Chapter 1

Appendix C

1.1 Demonstration of Roy's Model

The computation of the limits of agreement require that the variance of the difference of measurements. This variance is easily computable from the estimate of the Block - $\mathbf{\Omega}_i$ matrix. Lack of agreement can arise if there is a disagreement in overall variabilities. This may be due to due to the disagreement in either between-item variabilities or within-item variabilities, or both. Roy (2009) allows for a formal test of each.

Roy (2009) proposes a LME based approach with Kronecker product covariance structure with doubly multivariate setup to assess the agreement between 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. This approachs allows for unbalanced data, i.e. unequal numbers of replications for each subject.

Response for ith subject can be written as

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + b_{1i} z_{i1} + b_{2i} z_{i2} + \epsilon_i$$

- β_1 and β_2 are fixed effects corresponding to both methods. (β_0 is the intercept.)

- b_{1i} and b_{2i} are random effects corresponding to both methods.

1

Overall variability between the two methods (Ω) is sum of between-subject (G) and within-subject variability (Σ),

$$\text{Block } \boldsymbol{\Omega}_i = \begin{bmatrix} d_1^2 & d_{12} \\ d_{12} & d_2^2 \end{bmatrix} + \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}.$$

In order to express Roy's LME model in matrix notation we gather all $2n_i$ observations specific to item i into a single vector $\boldsymbol{y}_i = (y_{1i1}, y_{2i1}, y_{1i2}, \ldots, y_{mir}, \ldots, y_{1in_i}, y_{2in_i})'$. The LME model can be written

$$\boldsymbol{y_i} = \boldsymbol{X_i\beta} + \boldsymbol{Z_ib_i} + \boldsymbol{\epsilon_i},$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$ is a vector of fixed effects, and $\boldsymbol{X}_i$ is a corresponding $2n_i \times 3$ design matrix for the fixed effects. The random effects are expressed in the vector $\boldsymbol{b} = (b_1, b_2)'$, with $\boldsymbol{Z}_i$ the corresponding $2n_i \times 2$ design matrix. The vector $\boldsymbol{\epsilon}_i$ is a $2n_i \times 1$ vector of residual terms.

It is assumed that $\boldsymbol{b}_i \sim N(0, \boldsymbol{G})$, $\boldsymbol{\epsilon}_i$ is a matrix of random errors distributed as $N(0, \boldsymbol{R}_i)$ and that the random effects and residuals are independent of each other.

The maximum likelihood estimate of the between-subject variance covariance matrix of two methods is given as G. The estimate for the within-subject variance covariance matrix is $\hat{\Sigma}$. The estimated overall variance covariance matrix 'Block Ω_i' is the addition of $\hat{G}$ and $\hat{\Sigma}$.

$$\text{Block } \Omega_i = \hat{G} + \hat{\Sigma} \tag{1.1}$$

$\boldsymbol{G}$ is the variance covariance matrix for the random effects $\boldsymbol{b}$. i.e. between-item sources of variation. The between-item variance covariance matrix $\boldsymbol{G}$ is constructed as follows:

$$\text{Var} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \boldsymbol{G} = \begin{pmatrix} g_1^2 & g_{12} \\ g_{12} & g_2^2 \end{pmatrix}$$

It is important to note that no special assumptions about the structure of $\boldsymbol{G}$ are made. An example of such an assumption would be that $\boldsymbol{G}$ is the product of a scalar value and the identity matrix.

$\boldsymbol{R}_i$ is the variance covariance matrix for the residuals, i.e. the within-item sources of variation between both methods. Computational analysis of linear mixed effects models allow for the explicit analysis of both $\boldsymbol{G}$ and $\boldsymbol{R_i}$.

The overall variability between the two methods is the sum of between-item variability $\boldsymbol{G}$ and within-item variability $\boldsymbol{\Sigma}$. Roy (2009) denotes the overall variability as Block - $\boldsymbol{\Omega}_i$. The overall variation for methods 1 and 2 are given by

$$
\begin{pmatrix} \omega_1^2 & \omega_{12} \\ \omega_{12} & \omega_2^2 \end{pmatrix} = \begin{pmatrix} g_1^2 & g_{12} \\ g_{12} & g_2^2 \end{pmatrix} + \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}
$$

1.2 Model Fit

This is a simple model, this time with an interaction effect. There is a fixed effect for each method. This model has random effects at two levels b_i for the subject, and another, b_{ij}, for the respective method within each subject.

$$
y_{ijk} = \beta_j + b_i + b_{ij} + \epsilon_{ijk}, \qquad i = 1, \ldots, 2, j = 1, \ldots, 85, k = 1, \ldots, 3
$$

$$
b_i \sim \mathcal{N}(0, \sigma_1^2), \qquad b_{ij} \sim \mathcal{N}(0, \sigma_2^2), \qquad \epsilon_i \sim \mathcal{N}(0, \sigma^2)
$$

In this model, the random interaction terms all have the same variance σ_2^2. These terms are assumed to be independent of each other, even within the same subject.

This model treats the random interactions for each subject as a vector and allows the variance-covariance matrix for that vector to be estimated from the set of all positive-definite matrices. $\boldsymbol{y_i}$ is the entire response vector for the ith subject. $\boldsymbol{X_i}$ and $\boldsymbol{Z_i}$ are the fixed- and random-effects design matrices respectively.

$$
\boldsymbol{y_i} = \boldsymbol{X_i}\boldsymbol{\beta} + \boldsymbol{Z_i}\boldsymbol{b_i} + \boldsymbol{\epsilon_i}, \qquad i = 1, \ldots, 85
$$

$$
\boldsymbol{Z_i} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Psi}), \qquad \boldsymbol{\epsilon_i} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\sigma^2 \Lambda})
$$

Roy (2009) presents two nested models that specify the condition of equality as required, with a third nested model for an additional test. There three formulations share the same structure, and can be specified by making slight alterations of the code for the Reference Model.

3

For these tests, four candidate LME models are constructed. The differences in the models are specifically in how the the $\boldsymbol{G}$ and Λ matrices are constructed, using either an unstructured form or a compound symmetry form. To illustrate these differences, consider a generic matrix A,

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}.$$

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 $a_{11} = a_{22}$.

1.3   Roy's Variability Tests

For the blood pressure data used in Roy (2009), all four candidate models are implemented by slight variations of this piece of code, specifying either `pdSymm` or `pdCompSymm` in the second line, and either `corSymm` or `corCompSymm` in the fourth line.

Variability tests proposed by Roy (2009) affords the opportunity to expand upon Carstensen's approach.

The first test allows of the comparison the within-subject variability of two methods. Similarly, the second test assesses the between-subject variability of two methods. A third test is a test that compares the overall variability of the two methods.

The tests are implemented by fitting a specific LME model, and three variations thereof, to the data. These three variant models introduce equality constraints that act null hypothesis cases.

For the blood pressure data used in Roy (2009), all four candidate models are implemented by slight variations of this piece of code, specifying either `pdSymm` or `pdCompSymm` in the second line, and either `corSymm` or `corCompSymm` in the fourth line.

Using this `R` implementation for other data sets requires that the data set is structured appropriately (i.e. each case of observation records the index, response, method and replicate). Once formatted properly, implementation is simply a case of re-writing the first line of code, and computing the four candidate models accordingly.

Other important aspects of the method comparison study are consequent. The limits of agreement

are computed using the results of the first model.

1.4 Reference Model

The first of Roy's candidate model can be implemented using the following code;

```
> Ref.Fit = lme(y ~ meth-1, data = dat,   #Symm , Symm#
+      random = list(item=pdSymm(~ meth-1)),
+      weights=varIdent(form=~1|meth),
+      correlation = corSymm(form=~1 | item/repl),
+      method="ML")
```

Using this R implementation for other data sets requires that the data set is structured appropriately (i.e. each case of observation records the index, response, method and replicate). Once formatted properly, implementation is simply a case of re-writing the first line of code, and computing the four candidate models accordingly.

The fixed effects estimates are the same for all four candidate models. The inter-method bias can be easily determined by inspecting a summary of any model. The summary presents estimates for all of the important parameters, but not the complete variance-covariance matrices (although some simple R functions can be written to overcome this). The variance estimates for the random effects for NMW.fit is presented below.

```
Random effects:
Formula: ~method - 1 | subject
Structure: Compound Symmetry
StdDev Corr
methodJ  30.765
methodS  30.765 0.829
Residual  6.115
```

Similarly, for computing the limits of agreement the standard deviation of the differences is not explicitly given. Again, A simple `R` function can be written to calculate the limits of agreement directly.

1.4.1 Inter-method Bias

This bias is specified as a fixed effect in the LME model. For a practitioner who has a reasonable level of competency in R and undergraduate statistics (in particular simple linear regression model) this is a straight-forward procedure.

A formal test for inter-method bias can be implemented by examining the fixed effects of the model. This is common to well known classical linear model methodologies. The null hypotheses, that both methods have the same mean, which is tested against the alternative hypothesis, that both methods have different means.

The inter-method bias and necessary $t-$value and $p-$value are presented in computer output. A decision on whether the first of Roy's criteria is fulfilled can be based on these values.

The inter-method bias between the two method is found to be 15.62 , with a $t-$value of -7.64, with a $p-$value of less than 0.0001. Consequently there is a significant inter-method bias present between methods J and S, and the first of the Roy's three agreement criteria is unfulfilled.

1.5 Likelihood Ratio Tests

Conventionally LME models can be tested using Likelihood Ratio Tests, wherein a reference model is compared to a nested model.

1.6 Variability Test 1: Within-Item Variability

Next, the first variability test is carried out, yielding maximum likelihood estimates of the within-item variance covariance matrix, for both the null model, in compound symmetry (CS) form, and the alternative model in symmetric (symm) form.

```
NMW.fit = lme(y ~ meth-1, data = dat,    #Symm , CS#

    random = list(item=pdSymm(~ meth-1)),

    weights=varIdent(form=~1|meth),

    correlation = corCompSymm(form=~1 | item/repl),

    method="ML")
```

These matrices are determined to be as follows;

$$\hat{\boldsymbol{G}}_{CS} = \begin{pmatrix} 946.50 & 784.32 \\ 784.32 & 946.50 \end{pmatrix}, \qquad \hat{\boldsymbol{G}}_{Symm} = \begin{pmatrix} 923.98 & 785.24 \\ 785.24 & 971.30 \end{pmatrix}.$$

A likelihood ratio test is perform to compare both candidate models. The log-likelihood of the null model is -2030.7, and for the alternative model -2030.8. The test statistic, presented with greater precision than the log-likelihoods, is 0.1592. The $p-$value is 0.6958. Consequently we fail to reject the null model, and by extension, conclude that the hypothesis that methods J and S have the same between-subject variability. Thus the second of the criteria is fulfilled.

1.6.1 Variability Test 1

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

$$H_0: \quad \sigma_A = \sigma_B \tag{1.2}$$

$$H_A: \quad \sigma_A = \sigma_B \tag{1.3}$$

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

1.6.2 Test 1

The first variability test determines maximum likelihood estimates of the within-subject variance co-variance matrix, for both the reference model, in CS form, and the alternative model in symmetric form.

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

$$\hat{\Sigma}_{Symm} = \begin{pmatrix} 37.40 & 16.06 \\ 16.06 & 83.14 \end{pmatrix} \tag{1.4}$$

With the alternative model the MLE is as follows:

$$\hat{\Sigma}_{CS} = \begin{pmatrix} 60.27 & 16.06 \\ 16.06 & 60.27 \end{pmatrix} \tag{1.5}$$

$$\hat{\Sigma}_{CS} = \begin{pmatrix} 60.27 & 16.06 \\ 16.06 & 60.27 \end{pmatrix}, \qquad \hat{\Sigma}_{Symm} = \begin{pmatrix} 37.40 & 16.06 \\ 16.06 & 83.14 \end{pmatrix}.$$

The second candidate model 'MCS2' is implemented with the same code as MCS1, except for the term `pdCompSymm` in the second line, rather than `pdSymm`.

```
NMW.Fit = lme(y ~ meth-1, data = dat,   #Symm , CS#
  random = list(item=pdSymm(~ meth-1)),
  # weights=varIdent(form=~1|meth),
  correlation = corCompSymm(form=~1 | item/repl),
  method="ML")
```

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.

A likelihood ratio test is perform to compare both candidate models. The log-likelihood of the alternative model model is -2045.0. As before, the null model has a log-likelihood of -2030.7. The

test statistic is computed as 28.617, again presented with greater precision. The $p-$value is less than 0.0001. In this case we reject the null hypothesis of equal within-subject variability. Consequently the third of Roy's criteria is unfulfilled.

```
> 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; 0.153. The $p-$value is 0.696. Therefore we fail to reject the hypothesis that both have the same between-subject variabilities.

1.7 Variability Test 2: Between-Item Variances

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

$$H_0: \quad g_A = g_B \tag{1.6}$$

$$H_A: \quad g_A \neq g_B \tag{1.7}$$

When implemented using R, this test is facilitated by constructing a model specifying a symmetric form for $\boldsymbol{G}$ (i.e. the alternative model) and comparing it with a model that has compound symmetric form for $\boldsymbol{G}$ (i.e. the null model). For this test $\hat{\boldsymbol{\Sigma}}$ has a symmetric form for both models, and will result in the same estimate for both.

For example, the first test model 'NMB.fit' is implemented with the same code as Ref.Fit, except for the term pdCompSymm in the second line, rather than pdSymm.

1.7.1 Nested Model (Between-Item Variability)

```
NMB.fit  = lme(y ~ meth-1, data = dat,   #CS , Symm#
    random = list(item=pdCompSymm(~ meth-1)),
    correlation = corSymm(form=~1 | item/repl),
    method="ML")
```

With the alternative model, the MLE of the between-subject variance covariance matrix $\boldsymbol{G}$ is given by

$$\hat{\boldsymbol{G}}_{Symm} = \begin{pmatrix} 923.98 & 785.24 \\ 785.24 & 971.30 \end{pmatrix} \tag{1.8}$$

With the refence model the MLE is as follows:

$$\hat{\boldsymbol{G}}_{CS} = \begin{pmatrix} 946.50 & 784.32 \\ 784.32 & 946.50 \end{pmatrix} \tag{1.9}$$

For the blood pressure data used in Roy (2009), all four candidate models are implemented by slight variations of this piece of code, specifying either `pdSymm` or `pdCompSymm` in the second line, and either `corSymm` or `corCompSymm` in the fourth line.

The fixed effects estimates are the same for all four candidate models. The inter-method bias can be easily determined by inspecting a summary of any model. The summary presents estimates for all of the important parameters, but not the complete variance-covariance matrices (although some simple `R` functions can be written to overcome this). The variance estimates for the random effects for `NMW.Fit` is presented below.

```
Random effects:
Formula: ~method - 1 | subject
Structure: Compound Symmetry


        StdDev  Corr
methodJ  30.765
methodS  30.765  0.829
Residual  6.115
```

Similarly, for computing the limits of agreement the standard deviation of the differences is not explicitly given. Again, a simple `R` function can be written to calculate the limits of agreement directly.

1.8   Variability Test 3: Overall Variability

The last of the three variability tests is carried out to compare the overall variabilities of both methods. A third nested model is used to test overall variability, substantively a joint test for between-item and within-item variability.

The motivation for including such a test in the suite is not clear, although it does circumvent the need for multiple comparison procedures in certain circumstances, hence providing a simplified procedure for non-statisticians.

```
NMO.fit = lme(y ~ meth-1, data = dat,   #CS , CS#
  random = list(item=pdCompSymm(~ meth-1)),
  correlation = corCompSymm(form=~1 | item/repl),
  method="ML")
```

With the null model the MLE of the within-subject variance covariance matrix is given below. The overall variabilities for the null and alternative models, respectively, are determined to be as follows;

$$\hat{\boldsymbol{\Omega}}_{CS} = \begin{pmatrix} 1007.92 & 801.65 \\ 801.65 & 1007.92 \end{pmatrix}, \qquad \hat{\boldsymbol{\Omega}}_{Symm} = \begin{pmatrix} 961.38 & 801.40 \\ 801.40 & 1054.43 \end{pmatrix},$$

The log-likelihood of the alternative model model is -2045.2, and again, the null model has a log-likelihood of -2030.7. The test statistic is 28.884, and the $p-$value is less than 0.0001. The null hypothesis, that both methods have equal overall variability, is rejected. Further to the second variability test, it is known that this difference is specifically due to the difference of within-subject variabilities.

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

$$H_0: \quad \sigma_A = \sigma_B \tag{1.10}$$

$$H_A: \quad \sigma_A = \sigma_B \tag{1.11}$$

The null model is constructed a symmetric form for both $\hat{\boldsymbol{G}}$ and $\hat{\boldsymbol{\Sigma}}$ while the alternative model uses a compound symmetry form for both. With the null model the MLE of the within-subject variance covariance matrix is given below.

$$\hat{\boldsymbol{\Omega}}_{Symm} = \begin{pmatrix} 961.38 & 801.40 \\ 801.40 & 1054.43 \end{pmatrix} \tag{1.12}$$

With the alternative model the MLE is as follows:

$$\hat{\boldsymbol{\Omega}}_{CS} = \begin{pmatrix} 1007.92 & 801.65 \\ 801.65 & 1007.92 \end{pmatrix} \tag{1.13}$$

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.

1.9 Issue of Tractability

Bibliography

Roy, A. (2009). An application of the linear mixed effects model to ass the agreement between two methods with replicated observations. *Journal of Biopharmaceutical Statistics 19*, 150–173.