REML

One sample normal model

$y_1 \ldots y_n \sim \text{IID } N(\mu, \sigma^2) \iff y_i = \mu + \epsilon_i$

$\epsilon_1 \ldots \epsilon_n \sim \text{IID } N(0, \sigma^2)$

Vector-Matrix Form

$\underset{\sim}{y} \sim N_n(\underset{\sim}{1}_n \mu, \sigma^2 I_n) \iff \underset{\sim}{y} = \underset{\sim}{1}\mu + \underset{\sim}{\epsilon}, \ \underset{\sim}{\epsilon} \sim N_n(0, \sigma^2 I)$

Centered data

Let $C = I_n - \underset{\sim}{1}\underset{\sim}{1}^T/n$, the $n \times n$ "centering matrix"

$$
\begin{aligned}
C\underset{\sim}{y} &= (I - \underset{\sim}{1}\underset{\sim}{1}^T/n)\underset{\sim}{y} \\
&= \underset{\sim}{y} - \underset{\sim}{1}\,\underset{\sim}{1}^T\underset{\sim}{y}/n \\
&= \underset{\sim}{y} - \underset{\sim}{1}\bar{y} \quad = \begin{pmatrix} y_1 - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{pmatrix}
\end{aligned}
$$

Notice: $C\underset{\sim}{1} = \underset{\sim}{1} - \underset{\sim}{1}(\bar{1})$

$= 0$

what is the distribution of $\check{y} = Cy$?

$$\check{y} = C\underset{\sim}{y} = C(\underset{\sim}{1}\mu + \epsilon) = C\underset{\sim}{e}$$

$$e \sim N(0, \sigma^2 I) \Rightarrow \check{y} \sim N(0, \sigma^2 C)$$

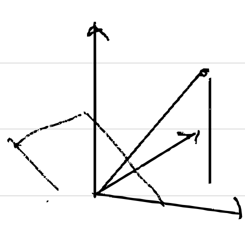
Idea: get MLE for σ^2 based on $\check{y}$, not y!

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

Dimension reduction

$$\underset{n \times n}{C} = \underset{n \times (n-1)}{N} \; \underset{(n-1) \times n}{N^T} \quad \text{where}$$

- $N \in \mathbb{R}^{n \times n-1}$
- $N^T N = I_{n-1}$
- $N^T 1 = 0$



N can be obtained in R:

N = evec(C)[, 1:(n-1)]

REML:

N can transform $\underset{\sim}{y}$ n-dim vec, mean $\mu \underset{\sim}{1}$

to

$N^T \underset{\sim}{y}$ (n-1)-dim vec, mean $\underset{\sim}{0}$:

$$\tilde{y} = N^T y = \mu N^T \underset{\sim}{1} + N^T \epsilon$$
$$= 0 + \tilde{\epsilon}$$

$$\tilde{\epsilon} = N^T \epsilon \sim N_{n-1}(0, N^T(\sigma^2 I) N)$$

$$= N_{n-1}(0, \sigma^2 N^T N) = N(0, \sigma^2 I)$$

Idea: find MLE of σ^2 using $\tilde{y}$

This is the REML estimate

$\tilde{y} \sim N_{n-1}(0, \sigma^2 I_{n-1})$

$p(\tilde{y}|\sigma^2) = (2\pi\sigma^2)^{-(n-1)/2} e^{-\frac{1}{2\sigma^2}\sum_{}^{n-1}\tilde{y}_i^2}$

$-2\log p(\tilde{y}|\sigma^2) = (n-1)\log\sigma^2 + \sum\tilde{y}_i^2/\sigma^2$

the MLE based on $\tilde{y}$ is

$$\tilde{\sigma}^2_{REML} = \arg\min -2\log p(\tilde{y}|\sigma^2)$$

$$= \sum\tilde{y}_i^2/(n-1)$$

<u>Bias</u>: $E[\hat{\sigma}^2_{REML}] = \frac{1}{n-1}\sum_1^{n-1} E[\tilde{y}_i^2]$

$$= \frac{1}{n-1}\sum_1^{n-1}\sigma^2 = \underline{\sigma^2}$$

$\Rightarrow$ The REML estimate is <u>unbiased</u>.

The magic trick

$$\hat{\sigma}^2_{REML} = \sum \tilde{y}_i^2 / n-1$$
$$= \tilde{y}^T \tilde{y} / n-1$$
$$= y^T N N^T y / n-1$$

$$= y^T C y / n-1$$

Lemma: C is symmetric and idempotent

$$C^T C = CC = (I - 11^T/n)(I - 11^T/n)$$
$$= I - 11^T/n - 11^T/n + 11^T/n$$
$$= I - 11^T/n = C$$

Therefore $y^T C y = y^T C^T C y$
$$= (Cy)^T (Cy)$$
$$= (y - \bar{y} 1)^T (y - \bar{y} 1)$$
$$= \sum_{i=1}^{n} (y_i - \bar{y})^2$$

$\Rightarrow \hat{\sigma}^2_{REML} = \dfrac{\sum (y_i - \bar{y})^2}{n-1}$, the (unbiased) sample variance.

REML for linear regression

$$y \sim N(X\beta, \sigma^2 I)$$

$$p(y \mid \beta, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}(y - X\beta)^T(y - X\beta)\right\}$$

$$-2\log p = n \log \sigma^2 + \frac{1}{\sigma^2}(y - X\beta)^T(y - X\beta)$$

For any σ^2, the minimizer in β is

$$\hat{\beta}_{MLE} = (X^T X)^{-1} X^T y$$
$$X\hat{\beta}_{MLE} = X(X^T X)^{-1} X^T y$$

To find MLE of σ^2, opt in σ^2:

$$\hat{\sigma}^2_{MLE} = \arg\min_{\sigma^2} n \log \sigma^2 + \frac{1}{\sigma^2}(y - X\hat{\beta})^T(y - X\hat{\beta})$$

$$= \frac{(y - X\hat{\beta})^T(y - X\hat{\beta})}{n}$$

$$= \frac{\sum (y_i - x_i^T\hat{\beta})^2}{n}$$

Can show $E[\hat{\sigma}^2_{MLE}] = \frac{n-p}{n} \sigma^2$

Thus bias could be large if p is large! Could be a serious problem.

REML solution

$y = X\beta + \epsilon \qquad X \in \mathbb{R}^{n \times p}$

If $n > p$, $\exists$ a matrix $N \in \mathbb{R}^{n \times (n-p)}$ s.t.

1. $N^T X = 0_{(n-p) \times p}$
2. $N^T N = I_{n-p}$

Idea: Let $\tilde{y} = N^T y$. Then

$$
\begin{aligned}
\tilde{y} = N^T y &= N^T X \beta + N^T \epsilon \\
&= 0 + N^T \epsilon \\
&= \tilde{\epsilon}
\end{aligned}
$$

where $\tilde{\epsilon} \sim N_{n-p}(0, N^T N \sigma^2 I) = \underline{N_{n-p}(0, \sigma^2 I_{n-p})}$

REML estimate of σ^2

$\hat{\sigma}^2_{REML}$ = $\hat{\sigma}^2_{MLE}$ based on $\tilde{y}$

$$= \sum_{i=1}^{n-p} \tilde{y}_i^2 / (n-p)$$

$$E[\hat{\sigma}^2_{REML}] = \frac{1}{n-p} \sum_i^{n-p} E[\tilde{y}_i^2] = \sigma^2$$

More Magic

$$\hat{\sigma}^2_{REML} = \frac{1}{n-p} \tilde{y}^T \tilde{y} = \frac{y^T N N^T y}{n-p}$$

Lemma: $NN^T = (I - X(X^TX)^{-1}X^T)$

This is symm & idempotent.

$$\frac{1}{n-p} Y^T NN^T y = \frac{1}{n-p} Y^T NN^T NN^T y)$$

But $NN^T y = Y - X(X^TX)^{-1}X^T\hat{\beta}$

$= Y - X\hat{\beta}$

so $\hat{\sigma}^2_{REML} = \dfrac{\tilde{Y}^T \tilde{Y}}{n-p} = \dfrac{(y - X\hat{\beta})^T (y - X\hat{\beta})}{n-p}$

$= \dfrac{\sum (y_i - x_i^T \hat{\beta})^2}{n-p}$, the usual unbiased est of σ^2.

REML for LME

$y_{ij} = x_{ij}^T \beta + z_{ij}^T a_j + \epsilon_{ij}$

$\underset{\sim}{y}_j = X_j \underset{\sim}{\beta} + Z_j \underset{\sim}{a}_j + \underset{\sim}{\epsilon}_j$

$$\underset{\sim}{y} = \begin{pmatrix} \underset{\sim}{y}_1 \\ \underset{\sim}{y}_2 \\ \vdots \\ \underset{\sim}{y}_m \end{pmatrix} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_m \end{pmatrix} \beta + \begin{pmatrix} Z_1 & 0 & & \\ 0 & Z_2 & & \\ & & \ddots & \\ & & & Z_m \end{pmatrix} \begin{pmatrix} \underset{\sim}{a}_1 \\ \vdots \\ \underset{\sim}{a}_m \end{pmatrix} + \underset{\sim}{\epsilon}$$

$= X\beta + Z\underset{\sim}{a} + \epsilon$

Variance Components:

$$\underset{\sim}{\epsilon} \sim N_{nm}(0, \sigma^2 I_{nm})$$

$$\underset{\sim}{a}_1, \ldots, \underset{\sim}{a}_m \sim \text{IID} \quad N_q(0, \Psi)$$

want to estimate σ^2, Ψ.

MLE: Directly optimize likelihood
Problem: $\hat{\Psi}, \hat{\sigma}^2$ can be very biased

Idea: Use REML to reduce bias

Let $N \in \mathbb{R}^{nm \times (nm-p)}$ be s.t.

① $N^T X = 0$

② $N^T N = I$

Let $\tilde{y} = N^T y$. Then

$$\tilde{y} = N^T y = N^T Z \underset{\sim}{a} + N^T \underset{\sim}{\epsilon}$$

$$\tilde{y} \sim N_{nm-p}\left(0,\ N^T Z \Psi Z^T N + N^T (I \sigma^2) N\right)$$

$$= N_{nm-p}\left(0,\ N^T Z \Psi Z^T N + \sigma^2 I_{nm-p}\right)$$

$$= N_{nm-p}\left(0,\ \tilde{Z} \Psi \tilde{Z}^T + \sigma^2 I_{nm-p}\right)$$

$$(\hat{\Psi}, \hat{\sigma}^2)_{REML} = \arg\min \ -2 \log p(\tilde{y} \mid \Psi, \sigma^2)$$

- May or may not be biased dep on structure of Ψ.
- Generally has lower bias than $(\hat{\Psi}, \hat{\sigma}^2)_{MLE}$

REML fixed effects estimates

LME model: $\quad y = X\underset{\sim}{\beta} + Z\underset{\sim}{a} + \underset{\sim}{\epsilon}$

$$\underset{\sim}{a} = \begin{pmatrix} \underset{\sim}{a}_1 \\ \vdots \\ \underset{\sim}{a}_m \end{pmatrix} \sim N\left(0, \begin{pmatrix} \Psi & 0 & 0 \\ 0 & \Psi & 0 \\ 0 & 0 & \Psi \ddots \end{pmatrix}\right) \qquad \underset{\sim}{\epsilon} \sim N_{mn}(0, \sigma^2 I)$$

$\tilde{\Psi}$

$$\Rightarrow \quad y \sim N(X\beta,\ Z\tilde{\Psi}Z' + \sigma^2 I) \equiv N(X\beta, \Sigma)$$

Suppose Σ were known:

$$\hat{\beta}_{MLE} = \arg\min -2\log p(y \mid \beta)$$

$$= \arg\min \quad (y - X\beta)\Sigma^{-1}(y - X\beta)$$

$$= \arg\min \quad -2\beta^T X^T \Sigma^{-1} y + \beta^T X^T \Sigma^{-1} X \beta$$

$$\frac{\partial}{\partial\beta} \text{ " } = -X^T \Sigma^{-1} y + X^T \Sigma^{-1} X \beta$$

$$\rightarrow \hat{\beta}_{MLE} = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} y$$

However, Σ not generally known

- MLEs of $(\Sigma, \beta) = (\hat{\Sigma}_{MLE}, (X^T \hat{\Sigma}_{MLE}^{-1} X)^{-1} X^T \hat{\Sigma}_{MLE}^{-1} y)$
- REMLEs of $(\Sigma, \beta) = (\hat{\Sigma}_{REML}, (X^T \hat{\Sigma}_{REML}^{-1} X)^{-1} X^T \hat{\Sigma}_{REML}^{-1} y)$

Both "plug in" their respective cov. estimates

Both estimates of β are unbiased.

Could be same or diff, dep on Σ.