

# Untitled

PHS 7010 Fall 2023 - Final Report

Kline Dubose, Haojia Li

April 25, 2024

## 1 Likelihood ratio tests in linear mixed models with one variance component

Crainiceanu and Ruppert (2004) derived the finite sample distribution of the LRT and RLRT statistics for testing the null hypothesis that the variance component is 0 in a linear mixed model (LMM) with one variance component. They also derived the asymptotic distribution of the LRT and RLRT statistics under the null with weak assumptions on eigenvalues of certain design matrices. The assumptions are different from the restrictive assumptions by Self and Liang (1987), that the response variable vector can be partitioned into independent and identically distributed (i.i.d.) subvectors and the number of independent subvectors tends to  $\infty$ . The downside of using the restrictive assumptions is that the result would not hold for a fixed number of subjects, even if the number of observations per subject increased to  $\infty$ .

Consider an LMM with one variance component

$$\mathbf{Y} = X\beta + Z\mathbf{b} + \epsilon, \quad E \begin{pmatrix} \mathbf{b} \\ \epsilon \end{pmatrix} = \begin{pmatrix} \mathbf{0}_K \\ \mathbf{0}_n \end{pmatrix}, \quad \text{cov} \begin{pmatrix} \mathbf{b} \\ \epsilon \end{pmatrix} = \begin{pmatrix} \sigma_b^2 \Sigma & \mathbf{0} \\ \mathbf{0} & \sigma_\epsilon^2 I_n \end{pmatrix}$$

where,

- $\mathbf{Y}$  is the  $n \times 1$  vector of observations,
- $X$  is the  $n \times p$  design matrix for the fixed effects,
- $Z$  is the  $n \times K$  design matrix for the random effects,
- $\beta$  is a  $p$ -dimensional vector of fixed effects parameters,
- $\mathbf{b}$  is a  $K$ -dimensional vector of random effects,
- $(\mathbf{b}, \epsilon)$  has a normal distribution.

Under these conditions it follows that

$$E(\mathbf{Y}) = X\beta, \\ \text{var}(\mathbf{Y}) = \sigma_\epsilon^2 V_\lambda$$

where

- $\lambda = \sigma_b^2 / \sigma_\epsilon^2$ , which can be considered a signal-to-noise ratio,
- $V_\lambda = I_n + \lambda Z \Sigma Z'$ .

## 2 Proofs of Lemma 3, Theorem 1, and Corollary 1

Before we begin, in this scenario, our null hypothesis is  $\lambda = 0$  where  $\lambda$  is the effect size of familial relatedness (FR) and the alternative is  $\lambda > 0$ .

### 2.1 Lemma 3

Suppose that  $\sup_{s \in \Omega_0} \{\rho_s\} \leq B$  are bounded. Under the null hypothesis, (i) both  $F_n(\lambda)$  and  $G_n(\lambda)$  uniformly converge in probability to  $F_0(\lambda) = \sum_{j=0}^m q_j \log(w_j) - \log(\sum_{j=0}^m q_j w_j)$  over  $\lambda \in [-\delta, T]$  for any  $0 < \delta < \frac{1}{B \vee \max_j \{\phi_j\}}$  and  $0 < T < +\infty$ ; (ii)  $F_0(\lambda)$  achieves its unique maximum at  $\lambda = 0$ ; (iii)  $\hat{\lambda}_n \rightarrow^p 0$  and  $\hat{\lambda}_n^r \rightarrow^p 0$ .

i) Essentially we will be showing that:

$$\begin{aligned} F_n(\lambda) &\rightarrow^p F_0(\lambda) \\ G_n(\lambda) &\rightarrow^p F_0(\lambda) \end{aligned}$$

where:

$$F_0(\lambda) = \sum_{j=0}^m \log(w_j) - \log \left( \sum_{j=0}^m q_j w_j \right)$$

Additionally, note that  $\phi_j$  is the set of all distinctive non-zero eigenvalues of the FR correlation matrix.

Let's start with rewriting  $F_n(\lambda)$ :

$$\begin{aligned} F_n(\lambda) &= \log \left( \frac{\sum_{i=1}^{n-p} u_i^2}{\sum_{j=0}^m f'_j w_j U_j + \sum_{s \in \Omega_0} v_s u_s^2} \right) + \frac{1}{n} \sum_{j=1}^m f_j \log(w_j) \\ &= \log \left( \frac{n-p}{n-p} \cdot \frac{\sum_{i=1}^{n-p} u_i^2}{\sum_{j=0}^m f'_j w_j U_j + \sum_{s \in \Omega_0} v_s u_s^2} \right) + \frac{1}{n} \sum_{j=1}^m f_j \log(w_j) \\ &= \log \left( \frac{1}{n-p} \sum_{i=1}^{n-p} u_i^2 \right) - \log \left( \frac{\sum_{j=0}^m f'_j w_j U_j}{n-p} + \frac{\sum_{s \in \Omega_0} v_s u_s^2}{n-p} \right) + \frac{1}{n} \sum_{j=1}^m f_j \log(w_j) \end{aligned}$$

Looking at the first term, let's note that  $u_i \sim N(0, 1)$  for  $i = 1, \dots, n-p$ . This allows us to rewrite the first term and note that:

$$\begin{aligned} \frac{1}{n-p} \sum_{i=1}^{n-p} u_i^2 &= \frac{1}{n-p} \sum_{i=1}^{n-p} (u_i - 0)^2 = S_u \rightarrow^p \sigma_u = 1 \\ \frac{1}{n-p} \sum_{i=1}^{n-p} u_i^2 &\rightarrow^p 1 \end{aligned}$$

by the weak law of large numbers.

$$\Rightarrow \log \left( \frac{1}{n-p} \sum_{i=1}^{n-p} u_i^2 \right) \rightarrow^p \log(1) = 0$$

Looking at the second term, we note that  $U_j = \frac{1}{f'_j} \sum_{i \in \Omega_j} u_i^2$  for  $j = 0, \dots, m$ . Additionally, we note here that  $f'_j = |\Omega_j|$  which is the count of the times that  $phi_j$  is replicated in non-zero eigenvalues. For the purposes of this, it is sufficient to say that  $|\Omega_j|$  is a count.

We can rewrite  $U_j$  as:

$$U_j = \frac{1}{|\Omega_j|} \sum_{i \in \Omega_j} (u_i - 0)^2 \rightarrow^p 1$$

since  $u_i \sim N(0, 1)$  as previously stated.

Additionally, this implies that as  $U_j \rightarrow^p 1$ :

$$\Rightarrow \sum_{s \in \Omega_0} \frac{v_s u_s^2}{n-p} \rightarrow^p 0$$

under the null hypothesis.

This is since  $v_s = 1$  under the null hypothesis:

$$v_s = \frac{1}{1 + \lambda \rho_s} = \frac{1}{1 + 0} = 1$$

Additionally, as long as  $\sum_{s \in \Omega_0} u_s^2$  is finite, then  $\sum_{s \in \Omega_0} \frac{v_s u_s^2}{n-p} = \sum_{s \in \Omega_0} \frac{u_s^2}{n-p} \rightarrow 0$  as  $(n-p) \rightarrow \infty$ .

Noted in the paper:

$$\begin{aligned} \frac{f'_j}{n-p} &\rightarrow q_j \\ \frac{f_j}{n} &\rightarrow q_j \end{aligned}$$

This suggests that the second term of the equation:

$$-\log \left( \frac{\sum_{j=0}^m f'_j w_j U_j}{n-p} + \frac{\sum_{s \in \Omega_0} v_s u_s^2}{n-p} \right) \rightarrow^p -\log \left( \sum_{j=0}^m q_j w_j + 0 \right) = -\log \left( \sum_{j=0}^m q_j w_j \right)$$

From here, we can see that  $\sup_{\lambda \in [-\delta, T]} \left| -\log \left( \frac{\sum_{j=0}^m f'_j w_j U_j}{n-p} + \frac{\sum_{s \in \Omega_0} v_s u_s^2}{n-p} \right) \right| \rightarrow^p 0$ .

The third element of the equation converges:

$$\sum_{j=1}^m \frac{f_j}{n} \log(w_j) \rightarrow^p \sum_{j=1}^m q_j \log(w_j)$$

Similar to the second element, the supremum of the third element should converge to 0. Using the continuous mapping theorem,  $\sup_{\lambda \in [-\delta, T]} |F_n(\lambda) - F_0(\lambda)| \rightarrow^p 0$ . Therefore,  $F_n(\lambda)$  converges uniformly to  $F_0(\lambda)$  over the interval  $\lambda \in [-\delta, T]$ .

Let's now look  $G_n(\lambda)$ . Let's start by rewriting  $G_n(\lambda)$ :

$$\begin{aligned} G_n(\lambda) &= \log \left( \frac{\sum_{i=1}^{n-p} u_i^2}{\sum_{j=0}^m f'_j w_j U_j + \sum_{s \in \Omega_0} v_s u_s^2} \right) + \frac{1}{n-p} \sum_{j=1}^m f'_j \log(w_j) + \frac{1}{n-p} \sum_{s \in \Omega_0} \log(v_s) \\ &= \log \left( \frac{n-p}{n-p} \cdot \frac{\sum_{i=1}^{n-p} u_i^2}{\sum_{j=0}^m f'_j w_j U_j + \sum_{s \in \Omega_0} v_s u_s^2} \right) + \frac{1}{n-p} \sum_{j=1}^m f'_j \log(w_j) + \frac{1}{n-p} \sum_{s \in \Omega_0} \log(v_s) \\ &= \log \left( \frac{1}{n-p} \sum_{j=1}^{n-p} u_i^2 \right) - \log \left( \frac{1}{n-p} \left( \sum_{j=0}^m f'_j w_j U_j + \sum_{s \in \Omega_0} v_s u_s^2 \right) \right) + \sum_{j=1}^m \frac{f'_j}{n-p} \log(w_j) + \frac{\sum_{s \in \Omega_0} \log(v_s)}{n-p} \end{aligned}$$

In the previous proof of  $F_n(\lambda)$  we showed the convergence of the first and second elements of  $G_n(\lambda)$ .

For the third element, recall that  $\frac{f'_j}{n-p} \rightarrow q_j$  which implies that  $\sum_{j=1}^m \frac{f'_j}{n-p} \log(w_j) \rightarrow^p \sum_{j=1}^m q_j \log(w_j)$  which has been shown that the supremum of this third element converges in probability to 0.

The final element of this equation converges to 0, since under the null hypothesis  $v_s \rightarrow 1$  as previously shown.

Therefore by the uniform convergence theorem,  $\sup_{\lambda \in [-\delta, T]} |G_n(\lambda) - F_0(\lambda)| \rightarrow^p 0$ .  $G_n(\lambda)$  therefore converges uniformly to  $F_0(\lambda)$  in probability over  $\lambda \in [-\delta, T]$ .