

04/12/2018 Thursday Page 579

①

Restricted ML (REML)

Consider a longitudinal regression model where responses are uncorrelated & have constant variance matrix.

$$\underline{y}_i \sim N(X_i \underline{\beta}, \Sigma_i)$$

Assume $\Sigma_i = \sigma^2 I_{n_i}$.

$$\underline{y} \sim N(X \underline{\beta}, \Sigma), \quad \Sigma = \sigma^2 I$$

The MLE of $\underline{\beta}$ is the LS estimator.

$$\hat{\underline{\beta}} = (X'X)^{-1} X' \underline{y} = \left(\sum_{i=1}^m X_i' X_i \right)^{-1} \sum_{i=1}^m X_i' \underline{y}_i$$

& the MLE of σ^2 is.

$$\hat{\sigma}^2 = \frac{1}{n} (\underline{y} - X \hat{\underline{\beta}})' (\underline{y} - X \hat{\underline{\beta}})$$

$$= \frac{1}{n} \sum_{i,j} (y_{ij} - X_{ij}' \hat{\underline{\beta}})^2$$

\uparrow \uparrow
jth row of \underline{y}_i & X_i

MLE of $\underline{\beta}$ is unbiased.

But estimation of σ^2 is biased.

Scale $\hat{\sigma}^2$ by .

$$\tilde{\sigma}^2 = \frac{1}{n-p} (y - X\beta)'(y - X\beta)$$

this is unbiased for σ^2 .

REML is a variant of ML that attempts to provide less biased estimators of variances, or more generally, of covariance matrices.

transfer. $\underline{y} \sim N(X\beta, \Sigma)$

$$\underline{z} = \begin{pmatrix} \underline{z}_1 \\ \underline{z}_2 \end{pmatrix} = \begin{pmatrix} L_1' \\ L_2' \end{pmatrix} \underline{y}$$

where $\underline{z}_i = L_i' \underline{y}$

distribution of \underline{z}_1 depends on Σ only.

distribution of \underline{z}_2 given \underline{z}_1 depends on

both β & Σ .

Likelihood function for (β, Σ) based on \underline{z}

$$L(\beta, \omega) \propto f_{\underline{z}_1, \underline{z}_2}(\underline{z}_1, \underline{z}_2)$$

$$\Sigma = \Sigma(\omega) \propto f_{\underline{z}_1}(\underline{z}_1) \cdot f_{\underline{z}_2|\underline{z}_1}(\underline{z}_2|\underline{z}_1)$$

$\downarrow \int_{\Lambda^*}(\omega)$ $\downarrow \int_{\Lambda^*}(\beta, \omega)$

the REML estimator of $\Sigma(\underline{\omega})$, or $\underline{\omega}$ is obtained by maximizing $L_A^*(\underline{\omega})$, this is inserted into $L_\beta^*(\beta, \underline{\omega})$, which is then maximized with respect to β , gives the REML estimator of β .

$$L_1' X = 0$$

the REML estimator of β has the same form as the MLE.

$$\begin{aligned} \hat{\beta}(\hat{\underline{\omega}}) &= (X' \Sigma^{-1}(\hat{\underline{\omega}}) X)^{-1} X' \Sigma^{-1}(\hat{\underline{\omega}}) Y \\ &= \left(\sum_{j=1}^m X_j' \Sigma_j^{-1}(\hat{\underline{\omega}}) X_j \right)^{-1} \left(\sum_{j=1}^m X_j' \Sigma_j^{-1}(\hat{\underline{\omega}}) y_j \right) \end{aligned}$$

except that $\hat{\underline{\omega}}$ is REML estimator of $\underline{\omega}$.

Remarks.

① REML tends to produce less biased estimates of variances.

— REML maximization for $\underline{\omega}$ on $L_A^*(\underline{\omega})$ is done under a lower dimensional space.

② Distribution theory for ML & REML ④
 estimators are identical.

Formulas for Wald, F, t & χ^2 statistics
 for MLTs apply to REML estimators.

Modeling individual trajectories

renew data again.

Distance & age is roughly linear
 within individuals.

Subjects	T1	T2	T3	T4
boys				
girls				

ignore gender, specify

the "individual" (1st stage) model. $j=1, 2, \dots, n_i$

$$\begin{array}{c}
 y_{ij} \\
 \swarrow \quad \searrow \\
 \text{child } i \quad \text{time}
 \end{array}
 = \beta_{i0} + \beta_{i1} t_{ij} + \epsilon_{ij}$$

$\uparrow \qquad \qquad \uparrow$
 Subject specific intercepts
 & slopes.

in general, child i is to be observed
 at times $t_{i1}, t_{i2}, \dots, t_{in_i}$.

Note:

5

1. each child, has its own regression model, with parameters.

$$\tilde{\beta}_i = \begin{bmatrix} \beta_{i0} \\ \beta_{i1} \end{bmatrix}$$
$$\begin{pmatrix} y_{i1} \\ \vdots \\ y_{in_i} \end{pmatrix} = \begin{pmatrix} 1 & t_{i1} \\ \vdots & \vdots \\ 1 & t_{in_i} \end{pmatrix} \begin{pmatrix} \beta_{i0} \\ \beta_{i1} \end{pmatrix} + \begin{pmatrix} \epsilon_{i1} \\ \vdots \\ \epsilon_{in_i} \end{pmatrix}$$

$$\tilde{y}_i = \tilde{z}_i \tilde{\beta}_i + \tilde{\epsilon}_i$$

2. ^{need} similar type of trend in each individual.

3. ϵ_{ij} : within individual variation or fluctuation above subject specific mean $\beta_{i0} + \beta_{i1} t_{ij}$.

4. ϵ_i might be written as when there are random effects.

$$\tilde{\epsilon}_i = \tilde{\epsilon}_{i1} + \tilde{\epsilon}_{i2}$$

with elements.

measurement error.

$$\epsilon_{ij} = \epsilon_{ij1} + \epsilon_{ij2}$$

within individual fluctuation.
above $\beta_{i0} + \beta_{i1} t_{ij}$ (in absence of measurement error)

usually assume, $\tilde{\epsilon}_{i1}$ & $\tilde{\epsilon}_{i2}$ independent. (6)

$$E(\tilde{\epsilon}_i) = E(\tilde{\epsilon}_{i1}) + E(\tilde{\epsilon}_{i2}) = 0$$

$$\begin{aligned} \text{Var}(\tilde{\epsilon}_i) &= \text{Var}(\tilde{\epsilon}_{i1}) + \text{Var}(\tilde{\epsilon}_{i2}) \\ &= R_i \end{aligned}$$

5. popular forms for R_i

(a). fluctuation over time independent of each other, with constant variance b_1^2 measurement errors independent over time with constant variance b_2^2 .

$$R_i = b_1^2 I_{n_i} + b_2^2 I_{n_i} = (b_1^2 + b_2^2) I_{n_i}$$

no fluctuation $R_i = b_2^2 I_{n_i}$

no measurement error, $R_i = b_1^2 I_{n_i}$.

Can't distinguish among these possibilities.

~~(b)~~ Independent fluctuations over time might make sense if times are well-separated.

(b). more generally, might assume.

⑦

$$R_i = \text{Var}(\tilde{\varepsilon}_{1i}) + \text{Var}(\tilde{\varepsilon}_{2i}) \\ = \sigma_1^2 \Gamma_i + \sigma_2^2 I_{n_i}$$

↓

Can be

AR(1),

CS

Markov, etc.

(c). if there are no measurement error.

$$R_i = \sigma_1^2 \Gamma_i$$

more generally, we might assume R_i is unstructured or is different across groups.

Population (second stage) model

1st stage: subject specific trajectory Φ
within subject fluctuation.

2nd stage: among subject variation

$$\begin{bmatrix} \beta_{i0} \\ \beta_{i1} \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} b_{i0} \\ b_{i1} \end{bmatrix}$$

⊕

$$\beta_i = \beta + b_i$$

average of
intercepts & slopes
across subjects in
population.

subject specific deviations
of intercept & slope
from population
average line.

b_i random. ∴ how much the i th
subject regression coefficients differ from
population values.

$$E(\underline{b}_i) = \mathbf{0} \quad E(\underline{\beta}_i) = \beta$$

$$\text{Var}(\underline{b}_i) = \mathbf{g} \quad (\text{or } g_i)$$

$$\mathbf{g} = \begin{bmatrix} \overset{\text{Var}(b_{i0})}{\beta_0^2} & \beta_0 \beta_1 \\ \beta_1 \beta_0 & \beta_1^2 \end{bmatrix} \begin{matrix} \rightarrow \text{Cov}(b_{i0}, b_{i1}) \\ \rightarrow \text{Var}(b_{i1}) \end{matrix}$$

\mathbf{g} specifies how regression effects vary
across subjects in population.

$$Y_{ij} = \beta_{i0} + \beta_{i1} t_{ij} + \epsilon_{ij} \quad \text{Stage 1 } \textcircled{9}$$

$$= (\beta_0 + b_{i0}) + (\beta_1 + b_{i1}) t_{ij} + \epsilon_{ij} \quad \text{Stage 2}$$

$$= \beta_0 + \beta_1 t_{ij} + b_{i0} + b_{i1} t_{ij} + \epsilon_{ij}$$

by matrix notation.

$$Y_i = Z_i \beta + \epsilon_i$$

$$= Z_i (\beta + b_i) + \epsilon_i$$

$$= Z_i \beta + Z_i b_i + \epsilon_i$$