Calculating Asymptotic Significance Levels of the Constrained Likelihood Ratio Test with Application to Multivariate Genetic Linkage Analysis

Nathan J. Morris[*]   Robert Elston[†]

Catherine M. Stein[‡]

[*]Case Western Reserve University, njm18@case.edu

[†]Case Western Reserve University, robert.elston@cwru.edu

[‡]Case Western Reserve University, cmstein@case.edu

Calculating Asymptotic Significance Levels of the Constrained Likelihood Ratio Test with Application to Multivariate Genetic Linkage Analysis*

Nathan J. Morris, Robert Elston, and Catherine M. Stein

Abstract

The asymptotic distribution of the multivariate variance component linkage analysis likelihood ratio test has provoked some contradictory accounts in the literature. In this paper we confirm that some previous results are not correct by deriving the asymptotic distribution in one special case. It is shown that this special case is a good approximation to the distribution in many situations. We also introduce a new approach to simulating from the asymptotic distribution of the likelihood ratio test statistic in constrained testing problems. It is shown that this method is very efficient for small p-values, and is applicable even when the constraints are not convex. The method is related to a multivariate integration problem. We illustrate how the approach can be applied to multivariate linkage analysis in a simulation study. Some more philosophical issues relating to one-sided tests in variance components linkage analysis are discussed.

KEYWORDS: multivariate linkage analysis, variance component testing, nonstandard conditions, constrained hypothesis testing, one-sided testing

1. Introduction

There are numerous statistical applications for which one-sided tests may be justified. Indeed, any quick search of the literature on one-sided tests or constrained tests will reveal applications in clinical studies, pharmacokinetics, genetic epidemiology and many other fields. This paper is motivated by a variance component testing problem that occurs in multivariate linkage analysis. In most one-sided situations we have a very simple hypothesis to test, such as $H_0 : \mu = 0$ vs $H_a : \mu > 0$. Here, something like the one-sided t-test is sufficient. However, sometimes the one-sided hypothesis is more complex, such as can be expressed as special cases of $H_0 : \boldsymbol{\theta} \in \Omega_1$ vs $H_a : \boldsymbol{\theta} \in \Omega_2$. The statistics available for this kind of test (Silvapulle and Sen 2005) do not necessarily follow simple asymptotic distributions. Perhaps the most obvious choice of statistic is the constrained maximum likelihood ratio test (CLRT) statistic discussed below. Numerous authors, such as Self and Liang (1987), have shown that in some cases this statistic asymptotically follows a mixture of chi-squared distributions, while in other cases the distribution is more complex. In this paper we discuss some results for constrained hypothesis testing that are relevant to multivariate linkage analysis.

In section 2 we briefly review some of the literature on the CLRT. In section 3 we suggest a simple and efficient approach to calculating the asymptotic significance levels of the CLRT even when the null distribution cannot be described as a mixture of chi-squared distributions. We demonstrate that our method has a strong computational advantage over another known method when small p-values are involved. We then show that the asymptotic distribution may be reduced to a hyperspherical integration problem. In section 4, we demonstrate our method for the multivariate variance component linkage analysis problem (Amos, de Andrade et al. 2001). This model has a particularly complex asymptotic distribution. The general consensus seems to be that "this issue warrants further detailed attention" (Marlow, Fisher et al. 2003), and, "there is an urgent need to characterize the asymptotic distribution associated with these multivariate tests" (Evans, Zhu et al. 2004). In a recent paper, Han and Chang (2008) have challenged the correctness of several relevant results in the literature. They also suggest a fairly simple way of simulating from the asymptotic null distribution. In this paper we confirm some of the findings of Han and Chang by showing that, under certain restrictive assumptions, analytical results are available. When these assumptions are not met, we show that much faster simulation methods are available. In section 5 we discuss some philosophical points concerning constrained testing in the variance component linkage setting, and review our results.

2. Asymptotics under Nonstandard Conditions

By definition, a cone C with its vertex at $\boldsymbol{\theta}_0$ is a set such that, if $\mathbf{x} \in C$, then $a(\mathbf{x} - \boldsymbol{\theta}_0) + \boldsymbol{\theta}_0 \in C$ for all $a \geq 0$. We say that a cone $C(\Omega, \boldsymbol{\theta}_0)$ approximates a set Ω at $\boldsymbol{\theta}_0$ if: $\inf_{C(\Omega, \boldsymbol{\theta}_0)} \|x - y\| = o(\|y - \boldsymbol{\theta}_0\|) \quad \forall y \in \Omega$ and $\inf_{\Omega} \|x - y\| = o(\|x - \boldsymbol{\theta}_0\|)$ $\forall x \in C(\Omega, \boldsymbol{\theta}_0)$. A function $f(z)$ is said to be $o(z)$ if $\lim_{z \to 0} f(z)/z = 0$. The above definition of approximating cones was introduced by Chernoff (1954). Geyer (1994) has discussed its relation to other approximating cones. The primary intuition behind the approximating cone is that, near $\boldsymbol{\theta}_0$, there is a similar shape for both $C(\Omega, \boldsymbol{\theta}_0)$ and Ω. We know from the (typical) consistency of the maximum likelihood estimate that, with a large enough sample size, estimates of a parameter $\boldsymbol{\theta}$ will be very close to the true value $\boldsymbol{\theta}_0$. At this point we may substitute $C(\Omega, \boldsymbol{\theta}_0)$ for the parameter space Ω. The beauty of a cone is that no matter how close we zoom in on the vertex, it looks the same. One fact that falls easily out of the definition of an approximating cone is that all cones approximate themselves. Furthermore, it is quite obvious that the space of positive semi-definite matrices is a cone with vertex at the zero matrix (if M is positive semi-definite, then so is aM for all $a \geq 0$). Indeed, the parameter spaces that are relevant to the present work are already cones and they do not need to be approximated.

Chernoff (1954) produced one of the original works on this subject. Perhaps the most well known work in this area was by Self and Liang (1987). It has been shown (Geyer 1994) that the regularity conditions used by Self and Liang are not sufficient to guarantee all the results given, but the results still hold with corrected regularity conditions. There have been numerous reformulations of the basic results presented below under relaxed regularity conditions (Geyer 1994; Vu and Zhou 1997; Delmas and Foulley 2007). The results have been extended to include the non-identically distributed case (Vu and Zhou 1997; Delmas and Foulley 2007). A thorough review of the literature on constrained testing is given by Silvapulle and Sen (2005).

When some of the parameters lie on the boundary of the parameter space, the asymptotic distribution of the CLRT quickly becomes complicated. We first reproduce the main result from Self and Liang (1987) and Chernoff (1954). However, we assume the regularity conditions used in Theorem 4.4 of Geyer (1994). Let $\boldsymbol{\theta}$ be a vector with p elements containing all model parameters, and let $\boldsymbol{\theta}_0$ be the actual parameters for the generating distribution of $\mathbf{y}$. Let $l(\boldsymbol{\theta}; \mathbf{y})$ be the log-likelihood for the data. Let us define the Fisher information

as $\mathbf{f}(\mathbf{\theta}) = -E\left[\frac{\partial^2}{\partial \mathbf{\theta}^2} l(\mathbf{\theta}; \mathbf{y})\right]$. Suppose we are performing a hypothesis test such that under the null hypothesis $\mathbf{\theta} \in \Omega_0$, and under the alternative hypothesis $\mathbf{\theta} \in \Omega_1$. The CLRT statistic is defined as:

$$\lambda = 2\left[\sup_{\mathbf{\theta} \in \Omega_1} l(\mathbf{\theta}; \mathbf{y}) - \sup_{\mathbf{\theta} \in \Omega_0} l(\mathbf{\theta}; \mathbf{y})\right].$$

We have the following well-known result:

> *Theorem 1:* Let $\mathbf{Z}$ be a multivariate normal (MVN) random variable with mean $\mathbf{\theta}_0$ and variance $\mathbf{f}^{-1}(\mathbf{\theta}_0)$. Let Ω_0 and Ω_1 be sets with non-empty approximating cones $C(\Omega_1, \mathbf{\theta}_0)$ and $C(\Omega_0, \mathbf{\theta}_0)$. Then, under certain regularity conditions, the asymptotic null distribution of the CLRT is the same as the distribution of the likelihood ratio test of $\mathbf{\theta} \in C(\Omega_0, \mathbf{\theta}_0)$ versus the alternative $\mathbf{\theta} \in C(\Omega_1, \mathbf{\theta}_0)$ based on a single realization of Z.

To restate Theorem 1, the asymptotic distribution of the CLRT statistic is equivalent to the distribution of a function of a multivariate random variable ($\mathbf{Z}$):

$$g(\mathbf{Z}) = \inf_{\theta \in C_{\Omega_0}} (\mathbf{Z} - \mathbf{\theta})^T \mathbf{f}(\mathbf{\theta}_0)(\mathbf{Z} - \mathbf{\theta}) - \inf_{\theta \in C_{\Omega_1}} (\mathbf{Z} - \mathbf{\theta})^T \mathbf{f}(\mathbf{\theta}_0)(\mathbf{Z} - \mathbf{\theta}). \quad (1)$$

This can be simplified somewhat. Let $\mathbf{f}^{1/2}$ be a matrix such that $\mathbf{f}^{1/2}\mathbf{f}^{1/2} = \mathbf{f}$ and let $\tilde{\mathbf{Z}}$ be a MVN random variable with mean $\mathbf{0}$ and variance the identity matrix. Theorem 1 then implies that the asymptotic distribution of the CLRT statistic is the same as that of:

$$\tilde{g}(\tilde{\mathbf{Z}}) = \inf_{\theta \in \tilde{C}(\Omega_0, \mathbf{\theta}_0)} \left\| \tilde{\mathbf{Z}} - \mathbf{\theta} \right\|^2 - \inf_{\theta \in \tilde{C}(\Omega_1, \mathbf{\theta}_0)} \left\| \tilde{\mathbf{Z}} - \mathbf{\theta} \right\|^2, \quad (2)$$

where $\tilde{C}(\Omega_i, \mathbf{\theta}_0) = \left\{ \mathbf{f}^{1/2}(\mathbf{\theta} - \mathbf{\theta}_0) \mid \mathbf{\theta} \in C(\Omega_i, \mathbf{\theta}_0) \right\}$. Note that $\tilde{C}(\Omega_i, \mathbf{\theta}_0)$ is a cone with vertex at the origin. Also note that $g(\bullet)$ and $\tilde{g}(\bullet)$ are related by the simple identity $\tilde{g}(\mathbf{V}) = g\left(\mathbf{f}^{-1/2}\mathbf{V} + \mathbf{\theta}_0\right)$.

In many cases the asymptotic distribution turns out to be a mixture of chi-squared distributions. It is fairly easy to understand this nice behavior when $g(\mathbf{Z})$ involves projections onto a set of flat surfaces in $C(\Omega_i, \mathbf{\theta}_0)$. In fact, it can be shown that when testing a hypothesis of the form $H_0 : \mathbf{\theta} = \mathbf{\theta}_0$ vs $H_0 : \mathbf{\theta} \in \Omega_1$,

where $C(\Omega_1, \boldsymbol{\theta}_0)$ is a piecewise smooth convex cone, the asymptotic distribution under the null hypothesis will always be a mixture of chi-squared distributions (Takemura and Kuriki 1997). However, deriving the mixing proportions is often a difficult task. Furthermore, many realistic situations violate these conditions. For example, in variance component linkage analysis it is desirable to use a non-convex space as the alternative hypothesis. This space is illustrated in Figure 1.b. Also, the existence of nuisance parameters may cause the hypothesis to be of a different form. In this paper we suggest that a direct derivation or estimation of mixing proportions is not necessary. To calculate statistical significance it is easy to implement efficient algorithms that do not assume the asymptotic distribution is a mixture of chi-squared random variables.

3. Calculating Statistical Significance

Method 1: Simple Simulation

As discussed earlier, the asymptotic distribution may be represented as a function ($\tilde{g}(\tilde{\mathbf{Z}})$) of standard normal variables. Suppose that *f* is the distribution function of $\tilde{\mathbf{Z}}$, and *S* is a likelihood ratio statistic. Then the p-value for *S* is:

$$P\left(\tilde{g}(\tilde{\mathbf{Z}}) > S\right) = \int_{g(\mathbf{z}) > S} f(\mathbf{z}) d\mathbf{z} . \qquad (3)$$

The simplest possible approach to finding the asymptotic p-value would be to sample from $\tilde{\mathbf{Z}}$ and find the fraction of samples for which $\tilde{g}(\tilde{\mathbf{Z}}) > S$. Various authors have suggested this (Silvapulle and Sen 2005, pp. 78-81). In multivariate linkage analysis this has been suggested by Han and Chang (2008). With very small p-values, often necessary to allow for multiple comparisons, accurate estimates require a large number of simulation samples.

Other simulation methods have been suggested that calculate mixing proportions (Silvapulle and Sen 2005 p 78-81). However, such methods are not applicable when the distribution is not a mixture of chi-squared distributions.

Method 2: Direction and Length Decomposition

We now suggest a more computationally efficient method. We can decompose the MVN distribution into two pieces: a directional p-vector $\mathbf{V} = \tilde{\mathbf{Z}} / \lVert \tilde{\mathbf{Z}} \rVert$ and a scalar $\lVert \tilde{\mathbf{Z}} \rVert$. Thus $\tilde{\mathbf{Z}} = \lVert \tilde{\mathbf{Z}} \rVert \mathbf{V}$. $\mathbf{V}$ can be visualized as a point on the unit hypersphere, while $\lVert \tilde{\mathbf{Z}} \rVert$ is the Euclidian length of $\tilde{\mathbf{Z}}$. It can be shown that

$\left\|\tilde{\mathbf{Z}}\right\|^2 \sim \chi_d^2$, where d is the number of elements in $\boldsymbol{\theta}$ and $\mathbf{V}$ has a uniform distribution on the surface of a unit hypersphere.

The function $\tilde{g}(\mathbf{z})$ has some very interesting properties, including being continuous. Let $\tilde{C}_i = \tilde{C}(\Omega_i, \boldsymbol{\theta}_0)$. The one property that we wish to exploit follows (in line 3 below) from the definition of a cone:

$$
\begin{aligned}
\tilde{g}(a\mathbf{z}) &= \inf_{\theta \in \tilde{C}_0} \left\| a\mathbf{z} - \boldsymbol{\theta} \right\|^2 - \inf_{\theta \in \tilde{C}_1} \left\| a\mathbf{z} - \boldsymbol{\theta} \right\|^2 \\
&= a^2 \inf_{\theta \in \tilde{C}_0} \left\| \mathbf{z} - \boldsymbol{\theta}/a \right\|^2 - a^2 \inf_{\theta \in \tilde{C}_1} \left\| \mathbf{z} - \boldsymbol{\theta}/a \right\|^2 \\
&= a^2 \inf_{\theta \in \tilde{C}_0} \left\| \mathbf{z} - \boldsymbol{\theta} \right\|^2 - a^2 \inf_{\theta \in \tilde{C}_1} \left\| \mathbf{z} - \boldsymbol{\theta} \right\|^2 \\
&= a^2 \tilde{g}(\mathbf{z}).
\end{aligned}
\tag{4}
$$

Let F_d be the cumulative distribution function of χ_d^2. Also, define $F_d(S/0) = 1$ for $S > 0$ and $F_d(0/0) = 0$. Using this notation, we can express the p-value as:

$$
\begin{aligned}
P\left(\tilde{g}(\tilde{\mathbf{Z}}) \geq S \right) &= P\left(\tilde{g}(\left\|\tilde{\mathbf{Z}}\right\| \mathbf{V}) \geq S \right) \\
&= P\left(\left\|\tilde{\mathbf{Z}}\right\|^2 \tilde{g}(\mathbf{V}) \geq S \right) \\
&= E_{g(\mathbf{V})} P\left(\left\|\tilde{\mathbf{Z}}\right\|^2 \tilde{g}(\mathbf{V}) \geq S \mid \tilde{g}(\mathbf{V}) \right) \\
&= E_{g(\mathbf{V})} \left[1 - F_d\left(S/\tilde{g}(\mathbf{V}) \right) \right].
\end{aligned}
\tag{5}
$$

The following simulation can be performed to calculate the p-value:

1. Simulate a d-vector from a multivariate normal distribution with mean 0, then divide it by its norm to obtain a sample from $\mathbf{V}$.
2. Calculate $\tilde{g}(\mathbf{V}) = g\left(\mathbf{f}^{1/2}\mathbf{V} + \boldsymbol{\theta}_0 \right)$. Note: this will usually be the most computationally demanding step.
3. Repeat steps 1 and 2 n times.
4. Find the estimated p-value:

$$
\hat{p}_n = 1 - \tfrac{1}{n} \sum_i F_d\left(S/\tilde{g}(\mathbf{V}_i) \right).
\tag{6}
$$

Implementing the simulation in this way can involve significant computational gains. Let us denote $1 - F_d(S)$ as p_U for the unconstrained p-value and let p denote the actual p-value. We point out the following inequality which is shown in Appendix A:

$$Var(\hat{p}_n) \le \frac{(p_U - p)p}{n}. \qquad (7)$$

As a crude estimate of what we gain in efficiency, we looked at the ratio of the variances achieved by using Method 2 (designated $\hat{p}_n$) versus that using Method 1 (designated $\tilde{p}_n$):

$$Var(\hat{p}_n)/Var(\tilde{p}_n) \le \frac{p_U - p}{1 - p}. \qquad (8)$$

Now note that $(p_U(S) - p(S))/(1 - p(S)) \to 0$ as $S \to \infty$. In other words, for very small p-values (and large S), $Var(\hat{p}_n)$ is a tiny fraction of $Var(\tilde{p}_n)$. This improvement is purchased at virtually no increased computational expenditure.

Consider the special case of testing $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ versus $H_a : \boldsymbol{\theta} \in \Omega_1$ where Ω is sufficiently regular to define the non-trivial cone $\tilde{C}(\Omega_1, \boldsymbol{\theta}_0)$. In this case we may improve upon (8) somewhat. We show in Appendix A that:

$$Var(\hat{p}_n)/Var(\tilde{p}_n) \le 1 - \frac{F_d\left[F_1^{-1}(1 - 2p)\right]}{1 - p}. \qquad (9)$$

Equation (9) shows again that for very small p-values $Var(\hat{p}_n)$ is a tiny fraction of $Var(\tilde{p}_n)$. These equations, however, show only worst case scenarios and in actual practice the benefits of Method 2 may be even greater. In section 4 we show an empirical evaluation of the relative variance of the two estimators in several situations.

One potential way to speed up the simulation is this: for each new $\mathbf{V}_i$ we can use $-\mathbf{V}_i$ as another observation. This leads to a negative correlation among the observations, thus reducing the variance. Another modification of the basic procedure is to implement a stopping rule to determine how many simulations to perform. In general, very large p-values do not need to be estimated accurately.

Method 3: Numerical Integration Approach

It is possible to avoid completely the use of Monte Carlo simulation. $\mathbf{V}$ can be represented in a hyperspherical coordinate system (Blumenson 1960). This involves a transformation from a series of d-1 angles to the d elements of $\mathbf{V}$. So:

$$
\mathbf{V} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ \vdots \\ v_{p-1} \\ v_p \end{pmatrix} = \begin{pmatrix} \cos(\varphi_1) \\ \sin(\varphi_1)\cos(\varphi_2) \\ \sin(\varphi_1)\sin(\varphi_2)\cos(\varphi_3) \\ \vdots \\ \sin(\varphi_1)\sin(\varphi_2)...\sin(\varphi_{p-2})\cos(\varphi_{p-1}) \\ \sin(\varphi_1)\sin(\varphi_2)...\sin(\varphi_{p-2})\sin(\varphi_{p-1}) \end{pmatrix},
\tag{10}
$$

where φ_{p-1} goes from 0 to 2π, and φ_1 through φ_{p-2} go from 0 to π. The Jacobian for this transformation is: $J = \sin^{p-2}(\phi_1)\sin^{p-3}(\phi_2)...\sin(\phi_{p-2})$. Using this transformation it is possible to evaluate the integral numerically:

$$
\begin{aligned}
P\left(\tilde{g}(\tilde{\mathbf{Z}}) > S\right) &= E_{g(\mathbf{V})}\left[1 - F_d\left(\frac{S}{\tilde{g}(\mathbf{V})}\right)\right] \\
&= \frac{1}{S_p}\int_{\varphi}\left[1 - F_d\left(\frac{S}{\tilde{g}(\mathbf{V})}\right)\right]J d\boldsymbol{\varphi}
\end{aligned}
\tag{11}
$$

Here $S_d = 2\pi^{d/2}/\Gamma(d/2)$ is the surface area of a d dimensional hypersphere of radius 1. We use the integral representation in our proof of theorem 2 below.

Nuisance Parameters

Suppose that we divide the parameter space into q parameters to be tested and d-q nuisance parameters: $\boldsymbol{\theta}^T = \begin{bmatrix} \boldsymbol{\psi}^T & \boldsymbol{\lambda}^T \end{bmatrix}$. Similarly, we partition the Fisher information matrix as:

$$
\mathbf{f}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{f}_{\psi\psi} & \mathbf{f}_{\psi\lambda}^T \\ \mathbf{f}_{\psi\lambda} & \mathbf{f}_{\lambda\lambda} \end{pmatrix}.
\tag{12}
$$

We wish to test a hypothesis of the form $H_0 : \mathbf{\theta} \in \Omega_0 = \{\mathbf{\theta} : \mathbf{\psi} = \mathbf{b} \ and \ \lambda \in \Omega_*\}$ versus $H_a : \mathbf{\theta} \in \Omega_1 = \{\mathbf{\theta} : \mathbf{\psi} \in \Omega_a \ and \ \lambda \in \Omega_*\}$. We define $\Omega_b = \{\mathbf{\theta} : \mathbf{\psi} = \mathbf{b}\}$ for notational convenience. In other words, we allow for the nuisance parameters to have constraints imposed upon them. This may be reasonable, for example, in the variance component setting. The first problem with having a large number of nuisance parameters is that the function $g(\mathbf{Z})$ involves optimizing two quadratic forms each with d^2 terms. Lemma 1, below, states two conditions under which we may reduce these quadratic forms to have only q^2 terms. Let $\mathbf{M}(\mathbf{\theta}_0) = \mathbf{f}_{\psi\psi} - \mathbf{f}_{\psi\lambda}\mathbf{f}_{\lambda\lambda}^{-1}\mathbf{f}_{\psi\lambda}^T$ be the Schur complement of the block $\mathbf{f}_{\lambda\lambda}$ of $\mathbf{f}$, and $h(\mathbf{Y}) = \inf_{\psi \in C(\Omega_b, \mathbf{b})} (\mathbf{Y} - \mathbf{\psi})^T \mathbf{M}(\mathbf{Y} - \mathbf{\psi}) - \inf_{\psi \in C(\Omega_a, \mathbf{b})} (\mathbf{Y} - \mathbf{\psi})^T \mathbf{M}(\mathbf{Y} - \mathbf{\psi})$. Also, define $\tilde{h}(\tilde{\mathbf{Y}}) = \inf_{\psi \in \tilde{C}(\Omega_b, \mathbf{b})} \|\tilde{\mathbf{Y}} - \mathbf{\psi}\|^2 - \inf_{\psi \in \tilde{C}(\Omega_a, \mathbf{b})} \|\tilde{\mathbf{Y}} - \mathbf{\psi}\|^2$.

> *Lemma 1:* If either (a) $\mathbf{f}_{\psi\lambda}(\mathbf{\theta}_0) = 0$ *or* (b) $C(\Omega_*, \mathbf{\theta}_0) = \mathbb{R}^{d-q}$, then $P_{\mathbf{Y}}[h(\mathbf{Y}) \geq S] = P_{\mathbf{Z}}[g(\mathbf{Z}) \geq S]$, where $\mathbf{Y}$ and $\mathbf{Z}$ are MVN variables with means $\mathbf{b}$ and $\mathbf{\theta}_0$ and variances $\mathbf{M}$ and $\mathbf{f}$, respectively.

Proof of Lemma 1 is given in Appendix A. Note that only one of the two conditions must be met for the results to follow. One situation where condition (a) is met is in the normal variance component setting where the fixed effects are being tested. In this situation, the part of the Fisher information corresponding to the covariances between the estimators of the main effects and the variance components is usually $\mathbf{0}$ and, as a result, enforcing constraints on variance components has no effect on the asymptotic distribution of the test. Note that if condition (a) is met, then we may set $\mathbf{M} = \mathbf{f}_{\psi\psi}$. Condition (b) requires that any constraints on the nuisance parameters disappear asymptotically. Condition (b) is satisfied when λ is an interior point of Ω_*.

A second problem with nuisance parameters is that using Theorem 1 and Lemma 1 requires us to know the actual parameter values ($\mathbf{\theta}_0$). One natural approach would be replace $\mathbf{\theta}_0$ and $\mathbf{f}$ in $g(\mathbf{Z}; \mathbf{f}(\mathbf{\theta}_0); C(\Omega_0, \mathbf{\theta}_0), C(\Omega_1, \mathbf{\theta}_0))$ with some consistent estimate ($\hat{\mathbf{\theta}}, \hat{\mathbf{f}}$). Unfortunately, this will not always work because $g(\mathbf{Z}; \mathbf{f}; C(\Omega_0, \mathbf{\theta}_0), C(\Omega_1, \mathbf{\theta}_0))$ is not a continuous function of $\mathbf{\theta}_0$. The approximating cone $C(\Omega, \theta_0)$ may behave badly as a function of $\mathbf{\theta}_0$. As a simple example, consider the space $\Omega = \{\theta \mid \theta \geq 0\}$. If $\theta_0 > 0$, then $C(\Omega, \theta_0) = \mathbb{R}$; but

if $\theta_0 = 0$, then $C(\Omega, \theta_0) = \{\theta \mid \theta \geq 0\}$. This problem has been discussed from a different point of view by Wolak (1991). Under the null hypothesis discussed above, $\boldsymbol{\psi}$ is specified and does not need to be estimated, while $\boldsymbol{\lambda}$ must be estimated. However, $C(\Omega_b, \mathbf{b})$ and $C(\Omega_a, \mathbf{b})$ are not functions of $\boldsymbol{\lambda}$ and hence the bad behavior of the approximating cone may be avoided when the conditions for Lemma 1 hold. Let $\hat{\mathbf{f}}_n$ be an estimate of $\mathbf{f}(\boldsymbol{\theta}_0)$, $\hat{\mathbf{M}}_n = \hat{\mathbf{f}}_{\psi\psi n} - \hat{\mathbf{f}}_{\psi\lambda n} \hat{\mathbf{f}}_{\lambda\lambda n}^{-1} \hat{\mathbf{f}}_{\psi\lambda n}^{T}$, $S_{jn} = \left\{ \hat{\mathbf{M}}_n^{1/2}(\boldsymbol{\theta} - \mathbf{b}) \mid \boldsymbol{\theta} \in C(\Omega_j, \mathbf{b}) \right\}$ and $\tilde{h}_n(\tilde{\mathbf{Y}}) = \inf_{\psi \in S_{bn}} \left\| \tilde{\mathbf{Y}} - \boldsymbol{\psi} \right\|^2 - \inf_{\psi \in S_{an}} \left\| \tilde{\mathbf{Y}} - \boldsymbol{\psi} \right\|^2$. Also, suppose that LR_n is an observed test statistic.

> *Theorem 2:* Suppose that $\hat{\mathbf{f}}_n \xrightarrow{P} \mathbf{f}(\boldsymbol{\theta}_0)$ under H_0. If (a) $\mathbf{f}_{\psi\lambda}(\boldsymbol{\theta}_0) = 0$ *or* (b) $C(\Omega_*, \boldsymbol{\theta}_0) = \mathbb{R}^{d-q}$, then for all $0 \leq \alpha < 1 - P(\tilde{g}(\tilde{\mathbf{Z}}) = 0)$
>
> $$P\left\{ P_{\tilde{\mathbf{Y}}}\left[\tilde{h}_n(\tilde{\mathbf{Y}}) \geq LR_n \mid \hat{\mathbf{f}}_n, LR_n \right] \leq \alpha \right\} \to \alpha \qquad (13)$$
>
> where the elements of $\tilde{\mathbf{Y}}$ are distributed as independent standard normal random variables.

A sketch of the proof for Theorem 2 is deferred to Appendix A. Hence simulating from $\tilde{h}_n(\tilde{\mathbf{Y}})$ is equivalent asymptotically to simulating from $\tilde{g}(\tilde{\mathbf{Z}})$. This approach has been criticized by some (Silvapulle and Sen 2005, p. 153) because the asymptotic distribution may be quite sensitive to the nuisance parameters. However, as we show below, there are at least some situations where the asymptotic distribution is relatively insensitive to the actual values of the nuisance parameters.

4. Multivariate Variance Component Linkage Analysis

Some Model Properties

First, we present the basic model for multivariate variance component linkage analysis (Amos, de Andrade et al. 2001). Let there be t traits (variates), n_k individuals in the kth pedigree, and K pedigrees. Let $\mathbf{y}_{ki}$ be a $t \times 1$ vector of trait values for individual i in pedigree k, and let $\mathbf{y}_k = \begin{bmatrix} \mathbf{y}_{k1}^T & \mathbf{y}_{k2}^T & \cdots & \mathbf{y}_{kn_k}^T \end{bmatrix}^T$. The

commonly used genetic model for quantitative traits assumes that $\mathbf{y}_{ki}$ is formed by adding together a set of independent random effects. For example, there may be a random environmental (including measurement error) effect and a random polygenic effect (i.e. the cumulative effect of alleles in many genes acting together additively). Because family members share on average a certain proportion of their alleles there will be some correlation between the random polygenic effects of family members. In linkage analysis, we are interested in testing whether a particular locus in the genome is linked to the traits of interest, and so we also have a random effect for a single gene of interest. Using genetic marker data we may calculate the proportion of alleles shared identical by descent (IBD) (Haseman and Elston 1972) for each pair of relatives at a particular genomic location. While there are many variations and additional random effects suggested in the literature, we will consider only these three random effects: linked locus, polygenic and environmental. First we define the following:

$\mathbf{\Pi}_k$ specifies the proportion of alleles shared identical by descent for each pair of relatives at a particular genomic location.

$\mathbf{\Phi}_k$ is a matrix of kinship coefficients.

$\mathbf{I}_k$ is an $n_k \times n_k$ identity matrix.

$\mathbf{A}$ represents the covariance matrix for the additive component of variance.

$\mathbf{P}$ represents a polygenic covariance matrix and

$\mathbf{E}$ represents an environmental (including measurement error) covariance matrix.

Note that $\mathbf{A}$, $\mathbf{P}$ and $\mathbf{E}$ are each $t \times t$ covariance matrices containing parameters, while $\mathbf{\Pi}_k$, $\mathbf{\Phi}_k$ and $\mathbf{I}_k$ are $n_k \times n_k$ matrices containing observed properties of the pedigree. In the multivariate variance component model, it is assumed that, given the number of alleles shared identical by descent for the pairs of members in each pedigree, $\mathbf{y}_k$ follows a MVN distribution with covariance matrix

$$\mathbf{C}_k = \mathbf{\Pi}_k \otimes \mathbf{A} + \mathbf{\Phi}_k \otimes \mathbf{P} + \mathbf{I}_k \otimes \mathbf{E}, \qquad (14)$$

where $\otimes$ represents the Kronecker product.

In general, we wish to test for linkage using the model above. This is a hypothesis test of the form $H_0 : \mathbf{A} = 0$ vs $H_a : \mathbf{A} \in \Omega_a$. For example, in the univariate case ($t = 1$) we have $H_0 : a = 0$ vs $H_a : a > 0$ because we expect variances to be positive. However, in the multivariate case there are several options for Ω_a. As we point out in the discussion, each of these options is valid if decided upon *a priori*. First, we may set $\Omega_a = \{\mathbf{A} \mid \mathbf{A} \geq 0\}$ where $\mathbf{A} \geq 0$ means that $\mathbf{A}$ is positive semi-definite. We will refer to this as constraint 1. Constraint 1

may be justified by the fact that covariance matrices must be positive semi-definite by definition. Second, we may choose $\Omega_a = \left\{ \boldsymbol{\varphi}\boldsymbol{\varphi}^T \mid \boldsymbol{\varphi} \in \mathbb{R}^t \right\}$ which we will refer to as constraint 2 (Todorov, Vogler et al. 1998; Amos, de Andrade et al. 2001; Marlow, Fisher et al. 2003). This is equivalent to constraining all the cross-trait correlations to be ± 1. Constraint 2 may be justified when the polymorphisms at a given genomic location responsible for the trait are effectively diallelic. Third, we may have $\Omega_a = \left\{ a\left(\mathbf{1}_t \mathbf{1}_t^T\right) \mid a > 0 \right\}$, where $\mathbf{1}_t$ is a column t-vector of ones (Evans, Zhu et al. 2004). This kind of constraint may be reasonable for repeated measures if we believe the gene will have the same effect on all measurements. Each of these constraints is illustrated in Figure 1. Wang (2003) and Amos et al. (2001) present some other possible constraints. A number of similar tests of variance components have been presented in contexts other than linkage analysis (Stram and Lee 1994; Stram and Lee 1995).
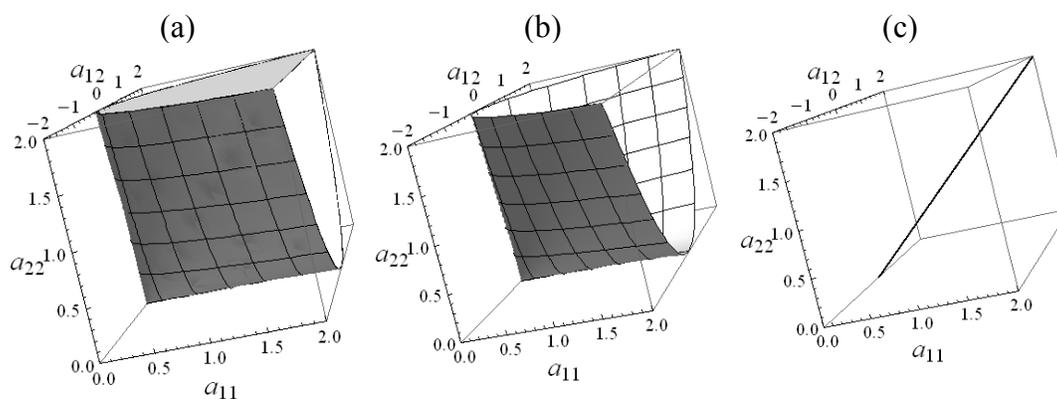
Figure 1: Geometric view of the alternative hypothesis in the bivariate case. (a) Constraint 1: $\Omega_a = \left\{ \mathbf{A} \mid \mathbf{A} \geq 0 \right\}$, (b) Constraint 2: $\Omega_a = \left\{ \boldsymbol{\varphi}\boldsymbol{\varphi}^T \mid \boldsymbol{\varphi} \in \mathbb{R}^t \right\}$, (c) Constraint 3: $\Omega_a = \left\{ \left(\mathbf{1}_t \mathbf{1}_t^T\right) a \mid a > 0 \right\}$.

Under the third constraint, the asymptotic distribution is a $\frac{1}{2} : \frac{1}{2}$ mixture of χ_0^2 and χ_1^2, where χ_0^2 represents a point mass at 0 (Evans, Zhu et al. 2004). To our knowledge, the literature does not directly address the asymptotic null distribution under constraint 1. However, constraint 2 has prompted some confusion and even contradiction within the literature. As mentioned earlier, Han and Chang (2008) have challenged several claims in the literature regarding constraint 2.

Although pedigrees come in various structures and sizes, it is still possible to force the variance component models discussed in the introduction into an

independent identically distributed context. We assume that there is an infinite population of non-overlapping pedigree structures from which we randomly sample. Although the whole human population can be considered as a single pedigree, in practice the way pedigrees are sampled can ensure that they are non-overlapping (Ginsburg, Elston et al. 2006). Thus, the probability $P\left(\mathbf{\Pi}_k, \mathbf{\Phi}_k, n_k\right)$ is not a function of the variance component parameters, and the likelihood function is of the form $\prod_k P\left(y_k \mid \mathbf{\Pi}_k, \mathbf{\Phi}_k, n_k\right) P\left(\mathbf{\Pi}_k, \mathbf{\Phi}_k, n_k\right)$. In the likelihood ratio, the factor $P\left(\mathbf{\Pi}_k, \mathbf{\Phi}_k, n_k\right)$ will cancel out. We have not proven in this work that the multivariate variance component model meets the regularity requirement for Theorems 1 and 2. However, our simulation results show the usefulness of our approach. In fact, the regularity requirements given by Geyer (1994) require global optima. The likelihood that needs to be optimized for our model is not convex, and it would be difficult to prove that the global maximum was truly achieved. Instead, we depend on the heuristic approach of using multiple starting points.

We have chosen below to enforce the constraint that $\mathbf{P}$ and $\mathbf{E}$ are positive definite under both the null and alternative hypotheses. However, we assume that the true parameters $\mathbf{P}$ and $\mathbf{E}$ do not lie on the boundary of the parameter space, and hence these constraints disappear asymptotically. Thus conditions (a) and (b) in Theorem 2 are both satisfied.

We are now ready to define a set of conditions:

(C1) $\mathbf{P}$ and $\mathbf{E}$ do not lie on the boundary of the parameter space.

(C2) Given the pedigree structure, the data generating process actually follows a MVN distribution with $Var\left(\mathbf{y}_k\right) = \mathbf{C}_k = \mathbf{\Phi}_k \otimes \mathbf{P} + \mathbf{I}_k \otimes \mathbf{E}$ and $E\left(\mathbf{y}_{ki}\right) = \mathbf{\mu}$.

(C3) The random variables $\mathbf{\Pi}$ and $\mathbf{\Phi}$ may only take on a finite number of possible values.

Next we define the following space:

$$\Gamma\left(\varepsilon, \eta\right) = \{\mathbf{\theta} \mid \lambda_{\min}\left(\mathbf{\Pi} \otimes \mathbf{A} + \mathbf{\Phi} \otimes \mathbf{P} + \mathbf{I} \otimes \mathbf{E}\right) \geq \varepsilon \text{ for all } \mathbf{\Pi}, \mathbf{\Phi}$$
$$and \ \|\mathbf{\theta}\| \leq \eta\} \tag{15}$$

Lemma 2: If (C1) through (C3) are met, then for constraints 1 and 2 Theorem 1 and 2 may be applied to the variance component likelihood ratio test for linkage. Here the likelihood ratio test is defined for arbitrarily large η and for $\varepsilon > 0$ arbitrarily close to 0, and the test statistic is

$$\lambda = 2\left[\sup_{\mathbf{\theta} \in \Omega_1 \cap \Gamma(\varepsilon, \eta)} l\left(\mathbf{\theta}; \mathbf{y}\right) - \sup_{\mathbf{\theta} \in \Omega_0 \cap \Gamma(\varepsilon, \eta)} l\left(\mathbf{\theta}; \mathbf{y}\right)\right].$$

We have not rigorously proven *Lemma 2*, but a very brief discussion of how these conditions may be related to the regularity conditions given by Geyer may be found in Appendix B. In fact, the regularity requirements given by Geyer (1994) require global optima. The likelihood function for our model is not convex, and it would be difficult to prove that the global maximum was truly achieved. Instead, we depend on the heuristic approach of using multiple starting points. Our simulation results show the usefulness of our approach.

It is convenient to reparametrize the covariance structure to be $\mathbf{\Sigma}_k = (\mathbf{\Pi}_k - \mathbf{\Phi}_k) \otimes \mathbf{A} + \mathbf{\Phi}_k \otimes \mathbf{P}^* + \mathbf{I}_k \otimes \mathbf{E}$, where $\mathbf{P}^* = \mathbf{P} - \mathbf{A}$. Let $\mathbf{B}_{ij} = \mathbf{e}_i \mathbf{e}_j^T + \mathbf{e}_j \mathbf{e}_i^T \, If [j \neq i]$, where $\mathbf{e}_i$ is the unit vector of length t in the ith direction, and let $If [\cdot]$ be the indicator function, In this case we can easily derive that, conditional on $\mathbf{\Pi}_k$ and the pedigree structure, the Fisher information for $\mathbf{A}$ and the cross parameter information between $\mathbf{A}$ and P* and that between $\mathbf{A}$ and $\mathbf{E}$ are respectively:

$$\hat{\mathbf{f}}_{a_{i_1 i_2} a_{j_1 j_2}} = \tfrac{1}{K} \sum_k tr \left[\mathbf{C}_k^{-1} \left((\mathbf{\Pi}_k - \mathbf{\Phi}_k) \otimes \mathbf{B}_{i_1 i_2} \right) \mathbf{C}_k^{-1} \left((\mathbf{\Pi}_k - \mathbf{\Phi}_k) \otimes \mathbf{B}_{j_1 j_2} \right) \right]$$

$$\hat{\mathbf{f}}_{p_{i_1 i_2} a_{j_1 j_2}} = \tfrac{1}{K} \sum_k tr \left[\mathbf{C}_k^{-1} \left(\mathbf{\Phi}_k \otimes \mathbf{B}_{i_1 i_2} \right) \mathbf{C}_k^{-1} \left((\mathbf{\Pi}_k - \mathbf{\Phi}_k) \otimes \mathbf{B}_{j_1 j_2} \right) \right] \qquad (16)$$

$$\hat{\mathbf{f}}_{e_{i_1 i_2} a_{j_1 j_2}} = \tfrac{1}{K} \sum_k tr \left[\mathbf{C}_k^{-1} \left(\mathbf{I}_k \otimes \mathbf{B}_{i_1 i_2} \right) \mathbf{C}_k^{-1} \left((\mathbf{\Pi}_k - \mathbf{\Phi}_k) \otimes \mathbf{B}_{j_1 j_2} \right) \right].$$

However, $E[\mathbf{\Pi}_k] - \mathbf{\Phi}_k = 0$ and therefore $\hat{\mathbf{f}}_{p_{i_1 i_2} a_{j_1 j_2}}$ and $\hat{\mathbf{f}}_{e_{i_1 i_2} a_{j_1 j_2}}$ will converge in probability to $\mathbf{0}$.

The above formulas show the Fisher information conditional on $\mathbf{\Pi}_k$ and pedigree structure, but $\hat{\mathbf{f}}_{a_{i_1 i_2} a_{j_1 j_2}}$ will converge in probability to some value $\mathbf{f}_{a_{i_1 i_2} a_{j_1 j_2}}$ that is dependent on the pedigree structures in the population. It may also be of interest to know some typical values for $\mathbf{f}_{a_{i_1 i_2} a_{j_1 j_2}}$. Assume we have a population of families containing only a father, mother and two siblings each with a fully informative marker. For this population, letting $\mathbf{\Delta} = \mathbf{e}_3 \mathbf{e}_4^T + \mathbf{e}_4 \mathbf{e}_3^T$, where $\mathbf{e}_i$ is the unit vector in the ith direction of length 4, the Fisher information is:

$$\mathbf{f}_{a_{i_1 i_2} a_{j_1 j_2}} = \frac{1}{8} tr \left[\mathbf{C}^{-1} \left(\mathbf{\Delta} \otimes \mathbf{B}_{i_1 i_2} \right) \mathbf{C}^{-1} \left(\mathbf{\Delta} \otimes \mathbf{B}_{j_1 j_2} \right) \right] \qquad (17)$$

Alternative Hypothesis Constraint 1

Constraint 2, as pictured in Figure 1, requires that **A** be positive semi-definite. Such a constraint is defensible because variances must be positive. To our knowledge there is no work suggesting an asymptotic distribution for the multivariate linkage model under this constraint. In Theorem 3 we show that in one special case the asymptotic distribution is available. We first define one more condition: (C4) $\mathbf{P} = \gamma\mathbf{E}$ for some positive scalar γ.

> *Theorem 3:* If (C1) through (C4) are met, then for an arbitrary number of traits the null asymptotic cumulative distribution function (CDF) of the multivariate linkage analysis CLRT statistic under constraint 1 is a mixture of chi-squared random variables with mixing proportions derivable by the recursion given in Kuriki (1993).

Theorem 3 follows immediately from Lemma 2 and Lemmas 3 and 5 in Appendix B and the work of Kuriki (1993). In Table 1 we give some of the mixing proportions derivable from the recursive formulas in Kuriki (1993).

Of course, it is not reasonable to assume that $\mathbf{P} = \gamma\mathbf{E}$, and so we might expect the asymptotic distribution derived in Theorem 3 to be of only theoretical interest. However, we may still use the methods suggested in section 3 if $\mathbf{P} \neq \gamma\mathbf{E}$. Our approach requires us to replace the nuisance parameters **P** and **E** with consistent estimates. As mentioned earlier, if the distribution is strongly dependent on **P** and **E**, then this approach may be subject to some criticism for finite samples. We briefly investigate this in Figure 2. We recalculated the asymptotic CDF under different possible values for r, where

$$P = \begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix} \text{ and } E = \begin{bmatrix} 1 & -r \\ -r & 1 \end{bmatrix}. \tag{18}$$

The results in Figure 2 were obtained from 200,000 simulation samples using Method 2 in section 3. We implemented this method in MATLAB and used the cutting-plane algorithm suggested by Shaw and Geyer (1997) for the optimization. As can be seen, in the cases investigated, the asymptotic distribution using different values for **P** and **E** appear to be nearly indistinguishable. While these results seem to indicate that Theorem 3 has some practical value, we have no guarantee that **P** and **E** will have such a small effect on the asymptotic distribution for other pedigree structures, numbers of traits or population parameter values. Therefore, we can only recommend using Theorem 3 as a first pass approximation.
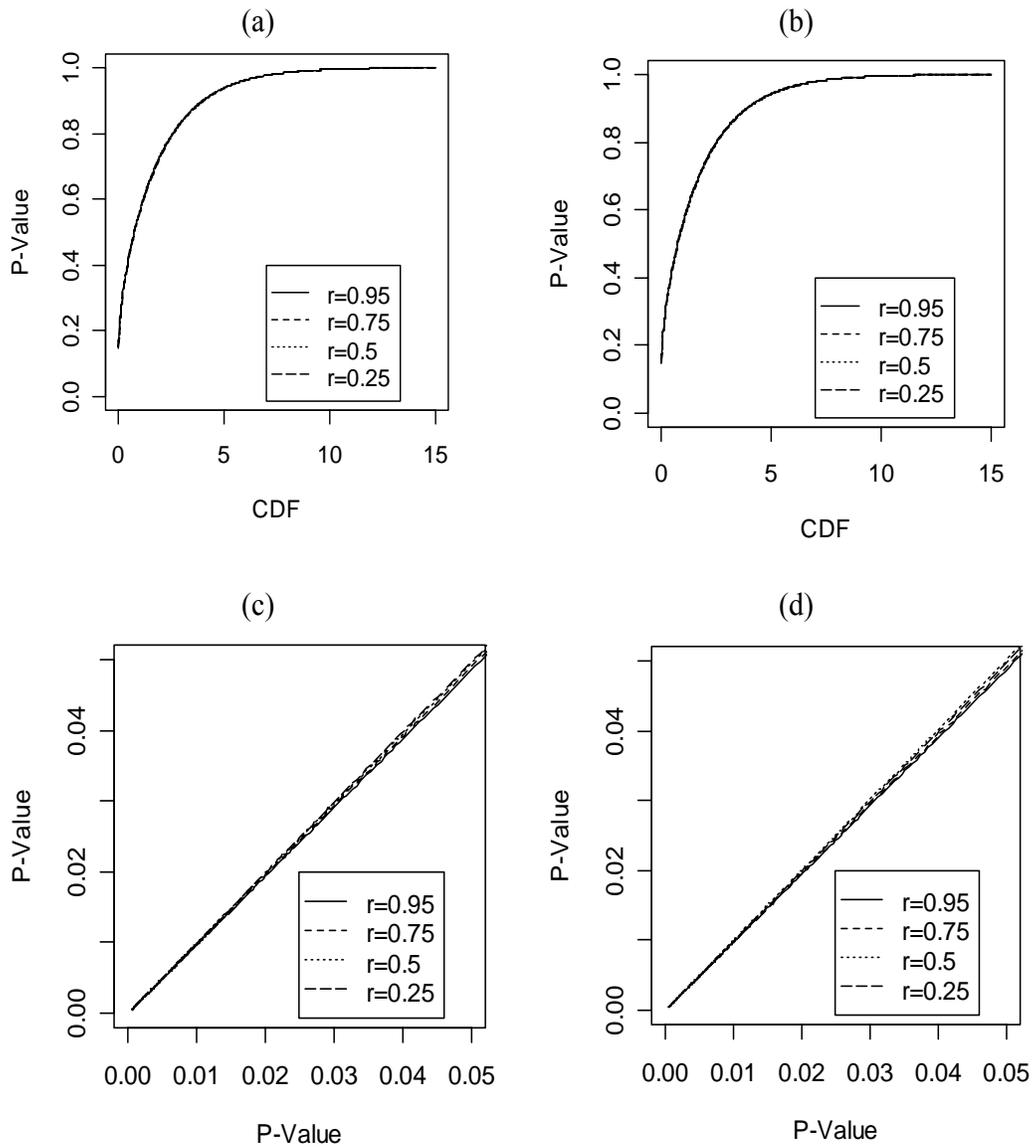
Figure 2: Comparison of the asymptotic distribution of the CLRT statistic for a population of pedigrees containing only a father, mother and two sibs when **P** and **E** are parameterized as in equation (18). The CDFs were calculated based on Method 2 in section 3. (a) CDF under constraint 1. (b) CDF under constraint 2. (c) This comparison is for constraint 1. The x-axis represents a p-value from the analytical formula derived in Theorem 3; The y-axis represents the p-value calculated as in (a). (d) This comparison is for constraint 2. The x-axis represents a p-value from the distribution derived in Theorem 4; The y-axis represents the p-value calculated as in (b).

Table 1: Mixing Proportions for Constraint 1

DF	$t=1$	$t=2$	$t=3$
0	2^{-1}	$2^{-1}-\left(2\sqrt{2}\right)^{-1}$	$2^{-2}-\left(\pi\sqrt{2}\right)^{-1}$
1	2^{-1}	$\left(2\sqrt{2}\right)^{-1}$	$2^{-2}\left(\sqrt{2}-1\right)$
2	0	$\left(2\sqrt{2}\right)^{-1}$	$\left(\pi\sqrt{2}\right)^{-1}$
3	0	$2^{-1}-\left(2\sqrt{2}\right)^{-1}$	$1-1/\sqrt{2}$
4	0	0	$\left(\pi\sqrt{2}\right)^{-1}$
5	0	0	$2^{-2}\left(\sqrt{2}-1\right)$
6	0	0	$2^{-2}-\left(\pi\sqrt{2}\right)^{-1}$

Alternative Hypothesis Constraint 2

Constraint 2, as pictured in Figure 1, requires that $\mathbf{A}$ be of the form $\boldsymbol{\varphi}\boldsymbol{\varphi}^{T}$ (Todorov, Vogler et al. 1998; Amos, de Andrade et al. 2001; Marlow, Fisher et al. 2003). This is equivalent to constraining all the cross-trait correlations to be ± 1. In the bivariate case, Evans (2002) seems to have claimed that the asymptotic distribution is $\frac{1}{2}\chi_1^2+\frac{1}{2}\chi_2^2$ while Amos et al. (2001) have claimed that the asymptotic distribution is $\frac{1}{4}\chi_0^2+\frac{1}{2}\chi_1^2+\frac{1}{4}\chi_2^2$. In fact, both of these claims are incorrect, although the distribution suggested by Evans closely approximates the true asymptotic distribution in the tails for some values of $\boldsymbol{\theta}_0$. Theorem 4 gives the asymptotic distribution in one special case, and Figure 3 displays a graphical comparison of the two previous suggestions and the true distribution in this special case.

> *Theorem 4:* If (C1) through (C3) are met, then the null asymptotic CDF of the bivariate ($t=2$) linkage analysis statistic under constraint 2 is:
> $$F(z)=\left(\Phi\left(\sqrt{2z}\right)-e^{-z/2}\,\Phi\left(\sqrt{z}\right)\big/\sqrt{2}\right)If\left[z>0\right].$$

Here $\Phi(x)$ represents the CDF of the standard normal distribution and *If* [.] is the indicator function. Proof of Theorem 4 is discussed in Appendix B.

Note that the asymptotic distribution in Theorem 4 does not appear to be a mixture of chi-squared random variables. This is precisely what we would expect, because the alternative hypothesis is not convex. Figure 3 shows that using the distribution suggested by Amos et al. (2001) may be asymptotically quite liberal in terms of type 1 error, as may be expected from the simulation results given by Amos et al. (2001). This also confirms the work of Han and Chang (2008).

As with constraint 1, we investigated the sensitivity of the asymptotic distribution to the values of $\mathbf{P}$ and $\mathbf{E}$. Figure 2 shows that the models investigated have nearly identical asymptotic distributions. Hence $\mathbf{P}$ and $\mathbf{E}$ have very little effect, as we may have suspected from the simulation work of Amos et al. (2001).
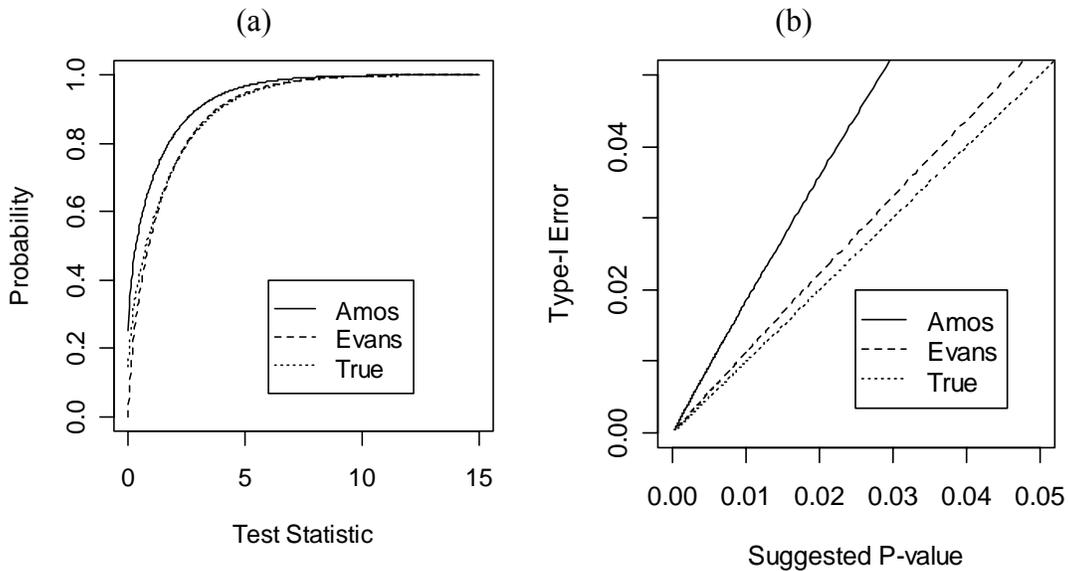

Figure 3: Comparison of some suggested distributions to the true asymptotic distribution. (a) Plot of the CDF function. (b) The x-axis of this plot represents a p-value from one of the suggested distributions; the y-axis represents the p-value that would have been calculated using the distribution in Theorem 4.

Empirical Evaluation of Method 2 versus Method 1

In equation (9) we gave an upper bound for the relative variances of estimation Methods 1 and 2. We now give an empirical evaluation of the relative variances. Consider the situation that $\mathbf{P} = \mathbf{E} = \mathbf{I}_2$. We may use the formulas derived in Appendix B to easily find the Fisher information in this case. We simulated 100,000 samples of $\mathbf{V}$ as in Method 2, and in each case we evaluated

$1 - F_d\left(S/\tilde{g}(\mathbf{V})\right)$ at a large number of possible values for *S*. At each of those values for *S* we estimated the associated p-value and its variance for Methods 1 and 2. Figure 4 shows the results. As may be seen, the relative variance is much better than the upper bound, and for small p-values Method 2 is much better than Method 1.
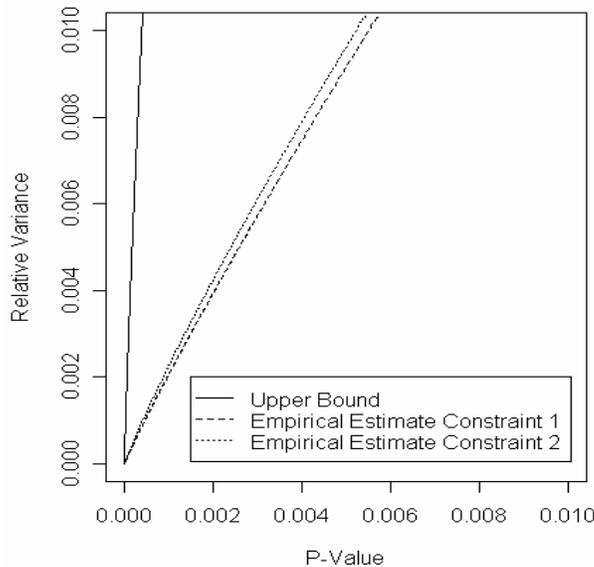
Figure 4: Empirical evaluation of the benefit of Method 2. The y-axis is the relative variance $(Var(\hat{p}_n)/Var(\tilde{p}_n))$. The x-axis is the estimated p-value. The upper bound is as derived in (9). The lines for constraints 1 and 2 are based on 100,000 simulations using Method 2.

Simulation Study of Convergence

To see how large a sample is needed for the statistic to approach the asymptotic distribution under the null hypothesis, we performed a small simulation study. Pedigrees with 2 parents and 2 siblings were simulated. A fully informative marker was assumed. The data were simulated to follow the actual variance component model. All simulations were performed using a program written in MATLAB release 2007b. For likelihood maximization, **A** , **P** and **E** were constrained to be positive semi-definite using the Cholesky method (Marlow, Fisher et al. 2003).

For our first set of simulations, we assumed that $\mathbf{E} = \mathbf{I}_2$ and $\mathbf{P}$ was either $\mathbf{I}_2$ or $\mathbf{0}$. We let the sample sizes be 25, 50, 100 and 250 pedigrees. We simulated 200,000 replicates for each situation. Note that when $\mathbf{P} = \mathbf{0}$, the assumptions used in this paper are violated because the nuisance parameters are not allowed to be on the boundary of the parameter space. Figure 5 shows the empirical type 1 error rates when Theorems 3 and 4 are used to calculate statistical significance. As can be seen, when $\mathbf{P} = \mathbf{I}_2$ the asymptotic distribution nicely approximates the empirical distribution with only around 100 pedigrees. However, when $\mathbf{P} = \mathbf{0}$ the type I error rate is unreasonably conservative. We give some additional comments on this problem in the discussion.

In a second set of simulations we parameterized $\mathbf{P}$ and $\mathbf{E}$ as in equation (18). for differing values of r. We then calculated the p-values using Method 2 in section 3. As stopping rule, for each simulation we tested whether the p-value was statistically different from 0.001 at an alpha level of 0.01. To our knowledge, the simulation algorithm failed to yield a p-value only once, probably due to a failed optimization. Missing values were ignored. For each parameter setup we used 50,000 simulated replicates. Table 2 shows the results. As can be seen, when r comes close to 1, the statistic approaches its asymptotic distribution more slowly. Again we see that using the asymptotic method described in this paper leads to conservative results when the nuisance parameters are near the boundary. We give some additional comments on this problem in the discussion.

Table 2: Type I error rates for various genetic models parameterized as in equation (18) when $\alpha = 0.001$. A value of 1 represents the theoretically correct Type 1 error. Each cell is estimated from 50,000 simulated replicates.

Constraint	Sample Size	Type I Error				
		r=0	r=0.25	r=0.5	r=0.75	r=0.95
1	50	0.32	0.36	0.20	0.20	0.36
	100	0.38	0.36	0.40	0.32	0.20
	250	0.82	0.84	0.50	0.28	0.44
	1000	1.18	1.06	1.18	0.50	0.40
2	50	0.18	0.24	0.44	0.58	0.48
	100	0.52	0.30	0.46	0.24	0.10
	250	0.90	0.66	0.64	0.50	0.44
	1000	1.02	1.06	0.84	0.50	0.50
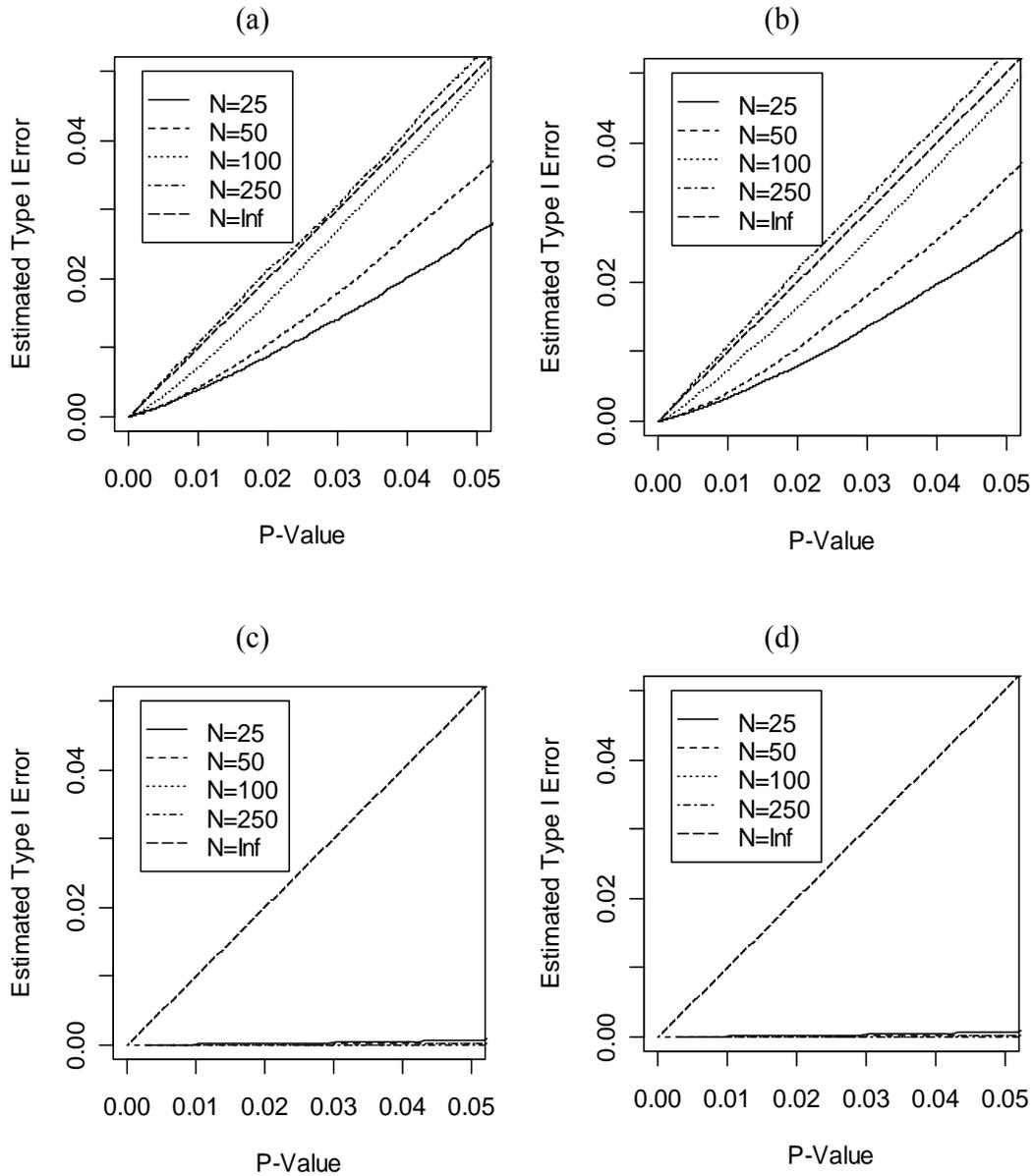
Figure 5: Empirical Type I error for different samples sizes assuming the statistic follows the distribution given in Theorems 3 and 4 for: (a) Constraint 1 and $\mathbf{P} = \mathbf{E} = \mathbf{I}_2$, (b) Constraint 2 and $\mathbf{P} = \mathbf{E} = \mathbf{I}_2$, (c) Constraint 1 and $\mathbf{E} = \mathbf{I}_2$, $\mathbf{P} = \mathbf{0}$ and (d) Constraint 2 and $\mathbf{E} = \mathbf{I}_2$, $\mathbf{P} = \mathbf{0}$.

5. Discussion

There are two possible motivations for a one-sided test. First, we may only be interested in the results of the study if the deviation from the null hypothesis is in a specific direction. In this case we really have a compound null hypothesis such as $H_0 : \mu \leq 0$ vs $H_a : \mu > 0$. Second, we may be convinced that the parameter could fall on only one side of the null hypothesis. It is this second reason that justifies the use of a one-sided test in the variance component setting. We test $H_0 : a = 0$ against $H_a : a > 0$ because we have a strong *a priori* commitment to the notion that a, as a variance, must be positive. In the univariate tests mentioned above, the one-sided p-value will be half the two-sided p-value if the parameter estimate is positive. Hence, one-sided tests cannot be performed after peeking at the data. We must justify the one-sided constraint by substantive knowledge, and not because it produces a lower p-value. Looking at the estimated covariance matrix before choosing a test is not an acceptable statistical procedure. In terms of type I error, any alternative hypothesis is defensible when decided upon *a priori*, although it may result in lower power. It is this last point that we wish to emphasize in the rather confusing set of possible alternative hypotheses.

We must make one additional cautionary note about one-sided tests in linkage analysis. While each of the constraints suggested above seems intuitively reasonable, we cannot say that **A** must always be positive semi-definite. It is only the total variance that is mathematically constrained to be positive semi-definite, and not the individual components of variance. Some bizarre scenarios are remotely possible. For example, suppose that there is a gene for contrarianism (i.e., family members intentionally seek to differentiate themselves). In such a situation, many behavioral traits would have a negative familial correlation induced by the locus for contrarianism. We would not be able to detect the linkage due to such behavioral traits using a one-sided test. There are also examples where environmental factors such as competition could result in negative familial correlations (Verbeke and Molenberghs 2003), but in such cases the one-sided test of linkage may still be valid. We also point out that estimation and testing are distinct issues. Within the context of estimation it may not be advisable to restrict variances to be positive because the accumulation of such results across studies may result in upwardly biased estimates (Lynch and Walsh 1998 pp. 563-564).

In this paper we have shown that it is possible to evaluate asymptotic significance levels using fairly simple methods, even in the absence of analytical results. While others (Silvapulle and Sen 2005, pp. 78-81) have discussed similar methods, our results are unique in that they are both efficient and applicable even to non-convex spaces. Our method may be easily transformed into a numerical integration problem. We have also shown that replacing the information matrix

with a consistent estimate is sometimes reasonable. With regard to variance component linkage analysis, we have shown that analytical formulas for the distribution function may be derived that can be used to obtain a first pass approximation to the statistical significance. However, we recommend that Method 2 above should generally be used and not the analytical formulas.

In this paper we have investigated a multivariate variance component likelihood ratio test that enforces constraints on the nuisance parameters. We do not recommend that this statistic be routinely used in practice for two reasons. First, when nuisance parameters are too near the boundary convergence to the asymptotic distribution may be very slow. A similar observation has been made in the univariate case (Shugart, O'Connell et al. 2002). However, in all cases simulated, the statistic was conservative. *The most natural approach to deal with this problem is to stop enforcing constraints on the nuisance parameters*. This may lead to increased computational difficulties, but it would improve the performance of the methods we have suggested. Second, tests of variance components that make use of the normal likelihood are not necessarily robust to the multivariate normal assumption. This is a well known problem in the univariate case (Allison, Neale et al. 1999). Our simulations assumed that the multivariate normal variance component model was actually correct. In reality, we expect a mixture distribution and many other sources of non-normality. Fortunately, there are many ways to make the variance component framework robust to the normality assumption (Chen, Broman et al. 2005). In future work we plan to address both these concerns by developing a robust score test that does not enforce constraints on the nuisance parameters. The asymptotic approach developed in this paper may also be used with these robust score tests.

Appendix A

Explanation of Equation (7):
Notice first that:

$$\inf_{\theta \in \tilde{C}(\Omega_1, \boldsymbol{\theta}_0)} \left\| \mathbf{V} - \boldsymbol{\theta} \right\|^2 \geq 0 \qquad (19)$$

because distance measures are greater than 0. Also:

$$\inf_{\theta \in \tilde{C}(\Omega_0, \boldsymbol{\theta}_0)} \left\| \mathbf{V} - \boldsymbol{\theta} \right\|^2 \leq \left\| \mathbf{V} - 0 \right\|^2 = 1, \qquad (20)$$

because $\mathbf{0} \in \tilde{C}(\Omega_0, \boldsymbol{\theta}_0)$. Putting (19) and (20) together we obtain:

$$\tilde{g}(\mathbf{V}_i) = \inf_{\theta \in \tilde{C}(\Omega_0, \mathbf{\theta}_0)} \|\mathbf{V} - \mathbf{\theta}\|^2 - \inf_{\theta \in \tilde{C}(\Omega_1, \mathbf{\theta}_0)} \|\mathbf{V} - \mathbf{\theta}\|^2 \tag{21}$$

$$\leq 1.$$

Furthermore, it is clear that $\inf_{\theta \in \tilde{C}(\Omega_1, \mathbf{\theta}_0)} \|\mathbf{V} - \mathbf{\theta}\|^2 \leq \inf_{\theta \in \tilde{C}(\Omega_0, \mathbf{\theta}_0)} \|\mathbf{V} - \mathbf{\theta}\|^2$, under the very general assumption that $\tilde{C}(\Omega_0, \mathbf{\theta}_0) \subset \tilde{C}(\Omega_1, \mathbf{\theta}_0)$. As a result,

$$0 \leq \tilde{g}(\mathbf{V}_i) \leq 1. \tag{22}$$

Now notice that $1 - F_d(S/x)$ is an increasing function because it is a decreasing function composed with a decreasing function. Note also that $1 - F_d(S/0) = 0$ by definition and $1 - F_d(S/1) = 1 - F_d(S)$. Therefore, from (22) and the fact that $1 - F_d(S/x)$ is increasing in x:

$$0 \leq 1 - F_d(S/\tilde{g}(\mathbf{V}_i)) \leq 1 - F_d(S). \tag{23}$$

We note that the variance of each simulated distribution is at its maximum when all its probability is concentrated at either end of the limits described above. That is, its variance is as large as possible when it is distributed as a point mass at 0 and a point mass at p_U. We let ζ be the amount of probability concentrated at p_U. Then $E[\hat{p}_n] = p = \zeta p_U$. So $\zeta = p / p_U$, and

$$Var(\hat{p}_n) \leq \frac{(1-\zeta)(p-0)^2 + \zeta(p-p_U)^2}{n} = \frac{(p_U - p)p}{n}. \tag{24}$$

Explanation of Equation (9):
Consider the cone $\tilde{C}(\Omega_1, \mathbf{\theta}_0)$ that has vertex at $\mathbf{0}$. Because it is nontrivial, there exists some non-zero vector $\mathbf{e} \in \tilde{C}(\Omega_1, \mathbf{\theta}_0)$. We define the cone $\Omega_2 = \{a\mathbf{e} \mid a > 0\}$. Define $\tilde{g}^*(\mathbf{V}) = \|\mathbf{\theta}\|^2 - \inf_{\theta \in \Omega_2} \|\mathbf{V} - \mathbf{\theta}\|^2$. Clearly $\Omega_2 \subseteq \tilde{C}(\Omega_1, \mathbf{\theta}_0)$ and therefore $\inf_{\theta \in \Omega_2} \|\mathbf{V} - \mathbf{\theta}\|^2 \geq \inf_{\theta \in \tilde{C}(\Omega_1, \mathbf{\theta}_0)} \|\mathbf{V} - \mathbf{\theta}\|^2$. Thus

$$\tilde{g}^*(\mathbf{V}) = \|\mathbf{\theta}\|^2 - \inf_{\theta \in \Omega_2} \|\mathbf{V} - \mathbf{\theta}\|^2 \leq \|\mathbf{\theta}\|^2 - \inf_{\theta \in \tilde{C}(\Omega_1, \mathbf{\theta}_0)} \|\mathbf{V} - \mathbf{\theta}\|^2 = \tilde{g}(\mathbf{V}). \tag{25}$$

But, because $1 - F_d\left(S/x\right)$ is increasing in x, $1 - F_d\left(S/\tilde{g}^*(\mathbf{V})\right) \le 1 - F_d\left(S/\tilde{g}(\mathbf{V})\right)$. Thus

$$E_{\mathbf{V}}\left[1 - F_d\left(S/\tilde{g}^*(\mathbf{V})\right)\right] \le E_{\mathbf{V}}\left[1 - F_d\left(S/\tilde{g}(\mathbf{V})\right)\right] = p. \qquad (26)$$

However, the test statistic for $H_0 : \boldsymbol{\theta} = 0$ versus $H_a : \boldsymbol{\theta} \in \Omega_2$ is asymptotically distributed as $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$, and therefore $E_{\mathbf{V}}\left[1 - F_d\left(S/\tilde{g}^*(\mathbf{V})\right)\right] = \frac{1}{2}\left(1 - F_1(S)\right)$. Combining this with (26) we get that $\frac{1}{2}\left(1 - F_1(S)\right) \le p$ and thus $S \ge F_1^{-1}\left(1 - 2p\right)$. We then have

$$p_U = 1 - F_d\left(S\right) \le 1 - F_d\left(F_1^{-1}\left(1 - 2p\right)\right). \qquad (27)$$

Equation (9) then follows by combining (27) with (8).

$$(28)$$

Proof of Lemma 1:
Define $\quad Q_2\left(\boldsymbol{\theta}\right) = \left(\mathbf{Z}_2 - \boldsymbol{\lambda} + \mathbf{f}_{\lambda\lambda}^{-1}\mathbf{f}_{\psi\lambda}^T\left(\mathbf{Z}_1 - \boldsymbol{\psi}\right)\right)^T \mathbf{f}_{\lambda\lambda}\left(\mathbf{Z}_2 - \boldsymbol{\lambda} + \mathbf{f}_{\lambda\lambda}^{-1}\mathbf{f}_{\psi\lambda}^T\left(\mathbf{Z}_1 - \boldsymbol{\psi}\right)\right) \quad$ and let $Q_1\left(\boldsymbol{\psi}\right) = \left(\mathbf{Z}_1 - \boldsymbol{\psi}\right)^T \mathbf{M}\left(\mathbf{Z}_1 - \boldsymbol{\psi}\right)$, where $\mathbf{Z}^T = \begin{bmatrix} \mathbf{Z}_1^T & \mathbf{Z}_2^T \end{bmatrix}^T$. It is easily verified that $\left(\mathbf{Z} - \boldsymbol{\theta}\right)^T \mathbf{f}\left(\mathbf{Z} - \boldsymbol{\theta}\right) = Q_1\left(\boldsymbol{\psi}\right) + Q_2\left(\boldsymbol{\theta}\right)$. Thus we see that

$$\inf_{\boldsymbol{\psi} \in C_i; \boldsymbol{\lambda} \in C_*} \left(\mathbf{Z} - \boldsymbol{\theta}\right)^T \mathbf{f}\left(\mathbf{Z} - \boldsymbol{\theta}\right) = \inf_{\boldsymbol{\psi} \in C_i}\left[Q_1\left(\boldsymbol{\psi}\right) + \inf_{\boldsymbol{\lambda} \in C_*} Q_2\left(\boldsymbol{\theta}\right)\right], \qquad (29)$$

where $C_i = C\left(\Omega_i, \boldsymbol{\theta}_0\right)$. Under condition (a), $Q_2\left(\boldsymbol{\theta}\right)$ is only a function of $\boldsymbol{\lambda}$, so $\inf_{\boldsymbol{\psi} \in C_i, \boldsymbol{\lambda} \in C_*} \left(\mathbf{Z} - \boldsymbol{\theta}\right)^T \mathbf{f}\left(\mathbf{Z} - \boldsymbol{\theta}\right) = \inf_{\boldsymbol{\psi} \in C_i} Q_1\left(\boldsymbol{\psi}\right) + \inf_{\boldsymbol{\lambda} \in C_*} Q_2\left(\boldsymbol{\theta}\right)$. Under condition (b) $\inf_{\boldsymbol{\psi} \in C_i, \boldsymbol{\lambda} \in C_*} \left(\mathbf{Z} - \boldsymbol{\theta}\right)^T \mathbf{f}\left(\mathbf{Z} - \boldsymbol{\theta}\right) = \inf_{\boldsymbol{\psi} \in C_i} Q_1\left(\boldsymbol{\psi}\right)$. Thus under either condition (a) or (b) $h\left(\mathbf{Z}_1\right) = g\left(\mathbf{Z}\right)$. It is a well known property of the Schur complement ($\mathbf{M}$) of the block $\mathbf{f}_{\psi\psi}$ of $\mathbf{f}$ that $\mathbf{M}^{-1}$ is the first block entry of $\hat{\mathbf{f}}^{-1}$. Thus the distribution of the first q entries of $\mathbf{f}^{-1/2}\tilde{\mathbf{Z}} + \boldsymbol{\theta}_0$, where $\tilde{\mathbf{Z}}$ is a vector of independent standard normal random variables, is identical in distribution to that of $\mathbf{M}^{-1/2}\tilde{\mathbf{Y}} + \mathbf{b}$ where $\tilde{\mathbf{Y}}$ is a vector of independent standard normal random variables. Therefore

$$P_{\mathbf{Y}}\left[h(\mathbf{Y}) \geq S\right] = P_{\tilde{\mathbf{Y}}}\left[h\left(\mathbf{M}^{-1/2}\tilde{\mathbf{Y}} + \mathbf{b}\right) \geq S\right]$$

$$= P_{\tilde{\mathbf{Z}}}\left[g\left(\mathbf{f}^{-1/2}\tilde{\mathbf{Z}} + \boldsymbol{\theta}_0\right) \geq S\right]. \tag{30}$$

$$= P_{\mathbf{Z}}\left[g(\mathbf{Z}) \geq S\right].$$

Proof of Theorem 2:

First we prove that $\tilde{h}_n\left(\tilde{\mathbf{Y}}\right) = h\left(\hat{\mathbf{M}}_n^{-1/2}\tilde{\mathbf{Y}} + \mathbf{b}\right)$ is a continuous function of $\hat{\mathbf{f}}_n$. Note that $\hat{\mathbf{M}}_n$ is a continuous function of $\hat{\mathbf{f}}_n$ because the matrix inverse is a continuous mapping (Lange 2004, p. 31). Also, suppose $\hat{\mathbf{M}}_n^{1/2}$ is the Cholesky decomposition, which is also a continuous mapping of $\hat{\mathbf{M}}_n$ (Schatzman and Taylor 2002, p. 295). It can be easily shown that $\inf_{\boldsymbol{\Psi} \in \Omega} \|\mathbf{x} - \mathbf{A}\boldsymbol{\Psi}\|^2$ is a continuous function of $\mathbf{x}$ and $\mathbf{A}$. So

$$h\left(\hat{\mathbf{M}}_n^{-1/2}\tilde{\mathbf{Y}} + \mathbf{b}\right) = \inf_{\boldsymbol{\Psi} \in C_b}\left\|\tilde{\mathbf{Y}} - \hat{\mathbf{M}}_n^{1/2}\left(\boldsymbol{\Psi} - \mathbf{b}\right)\right\|^2 - \inf_{\boldsymbol{\Psi} \in C_a}\left\|\tilde{\mathbf{Y}} - \hat{\mathbf{M}}_n^{1/2}\left(\boldsymbol{\Psi} - \mathbf{b}\right)\right\|^2 \tag{31}$$

is a continuous function of $\hat{\mathbf{f}}_n$.

Let $\beta(\mathbf{B}, l) = P_{\tilde{\mathbf{Y}}}\left[\tilde{h}_n\left(\tilde{\mathbf{Y}}\right) \geq LR_n \middle| \hat{\mathbf{f}}_n = \mathbf{B}, LR_n = l\right]$. By inspection of equation (11), we have $\beta(\mathbf{B}, l) = \int_{\varphi}\left[1 - F_p\left(l / h\left(\mathbf{M}(\mathbf{B})^{-1/2}\tilde{\mathbf{Y}} + \mathbf{b}\right)\right)\right] J d\varphi / S_p$. Let $(\mathbf{B}_i, l_i) \to (\mathbf{B}, l)$ be a convergent sequence with $l > 0$ and $\mathbf{B} > 0$. Note that $0 \leq F_p(x) \leq 1$, so, by the dominated convergence theorem, $\beta(\mathbf{B}_i, l_i) \to \beta(\mathbf{B}, l)$ and β is thus continuous for $l > 0$ and $\mathbf{B} > 0$. A discontinuity at $l = 0$ occurs because by our definition $F_p(S/0) = 1$ for $S > 0$ and $F_p(0/0) = 0$. To fix this discontinuity define:

$$\gamma(\mathbf{B}, l) = \begin{cases} P_{\tilde{\mathbf{Y}}}\left[\tilde{h}_n\left(\tilde{\mathbf{Y}}\right) > 0 \middle| \hat{\mathbf{f}}_n = \mathbf{B}, LR_n = l\right] & l \leq 0 \\ \beta(\mathbf{B}, l) & l > 0 \end{cases}. \tag{32}$$

But $\gamma(\mathbf{B}, l)$ is fairly clearly a continuous function for $\mathbf{B} > 0$ and $l \in \mathbb{R}$. Also $LR_n \xrightarrow{L} \tilde{h}(\tilde{\mathbf{Y}})$ and $\hat{\mathbf{f}}_n \xrightarrow{P} \mathbf{f}(\boldsymbol{\theta}_0)$. From Slutsky's Theorem,

$\gamma\left(\hat{\mathbf{f}}_n, LR_n\right) \xrightarrow{\quad L \quad} \gamma\left(\mathbf{f}(\boldsymbol{\theta}_0), \tilde{h}(\tilde{\mathbf{Y}})\right)$. Theorem 2 follows from the definition of convergence in distribution.

Appendix B

Discussion of Lemma 2:

For sufficiently large η and sufficiently small ε the true parameter will be an interior point of $\Gamma(\varepsilon, \eta)$ and hence this constraint will asymptotically disappear. For $i \in \{0, 1\}$, $\Omega_i \cap \Gamma(\varepsilon, \eta)$ is a compact space with bounded derivatives for the likelihood up to the 3$^\text{rd}$ derivative. As mentioned earlier, our problem fits into the independent identically distributed framework if we consider pedigree structure and IBD values to be random. Geyer (1994) gives four regularity conditions (excluding the Chernoff regularity of the cone, which is clearly satisfied in our case) for establishing theorem 1. See Theorem 4.4 and the remark at the end of section 4 in Geyer (1994). The first regularity condition is the existence of a local quadratic approximation of the likelihood. This appears to be obviously satisfied in our case. The second condition is termed stochastic equicontinuity. This condition is more difficult to verify, but it appears that an application of the Equicontinuity Lemma on p. 150 of Pollard (1984) could be used. The third regularity condition is the existence of a central limit theorem for the first derivative, and this again seems to be fairly obviously satisfied. The final condition is the use of a consistent root-K minimizer. Consistency may be established for example by Theorem 17 of Ferguson (1996), and the root-K minimizer is guaranteed by how we have defined the likelihood ratio.

Let $\quad vec\mathbf{A}_{t \times t} = \begin{bmatrix} a_{11} & \cdots & a_{1t} & a_{21} & \cdots & a_{tt} \end{bmatrix}$ and $vech\mathbf{A}_{t \times t} = \begin{bmatrix} a_{11} & \cdots & a_{1t} & a_{22} & \cdots & a_{tt} \end{bmatrix}$ as defined for example in (Henderson and Searle 1979). There exists matrices $\mathbf{G}$ and $\mathbf{H}$ such that $vec\mathbf{A} = \mathbf{G}vech\mathbf{A}$ and $vech\mathbf{A} = \mathbf{H}vec\mathbf{A}$, again defined for example in Henderson and Searle (1979).

Lemma 3: If $\mathbf{A} = 0$ and $\mathbf{P} = \gamma\mathbf{E}$ for some positive scalar γ, then $\mathbf{M} = \kappa\left[\mathbf{G}^T\left(\mathbf{E}^{-1} \otimes \mathbf{E}^{-1}\right)\mathbf{G}\right]$, where $\mathbf{M}$ is the Schur complement of the block $\mathbf{f}_{\psi\psi}$ of $\mathbf{f}$ and κ is some constant dependent upon the population.

Proof of Lemma 3:

Let τ represent a random pedigree structure sampled from a population. The covariance structure for a given τ is $\mathbf{C}_\tau = \boldsymbol{\Phi}_\tau \otimes \mathbf{P} + \mathbf{I}_\tau \otimes \mathbf{E} = \left(\gamma\boldsymbol{\Phi}_\tau + \mathbf{I}_\tau\right) \otimes \mathbf{E}$. Let

$\boldsymbol{\Delta}_\tau = \boldsymbol{\Pi}_\tau - \boldsymbol{\Phi}_\tau$ and $\mathbf{B}_{ij} = \mathbf{e}_i \mathbf{e}_j^T + \mathbf{e}_j \mathbf{e}_i^T \, If \left[j \neq i \right]$, where $\mathbf{e}_i$ is the unit vector of length t in the ith direction. Then:

$$
\begin{aligned}
\mathbf{f}_{a_{i_1 i_2} a_{j_1 j_2}} &= E_\tau E_{\boldsymbol{\Pi}_\tau} tr \left[\mathbf{C}_\tau^{-1} \left(\boldsymbol{\Delta}_\tau \otimes \mathbf{B}_{i_1 i_2} \right) \mathbf{C}_\tau^{-1} \left(\boldsymbol{\Delta}_\tau \otimes \mathbf{B}_{j_1 j_2} \right) \right] \\
&= E_\tau E_{\boldsymbol{\Pi}_\tau} tr \left[\left(\gamma \boldsymbol{\Phi}_\tau + \mathbf{I}_\tau \right)^{-1} \boldsymbol{\Delta}_\tau \left(\gamma \boldsymbol{\Phi}_\tau + \mathbf{I}_\tau \right)^{-1} \boldsymbol{\Delta}_\tau \otimes \mathbf{E}^{-1} \mathbf{B}_{i_1 i_2} \mathbf{E}^{-1} \mathbf{B}_{j_1 j_2} \right] \\
&= tr \left[\mathbf{E}^{-1} \mathbf{B}_{i_1 i_2} \mathbf{E}^{-1} \mathbf{B}_{j_1 j_2} \right] \kappa \\
&= \kappa \left(vec \mathbf{B}_{i_1 i_2} \right)^T \left(\mathbf{E}^{-1} \otimes \mathbf{E}^{-1} \right) vec \mathbf{B}_{i_1 i_2},
\end{aligned}
\tag{33}
$$

where $\kappa = E_\tau E_{\boldsymbol{\Pi}_\tau} tr \left[\left(\gamma \boldsymbol{\Phi}_\tau + \mathbf{I}_\tau \right)^{-1} \boldsymbol{\Delta}_\tau \left(\gamma \boldsymbol{\Phi}_\tau + \mathbf{I}_\tau \right)^{-1} \boldsymbol{\Delta}_\tau \right]$. Line 4 above follows from line 9 of Henderson and Searle (1979). Applying the same algebraic steps and using the fact that $E_{\boldsymbol{\Pi}_\tau} \boldsymbol{\Delta}_\tau = 0$, we obtain $\mathbf{f}_{a_{i_1 i_2} p_{j_1 j_2}} = 0$ and $\mathbf{f}_{a_{i_1 i_2} e_{j_1 j_2}} = 0$. Note that $\mathbf{G} = \left[vec \mathbf{B}_{11} \quad \cdots \quad vec \mathbf{B}_{1t} \quad vec \mathbf{B}_{22} \quad \cdots \quad vec \mathbf{B}_{tt} \right]$. Lemma 3 follows immediately.

> *Lemma 4:* If $\mathbf{M}$ can be factored as $\left[\mathbf{G}^T \left(\mathbf{T}^{-1} \otimes \mathbf{T}^{-1} \right) \mathbf{G} \right]$ and $\mathbf{P}$ and $\mathbf{E}$ are not on the boundary of the parameter space, then the asymptotic null distribution of the variance component CLRT statistic (λ) for $H_0 : \mathbf{A} = 0$ vs. $H_a : \mathbf{A} \geq 0$ is equivalent to the distribution of $\lambda^* = \sum_{i=1}^{t} \left[\max \left(b_i, 0 \right) \right]^2$. Here b_i represents the eigenvalues of a symmetric random matrix $\tilde{\mathbf{Z}}_{t \times t}$, and the elements $\tilde{\mathbf{Z}}$ are independent with distribution $N(0,1)$ on the diagonals and $N\left(0, \tfrac{1}{2}\right)$ on the off diagonals.

Proof of Lemma 4:
Let $\mathbf{X}$ and $\boldsymbol{\theta}$ be $t \times t$ symmetric matrices with $\mathbf{Z} = vech \mathbf{X}$ distributed as a MVN random variable with mean 0 and variance $\mathbf{M}^{-1}$. Then

$$vech(\mathbf{X}-\boldsymbol{\theta})^{T}\mathbf{M}vech(\mathbf{X}-\boldsymbol{\theta})=vec(\mathbf{X}-\boldsymbol{\theta})^{T}\left(\mathbf{T}^{-1}\otimes\mathbf{T}^{-1}\right)vec(\mathbf{X}-\boldsymbol{\theta})$$

$$=vec(\mathbf{X}-\boldsymbol{\theta})^{T}vec\left[\mathbf{T}^{-1}(\mathbf{X}-\boldsymbol{\theta})\mathbf{T}^{-1}\right] \quad (34)$$

$$=tr\left[\left(\mathbf{T}^{-1/2}\mathbf{X}\mathbf{T}^{-1/2}-\mathbf{T}^{-1/2}\boldsymbol{\theta}\mathbf{T}^{-1/2}\right)^{2}\right].$$

Let $\tilde{\mathbf{X}}=\mathbf{T}^{-1/2}\mathbf{X}\mathbf{T}^{-1/2}$. Note that $\mathbf{M}^{-1}=\mathbf{H}(\mathbf{T}\otimes\mathbf{T})\mathbf{H}^{T}$. This may be easily verified using line (24) and (20) of Henderson and Searle (1979). Note also that $vech\tilde{\mathbf{X}}=\mathbf{H}\left(\mathbf{T}^{-1/2}\otimes\mathbf{T}^{-1/2}\right)\mathbf{G}vech\mathbf{X}$. Hence $vech\tilde{\mathbf{X}}$ is distributed as MVN with mean 0 and variance

$$\mathrm{var}\left(vech\tilde{\mathbf{X}}\right)=\mathbf{H}\left(\mathbf{T}^{-1/2}\otimes\mathbf{T}^{-1/2}\right)\mathbf{G}\mathbf{H}(\mathbf{T}\otimes\mathbf{T})\mathbf{H}^{T}\mathbf{G}^{T}\left(\mathbf{T}^{-1/2}\otimes\mathbf{T}^{-1/2}\right)\mathbf{H}^{T}$$

$$=\mathbf{H}\mathbf{H}^{T} \qquad . \; (35)$$

But $\mathbf{H}\mathbf{H}^{T}$ is a diagonal matrix with 1 corresponding to the diagonal elements of $\tilde{\mathbf{X}}$ and ½ for the off diagonal elements. $\tilde{S}=\left\{\tilde{\boldsymbol{\theta}}:\tilde{\boldsymbol{\theta}}=\mathbf{T}^{-1/2}\boldsymbol{\theta}\mathbf{T}^{-1/2}\;for\;some\;\boldsymbol{\theta}\geq0\right\}=\left\{\tilde{\boldsymbol{\theta}}:\tilde{\boldsymbol{\theta}}\geq0\right\}$. From Theorem 1 and 2, λ has the same asymptotic distribution as:

$$g(\mathbf{Z})=vech(\mathbf{X})^{T}\mathbf{M}vech(\mathbf{X})-\inf_{\theta\geq0}vech(\mathbf{X}-\boldsymbol{\theta})^{T}\mathbf{M}vech(\mathbf{X}-\boldsymbol{\theta})$$

$$=tr\left[\tilde{\mathbf{X}}^{2}\right]-\inf_{\tilde{\theta}\in\tilde{S}}tr\left[\left(\tilde{\mathbf{X}}-\tilde{\boldsymbol{\theta}}\right)^{2}\right]. \qquad (36)$$

Lemma 4 then follows from Theorem 8 in Chapter 2 of Dimedenco (2005). Note that Lemmas 3 and 4 show the asymptotic distribution to be identical to the traditional multivariate variance component model in one special case.

Lemma 5: If $\mathbf{M}$ can be factored as $\left[\mathbf{G}^{T}\left(\mathbf{T}^{-1}\otimes\mathbf{T}^{-1}\right)\mathbf{G}\right]$, where $\mathbf{T}>0$, then the asymptotic null distribution of the CLRT statistic (λ) for a variance component linkage model, testing $H_{0}:\mathbf{A}=0$ vs. $H_{a}:\mathbf{A}\geq0$, is equivalent to the distribution of: $\lambda^{*}=\left[\max(b_{1},0)\right]^{2}$. Here $b_{1}\geq\ldots\geq b_{t}$ represent the eigenvalues of a symmetric random matrix $\tilde{\mathbf{X}}_{t\times t}$, and the elements $\tilde{\mathbf{Z}}$ are independent with distribution $N(0,1)$ on the diagonals and $N\left(0,\frac{1}{2}\right)$ on the off-diagonals.

Proof of Lemma 5:

Define $\tilde{S} = \left\{ \mathbf{T}^{-1/2}\boldsymbol{\varphi}\boldsymbol{\varphi}^T\mathbf{T}^{-1/2} \mid \boldsymbol{\varphi} \in \mathbb{R}^t \right\} = \left\{ \boldsymbol{\varphi}\boldsymbol{\varphi}^T \mid \boldsymbol{\varphi} \in \mathbb{R}^t \right\}$. Following the same steps as in Lemma 3, we have:

$$\tilde{g}\left(\tilde{\mathbf{X}}\right) = tr\tilde{\mathbf{X}}^2 - \inf_{\tilde{\theta}\in\tilde{S}} tr\left[\left(\tilde{\mathbf{X}} - \tilde{\boldsymbol{\theta}}\right)^2\right] = \sup_{\boldsymbol{\varphi}\in\mathbb{R}^t}\left(tr\left(\tilde{\mathbf{X}}\boldsymbol{\varphi}\boldsymbol{\varphi}^T\right) - tr\left[\boldsymbol{\varphi}\boldsymbol{\varphi}^T\boldsymbol{\varphi}\boldsymbol{\varphi}^T\right]\right)$$

$$= \sup_{\boldsymbol{\varphi}\in\mathbb{R}^t}\left(\boldsymbol{\varphi}^T\tilde{\mathbf{X}}\boldsymbol{\varphi} - \left(\boldsymbol{\varphi}^T\boldsymbol{\varphi}\right)^2\right) = \sup_{\|\mathbf{V}\|=1}\sup_{r}\left(r^2\mathbf{V}^T\tilde{\mathbf{X}}\mathbf{V} - r^4\|\mathbf{V}\|^4\right) \qquad .$$

$$= \sup_{\|\mathbf{V}\|=1}\left(\frac{\mathbf{V}^T\tilde{\mathbf{X}}\mathbf{V}}{\mathbf{V}^T\mathbf{V}}I\left[\mathbf{V}^T\tilde{\mathbf{X}}\mathbf{V} > 0\right]\right)^2 .$$

Lemma 5 follows from this.

Sketch of Proof for Theorem 4:

By using Lemmas 3 and 5 we may conclude that the asymptotic distribution is a function of the maximum eigenvalue of a random matrix. The eigenvalue distribution has been given in numerous places, such as Amemiya, Anderson et al. (1990). The distribution function for $t = 2$ is: $\left(b_1 - b_2\right)\exp\left[-\tfrac{1}{2}\left(b_1^2 + b_2^2\right)\right]/2\sqrt{\pi}$.

It is a fairly straightforward application of calculus to derive the distribution in Theorem 4 using a symbolic computational system such as Mathematica.

References

Allison, D. B., M. C. Neale, et al. (1999). "Testing the Robustness of the Likelihood-Ratio Test in a Variance-Component Quantitative-Trait Loci–Mapping Procedure." <u>The American Journal of Human Genetics</u> **65**(2): 531-544.

Amemiya, Y., T. W. Anderson, et al. (1990). "Percentage points for a test of rank in multivariate components of variance." <u>Biometrika</u> **77**: 637-641.

Amos, C., M. de Andrade, et al. (2001). "Comparison of multivariate tests for genetic linkage." <u>Hum Hered</u> **51**(3): 133-44.

Blumenson, L. E. (1960). "A Derivation of n-Dimensional Spherical Coordinates." <u>The American Mathematical Monthly</u> **67**(1): 63-66.

Chen, W. M., K. W. Broman, et al. (2005). "Power and robustness of linkage tests for quantitative traits in general pedigrees." <u>Genetic Epidemiology</u> **28**(1): 11-23.

Chernoff, H. (1954). "On the Distribution of the Likelihood Ratio." <u>The Annals of Mathematical Statistics</u> **25**(3): 573-578.

Delmas, C. and J. L. Foulley (2007). "On testing a class of restricted hypotheses." <u>Journal of Statistical Planning and Inference</u> **137**(4): 1343-1361.

Demidenko, E. (2005). <u>Mixed Models: Theory and Applications</u>: Wiley-Interscience.

Evans, D. M. (2002). "The power of multivariate quantitative-trait loci linkage analysis is influenced by the correlation between variables." <u>American journal of human genetics</u> **70**(6): 1599-1602.

Evans, D. M., G. Zhu, et al. (2004). "Multivariate QTL linkage analysis suggests a QTL for platelet count on chromosome 19q." <u>European Journal of Human Genetics</u> **12**: 835-842.

Ferguson, T. S. (1996). <u>A Course in Large Sample Theory</u>: Chapman & Hall/CRC.

Geyer, C. J. (1994). "On the Asymptotics of Constrained M-Estimation." <u>The Annals of Statistics</u> **22**(4): 1993-2010.

Ginsburg, E., R. C. Elston, et al. (2006). <u>Theoretical aspects of pedigree analysis</u>: Ramot-tel Aviv University.

Han, S. S. and J. T. Chang (2008). "Reconsidering the null distribution of likelihood ratio tests for genetic linkage in multivariate variance components models." <u>eprint arXiv: 0808.2000</u>.

Haseman, J. K. and R. C. Elston (1972). "The investigation of linkage between a quantitative trait and a marker locus." <u>Behavior Genetics</u> **2**(1): 3-19.

Henderson, H. V. and S. R. Searle (1979). "Vec and vech operators for matrices, with some uses in Jacobians and multivariate statistics." <u>Canadian Journal of Statistics</u> **7**: 65-81.

Kuriki, S. (1993). "One-sided test for the equality of two covariance matrices." <u>Annals of Statistics</u> **21**: 1379-1379.

Lange, K. (2004). <u>Optimization</u>. New York: Springer.

Lynch, M. and B. Walsh (1998). <u>Genetics and Analysis of Quantitative Traits</u>: Sinauer Associates, Sunderland, MA.

Marlow, A. J., S. E. Fisher, et al. (2003). "Use of Multivariate Linkage Analysis for Dissection of a Complex Cognitive Trait." The American Journal of Human Genetics **72**(3): 561-570.

Pollard, D. (1984). Convergence of stochastic processes: Springer.

Schatzman, M. and J. Taylor (2002). Numerical Anlysis: A Mathematicl Introduction. New York: Oxford University Press.

Self, S. G. and K. Y. Liang (1987). "Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests Under Nonstandard Conditions." Journal of the American Statistical Association **82**(398): 605-610.

Shaw, F. H. and C. J. Geyer (1997). "Estimation and testing in constrained covariance component models." Biometrika **84**(1): 95-102.

Shugart, Y. Y., J. R. O'Connell, et al. (2002). "An evaluation of the variance components approach: type I error, power and size of the estimated effect." European journal of human genetics: EJHG **10**(2): 133.

Silvapulle, M. and P. Sen (2005). Constrained Statistical Inference: Inequality, Order, and Shape Restrictions. Hoboken, New Jersey: Wiley-Interscience.

Stram, D. O. and J. W. Lee (1994). "Variance Components Testing in the Longitudinal Mixed Effects Model." Biometrics **50**(4): 1171-1177.

Stram, D. O. and J. W. Lee (1995). "Correction to: variance components testing in the longitudinal mixed effects model." Biometrics **51**(1): 196.

Takemura, A. and S. Kuriki (1997). "Weights of Chi-Bar Squared Distribution for Smooth or Piecewise Smooth Cone Alternatives." The Annals of Statistics **25**(6): 2368-2387.

Todorov, A. A., G. P. Vogler, et al. (1998). "Testing causal hypotheses in multivariate linkage analysis of quantitative traits: general formulation and application to sibpair data." Genet Epidemiol **15**(3): 263-78.

Verbeke, G. and G. Molenberghs (2003). "The Use of Score Tests for Inference on Variance Components." Biometrics **59**(2): 254-262.

Vu, H. T. V. and S. Zhou (1997). "Generalization of Likelihood Ratio Tests under Nonstandard Conditions." The Annals of Statistics **25**(2): 897-916.

Wang, K. (2003). "Mapping Quantitative Trait Loci Using Multiple Phenotypes in General Pedigrees." <u>Hum Hered</u> **55**(1): 1-15.

Wolak, F. A. (1991). "The local nature of hypothesis tests involving inequality constraints in nonlinear models." <u>Econometrica</u> **59**(4): 981-995.