

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 ∞ . 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 ∞ .

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,
- β is a p -dimensional vector of fixed effects parameters,
- \mathbf{b} is a K -dimensional vector of random effects,
- (\mathbf{b}, ϵ) 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 λ 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 ϕ_j is the set of all distinct non-zero eigenvalues of the FR correlation matrix.

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

$$\begin{aligned} F_n(\lambda) &= \log \left(\frac{\sum_{i=1}^{n-p} u_i^2}{\sum_{j=0}^m f'_j w_j U_j + \sum_{s \in \Omega_0} v_s u_s^2} \right) + \frac{1}{n} \sum_{j=1}^m f_j \log(w_j) \\ &= \log \left(\frac{n-p}{n-p} \cdot \frac{\sum_{i=1}^{n-p} u_i^2}{\sum_{j=0}^m f'_j w_j U_j + \sum_{s \in \omega_0} v_s u_s^2} \right) + \frac{1}{n} \sum_{j=1}^m f_j \log(w_j) \\ &= \log \left(\frac{1}{n-p} \sum_{i=1}^{n-p} u_i^2 \right) - \log \left(\frac{\sum_{j=0}^m f'_j w_j U_j + \sum_{s \in \Omega_0} v_s u_s^2}{n-p} \right) + \frac{1}{n} \sum_{j=1}^m f_j \log(w_j) \end{aligned}$$

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:

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

The third element of the equation converges:

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

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 $\lambda \in [-\delta, T]$.

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

The proof that the authors used involved Jensen's inequality to show that $F_0(\lambda)$ achieves its unique maximum at $\lambda = 0$. I was confused by this and will include my thought process in coming to the conclusion that I did.

When evaluating under the null hypothesis at $\lambda = 0$, we really only need to worry about the second term $-\log(\sum_{j=0}^m q_j w_j)$. Additionally, when considering with the constrain of the bounded parameter space $\lambda \geq 0$, we can see that $F_0(\lambda) \leq 0$ when $\lambda \geq 0$. We also note that q_j is a proportion such that $\sum_{j=0}^m q_j \leq 1$.

Under these constraints.

$$\begin{aligned}
F_0(0) &= -\log \left(\sum_{j=0}^m q_j \right) \mid \sum_{j=0}^m q_j \leq 1 \\
&\Rightarrow F_0(0) \geq 0 \text{ when } \lambda = 0
\end{aligned}$$

Since λ can only be non-negative with the way it has been defined in this paper:

$$\begin{aligned}
\left| \sum_{j=0}^m q_j \log \left(\frac{1}{1 + \lambda \phi_j} \right) \right| &> \left| \log \left(\sum_{j=0}^m q_j \frac{1}{1 + \lambda \phi_j} \right) \right| \\
&\Rightarrow F_0(\lambda) \leq 0
\end{aligned}$$

Which suggests that $\lambda = 0$ is the unique maximum of $F_0(\lambda)$.

iii) Since $\lambda = 0$ has been established as the unique global maximum, we can reference theorem 5.7 of *Asymptotic Statistics* (1998) by A. van der Vaart, which states that if there is a unique global

estimator with the properties we have already established, then $\hat{\theta} \rightarrow^p \theta_0$. In this case, both the MLE $\hat{\lambda}_n$ and the REML $\hat{\lambda}_n^r$ have the unique global maximum at $\lambda = 0$ and would both converge in probability to 0 according to this theorem.