DOI: 10.1111/j.1541-0420.2009.01373.x
Handle: http://hdl.handle.net/1942/467

The Use of Score Tests for Inference on Variance Components

Geert Molenberghs[1] and Geert Verbeke[2]

[1] Limburgs Universitair Centrum, Center for Statistics, Universitaire Campus, Building D, B-3590 Diepenbeek, Belgium; geert.molenberghs@luc.ac.be
[2] Katholieke Universiteit Leuven, Biostatistical Center, UZ Sint-Rafaël, Kapucijnenvoer 35, B-3000 Leuven, Belgium; geert.verbeke@med.kuleuven.ac.be

Abstract: Whenever inference for variance components is required, the choice between one-sided and two-sided tests is crucial. This choice is usually driven by whether or not negative variance components are permitted. For two-sided tests, classical inferential procedures can be followed, based on likelihood ratios, score statistics, or Wald statistics. For one-sided tests, however, one-sided test statistics need to be developed, and their null distribution derived. While this has received considerable attention in the context of the likelihood ratio test, there appears to be much confusion about the related problem for the score test.

Keywords: Boundary condition; Likelihood ratio test; Linear mixed model; One-sided test; Variance component.

1 Introduction

The linear mixed-effects model (Laird and Ware 1982, Verbeke and Molenberghs 2000) is a commonly used tool for variance component models and for longitudinal data. Let $\boldsymbol{Y_i}$ denote the n_i-dimensional vector of measurements available for subject $i = 1, \ldots, N$. A general linear mixed model then assumes that $\boldsymbol{Y_i}$ satisfies

$$\boldsymbol{Y_i} = X_i\boldsymbol{\beta} + Z_i\boldsymbol{b_i} + \boldsymbol{\varepsilon_i}, \tag{1}$$

in which $\boldsymbol{\beta}$ is a vector of population-averaged regression coefficients called fixed effects, and where $\boldsymbol{b_i}$ is a vector of subject-specific regression coefficients. The $\boldsymbol{b_i}$ describe how the evolution of the ith subject deviates from the average evolution in the population. The matrices X_i and Z_i are $(n_i \times p)$ and $(n_i \times q)$ matrices of known covariates. The random effects $\boldsymbol{b_i}$ and residual components $\boldsymbol{\varepsilon_i}$ are assumed to be independent with distributions $N(\mathbf{0}, D)$, and $N(\mathbf{0}, \Sigma_i)$, respectively. Inference for linear mixed models is usually based on maximum likelihood or REML *under the marginal model.* Thus, we can adopt two *different* views on the linear mixed model. The fully *hierarchical* model is specified by

$$\begin{array}{rcl} \boldsymbol{Y_i}|\boldsymbol{b}_i & \sim & N_{n_i}(X_i\boldsymbol{\beta} + Z_i\boldsymbol{b}_i, \Sigma_i), \\ \boldsymbol{b}_i & \sim & N(0, D), \end{array} \tag{2}$$

while the marginal model is given by

$$\boldsymbol{Y_i} \sim N_{n_i}(X_i\boldsymbol{\beta}, V_i = Z_i D Z_i' + \Sigma_i). \tag{3}$$

Even though they are often treated as equivalent, there are important differences between both views. Obviously, (2) requires the covariance matrices Σ_i and D to be positive definite, while in (3) it is sufficient for the resulting matrix V_i to be positive definite. Different hierarchical models can produce the same marginal model and some marginal models are not implied by any hierarchical model.

The simplest example to illustrate differences between the marginal and hierarchical views is found by restricting the random effects in (1) to a random intercept, producing the marginal model:

$$\boldsymbol{Y}_i \sim N(X_i\boldsymbol{\beta}, \tau^2 J_{n_i} + \sigma^2 I_{n_i}) \tag{4}$$

where J_{n_i} equals the $n_i \times n_i$ matrix containing only ones. In the marginal view, negative values for τ^2 are perfectly acceptable (Nelder 1954, Verbeke and Molenberghs 2000, Sec. 5.6.2), since this merely corresponds to negative within-cluster correlation $\rho = \tau^2/(\tau^2 + \sigma^2)$. In the hierarchical view, it is clearly imperative to restrict τ^2 to nonnegative values.

2 Inference for Variance Components

While each of the two views are possible, there are important differences regarding statistical inference for variance components. The first, *unconstrained case*, is classical regarding inference for the variance component τ^2 since the usual two-sided alternative $H_0 : \tau^2 = 0$ *versus* $H_{A2} : \tau^2 \neq 0$ is then used. Wald, likelihood ratio, and score tests are then asymptotically equivalent, and the asymptotic null distribution is well known to be χ^2_1. In the *constrained case*, one typically needs one-sided tests of the null-hypothesis

$$H_0 : \tau^2 = 0 \qquad \textit{versus} \qquad H_{A1} : \tau^2 > 0. \tag{5}$$

As the null-hypothesis is now on the boundary of the parameter space, classical inference no longer holds, appropriate tailored test statistics need to be developed, and the corresponding (asymptotic) null distributions derived. We will briefly review the likelihood-ratio case and then turn to score tests in the next section.

Suppressing dependence on the other parameters, let $\ell(\tau^2)$ denote the log-likelihood, as a function of the random-intercepts variance τ^2. Further, let

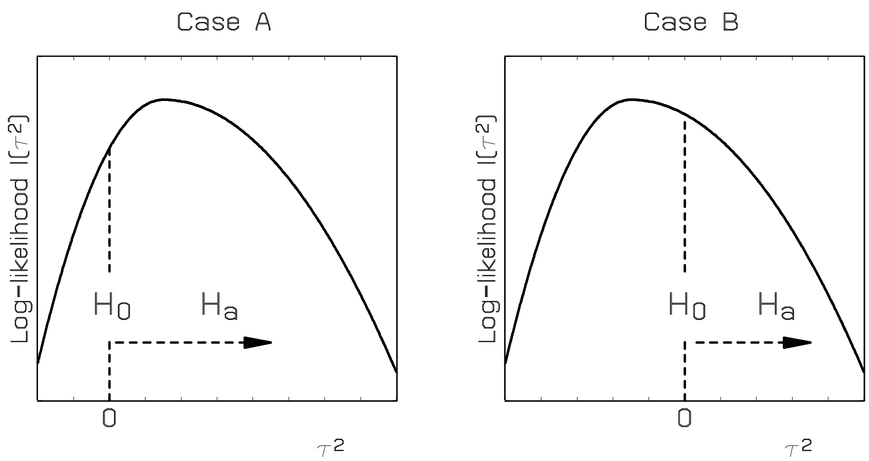

FIGURE 1. Graphical representation of two different situations, when developing one-sided tests for the variance τ^2 of the random intercepts b_i in model.

$\widehat{\tau}^2$ denote the maximum likelihood estimate of τ^2 under the unconstrained parameterization. We first consider the likelihood ratio test, with statistic:

$$T_{LR} \quad = \quad 2\ln\left[\frac{\max_{H_{1A}} \ell(\tau^2)}{\max_{H_0} \ell(\tau^2)}\right].$$

Two cases, graphically represented in Figure 1, can now be distinguished. Under Case A, $\widehat{\tau}^2$ is positive, and the likelihood ratio test statistic is identical to the one that would be obtained under the unconstrained parameter space for τ^2. Hence, conditionally on $\widehat{\tau}^2 \geq 0$, T_{LR} has asymptotic null distribution equal to the classical χ^2_1. Under Case B, $\ell(\tau^2)$ is maximized at $\tau^2 = 0$ under H_{1A} as well as under H_0, yielding $T_{LR} = 0$. Both cases are equally probable to occur, under the null. Hence, the asymptotic null distribution of T_{LR} is easily seen to follow a $0.5P(\chi^2_1 > c) + 0.5P(\chi^2_0 > c)$ null distribution. This was one of Stram and Lee's (1994) special cases. Note that, whenever $\widehat{\tau}^2 \geq 0$, the observed likelihood ratio test statistic is equal to the one under the unconstrained model, but the p-value is half the size of the one obtained from the classical χ^2_1 approximation to the null distribution.

In general, inference under the unconstrained model for the variance components in D can be based on the classical chi-squared approximation to the null distribution for the likelihood ratio test statistic. Under the constrained model, Stram and Lee (1994) have shown that the asymptotic null distribution for the likelihood ratio test statistic for testing a null hypothesis which allows for k correlated random effects versus an alternative of $k + 1$ correlated random effects (with positive semi-definite covariance matrix D_{k+1}), is a mixture of a χ^2_k and a χ^2_{k+1}, with equal probability 1/2. For more general settings, e.g., comparing models with k and $k+k'$ ($k'>1$) random effects, the null distribution is a mixture of χ^2 random variables (Shapiro 1988), the weights of which can only be calculated analytically in a number of special cases. Shapiro's (1988) results provide a few important special cases, not studied by Stram and Lee (1994). For example, if the null hypothesis allows for k *un*correlated random effects (with a diagonal covariance matrix D_k) versus the alternative of $k+k'$ uncorrelated random effects (with diagonal covariance matrix $D_{k+k'}$), the null distribution is a mixture of the form

$$\sum_{m=0}^{k'} 2^{-k'} \binom{k'}{m} \chi^2_m.$$

Shapiro (1988) shows that, for a broad number of cases, determining the mixture's weights is a complex and perhaps numerical task.

3 The Score Test

Verbeke and Molenberghs (2003), using results by Silvapulle and Silvapulle (1995), have shown that similar results are obtained when a score test is used instead of a likelihood ratio test. The use of score tests for testing variance components under a constrained parameterization requires replacing the classical score test statistic by an appropriate one-sided version. This is where the general theory of Silvapulle and Silvapulle (1995) on one-sided score tests proves very useful. They consider models parameterized through a vector $\boldsymbol{\theta} = (\boldsymbol{\lambda}', \boldsymbol{\psi}')'$, where testing a general hypothesis of the form $H_0 : \boldsymbol{\psi} = \mathbf{0}$ *versus* $H_A : \boldsymbol{\psi} \in \mathcal{C}$ is of interest. Silvapulle and Silvapulle (1995) allow $\mathcal{C}$ to be a closed and convex cone in Euclidean space, with vertex at the origin. The advantage of such a general definition is that one-sided, two-sided, and combinations of one-sided and two-sided hypotheses are included.

Adopt the following notation. Let $\boldsymbol{S}_N(\boldsymbol{\theta})$ and $H\boldsymbol{\theta})$ be the score vector and Hessian matrix of the log-likelihood function. Further, decompose $\boldsymbol{S}_N$ as $\boldsymbol{S}_N = (\boldsymbol{S}'_{N\lambda}, \boldsymbol{S}'_{N\psi})'$, let $H_{\lambda\lambda}(\boldsymbol{\theta})$, $H_{\lambda\psi}(\boldsymbol{\theta})$ and $H_{\psi\psi}(\boldsymbol{\theta})$ be the corresponding blocks in $H(\boldsymbol{\theta})$, and define $\boldsymbol{\theta}_H = (\boldsymbol{\lambda}', \mathbf{0}')'$. $\boldsymbol{\theta}_H$ can be estimated by $\widehat{\boldsymbol{\theta}}_H = (\widehat{\boldsymbol{\lambda}}', \mathbf{0}')'$, in which $\widehat{\boldsymbol{\lambda}}$ is the maximum likelihood estimate of $\boldsymbol{\lambda}$, under H_0. Finally, let $\boldsymbol{Z}_N$ be equal to $\boldsymbol{Z}_N = N^{-1/2}\boldsymbol{S}_{N\psi}(\widehat{\boldsymbol{\theta}}_H)$. A one-sided score statistic can now be defined as

$$T_S := \boldsymbol{Z}'_N H^{-1}_{\psi\psi}(\widehat{\boldsymbol{\theta}}_H)\boldsymbol{Z}_N \;-\; \inf\left\{(\boldsymbol{Z}_N - \boldsymbol{b})' H^{-1}_{\psi\psi}(\widehat{\boldsymbol{\theta}}_H)(\boldsymbol{Z}_N - \boldsymbol{b}) | \boldsymbol{b} \in \mathcal{C}\right\}. \quad (6)$$

Note that the score statistic, heuristically defined in the case of the random-intercepts model is a special case of (6). Indeed, when $\widehat{\tau}^2$ is positive, the score at zero is positive, and therefore in $\mathcal{C}$, such that the infimum in (6)

becomes zero. For $\widehat{\tau}^2$ negative, the score at zero is negative as well and the infimum in (6) is attained for $\boldsymbol{b} = \mathbf{0}$, resulting in $T_S = 0$.
It follows from Silvapulle and Silvapulle (1995) that, under suitable regularity conditions, for $N \to \infty$, the likelihood ratio and score test statistics satisfy $T_{LR} = T_S + o_p(1)$. This indicates that the equivalence of the score and likelihood ratio tests not only holds in the two-sided but also in the one-sided cases. Moreover, what is known about the null distribution in the case of the likelihood ratio test, immediately carries over to the score test case. This result corrects the common belief that, even when variance components are on the boundary of the parameter space, the score test deserved no special treatment. Verbeke and Molenberghs (2003) provide an empirical illustration. In practice, calculation of (6) requires some extra programming work and, even though it is not insurmountable, in most situations one may therefore be inclined to resort to likelihood ratio testing.

Acknowledgments: We acknowledge support from FWO-Vlaanderen Research Project "Sensitivity Analysis for Incomplete and Coarse Data" and Belgian IUAP/PAI network "Statistical Techniques and Modeling for Complex Substantive Questions with Complex Data".

References

Laird, N.M. and Ware, J.H. (1982). Random effects models for longitudinal data. *Biometrics* **38**, 963–974.

Nelder, J.A. (1954). The interpretation of negative components of variance. *Biometrika* **41**, 544–548.

Shapiro, A. (1988). Towards a unified theory of inequality constrained testing in multivariate analysis. *International Statistical Review* **56**, 49–62.

Silvapulle, M.J. and Silvapulle, P. (1995). A score test against one-sided alternatives. *Journal of the American Statistical Association* **90**, 342–349.

Stram, D.O. and Lee, J.W. (1994). Variance components testing in the longitudinal mixed effects model. *Biometrics* **50**, 1171–1177 (correction in 1995).

Verbeke, G. and Molenberghs, G. (2000). *Linear mixed models for longitudinal data*, Springer Series in Statistics, Springer-Verlag, New-York.

Verbeke, G. and Molenberghs, G. (2003). The use of score tests for inference on variance components. *Biometrics*, **59**, 254–262.