j) = Tau0((i-1)*4+j) - t - column(j) - row(i)
c
      write (ksout,23) t,      column(2), column(3), column(4),
     !             row(2), intr(2,2), intr(2,3), intr(2,4),
     !             row(3), intr(3,2), intr(3,3), intr(3,4)
   23 format(/' GLIM Parametrisation'
     ! /' t ',f6.1,' |   c2',f6.1,'   c3',f6.1,'   c4',f6.1
     ! /' ----------+-----------------------------------'
     ! /' r2',f6.1,' | i2,2',f6.1,' i2,3',f6.1,' i2,4',f6.1
     ! /' r3',f6.1,' | i3,2',f6.1,' i3,3',f6.1,' i3,4',f6.1//)
c
```

```
c     calculate the sums of squares of y
c
      yijk2 = 0.0
      do 25 i=1,72
   25   yijk2 = yijk2 + y(i) * y(i)
c
      yijd2 = 0.0
      do 27 i=1,18
        sum = 0.0
        do 26 j=1,4
   26     sum = sum + y((i-1)*4+j)
   27   yijd2 = yijd2 + sum * sum
c
      yidd2 = 0.0
      do 29 i=1,6
        sum = 0.0
        do 28 j=1,12
   28     sum = sum + y((i-1)*12+j)
   29   yidd2 = yidd2 + sum * sum
c
      Y32 =  yidd2 / 12
      Y22 = (yijd2 - yidd2 / 3) / 4
      Y12 =  yijk2 - yijd2 / 4
c
      return
      end
c
c
      real function PRIOR(Lambda,ndim)
      implicit none
      common /probpr/nprior,priorm,priorv
      real Lambda(3), lambd1,lambd2,lambd3
      integer ndim
c
      integer nprior
      real priorm(3), priorv(3)
c
      integer i
      real sum
c
```

```
c     Prior uniform on log Eigenvalues
      if (nprior.eq.1) then
        prior = 1.0
c
c     Inverse Chi-squared Prior
c
      else if (nprior.eq.2) then
       lambd1 = exp(Lambda(1))
       lambd2 = exp(Lambda(2))
       lambd3 = exp(Lambda(3))
       prior = exp(-22.5 * Lambda(1) -22.5 *  6625 / lambd1
     +              -5.0 * Lambda(2) - 5.0 * 17613 / lambd2
     +              -2.5 * Lambda(3) - 2.5 * 16618 / lambd3 + 330)
     +         / (lambd1 * lambd2 * lambd3)
c
c     Log-normal prior
c
      else if (nprior.eq.3) then
        sum = 0.0
        do 2 i=1,ndim
    2    sum = sum + (Lambda(i) - priorm(i)) ** 2 / priorv(i)
        prior = exp( -0.5 * sum)
      endif
c
      return
      end
c
c
      subroutine LOGLIK(Lambda,ndim,result,ok)
c
c     subroutine to calculate the log-likelihood.
c
      implicit none
      common /problm/Y12,Y22,Y32
      common /problm1/resul,yVy,lDetV
      common /problm2/normal, nu2,nun10,nun11,nun12
c
      real Lambda(3), result, Y12,Y22,Y32
      integer ndim
      real nu2,nun10,nun11,nun12
```

```
      logical ok, normal
      real yVy,resul,lDetV
c
c     Calculate the log of the Determinant of V and X' Vinv X
c     = log |V| |X' Vinv X|
c
      lDetV = 45 * Lambda(1) + 10 * Lambda(2) + 5 * Lambda(3)
c
      yVy = exp(-Lambda(1)) * Y12
     !  + exp(-Lambda(2)) * Y22
     !  + exp(-Lambda(3)) * Y32
c
      if (normal) then
        result = -0.5 * (lDetV + yVy)
      else
        result = -0.5 * (lDetV + log(1.0 + yVy / nu2) * nun12)
      endif
      resul = result
c
      ok = .true.
c
      return
      end
c
c
c     This subroutine does the initialiation for Special Functions
c
      subroutine bxinit(nofun)
      common /probsf/x,y,predx
      common /probsf2/npoint
      integer i,nofun
      real x(12), y(12), predx(12)
c
c     Specify Number of functions
c     npoint          for a Univariate Marginal on Tauij
c     npoint by npoint for a Bi-variate Marginal on Tauij,Tauik     j <> k
c     npoint by npoint for a Bi-variate Marginal on Tauij,Tauim     i <> 1
c     npoint          for a Predicitive Density for y111
c     3 for Expectations of the Variance Components
c
```

```
      nofun = npoint + npoint*npoint + npoint*npoint + npoint + 3
c
      do 1 i=1,npoint
       x(i) = (i-0.5*(npoint+1)) * 300.0 / (npoint-1)
       y(i) = (i-0.5*(npoint+1)) * 300.0 / (npoint-1)
      print *,x(i)
    1 predx(i) = (i-0.5*(npoint+1)) * 800.0 / (npoint-1)
c
      return
      end
c
c
c     This subroutine defines the Special Functions
c
c     It returns:
c     npoint          spot heights on the Tau11 distribution,
c     npoint by npoint spot heights on the (Tau11,Tau12) distribution
c     npoint by npoint spot heights on the (Tau11,Tau21) distribution
c     npoint spot      heights for the predictive y111 density
c     and the 3 variances and covariances for the fixed effects
c
      subroutine bxfun(Lambda,ndim,funs,nofun)
      implicit none
      common /probsf/x,y,predx
      common /probsf2/npoint
      common /problm1/result,yVy,lDetV
      common /problm2/normal, nu2,nun10,nun11,nun12
      integer ndim,nofun, npoint
      real Lambda(3), funs(nofun)
      real x(12),y(12), predx(12)
      real nu2,nun10,nun11,nun12
      logical normal
      real result,yVy,lDetV, C
c
      real lambd1,lambd2,lambd3, sig2,siga2,sigb2
      real sqDet, var, cov
      integer index, i
c
c
      lambd1 = exp(Lambda(1))
```

```
      lambd2 = exp(Lambda(2))
      lambd3 = exp(Lambda(3))
c
      sig2  =  lambd1
      siga2 = (lambd2-lambd1)/4
      sigb2 = (lambd3-lambd2)/12
c
c     Calculate the univariate Tau11 distribution
c
      var = (sig2 + siga2 + sigb2) / 6.0
      sqDet = sqrt(var)
c
      if (normal) then
       do 1 i=1,npoint
    1     funs(i) = exp(-0.5 * x(i)**2 / var) / sqDet
       else
       do 2 i=1,npoint
          C = yVy + x(i)**2 / var
          funs(i) = exp(-0.5*(lDetV + log(1.0+C/nu2)*nun11) - result)
     !          / sqDet
    2  continue
       endif
c
c     Calculate Bivariate Marginal (Tauij, Tauik)    j <> k
c
      cov = (siga2 + sigb2) / 6.0
      index = 0
c
      call CalBiv(var,cov, index,funs,nofun)
c
c     Calculate Bivariate Marginal (Tauij, Tauim)    l <> m
c
      cov = sigb2 / 6.0
      index = npoint * npoint
c
      call CalBiv(var,cov, index,funs,nofun)
c
c     Calculate Predictive Density for observation from cell 11
c
      if (normal) then
```

```
      do 3 i=1,npoint
    3    funs(2*npoint*npoint+npoint+i)
     !       = exp(-0.5 * predx(i)**2 / var) / sqDet
      else
       do 4 i=1,npoint
    4    funs(2*npoint*npoint+npoint+i)
     !       = (1.0+predx(i)**2/nu2/var) ** (-0.5*(nu2+3)) / sqDet
      endif
c
      funs(nofun-2) = var
      funs(nofun-1) = (siga2 + sigb2) / (sig2 + siga2 + sigb2)
      funs(nofun)   =          sigb2  / (sig2 + siga2 + sigb2)
c
      return
      end
c
c     Subroutine to calculate Bivariate Marginals
c
      Subroutine CalBiv(var,cov, index,funs,nofun)
      implicit none
      common /probsf/x,y,predx
      common /probsf2/npoint
      common /problm1/result,yVy,lDetV
      common /problm2/normal, nu2,nun10,nun11,nun12
      integer npoint,np2, index,indx,indexi,indxi, nofun
      real var,cov, funs(nofun), yVy, result,lDetV, f
      real x(12),y(12),predx(12)
      real nu2,nun10,nun11,nun12, val
      logical normal
c
      integer i,j
      real Det,sqDet, xi,yj, C
c
      Det = var*var - cov*cov
      sqDet = sqrt(Det)
c
      np2 = (npoint + 1) / 2
      indx = index + (npoint+1)*(npoint+1)
      if (normal) then
       f = -0.5 / Det
```

```
      do 1 i=1,np2
         xi = x(i)
        indexi = index + npoint*i
        indxi  = indx  - npoint*i
        do 1 j=1,npoint
           yj = y(j)
          val = exp(f * (var*(xi**2+yj**2) - 2*cov*xi*yj)) / sqDet
          funs(indexi+j) = val
     1    funs(indxi -j) = val
      else
       do 2 i=1,np2
         xi = x(i)
         indexi = index + npoint*i
         indxi  = indx  - npoint*i
         do 2 j=1,npoint
            yj = y(j)
          C = yVy + (var * (xi**2+yj**2) - 2.0*cov * xi*yj) / Det
          val = exp(-0.5*(lDetV + log(1.0+C/nu2)*nun10) - result)
     !      / sqDet
          funs(indexi+j) = val
     2    funs(indxi -j) = val
      endif
c
      return
      end
c
c     Subroutine to display graphically the marginal distribution
c
      subroutine bxout(nofun)
      implicit none
      common /probsf/x,y,predx
      common /probsf2/npoint
      common /probDr1/Tau0,kpin
      common /probDr2/ksout,type,type2
      real x(12),y(12),predx(12)
      real Tau0(12), bxval
      integer nofun,i, kpin,ksout, npoint,index
      character*7 type,type2
      character*17 filnam
      character*40 title
```

```
c
c    Draw Univariate Marginal for Tau11
c
     filnam = 'MT12.'//type2
     call DrwUni(Tau0(2),filnam)
c
c    Draw Bivariate Marginal for Tau11 v Tau12
c
     filnam = 'MT11T12.'//type2
     index = npoint
     call DrwBiv(index,Tau0(1),Tau0(2),'Tau12',filnam)
c
c    Draw Bivariate Marginal for Tau11 v Tau21
c
     filnam = 'MT11T21.'//type2
     index = npoint + npoint*npoint
     call DrwBiv(index,Tau0(1),Tau0(5),'Tau21',filnam)
c
     filnam = 'Predy111.'//type2
     open (unit=42,status='UNKNOWN',name=filnam)
     title = ' Predictive y111  '//type
     write (42,1) title,npoint,(predx(i)+Tau0(1),i=1,npoint)
   1 format(a40/,' 1'/' y111'/i3/12f7.1/)
c
     index = npoint + 2*npoint*npoint
     do 2 i=1,npoint
   2  write (42,3) bxval(index+i)
   3 format (1h ,e16.6)
     close(42)
c
     write (*,4)     bxval(nofun-2), bxval(nofun-1), bxval(nofun)
     write (ksout,4) bxval(nofun-2), bxval(nofun-1), bxval(nofun)
   4 format(' var(Tauij)',f8.3,
    !       '  corr(Tauij,Tauik)',f7.4,
    !       '  corr(Tauij,Taulm)',f7.4)
c
     return
     end
c
c    Subroutine to Draw Univariate Marginal
```

```
c
      Subroutine DrwUni(Tau1,fname)
      implicit none
      common /probsf/x,y,predx
      common /probsf2/npoint
      common /probDr1/Tau0,kpin
      common /probDr2/ksout,type,type2
      real x(12),y(12),predx(12)
      real Tau0(12)
      character*17 fname
      character*8 type,type2
      character*40 title
      integer npoint,i,j,nstars, kpin,ksout
      real area,bxval,fract, max, Tau1,Tau11
      character*1 star(100)
c
      write (ksout,1)
    1 format(//' Univariate Marginal'//)
      area = 0.5*(bxval(1)-bxval(npoint))
      max = bxval(1)
      do 2 i=2,npoint
        if (bxval(i).gt.max) max=bxval(i)
    2   area = area + bxval(i)
c
      do 3 j=1,79
    3   star(j) = '*'
      do 4 i=1,npoint
        Tau11 = (i-6) * 30.0
        nstars = bxval(i) * 79 / max
        fract = bxval(i) / area * 100
    4   write (ksout,5) Tau11,fract,(star(j),j=1,nstars)
    5 format(f7.1,' ',f7.2,' ',100a1)
c
      open (unit=42,status='UNKNOWN',name=fname)
      title = ' Marginal Tau12 '//type
      write (42,6) title,npoint,(x(i)+Tau1,i=1,npoint)
    6 format(a40/' 1'/' Tau12'/i3/11f7.1/)
      do 7 i=1,npoint
    7   write (42,8) bxval(i)
    8 format (1h ,e16.6)
```

```
      close (42)
c
      return
      end
c
c     Subroutine to Draw Bivariate Marginal
c
      Subroutine DrwBiv(istart,Tau1,Tau2,name,fname)
      implicit none
      common /probsf/x,y,predx
      common /probsf2/npoint
      common /probDr2/ksout,type,type2
      real bxval,Tau1,Tau2
      integer npoint, istart,ifinish, ksout
      character*5 name
      character*7 type,type2
      character*40 title
      character*17 fname
      real x(12),y(12),predx(12)
      real max,temp(121)
      integer i,j,index
c
      open (unit=42,status='UNKNOWN',name=fname)
      title = ' Marginal Tau11,'//name//' '//type
      write (42,1) title,name,npoint,npoint
    1 format(a40/,' 2'/' Tau11'/1h ,a5/i3/i3)
      write (42,2) (x(i)+Tau1,i=1,npoint)
      write (42,2) (y(i)+Tau2,i=1,npoint)
    2 format(12f7.1)
      do 3 i=1,npoint
       do 3 j=1,npoint
         index = istart+(i-1)*npoint+j
    3    write (42,4) bxval(index)
    4 format(1h ,e16.6)
      close (42)
c
      write (ksout,5) name
    5 format(/' Bivariate Marginal for Tau11 v ',a5/)
      max = 0.0
      ifinish = istart + npoint*npoint
```

```
      do 6 i=istart+1,ifinish
    6 if (bxval(i).gt.max) max = bxval(i)
c
      ifinish = npoint*npoint
      do 7 i=1,ifinish
    7 temp(i) = bxval(istart+i) / max
c
      do 8 i=npoint,1,-1
    8 write (ksout,9) y(i),(temp((i-1)*npoint+j),j=1,npoint)
    9 format(1h ,f7.1,' |',12f6.3)
      write (ksout,10) (x(i),i=1,npoint)
   10 format('      +',66('-')/'       ',12f6.0/)
      write (ksout,11) fname,Tau1,Tau2
   11 format (' nb all (x,y) values in',a12,' have been translated by ('
     !     f7.1,',',f7.1,').'//)
c
      return
      end
```

The MAPPLES program

```
c     program MAPPLES
c
c     Program to analyse the Knuiman apples data set.
c
c     This program considers the 12 fixed effects as being of one of 2 forms:
c     1)  12 fixed effects - equivalent to row effects + column effects
c        + interaction effects  (+ 3 variance comps)
c     2)  4 row effects plus 3 column effects but without interaction.
c        Thus there are 1+(4-1)+(3-1)=6 dimensions + 3 variance components.
c     BTFTRN is called to convert the 12, 6 vector into a 12 vector
c
c     The Prior may be one of two types:
c     1)  Uniform Prior on Fixed effects and log eigenvalues
c        ie. Jeffreys prior
c     2)  Uniform Prior on the Fixed effects and an Inverse Chi-squared
c        prior on the eigenvalues.
c
c     The error distribution may be:
```

```
c    1) Normal
c    2) t errors
c
c    Finally special function analysis can be done to get a predective
c    density on any observation (eg y111) which typifies all the others.
c
     print *,'MAPPLES (15 or 9 dimensions)'
     call bayld
     call bayes
     call bayend
     stop
     end
c
c
     Subroutine PROBLD(vnam,ndim)
c
c    Subroutine to read the data file, determine the type of problem
c    calculate sufficient statistics (various sums of y), and perform
c    other housekeeping to enable fast execution of Log_Lik
c
     IMPLICIT NONE
     COMMON /problm/ydjd,ydjk,yidd2,yijd2,yijk2
     COMMON /probl2/nu
     COMMON /probpr/nprior
c
     REAL y(72), ydjd(3), ydjk(12), yidd2, yijd2, yijk2
     INTEGER ndata, index, i,j,k,nfe,ndim
     REAL sum,ymean
     INTEGER kpin,ksout, nu, nprior
     CHARACTER*8 vnam(15)
     CHARACTER*40 string
c
c    Get I/O Channel numbers
     call bfgpio(kpin,ksout)
c
   1 write (*,2)
   2 format(/' Please indicate which model you wish to use'
    +      /' The full model with 12 fixed effects'
    +       ' (row, column and interaction terms),'
    +      /'     or a model with row effects and column effects'
```

```
     +       ' but no interaction (6 f.e.),'
     +     //' Type 12 or 6 ')
      read (*,'(i6)') nfe
c
      read (kpin,*) string
c
      if (nfe.eq.12) then
         write (ksout,3) string
      else if (nfe.eq.6) then
         write (ksout,4) string
      else
         print *, ' Error - please type 12 or 6'
         goto 1
      endif
    3   format(//' <M12APPLES>  Analysis of ',a40/)
    4   format(//' <M6APPLES>  Analysis of ',a40/)
c
      write (*,6)
    6 format(/' Choose between a multivariate Normal error distribution'
     +      /' and a multivariate t error distribution.'
     +      /' Type 0 for a Normal, or the number of degrees of'
     +       ' freedom for a t')
      read (*,*) nu
      if (nu.lt.1) then
        write (*,7)
        write (ksout,7)
    7   format(' Errors have a Normal distribution')
        nu = -1
      else
        write (*,8) nu
        write (ksout,8) nu
    8   format(' Errors have a t distribution with',i4,' d.o.f.')
      endif
c
    9 write (*,10)
   10 format(/' Choose between a Jeffreys Prior and an Inverse Chi'
     +       ' Prior on the vc'
     +      /' Type 1 for the Jeffreys Prior'
     +      /'  or  2 for the Inverse Chi squared Prior')
      read (*,*) nprior
```

```
      if (nprior.ne.1.and.nprior.ne.2) goto 9
c
      if (nprior.eq.1) then
        write (*,11)
        write (ksout,11)
   11   format(' Jeffreys Prior')
      else
        write (*,12)
        write (ksout,12)
   12   format(' Inverse Chi-squared Prior on the variance components')
      endif
c
c     6 blocks of 3 plots       of 4 trees
c     6 blocks of 3 irrigations of 4 thinnings
c
      ndata = 6 * 3 * 4
c
c     set number of dimensions for integration routine
      ndim = nfe + 3
c
c     set names for each dimension
      if (nfe.eq.6) then
        vnam(1) = 't'
        vnam(2) = 'c2'
        vnam(3) = 'c3'
        vnam(4) = 'c4'
        vnam(5) = 'r2'
        vnam(6) = 'r3'
      else
        vnam(1) = 'Tau11'
        vnam(2) = 'Tau12'
        vnam(3) = 'Tau13'
        vnam(4) = 'Tau14'
        vnam(5) = 'Tau21'
        vnam(6) = 'Tau22'
        vnam(7) = 'Tau23'
        vnam(8) = 'Tau24'
        vnam(9) = 'Tau31'
        vnam(10)= 'Tau32'
        vnam(11)= 'Tau33'
```

```
        vnam(12)= 'Tau34'
      endif
      vnam(nfe+1) = 'l.eps3'
      vnam(nfe+2) = 'l.eps2'
      vnam(nfe+3) = 'l.eps1'
c
c     get all the raw data y values
      read (kpin,*) (y(i),i=1,ndata)
c
c     calculate ymean and subtract it from all the data
c     the mean is supposed to be a function of the year
c     and we are interested in the effects of the treatments
c     after the year effect has been discarded.
c
      sum = 0.0
      do 13 i=1,ndata
   13   sum = sum + y(i)
      ymean = sum / ndata
      do 14 i=1,ndata
   14   y(i) = y(i) - ymean
c
      write (ksout,*) ' The Mean of the data values is ', ymean
c
c     calculate the sums of squares of y
c
      yijk2 = 0.0
      do 15, i=1,72
   15   yijk2 = yijk2 + y(i)*y(i)
c
      yijd2 = 0.0
      do 17, i=1,18
        sum = 0.0
        do 16, j=1,4
   16     sum = sum + y((i-1)*4+j)
   17   yijd2 = yijd2 + sum * sum
c
      yidd2 = 0.0
      do 19, i=1,6
        sum = 0.0
        do 18, j=1,12
```

```
   18      sum = sum + y((i-1)*12+j)
   19    yidd2 = yidd2 + sum * sum
 c
      do 23 j=1,3
        ydjd(j) = 0.0
        do 23 i=1,6
          do 23 k=1,4
   23       ydjd(j) = ydjd(j) + y((i-1)*12+(j-1)*4+k)
 c
      do 24 j=1,3
        do 24 k=1,4
          index = (j-1)*4 + k
          ydjk(index) = 0.0
          do 24 i=1,6
   24       ydjk(index) = ydjk(index) + y((i-1)*12 + index)
 c
      return
      end
 c
 c
      Subroutine BTFTRN(theta,ndim,rcon)
      COMMON /probth/newtheta
      INTEGER I, ndim
      REAL theta(ndim),rcon, newtheta(15)
 c
 c    This Subroutine converts the 9 dimensional problem back
 c    up to the original 15 dimensional vector of parameters
 c
 c    theta( ) is t c2 c3 c4 r2 r3 t1 t2 t3
 c    tau( ) is t    t   +c2 t   +c3 t   +c4
 c              t+r2 t+r2+c2 t+r2+c3 t+r2+c4
 c              t+r3 t+r3+c2 t+r3+c3 t+r3+c4
 c    newtheta( ) is tau( ) t1 t2 t3
 c
      if (ndim.eq.9) then
        newtheta(1) = theta(1)
        newtheta(2) = theta(1) +          theta(2)
        newtheta(3) = theta(1) +          theta(3)
        newtheta(4) = theta(1) +          theta(4)
        newtheta(5) = theta(1) + theta(5)
```

```
      newtheta(6) = theta(1) + theta(5) + theta(2)
      newtheta(7) = theta(1) + theta(5) + theta(3)
      newtheta(8) = theta(1) + theta(5) + theta(4)
      newtheta(9) = theta(1) + theta(6)
      newtheta(10)= theta(1) + theta(6) + theta(2)
      newtheta(11)= theta(1) + theta(6) + theta(3)
      newtheta(12)= theta(1) + theta(6) + theta(4)
    else
      do 1 i=1,12
    1     newtheta(i) = theta(i)
    endif
c
c    Copy Log Eigenvalues
    newtheta(13) = theta(ndim-2)
    newtheta(14) = theta(ndim-1)
    newtheta(15) = theta(ndim)
c
    rcon=1.0
    return
    end
c
c
    REAL function PRIOR(junk,ndim)
    IMPLICIT NONE
    COMMON /probpr/nprior
    COMMON /probth/theta
    REAL theta(15), junk(15), epst, epsp, epsb
    INTEGER ndim, nprior
c
    prior = 1.0
    if (theta(13).lt.-40.0.or.theta(13).gt.40.0 .or.
   +    theta(14).lt.-40.0.or.theta(14).gt.40.0 .or.
   +    theta(15).lt.-40.0.or.theta(15).gt.40.0) then
      prior = 0.0
      goto 1
    endif
    if (nprior.eq.1) then
      prior = 1.0
    else
      epsb = exp(theta(13))
```

```
      epsp = exp(theta(14))
      epst = exp(theta(15))
      prior = exp( -2.5*theta(13) - 2.5*16618/epsb
     +             -5.0*theta(14) - 5.0*17613/epsp
     +            -22.5*theta(15) -22.5* 6625/epst + 330)
     +          / (epsb * epsp * epst)
      endif
c
    1 return
      end
c
c
      Subroutine LOGLIK(junk,ndim,result,ok)
c
c     Subroutine to calculate the log-likelihood.
c
c     Log P(y|sigma,tau) = -0.5 * log|V|
c                          -0.5 * (y - X tau)' Vinv (y - X tau)
c
c     the first 12 entries in the theta Vector correspond to the tau s
c     and the last 3 entries correspond to the sigma s
c     theta(13) = log eps_b  theta(14) = log eps_p  theta(15) = log eps_t
c
c     now (y - X tau)' Vinv (y - X tau)
c           = y' Vinv y - 2 y' Vinv X tau + tau' X' Vinv X tau
c
c     y' Vinv y' = beta sum_i yi..2
c             + gamma sum_ij yij.2
c             + delta sum_ijk yijk2
c     y' Vinv X tau = beta tau.. y...
c                + gamma sum_i taui. y.i.
c                + delta sum_ij tauij y.ij
c     tau' X' Vinv X tau = 6 ( beta tau..2
c                          + gamma sum_i taui.2
c                          + delta sum_ij tauij2 )
c
      IMPLICIT NONE
      COMMON /problm/yd1d,yd2d,yd3d,
     +               yd11,yd12,yd13,yd14,
     +               yd21,yd22,yd23,yd24,
```

```
     +            yd31,yd32,yd33,yd34,
     +            yidd2, yijd2, yijk2
      COMMON /probl2/nu
c
      COMMON /probth/theta
c
      REAL junk(15),theta(15), result
      INTEGER ndim
      LOGICAL ok
      REAL epsb1,epsp1,epst1, beta, gamma, delta
      REAL yd1d,yd2d,yd3d, yidd2,yijd2,yijk2
      REAL yd11,yd12,yd13,yd14,yd21,yd22,yd23,yd24,yd31,yd32,yd33,yd34
      REAL lDetV, tau1d,tau2d,tau3d, taudd, tauid2, tauij2
      REAL yVy,yVXt,tXVXt, quadf
      INTEGER i, nu
c
      epsb1 = exp(-theta(13))
      epsp1 = exp(-theta(14))
      epst1 = exp(-theta(15))
c
      beta  = (epsb1 - epsp1) / 12.0
      gamma = (epsp1 - epst1) /  4.0
      delta =  epst1
c
c     Calculate the log of the Determinant of V
c              6.3.(4-1)      6.(3-1)      6
c     |V| = eps_t        . eps_p      . eps_b
c
      lDetV = 54 * theta(15) + 12 * theta(14) + 6 * theta(13)
c
      yVy = beta * yidd2 + gamma * yijd2 + delta * yijk2
c
      tau1d = theta(1) + theta(2) + theta(3) + theta(4)
      tau2d = theta(5) + theta(6) + theta(7) + theta(8)
      tau3d = theta(9) + theta(10)+ theta(11)+ theta(12)
c
      taudd = tau1d + tau2d + tau3d
      tauid2 = tau1d**2 + tau2d**2 + tau3d**2
c
      tauij2 = 0.0
```

```
      do 1 i=1,12
     1    tauij2 = tauij2 + theta(i) **2
c
      yVXt = gamma * (tau1d*yd1d + tau2d*yd2d + tau3d*yd3d)
     +  + delta * ( theta(1)*yd11 + theta(2)*yd12
     +           + theta(3)*yd13 + theta(4)*yd14
     +           + theta(5)*yd21 + theta(6)*yd22
     +           + theta(7)*yd23 + theta(8)*yd24
     +           + theta(9)*yd31 + theta(10)*yd32
     +           + theta(11)*yd33+ theta(12)*yd34)
c
      tXVXt = 6 * (beta*taudd**2 + gamma*tauid2 + delta*tauij2)
      quadf = yVy - 2.0 * yVXt + tXVXt
c
      if (nu.eq.-1) then
c       Normal distribution
        result = -0.5 * lDetV - 0.5 * quadf
      else
c       t distribution
        result = -0.5 * lDetV - 0.5*(nu+72) * log(1.0 + quadf/(nu-2))
      endif
c
      ok = .true.
c
      return
      end
c
c
c     This subroutine does the initialiation for Special Functions
c
      Subroutine bxinit(nofun)
      COMMON /probsf/predx
      INTEGER i,nofun
      REAL predx(11)
c
c     Specify Number of functions
c     11 for Predicitive Density for y111
c
      nofun = 11
c
```

```
      do 1 i=1,11
    1   predx(i) = (i-6) * 90.0 - 75.0
c
      return
      end
c
c
c     This subroutine defines the Special Functions
c
c     Returns 11 spot heights for the predictive y111 density
c
      Subroutine bxfun(junk,ndim,funs,nofun)
      IMPLICIT none
      COMMON /probl2/nu
      COMMON /probth/theta
      COMMON /probsf/predx
      INTEGER ndim,nofun, nu
      REAL junk(15), theta(15), funs(nofun)
      REAL predx(11), nu1
c
      REAL epsb,epsp,epst, sig2,siga2,sigb2
      REAL Kinv,sqKinv,K2
      INTEGER i
c
      epsb = exp(theta(13))
      epsp = exp(theta(14))
      epst = exp(theta(15))
c
      sig2 = epst
      siga2 = (epsp-epst)/4
      sigb2 = (epsb-epsp)/12
      Kinv = 1/ (sig2 + siga2 + sigb2)
      sqKinv = sqrt(Kinv)
      K2 = Kinv / (nu-2)
      nu1 = -0.5 * (nu + 1.0)
      if (nu.eq.-1.0) then
        do 4 i=1,11
    4     funs(i) = sqKinv * exp(-0.5 * Kinv * (predx(i)-theta(1))**2)
      else
        do 5 i=1,11
```

```
   5     funs(i) = sqKinv * (1.0 + K2 * (predx(i)-theta(1))**2)** nu1
      endif
c
      return
      end
c
c     Subroutine to display graphically the marginal distribution
c
      Subroutine bxout(nofun)
      IMPLICIT none
      COMMON /probsf/predx
      INTEGER nofun,i
      REAL predx(11), bxval
c
c     Predective Marginal for y111
c
      open (unit=42,status='UNKNOWN',name='PREDY111.DAT')
      write (42,1) (predx(i),i=1,11),(bxval(i),i=1,11)
    1 format(' Predictive Density'/,
     +       ' 1'/' pred. y111'/' 11'/
     +       11f7.1/(11(e13.6/)))
      close(42)
c
      return
      end
```