# ML for Estimating Theta

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{b} + \epsilon,$$

where

$$\begin{aligned} \mathbf{b} &\sim N\{\mathbf{0}, \mathbf{D}(\theta)\} \\ \epsilon &\sim N(\mathbf{0}, \mathbf{R}(\theta)) \\ \text{and } \mathbf{V} &= cov(\mathbf{Y}) = \mathbf{Z}\mathbf{D}\mathbf{Z}^{\mathbf{T}} + \mathbf{R} \end{aligned}$$

ML score equation for $\boldsymbol{\theta}$:

$$U_{\theta_j} = -\frac{1}{2} tr(\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_j}) + \frac{1}{2}(\mathbf{Y} - \mathbf{X}\hat{\beta})^{\mathbf{T}}\mathbf{V}^{-\mathbf{1}}\frac{\partial \mathbf{V}}{\partial \theta_{\mathbf{j}}}\mathbf{V}^{-\mathbf{1}}(\mathbf{Y} - \mathbf{X}\hat{\beta})$$

# Disadvantages to using ML for Theta

- The MLE of $\boldsymbol{\theta}$ fails to account for the loss of degrees of freedom from estimating $\boldsymbol{\beta}$.
- $\hat{\boldsymbol{\theta}}$ is biased for finite samples.

REML for linear regression:

$$Y_i = \mathbf{X}_i^T \beta + \epsilon_i, \quad where \;\; \epsilon_i \sim N(0, \sigma^2)$$

$$\Rightarrow \qquad \widehat{\sigma^2} = \frac{1}{n} \sum_i (Y_i - \mathbf{X}_i^T \hat{\beta})^2 = ML$$

$$\widehat{\sigma^2} = \frac{1}{n-p} \sum_i (Y_i - \mathbf{X}_i^T \hat{\beta})^2 = REML$$

# How to Obtain REML Estimates

Objective: Construct an estimator of $\boldsymbol{\theta}$ by accounting for the loss of degrees of freedom from estimating $\boldsymbol{\beta}$.

How? Inference on $\boldsymbol{\theta}$ proceeds by maximizing the log-likelihood of an error contrast whose distribution is free of $\boldsymbol{\beta}$.

Approach **Error contrasts**

Note: For variance of normal population and linear regression, we can just calculate the bias and transform $\rightarrow$ doesn't work for LMM.

# Error Contrasts

In matrix form:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{b} + \epsilon$$

and

$$\mathbf{Y} \sim N(\mathbf{X}\beta, \mathbf{V}(\boldsymbol{\theta}))$$

Define $\mathbf{U} = \mathbf{A}^T\mathbf{Y}$ where $\mathbf{A} = n \times (n-p)$ full rank matrix orthogonal to $\mathbf{X}$:

$$\mathbf{U} \sim N(\mathbf{0}, \mathbf{A}^T\mathbf{V}(\theta)\mathbf{A})$$

which doesn't depend on $\beta$. ($\mathbf{A}$ can be $n \times n$ if used generalized inverses)

# REML Log-Likelihood

Harville (1974) shows that:

$$\begin{aligned} L(\theta) &= (2\pi)^{-(n-p)/2} \left| \sum_{i=1}^{N} \mathbf{X}_i^T \mathbf{X}_i \right|^{1/2} \left| \sum_{i=1}^{N} \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i \right|^{-1/2} \\ &\times \prod_{i=1}^{N} |\mathbf{V}_i|^{-1/2} exp \left\{ -\frac{1}{2} \sum_{i=1}^{N} (\mathbf{Y}_i - \mathbf{X}_i \widehat{\beta})^T \mathbf{V}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \widehat{\beta}) \right\} \end{aligned}$$

and

$$\ell_R(\boldsymbol{\theta}) = -\frac{1}{2} ln|\mathbf{X}^\mathbf{T}\mathbf{V}^{-\mathbf{1}}\mathbf{X}| - \frac{1}{2} ln|\mathbf{V}| - \frac{1}{2}(\mathbf{Y} - \mathbf{X}\hat{\beta})^\mathbf{T}\mathbf{V}^{-\mathbf{1}}(\mathbf{Y} - \mathbf{X}\hat{\beta})$$

$$\widehat{\beta} = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{Y}$$

# REML Score and Information

- REML score equation for $\boldsymbol{\theta}$:

$$U_{R\theta_j} = -\frac{1}{2} tr(\mathbf{P} \frac{\partial \mathbf{V}}{\partial \theta_j}) + \frac{1}{2} (\mathbf{Y} - \mathbf{X}\hat{\beta})^{\mathbf{T}} \mathbf{V}^{-\mathbf{1}} \frac{\partial \mathbf{V}}{\partial \theta_{\mathbf{j}}} \mathbf{V}^{-\mathbf{1}} (\mathbf{Y} - \mathbf{X}\hat{\beta})$$

- REML Fisher Information:

$$I_{\theta\theta,jk} = \frac{1}{2} tr(\mathbf{P} \frac{\partial \mathbf{V}}{\partial \theta_j} \mathbf{P} \frac{\partial \mathbf{V}}{\partial \theta_k})$$

$$\mathbf{P} = \mathbf{V}^{-\mathbf{1}} - \mathbf{V}^{-\mathbf{1}} \mathbf{X} (\mathbf{X}^{\mathbf{T}} \mathbf{V}^{-\mathbf{1}} \mathbf{X})^{-\mathbf{1}} \mathbf{X}^{\mathbf{T}} \mathbf{V}^{-\mathbf{1}}$$

# ML Score and Information

- ML score equation for $\boldsymbol{\theta}$:

$$U_{\theta_j} = -\frac{1}{2} tr(\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_j}) + \frac{1}{2}(\mathbf{Y} - \mathbf{X}\hat{\beta})^{\mathbf{T}}\mathbf{V}^{-\mathbf{1}}\frac{\partial \mathbf{V}}{\partial \theta_{\mathbf{j}}}\mathbf{V}^{-\mathbf{1}}(\mathbf{Y} - \mathbf{X}\hat{\beta})$$

- ML Fisher Information:

$$I_{\theta\theta,jk} = \frac{1}{2} tr(\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_j}\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \theta_k})$$

if $n \to \infty$, $\mathbf{P} = \mathbf{V}^{-\mathbf{1}}$, no need to adjust for the degrees of freedom.

# Remarks

- $\ell_R(\boldsymbol{\theta}) \to \widehat{\boldsymbol{\theta}}_R$.
- $\widehat{\beta}_R = (\mathbf{X}^T\mathbf{V}(\widehat{\boldsymbol{\theta}}_R)^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}(\widehat{\boldsymbol{\theta}}_R)^{-1}\mathbf{Y}$
- REML Estimator for $\boldsymbol{\theta}$ has smaller bias and larger variance compared to MLE counterpart
- Error contrasts will work for variance of normal population and linear regression regression.

Estimation is important, but what about statistical inference?

## Inference and Hypothesis Testing

- Inference: process in which one draws conclusions (from data)
- **Statistical** Inference: putting a measure of confidence around our conclusions
- We are often interested in inference on Fixed Effects and Variance components:
  - What is the effect of treatment?
  - Do we need a particular variance component? E.g. is the curvature quadratic?
  - Do we need ANY random effects?

# Inference for Fixed Effects

Recall:

$$\widehat{\beta}(\widehat{\boldsymbol{\theta}}) = \left(\sum_{i=1}^{m} \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i\right)^{-1} \sum_{i=1}^{m} \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{Y}_i = (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}^{-1} \mathbf{Y},$$

$\widehat{\boldsymbol{\theta}}$ estimated by either REML or ML.

Objective:

$$H_0 : \mathbf{L}\beta = \mathbf{0}, \quad \text{vs} \quad H_A : \mathbf{L}\beta \neq \mathbf{0}$$

Some Tests:

- Wald
- Robust
- Likelihood Ratio

# Inference for Fixed Effects - Wald Test

$$
\begin{aligned}
var(\widehat{\beta}) &= var((\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{Y}) \\
&= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\left[var(\mathbf{Y})\right]\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \\
&= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}(\text{matrix form}) \\
&= \left(\sum_{i=1}^{m}\mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i\right)^{-1} (\text{with summations})
\end{aligned}
$$

which is from information ($\mathbf{I}_{\beta\beta}$).

$$
(\widehat{\beta}-\beta)'\mathbf{L}'\left[\mathbf{L}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{L}'\right]^{-1}\mathbf{L}(\widehat{\beta}-\beta) \xrightarrow{\mathcal{D}} \chi^2_{rank(\mathbf{L})}.
$$

CIs obtained also from normal assumption.

Note: Wald test statistics based on SE which are downward biased (variability of $\widehat{\theta}$)

# Inference for Fixed Effects - F and robust tests

$$F = \frac{(\widehat{\beta} - \beta)'\mathbf{L}' \left[\mathbf{L}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{L}'\right]^{-1} \mathbf{L}(\widehat{\beta} - \beta)}{rank(\mathbf{L})}$$

Numerator df $= rank(\mathbf{L})$, but denominator df is estimated from the data (and depends on software used!)

## Robust Inference

Use of $(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}$ assumes that marginal covariance is correctly specified. Alternative: set $var(\mathbf{Y}) = \mathbf{rr}'$ where $\mathbf{r} = (\mathbf{Y} - \mathbf{X}\beta)$

- Robust to misspecification of marginal covariance
- Ideal is general: as long as you get mean model, you are good (could even use OLS)
- Missing data is a problem
- Closely related to GEEs

# Inference for Fixed Effects - Likelihood Ratio Test

$$H_0 : \beta \in \mathcal{B}_{\beta,\mathbf{0}} \quad \text{vs} \quad H_A : \beta \notin \mathcal{B}_{\beta,\mathbf{0}}$$

$\mathcal{B}_{\beta,\mathbf{0}}$ is a subspace of the parameter space ($\mathcal{B}_\beta$) Usual LR statistic:

$$-2\log\lambda_N = -2log\left[\frac{L_{ML}(\widehat{\beta}_{ML,0})}{L_{ML}(\widehat{\beta}_{ML})}\right] \xrightarrow{\mathcal{D}} \chi^2_{df}$$

where *df* is difference in dimensions of $\mathcal{B}_{\beta,0}$ and $\mathcal{B}_\beta$

Note: this is **only valid when using ML**!

- Cannot use REML to fit models: mean structure of the model is different under $H_0$ and $H_A$ such that they would result in different error contrasts.

# Inference for Variance Components - Wald

ML Fisher Information:

$$I_{\theta\theta,jk} = \frac{1}{2} tr(\mathbf{V}^{-\mathbf{1}} \frac{\partial \mathbf{V}}{\partial \theta_j} \mathbf{V}^{-\mathbf{1}} \frac{\partial \mathbf{V}}{\partial \theta_k})$$

REML Fisher Information:

$$I_{\theta\theta,jk} = \frac{1}{2} tr(\mathbf{P} \frac{\partial \mathbf{V}}{\partial \theta_j} \mathbf{P} \frac{\partial \mathbf{V}}{\partial \theta_k})$$

Can we construct CI and Test via Normal approximations?
**Sometimes**

# Inference for Variance Components - Wald

In general

- If $\boldsymbol{\theta}$ is far from the boundary of parameter space $\boldsymbol{\Theta}_\theta$, then normal approximation OK!
- If $\boldsymbol{\theta}$ is on the boundary of parameter space $\boldsymbol{\Theta}_\theta$, then **normal approximation fails**!

Distinction between mixed model and marginal model:

- Under hierarchical model $\mathbf{D}(\boldsymbol{\theta})$ must be PSD $\rightarrow d_{kk} \geq 0$ (0 on boundary)
- Under marginal model, then we assume $\mathbf{V}_i = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i' + \sigma^2 I_{n_i}$, then $d_{kk} = 0$ is on the interior.

In general, we say that the usual Wald tests fail (for boundary case) in this case since we are interested in having $\mathbf{D}$ have real interpretations. Confidence intervals are OK when truth is far from bounds.

# Inference for Variance Components - LRT

LR Statistic

$$-2\log\lambda_N = -2log\left[\frac{L_{ML/REML}(\widehat{\boldsymbol{\theta}}_{ML/REML,0})}{L_{ML/REML}(\widehat{\boldsymbol{\theta}}_{ML/REML})}\right] \xrightarrow{\mathcal{D}} \chi^2_{df}$$

Key Remarks:

- Key regularity condition for this convergence is that we are NOT on the boundary $\rightarrow$ have same problem as Wald test
- We can use either ML or REML in this case

Frequently: testing $\theta = 0$ via LRT converges to mixture of $\chi^2$ distributions. (Homework!)