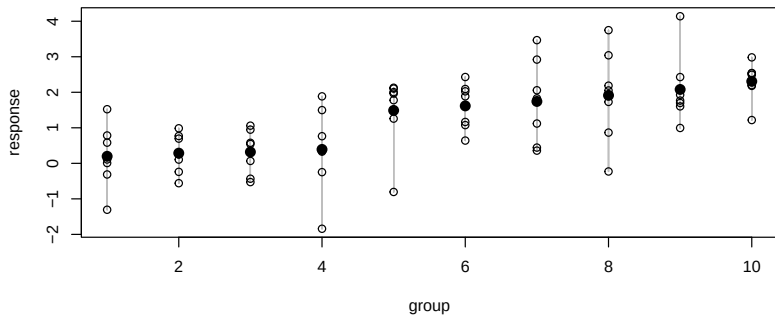


Variance component estimation methods

Peter Hoff
Duke STA 610



REML = TRUE

```
fitREML<-lmer( y ~ 1 + (1 | g) )

summary(fitREML)

## Linear mixed model fit by REML ['lmerMod']
## Formula: y ~ 1 + (1 | g)
##
## REML criterion at convergence: 208.2
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.51361 -0.48196  0.09502  0.52383  2.33192
##
## Random effects:
##   Groups      Name      Variance Std.Dev.
##    g      (Intercept)  0.5727   0.7568
##   Residual              0.9026   0.9500
## Number of obs: 70, groups:  g, 10
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)   1.2346    0.2649   4.661
```

REML = FALSE

```
fitML<-lmer( y ~ 1 + (1 | g), REML=FALSE )

summary(fitML)

## Linear mixed model fit by maximum likelihood ['lmerMod']
## Formula: y ~ 1 + (1 | g)
##
##           AIC          BIC    logLik deviance df.resid
##      213.4       220.1   -103.7    207.4         67
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.5317 -0.4807  0.1076  0.5275  2.3501
##
## Random effects:
##  Groups   Name                Variance Std.Dev.
##  g        (Intercept)  0.5025     0.7089
## Residual                    0.9026     0.9500
## Number of obs: 70, groups:  g, 10
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)   1.2346     0.2513   4.913
```

Estimator and calculation methods

$$\begin{aligned}y_{i,j} &= \mu + a_j + \epsilon_{i,j} \\ a_1, \dots, a_m &\sim \text{i.i.d. } N(0, \tau^2) \\ \{\epsilon_{i,j}\} &\sim \text{i.i.d. } N(0, \sigma^2)\end{aligned}$$

Parameters: μ, σ^2, τ^2 .

Variance component estimators: Estimators of σ^2, τ^2 .

- Unbiased moment estimators (ANOVA).
- Maximum likelihood estimators.
- Restricted maximum likelihood estimators (?)

Moment estimators

Recall the unbiased ANOVA-based method of moments estimators for μ and σ^2 :

$$\hat{\mu} = \bar{y} = \sum_j \sum_i y_{i,j} / (n \times m)$$
$$\hat{\sigma}^2 = \sum_j s_j^2 / m.$$

We showed that $E[\hat{\mu}] = \mu$ and $E[\hat{\sigma}^2] = \sigma^2$. Also,

$$MSA = n \times \text{sample variance}(\bar{y}_1, \dots, \bar{y}_m)$$
$$E[MSA] = \sigma^2 + n \times \tau^2$$

This suggests we estimate τ^2 as

$$\hat{\tau}^2 = (MSA - \hat{\sigma}^2) / n,$$

in which case we have $E[\hat{\tau}^2] = \tau^2$.

These are *moment-based estimators*. Normality not required for unbiasedness.

```

mean(y)

## [1] 1.234627

anova(lm(y ~ as.factor(g)))

## Analysis of Variance Table
##
## Response: y
##           Df Sum Sq Mean Sq F value    Pr(>F)
## as.factor(g)  9 44.203   4.9115   5.4416 1.924e-05 ***
## Residuals    60 54.155   0.9026
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

MSA<-n*var( tapply(y,g,mean) )

MSW<-mean( tapply(y,g,var) )

MSA

## [1] 4.91148

MSW

## [1] 0.9025845

t2hat<-(MSA-MSW)/n

t2hat

## [1] 0.5726993

```

Maximum likelihood estimation

$$\begin{aligned}y_{i,j} &= \mu + a_j + \epsilon_{i,j} \\ \epsilon_{1,j}, \dots, \epsilon_{n_j,j} &\sim \text{iid } N(0, \sigma^2) \\ a_j &\sim N(0, \tau^2)\end{aligned}$$

As we've discussed, the $y_{i,j}$'s are normal with

- $E[y_{i,j}|\mu] = \mu$
- $\text{Var}[y_{i,j}|\mu] = \sigma^2 + \tau^2$
- $\text{Cov}[y_{i_1,j}, y_{i_2,j}|\mu] = \tau^2$

In vector form, we can express this as follows:

$$E[\mathbf{y}_j|\mu] = \begin{pmatrix} \mu \\ \mu \\ \vdots \\ \mu \end{pmatrix} = \mu \mathbf{1} \quad \text{Cov}[\mathbf{y}_j|\mu] = \begin{pmatrix} \sigma^2 + \tau^2 & \tau^2 & \dots & \tau^2 \\ \tau^2 & \sigma^2 + \tau^2 & \dots & \tau^2 \\ \vdots & \vdots & \dots & \vdots \\ \tau^2 & \tau^2 & \dots & \sigma^2 + \tau^2 \end{pmatrix}$$

Computing the log-likelihood

MLEs of (μ, σ^2, τ^2) can be found by maximizing the log likelihood.

Log likelihood:

$$\begin{aligned} L(\mathbf{y} : \mu, \sigma^2, \tau^2) &= p(\mathbf{y}_1, \dots, \mathbf{y}_m | \mu, \sigma^2, \tau^2) \\ l(\mathbf{y} : \mu, \sigma^2, \tau^2) &= \log p(\mathbf{y}_1, \dots, \mathbf{y}_m | \mu, \sigma^2, \tau^2) \\ &= \log \prod_{j=1}^m p(\mathbf{y}_j | \mu, \sigma^2, \tau^2) \\ &= \sum_{j=1}^m \log p(\mathbf{y}_j | \mu, \sigma^2, \tau^2), \end{aligned}$$

where $\log p(\mathbf{y}_j | \mu, \sigma^2, \tau^2)$ is the log of a multivariate normal density.

Maximum likelihood estimates

```
fitML<-lmer(y ~ 1 + (1|g), REML=FALSE )  
  
fixef(fitML)  
  
## (Intercept)  
##      1.234627  
  
as.data.frame( VarCorr(fitML) )  
  
##      grp      var1 var2      vcov      sdcor  
## 1      g (Intercept) <NA> 0.5025353 0.7088973  
## 2 Residual      <NA> <NA> 0.9025845 0.9500445
```

Notice:

- The estimate of μ is the same as the unbiased moment estimator.
- The estimate of τ^2 is different.

In fact, the MLE is *biased* for this variance components:

$$E[\hat{\tau}_{MLE}^2] \neq \tau^2.$$

Restricted maximum likelihood

The default estimates in `lme4` are the REML estimates.

```
fit0<-lmer(y ~ 1 + (1|g) )  
  
fixef(fit0)  
  
## (Intercept)  
##      1.234627  
  
as.data.frame( VarCorr(fit0) )  
  
##           grp          var1 var2          vcov          sdcor  
## 1           g (Intercept) <NA> 0.5726994 0.7567690  
## 2 Residual          <NA> <NA> 0.9025845 0.9500445
```

Notice that for this model, the estimates are the same as the unbiased moment estimates we got from ANOVA:

```
mean(y)  
  
## [1] 1.234627  
  
MSW  
  
## [1] 0.9025845  
  
t2hat  
  
## [1] 0.5726993
```

REML

REML estimates maximize an alternative “restricted” likelihood function.

- In some cases they are equivalent to unbiased moment estimators.
- In others they are different, but generally have lower bias than MLEs.

The main idea:

- Bias in MLEs of variances arise from “plugging in” the mean estimates.
- Restricted likelihoods don’t involve mean parameters, so avoid “plug-in” bias.

Simplest REML example

One sample normal model

$$y_1, \dots, y_n \sim \text{i.i.d. } N(\mu, \sigma^2)$$

MLEs

$$p(\mathbf{y}|\mu, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp(-\sum (y_i - \mu)^2 / [2\sigma^2])$$
$$-2 \times \log p(\mathbf{y}|\mu, \sigma^2) + c = n \log \sigma^2 + \sum (y_i - \mu)^2 / \sigma^2$$

Exercise: Show that the MLEs $(\hat{\mu}_{MLE}, \hat{\sigma}_{MLE}^2)$ of (μ, σ^2) are

- $\hat{\mu}_{MLE} = \bar{y}$;
- $\hat{\sigma}_{MLE}^2 = \sum_i (y_i - \bar{y})^2 / n$.

Bias of MLE variance estimates

Exercise: Show that

$$E[\hat{\mu}_{MLE}] = \mu$$

$$E[\hat{\sigma}_{MLE}^2] = \frac{(n-1)}{n} \sigma^2.$$

- The MLE of σ^2 is biased.
- The bias is small for large n , so maybe not a big deal.

Plug-in bias

Note that

$$\begin{aligned}E\left[\sum_i (y_i - \bar{y})^2 / n\right] &= \sigma^2(n-1)/n \\E\left[\sum_i (y_i - \mu)^2 / n\right] &= \sigma^2\end{aligned}$$

So the bias in the MLE can be thought of as “plug-in” bias:

- If we knew μ , then we would use the second equation to estimate σ^2 .
- We don't know μ , so we simply “plug-in” an estimate $\bar{y}$.
- But $\bar{y}$ is closer to $(y_1, \dots, y_n)$ than μ is.
- The MLE underestimates the variance.

Restricted likelihood

REML:

- Obtain an estimate of σ^2 where there is no plug-in bias.
- We need a “likelihood” that doesn’t depend on μ .
- Such a likelihood can be obtained from a transformed data set.

Centering to eliminate fixed effects

The model $y_1, \dots, y_n \sim \text{i.i.d. } N(\mu, \sigma^2)$ can be written

$$\begin{aligned}\mathbf{y} &= \mu \mathbf{1} + \mathbf{e} \\ \mathbf{e} &\sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)\end{aligned}$$

Let $\mathbf{C} = \mathbf{I} - \mathbf{1}\mathbf{1}^\top/n$. Then

$$\begin{aligned}\mathbf{C}\mathbf{1} &= \mathbf{0} \\ \mathbf{C}\mathbf{y} &= \mathbf{0} + \mathbf{C}\mathbf{e}.\end{aligned}$$

Letting $\check{\mathbf{y}} = \mathbf{C}\mathbf{y}$, we have

$$\check{\mathbf{y}} \sim N(\mathbf{0}, \sigma^2 \mathbf{C}).$$

Idea: Estimate σ^2 as the MLE of σ^2 based on $\check{\mathbf{y}}$.

Problem: $\mathbf{C}$ is a singular matrix ($\text{rank}(\mathbf{C}) = n - 1$).

Dimension-reduced data

The matrix $\mathbf{C}$ can be represented as $\mathbf{C} = \mathbf{N}\mathbf{N}^\top$ where

- $\mathbf{N} \in \mathbb{R}^{n \times (n-1)}$;
- $\mathbf{N}^\top \mathbf{N} = \mathbf{I}_{n-1}$;
- $\mathbf{N}^\top \mathbf{1} = \mathbf{0}$.

Dimension-reduced data

The columns of **N** are the eigenvectors of **C** having nonzero eigenvalues.

```
n<-5
C<-diag(n) - matrix(1/n,n,n)

eC<-eigen(C)
eC$val

## [1] 1.000000e+00 1.000000e+00 1.000000e+00 1.000000e+00 8.881784e-16

N<-eC$vec[,1:(n-1)]

t(N)%*%N

##           [,1]           [,2]           [,3]           [,4]
## [1,] 1.000000e+00 -1.220637e-16 -4.507497e-17 -2.922069e-17
## [2,] -1.220637e-16 1.000000e+00 1.665335e-16 -1.387779e-17
## [3,] -4.507497e-17 1.665335e-16 1.000000e+00 5.551115e-17
## [4,] -2.922069e-17 -1.387779e-17 5.551115e-17 1.000000e+00

t(N)%*%rep(1,n)

##           [,1]
## [1,] -1.665335e-16
## [2,] -5.551115e-17
## [3,] 0.000000e+00
## [4,] -6.106227e-16
```

REML

$\mathbf{N}$ can be used to transform an

- n -dimensional normal vector $\mathbf{y}$ with mean $\mu\mathbf{1}$ to an
- $(n - 1)$ -dimensional normal vector $\tilde{\mathbf{y}}$ with mean $\mathbf{0}$.

$$\begin{aligned}\mathbf{N}^\top \mathbf{y} &= \tilde{\mathbf{y}} = \mu \mathbf{N}^\top \mathbf{1} + \mathbf{N}^\top \mathbf{e} \\ \tilde{\mathbf{y}} &= \mathbf{0} \quad + \quad \tilde{\mathbf{e}}\end{aligned}$$

where $\tilde{\mathbf{e}} \sim N_{n-1}(\mathbf{0}, \sigma^2 \mathbf{N}^\top \mathbf{N}) = N_{n-1}(\mathbf{0}, \sigma^2 \mathbf{I}_{n-1})$.

The -2 log likelihood $-2 \log p(\tilde{\mathbf{y}}|\sigma^2)$ is

$$-2 \log p(\tilde{\mathbf{y}}|\sigma^2) = (n - 1)\sigma^2 + \sum_{i=1}^{n-1} \tilde{y}_i^2.$$

This is the “restricted likelihood.” The (R)MLE of σ^2 based on $\tilde{\mathbf{y}}$ is

$$\hat{\sigma}_{REML}^2 = \sum_i \tilde{y}_i^2 / (n - 1).$$

The final step

$$\begin{aligned}\hat{\sigma}_{REML}^2 &= \sum_i \tilde{y}_i^2 / (n - 1) \\ &= \tilde{\mathbf{y}}^\top \tilde{\mathbf{y}} / (n - 1) \\ &= \mathbf{y}^\top \mathbf{C}^\top \mathbf{C} \mathbf{y} / (n - 1) \\ &= \sum_{i=1}^n (y_i - \bar{y})^2 / (n - 1) = s^2.\end{aligned}$$

So in this case,

- the REML estimate of σ^2 based $\tilde{\mathbf{y}}$ is equal to
- the (unbiased) sample variance estimate s^2 .

REML for normal linear regression