

1 Linear Mixed effects Models

A linear mixed effects (LME) model is a statistical model containing both fixed effects and random effects (random effects are also known as variance components). LME models are a generalization of the classical linear model, which contain fixed effects only. When the levels of factors are considered to be sampled from a population, and each level is not of particular interest, they are considered random quantities with associated variances. The effects of the levels, as described, are known as random effects. Random effects are represented by unobservable normally distributed random variables. Conversely fixed effects are considered non-random and the levels of each factor are of specific interest.

? introduced variance components models for use in genetical studies. Whereas an estimate for variance must take a non-negative value, an individual variance component, i.e. a component of the overall variance, may be negative.

The methodology has developed since, including contributions from ?, who extend the use of variance components into linear models, and ?, who introduced the ‘mixed model’ terminology and formally distinguished between mixed and random effects models. ? devised a methodology for deriving estimates for both the fixed effects and the random effects, using a set of equations that would become known as ‘mixed model equations’ or ‘Henderson’s equations’. LME methodology is further enhanced by Henderson’s later works (????). The key features of Henderson’s work provide the basis for the estimation techniques.

? demonstrated that unique estimates of the variance components could be obtained using maximum likelihood methods. However these estimates are known to be biased ‘downwards’ (i.e. underestimated), because of the assumption that the fixed estimates are known, rather than being estimated from the data. ? produced an alternative set of estimates, known as the restricted maximum likelihood (REML) estimates, that

do not require the fixed effects to be known. Thusly there is a distinction the REML estimates and the original estimates, now commonly referred to as ML estimates.

? provides a form of notation for notation for LME models that has since become the standard form, or the basis for more complex formulations. Due to computation complexity, linear mixed effects models have not seen widespread use until many well known statistical software applications began facilitating them. SAS Institute added PROC MIXED to its software suite in 1992 (?). ? described how to compute LME models in the S-plus environment.

Using Laird-Ware form, the LME model is commonly described in matrix form,

$$y = X\beta + Zb + \epsilon \quad (1)$$

where y is a vector of N observable random variables, β is a vector of p fixed effects, X and Z are $N \times p$ and $N \times q$ known matrices, and b and ϵ are vectors of q and N , respectively, random effects such that $E(b) = 0$, $E(\epsilon) = 0$ and

$$\text{var} \begin{pmatrix} b \\ \epsilon \end{pmatrix} = \begin{pmatrix} D & 0 \\ 0 & \Sigma \end{pmatrix}$$

where D and Σ are positive definite matrices parameterized by an unknown variance component parameter vector θ . The variance-covariance matrix for the vector of observations y is given by $V = ZDZ' + \Sigma$. This implies $y \sim (X\beta, V) = (X\beta, ZDZ' + \Sigma)$. It is worth noting that V is an $n \times n$ matrix, as the dimensionality becomes relevant later on. The notation provided here is generic, and will be adapted to accord with complex formulations that will be encountered in due course.

1.1 Estimation

Estimation of LME models involve two complementary estimation issues'; estimating the vectors of the fixed and random effects estimates $\hat{\beta}$ and \hat{b} and estimating the variance covariance matrices D and Σ . Inference about fixed effects have become known as 'estimates', while inferences about random effects have become known as 'predictions'.

The most common approach to obtain estimators are Best Linear Unbiased Estimator (BLUE) and Best Linear Unbiased Predictor (BLUP). For an LME model given by (1), the BLUE of $\hat{\beta}$ is given by

$$\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y,$$

whereas the BLUP of \hat{b} is given by

$$\hat{b} = DZ'V^{-1}(y - X\hat{\beta}).$$

1.1.1 Henderson's equations

Because of the dimensionality of V (i.e. $n \times n$) computing the inverse of V can be difficult. As a way around this problem, Henderson offered a more simpler approach of jointly estimating $\hat{\beta}$ and \hat{b} . Henderson made the (ad-hoc) distributional assumptions $y|b \sim N(X\beta + Zb, \Sigma)$ and $b \sim N(0, D)$, and proceeded to maximize the joint density of y and b

$$\left| \begin{matrix} D & 0 \\ 0 & \Sigma \end{matrix} \right|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \begin{pmatrix} b \\ y - X\beta - Zb \end{pmatrix}' \begin{pmatrix} D & 0 \\ 0 & \Sigma \end{pmatrix}^{-1} \begin{pmatrix} b \\ y - X\beta - Zb \end{pmatrix} \right\}, \quad (2)$$

with respect to β and b , which ultimately requires minimizing the criterion

$$(y - X\beta - Zb)' \Sigma^{-1} (y - X\beta - Zb) + b' D^{-1} b. \quad (3)$$

This leads to the mixed model equations

$$\begin{pmatrix} X'\Sigma^{-1}X & X'\Sigma^{-1}Z \\ Z'\Sigma^{-1}X & Z'\Sigma^{-1}Z + D^{-1} \end{pmatrix} \begin{pmatrix} \beta \\ b \end{pmatrix} = \begin{pmatrix} X'\Sigma^{-1}y \\ Z'\Sigma^{-1}y \end{pmatrix}. \quad (4)$$

Using these equations, obtaining the estimates requires the inversion of a matrix of dimension $p + q \times p + q$, considerably smaller in size than V . Henderson shows that these mixed model equations do not depend on normality and that $\hat{\beta}$ and \hat{b} are the BLUE and BLUP under general conditions, provided D and Σ are known.

Henderson points out that although Henderson initially referred to the estimates $\hat{\beta}$ and \hat{b} from (57) as “joint maximum likelihood estimates”, Henderson later advised that these estimates should

not be referred to as “maximum likelihood” as the function being maximized in (56) is a joint density rather than a likelihood function. ? remarks that it is clear that Henderson used joint estimation for computational purposes, without recognizing the theoretical implications.

1.1.2 Estimation of the fixed parameters

The vector y has marginal density $y \sim N(X\beta, V)$, where $V = \Sigma + ZDZ'$ is specified through the variance component parameters θ . The log-likelihood of the fixed parameters (β, θ) is

$$\ell(\beta, \theta | y) = -\frac{1}{2} \log |V| - \frac{1}{2} (y - X\beta)' V^{-1} (y - X\beta), \quad (5)$$

and for fixed θ the estimate $\hat{\beta}$ of β is obtained as the solution of

$$(X'V^{-1}X)\beta = X'V^{-1}y. \quad (6)$$

Substituting $\hat{\beta}$ from (59) into $\ell(\beta, \theta | y)$ from (58) returns the *profile* log-likelihood

$$\begin{aligned} \ell_P(\theta | y) &= \ell(\hat{\beta}, \theta | y) \\ &= -\frac{1}{2} \log |V| - \frac{1}{2} (y - X\hat{\beta})' V^{-1} (y - X\hat{\beta}) \end{aligned}$$

of the variance parameter θ . Estimates of the parameters θ specifying V can be found by maximizing $\ell_P(\theta | y)$ over θ . These are the ML estimates.

For REML estimation the *restricted* log-likelihood is defined as

$$\ell_R(\theta | y) = \ell_P(\theta | y) - \frac{1}{2} \log |X'VX|.$$

The REML approach does not base estimates on a maximum likelihood fit of all the information, but instead uses a likelihood function derived from a data set, transformed to remove the irrelevant influences (?). Restricted maximum likelihood is often preferred to maximum likelihood because REML estimation reduces the bias in the variance component by taking into account the loss of degrees of freedom that results from estimating the fixed effects in β . Restricted maximum likelihood also handles

high correlations more effectively, and is less sensitive to outliers than maximum likelihood. The problem with REML for model building is that the likelihoods obtained for different fixed effects are not comparable. Hence it is not valid to compare models with different fixed effects using a likelihood ratio test or AIC when REML is used to estimate the model. Therefore models derived using ML must be used instead.

1.1.3 Estimation of the random effects

The established approach for estimating the random effects is to use the best linear predictor of b from y , which for a given β equals $DZ'V^{-1}(y - X\beta)$. In practice β is replaced by an estimator such as $\hat{\beta}$ from (59) so that $\hat{b} = DZ'V^{-1}(y - X\hat{\beta})$. Pre-multiplying by the appropriate matrices it is straightforward to show that these estimates $\hat{\beta}$ and \hat{b} satisfy the equations in (57).

1.1.4 Algorithms for likelihood function optimization

Iterative numerical techniques are used to optimize the log-likelihood function and estimate the covariance parameters θ . The procedure is subject to the constraint that R and D are both positive definite. The most common iterative algorithms for optimizing the likelihood function are the Newton-Raphson method, which is the preferred method, the expectation maximization (EM) algorithm and the Fisher scoring methods.

The EM algorithm, introduced by ?, is an iterative technique for maximizing complicated likelihood functions. The algorithm alternates between performing an expectation (E) step and the maximization (M) step. The ‘E’ step computes the expectation of the log-likelihood evaluated using the current estimate for the variables. In the ‘M’ step, parameters that maximize the expected log-likelihood, found on the previous ‘E’ step, are computed. These parameter estimates are then used to determine the distribution of the variables in the next ‘E’ step. The algorithm alternatives between these two steps until convergence is reached.

The main drawback of the EM algorithm is its slow rate of convergence. Consequently the EM algorithm is rarely used entirely in LME estimation, instead providing an initial set of values that can be passed to other optimization techniques.

The Newton Raphson (NR) method is the most common, and recommended technique for ML and REML estimation. The NR algorithm minimizes an objective function defined as -2 times the log likelihood for the covariance parameters θ . At every iteration the NR algorithm requires the calculation of a vector of partial derivatives, known as the gradient, and the second derivative matrix with respect to the covariance parameters. This is known as the observed Hessian matrix. Due to the Hessian matrix, the NR algorithm is more time-consuming, but convergence is reached with fewer iterations compared to the EM algorithm. The Fisher scoring algorithm is a variant of the NR algorithm that is more numerically stable and likely to converge, but not recommended to obtain final estimates.

1.1.5 The extended likelihood

The desire to have an entirely likelihood-based justification for estimates of random effects, in contrast to Henderson's equation, has motivated ?, page 429 to define the *extended likelihood*. He remarks "In mixed effects modelling the extended likelihood has been called *h-likelihood* (for hierarchical likelihood) by ?, while in smoothing literature it is known as the *penalized likelihood* (e.g. ? ?)." The extended likelihood can be written $L(\beta, \theta, b|y) = p(y|b; \beta, \theta)p(b; \theta)$ and adopting the same distributional assumptions used by ? yields the log-likelihood function

$$\begin{aligned} \ell_h(\beta, \theta, b|y) = & -\frac{1}{2} \{ \log |\Sigma| + (y - X\beta - Zb)' \Sigma^{-1} (y - X\beta - Zb) \\ & + \log |D| + b' D^{-1} b \}. \end{aligned}$$

Given θ , differentiating with respect to β and b returns Henderson's equations in (57).

2 Repeated measurements in LME models

In many statistical analyzes, the need to determine parameter estimates where multiple measurements are available on each of a set of variables often arises. Further to ?, ? performs an analysis of the correlation of replicate measurements, for two variables of interest, using LME models.

Let y_{Aij} and y_{Bij} be the j th repeated observations of the variables of interest A and B taken on the i th subject. The number of repeated measurements for each variable may differ for each individual. Both variables are measured on each time points. Let n_i be the number of observations for each variable, hence $2 \times n_i$ observations in total.

It is assumed that the pair y_{Aij} and y_{Bij} follow a bivariate normal distribution.

$$\begin{pmatrix} y_{Aij} \\ y_{Bij} \end{pmatrix} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \text{ where } \boldsymbol{\mu} = \begin{pmatrix} \mu_A \\ \mu_B \end{pmatrix}$$

The matrix $\boldsymbol{\Sigma}$ represents the variance component matrix between response variables at a given time point j .

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_A^2 & \sigma_{AB} \\ \sigma_{AB} & \sigma_B^2 \end{pmatrix}$$

σ_A^2 is the variance of variable A , σ_B^2 is the variance of variable B and σ_{AB} is the covariance of the two variable. It is assumed that $\boldsymbol{\Sigma}$ does not depend on a particular time point, and is the same over all time points.

2.1 Formulation of the response vector

Information of individual i is recorded in a response vector \mathbf{y}_i . The response vector is constructed by stacking the response of the 2 responses at the first instance, then the 2 responses at the second instance, and so on. Therefore the response vector is a $2n_i \times 1$ column vector. The covariance matrix of \mathbf{y}_i is a $2n_i \times 2n_i$ positive definite matrix $\boldsymbol{\Omega}_i$.

Consider the case where three measurements are taken by both methods A and B ,

\mathbf{y}_i is a 6×1 random vector describing the i th subject.

$$\mathbf{y}_i = (y_i^{A1}, y_i^{B1}, y_i^{A2}, y_i^{B2}, y_i^{A3}, y_i^{B3})'$$

The response vector \mathbf{y}_i can be formulated as an LME model according to Laird-Ware form.

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i$$

$$\mathbf{b}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{D})$$

$$\boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_i)$$

Information on the fixed effects are contained in a three dimensional vector $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$. For computational purposes β_2 is conventionally set to zero. Consequently $\boldsymbol{\beta}$ is the solutions of the means of the two methods, i.e. $E(\mathbf{y}_i) = \mathbf{X}_i\boldsymbol{\beta}$. The variance covariance matrix \mathbf{D} is a general 2×2 matrix, while \mathbf{R}_i is a $2n_i \times 2n_i$ matrix.

2.2 Decomposition of the response covariance matrix

The variance covariance structure can be re-expressed in the following form,

$$\text{Cov}(\mathbf{y}_i) = \boldsymbol{\Omega}_i = \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i' + \mathbf{R}_i.$$

\mathbf{R}_i can be shown to be the Kronecker product of a correlation matrix \mathbf{V} and $\mathbf{\Lambda}$. The correlation matrix \mathbf{V} of the repeated measures on a given response variable is assumed to be the same for all response variables. Both ? and ? use the identity matrix, with dimensions $n_i \times n_i$ as the formulation for \mathbf{V} . ? remarks that, with repeated measures, the response for each subject is correlated for each variable, and that such correlation must be taken into account in order to produce a valid inference on correlation estimates. ? proposes various correlation structures may be assumed for repeated measure correlations, such as the compound symmetry and autoregressive structures, as alternative to the identity matrix.

However, for the purposes of method comparison studies, the necessary estimates are currently only determinable when the identity matrix is specified, and the results in ? indicate its use.

For the response vector described, ? presents a detailed covariance matrix. A brief summary shall be presented here only. The overall variance matrix is a 6×6 matrix composed of two types of 2×2 blocks. Each block represents one separate time of measurement.

$$\mathbf{\Omega}_i = \begin{pmatrix} \mathbf{\Sigma} & \mathbf{D} & \mathbf{D} \\ \mathbf{D} & \mathbf{\Sigma} & \mathbf{D} \\ \mathbf{D} & \mathbf{D} & \mathbf{\Sigma} \end{pmatrix}$$

The diagonal blocks are $\mathbf{\Sigma}$, as described previously. The 2×2 block diagonal matrix in $\mathbf{\Omega}$ gives $\mathbf{\Sigma}$. $\mathbf{\Sigma}$ is the sum of the between-subject variability \mathbf{D} and the within subject variability $\mathbf{\Lambda}$.

$\mathbf{\Omega}_i$ can be expressed as

$$\mathbf{\Omega}_i = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i' + (\mathbf{I}_{n_i} \otimes \mathbf{\Lambda}).$$

The notation dim_{n_i} means an $n_i \times n_i$ diagonal block.

2.3 Correlation terms

? demonstrated how the between-subject and within subject variabilities can be expressed in terms of correlation terms.

$$\mathbf{D} = \begin{pmatrix} \sigma_A^2 \rho_A & \sigma_A \sigma_B \rho_{AB} \delta \\ \sigma_A \sigma_B \rho_{AB} \delta & \sigma_B^2 \rho_B \end{pmatrix}$$

$$\mathbf{\Lambda} = \begin{pmatrix} \sigma_A^2 (1 - \rho_A) & \sigma_{AB} (1 - \delta) \\ \sigma_{AB} (1 - \delta) & \sigma_B^2 (1 - \rho_B) \end{pmatrix}.$$

ρ_A describe the correlations of measurements made by the method A at different times. Similarly ρ_B describe the correlation of measurements made by the method B at different times. Correlations among repeated measures within the same method are known as intra-class correlation coefficients. ρ_{AB} describes the correlation of measurements taken at the same same time by both methods. The coefficient δ is added for

when the measurements are taken at different times, and is a constant of less than 1 for linked replicates. This is based on the assumption that linked replicates measurements taken at the same time would have greater correlation than those taken at different times. For unlinked replicates δ is simply 1. ? provides a useful graphical depiction of the role of each correlation coefficients.

3 Limits of agreement in LME models

Limits of agreement are used extensively for assessing agreement, because they are intuitive and easy to use. Necessarily their prevalence in literature has meant that they are now the best known measurement for agreement, and therefore any newer methodology would benefit by making reference to them.

? uses LME models to determine the limits of agreement. Between-subject variation for method m is given by d_m^2 and within-subject variation is given by λ_m^2 . ? remarks that for two methods A and B , separate values of d_A^2 and d_B^2 cannot be estimated, only their average. Hence the assumption that $d_x = d_y = d$ is necessary. The between-subject variability \mathbf{D} and within-subject variability $\mathbf{\Lambda}$ can be presented in matrix form,

$$\mathbf{D} = \begin{pmatrix} d_A^2 & 0 \\ 0 & d_B^2 \end{pmatrix} = \begin{pmatrix} d^2 & 0 \\ 0 & d^2 \end{pmatrix}, \quad \mathbf{\Lambda} = \begin{pmatrix} \lambda_A^2 & 0 \\ 0 & \lambda_B^2 \end{pmatrix}.$$

The variance for method m is $d_m^2 + \lambda_m^2$. Limits of agreement are determined using the standard deviation of the case-wise differences between the sets of measurements by two methods A and B , given by

$$\text{var}(y_A - y_B) = 2d^2 + \lambda_A^2 + \lambda_B^2. \quad (7)$$

Importantly the covariance terms in both variability matrices are zero, and no covariance component is present.

? presents a data set ‘fat’, which is a comparison of measurements of subcutaneous fat by two observers at the Steno Diabetes Center, Copenhagen. Measurements are in millimeters (mm). Each person is measured three times by each observer. The observations are considered to be ‘true’ replicates.

A linear mixed effects model is formulated, and implementation through several software packages is demonstrated. All of the necessary terms are presented in the computer output. The limits of agreement are therefore,

$$0.0449 \pm 1.96 \times \sqrt{2 \times 0.0596^2 + 0.0772^2 + 0.0724^2} = (-0.220, 0.309). \quad (8)$$

? has demonstrated a methodology whereby d_A^2 and d_B^2 can be estimated separately. Also covariance terms are present in both \mathbf{D} and $\mathbf{\Lambda}$. Using Roy's methodology, the variance of the differences is

$$\text{var}(y_{iA} - y_{iB}) = d_A^2 + \lambda_B^2 + d_A^2 + \lambda_B^2 - 2(d_{AB} + \lambda_{AB}) \quad (9)$$

All of these terms are given or determinable in computer output. The limits of agreement can therefore be evaluated using

$$\bar{y}_A - \bar{y}_B \pm 1.96 \times \sqrt{\sigma_A^2 + \sigma_B^2 - 2(\sigma_{AB})}. \quad (10)$$

For Carstensen's 'fat' data, the limits of agreement computed using Roy's method are consistent with the estimates given by ?; $0.044884 \pm 1.96 \times 0.1373979 = (-0.224, 0.314)$.

3.1 Linked replicates

? proposes the addition of an random effects term to their model when the replicates are linked. This term is used to describe the 'item by replicate' interaction, which is independent of the methods. This interaction is a source of variability independent of the methods. Therefore failure to account for it will result in variability being wrongly attributed to the methods.

? introduces a second data set; the oximetry study. This study done at the Royal Childrens Hospital in Melbourne to assess the agreement between co-oximetry and pulse oximetry in small babies.

In most cases, measurements were taken by both method at three different times. In some cases there are either one or two pairs of measurements, hence the data is unbalanced. ? describes many of the children as being very sick, and with very low oxygen saturations levels. Therefore it must be assumed that a biological change can occur in interim periods, and measurements are not true replicates.

? demonstrate the necessity of accounting for linked replicated by comparing the limits of agreement from the 'oximetry' data set using a model with the additional term, and one without. When the interaction is accounted for the limits of agreement

are (-9.62,14.56). When the interaction is not accounted for, the limits of agreement are (-11.88,16.83). It is shown that the failure to include this additional term results in an over-estimation of the standard deviations of differences.

Limits of agreement are determined using Roy's methodology, without adding any additional terms, are found to be consistent with the 'interaction' model; (-9.562, 14.504). Roy's methodology assumes that replicates are linked. However, following Carstensen's example, an addition interaction term is added to the implementation of Roy's model to assess the effect, the limits of agreement estimates do not change. However there is a conspicuous difference in within-subject matrices of Roy's model and the modified model (denoted 1 and 2 respectively);

$$\hat{\mathbf{\Lambda}}_1 = \begin{pmatrix} 16.61 & 11.67 \\ 11.67 & 27.65 \end{pmatrix} \quad \hat{\mathbf{\Lambda}}_2 = \begin{pmatrix} 7.55 & 2.60 \\ 2.60 & 18.59 \end{pmatrix}. \quad (11)$$

(The variance of the additional random effect in model 2 is 3.01.)

? introduces the Akaike information criterion (*AIC*), a model selection tool based on the likelihood function. Given a data set, candidate models are ranked according to their AIC values, with the model having the lowest AIC being considered the best fit. Two candidate models can said to be equally good if there is a difference of less than 2 in their AIC values.

The Akaike information criterion (AIC) for both models are $AIC_1 = 2304.226$ and $AIC_2 = 2306.226$, indicating little difference in models. The AIC values for the Carstensen 'unlinked' and 'linked' models are 1994.66 and 1955.48 respectively, indicating an improvement by adding the interaction term.

The $\hat{\mathbf{\Lambda}}$ matrices are informative as to the difference between Carstensen's unlinked and linked models. For the oximetry data, the covariance terms (given above as 11.67 and 2.6 respectively) are of similar magnitudes to the variance terms. Conversely for the 'fat' data the covariance term (-0.00032) is negligible. When the interaction term is added to the model, the covariance term remains negligible. (For the 'fat' data, the difference in AIC values is also approximately 2).

To conclude, Carstensen's models provided a rigorous way to determine limits of agreement, but don't provide for the computation of \hat{D} and $\hat{\Lambda}$. Therefore the test's proposed by ? can not be implemented. Conversely, accurate limits of agreement as determined by Carstensen's model may also be found using Roy's method. Addition of the interaction term erodes the capability of Roy's methodology to compare candidate models, and therefore shall not be adopted.

Finally, to complement the blood pressure (i.e. 'J vs S') method comparison from the previous section (i.e. 'J vs S'), the limits of agreement are $15.62 \pm 1.96 \times 20.33 = (-24.22, 55.46)$.

4 Extension of Roy's methodology

Roy's methodology is constructed to compare two methods in the presence of replicate measurements. Necessarily it is worth examining whether this methodology can be adapted for different circumstances.

An implementation of Roy's methodology, whereby three or more methods are used, is not feasible due to computational restrictions. Specifically there is a failure to reach convergence before the iteration limit is reached. This may be due to the presence of additional variables, causing the problem of non-identifiability. In the case of two variables, it is required to estimate two variance terms and four correlation terms, six in all. For the case of three variabilities, three variance terms must be estimated as well as nine correlation terms, twelve in all. In general for n methods has $2 \times T_n$ variance terms, where T_n is the triangular number for n , i.e. the addition analogue of the factorial. Hence the computational complexity quite increases substantially for every increase in n .

Should an implementation be feasible, further difficulty arises when interpreting the results. The fundamental question is whether two methods have close agreement so as to be interchangeable. When three methods are present in the model, the null hypothesis is that all three methods have the same variability relevant to the respective tests. The outcome of the analysis will either be that all three are interchangeable or that all three are not interchangeable.

The tests would not be informative as to whether any two of those three were interchangeable, or equivalently if one method in particular disagreed with the other two. Indeed it is easier to perform three pair-wise comparisons separately and then to combine the results.

Roy's methodology is not suitable for the case of single measurements because it follows from the decomposition for the covariance matrix of the response vector y_i , as presented in ?. The decomposition depends on the estimation of correlation terms, which would be absent in the single measurement case. Indeed there can be no within-

subject variability if there are no repeated terms for it to describe. There would only be the covariance matrix of the measurements by both methods, which doesn't require the use of LME models. To conclude, simpler existing methodologies, such as Deming regression, would be the correct approach where there only one measurements by each method.

5 Conclusion

? and ? highlight the need for method comparison methodologies suitable for use in the presence of replicate measurements. ? presents a comprehensive methodology for assessing the agreement of two methods, for replicate measurements. This methodology has the added benefit of overcoming the problems of unbalanced data and unequal numbers of replicates. Implementation of the methodology, and interpretation of the results, is relatively easy for practitioners who have only basic statistical training. Furthermore, it can be shown that widely used existing methodologies, such as the limits of agreement, can be incorporated into Roy's methodology.

6 Linear Mixed effects Models

A linear mixed effects (LME) model is a statistical model containing both fixed effects and random effects, also known as variance components. LME models are a generalization of the classical linear model, which contain fixed effects only. When the levels of factors are considered to be sampled from a population, and each such level is not of particular interest, they are considered random quantities with associated variances. The effects of the levels, as described, are known as random effects. Random effects are represented by unobservable normally distributed random variables. Conversely fixed effects are considered non-random and the levels of each factor are of specific interest.

? introduces variance components models for use in genetical studies. Whereas an estimate for variance must take a non-negative value, an individual variance component, i.e. a component of the overall variance, may be negative.

The methodology has developed since, including contributions from ?, who extend the use of variance components into linear models, and ?, who introduced the ‘mixed model’ terminology and formally distinguished between mixed and random effects models. ? devised a methodology for deriving estimates for both the fixed effects and the random effects, using a set of equations that would become known as ‘mixed model equations’ or ‘Henderson’s equations’. LME methodology is further enhanced by Henderson’s later works (?????). The key features of Henderson’s work shall be discussed in depth in due course.

? demonstrated that unique estimates of the variance components could be obtained using maximum likelihood methods. These estimates would later become known as the maximum likelihood (ML) estimates. However these estimates are known to be biased downwards, because of the assumption that the fixed estimates are known, rather than being estimated from the data. ? produced an alternative set of estimates, known as the restricted maximum likelihood (REML) estimates, that do not require the fixed effects to be known. Both ML and REML estimation shall be reverted to in due course.

? provides a form of notation for notation for LME models that has since become

the standard form, or the basis for more complex formulations. Due to computation complexity, linear mixed effects models have not seen widespread use until many well known statistical software applications began facilitating them. SAS Institute added PROC MIXED to its software suite in 1992 (?). ? described how to compute LME models in the R environment.

Using Laird-Ware form, the LME model is commonly described in matrix form,

$$y = X\beta + Zb + \epsilon \quad (12)$$

where y is a vector of N observable random variables, β is a vector of p fixed effects, X and Z are $N \times p$ and $N \times q$ known matrices, and b and ϵ are vectors of q and N , respectively, random effects such that $E(b) = 0$, $E(\epsilon) = 0$ and

$$\text{var} \begin{pmatrix} b \\ \epsilon \end{pmatrix} = \begin{pmatrix} D & 0 \\ 0 & \Sigma \end{pmatrix}$$

where D and Σ are positive definite matrices parameterized by an unknown variance component parameter vector θ . The notation provided here is generic, and will be adapted to accord with complex formulations that will be encountered in due course.

6.1 Likelihood and estimation

Likelihood is the hypothetical probability that an event that has already occurred would yield a specific outcome. Likelihood differs from probability in that probability refers to future occurrences, while likelihood refers to past known outcomes.

The likelihood function ($L(\theta)$) is a fundamental concept in statistical inference. It indicates how likely a particular population is to produce an observed sample. The set of values that maximize the likelihood function are considered to be optimal, and are used as the estimates of the parameters. For computational ease, it is common to use the logarithm of the likelihood function, known simply as the log-likelihood ($\ell(\theta)$).

6.1.1 Likelihood-based tools

Likelihood functions provide the basis for two important statistical concepts that shall be further referred to; the likelihood ratio test and the Akaike information criterion.

Likelihood ratio tests are a class of tests based on the comparison of the values of the likelihood functions of two candidate models. LRTs can be used to test hypotheses about covariance parameters or fixed effects parameters in the context of LMEs. The test statistic for the likelihood ratio test is the difference of the log-likelihood functions, multiplied by -2 . The probability distribution of the test statistic is approximated by the χ^2 distribution with $(\nu_1 - \nu_2)$ degrees of freedom, where ν_1 and ν_2 are the degrees of freedom of models 1 and 2 respectively.

? introduces the Akaike information criterion (AIC), a model selection tool based on the likelihood function. Given a data set, candidate models are ranked according to their AIC values, with the model having the lowest AIC being considered the best fit.

6.1.2 Likelihood estimation techniques

Maximum likelihood and restricted maximum likelihood have become the most common strategies for estimating the variance component parameter θ . Maximum likelihood estimation obtains parameter estimates by optimizing the likelihood function. To obtain ML estimate the likelihood is constructed as a function of the parameters in the specified LME model. The maximum likelihood estimates (MLEs) of the parameters are the values of the arguments that maximize the likelihood function. The REML approach is a variant of maximum likelihood estimation which does not base estimates on a maximum likelihood fit of all the information, but instead uses a likelihood function derived from a data set, transformed to remove the irrelevant influences (?).

Restricted maximum likelihood is often preferred to maximum likelihood because REML estimation reduces the bias in the variance component by taking into account the loss of degrees of freedom that results from estimating the fixed effects in β . Re-

restricted maximum likelihood also handles high correlations more effectively, and is less sensitive to outliers than maximum likelihood. The problem with REML for model building is that the likelihoods obtained for different fixed effects are not comparable. Hence it is not valid to compare models with different fixed effects using a likelihood ratio test or AIC when REML is used to estimate the model. Therefore models derived using ML must be used instead.

6.2 Henderson's equations

? made the (ad-hoc) distributional assumptions $y|b \sim N(X\beta + Zb, \Sigma)$ and $b \sim N(0, D)$, and proceeded to maximize the joint density of y and b

$$\left| \begin{pmatrix} D & 0 \\ 0 & \Sigma \end{pmatrix} \right|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \begin{pmatrix} b \\ y - X\beta - Zb \end{pmatrix}' \begin{pmatrix} D & 0 \\ 0 & \Sigma \end{pmatrix}^{-1} \begin{pmatrix} b \\ y - X\beta - Zb \end{pmatrix} \right\}, \quad (13)$$

with respect to β and b , which ultimately requires minimizing the criterion

$$(y - X\beta - Zb)' \Sigma^{-1} (y - X\beta - Zb) + b' D^{-1} b. \quad (14)$$

This leads to the mixed model equations

$$\begin{pmatrix} X' \Sigma^{-1} X & X' \Sigma^{-1} Z \\ Z' \Sigma^{-1} X & X' \Sigma^{-1} X + D^{-1} \end{pmatrix} \begin{pmatrix} \beta \\ b \end{pmatrix} = \begin{pmatrix} X' \Sigma^{-1} y \\ Z' \Sigma^{-1} y \end{pmatrix}. \quad (15)$$

? points out that although ? initially referred to the estimates $\hat{\beta}$ and \hat{b} from (57) as “joint maximum likelihood estimates”. ? later advised that these estimates should not be referred to as “maximum likelihood” as the function being maximized in (56) is a joint density rather than a likelihood function.

6.2.1 Estimation of the fixed parameters

The vector y has marginal density $y \sim N(X\beta, V)$, where $V = \Sigma + ZDZ'$ is specified through the variance component parameters θ . The log-likelihood of the fixed parameters (β, θ) is

$$\ell(\beta, \theta | y) = -\frac{1}{2} \log |V| - \frac{1}{2} (y - X\beta)' V^{-1} (y - X\beta), \quad (16)$$

and for fixed θ the estimate $\hat{\beta}$ of β is obtained as the solution of

$$(X'V^{-1}X)\beta = X'V^{-1}y. \quad (17)$$

Substituting $\hat{\beta}$ from (59) into $\ell(\beta, \theta|y)$ from (58) returns the *profile* log-likelihood

$$\begin{aligned} \ell_P(\theta | y) &= \ell(\hat{\beta}, \theta | y) \\ &= -\frac{1}{2} \log |V| - \frac{1}{2} (y - X\hat{\beta})' V^{-1} (y - X\hat{\beta}) \end{aligned}$$

of the variance parameter θ . Estimates of the parameters θ specifying V can be found by maximizing $\ell_P(\theta | y)$ over θ . These are the ML estimates. For REML estimation the *restricted* log-likelihood is defined as

$$\ell_R(\theta | y) = \ell_P(\theta | y) - \frac{1}{2} \log |X'VX|.$$

6.2.2 Estimation of the random effects

The established approach for estimating the random effects is to use the best linear predictor of b from y , which for a given β equals $DZ'V^{-1}(y - X\beta)$. In practice β is replaced by an estimator such as $\hat{\beta}$ from (59) so that $\hat{b} = DZ'V^{-1}(y - X\hat{\beta})$. Pre-multiplying by the appropriate matrices it is straightforward to show that these estimates $\hat{\beta}$ and \hat{b} satisfy the equations in (57).

6.3 Algorithms for likelihood function optimization

Iterative numerical techniques are used to optimize the log-likelihood function and estimate the covariance parameters θ . The procedure is subject to the constraint that R and D are both positive definite. The most common iterative algorithms for optimizing the likelihood function are the Newton-Raphson method, which is the preferred method, the expectation maximization (EM) algorithm and the Fisher scoring methods.

The EM algorithm, introduced by ?, is an iterative technique for maximizing complicated likelihood functions. The algorithm alternates between performing an expectation (E) step and the maximization (M) step. The ‘E’ step computes the expectation

of the log-likelihood evaluated using the current estimate for the variables. In the ‘M’ step, parameters that maximize the expected log-likelihood, found on the previous ‘E’ step, are computed. These parameter estimates are then used to determine the distribution of the variables in the next ‘E’ step. The algorithm alternatives between these two steps until convergence is reached.

The main drawback of the EM algorithm is its slow rate of convergence. Consequently the EM algorithm is rarely used entirely in LME estimation, instead providing an initial set of values that can be passed to other optimization techniques.

The Newton Raphson (NR) method is the most common, and recommended technique for ML and REML estimation. The NR algorithm minimizes an objective function defines as -2 times the log likelihood for the covariance parameters θ . At every iteration the NR algorithm requires the calculation of a vector of partial derivatives, known as the gradient, and the second derivative matrix with respect to the covariance parameters. This is known as the observed Hessian matrix. Due to the Hessian matrix, the NR algorithm is more time-consuming, but convergence is reached with fewer iterations, compared to the EM algorithm. The Fisher scoring algorithm is an variant of the NR algorithm, that is more numerically stable and likely to converge, but not recommended to obtain final estimates.

6.4 The extended likelihood

The desire to have an entirely likelihood-based justification for estimates of random effects has motivated ?, page 429 to define the *extended likelihood*. He remarks “In mixed effects modelling the extended likelihood has been called *h-likelihood* (for hierarchical likelihood) by ?, while in smoothing literature it is known as the *penalized likelihood* (e.g. ? ?).” The extended likelihood can be written $L(\beta, \theta, b|y) = p(y|b; \beta, \theta)p(b; \theta)$ and adopting the same distributional assumptions used by ? yields the log-likelihood function

$$\ell_h(\beta, \theta, b|y) = -\frac{1}{2} \{ \log |\Sigma| + (y - X\beta - Zb)' \Sigma^{-1} (y - X\beta - Zb) \}$$

$$+ \log |D| + b'D^{-1}b\}.$$

Given θ , differentiating with respect to β and b returns Henderson's equations in (57).

Henderson's equations in (57) can be rewritten $(T'W^{-1}T)\delta = T'W^{-1}y_a$ using

$$\delta = \begin{pmatrix} \beta \\ b \end{pmatrix}, \quad y_a = \begin{pmatrix} y \\ \psi \end{pmatrix}, \quad T = \begin{pmatrix} X & Z \\ 0 & I \end{pmatrix}, \quad \text{and } W = \begin{pmatrix} \Sigma & 0 \\ 0 & D \end{pmatrix},$$

where ? describe $\psi = 0$ as quasi-data with mean $E(\psi) = b$. Their formulation suggests that the joint estimation of the coefficients β and b of the linear mixed effects model can be derived via a classical augmented general linear model $y_a = T\delta + \varepsilon$ where $E(\varepsilon) = 0$ and $\text{var}(\varepsilon) = W$, with *both* β and b appearing as fixed parameters.

7 Repeated measurements in LME models

? used ML estimation to estimate the true correlation between the variables when the measurements are linked over time. The methodology relies on the assumption that the two variables with repeated measures follow a multivariate normal distribution. The methodology currently does not extend to any more than two cases. The MLE of the correlation takes into account the dependency among repeated measures.

The true correlation ρ_{xy} is repeated measurements can be considered as having two components: between subject and within-subject correlation. The usefulness of estimating repeated measure correlation coefficients is the calculation of between-method and within-method variabilities are produced as by-products.

7.1 Statistical model

Let y_{Aij} and y_{Bij} be the j th repeated observations of the variables of interest A and B taken on the i th subject. The number of repeated measurements for each variable may differ for each individual. Both variables are measured on each time points. Let n_i be the number of observations for each variable, hence $2 \times n_i$ observations in total.

It is assumed that the pair y_{Aij} and y_{Bij} follow a bivariate normal distribution.

$$\begin{pmatrix} y_{Aij} \\ y_{Bij} \end{pmatrix} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \text{ where } \boldsymbol{\mu} = \begin{pmatrix} \mu_A \\ \mu_B \end{pmatrix} \quad (18)$$

The matrix $\boldsymbol{\Sigma}$ represents the variance component matrix between response variables at a given time point j .

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_A^2 & \sigma_{AB} \\ \sigma_{AB} & \sigma_B^2 \end{pmatrix} \quad (19)$$

σ_A^2 is the variance of variable A , σ_B^2 is the variance of variable B and σ_{AB} is the covariance of the two variable. It is assumed that $\boldsymbol{\Sigma}$ does not depend on a particular time point, and is the same over all time points.

7.2 Formulation of the response vector

Information of individual i is recorded in a response vector \mathbf{y}_i . The response vector is constructed by stacking the response of the 2 responses at the first time point, then the 2 responses at the second time point, and so on. Therefore the response vector is a $2n_i \times 1$ column vector. The covariance matrix of \mathbf{y}_i is a $2n_i \times 2n_i$ positive definite matrix $\mathbf{\Omega}$.

Consider the case where three measurements are taken by both methods A and B , \mathbf{y}_i is a 6×1 random vector describing the i th subject.

$$\mathbf{y}_i = (y_i^{A1}, y_i^{B1}, y_i^{A2}, y_i^{B2}, y_i^{A3}, y_i^{B3})' \quad (20)$$

The response vector \mathbf{y}_i can be formulated as an LME model according to Laird-Ware form.

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i \quad (21)$$

$$\mathbf{b}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{D}) \quad (22)$$

$$\boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_i) \quad (23)$$

$\boldsymbol{\beta}$ is a three dimensional vector containing the fixed effects. $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$. β_2 is usually set to zero. Consequently $\boldsymbol{\beta}$ is the solutions of the means of the two methods, i.e. $E(\mathbf{y}_i) = \mathbf{X}_i\boldsymbol{\beta}$. The variance covariance matrix \mathbf{D} is a general 2×2 matrix, while \mathbf{R}_i is a $2n_i \times 2n_i$ matrix.

7.3 Decomposition of the response covariance matrix

The variance covariance structure can be re-expressed in the following form,

$$\text{Cov}(\mathbf{y}_i) = \mathbf{\Omega}_i = \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i' + \mathbf{R}_i. \quad (24)$$

\mathbf{R}_i can be shown to be the Kronecker product of a correlation matrix \mathbf{V} and $\mathbf{\Lambda}$. The correlation matrix \mathbf{V} of the repeated measures on a given response variable is assumed to be the same for all response variables. ? remarks that, with repeated

measures, the response for each subject is correlated for each variable, and that such correlation must be taken into account in order to produce a valid inference. Various error structures may be assumed for repeated measure correlations. However both ? and ? use the identity matrix, with dimensions $n_i \times n_i$ as the formulation for \mathbf{V} . This has the benefit of being simple to implement. Roy's results also indicate the use of the identity matrix.

For the response vector described, ? presents a detailed covariance matrix. A brief summary shall be presented here only. The overall variance matrix is a 6×6 matrix composed of two types of 2×2 blocks. Each block represents one separate time of measurement.

$$\mathbf{\Omega}_i = \begin{pmatrix} \mathbf{\Sigma} & \mathbf{D} & \mathbf{D} \\ \mathbf{D} & \mathbf{\Sigma} & \mathbf{D} \\ \mathbf{D} & \mathbf{D} & \mathbf{\Sigma} \end{pmatrix} \quad (25)$$

The diagonal blocks are Σ , as described previously. The 2×2 block diagonal matrix in $\mathbf{\Omega}$ gives Σ . Σ is the sum of the between-subject variability \mathbf{D} and the within subject variability $\mathbf{\Lambda}$.

$\mathbf{\Omega}_i$ can be expressed as

$$\mathbf{\Omega}_i = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i' + (\mathbf{I}_{n_i} \otimes \mathbf{\Lambda}). \quad (26)$$

The notation dim_{n_i} means an $n_i \times n_i$ diagonal block.

7.4 Correlation terms

? demonstrated how the between-subject and within subject variabilities can be expressed in terms of correlation terms.

$$\mathbf{D} = \begin{pmatrix} \sigma_A^2 \rho_A & \sigma_A \sigma_b \rho_{AB} \delta \\ \sigma_A \sigma_b \rho_{AB} \delta & \sigma_B^2 \rho_B \end{pmatrix} \quad (27)$$

$$\mathbf{\Lambda} = \begin{pmatrix} \sigma_A^2(1 - \rho_A) & \sigma_{AB}(1 - \delta) \\ \sigma_{AB}(1 - \delta) & \sigma_B^2(1 - \rho_B) \end{pmatrix}. \quad (28)$$

ρ_A describe the correlations of measurements made by the method A at different times. Similarly ρ_B describe the correlation of measurements made by the method B at different times. Correlations among repeated measures within the same method are known as intra-class correlation coefficients. ρ_{AB} describes the correlation of measurements taken at the same same time by both methods. The coefficient δ is added for when the measurements are taken at different times, and is a constant of less than 1 for linked replicates. This is based on the assumption that linked replicates measurements taken at the same time would have greater correlation than those taken at different times. For unlinked replicates δ is simply 1. ? provides a useful graphical depiction of the role of each correlation coefficients.

8 Using LMEs for method comparison

? remarks upon ‘by-hand’ approaches advocated in ? discouragingly, describing them as tedious, unnecessary and ‘outdated’. Due to the prevalence of modern statistical software, ? advocates the adoption of computer based approaches to method comparison studies, allowing the use of LME models that would not have been feasible otherwise. Rather than using the ‘by hand’ methods, estimates for required parameters can be gotten directly from output code. Furthermore, using computer approaches removes constraints, such as the need for the design to be perfectly balanced.

9 Roy's variability tests

For the purposes of method comparison, Roy presents a methodology utilising linear mixed effects model. The formulation contains a Kronecker product covariance structure in a doubly multivariate setup. This methodology provides for the formal testing of inter-method bias, between-subject variability and within-subject variability of two methods. By doubly multivariate set up, Roy means that the information on each patient or item is multivariate at two levels, the number of methods and number of replicated measurements. Further to ?, it is assumed that the replicates are linked over time. However it is easy to modify to the unlinked case.

Roy sets out three conditions for two methods to be considered in agreement. Firstly that there be no significant bias. Second that there is no difference in the between-subject variabilities, and lastly that there is no significant difference in the within-subject variabilities. Should both the second and third conditions be fulfilled, then the overall variabilities of both methods would be equal. Roy additionally uses the overall correlation coefficient to provide extra information about the comparison, with a minimum of 0.82 being required.

Roy proposes a series of three tests on the variance components of an LME model. For these tests, four candidate models are constructed. The difference in the models are specifically in how the D and Λ matrices are constructed, using either an unstructured form or a compound symmetry form. The first model is compared against each of three other models successively.

9.1 Testing Procedures

Roy's methodology requires the construction of four candidate models. The first candidate model is compared to each of the three other models successively. It is the alternative model in each of the three tests, with the other three models acting as the respective null models.

The likelihood ratio test is the procedure used to compare the fit of two models. For

each candidate model, the ‘-2 log likelihood’ ($M2LL$) is computed. The test statistic for each of the three hypothesis tests is the difference of the $M2LL$ for each pair of models. If the p -value in each of the respective tests exceed as significance level chosen by the analyst, then the null model must be rejected.

$$-2 \ln \Lambda_d = [M2LL \text{ under } H_0 \text{ model}] - [M2LL \text{ under } H_A \text{ model}] \quad (29)$$

These test statistics follow a chi-square distribution with the degrees of freedom computed as the difference of the LRT degrees of freedom.

$$\nu = [\text{LRT df under } H_0 \text{ model}] - [\text{LRT df under } H_A \text{ model}] \quad (30)$$

The probability distribution of the test statistic can be approximated by a chi-square distribution with $(\nu_1 - \nu_2)$ degrees of freedom, where ν_1 and ν_2 are the degrees of freedom of models 1 and 2 respectively.

Likelihood ratio tests are very simple to implement in **R**, simply use the ‘`anova()`’ commands. Sample output will be given for each variability test.

9.2 Test for inter-method bias

Bias is determinable by examination of the ‘t-table’. Estimate for both methods are given, and the bias is simply the difference between the two. Because the *R* implementation does not account for an intercept term, a p -value is not given. Should a p -value be required specifically for the bias, and simple restructuring of the model is required wherein an intercept term is included. Output from a second implementation will yield a p -value.

9.3 Correlation coefficient

In addition to the variability tests, Roy advises that it is preferable that a correlation of greater than 0.82 exist for two methods to be considered interchangeable. However if two methods fulfil all the other conditions for agreement, failure to comply with this

one can be overlooked. Indeed Roy demonstrates that placing undue importance to it can lead to incorrect conclusions. ? remarks that PROC MIXED only gives overall correlation coefficients, but not their variances. Similarly variance are not determinable in R as yet either. Consequently it is not possible to carry out inferences based on all overall correlation coefficients.

10 Demonstration of Roy's testing

Roy provides three case studies, using data sets well known in method comparison studies, to demonstrate how the methodology should be used. The first two examples used are from the ‘blood pressure’ data set introduced by ?. The data set is a tabulation of simultaneous measurements of systolic blood pressure were made by each of two experienced observers (denoted ‘J’ and ‘R’) using a sphygmomanometer and by a semi-automatic blood pressure monitor (denoted ‘S’). Three sets of readings were made in quick succession. Roy compares the ‘J’ and ‘S’ methods in his first example, and the ‘R’ and ‘S’ methods in his second.

10.1 Matrix structures

Before discussing the tests, it is useful to point out the difference between symmetric form and compound symmetry form. Consider a generic matrix A ,

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}. \quad (31)$$

A symmetric matrix allows the diagonal terms a_{11} and a_{22} to differ. The compound symmetry structure requires that both of these terms be equal, i.e $a_{11} = a_{22}$.

10.2 Variability test 1

This is a test on whether both methods A and B have the same between-subject variability or not.

$$H_0 : d_A = d_B \quad (32)$$

$$H_A : d_A \neq d_B \quad (33)$$

When implemented using **R**, this test is facilitated by constructing a model specifying a symmetric form for D (i.e. the alternative model) and comparing it with a model

that has compound symmetric form for D (i.e. the null model). For this test $\hat{\Lambda}$ has a symmetric form for both models, and will be the same for both.

10.2.1 Bland-Altman's blood data

With the alternative model, the MLE of the between-subject variance covariance matrix is given by

$$\hat{D}_{Symm} = \begin{pmatrix} 923.98 & 785.24 \\ 785.24 & 971.30 \end{pmatrix} \quad (34)$$

With the null model the MLE is as follows:

$$\hat{D}_{CS} = \begin{pmatrix} 946.50 & 784.32 \\ 784.32 & 946.50 \end{pmatrix} \quad (35)$$

A likelihood ratio test is perform to determine which model is more suitable. The outcome of this test is presented in the following **R** code.

```
> anova(MCS1,MCS2)
>
>
      Model df      AIC      BIC  logLik   Test L.Ratio p-value
MCS1      1  8 4077.5 4111.3 -2030.7
MCS2      2  7 4075.6 4105.3 -2030.8 1 vs 2 0.15291 0.6958
```

The test statistic is the difference of the -2 log likelihoods; 0.15291. The p -value is 0.6958. Therefore we fail to reject the hypothesis that both have the same between-subject variabilities.

10.3 Variability test 2

This is a test on whether both methods A and B have the same within-subject variability or not.

$$H_0 : \lambda_A = \lambda_B \quad (36)$$

$$H_A : \lambda_A = \lambda_B \quad (37)$$

This model is performed in the same manner as the first test, only reversing the roles of $\hat{\mathbf{D}}$ and $\hat{\mathbf{\Lambda}}$. The null model is constructed a symmetric form for $\hat{\mathbf{\Lambda}}$ while the alternative model uses a compound symmetry form. This time $\hat{\mathbf{D}}$ has a symmetric form for both models, and will be the same for both.

10.3.1 Bland-Altman's blood data

For the null model the MLE of the within-subject variance covariance matrix is given below.

$$\hat{\mathbf{\Lambda}}_{Symm} = \begin{pmatrix} 37.40 & 16.06 \\ 16.06 & 83.14 \end{pmatrix} \quad (38)$$

With the alternative model the MLE is as follows:

$$\hat{\mathbf{\Lambda}}_{CS} = \begin{pmatrix} 60.27 & 16.06 \\ 16.06 & 60.27 \end{pmatrix} \quad (39)$$

A likelihood ratio test is perform to determine which model is more suitable. The outcome of this test is that it can be assumed that they have equal The test statistic is the difference of the $-2 \log$ likelihoods; 28.617. The p -value is less than 0.0001. In this case we reject the null hypothesis that both models have the same within-subject variabilities.

10.4 Variability test 3

This is a test on whether both methods A and B have the same overall variability or not.

$$H_0 : \sigma_A = \sigma_B \quad (40)$$

$$H_A : \sigma_A = \sigma_B \quad (41)$$

The null model is constructed a symmetric form for both $\hat{\mathbf{D}}$ and $\hat{\mathbf{\Lambda}}$ while the alternative model uses a compound symmetry form for both.

10.4.1 Bland-Altman's blood data

With the null model the MLE of the within-subject variance covariance matrix is given below.

$$\hat{\Sigma}_{Symm} = \begin{pmatrix} 961.38 & 801.40 \\ 801.40 & 1054.43 \end{pmatrix} \quad (42)$$

With the alternative model the MLE is as follows:

$$\hat{\Sigma}_{CS} = \begin{pmatrix} 1007.92 & 801.65 \\ 801.65 & 1007.92 \end{pmatrix} \quad (43)$$

Again a likelihood ratio test is used to determine the most suitable of the two candidate models. The test statistic is the difference of the $-2 \log$ likelihoods; 28.884. The p -value is less than 0.0001. We again reject the null hypothesis. Each model has a different overall variability, a foregone conclusion from the second variability test.

10.5 Test for inter-method bias

The inter-method bias between the two method is found to be 15.62 , with a p -value of

10.6 Correlation Test

$$\hat{\mathbf{r}}_{\Omega_{ii}} = \begin{pmatrix} 1 & 0.7959 \\ 0.7959 & 1 \end{pmatrix} \quad (44)$$

The diagonal blocks $\hat{\mathbf{r}}_{\Omega_{ii}}$ of the correlation matrix indicate an overall coefficient of 0.7959. This is less than the threshold of 0.82 that Roy recommends.

The off diagonal blocks of the overall correlation matrix $\hat{\mathbf{r}}_{\Omega_{ii'}}$ are

$$\hat{\mathbf{r}}_{\Omega_{ii'}} = \begin{pmatrix} 0.9611 & 0.7799 \\ 0.7799 & 0.9212 \end{pmatrix}. \quad (45)$$

10.7 Conclusion of procedure

The overall conclusion of the procedure is that the two methods are not in agreement, specifically due to the within-subject variability, and the inter-method bias. The repeatability coefficients are substantially different, one being 49% larger than the other. Additionally the overall correlation coefficient did not exceed the recommended threshold of 0.82.

11 Limits of agreement in LME models

Limits of agreement are used extensively for assessing agreement, due to they're being intuitive and easy to use. Necessarily their prevalence in literature has meant that they are now the best known measurement for agreement, and that any newer methodology would benefit by making reference to them.

? uses LME models to determine the limits of agreement. Between-subject variation for method m is given by d_m^2 and within-subject variation is given by λ_m^2 . ? remarks that for two methods A and B , separate values of d_A^2 and d_B^2 cannot be estimated, only their average. Hence the assumption that $d_x = d_y = d$ is necessary. The between-subject variability \mathbf{D} and within-subject variability $\mathbf{\Lambda}$ can be presented in matrix form,

$$\mathbf{D} = \begin{pmatrix} d_A^2 & 0 \\ 0 & d_B^2 \end{pmatrix} = \begin{pmatrix} d^2 & 0 \\ 0 & d^2 \end{pmatrix}, \quad \mathbf{\Lambda} = \begin{pmatrix} \lambda_A^2 & 0 \\ 0 & \lambda_B^2 \end{pmatrix}. \quad (46)$$

The variance for method m is $d_m^2 + \lambda_m^2$. Limits of agreement are determined using the standard deviation of the case-wise differences between the sets of measurements by two methods A and B , given by

$$\text{var}(y_A - y_B) = 2d^2 + \lambda_A^2 + \lambda_B^2. \quad (47)$$

Importantly the covariance terms in both variability matrices are zero, and no covariance component is present.

? presents a data set ‘fat’, which is a comparison of measurements of subcutaneous fat by two observers at the Steno Diabetes Center, Copenhagen. Measurements are in millimeters (mm). Each person is measured three times by each observer. The observations are considered to be ‘true’ replicates.

A linear mixed effects model is formulated, and implementation through several software packages is demonstrated. All of the necessary terms are presented in the computer output. The limits of agreement are therefore,

$$0.0449 \pm 1.96 \times \sqrt{2 \times 0.0596^2 + 0.0772^2 + 0.0724^2} = (-0.220, 0.309). \quad (48)$$

? has demonstrated a methodology whereby d_A^2 and d_B^2 can be estimated separately. Also covariance terms are present in both \mathbf{D} and $\mathbf{\Lambda}$. Using Roy's methodology, the variance of the differences is

$$\text{var}(y_{iA} - y_{iB}) = d_A^2 + \lambda_B^2 + d_A^2 + \lambda_B^2 - 2(d_{AB} + \lambda_{AB}) \quad (49)$$

All of these terms are given or determinable in computer output. The limits of agreement can therefore be evaluated using

$$\bar{y}_A - \bar{y}_B \pm 1.96 \times \sqrt{\sigma_A^2 + \sigma_B^2 - 2(\sigma_{AB})}. \quad (50)$$

For Carstensen's 'fat' data, the limits of agreement computed using Roy's method are consistent with the estimates given by ?; $0.044884 \pm 1.96 \times 0.1373979 = (-0.224, 0.314)$.

11.1 Linked replicates

? proposes the addition of an random effects term to their model when the replicates are linked. This term is used to describe the 'item by replicate' interaction, which is independent of the methods. This interaction is a source of variability independent of the methods. Therefore failure to account for it will result in variability being wrongly attributed to the methods.

? introduces a second data set; the oximetry study. This study done at the Royal Childrens Hospital in Melbourne to assess the agreement between co-oximetry and pulse oximetry in small babies.

In most cases, measurements were taken by both method at three different times. In some cases there are either one or two pairs of measurements, hence the data is unbalanced. ? describes many of the children as being very sick, and with very low oxygen saturations levels. Therefore it must be assumed that a biological change can occur in interim periods, and measurements are not true replicates.

? demonstrate the necessity of accounting for linked replicated by comparing the limits of agreement from the 'oximetry' data set using a model with the additional term, and one without. When the interaction is accounted for the limits of agreement

are $(-9.62, 14.56)$. When the interaction is not accounted for, the limits of agreement are $(-11.88, 16.83)$. It is shown that the failure to include this additional term results in an over-estimation of the standard deviations of differences.

Limits of agreement are determined using Roy's methodology, without adding any additional terms, are found to be consistent with the 'interaction' model; $(-9.562, 14.504)$. Roy's methodology assumes that replicates are linked. However, following Carstensen's example, an additional interaction term is added to the implementation of Roy's model to assess the effect, the limits of agreement estimates do not change. However there is a conspicuous difference in within-subject matrices of both models (denoted 1 and 2 respectively);

$$\hat{\mathbf{\Lambda}}_1 = \begin{pmatrix} 16.61 & 11.67 \\ 11.67 & 27.65 \end{pmatrix} \quad \hat{\mathbf{\Lambda}}_2 = \begin{pmatrix} 7.55 & 2.60 \\ 2.60 & 18.59 \end{pmatrix} \quad (51)$$

The variance of the additional random effect in model 2 is 3.01.

The Akaike information criterion (AIC) for both of models are $AIC_1 = 2304.226$ and $AIC_2 = 2306.226$. Having a difference of AIC values of 2 is equivalent to both models being equally as good as the other. The AIC values for the Carstensen 'unlinked' and 'linked' models are 1994.66 and 1955.48 respectively.

The $\hat{\mathbf{\Lambda}}$ matrices are informative as to the difference between Carstensen's unlinked and linked models. For the oximetry data, the covariance terms (given above as 11.67 and 2.6 respectively) are of similar magnitudes to the variance terms. Conversely for the 'fat' data the covariance term (-0.00032) is negligible. When the interaction term is added to the model, the covariance term remains negligible. (For the 'fat' data, the difference in AIC values is also 2).

To conclude, Carstensen's models provided a rigorous way to determine limits of agreement, but don't provide for the computation of $\hat{\mathbf{D}}$ and $\hat{\mathbf{\Lambda}}$. Therefore the test's proposed by ? can not be implemented. Conversely, accurate limits of agreement can be found using Roy's method.

Addition of the interaction term erodes the capability of Roy's methodology to compare candidate models, and therefore shall not be adopted.

(N.B. To complement the blood pressure ‘J vs S’ analysis, the limits of agreement are $15.62 \pm 1.96 \times 20.33 = (-24.22, 55.46)$.)

12 Extension of Roy's methodology

An implementation of Roy's methodology whereby three or more methods are used is not feasible in R due to computational restrictions. Specifically there is a failure of R compiler to reach convergence before the iteration limit is reached. This may be due to the presence of additional variables, causing the problem of non-identifiability. In the case of two variables, it is required to estimate two variance terms and four correlation terms, six in all. For the case of three variabilities, three variance terms must be estimated as well as nine correlation terms, twelve in all. In general for n methods has $n + T_{n-1} + T_n$ variance terms, where T_n is the triangular number for n . Hence the computational complexity quite increases substantially for every increase in n .

Should an implementation be feasible, further difficulty arises when interpreting the results. The fundamental question is whether two methods have close agreement so as to be interchangeable. When three methods are present in the model, the null hypothesis is that all three methods have the same variability relevant to the respective tests. The outcome of the analysis will either be that all three are interchangeable or that all three are not interchangeable.

The tests would not be informative as to whether any two of those three were interchangeable, or equivalently if one method in particular disagreed with the other two. Indeed it is easier to perform three pair-wise comparisons separately and then to combine the results.

13 Roy's methodology for single measurements

Roy's methodology follows from the decomposition for the covariance matrix of the response vector y_i , as presented in ?. The decomposition depends on the estimation of correlation terms, which would be absent in the single measurement case. Indeed there can be no within-subject variability if there are no repeated terms for it to

describe. There would only be the covariance matrix of the measurements by both methods, which doesn't require the use of LME models. To conclude, simple existing methodologies would be the correct approach where there only one measurements by each method.

LME Estimation and Algorithms

14 Algorithms

Maximum likelihood estimation is a method of obtaining estimates of unknown parameters by optimizing a likelihood function. The ML parameter estimates are the values of the argument that maximise the likelihood function, i.e. the estimates that make the observed values of the dependent variable most likely, given the distributional assumptions

The most common iterative algorithms used for the optimization problem in the context of LMEs are the EM algorithm, fisher scoring algorithm and NR algorithm, which [cite:West] commends as the preferred method.

A mixed model is an extension of the general linear models that can specify additional random effects terms.

Parameter of the mixed model can be estimated using either ML or REML, while the AIC and the BIC can be used as measures of "goodness of fit" for particular models, where smaller values are considered preferable.

14.1 ML v REML

(***Wikipedia***)The restricted (or residual, or reduced) maximum likelihood (REML) approach is a particular form of maximum likelihood estimation which does not base estimates on a maximum likelihood fit of all the information, but instead uses a likelihood function calculated from a transformed set of data, so that nuisance parameters have no effect.

In contrast to the earlier maximum likelihood estimation, REML can produce unbiased estimates of variance and covariance parameters.

15 ML procedures for LME

The maximum likelihood procedure of Hartley and Rao yields simultaneous estimates for both the fixed effects and the random effect, by maximising the likelihood of \mathbf{y} with respect to each element of $\boldsymbol{\beta}$ and \mathbf{b} .

16 Estimation of random effects

Estimation of random effects for LME models in the NLME package is accomplished through use of both EM (Expectation-Maximization) algorithms and Newton-Raphson algorithms.

- EM iterations bring estimates of the parameters into the region of the optimum very quickly, but convergence to the optimum is slow when near the optimum.
- Newton-Raphson iterations are computationally intensive and can be unstable when far from the optimum. However, close to the optimum they converge quickly.
- The LME function implements a hybrid approach, using 25 EM iterations to quickly get near the optimum, then switching to Newton-Raphson iterations to quickly converge to the optimum.
- If convergence problems occur, the “control argument in LME can be used to change the way the model arrives at the optimum.

References

New LME Notes

17 Linear Mixed Effects Models

17.1 Why use LMEs for Method Comparison?

The well-known “Limits of Agreement”, as developed by Bland and Altman (1986) are easily computable using the LME framework, proposed by Roy. While we will not be considering this analysis, a demonstration will be provided in the example.

Further to this, Roy(2009) demonstrates an suite of tests that can be used to determine how well two methods of measurement, in the presence of repeated measures, agree with each other.

- No Significant inter-method bias
- No difference in the between-subject variabilities of the two methods
- No difference in the within-subject variabilities of the two methods

1. Classical models
2. Grouped data sets
3. variance components
 - (a) Fixed effects
 - (b) Random effects

18 Classical Models

18.1 With non-iid residual errors

The assumption may be violated in some situations. This often happens when repeated measurements are made on each subject. In the *growth* study dataset, for example, the response variable of each subject is measured at various ages. We may suspect that error terms within a subject are correlated. A reasonable choice of the residual error covariance will therefore be a block diagonal matrix, where each block is a first-order autoregressive (AR1) covariance matrix.

Old Notes

1. Classical models
2. Grouped data sets
3. variance components
 - (a) Fixed effects
 - (b) Random effects

19 Classical Models

19.1 With non-iid residual errors

The assumption may be violated in some situations. This often happens when repeated measurements are made on each subject. In the *growth* study dataset, for example, the response variable of each subject is measured at various ages. We may suspect that error terms within a subject are correlated. A reasonable choice of the residual error covariance will therefore be a block diagonal matrix, where each block is a first-order autoregressive (AR1) covariance matrix.

20 Grouped Data Sets

In modern statistical analysis , data sets have very complex structures, such as clustered data, repeated data and hierarchical data (henceforth referred to as grouped data).

Repeated data considers various observations periodically taken from the same subjects. ‘Before and after’ measurements, as used in paired t tests, are a well known example of repeated measurements. Clustered data is simply the grouping of observations according to common characteristics. For example, an study of pupils of a school would account for the fact that they are grouped according to their classes.

Hierarchical structures organize data into a tree-like structure, i.e. groups within groups. Using the previous example, the pupils would be categorized according to their years (i.e the parent group) and then their classes (i.e the child group). This can be extended again to multiple schools, where each school would be the parent group of each year.

An important feature of such data sets is that there is correlation between observations within each of the groups (?). Observations in different groups may be independent, but any assumption that these observations within the same group are independent is inappropriate . Consequently ? states that there is two sources of variations to be considered, ‘within groups’ and ‘between groups’.

21 Grouped Data Sets

In modern statistical analysis, data sets have very complex structures, such as clustered data, repeated data and hierarchical data (henceforth referred to as grouped data).

Repeated data considers various observations periodically taken from the same subjects. ‘Before and after’ measurements, as used in paired t tests, are a well known example of repeated measurements. Clustered data is simply the grouping of observations according to common characteristics. For example, an study of pupils of a school would account for the fact that they are grouped according to their classes.

Hierarchical structures organize data into a tree-like structure, i.e. groups within groups. Using the previous example, the pupils would be categorized according to their years (i.e the parent group) and then their classes (i.e the child group). This can be extended again to multiple schools, where each school would be the parent group of each year.

An important feature of such data sets is that there is correlation between observations within each of the groups (?). Observations in different groups may be independent, but any assumption that these observations within the same group are independent is inappropriate. Consequently ? states that there is two sources of variations to be considered, ‘within groups’ and ‘between groups’.

22 Laird Ware Formulation

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, 85 \quad (52)$$

$$\mathbf{Z}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}), \quad \boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\sigma}^2\boldsymbol{\Lambda}) \quad (53)$$

23 Normal linear mixed models

The standard linear mixed effects model specifies

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon}, \quad (54)$$

where y is a vector of N observable random variables, β is a vector of p unknown parameters having fixed values (fixed effects), X and Z are $N \times p$ and $N \times q$ known matrices, and b and ϵ are vectors of q and N , respectively, unobservable random variables (random effects) such that $E(b) = 0$, $E(\epsilon) = 0$ and

$$\text{var} \begin{pmatrix} b \\ \epsilon \end{pmatrix} = \begin{pmatrix} D & 0 \\ 0 & \Sigma \end{pmatrix}$$

where D and Σ are positive definite matrices parameterized by an unknown variance component parameter vector θ .

Henderson's equations

? made the (ad-hoc) distributional assumptions $y|b \sim N(X\beta + Zb, \Sigma)$ and $b \sim N(0, D)$, and proceeded to maximize the joint density of y and b

$$\left| \begin{pmatrix} D & 0 \\ 0 & \Sigma \end{pmatrix} \right|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \begin{pmatrix} b \\ y - X\beta - Zb \end{pmatrix}' \begin{pmatrix} D & 0 \\ 0 & \Sigma \end{pmatrix}^{-1} \begin{pmatrix} b \\ y - X\beta - Zb \end{pmatrix} \right\}, \quad (55)$$

with respect to β and b , which ultimately requires minimizing the criterion

$$(y - X\beta - Zb)' \Sigma^{-1} (y - X\beta - Zb) + b' D^{-1} b. \quad (56)$$

This leads to the solutions

$$\begin{pmatrix} X' \Sigma^{-1} X & X' \Sigma^{-1} Z \\ Z' \Sigma^{-1} X & Z' \Sigma^{-1} Z + D^{-1} \end{pmatrix} \begin{pmatrix} \beta \\ b \end{pmatrix} = \begin{pmatrix} X' \Sigma^{-1} y \\ Z' \Sigma^{-1} y \end{pmatrix}. \quad (57)$$

? points out that although ? initially referred to the estimates $\hat{\beta}$ and \hat{b} from (57) as “joint maximum likelihood estimates” ? later advised that these estimates should not be referred to as “maximum likelihood” as the function being maximized in (56) is a joint density rather than a likelihood function.

Estimation of the fixed parameters

The vector y has marginal density $y \sim N(X\beta, V)$, where $V = \Sigma + ZDZ'$ is specified through the variance component parameters θ . The log-likelihood of the fixed parameters (β, θ) is

$$\ell(\beta, \theta|y) = -\frac{1}{2} \log |V| - \frac{1}{2} (y - X\beta)' V^{-1} (y - X\beta), \quad (58)$$

and for fixed θ the estimate $\hat{\beta}$ of β is obtained as the solution of

$$(X'V^{-1}X)\beta = X'V^{-1}y. \quad (59)$$

Maximum likelihood and restricted maximum likelihood have become the most common strategies for estimating the variance component parameter θ . Substituting $\hat{\beta}$ from (59) into $\ell(\beta, \theta|y)$ from (58) returns the *profile* log-likelihood

$$\begin{aligned} \ell_P(\theta | y) &= \ell(\hat{\beta}, \theta | y) \\ &= -\frac{1}{2} \log |V| - \frac{1}{2} (y - X\hat{\beta})' V^{-1} (y - X\hat{\beta}) \end{aligned}$$

of the variance parameter θ . Estimates of the parameters θ specifying V can be found by maximizing $\ell_P(\theta | y)$ over θ . In practice the *restricted* log-likelihood

$$\ell_R(\theta | y) = \ell_P(\theta | y) - \frac{1}{2} \log |X'VX|$$

is preferred. This approach is based on maximizing the likelihood of linear combinations of y that do not depend on β , and in this way takes into account the estimation of β .

Estimation of the random effects

The established approach for estimating the random effects is to use the best linear predictor of b from y , which for a given β equals $DZ'V^{-1}(y - X\beta)$. In practice β is replaced by an estimator such as $\hat{\beta}$ from (59) so that $\hat{b} = DZ'V^{-1}(y - X\hat{\beta})$. Pre-multiplying by the appropriate matrices it is straightforward to show that these estimates $\hat{\beta}$ and \hat{b} satisfy the equations in (57).

The extended likelihood

The desire to have an entirely likelihood-based justification for estimates of random effects has motivated ?, page 429 to define the *extended likelihood*. He remarks “In mixed effects modelling the extended likelihood has been called *h-likelihood* (for hierarchical likelihood) by ?, while in smoothing literature it is known as the *penalized likelihood* (e.g. ? ?).” The extended likelihood can be written $L(\beta, \theta, b|y) = p(y|b; \beta, \theta)p(b; \theta)$

and adopting the same distributional assumptions used by ? yields the log-likelihood function

$$\begin{aligned}\ell_h(\beta, \theta, b|y) = & -\frac{1}{2} \{ \log |\Sigma| + (y - X\beta - Zb)' \Sigma^{-1} (y - X\beta - Zb) \\ & + \log |D| + b' D^{-1} b \}.\end{aligned}$$

Given θ , differentiating with respect to β and b returns Henderson's equations in (57).

Henderson's equations in (57) can be rewritten $(T'W^{-1}T)\delta = T'W^{-1}y_a$ using

$$\delta = \begin{pmatrix} \beta \\ b \end{pmatrix}, \quad y_a = \begin{pmatrix} y \\ \psi \end{pmatrix}, \quad T = \begin{pmatrix} X & Z \\ 0 & I \end{pmatrix}, \quad \text{and } W = \begin{pmatrix} \Sigma & 0 \\ 0 & D \end{pmatrix},$$

where ? describe $\psi = 0$ as quasi-data with mean $E(\psi) = b$. Their formulation suggests that the joint estimation of the coefficients β and b of the linear mixed effects model in (54) can be derived via a classical augmented general linear model $y_a = T\delta + \varepsilon$ where $E(\varepsilon) = 0$ and $\text{var}(\varepsilon) = W$, with *both* β and b appearing as fixed parameters.

24 Note 2: Model terms

It is important to note the following characteristics of this model.

- Let the number of replicate measurements on each item i for both methods be n_i , hence $2 \times n_i$ responses. However, it is assumed that there may be a different number of replicates made for different items. Let the maximum number of replicates be p . An item will have up to $2p$ measurements, i.e. $\max(n_i) = 2p$.
- Later on \mathbf{X}_i will be reduced to a 2×1 matrix, to allow estimation of terms. This is due to a shortage of rank. The fixed effects vector can be modified accordingly.
- \mathbf{Z}_i is the $2n_i \times 2$ model matrix for the random effects for measurement methods on item i .
- \mathbf{b}_i is the 2×1 vector of random-effect coefficients on item i , one for each method.
- $\boldsymbol{\epsilon}$ is the $2n_i \times 1$ vector of residuals for measurements on item i .

- \mathbf{G} is the 2×2 covariance matrix for the random effects.
- \mathbf{R}_i is the $2n_i \times 2n_i$ covariance matrix for the residuals on item i .
- The expected value is given as $E(\mathbf{y}_i) = \mathbf{X}_i\boldsymbol{\beta}$. (?)
- The variance of the response vector is given by $\text{Var}(\mathbf{y}_i) = \mathbf{Z}_i\mathbf{G}\mathbf{Z}_i' + \mathbf{R}_i$ (?).

References