Random intercept model in a balanced design

Brice Ozenne

October 22, 2024

1 Notations

Consider an outcome variable Y measured in n subjects at p occasions. We will index the subjects by $i \in \{1, \ldots, n\}$ and the occasions by $j \in \{1, \ldots, p\}$. During their follow-up each subject is subject to an active ($T = 1$) and a control treatment ($T = 0$) respectively p_1 and p_0 times. We will use the bold notation to denote vector of random variables, e.g. $\mathbf{T}_i = \{T_{i,1}, \ldots, T_{i,p}\}$.

As a working model we will consider the following random intercept model:

$$Y_{i,j} = \alpha + \beta T_{i,j} + u_i + \boldsymbol{\varepsilon}_{i,j}$$

where $u_i \sim \mathcal{N}(0, \tau)$ and $\varepsilon_{i,j} \sim \mathcal{N}(0, \delta)$. Introducing $\rho = \frac{\tau}{\tau+\delta}$ and $\sigma^2 = \tau + \delta$, we can then express the residual variance-covariance matrix as:

$$\mathbb{V}ar\left[\mathbf{Y}_i|\mathbf{T}_i\right] = \mathbb{V}ar\left[u_i + \boldsymbol{\varepsilon}_i|T_i\right] = \Omega = \sigma^2 R = \sigma^2((1-\rho)I + \rho\mathbf{e}\mathbf{e}^\intercal)$$

where I denotes the $p \times p$ identity matrix and $\mathbf{e}$ a corresponding of size p containing only 1. $\Theta = (\alpha, \beta, \delta, \tau)$ or equivalently $(\alpha, \beta, \rho, \sigma)$ will denote the vector of model parameters and $\boldsymbol{\mu}_i = (\alpha + \beta T_{i1}, \ldots, \alpha + \beta T_{ip})$ the vector of fitted values. Note that since we assume a balanced design and since Ω is unchanged by re-ordering, we can re-order the data such that $\mathbf{T}_i = \mathbf{T}_{i'} = \mathbf{T}$ for all $(i, i') \in \{1, \ldots, n\}^2$.

2 Estimates in a random intercept model

2.1 Theory

Appendix B shows that the Maximum Likelihood estimate of Θ are:

- **mean parameters**: $\widehat{\alpha} = \frac{1}{np}\sum_{i=1}^{n}\sum_{j=1}^{p}(1-T_{i,j})Y_{i,j}$
  $\widehat{\beta} = \frac{1}{np}\sum_{i=1}^{n}\sum_{t=1}^{p}(2T_{i,j}-1)Y_{i,j}$
  $\widehat{\mu}_{i,j} = \widehat{\alpha} - T_{i,j}\widehat{\beta}$
- **variance parameter**: $\widehat{\sigma}^2 = \frac{1}{np}\sum_{i=1}^{n}\sum_{j=1}^{p}(Y_{i,j}-\widehat{\mu}_{i,j})^2$
- **correlation parameter**: $\widehat{\rho} = \frac{1}{p(p-1)/2}\sum_{j=1}^{p}\sum_{j'\in\{1,\ldots,j-1\}}\widehat{\rho}_{j,j'}$
  where for $j \in \{1,\ldots,p\}$, $j' \in \{1,\ldots,j-1\}$, $\widehat{\rho}_{j,j'} = \frac{1}{n}\sum_{i=1}^{n}\frac{(Y_{i,j}-\widehat{\mu}_{i,j})}{\widehat{\sigma}}\frac{(Y_{i,j'}-\widehat{\mu}_{i,j'})}{\widehat{\sigma}}$

i.e. the empirical mean of the outcome, the empirical residual variance, and the average empirical residual correlation.

2.2 Numerical example

We will illustrate the previous result on an example. First we simulate some data in the long format:

```
library(LMMstar)
library(lava)
library(Matrix)

set.seed(1)
n.obs <- 1e3
n.times <- 4
dfL <- sampleRem(n.obs, n.times = n.times, format = "long",
                 mu = c(1,0,0,0), sigma = 1:4, lambda = c(0.5,0.25,2,1))
dfL <- dfL[order(dfL$id),c("id","visit","Y")]
dfL$treatment <- as.numeric(dfL$visit) %% 2
head(dfL)
```

```
  id visit          Y treatment
1  1     1  0.2551614         1
2  1     2  0.7913185         0
3  1     3 -2.1031314         1
4  1     4 -0.4489691         0
5  2     1  1.5637433         1
6  2     2 -0.1637081         0
```

Converting to the wide format facilitate the calculation of the time specific mean, variance, and correlation:

```
dfW <- reshape(dfL[,c("id","visit","Y")],
               direction = "wide", idvar = "id", timevar = "visit")
rbind(mean = colMeans(dfW[,-1]),
      var = apply(dfW[,-1],2,var))
cor(dfW[,-1])
```

```
          Y.1       Y.2       Y.3       Y.4
mean 1.534321 0.2534847  2.101116  1.040294
var  3.770008 1.6149499 47.740082 12.611689
          Y.1       Y.2       Y.3       Y.4
Y.1 1.0000000 0.5515201 0.8579057 0.8330143
Y.2 0.5515201 1.0000000 0.6468049 0.6131780
Y.3 0.8579057 0.6468049 1.0000000 0.9503735
Y.4 0.8330143 0.6131780 0.9503735 1.0000000
```

2.2.1 Maximum likelihood

A random intercept model estimated by Maximum Likelihood (ML) leads to the following results:

```
eML.RI <- lmm(Y~treatment+(1|id), data = dfL, method.fit = "ML")
coef(eML.RI, effects = "all")
```

```
(Intercept)   treatment       sigma     rho(id)
  0.6468893   1.1708292   4.0663607   0.5049300
```

We retrive the empirical means for the intercept and treatment effects:

```
alphaHat <- mean(dfL$Y[dfL$treatment == 0])
betaHat <- mean(dfL$Y[dfL$treatment == 1]) - alphaHat
c(alphaHat, betaHat)
```

```
[1] 0.6468893 1.1708292
```

the empirical squared residuals for the variance:

```
dfL$res <- dfL$Y - alphaHat - dfL$treatment*+betaHat
sqrt(mean(dfL$res^2))
```

```
[1] 4.066361
```

and the empirical residual correlation:

```
dfL$res.normML <- dfL$res/sqrt(mean(dfL$res^2))
dfWres.normML <- reshape(dfL[,c("id","visit","res.normML")],
                         direction = "wide", idvar = "id", timevar = "
  visit")
M.MLcor <- crossprod(as.matrix(dfWres.normML[,-1]))/n.obs
mean(M.MLcor[lower.tri(M.MLcor)])
```

```
[1] 0.50493
```

2.2.2 Restricted maximum likelihood

When fitting a random intercept model estimated by Maximum Likelihood (REML):

```
eREML.RI <- lmm(Y~treatment+(1|id), data = dfL, method.fit = "REML")
coef(eREML.RI, effects = "all")
```

```
(Intercept)   treatment       sigma     rho(id)
  0.6468893   1.1708292   4.0678916   0.5051376
```

We retrive the empirical means for the intercept and treatment effects. However we do not 'exactly' retrieve the REML estimate of the residual standard deviation using:

```
sd(dfL$res)
```

```
[1] 4.066869
```

To closer we can get would be using 3 degrees of freedom:

```
NROW(dfL)-sum(tapply(dfL$res^2, dfL$visit, sum))/(coef(eREML.RI, effects =
   "variance"))^2
```

```
   sigma
3.010256
```

We do not 'exactly' retrieve the REML estimate of the residual correlation using the Pearson correlation:

```
dfL$res.normREML <- dfL$res/coef(eREML.RI, effects = "variance")
dfWres.normREML <- reshape(dfL[,c("id","visit","res.normREML")],
                           direction = "wide", idvar = "id", timevar = "
  visit")
M.REMLcor <- crossprod(as.matrix(dfWres.normREML[,-1]))/(NROW(dfWres.
  normREML)-1)
mean(M.REMLcor[lower.tri(M.REMLcor)])
```

```
[1] 0.505055
```

3 Standard error in a random intercept model

3.1 Theory

Appendix C shows that the standard error of the treatment effect estimator can be expressed as:

$$\sigma_{\widehat{\beta}} = \sqrt{\frac{\delta}{n}\frac{p}{p_1(p-p_1)}}$$

It also shows that in the special case of Maximum Likelihood estimation with as many observations under treatment as under control ($p_1 = p/2$) it simplifies to:

$$\sigma_{\widehat{\beta}} = \sqrt{\frac{2\left(1 - \frac{1}{p(p-1)/2}\sum_{j<j'}\rho_{j,j'}\right)\sigma^2}{np_1}} \tag{1}$$

where for $j \in \{1, \dots, p\}$, $j' \in \{1, \dots, j-1\}$:

$$\begin{aligned}
\rho_{j,j'} &= \mathbb{E}\left[\frac{(Y_{.,j} - \mu_{.,j})}{\sigma}\frac{(Y_{.,j'} - \mu_{.,j'})}{\sigma}\right] \\
&\approx \frac{1}{n}\sum_{i=1}^{n}\frac{(Y_{i,j} - \mu_{i,j})}{\sigma}\frac{(Y_{i,j'} - \mu_{i,j'})}{\sigma}
\end{aligned}$$

which can be understood as the data generating within-subject correlation between observation j and j', provided that the mean and variance structure of the model are correctly specified.

3.2 Comparison to a t-test on the first change

When using a t-test on the change based **only on the first observation under each treatment**, the variance is:

$$\sigma_{\widehat{\Delta}(1)} = \sqrt{\frac{2(1-\rho_{1,p_1+1})\sigma^2}{n}}$$

where for convenience the first p_1 observations are under one treatment condition and the last p_1 observations under the other treatment condition. This strategy controls the type 1 error and is optimal if observations from the same treatment are (nearly) perfectly correlated i.e. $\rho_{j,j'} \approx 1$ if $(j,j') \in \{1, \dots, p_1\}^2$ or if $(j,j') \in \{p_1+1, \dots, p_1^2\}$. In such a case the cross-correlation must be (nearly) constant for the correlation matrix to be positive definite. We thus have:

$$\frac{1}{p(p-1)/2}\sum_{j<j'}\rho_{j,j'} = \frac{2(p/2)(p/2-1)/2\rho_{1,1} + (p/2)^2\rho_{1,p_1+1}}{p(p-1)/2}$$

$$
\begin{aligned}
&= \frac{(p/2-1)\rho_{1,1} + p/2\rho_{1,p_1+1}}{p-1} = \frac{p\frac{\rho_{1,p_1+1}+\rho_{1,1}}{2} - \rho_{1,1}}{p-1} \\
&= \rho_{1,1} - \frac{p}{2(p-1)}(\rho_{1,1} - \rho_{1,p_1+1}) \approx 1 - \frac{p}{2(p-1)}(1 - \rho_{1,p_1+1})
\end{aligned}
$$

The random intercept model will not control the type 1 error whenever $\sigma_{\widehat{\Delta}(1)} > \sigma_{\widehat{\beta}}$:

$$
1 - \rho_{1,p_1+1} > \frac{1 - \frac{1}{p(p-1)/2}\sum_{j<j'} \rho_{j,j'}}{p_1} = \frac{1}{p-1}(1 - \rho_{1,p_1+1})
$$

which is always true unless $p = 1$ (no repetition) or $\rho_{1,p_1+1} \approx 1$ (compound symmetry structure).

3.3 Comparison to a t-test on changes in a paired design

Consider now computing p_1 changes per patient, e.g. $p_1 + 1$ vs. 1, $p_1 + 2$ vs. 2, ..., using only distinct observations (no observation is used twice when computing changes), and stacking all changes into a t-test. This strategy will generally not control the type 1 error (as it disregard within-individual correlation). The variance of the corresponding estimator can be expressed as:

$$
\sigma_{\widehat{\Delta}(p_1)} = \sqrt{\frac{2(1-\rho_{1,p_1+1})\sigma^2}{np_1}}
$$

Assuming constant within (no necessarily close to one) and constant cross-correlation we can compare this variance with the mixed model variance:

$$
\sigma_{\widehat{\beta}} - \sigma_{\widehat{\Delta}(p)} = \sqrt{\frac{4\left(1 - \rho_{1,1} + \frac{p}{2(p-1)}(\rho_{1,1} - \rho_{1,p_1+1})\right)\sigma^2}{np}} - \sqrt{\frac{4(1-\rho_{1,p_1+1})\sigma^2}{np}}
$$

which has the same sign as:

$$
\begin{aligned}
\left(1 - \rho_{1,1} + \frac{p}{2(p-1)}(\rho_{1,1} - \rho_{1,p_1+1})\right) - (1 - \rho_{1,p_1+1}) &= (\rho_{1,1} - \rho_{1,p_1+1})\left(\frac{p}{2(p-1)} - 1\right) \\
&= (\rho_{1,1} - \rho_{1,p_1+1})\frac{2-p}{2(p-1)}
\end{aligned}
$$

Because $\frac{2-p}{2(p-1)} < 0$ the mixed model will be more liberal (i.e. provide a worse type 1 error control) if $\rho_{1,1} > \rho_{1,p_1+1}$ otherwise it will be less liberal.

3.4 Numerical example

We can retrieve the standard error estimated by the linear mixed model:

```
model.tables(eML.RI)["treatment",]
```

```
          estimate         se       df     lower    upper p.value
treatment 1.170829 0.09047721 2999.845 0.9934256 1.348233       0
```

pluging in formula Equation 1 the variance and correlation estimates based on the residuals:

```
sqrt(2*(1-mean(M.MLcor[lower.tri(M.MLcor)]))*mean(dfL$res^2)/(n.obs*n.
   times/2))
```

```
[1] 0.09047721
```

A Inverse of a compound symmetry matrix

Consider the compound symmetry matrix:

$$R = (1-\rho)I + \rho \mathbf{e}\mathbf{e}^\intercal = \rho\left(\frac{1-\rho}{\rho}I + \mathbf{e}\mathbf{e}^\intercal\right)$$

The Sherman-Morrison formula indicates that:

$$\begin{aligned} R^{-1} &= \rho^{-1}\left(\frac{\rho}{1-\rho}I - \frac{\rho^2}{(1-\rho)^2}\frac{\mathbf{e}\mathbf{e}^\intercal}{1+\frac{\rho}{1-\rho}\mathbf{e}^\intercal\mathbf{e}}\right) = \frac{1}{1-\rho}I - \frac{\rho}{(1-\rho)^2}\frac{\mathbf{e}\mathbf{e}^\intercal}{1+\frac{\rho}{1-\rho}p} \\ &= \frac{1}{1-\rho}I - \frac{\rho\mathbf{e}\mathbf{e}^\intercal}{(1-\rho)^2+\rho(1-\rho)p} = \frac{1}{1-\rho}\left(I - \frac{\rho\mathbf{e}\mathbf{e}^\intercal}{1+\rho(p-1)}\right) \end{aligned}$$

B Estimates in a random intercept model

The log-likelihood of a random intercept model can be written:

$$\mathcal{L}(\Theta|\mathbf{Y},\mathbf{T}) = \sum_{i=1}^{n}\left(-\frac{m}{2}\log(2\pi) - \frac{1}{2}\log|\Omega| - \frac{1}{2}(\mathbf{Y}_i-\boldsymbol{\mu}_i)^\intercal\Omega^{-1}(\mathbf{Y}_i-\boldsymbol{\mu}_i)\right)$$

and the corresponding restricted likelihood:

$$\mathcal{L}^R(\Theta|\mathbf{Y},\mathbf{T}) = \mathcal{L}(\Theta|\mathbf{Y},\mathbf{T}) + \frac{p}{2}\log(2\pi) - \frac{1}{2}\log\left(\left|\sum_{i=1}^{n}\mathbf{Z}_i^\intercal\Omega^{-1}\mathbf{Z}_i\right|\right)$$

where $\mathbf{Z}_i = (1,\mathbf{T}_i)$ is the design matrix w.r.t. subject i.

B.1 Mean parameters

The score equation w.r.t. the mean parameters is identical when considering the log-likelihood or the restricted log-likelihood. Using the expression of R^{-1} found in appendix B we get:

$$\begin{bmatrix}0\\0\end{bmatrix} = \begin{bmatrix}\sum_{i=1}^n e^\intercal\Omega^{-1}(\mathbf{Y}_i-\boldsymbol{\mu}_i))\\ \sum_{i=1}^n \mathbf{T}^\intercal\Omega^{-1}(\mathbf{Y}_i-\boldsymbol{\mu}_i)\end{bmatrix} = \begin{bmatrix}\frac{1}{\sigma^2(1-\rho)}\sum_{i=1}^n e^\intercal\left(I-\frac{\rho\mathbf{e}\mathbf{e}^\intercal}{1+\rho(p-1)}\right)(\mathbf{Y}_i-\boldsymbol{\mu}_i)\\ \frac{1}{\sigma^2(1-\rho)}\sum_{i=1}^n \mathbf{T}^\intercal\left(I-\frac{\rho\mathbf{e}\mathbf{e}^\intercal}{1+\rho(p-1)}\right)(\mathbf{Y}_i-\boldsymbol{\mu}_i)\end{bmatrix}$$

which is equivalent to:

$$\begin{aligned} \begin{bmatrix}0\\0\end{bmatrix} &= \begin{bmatrix}\sum_{i=1}^n\left(e^\intercal(\mathbf{Y}_i-\boldsymbol{\mu}_i) - \frac{\rho p\mathbf{e}^\intercal(\mathbf{Y}_i-\boldsymbol{\mu}_i)}{1+\rho(p-1)}\right)\\ \sum_{i=1}^n\left(\mathbf{T}^\intercal(\mathbf{Y}_i-\boldsymbol{\mu}_i) - \frac{\rho p_1\mathbf{e}^\intercal(\mathbf{Y}_i-\boldsymbol{\mu}_i)}{1+\rho(p-1)}\right)\end{bmatrix} \\ &= \begin{bmatrix}\left(1-\frac{\rho p}{1+\rho(p-1)}\right)\sum_{i=1}^n e^\intercal(\mathbf{Y}_i-\boldsymbol{\mu}_i)\\ \sum_{i=1}^n\mathbf{T}^\intercal(\mathbf{Y}_i-\boldsymbol{\mu}_i) - \frac{\rho p_1}{1+\rho(p-1)}\sum_{i=1}^n\mathbf{e}^\intercal(\mathbf{Y}_i-\boldsymbol{\mu}_i)\end{bmatrix} \end{aligned}$$

Using that $1 - \frac{\rho p}{1+\rho(p-1)} = 1 + \rho(p-1) - \rho p = 1 - \rho > 0$ and substracting p_1/p times equation 1 from equation 2 we get:

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n e^\intercal(\mathbf{Y}_i - \boldsymbol{\mu}_i) \\ \sum_{i=1}^n \mathbf{T}^\intercal(\mathbf{Y}_i - \boldsymbol{\mu}_i) - \frac{p_1}{p}\sum_{i=1}^n \mathbf{e}^\intercal(\mathbf{Y}_i - \boldsymbol{\mu}_i) \end{bmatrix}$$

Denoting the by $\widehat{\alpha} = \frac{1}{np}\sum_{i=1}^n\sum_{t=1}^p(1-T_{it})Y_{it}$ and $\widehat{\beta} = \frac{1}{np}\sum_{i=1}^n\sum_{t=1}^p T_{it}Y_{it} - \widehat{\alpha}$ the empirical mean over timepoints and patients under control and under treatment. The former equations are equivalent to:

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \widehat{\alpha} - \alpha + p_1(\widehat{\beta} - \beta) \\ p_1(\widehat{\alpha} + \widehat{\beta} - \alpha - \beta) - \frac{p_1}{p}(\widehat{\alpha} - \alpha + p_1(\widehat{\beta} - \beta)) \end{bmatrix}$$
$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \widehat{\alpha} - \alpha + (\widehat{\beta} - \beta) \\ (\widehat{\alpha} - \alpha + \widehat{\beta} - \beta) - \frac{1}{p}(\widehat{\alpha} - \alpha + p_1(\widehat{\beta} - \beta)) \end{bmatrix}$$

So $\widehat{\beta} - \beta = -\frac{1}{p_1}(\widehat{\alpha} - \alpha)$ and:

$$0 = (\widehat{\alpha} - \alpha)\left(1 - \frac{1}{p_1} - \frac{1}{p} + 1\right)$$

Since design $p_0 \geq 1$ and $p \geq 2$ so $2 - \frac{1}{p_1} - \frac{1}{p} \geq 0.5$. It follows that $\alpha = \widehat{\alpha}$ and therefore $\beta = \widehat{\beta}$: the maximum likelihood (ML) and restricted maximum likelihood (REML) estimates of the mean parameters are the empirical means in the appropriate subgroups.

B.2 Correlation parameter (ML)

The ML score equation w.r.t the correlation parameter is:

$$\begin{aligned} 0 &= -\frac{n}{2}tr\left(\Omega^{-1}\frac{\partial\Omega}{\partial\rho}\right) + \frac{1}{2}\sum_{i=1}^n(\mathbf{Y}_i - \widehat{\boldsymbol{\mu}}_i)^\intercal\Omega^{-1}\frac{\partial\Omega}{\partial\rho}\Omega^{-1}(\mathbf{Y}_i - \widehat{\boldsymbol{\mu}}_i) \\ &= -\frac{n}{2}tr\left(R^{-1}\frac{\partial R}{\partial\rho}\right) + \frac{1}{2\sigma^2}tr\left(R^{-1}\frac{\partial R}{\partial\rho}R^{-1}\sum_{i=1}^n(\mathbf{Y}_i - \widehat{\boldsymbol{\mu}}_i)(\mathbf{Y}_i - \widehat{\boldsymbol{\mu}}_i)^\intercal\right) \end{aligned}$$

We first explicit the first term:

$$\begin{aligned} R^{-1}\frac{\partial R}{\partial\rho} &= \frac{1}{1-\rho}\left(I - \frac{\rho\mathbf{e}\mathbf{e}^\intercal}{1+\rho(p-1)}\right)(-I + \mathbf{e}\mathbf{e}^\intercal) \\ &= \frac{1}{1-\rho}\left(-I + \mathbf{e}\mathbf{e}^\intercal + \frac{\rho\mathbf{e}\mathbf{e}^\intercal}{1+\rho(p-1)} - \frac{\rho p\mathbf{e}\mathbf{e}^\intercal}{1+\rho(p-1)}\right) \\ &= \frac{1}{1-\rho}\left(-I + \mathbf{e}\mathbf{e}^\intercal\frac{1+\rho(p-1)+\rho-\rho p}{1+\rho(p-1)}\right) \end{aligned}$$

$$= \frac{1}{1-\rho}\left(-I + \frac{\mathbf{e}\mathbf{e}^\intercal}{1+\rho(p-1)}\right)$$

Thus:

$$tr\left(R^{-1}\frac{\partial R}{\partial \rho}\right) = \frac{p}{1-\rho}\left(-1 + \frac{1}{1+\rho(p-1)}\right) = -\frac{p\rho(p-1)}{(1-\rho)(1+\rho(p-1))}$$

We now consider:

$$\begin{aligned}
R^{-1}\frac{\partial R}{\partial \rho}R^{-1} &= \frac{1}{(1-\rho)^2}\left(-I + \frac{\mathbf{e}\mathbf{e}^\intercal}{1+\rho(p-1)}\right)\left(I - \frac{\rho\mathbf{e}\mathbf{e}^\intercal}{1+\rho(p-1)}\right) \\
&= \frac{1}{(1-\rho)^2}\left(-I + \frac{\rho\mathbf{e}\mathbf{e}^\intercal}{1+\rho(p-1)} + \frac{\mathbf{e}\mathbf{e}^\intercal}{1+\rho(p-1)} - \frac{\rho p\mathbf{e}\mathbf{e}^\intercal}{(1+\rho(p-1))^2}\right) \\
&= \frac{1}{(1-\rho)^2}\left(-I + \mathbf{e}\mathbf{e}^\intercal\frac{\rho+\rho^2(p-1)+1+\rho(p-1)-\rho p}{(1+\rho(p-1))^2}\right) \\
&= \frac{1}{(1-\rho)^2}\left(-I + \mathbf{e}\mathbf{e}^\intercal\frac{\rho^2(p-1)+1}{(1+\rho(p-1))^2}\right)
\end{aligned}$$

We now consider the matrix $\frac{1}{n}\sum_{i=1}^n(\mathbf{Y}_i - \widehat{\boldsymbol{\mu}}_i)(\mathbf{Y}_i - \widehat{\boldsymbol{\mu}}_i)^\intercal$ and denote by:

- $\left(\widetilde{\sigma}_1^2, \ldots, \widetilde{\sigma}_p^2\right) = \left(\frac{1}{n}\sum_{i=1}^n(Y_{i,1} - \widehat{\mu}_{i,1})^2, \ldots, \frac{1}{n}\sum_{i=1}^n(Y_{i,p} - \widehat{\mu}_{i,p})^2\right)$ its diagonal elements. The tilde notation is used instead of the hat notation to stress that they generally differ from the time-specific empirical variance estimator (which would center the residuals at each timepoint). Note that their average equal the empirical residual variance: $\widehat{\sigma}^2 = \frac{1}{p}\sum_{j=1}^p \widetilde{\sigma}_j^2$.
- $\forall (j,j') \in \{1,\ldots,p\}$ such that $j \neq j'$, we denote the off diagonal elements by $\widehat{\sigma}^2\widehat{\rho}_{j,j'}$ where $\widehat{\rho}_{j,j'} = \widehat{\rho}_{j',j} = \frac{1}{n}\sum_{i=1}^n \frac{Y_{i,j}-\widehat{\mu}_{i,j}}{\widehat{\sigma}}\frac{Y_{i,j'}-\widehat{\mu}_{i,j'}}{\widehat{\sigma}}$ its off diagonal elements.

Then:

$$\begin{aligned}
&tr\left(R^{-1}\frac{\partial R}{\partial \rho}R^{-1}\sum_{i=1}^n(\mathbf{Y}_i - \widehat{\boldsymbol{\mu}}_i)(\mathbf{Y}_i - \widehat{\boldsymbol{\mu}}_i)^\intercal\right) \\
&= \frac{n\widehat{\sigma}^2}{(1-\rho)^2}\left(p\left(-1 + \frac{\rho^2(p-1)+1}{(1+\rho(p-1))^2}\right) + \frac{2\rho^2(p-1)+2}{(1+\rho(p-1))^2}\sum_{j<j'}\widehat{\rho}_{j,j'}\right) \\
&= \frac{n\widehat{\sigma}^2}{(1-\rho)^2}\left(p\left(\frac{-2\rho(p-1)-\rho^2(p-1)^2+\rho^2(p-1)}{(1+\rho(p-1))^2}\right) + \frac{2\rho^2(p-1)+2}{(1+\rho(p-1))^2}\sum_{j<j'}\widehat{\rho}_{j,j'}\right) \\
&= \frac{n\widehat{\sigma}^2}{(1-\rho)^2(1+\rho(p-1))^2}\left(p\rho(p-1)\left(-2-\rho(p-2)\right) + \left(2\rho^2(p-1)+2\right)\sum_{j<j'}\widehat{\rho}_{j,j'}\right)
\end{aligned}$$

The score equation becomes:

$$0 = \frac{np(p-1)}{2(1-\rho)^2(1+\rho(p-1))^2}\left(\rho(1-\rho)(1+\rho(p-1)) - \frac{\widehat{\sigma}^2}{\sigma^2}\rho\left(2+\rho(p-2)\right) + \frac{\rho^2(p-1)+1}{p(p-1)/2}\sum_{j<j'}\widehat{\rho}_{j,j'}\right)$$

$$= \frac{np(p-1)(\rho^2(p-1)+1)}{2(1-\rho)^2(1+\rho(p-1))^2}\left(\rho\frac{(1-\rho)(1+\rho(p-1)) - \frac{\widehat{\sigma}^2}{\sigma^2}\left(2+\rho(p-2)\right))}{\rho^2(p-1)+1} + \frac{1}{p(p-1)/2}\sum_{j<j'}\widehat{\rho}_{j,j'}\right)$$

Using that $(1-\rho)(1+\rho(p-1)) = 1-\rho+\rho(p-1)-\rho^2(p-1) = -\rho^2(p-1) + \rho(p-2)+1 = -(\rho^2(p-1)+1)+\rho(p-2)+2$, it follows that:

$$0 = \frac{np(p-1)\rho^2(p-1)+1}{2(1-\rho)^2(1+\rho(p-1))^2}\left(-\rho+\rho\frac{\rho(p-2)+2-\frac{\widehat{\sigma}^2}{\sigma^2}\left(2+\rho(p-2)\right))}{\rho^2(p-1)+1} + \frac{1}{p(p-1)/2}\sum_{j<j'}\widehat{\rho}_{j,j'}\right)$$

$$= \frac{np(p-1)\rho^2(p-1)+1}{2(1-\rho)^2(1+\rho(p-1))^2}\left(-\rho-\rho\left(\frac{\widehat{\sigma}^2}{\sigma^2}-1\right)\frac{2+\rho(p-2)}{1+\rho^2(p-1)} + \frac{1}{p(p-1)/2}\sum_{j<j'}\widehat{\rho}_{j,j'}\right)$$

Since the first term is strictly positive ($0<\rho<1$ and $p>1$) we can simplify and get that:

$$\frac{1}{p(p-1)/2}\sum_{j<j'}\widehat{\rho}_{j,j'} = \rho+\rho\left(\frac{\sigma^2}{\widehat{\sigma}^2}-1\right)\frac{2+\rho(p-2)}{1+\rho^2(p-1)} \tag{2}$$

B.3 Variance parameter (ML)

The ML score equation w.r.t the variance parameter is:

$$\begin{aligned}
0 &= -\frac{n}{2}tr\left(\Omega^{-1}\frac{\partial\Omega}{\partial\sigma^2}\right) + \frac{1}{2}\sum_{i=1}^{n}(\mathbf{Y}_i-\widehat{\boldsymbol{\mu}}_i)^\intercal\Omega^{-1}\frac{\partial\Omega}{\partial\sigma^2}\Omega^{-1}(\mathbf{Y}_i-\widehat{\boldsymbol{\mu}}_i)\\
&= -\frac{n}{2}tr\left(\sigma^{-2}R^{-1}R\right) + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(\mathbf{Y}_i-\widehat{\boldsymbol{\mu}}_i)^\intercal R^{-1}RR^{-1}(\mathbf{Y}_i-\widehat{\boldsymbol{\mu}}_i)\\
&= -\frac{np}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(\mathbf{Y}_i-\widehat{\boldsymbol{\mu}}_i)^\intercal R^{-1}(\mathbf{Y}_i-\widehat{\boldsymbol{\mu}}_i)
\end{aligned}$$

Using the expression of R^{-1} found in appendix B we get:

$$\begin{aligned}
0 &= -\frac{np}{2\sigma^2} + \frac{1}{2\sigma^4(1-\rho)}\sum_{i=1}^{n}(\mathbf{Y}_i-\widehat{\boldsymbol{\mu}}_i)^\intercal\left(I-\frac{\rho\mathbf{e}\mathbf{e}^\intercal}{(1-\rho)+\rho p}\right)(\mathbf{Y}_i-\widehat{\boldsymbol{\mu}}_i)\\
&= -\frac{np}{2\sigma^2} + \frac{1}{2\sigma^4(1-\rho)}\sum_{i=1}^{n}(\mathbf{Y}_i-\widehat{\boldsymbol{\mu}}_i)^\intercal(\mathbf{Y}_i-\widehat{\boldsymbol{\mu}}_i)\\
&\quad - \frac{\rho}{2\sigma^4(1-\rho)((1-\rho)+\rho p)}\sum_{i=1}^{n}(\mathbf{Y}_i-\widehat{\boldsymbol{\mu}}_i)^\intercal\mathbf{e}\mathbf{e}^\intercal(\mathbf{Y}_i-\widehat{\boldsymbol{\mu}}_i)\\
&= -\frac{np}{2\sigma^2} + \frac{np\widehat{\sigma}^2}{2\sigma^4(1-\rho)} - \frac{\rho np^2}{2\sigma^4(1-\rho)((1-\rho)+\rho p)}\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{p}\sum_{i=1}^{n}Y_{i,j}-\widehat{\mu}_{i,j}\right)^2
\end{aligned}$$

Since:

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{p}\sum_{j=1}^{p}Y_{i,j}-\widehat{\mu}_{i,j}\right)^2 = \frac{1}{np^2}\sum_{i=1}^{n}\sum_{j=1}^{p}\sum_{j'=1}^{p}\left(Y_{i,j}-\widehat{\mu}_j\right)\left(Y_{i,j'}-\widehat{\mu}_{j'}\right)$$

$$=\frac{\widehat{\sigma}^2}{p^2}\left(p+2\sum_{j<j'}\widehat{\rho}_{j,j'}\right)=\frac{\widehat{\sigma}^2}{p}\left(1+(p-1)\frac{1}{p(p-1)/2}\sum_{j<j'}\widehat{\rho}_{j,j'}\right)$$

we obtain:

$$0=-\frac{np}{2\sigma^2}+\frac{np\widehat{\sigma}^2}{2\sigma^4(1-\rho)}-\frac{\rho np\widehat{\sigma}^2}{2\sigma^4(1-\rho)((1-\rho)+\rho p)}\left(1+(p-1)\frac{1}{p(p-1)/2}\sum_{j<j'}\widehat{\rho}_{j,j'}\right)$$

$$0=-\frac{np\widehat{\sigma}^2}{2\sigma^4}\left(\frac{\sigma^2}{\widehat{\sigma}^2}-\frac{1}{(1-\rho)}+\frac{\rho}{(1-\rho)((1-\rho)+\rho p)}\left(1+(p-1)\frac{1}{p(p-1)/2}\sum_{j<j'}\widehat{\rho}_{j,j'}\right)\right)$$

Since

$$\begin{aligned}
&-\frac{1}{(1-\rho)}+\frac{\rho}{(1-\rho)^2+\rho(1-\rho)p}\left(1+(p-1)\frac{1}{p(p-1)/2}\sum_{j<j'}\widehat{\rho}_{j,j'}\right)\\
=&-\frac{1}{(1-\rho)}\left(1-\frac{\rho}{1+\rho(p-1)}-\frac{\rho(p-1)}{1+\rho(p-1)}\frac{1}{p(p-1)/2}\sum_{j<j'}\widehat{\rho}_{j,j'}\right)\\
=&-\frac{1}{(1-\rho)(1+\rho(p-1))}\left(1+\rho(p-2)-\rho(p-1)\frac{1}{p(p-1)/2}\sum_{j<j'}\widehat{\rho}_{j,j'}\right)
\end{aligned}$$

We get:

$$\begin{aligned}
0=&-\frac{np\widehat{\sigma}^2}{2\sigma^4}\left(\frac{\sigma^2}{\widehat{\sigma}^2}-\frac{1}{(1-\rho)(1+\rho(p-1))}\left(1+\rho(p-2)-\rho(p-1)\frac{1}{p(p-1)/2}\sum_{j<j'}\widehat{\rho}_{j,j'}\right)\right)\\
=&-\frac{np\widehat{\sigma}^2}{2\sigma^4}\left(\frac{\sigma^2}{\widehat{\sigma}^2}-\frac{1}{1+\rho(p-2)-\rho^2(p-1)}\left(1+\rho(p-2)-\rho(p-1)\frac{1}{p(p-1)/2}\sum_{j<j'}\widehat{\rho}_{j,j'}\right)\right)\\
=&-\frac{np\widehat{\sigma}^2}{2\sigma^4}\left(\frac{\sigma^2}{\widehat{\sigma}^2}-1-\frac{\rho(p-1)\left(\rho-\frac{1}{p(p-1)/2}\sum_{j<j'}\widehat{\rho}_{j,j'}\right)}{1+\rho(p-2)-\rho^2(p-1)}\right)
\end{aligned}$$

Since $\frac{np\widehat{\sigma}^2}{2\sigma^4}\neq 0$ and using equation Equation 2, we obtain:

$$0=\frac{\sigma^2}{\widehat{\sigma}^2}-1+\frac{\rho^2(p-1)\left(\frac{\sigma^2}{\widehat{\sigma}^2}-1\right)\frac{2+\rho(p-2)}{1+\rho^2(p-1)}}{1+\rho(p-2)-\rho^2(p-1)}=\left(\frac{\sigma^2}{\widehat{\sigma}^2}-1\right)\left(1+\frac{\rho^2(p-1)\frac{2+\rho(p-2)}{1+\rho^2(p-1)}}{(1-\rho)(1+\rho(p-1))}\right)$$

The second term is strictly positive: it is clear when $p>2$ because all terms are positive or null and one is added. When $p=1$ then $2+\rho(p-2)=2-\rho>0$ because $\rho<1$. So we must have $\sigma^2=\widehat{\sigma}^2$. Plugging this value in the score equation for the correlation parameter leads to:

$$\rho=\frac{1}{p(p-1)/2}\sum_{j<j'}\widehat{\rho}_{j,j'}$$

C Standard error of the treatment effect in a balanced random intercept model

The standard error of the treatment effect based on the expected information is the last element (i.e. second row, second column) of the variance-covariance matrix $(X^\intercal \Omega^{-1} X)^{-1}$.

Using the expression of the inverse of R given in appendix A, we get that its matrix product with the $p \times 2$ matrix $X = (1, T)$ where T is either 0 or 1 (respectively p_0 and p_1 times) is:

$$
\begin{aligned}
X^\intercal R^{-1} X &= \frac{1}{1-\rho} X^\intercal X - \frac{\rho X^\intercal \mathbf{e}\mathbf{e}^\intercal X}{(1-\rho)^2 + \rho(1-\rho)p} \\
&= \frac{1}{1-\rho}\left(X^\intercal X - \frac{\rho X^\intercal \mathbf{e}\mathbf{e}^\intercal X}{1+\rho(p-1)}\right) \\
&= \frac{1}{1-\rho}\left(\begin{bmatrix} p & p_1 \\ p_1 & p_1 \end{bmatrix} - \frac{\rho}{1+\rho(p-1)}\begin{bmatrix} p^2 & pp_1 \\ pp_1 & p_1^2 \end{bmatrix}\right) \\
&= \frac{1}{(1-\rho)(1+\rho(p-1))}\begin{bmatrix} p + p\rho(p-1) - \rho p^2 & p_1 + p_1\rho(p-1) - \rho p p_1 \\ p_1 + p_1\rho(p-1) - \rho p p_1 & p_1 + p_1\rho(p-1) - \rho p_1^2 \end{bmatrix} \\
&= \frac{1}{(1-\rho)(1+\rho(p-1))}\begin{bmatrix} p(1-\rho) & p_1(1-\rho) \\ p_1(1-\rho) & p_1(1+\rho(p-p_1-1)) \end{bmatrix}
\end{aligned}
$$

whose inverse is:

$$
\begin{aligned}
\left(X^\intercal R^{-1} X\right)^{-1} &= \frac{(1-\rho)(1+\rho(p-1))}{p_1 p(1-\rho)(1+\rho(p-p_1-1)) - p_1^2(1-\rho)^2}\begin{bmatrix} p_1(1+\rho(p-p_1-1)) & -p_1(1-\rho) \\ -p_1(1-\rho) & p(1-\rho) \end{bmatrix} \\
&= \frac{1+\rho(p-1)}{p_1 p(1+\rho(p-p_1-1)) - p_1^2(1-\rho)}\begin{bmatrix} p_1(1+\rho(p-p_1-1)) & -p_1(1-\rho) \\ -p_1(1-\rho) & p(1-\rho) \end{bmatrix} \\
&= \frac{1+\rho(p-1)}{(p-p_1) + \rho(p^2 - pp_1 - p + p_1)}\begin{bmatrix} 1+\rho(p-p_1-1) & -(1-\rho) \\ -(1-\rho) & \frac{p}{p_1}(1-\rho) \end{bmatrix} \\
&= \frac{1}{p-p_1}\begin{bmatrix} 1+\rho(p-p_1-1) & -(1-\rho) \\ -(1-\rho) & \frac{p}{p_1}(1-\rho) \end{bmatrix}
\end{aligned}
$$

So in the random intercept model, the standard error of the treatment estimator will be:

$$
\sigma_{\widehat{\beta}} = \sqrt{\sigma_0^2(1-\rho)\frac{p}{np_1(p-p_1)}} = \sqrt{\frac{\delta}{n}\frac{p}{p_1(p-p_1)}}
$$

In a design with as many observations under treatment as under control $p_1 = p/2$

and the expression simplifies into.

$$\sigma_{\widehat{\beta}} = \sqrt{\frac{4\delta}{np}} = \sqrt{\frac{2\delta}{np_1}}$$

From appendix B we deduce that in a balanced design the standard error of the Maximum Likelihood estimator is:

$$\sigma_{\widehat{\beta}} = \sqrt{\frac{\left(1 - \frac{1}{p(p-1)/2} \sum_{j<j'} \rho_{j,j'}\right) \sigma^2}{n} \frac{p}{p_1(p-p_1)}}$$

which in a design with as many observations under treatment as under control simplifies to:

$$\sigma_{\widehat{\beta}} = \sqrt{\frac{2\left(1 - \frac{1}{p(p-1)/2} \sum_{j<j'} \rho_{j,j'}\right) \sigma^2}{np_1}}$$