GLS and LME

Peter Hoff
Duke STA 610

LME - Is it worth it?

```
fitOLS<-lm(y.nels ~ flp.nels + ses.nels + flp.nels*ses.nels)
summary(fitOLS)

##
## Call:
## lm(formula = y.nels ~ flp.nels + ses.nels + flp.nels * ses.nels)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -36.107  -5.758   0.142   5.977  33.538
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        54.8442     0.2280  240.50   <2e-16 ***
## flp.nels           -2.0809     0.1075  -19.36   <2e-16 ***
## ses.nels            4.9058     0.2810   17.46   <2e-16 ***
## flp.nels:ses.nels  -0.1279     0.1361   -0.94    0.347
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.754 on 12970 degrees of freedom
## Multiple R-squared:  0.2028, Adjusted R-squared:  0.2026
## F-statistic:  1100 on 3 and 12970 DF,  p-value: < 2.2e-16
```

LME - Is it worth it?

```
fitLME<-lmer(y.nels ~ flp.nels + ses.nels + flp.nels:ses.nels + (ses.nels|g.nels)  )
summary(fitLME)

## Linear mixed model fit by REML ['lmerMod']
## Formula: y.nels ~ flp.nels + ses.nels + flp.nels:ses.nels + (ses.nels |
##     g.nels)
##
## REML criterion at convergence: 92388.1
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.9769 -0.6415  0.0198  0.6659  4.5206
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  g.nels   (Intercept)  9.056   3.009
##           ses.nels     1.602   1.266    0.06
##  Residual             67.258   8.201
## Number of obs: 12974, groups:  g.nels, 684
##
## Fixed effects:
##                   Estimate Std. Error t value
## (Intercept)        55.3989     0.3866 143.285
## flp.nels           -2.4070     0.1822 -13.212
## ses.nels            4.4899     0.3333  13.472
## flp.nels:ses.nels  -0.1931     0.1590  -1.214
##
## Correlation of Fixed Effects:
##             (Intr) flp.nl ss.nls
## flp.nels    -0.930
## ses.nels    -0.157  0.088
## flp.nls:ss.  0.085 -0.007 -0.926
```

LME - Is it worth it?

For the mixed effects model

$$\mathbf{y}_j = \mathbf{X}_j\boldsymbol{\beta} + \mathbf{Z}_j\mathbf{a}_j + \boldsymbol{\epsilon}_j,$$

the OLS estimate is still *unbiased*. However,

- it is no longer the BLUE;
- its variance is no longer $\sigma^2(\mathbf{X}^\top\mathbf{X})^{-1}$.

For this model,

- the BLUE is (approximately) the MLE returned by `lmer`;
- the standard errors for fixed effects account for within-group correlation;
- the standard errors for the macro-level fixed effects can be very different.

Linear unbiased estimators

Model:

- $\mathbf{y} = \mathbf{X}\beta + \epsilon$
- $\mathrm{E}[\epsilon|\mathbf{X}] = \mathbf{0}, \mathrm{Var}[\epsilon|\mathbf{X}] = \sigma^2\mathbf{I}$.

or equivalently,

- $\mathrm{E}[\mathbf{y}|\mathbf{X}] = \mathbf{X}\beta$
- $\mathrm{Var}[\mathbf{y}|\mathbf{X}] = \sigma^2\mathbf{I}$

OLS Estimator: $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}$

Linear unbiased estimators: $\check{\boldsymbol{\beta}} = [(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top + \mathbf{H}^\top]\mathbf{y}$, where $\mathbf{H}^\top\mathbf{X} = \mathbf{0}$.

Exercise: Show that $\check{\boldsymbol{\beta}}$, and hence $\hat{\boldsymbol{\beta}}$, are unbiased.

Variance of linear unbiased estimators

Let $\mathbf{X}^+ = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$

$$\begin{aligned}
\text{Var}[\check{\beta}] &= (\mathbf{X}^+ + \mathbf{H}^\top)\text{Var}[\epsilon](\mathbf{X}^+ + \mathbf{H}^\top)^\top \\
&= \sigma^2(\mathbf{X}^+ + \mathbf{H}^\top)(\mathbf{X}^+ + \mathbf{H}^\top)^\top \\
&= \sigma^2\left(\mathbf{X}^+(\mathbf{X}^+)^\top + \mathbf{X}^+\mathbf{H} + \mathbf{H}^\top(\mathbf{X}^+)^\top + \mathbf{H}^\top\mathbf{H}\right).
\end{aligned}$$

Now calculate the individual terms:

$$\begin{aligned}
\mathbf{X}^+(\mathbf{X}^+)^\top &= (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1} \\
&= (\mathbf{X}^\top\mathbf{X})^{-1}, \\
\mathbf{H}^\top(\mathbf{X}^+)^\top &= \mathbf{H}^\top\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1} \\
&= \mathbf{0}.
\end{aligned}$$

So

$$\begin{aligned}
\text{Var}[\check{\beta}] &= \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1} + \sigma^2\mathbf{H}^\top\mathbf{H} \\
&= \text{Var}[\hat{\beta}] + \sigma^2\mathbf{H}^\top\mathbf{H}.
\end{aligned}$$

Gauss-Markov theorem

Definition (Loewner order)

For two positive semidefinite matrices Σ_1 and Σ_2 of the same size, we say that $\Sigma_1 > \Sigma_2$ if $\Sigma_1 - \Sigma_2$ is positive definite, and that $\Sigma_1 \geq \Sigma_2$ if $\Sigma_1 - \Sigma_2$ is positive semidefinite.

Theorem

Let $\check{\boldsymbol{\beta}}$ be a linear unbiased estimator of $\boldsymbol{\beta}$ in a linear model where $E[\mathbf{y}] = \mathbf{X}\boldsymbol{\beta}, \boldsymbol{\beta} \in \mathbb{R}^p$ and $Var[\mathbf{y}] = \sigma^2\mathbf{I}$, $\sigma^2 > 0$. Then

$$Var[\check{\boldsymbol{\beta}}] \geq Var[\hat{\boldsymbol{\beta}}],$$

where $\hat{\boldsymbol{\beta}}$ is the OLS estimator.

The OLS estimator is the BLUE in this case.

Non-isotropic variance

What if $\text{Var}[\mathbf{y}] \neq \sigma^2 \mathbf{I}$?

- Heteroscedasticity: $\text{Var}[\mathbf{y}_i] = w_i \sigma^2$ for some known $w_1, \ldots, w_n$.
- Time series: $\text{Var}[\mathbf{y}] = \sigma^2 \mathbf{A}$, where $a_{i,j} = \rho^{|i-j|}$.

LME models can be viewed as models for correlated data. Let

$$\mathbf{y}_j = \mathbf{X}_j \boldsymbol{\beta} + \mathbf{Z}_j \mathbf{a}_j + \boldsymbol{\epsilon}_j$$

where

- $\text{E}[\mathbf{a}_j] = \mathbf{0}, \text{Var}[\mathbf{a}_j] = \Psi$.
- $\text{E}[\boldsymbol{\epsilon}_j] = \mathbf{0}, \text{Var}[\boldsymbol{\epsilon}_j] = \sigma^2 \mathbf{I}$.
- $\text{E}[\boldsymbol{\epsilon}_j \mathbf{a}_j^\top] = \mathbf{0}$.

Then

$$\begin{aligned} \text{E}[\mathbf{y}_j] &= \mathbf{X}_j \boldsymbol{\beta} \\ \text{Var}[\mathbf{y}_j] &= \mathbf{Z}_j \Psi \mathbf{Z}_j^\top + \sigma^2 \mathbf{I}. \end{aligned}$$

OLS with dependent data

The OLS estimator is still unbiased when data are correlated:

$$\begin{aligned} \mathrm{E}[\hat{\boldsymbol{\beta}}] = \mathrm{E}[(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}] &= (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathrm{E}[\mathbf{X}^\top\boldsymbol{\beta} + \boldsymbol{\epsilon}] \\ &= \boldsymbol{\beta} + \mathbf{0} = \boldsymbol{\beta}. \end{aligned}$$

However, its variance in this case is complicated:

$$\begin{aligned} \mathrm{Var}[\hat{\boldsymbol{\beta}}] = \mathrm{Var}[(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}] &= (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathrm{Var}[\mathbf{y}]\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{V}\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}. \end{aligned}$$

This is quite messy, and not equal to $\sigma^2(\mathbf{X}^\top\mathbf{X})^{-1}$ unless $\mathrm{Var}[\mathbf{y}]$ is special.

GLS estimator

Let $\text{Var}[\mathbf{y}] = \text{Var}[\epsilon] = \sigma^2\mathbf{V}$. We define the *symmetric square root* $\mathbf{V}^{1/2}$ of $\mathbf{V}$ as

$$\mathbf{V}^{1/2} = \mathbf{E}\mathbf{\Lambda}^{1/2}\mathbf{E}^\top.$$

where $(\mathbf{E}, \mathbf{\Lambda})$ are the eigenvectors and values of Σ. Note that$\mathbf{V}^{1/2}\mathbf{V}^{1/2} = \mathbf{V}$.

$\mathbf{V}^{-1/2}$ is a *whitening matrix* for $\mathbf{y}$:

$$\begin{aligned}\text{Var}[\mathbf{V}^{-1/2}\mathbf{y}] &= \mathbf{V}^{-1/2}\text{Var}[\mathbf{y}]\mathbf{V}^{-\top/2} \\ &= \mathbf{V}^{-1/2}(\sigma^2\mathbf{V})\mathbf{V}^{-1/2} = \sigma^2\mathbf{I}.\end{aligned}$$

OLS with whitened data

Let $\tilde{\mathbf{y}} = \mathbf{V}^{-1/2}\mathbf{y}$. The linear model for $\tilde{\mathbf{y}}$ is then

$$\mathbf{V}^{-1/2}\mathbf{y} = \mathbf{V}^{-1/2}\mathbf{X}\boldsymbol{\beta} + \mathbf{V}^{-1/2}\boldsymbol{\epsilon}$$
$$\tilde{\mathbf{y}} = \tilde{\mathbf{X}}\boldsymbol{\beta} + \tilde{\boldsymbol{\epsilon}},$$

where $\mathrm{E}[\tilde{\epsilon}] = \mathbf{0}$ and

$$\mathrm{Var}[\tilde{\epsilon}] = \sigma^2\mathbf{V}^{-1/2}\mathbf{V}\mathbf{V}^{-1/2} = \sigma^2\mathbf{I}.$$

The BLUE based on $\tilde{\mathbf{y}}$, $\tilde{\mathbf{X}}$ is

$$\hat{\boldsymbol{\beta}}_V = (\tilde{\mathbf{X}}^\top\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}^\top\tilde{\mathbf{y}}$$

GLS via whitened OLS

$\hat{\boldsymbol{\beta}}_V$ is linear in $\mathbf{y}$! So $\hat{\boldsymbol{\beta}}_V$ is the BLUE of $\boldsymbol{\beta}$, based on either $\tilde{\mathbf{y}}$ or $\mathbf{y}$.

On the original scale of the data, we have

$$\begin{aligned}\hat{\boldsymbol{\beta}}_V &= (\tilde{\mathbf{X}}^\top \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^\top \tilde{\mathbf{y}} \\ &= (\mathbf{X}^\top \mathbf{V}^{-1/2} \mathbf{V}^{-1/2} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{V}^{-1/2} \mathbf{V}^{-1/2} \mathbf{y} \\ &= (\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{V}^{-1} \mathbf{y}.\end{aligned}$$

This estimator is the *generalized least squares* (GLS) estimator of $\boldsymbol{\beta}$. Its variance is

$$\text{Var}[\hat{\boldsymbol{\beta}}_V] = \sigma^2 (\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1}.$$

Gauss-Markov-Aitkin theorem

Theorem

Let $\check{\boldsymbol{\beta}}$ be a linear unbiased estimator of $\boldsymbol{\beta}$ in a linear model with $E[\mathbf{y}] = \mathbf{X}\beta$, $Var[\mathbf{y}] = \sigma^2\mathbf{V}$ for $(\boldsymbol{\beta}, \sigma^2) \in \mathbb{R}^p \times \mathbb{R}^+$ with $\mathbf{X}$ and $\mathbf{V}$ known. Then

$$Var[\check{\boldsymbol{\beta}}] \geq \sigma^2(\mathbf{X}\mathbf{V}^{-1}\mathbf{X}^\top)^{-1} = Var[\hat{\boldsymbol{\beta}}_V],$$

where $\hat{\boldsymbol{\beta}}_V = (\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{y}$.

LME model as a GLM

Within-groups model: $\mathbf{y}_j = \mathbf{X}_j\boldsymbol{\beta} + \mathbf{Z}_j\mathbf{a}_j + \boldsymbol{\epsilon}_j$.

- $\mathrm{E}[\mathbf{y}_j] = \mathbf{X}_j\boldsymbol{\beta}$;
- $\mathrm{Var}[\mathbf{y}_j] = \mathbf{Z}_j\Psi\mathbf{Z}_j^\top + \sigma^2 I_{n_j}$.

Let

- $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_m) \in \mathbb{R}^{\sum n_j}$;
- $\mathbf{X} = (\mathbf{X}_1^\top \cdots \mathbf{X}_m^\top)^\top \in \mathbb{R}^{\sum n_j \times p}$.

Then

$$\mathrm{E}[\mathbf{y}] = \mathbf{X}\beta$$

$$\mathrm{Var}[\mathbf{y}] = \begin{pmatrix} \mathbf{Z}_1\Psi\mathbf{Z}_1^\top & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2\Psi\mathbf{Z}_2^\top & \cdots & \mathbf{0} \\ \vdots & & & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{Z}_m\Psi\mathbf{Z}_m^\top \end{pmatrix} + \sigma^2\mathbf{I} \equiv \Gamma.$$

Iterative estimation procedures

Typically estimation of $(\boldsymbol{\beta}, \Psi, \sigma^2)$ is done in two or more stages:

Feasible GLS:

1. Estimate Ψ and $\Omega = \text{Var}[\mathbf{y}]$
   1.1 Find $\mathbf{N}$ so that $\mathbf{N}^\top \mathbf{X} = \mathbf{0}$;
   1.2 Let $\mathbf{e} = \mathbf{N}^\top \mathbf{y}$ so that $\text{E}[\mathbf{e}] = \mathbf{0}$, $\text{Var}[\mathbf{e}] = \mathbf{N}^\top \Omega \mathbf{N}$
   1.3 Obtain estimate $\hat{\Psi}$ of Ψ from $\mathbf{e}$.
   1.4 Construct the estimate $\hat{\Omega}$ of Ω using $\hat{\Psi}$.
2. Estimate $\hat{\boldsymbol{\beta}}$ using feasible GLS:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \hat{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \hat{\Omega}^{-1} \mathbf{y}.$$

Simulation study

```
m<-11
n<-7

g<-rep(1:m,times=rep(n,m))
xg<-rnorm(m)[g]
xn<-rnorm(m*n)
X<-cbind(1,xg,xn)

X[1:25,]

##               xg          xn
##  [1,] 1 -0.6264538  0.38984324
##  [2,] 1 -0.6264538 -0.62124058
##  [3,] 1 -0.6264538 -2.21469989
##  [4,] 1 -0.6264538  1.12493092
##  [5,] 1 -0.6264538 -0.04493361
##  [6,] 1 -0.6264538 -0.01619026
##  [7,] 1 -0.6264538  0.94383621
##  [8,] 1  0.1836433  0.82122120
##  [9,] 1  0.1836433  0.59390132
## [10,] 1  0.1836433  0.91897737
## [11,] 1  0.1836433  0.78213630
```

Simulation study

```
tau<-2 ; beta<-c(1,2,3)

y<- X%*%beta + tau*rnorm(m)[g] + rnorm(m*n)
```

If we (incorrectly) ignored grouping, we would assume

- $\hat{\boldsymbol{\beta}}_{OLS}$ is optimal;
- $\text{Var}[\hat{\beta}_{OLS}] = (\mathbf{X}^\top\mathbf{X})^{-1}$.

```
VBI<-solve(t(X)%*%X)
VBI

##                          xg           xn
##     0.014382218 -0.005053766 -0.001140851
## xg -0.005053766  0.019830831 -0.000673228
## xn -0.001140851 -0.000673228  0.016077814

sqrt(diag(VBI))

##                  xg        xn
## 0.1199259 0.1408220 0.1267983
```

Simulation study

```
beta<-c(1,2,3)
BOLS<-BGLS<-NULL
for(s in 1:1000){

  y<- X%*%beta + tau*rnorm(m)[g] + rnorm(m*n)

  BOLS<-rbind(BOLS,lm(y ~ -1 + X )$coef)
  BGLS<-rbind(BGLS,fixef(lmer(y ~ -1 + X + (1|g)  )))
}

apply(BOLS,2,sd)

##         X       Xxg       Xxn
## 0.6288288 0.7527230 0.2116257

apply(BGLS,2,sd)

##         X       Xxg       Xxn
## 0.6287044 0.7525413 0.1355320
```

Simulation study

The simulation results match the theory:

```
V<-diag(n*m) + tau^2*kronecker(diag(m),matrix(1,n,n))

## Actual variance of betaOLS
IX<-solve(t(X)%*%X)%*%t(X)
VOLS<-IX%*%V%*%t(IX)
sqrt(diag(VOLS))

##                  xg        xn
## 0.6441631 0.7578585 0.2039284

## Actual variance of betaGLS
VGLS<-solve(t(X)%*%solve(V)%*%X)
sqrt(diag(VGLS))

##                  xg        xn
## 0.6440670 0.7578301 0.1304160
```

MLEs and Bayes

Recall the HNM:

$$y_{i,j} = \mu + a_j + \epsilon_{i,j}$$

Our estimators for μ and the a_j's were

- the MLE/WLS estimate for μ;
- Bayes estimators for a_j's (or equvalently, $\theta_j = \mu + a_j$).

Similarly, for the HLM

$$y_{i,j} = \mathbf{x}_{i,j}^\top \boldsymbol{\beta} + \mathbf{z}_{i,j}^\top \mathbf{a}_j + \epsilon_{i,j}$$

the estimators we've discussed are

- MLE/GLS for $\boldsymbol{\beta}$;
- Bayes estimators for $\mathbf{a}_j$'s (or equvalently, $\boldsymbol{\beta}_j = \boldsymbol{\beta} + \mathbf{a}_j$).

OLS, MLE and improper Bayes

Recall that for the *non-hierarchical model*

$$y_i = \mu + \epsilon_i$$

$\hat{\mu} = \bar{y}$ is the

- OLS estimator;
- MLE under normality;
- Bayes estimator under normality and a flat prior.

Similarly, for the *non-hierarchical model*

$$y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i$$

$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$ is the

- OLS estimator;
- MLE under normality;
- Bayes estimator under normality and a flat prior.

BLUEs BLUPs and Bayes

For the HLM

$$y_{i,j} = \mathbf{x}_{i,j}^\top \boldsymbol{\beta} + +\mathbf{z}_{i,j}^\top \mathbf{a}_j \ \epsilon_{i,j}$$

we want to get the

- MLE for $\boldsymbol{\beta}$;
- the Bayes estimator for $\mathbf{a}_j$, under the prior $\mathbf{a}_j \sim N(\mathbf{0}, \Psi)$.

So consider the Bayes estimator of $(\boldsymbol{\beta}, \mathbf{a})$ under the "semi-improper" prior

$$\begin{pmatrix} \boldsymbol{\beta} \\ \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_m \end{pmatrix} \sim N(\mathbf{0}, \mathbf{V})$$

where

$$\mathbf{V} = \begin{pmatrix} \infty\mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \Psi & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \cdots & \cdots & \Psi \end{pmatrix}.$$

Henderson's equations

Surprisingly (?) it turns out the BLUE/BLUP of $(\boldsymbol{\beta}, \mathbf{a}_1, \ldots, \mathbf{a}_m)$ is given by the posterior mean estimator under the semi-improper prior.

$$\begin{pmatrix} \hat{\boldsymbol{a}} \\ \hat{\boldsymbol{\beta}} \end{pmatrix} = \begin{pmatrix} \mathbf{Z}^\top \mathbf{Z}/\sigma^2 + \boldsymbol{\Psi}^{-1} & \mathbf{Z}^\top \mathbf{X}/\sigma^2 \\ \mathbf{X}^\top \mathbf{Z}/\sigma^2 & \mathbf{X}^\top \mathbf{X}/\sigma^2 \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{Z}^\top \mathbf{y}/\sigma^2 \\ \mathbf{X}^\top \mathbf{y}/\sigma^2 \end{pmatrix}. \tag{1}$$

This result has practical implications: Computing $\hat{\boldsymbol{\beta}}$ via the GLS equation seems to require inversion of $\text{Var}[\mathbf{y}]$, which is $nm \times nm$. The matrix here is of dimension $(p + mq) \times (p + mq)$