BIOMETRICS 38, 963–974
December 1982

Random-Effects Models for Longitudinal Data

Nan M. Laird[1,2] and James H. Ware[1]

[1]Department of Biostatistics, Harvard School of Public Health, 677 Huntington Avenue, Boston, Massachusetts 02115, U.S.A. and [2]Department of Statistics, Harvard University, Science Center 603, 1 Oxford Street, Cambridge, Massachusetts 02138, U.S.A.

SUMMARY

Models for the analysis of longitudinal data must recognize the relationship between serial observations on the same unit. Multivariate models with general covariance structure are often difficult to apply to highly unbalanced data, whereas two-stage random-effects models can be used easily. In two-stage models, the probability distributions for the response vectors of different individuals belong to a single family, but some random-effects parameters vary across individuals, with a distribution specified at the second stage. A general family of models is discussed, which includes both growth models and repeated-measures models as special cases. A unified approach to fitting these models, based on a combination of empirical Bayes and maximum likelihood estimation of model parameters and using the EM algorithm, is discussed. Two examples are taken from a current epidemiological study of the health effects of air pollution.

1. Introduction

Many longitudinal studies are designed to investigate changes over time in a characteristic which is measured repeatedly for each study participant. In medical studies, the measurement might be blood pressure, cholesterol level, lung volume, or serum glucose. Multiple measurements are obtained from each individual, at different times and possibly under changing experimental conditions. Often, we cannot fully control the circumstances under which the measurements are taken, and there may be considerable variation among individuals in the number and timing of observations. The resulting unbalanced data sets are typically not amenable to analysis using a general multivariate model with unrestricted covariance structure.

Statisticians have often analyzed data of this form using some variant of a two-stage model. In this formulation, the probability distribution for the multiple measurements has the same form for each individual, but the parameters of that distribution vary over individuals. The distribution of these parameters, or 'random effects', in the population constitutes the second stage of the model. In a study of changes in lung volume during childhood, for instance, it may be reasonable to assume that the relationship between lung volume and the cube of height is linear for each child, but with linear regression parameters that vary among children. If we assume that the usual linear regression model applies for each child, conditional on the child's individual parameters, and that the regression parameters have a bivariate normal distribution in the population, the marginal distribution of the serial measurements is multivariate normal with a special covariance structure.

Key words: Two-stage models; Variance components; Growth models; Repeated measures; Empirical Bayes; EM algorithm; Restricted maximum likelihood; Air pollution; Pulmonary function.

Such two-stage models have several desirable features. There is no requirement for balance in the data. They allow explicit modelling and analysis of between- and within-individual variation. Often, the individual parameters have a natural interpretation which is relevant to the goals of the study, and their estimates can be used for exploratory analysis. These models also facilitate the study of the effects of background variables on the response. The major limitation of these models relative to the general multivariate model is the special form assumed for the covariance structure.

Despite wide recognition of the value of two-stage models, many statisticians are unaware of recent methodologic developments that allow a unified approach to the formulation and fitting of these models. This paper describes that unified methodology and illustrates its value in two problems of data analysis arising in an epidemiologic study of air pollution.

In §2, we introduce a family of two-stage models for repeated measurements, based on the work of Harville (1977). This family includes growth models and repeated-measures models as special cases. In §3, §4 and §5, we describe a unified approach to inference using these models. Both maximum likelihood and empirical Bayes estimation are discussed. The relationships, both conceptual and algorithmic, between these two approaches to inference are stressed. The EM algorithm is shown to offer a conceptually straightforward method for parameter estimation in this setting. Section 6 describes two applications of the method.

2. Models

Most stochastic models for serial measurements can be classified either as full multivariate models or multi-stage random-effects models. In the full multivariate model, we assume that each vector of responses, $\mathbf{y}_i$, is multivariate normal with mean $\boldsymbol{\mu}_i(n_i \times 1)$ and an arbitrary $n_i \times n_i$ dispersion matrix $\boldsymbol{\Sigma}$. Here n_i is the number of observations for the ith individual, $i = 1, 2, \ldots, m$. The mean vector may depend upon the pattern of observations and also upon covariates.

When the design is balanced, but observations are missing at random, the full multivariate model can be applied by use of multivariate methods for missing observations (Orchard and Woodbury, 1972; Beale and Little, 1975; Dempster, Laird and Rubin, 1977). However, when individuals are measured at arbitrary or unique times, or when the dimension of $\boldsymbol{\Sigma}$ is large, this approach becomes unattractive, since a full multivariate model with an unrestricted dispersion matrix requires a proliferation of variance parameters, many of which will be poorly estimated. In addition, the full multivariate model does not permit the definition and estimation of (random) individual characteristics.

Two-stage random-effects models are based on explicit identification of individual and population characteristics, and their form extends naturally to the unbalanced situation. Most of the two-stage models in the literature can be described either as growth models or as repeated-measures models.

As the name implies, growth-curve analyses emphasize the explanation of within-person variation by the natural developmental or aging process (Rao, 1965; Fearn, 1975; Ware, 1983). These analyses often compare growth characteristics for different populations, emphasizing the contribution of experimental conditions to between-individual variability.

In constrast, repeated-measures models, as described in the design literature, typically assume that individual effects remain constant over the time period of interest (Hayes, 1973). Experimental conditions are changed during the course of observation, either by

design or circumstance, so that the experimental effects contribute to the within-person variation.

In this section, we utilize ideas introduced by Harville (1977) to define a family of models for serial measurements that includes both growth models and repeated-measures models as special cases. Population parameters, individual effects, and within-person variation are introduced at Stage 1, and between-person variation at Stage 2. In §6, we illustrate how growth and repeated measures can be represented in this family, and demonstrate some of the advantages of the more general formulation.

Let $\boldsymbol{\alpha}$ denote a $p \times 1$ vector of unknown population parameters and $\mathbf{X}_i$ be a known $n_i \times p$ design matrix linking $\boldsymbol{\alpha}$ to $\mathbf{y}_i$. Let $\mathbf{b}_i$ denote a $k \times 1$ vector of unknown individual effects and $\mathbf{Z}_i$ a known $n_i \times k$ design matrix linking $\mathbf{b}_i$ to $\mathbf{y}_i$. For measured, multivariate normal data, we propose the following model:

Stage 1. For each individual unit, i,

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\alpha} + \mathbf{Z}_i\mathbf{b}_i + \mathbf{e}_i, \tag{2.1}$$

where $\mathbf{e}_i$ is distributed as $N(\mathbf{0}, \mathbf{R}_i)$ (normal with mean $\mathbf{0}$ and covariance matrix $\mathbf{R}_i$). Here $\mathbf{R}_i$ is an $n_i \times n_i$ positive-definite covariance matrix; it depends on i through its dimension n_i, but the set of unknown parameters in $\mathbf{R}_i$ will not depend upon i. At this stage, $\boldsymbol{\alpha}$ and $\mathbf{b}_i$ are considered fixed, and the $\mathbf{e}_i$ are assumed to be independent.

Stage 2. The $\mathbf{b}_i$ are distributed as $N(\mathbf{0}, \mathbf{D})$, independently of each other and of the $\mathbf{e}_i$. Here $\mathbf{D}$ is a $k \times k$ positive-definite covariance matrix. The population parameters, $\boldsymbol{\alpha}$, are treated as fixed effects.

Marginally, the $\mathbf{y}_i$ are independent normals with mean $\mathbf{X}_i\boldsymbol{\alpha}$ and covariance matrix $\mathbf{R}_i + \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^T$. Further simplification of this model arises when $\mathbf{R}_i = \sigma^2\mathbf{I}$, where $\mathbf{I}$ denotes an identity matrix. In that case we call the model the 'conditional-independence model', since it implies that the n_i responses on Individual i are independent, conditional on $\mathbf{b}_i$ and $\boldsymbol{\alpha}$.

Inference for this general linear model can be based either on least squares and maximum likelihood methods, or on empirical Bayes methodology. If $\boldsymbol{\theta}$ is a q-vector of variance and covariance parameters found in $\mathbf{R}_i$, $i = 1, \ldots, m$, and $\mathbf{D}$, the classical approach is based on maximum likelihood estimation of $\boldsymbol{\alpha}$ and $\boldsymbol{\theta}$ from the marginal distribution of $\mathbf{y}^T = (\mathbf{y}_1^T, \ldots, \mathbf{y}_m^T)$. An estimate for $\mathbf{b}^T = (\mathbf{b}_1^T, \ldots, \mathbf{b}_m^T)$ can be obtained by use of an extended version of the Gauss–Markov theorem for random effects (Harville, 1976). This approach is reviewed by Harville (1977). We apply it to the model (2.1) in §3, discussing some of the limitations of the maximum likelihood approach, and possible alternatives.

One alternative can be derived using a Bayesian formulation of the model. Here we introduce a flat prior for the location parameters, $\boldsymbol{\alpha}$, and estimate $\boldsymbol{\theta}$ from the marginal likelihood of $\mathbf{y}$ after integrating out $\boldsymbol{\alpha}$ and $\mathbf{b}_i$, $i = 1, 2, \ldots, m$. This approach was considered by Harville (1974, 1976) and Dempster, Rubin and Tsutakawa (1981). This modification yields restricted maximum likelihood (REML) estimates for $\boldsymbol{\theta}$. The empirical Bayes estimates of $\boldsymbol{\alpha}$ and the $\mathbf{b}_i$ are the estimated means of the posterior distributions.

In §5, we show that this Bayesian approach leads to estimates of parameters and their variances which are identical to those proposed in a sampling-theory context as alternatives to maximum likelihood estimates. The Bayesian formulation is emphasized in this paper because it provides both a conceptual and computational unity to the estimation

methods which we discuss. For either formulation, the EM algorithm provides a convenient approach to computation for random-effects models, since the individual characteristics can be viewed as missing data. Implementation of the EM algorithm for parameter estimation is described in §4.

3. Estimation and Inference with Measured Response Data

3.1 *Known Variance*

When all covariance parameters are known, and $\boldsymbol{\alpha}$ is treated as a fixed effect, expressions for the estimates of the population and individual effects and their standard errors are well-known. Writing $\text{var}(\mathbf{y}_i) = \mathbf{V}_i = \mathbf{R}_i + \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^T$, and $\mathbf{W}_i = \mathbf{V}_i^{-1}$, we have

$$\hat{\boldsymbol{\alpha}} = \left(\sum_1^m \mathbf{X}_i^T\mathbf{W}_i\mathbf{X}_i\right)^{-1} \sum_1^m \mathbf{X}_i^T\mathbf{W}_i\mathbf{y}_i \tag{3.1}$$

and

$$\hat{\mathbf{b}}_i = \mathbf{D}\mathbf{Z}_i^T\mathbf{W}_i(\mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\alpha}}). \tag{3.2}$$

We assume throughout that, whenever this is implied, the necessary matrix inverses exist. The equivalent formulas using generalized inverses can be worked out for cases of less than full rank.

The estimate of $\boldsymbol{\alpha}$ maximizes the likelihood based on the marginal distribution of the data and is also the minimum variance unbiased estimate. The expression for $\hat{\mathbf{b}}_i$ is of course not maximum likelihood but can be derived by an extension of the Gauss–Markov theorem to cover random effects (Harville, 1976).

The estimate for $\mathbf{b}_i$ is also empirical Bayes, since it has the form $\hat{\mathbf{b}}_i = \text{E}(\hat{\mathbf{b}}_i \mid \mathbf{y}_i, \hat{\boldsymbol{\alpha}}, \boldsymbol{\theta})$. Because the prior mean of $\mathbf{b}_i$ is zero, $\hat{\mathbf{b}}_i$ is a weighted combination of $\mathbf{0}$ and $\bar{\mathbf{b}}_i$, where $\bar{\mathbf{b}}_i$ is the ordinary weighted least squares estimate obtained by treating $\mathbf{b}_i$ as a fixed effect. As such, $\hat{\mathbf{b}}_i$ is related to Stein-type estimates obtained by 'shrinking towards the origin'. Similar empirical Bayes estimates of individual parameters are discussed by Rosenberg (1973) and Rao (1975) in the context of growth curves.

Since both $\hat{\boldsymbol{\alpha}}$ and $\hat{\mathbf{b}}_i$ are linear functions of $\mathbf{y}$, expressions for their standard errors are easily derived as

$$\text{var}(\hat{\boldsymbol{\alpha}}) = \left(\sum_1^m \mathbf{X}_i^T\mathbf{W}_i\mathbf{X}_i\right)^{-1} \tag{3.3}$$

and

$$\text{var}(\hat{\mathbf{b}}_i) = \mathbf{D}\mathbf{Z}_i^T\left\{\mathbf{W}_i - \mathbf{W}_i\mathbf{X}_i\left(\sum_1^m \mathbf{X}_i^T\mathbf{W}_i\mathbf{X}_i\right)^{-1}\mathbf{X}_i^T\mathbf{W}_i\right\}\mathbf{Z}_i\mathbf{D}. \tag{3.4}$$

If (3.4) is used to assess the error of estimation, the variation in $\hat{\mathbf{b}}_i - \mathbf{b}_i$ will be understated, because this expression ignores the variation of $\mathbf{b}_i$. We use instead

$$\text{var}(\hat{\mathbf{b}}_i - \mathbf{b}_i) = \mathbf{D} - \mathbf{D}\mathbf{Z}_i^T\mathbf{W}_i\mathbf{Z}_i\mathbf{D} + \mathbf{D}\mathbf{Z}_i^T\mathbf{W}_i\mathbf{X}_i\left(\sum_1^m \mathbf{X}_i^T\mathbf{W}_i\mathbf{X}_i\right)^{-1}\mathbf{X}_i^T\mathbf{W}_i\mathbf{Z}_i\mathbf{D}. \tag{3.5}$$

These expressions for the variances, and related ones for covariances, are special cases of the general formulas given by Harville (1976).

3.2 *Unknown Variance*

When the covariance matrices are unknown, but an estimate of $\boldsymbol{\theta}$, and thus of $\mathbf{R}_i$ and $\mathbf{D}$, is available, it is natural to set $\hat{\mathbf{V}}_i = \hat{\mathbf{R}}_i + \mathbf{Z}_i\hat{\mathbf{D}}\mathbf{Z}_i^T = \hat{\mathbf{W}}_i^{-1}$, and estimate $\boldsymbol{\alpha}$ and $\mathbf{b}_i$ by using the

weighted least squares equations (3.1) and (3.2), replacing each $\mathbf{W}_i$ by $\hat{\mathbf{W}}_i$. We denote these estimates by $\hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\theta}})$ and $\hat{\mathbf{b}}_i(\hat{\boldsymbol{\theta}})$.

This approach arises naturally when we consider the estimation of $\boldsymbol{\alpha}$ and $\boldsymbol{\theta}$ simultaneously by maximizing their joint likelihood based on the marginal distribution of $\mathbf{y}$. It is easily shown that the ML estimates $(\hat{\boldsymbol{\alpha}}_M, \hat{\boldsymbol{\theta}}_M)$ satisfy $\hat{\boldsymbol{\alpha}}_M = \hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\theta}}_M)$. Further, let $\mathbf{b}^T = (\mathbf{b}_1^T, \mathbf{b}_2^T, \ldots, \mathbf{b}_m^T)$. Setting $\hat{\mathbf{b}}_M = \mathrm{E}(\mathbf{b} \mid \mathbf{y}, \hat{\boldsymbol{\alpha}}_M, \hat{\boldsymbol{\theta}}_M)$ gives $\hat{\mathbf{b}}_M = \hat{\mathbf{b}}(\hat{\boldsymbol{\theta}}_M)$, which is the empirical Bayes estimate for $\mathbf{b}$ when $\boldsymbol{\theta}$ is estimated by maximum likelihood. Thus, the intuitive approach is maximum likelihood for $\boldsymbol{\alpha}$ and empirical Bayes for $\mathbf{b}$.

Estimates of the standard errors of $\hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\theta}})$ and $\hat{\mathbf{b}}(\hat{\boldsymbol{\theta}})$ can be obtained by substituting $\hat{\boldsymbol{\theta}}$ in (3.3), (3.4) and (3.5). As noted by Dempster *et al.* (1981), methods of adjusting the resulting expressions to reflect the uncertainty arising from the use of $\hat{\boldsymbol{\theta}}$ rather than $\boldsymbol{\theta}$ would be useful, but such methods are not available at present.

Except for this problem of adjusting the standard errors of $\hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\theta}})$ and $\hat{\mathbf{b}}(\hat{\boldsymbol{\theta}})$, inference about $\boldsymbol{\alpha}$ and $\mathbf{b}$ is relatively straightforward. There seems to be general agreement in the literature on the use of $\hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\theta}})$ and $\hat{\mathbf{b}}(\hat{\boldsymbol{\theta}})$ for some choice of $\hat{\boldsymbol{\theta}}$. In addition, once $\hat{\boldsymbol{\theta}}$ is available, $\hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\theta}})$ and $\hat{\mathbf{b}}(\hat{\boldsymbol{\theta}})$ and expressions for their standard errors are obtained noniteratively. The more difficult problem is to obtain a good and easily computed estimate of $\boldsymbol{\theta}$.

3.3 *Estimating the Covariance Matrix*

The literature on the estimation of variance components is extensive, most of it in the context of ANOVA models. Harville (1977) reviews the state of the art, treating both the optimality of various estimates and their computation. We focus here on two competitive estimates, the ML estimate, $\hat{\boldsymbol{\theta}}_M$, and a relative, the restricted ML estimate (REML), $\hat{\boldsymbol{\theta}}_R$. We do so not only because they are leading candidates ($\hat{\boldsymbol{\theta}}_R$ especially), but also because (i) the use of either of these estimates leads to a unified approach to estimation for $\boldsymbol{\alpha}$, $\mathbf{b}$ and $\boldsymbol{\theta}$, and (ii) the use of the EM algorithm unifies the computation of $\hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\theta}})$, $\hat{\mathbf{b}}(\hat{\boldsymbol{\theta}})$ and $\hat{\boldsymbol{\theta}}$, whenever either $\hat{\boldsymbol{\theta}}_M$ or $\hat{\boldsymbol{\theta}}_R$ is used.

First consider the ML estimate, $\hat{\boldsymbol{\theta}}_M$. As pointed out previously, $\hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\theta}}_M)$ and $\hat{\boldsymbol{\theta}}_M$ jointly maximize the marginal likelihood of $(\boldsymbol{\alpha}, \boldsymbol{\theta})$, and $\hat{\mathbf{b}}(\hat{\boldsymbol{\theta}}_M)$ is the empirical Bayes estimate of $\mathbf{b}$, appropriate when we use $\hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\theta}}_M)$ and $\hat{\boldsymbol{\theta}}_M$ to estimate $\boldsymbol{\alpha}$ and $\boldsymbol{\theta}$. Thus, using ML for $\boldsymbol{\theta}$ leads to a unified approach for estimating $\boldsymbol{\alpha}$ and $\mathbf{b}$ as well. In §4 we discuss the use of the EM algorithm to calculate $\hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\theta}}_M)$, $\hat{\mathbf{b}}(\hat{\boldsymbol{\theta}}_M)$ and $\hat{\boldsymbol{\theta}}_M$.

In balanced ANOVA models, ML estimates of variance components fail to take into account the degrees of freedom lost in estimating $\boldsymbol{\alpha}$, and are thus biased downwards. The REML estimates are not biased. The REML estimate is obtained by maximizing the likelihood of $\boldsymbol{\theta}$ based, not on $\mathbf{y}$ as in maximum likelihood, but on any full-rank set of error contrasts, $\mathbf{u}^T\mathbf{y}$, chosen so that

$$\mathrm{E}(\mathbf{u}^T\mathbf{y}) = \mathbf{0}.$$

In balanced ANOVA models, the REML likelihood equations have the standard ANOVA (unbiased) estimates as their solution. Patterson and Thompson (1971) justify their use by giving sufficiency arguments of the type subsequently formalized by Sprott (1971).

It is not so straightforward to see that the use of $\hat{\boldsymbol{\theta}}_R$ leads to a unified approach to both estimation and computation of estimates of $\boldsymbol{\alpha}$ and $\mathbf{b}$. The REML estimate can be derived in at least two completely unrelated ways, one relying on the sampling theoretic arguments given above, and the other on a Bayesian approach. The sampling theoretic approach gives the much more well-known justification for REML, but the Bayesian

approach leads to the unified treatment of estimation and computation. Strictly speaking, it is unnecessary to understand the Bayesian approach to REML, but this approach is attractive because it clarifies the theoretical justification for $\hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\theta}}_R)$, $\hat{\mathbf{b}}(\hat{\boldsymbol{\theta}}_R)$ and $\hat{\boldsymbol{\theta}}_R$, and, more importantly, it leads to a simplified derivation of the likelihood equations and shows how to handle their computation. For this reason, we discuss the Bayesian approach to REML in §5, after a discussion of the use of the EM algorithm for ML estimation.

4. Using the EM Algorithm for ML Estimates

In their paper on maximum likelihood estimation with incomplete data, Dempster *et al.* (1977) noted that many iterative algorithms for computing maximum likelihood estimates are merely special cases of a very general computing algorithm called EM, applicable in a broadly defined incomplete-data setting. Both variance-component models and empirical Bayes models were discussed as incomplete-data problems in which the algorithm can be applied. Dempster *et al.* (1981) consider its application to covariance-component models as well. We outline here its use in the calculation of $\hat{\boldsymbol{\theta}}_M$, and also of $\hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\theta}}_M)$ and $\hat{\mathbf{b}}(\hat{\boldsymbol{\theta}}_M)$.

Laird (1981) shows that, when $\boldsymbol{\theta}$ is a vector of variance components ($\mathbf{R}_i = \sigma^2\mathbf{I}_{n_i \times n_i}$ and $\mathbf{D}$ is diagonal), the EM algorithm is equivalent to Henderson's algorithm for maximum likelihood, as described in Harville (1977). As Dempster *et al.* (1981) point out, the reasons for viewing the particular algorithm (Henderson's) in the more general form (EM) include the following: (i) special derivations are not required for other cases, such as covariance-components models; (ii) the general EM theory, which shows that each iteration increases the likelihood, can be applied; and (iii) the general expressions for defining the iterative steps of the algorithm have meaningful statistical interpretations which, as we show in §5, help to elucidate the distinction between REML and maximum likelihood estimates of $\boldsymbol{\theta}$.

To put the longitudinal-data problem in the context of incomplete-data, note that if we were to observe $\mathbf{b}_i$ and $\mathbf{e}_i$, in addition to $\mathbf{y}_i$, we could easily find simple closed-form maximum likelihood estimates of the components of $\boldsymbol{\theta}$, based on quadratic forms in $\mathbf{b}_i$ and $\mathbf{e}_i$, $i = 1, \ldots, m$. For example, if $\mathbf{R}_i = \sigma^2\mathbf{I}_{n_i \times n_i}$ and $\mathbf{D}$ is an arbitrary $k \times k$ nonnegative-definite matrix, we would use

$$\hat{\sigma}^2 = \sum_1^m \mathbf{e}_i^T\mathbf{e}_i \Big/ \sum_1^m n_i = t_1 \Big/ \sum_1^m n_i \tag{4.1}$$

and

$$\hat{\mathbf{D}} = m^{-1}\sum_1^m \mathbf{b}_i\mathbf{b}_i^T = \mathbf{t}_2/m, \tag{4.2}$$

the 'sufficient statistics' for $\boldsymbol{\theta}$ being t_1 and the $\frac{1}{2}k(k+1)$ nonredundant components of $\mathbf{t}_2$.

If an estimate of $\boldsymbol{\theta}$ is available, we can use it to calculate 'estimates' of the missing 'sufficient statistics', by setting them equal to their expectations, conditional on the observed data vector $\mathbf{y}$. Letting $\hat{\boldsymbol{\theta}}$ denote the estimate of $\boldsymbol{\theta}$, $\hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\theta}})$ and $\hat{\mathbf{b}}_i(\hat{\boldsymbol{\theta}})$ the corresponding estimates of $\boldsymbol{\alpha}$ and $\mathbf{b}_i$, and $\hat{t}_1$ and $\hat{\mathbf{t}}_2$ the 'estimated sufficient statistics', we have

$$\begin{aligned}\hat{t}_1 &= \mathrm{E}\left\{\sum_1^m \mathbf{e}_i^T\mathbf{e}_i \mid \mathbf{y}_i, \hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\theta}}), \hat{\boldsymbol{\theta}}\right\} \\ &= \sum_1^m [\hat{\mathbf{e}}_i(\hat{\boldsymbol{\theta}})^T\hat{\mathbf{e}}_i(\hat{\boldsymbol{\theta}}) + \operatorname{tr} \operatorname{var}\{\mathbf{e}_i \mid \mathbf{y}_i, \hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\theta}}), \hat{\boldsymbol{\theta}}\}],\end{aligned} \tag{4.3}$$

and

$$\hat{\mathbf{t}}_2 = \mathrm{E}\left\{\sum_1^m \mathbf{b}_i\mathbf{b}_i^T \mid \mathbf{y}_i, \hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\theta}}), \hat{\boldsymbol{\theta}}\right\}$$

$$= \sum_{i=1}^m \{\hat{\mathbf{b}}_i(\hat{\boldsymbol{\theta}})\hat{\mathbf{b}}_i(\hat{\boldsymbol{\theta}})^T + \mathrm{var}(\mathbf{b}_i \mid \mathbf{y}_i, \hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\theta}}), \hat{\boldsymbol{\theta}})\}, \tag{4.4}$$

where $\hat{\mathbf{e}}_i(\hat{\boldsymbol{\theta}}) = \mathrm{E}(\mathbf{e}_i \mid \mathbf{y}_i, \hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\theta}}), \hat{\boldsymbol{\theta}}) = \mathbf{y}_i - \mathbf{X}_i\hat{\alpha}(\hat{\boldsymbol{\theta}}) - \mathbf{Z}_i\hat{\mathbf{b}}_i(\hat{\boldsymbol{\theta}})$.

To obtain the ML estimate, $\hat{\boldsymbol{\theta}}_M$, we start with any suitable initial value for $\hat{\boldsymbol{\theta}}$, and thus for $\hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\theta}})$, then iterate between (4.3) and (4.4), which define the E-step, and (4.1) and (4.2), which define the M-step. At convergence we have not only $\hat{\boldsymbol{\theta}}_M$, but $\hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\theta}}_M)$ and $\mathbf{b}(\hat{\boldsymbol{\theta}}_M)$, from calculation of the last E-step.

For other models applied to longitudinal data, $\sum_1^m \mathbf{e}_i^T\mathbf{e}_i$ and $\sum_1^m \mathbf{b}_i\mathbf{b}_i^T$ are replaced by the appropriate 'sufficient statistics', which, depending on $\boldsymbol{\theta}$, are generally quadratic forms in $\mathbf{b}_i$ and $\mathbf{e}_i$. Denoting these sufficient statistics by $\mathbf{t}$, (4.1) and (4.2) (defining the M-step of the EM algorithm) become

$$\text{M-step: } \hat{\boldsymbol{\theta}} = M(\mathbf{t}), \tag{4.5}$$

where M is the appropriately defined mapping which gives maximum likelihood estimates of $\boldsymbol{\theta}$ when $\mathbf{t}$ is observed. The general form of the E-step [(4.3) and (4.4)] becomes

$$\text{E-step: } \hat{\mathbf{t}} = \mathrm{E}\{\mathbf{t} \mid \mathbf{y}, \hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\theta}}), \hat{\boldsymbol{\theta}}\}. \tag{4.6}$$

A different application of the EM algorithm arises in multivariate normal missing-data problems (Dempster *et al.*, 1977). Here one proceeds by 'filling in' the missing observations. This technique is sometimes employed to create artificially a balanced design even when no data are missing (Healy and Westmacott, 1956; Kleinbaum, 1973). Our approach here both to modelling and computation is fundamentally different. We regard no data as missing; we use the EM algorithm to 'estimate' unobservable (i.e. random) parameters, not missing observations.

In the discussion of Dempster *et al.* (1977), concerns were raised about the slow convergence of the EM algorithm, its sensitivity to starting values, and its convergence to a local, rather than a global, maximum. Examples involving its application to missing-data problems, mixture problems and factor analysis were given. In applications of the two-stage model, we have often experienced slow convergence of the estimates of variance components. The problem of slow convergence will be especially severe if the maximum likelihood occurs on or near a boundary of the parameter space. Further work on the convergence properties of the algorithm is desirable before it can be universally endorsed.

5. REML Estimation and Computation

For reasons noted earlier, REML estimates of variance components are generally preferable to ML estimates. In this section, we discuss the Bayesian interpretation of REML, showing how it leads to a unified approach to estimation of $\boldsymbol{\alpha}$, $\mathbf{b}$ and $\boldsymbol{\theta}$, and to computation of their estimates.

Consider the following 'Bayesian' formulation of the general two-stage model presented in §2. Stage 1 remains unchanged so that, conditional on $\boldsymbol{\alpha}$ and $\mathbf{b}_i$, $\mathbf{y}_i$ is normal with mean $\mathbf{X}_i\boldsymbol{\alpha} + \mathbf{Z}_i\mathbf{b}_i$ and variance $\mathbf{R}_i$. At Stage 2, we let $\boldsymbol{\alpha}$ and each $\mathbf{b}_i$ be independent and normally distributed with mean vectors equal to $\mathbf{0}$ and $\mathrm{var}(\boldsymbol{\alpha}) = \boldsymbol{\Gamma}$, $\mathrm{var}(\mathbf{b}_i) = \mathbf{D}$ and $\mathrm{cov}(\boldsymbol{\alpha}, \mathbf{b}_i) = \mathbf{0}$, for $i = 1, \ldots, m$. Marginally, we now have

$$\mathbf{y}_i \sim \mathrm{N}(\mathbf{0}, \mathbf{X}_i\boldsymbol{\Gamma}\mathbf{X}_i^T + \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^T + \mathbf{R}_i).$$

We continue to let $\boldsymbol{\theta}$ denote the unknown parameters in $\mathbf{D}$ and $\mathbf{R}_i$, $i = 1, \ldots, m$.

If $\boldsymbol{\theta}$ and $\boldsymbol{\Gamma}$ were known, 'Bayesian' estimates for $\boldsymbol{\alpha}$ and $\mathbf{b}$ could be obtained as their posterior expectations, given $\mathbf{y}$, $\boldsymbol{\theta}$ and $\boldsymbol{\Gamma}$. With $\boldsymbol{\Gamma}$ and $\boldsymbol{\theta}$ unknown, an empirical Bayes approach would replace $\boldsymbol{\Gamma}$ and $\boldsymbol{\theta}$ with estimates obtained by maximizing their marginal normal likelihood based on $\mathbf{y}$, integrating over $\boldsymbol{\alpha}$ and $\mathbf{b}$.

We have information in the data about $\boldsymbol{\theta}$, which models both between- and within-individual variation, and thus we can obtain an estimate of $\boldsymbol{\theta}$. Typically, however, we have no information about $\boldsymbol{\Gamma}$, which models variation in the population characteristics. Thus, we cannot estimate $\boldsymbol{\Gamma}$. A reasonable strategy is to let $\boldsymbol{\Gamma}^{-1} = \mathbf{0}$, indicating vague prior information about $\boldsymbol{\alpha}$, and use an estimate of $\boldsymbol{\theta}$ obtained by maximizing the limiting (as $\boldsymbol{\Gamma}^{-1} \to \mathbf{0}$) marginal likelihood of $\boldsymbol{\theta}$ given $\mathbf{y}$. Harville (1976) demonstrates that this limiting likelihood is precisely equivalent to the REML likelihood. Thus, the estimates of $\boldsymbol{\theta}$ obtained in this setting are the REML estimates.

The expressions for the posterior means of $\boldsymbol{\alpha}$ and $\mathbf{b}$, given that $\boldsymbol{\Gamma}^{-1} = \mathbf{0}$ and $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_R$, are simply $\hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\theta}}_R)$ and $\hat{\mathbf{b}}(\hat{\boldsymbol{\theta}}R)$. That is,

$$\mathrm{E}(\boldsymbol{\alpha} \mid \mathbf{y}, \boldsymbol{\Gamma}^{-1} = \mathbf{0}, \hat{\boldsymbol{\theta}}_R) = \hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\theta}}_R)$$

and

$$\mathrm{E}(\mathbf{b}_i \mid \mathbf{y}_i, \boldsymbol{\Gamma}^{-1} = \mathbf{0}, \hat{\boldsymbol{\theta}}_R) = \mathrm{E}\{\mathbf{b}_i \mid \mathbf{y}_i, \hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\theta}}_R), \hat{\boldsymbol{\theta}}_R\} = \hat{\mathbf{b}}_i(\hat{\boldsymbol{\theta}}_R).$$

These equations follow from the equivalence of ML estimates and posterior means for parameters with flat priors. The estimators of $\boldsymbol{\alpha}$ and $\mathbf{b}$ have the algebraic form previously derived, but the estimates may differ slightly because of the use of $\hat{\boldsymbol{\theta}}_R$ rather than $\hat{\boldsymbol{\theta}}_M$ in the matrix of weights.

Using this empirical Bayes approach to REML, it is straightforward to show how to use EM to calculate $\hat{\boldsymbol{\theta}}_R$, $\hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\theta}}_R)$ and $\hat{\mathbf{b}}(\hat{\boldsymbol{\theta}}_R)$. For estimating $\boldsymbol{\theta}$, the 'complete data' still consist of $\mathbf{y}, \mathbf{b}$ and $\mathbf{e}$; thus, the M-step [(4.1) and (4.2), or (4.5)] remains the same. The difference between ML and REML comes in the E-step [(4.3) and (4.4), or (4.6)]. Here, expectations for ML were taken conditional on $\mathbf{y}$ and $\boldsymbol{\alpha}$. With REML, these expectations are conditional on $\mathbf{y}$ only, because $\boldsymbol{\alpha}$ has been integrated out of the likelihood. Thus, (4.3) and (4.4) become

$$\hat{t}_1 = \mathrm{E}\left(\sum_1^m \mathbf{e}_i^T \mathbf{e}_i \mid \mathbf{y}_i, \hat{\boldsymbol{\theta}}\right) = \sum_1^m \{\hat{\mathbf{e}}_i(\hat{\boldsymbol{\theta}})^T \mathbf{e}_i(\hat{\boldsymbol{\theta}}) + \mathrm{tr}\,\mathrm{var}(\mathbf{e}_i \mid \mathbf{y}_i, \hat{\boldsymbol{\theta}})\} \tag{5.1}$$

and

$$\hat{\mathbf{t}}_2 = \mathrm{E}\left(\sum_1^m \mathbf{b}_i \mathbf{b}_i^T \mid \mathbf{y}_i, \hat{\boldsymbol{\theta}}\right) = \sum_1^m \{\hat{\mathbf{b}}_i(\hat{\boldsymbol{\theta}})\hat{\mathbf{b}}_i(\hat{\boldsymbol{\theta}})^T + \mathrm{var}(\mathbf{b}_i \mid \mathbf{y}_i, \hat{\boldsymbol{\theta}})\}. \tag{5.2}$$

Here $\hat{\mathbf{e}}_i(\hat{\boldsymbol{\theta}})$ still equals $\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\alpha}(\hat{\boldsymbol{\theta}}) - \mathbf{Z}_i\hat{\mathbf{b}}_i(\hat{\boldsymbol{\theta}})$. In general, the E-step becomes

$$\hat{\mathbf{t}} = \mathrm{E}(\mathbf{t} \mid \mathbf{y}, \hat{\boldsymbol{\theta}}). \tag{5.3}$$

Note that the expectations computed at the E-step involve the conditional means and variances of $\mathbf{b}_i$ and $\mathbf{e}_i$. As noted earlier, the conditional means of $\mathbf{b}_i$ and $\mathbf{e}_i$ are the same, regardless of whether we take $\boldsymbol{\alpha}$ fixed, or random with infinite variance. The conditional variances are different, and this fact illustrates why ML estimates of variance components are biased downwards. Because we are dealing with normal expectations, neither $\mathrm{var}(\mathbf{b}_i \mid \mathbf{y}_i, \boldsymbol{\alpha}, \boldsymbol{\theta})$ nor $\mathrm{var}(\mathbf{e}_i \mid \mathbf{y}_i, \boldsymbol{\alpha}, \boldsymbol{\theta})$ depends upon $\boldsymbol{\alpha}$. This implies that

$$\mathrm{var}(\mathbf{b}_i \mid \mathbf{y}_i, \boldsymbol{\theta}) = \mathrm{var}(\mathbf{b}_i \mid \mathbf{y}_i, \boldsymbol{\alpha}, \boldsymbol{\theta}) + \mathrm{var}\{\mathrm{E}(\mathbf{b}_i \mid \mathbf{y}_i, \boldsymbol{\alpha}, \boldsymbol{\theta})\},$$

and likewise for the $\mathbf{e}_i$. Thus the expectations computed at the E-step for ML, (4.3) and (4.4), are smaller than the corresponding quantities for REML, (5.1) and (5.2), which will lead to smaller estimated variance components.

There is one additional feature of this Bayesian approach which enhances its attractiveness, and also that of using the EM algorithm. Harville (1976) has shown that, if we use

$\text{var}(\boldsymbol{\alpha} \mid \mathbf{y}, \hat{\boldsymbol{\theta}}_R)$ and $\text{var}(\mathbf{b} \mid \mathbf{y}, \hat{\boldsymbol{\theta}}_R)$ to assign estimated variances to $\hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\theta}}_R)$ and $\hat{\mathbf{b}}(\hat{\boldsymbol{\theta}}_R)$, then these quantities are the same as the sampling-theory variances, $\text{var}(\hat{\boldsymbol{\alpha}})$ and $\text{var}(\hat{\mathbf{b}}-\mathbf{b})$, given in (3.3) and (3.5), with $\hat{\boldsymbol{\theta}}_R$ substituted for $\boldsymbol{\theta}$. Since $\text{var}(\mathbf{b} \mid \mathbf{y}, \hat{\boldsymbol{\theta}})$ must be computed at the E-step, it is available at the final iteration, as is $\hat{\mathbf{b}}(\hat{\boldsymbol{\theta}}_R)$. The estimate of $\boldsymbol{\alpha}$, $\hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\theta}}_R)$, and its variance, $\text{var}(\boldsymbol{\alpha} \mid \mathbf{y}, \hat{\boldsymbol{\theta}}_R)$ need not be explicitly computed at the E-step, but are readily available from the other quantities computed at this step. Thus $\hat{\boldsymbol{\theta}}_R$, $\hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\theta}}_R)$, $\hat{\mathbf{b}}(\hat{\boldsymbol{\theta}}_R)$, $\text{var}_{\text{est}}(\hat{\boldsymbol{\alpha}})$ and $\text{var}_{\text{est}}(\hat{\mathbf{b}}-\mathbf{b})$ are calculable, either directly or indirectly, using the EM algorithm.

6. Examples

In this section, we illustrate the application of the general model to two data-analysis problems arising in the study of effects of atmospheric pollutants on pulmonary function. The first problem has features of a repeated-measures design, and the second is closely related to growth-curve analysis. We show how each of these models is represented in the general linear model and demonstrate the advantages of a more general formulation. A detailed comparison of several approaches to the analysis of Example 1 is the subject of a forthcoming report.

Example 1: *Analysis of the effect of air pollution episodes on pulmonary function.* Approximately 200 school children were examined under normal conditions, then during an air pollution alert and on three successive weeks following the alert. The objective was to determine whether FEV_1, the volume of air exhaled in the first second of a forced exhalation, was depressed during the alert. A secondary objective was the identification of sensitive subgroups or individuals most severely affected by the pollution episode.

The simplest model for these data is a two-stage mixed model, corresponding to the repeated-measures design. If $\mathbf{y}_i$ is the 5×1 vector of FEV_1 values for the ith child,

$$\mathbf{y}_i = \mathbf{I}\boldsymbol{\alpha} + \mathbf{1}\mathbf{b}_i + \mathbf{e}_i, \qquad i = 1, 2, \ldots, m, \tag{6.1}$$

where $\mathbf{I}$ is the 5×5 identity matrix and $\mathbf{1}$ is a 5×1 vector of '1's. The vector $\boldsymbol{\alpha}$ contains the population mean FEV_1 values on the five days, and b_i is the random deviation in average FEV_1 value for the ith child. We assume $\mathbf{e}_i \sim \text{N}(\mathbf{0}, \sigma^2\mathbf{I})$ and $b_i \sim \text{N}(0, \tau^2)$. When some measurements are missed, the model for $\mathbf{y}_i$ ($n_i \times 1$) is the natural modification of (6.1).

The data were analyzed using the methods described in §5 and by standard multivariate methods, with the following findings: (i) a decline in mean FEV_1 was observed on and after the alert day; (ii) the variances and covariances for the last four measurements were larger than those involving the baseline day.

The increased variability on and after the alert day is consistent with the hypothesis that individuals respond differently to the exposure. The notion of sensitive individuals is an important idea in the air pollution literature. We can introduce a second random effect to quantify the average decline in FEV_1 for each child. Let

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\alpha} + \mathbf{Z}_i\mathbf{b}_i + \mathbf{e}_i,$$

where $\mathbf{X}_i$, $\boldsymbol{\alpha}$ and $\mathbf{e}_i$ are defined as before, but $\mathbf{b}_i^T = (b_{1i}, b_{2i})$ and

$$\mathbf{Z}_i = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \end{bmatrix}$$

is $n_i \times 2$.

The second stage of the model is $\boldsymbol{\alpha} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Gamma})$, $\mathbf{b}_i \sim \mathrm{N}(\mathbf{0}, \mathbf{D})$ with $\boldsymbol{\Gamma}^{-1} \to \mathbf{0}$. Then b_{1i} is the individual effect for the ith child at the baseline examination, and b_{2i} measures the average deviation from that value for the alert and post-alert examinations. A negative value for $\hat{b}_{2i}$ implies a larger decline in FEV_1 for the ith child than for the sample as a whole. The estimates of b_{2i} were used to identify children who showed greatest declines, and 20 children were identified for further study and review of previous examinations. Many of these children had developed a cold during the follow-up period. The ability to summarize individual responses in a way that is useful for exploratory analysis is an important feature of random-effects models.

To study the influence of sex, race, location of residence and other individual characteristics on response, we defined a model including these factors. If $\mathbf{v}_i$ is the vector of individual characteristics for the ith individual, let b_{1i} and b_{2i} have linear regressions on these factors, $b_{1i} = \mathbf{v}_i^T \boldsymbol{\gamma}_1 + b'_{1i}$ and $b_{2i} = \mathbf{v}_i^T \boldsymbol{\gamma}_2 + b'_{2i}$, where b'_{1i} and b'_{2i} are the individual deviations from the regression model. Then

$$\begin{aligned}\mathbf{b}_i &= \begin{bmatrix} \mathbf{v}_i^T & \mathbf{0} \\ \mathbf{0} & \mathbf{v}_i^T \end{bmatrix} \begin{bmatrix} \boldsymbol{\gamma}_1 \\ \boldsymbol{\gamma}_2 \end{bmatrix} + \mathbf{b}'_i \\ &= \mathbf{V}_i \boldsymbol{\gamma} + \mathbf{b}'_i,\end{aligned}$$

and the general model for $\mathbf{y}_i$ is

$$\begin{aligned}\mathbf{y}_i &= \mathbf{X}_i \boldsymbol{\alpha} + \mathbf{Z}_i \mathbf{b}_i + \mathbf{e}_i \\ &= \mathbf{X}_i \boldsymbol{\alpha} + \mathbf{Z}_i (\mathbf{V}_i \boldsymbol{\gamma} + \mathbf{b}'_i) + \mathbf{e}_i \\ &= [\mathbf{X}_i, \mathbf{Z}_i \mathbf{V}_i] \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\gamma} \end{bmatrix} + \mathbf{Z}_i \mathbf{b}'_i + \mathbf{e}_i,\end{aligned}$$

a new model in the general family. This development illustrates how we can begin with a repeated-measures analysis and continue naturally to models that require the more general family.

Example 2: *Analyzing the effect of air pollutants on pulmonary function development.* In a study related to that of Example 1, about 2000 children are examined annually in each of six cities to determine the influence of tobacco smoke and fossil-fuel combustion products on the level and rate of development of pulmonary function. For specificity, we continue with FEV_1 as the response.

The principal factor influencing FEV_1 is body size, which we represent here by height. If the ith child has a vector, $\mathbf{y}_i$, of responses, the growth-curve model assumes that each child also has a vector of growth-curve parameters, $\mathbf{b}_i^*$, and a growth model $\mathbf{y}_i \mid \mathbf{b}_i^* = \mathbf{X}_i \mathbf{b}_i^* + \mathbf{e}_i$ at Stage 1, and $\mathbf{b}_i^* \sim \mathrm{N}(\boldsymbol{\beta}, \mathbf{D})$ at Stage 2. If a polynomial growth curve is assumed, the matrix $\mathbf{X}_i$ will contain a column of '1's and various powers of the heights at different examinations. Since the pattern of growth, and the number and timing of visits, vary among children, the design will be unbalanced.

If $\mathbf{b}_i = \mathbf{b}_i^* - \boldsymbol{\beta}$ centers the individual effects at $\mathbf{0}$, the model can be expressed as

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{X}_i \mathbf{b}_i + \mathbf{e}_i.$$

Relating this to the model (2.1), growth models implicitly require that the column space of $\mathbf{X}_i$, the matrix linking population parameters to $\mathbf{y}_i$, is contained in the column space of $\mathbf{Z}_i$, the matrix linking individual effects to $\mathbf{y}_i$. This is not always desirable. For instance, we may want to fit a saturated model to the population growth curve and a very simple model to the individual deviations.

We may assume further, following Grizzle and Allen (1969), that the values of the growth parameters depend linearly on a vector of individual characteristics $\mathbf{v}_i$, through the relationship $\mathbf{b}_i^* = \mathbf{V}_i\boldsymbol{\gamma} + \mathbf{b}_i$, where

$$\mathbf{V}_i = \begin{bmatrix} \mathbf{v}_i^T & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{v}_i^T & \mathbf{0} \cdots & \\ \vdots & \vdots & & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{v}_i^T \end{bmatrix}.$$

Then the model can be expressed as

$$\mathbf{Y}_i = \mathbf{X}_i\mathbf{V}_i\boldsymbol{\gamma} + \mathbf{X}_i\mathbf{b}_i + \mathbf{e}_i,$$

and still belongs to the family of growth curves. However, if changes in some of the individual characteristics, such as air pollution concentration, stove type, or smoking in the home, can alter the expected rate of growth, the general linear model is required for representation of this relationship.

7. Discussion

The principal advantage of the general linear model (2.1) is the ability to treat a variety of important special problems in a unified way. Although readers may disagree about the relative appeal of the Bayes and non-Bayes approaches to parameter estimation, practically speaking this choice is often of secondary importance and either approach can be accommodated in the general theory.

We do see room for further technical development in methods of implementation. Methods are needed for adjusting the estimates of var($\hat{\boldsymbol{\alpha}}$) and var($\hat{\mathbf{b}}$) to account for the uncertainty in $\hat{\boldsymbol{\theta}}$. The EM algorithm is a powerful concept for simplifying the computation of parameter estimates and likelihoods, but more information on its speed of convergence, sensitivity to initial estimates, and convergence to boundary points or nonglobal maxima is necessary before it can be universally endorsed. Since the main computational burden is the iterative computation of $\hat{\boldsymbol{\theta}}$, investigation of the properties of noniterative alternatives could be useful.

Résumé

Les modèles d'analyse de données longitudinales doivent prendre en compte la relation entre les observations faites en série sur une même unité expérimentale. Les modèles multivariates avec une structure de covariance générale sont souvent difficiles à appliquer à des données fortement déséquilibrées, alors que les modèles à deux niveaux d'effets aléatoires sont d'un emploi aisé. Dans ces modèles, les distributions de probabilité des vecteurs réponse d'individus différents appartiennent à une seule famille, mais quelques paramètres des effets aléatoires varient entre les individus, avec une distribution qui est spécifiée pour le second niveau. Une famille générale de modèles est discutée, comprenant les modèles de croissance aussi que les modèles des mesures répétées comme des cas particuliers. On discute une approche unifiée à l'adjustement de ces modèles, fondée sur une combinaison de l'estimation bayesienne empirique et la méthode du maximum de vraisemblance, utilisant l'algorithme EM. Deux exemples sont pris à un travail épidémiologique en cours, concernant les effets sur la santé de la pollution atmosphérique.

References

Beale, E. M. L. and Little, R. J. A. (1975). Missing values in multivariate analysis. *Journal of the Royal Statistical Society, Series B* **37,** 129–145.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood with incomplete data via the E–M algorithm. *Journal of the Royal Statistical Society, Series B* **39,** 1–38.

Dempster, A. P., Rubin, D. B. and Tsutakawa, R. K. (1981). Estimation in covariance component models. *Journal of the American Statistical Association* **76,** 341–353.

Fearn, T. (1975). A Bayesian approach to growth curves. *Biometrika* **62,** 89–100.

Grizzle, J. E. and Allen, D. M. (1969). Analysis of growth and dose response curves. *Biometrics* **25,** 357–382.

Harville, D. A. (1974). Bayesian inference for variance components using only error contrasts. *Biometrika* **61,** 383–385.

Harville, D. A. (1976). Extension of the Gauss–Markov theorem to include the estimation of random effects. *Annals of Statistics* **4,** 384–395.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association* **72,** 320–340.

Hayes, W. L. (1973). *Statistics for the Social Sciences.* New York: Holt, Rinehart and Winston.

Healy, M. and Westmacott, M. (1956). Missing values in experiments analyzed on automatic computers. *Applied Statistics* **5,** 203–206.

Kleinbaum, D. (1973). A generalization of the growth curve model which allows missing data. *Journal of Multivariate Analysis* **3,** 117–124.

Laird, N. M. (1982). Computation of variance components using the E–M algorithm. *Journal of Statistical Computation and Simulation* **14,** 295–303.

Orchard, T. and Woodbury, M. A. (1972). A missing information principle: theory and applications. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Vol.* I, L. M. LeCam, J. Neyman and E. L. Scott (eds), 697–715. Berkeley: University of California Press.

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

Rao, C. R. (1965). The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves. *Biometrika* **52,** 447–458.

Rao, C. R. (1967). Least squares theory using an estimated dispersion matrix and its application to measurement of signals. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Vol.* I, L. M. LeCam and J. Neyman (eds), 355–72. Berkeley: University of California Press.

Rao, C. R. (1975). Simultaneous estimation of parameters in different linear models and applications to biometric problems. *Biometrics* **31,** 545–554.

Rosenberg, B. (1973). Linear regression with randomly dispersed parameters. *Biometrika* **60,** 65–72.

Sprott, D. A. (1971). Marginal conditional sufficiency. *Biometrika* **62,** 599–605.

Ware, J. H. (1983). Growth curves. In *Encyclopedia of Statistical Sciences.*, S. Kotz, N. L. Johnson and C. B. Read (eds). New York: Wiley. To appear.

Received April 1981; *revised October* 1981