



Mathematical Statistics
Stockholm University

Asymptotics of LR tests at Boundary
of Parameter Space Under
Non-Identifiability

Ola Hössjer

Research Report 2007:3

ISSN 1650-0377

Postal address:

Mathematical Statistics
Dept. of Mathematics
Stockholm University
SE-106 91 Stockholm
Sweden

Internet:

<http://www.math.su.se/matstat>



Asymptotics of LR tests at Boundary of Parameter Space Under Non-Identifiability

Ola Hössjer

Abstract

In this paper we provide a general theory for likelihood ratio tests, when, under the null hypothesis, i) a q -dimensional subset of the parameter is non-identifiable and ii) a p -dimensional subset of the parameter is at the boundary of the parameter space. The asymptotic null distribution of twice the log likelihood ratio statistic is the supremum of a so called $\bar{\chi}^2$ -process $Y : \mathbb{T} \rightarrow \mathbb{R}^p$, where the index set $\mathbb{T} \subset \mathbb{R}^q$ corresponds to the unidentifiable parameter. The limit distribution involves a Gaussian process $Z : \mathbb{T} \rightarrow \mathbb{R}^p$ with standard normal marginals, so that, for each $t \in \mathbb{T}$, $Y(t)$ is the squared norm of the projection of $Z(t)$ onto a (possibly t -varying) cone in \mathbb{R}^p . This cone is derived from the local shape of the boundary region around the null parameter in ii). The marginal distribution of $Y(t)$ is $\bar{\chi}^2$, i.e. a mixture of χ^2 -distributions with $p-r, \dots, p$ degrees of freedoms, where r is the number of boundary conditions imposed in ii).

We illustrate the theory with several examples, including LR-tests for components of normal mixtures and LR-tests for the presence of disease genes along a chromosome in genetic linkage analysis (MOD scores and MLS scores). The examples involve a wide range of different $\bar{\chi}^2$ -processes, with varying dimensionality (q, p, r) .

KEY WORDS: Asymptotic distribution, boundary of parameter space, likelihood ratio tests, linkage analysis, non-identifiability, normal mixtures, $\bar{\chi}^2$ -process.

1 Introduction

Consider a probability space (Ω, \mathcal{F}, P) and J independent sequences $\{X_{ji}\}_{i=1}^{\infty}$ of independent and identically distributed random variables $\Omega \rightarrow \mathcal{X}_j$. Let $f_j(x; \theta)$ be the common density of all $\{X_{ji}\}_{i=1}^{\infty}$ with respect to a measure μ_j on $(\mathcal{X}_j, \mathcal{B}(\mathcal{X}_j))$, where $\mathcal{B}(\mathcal{X}_j)$ is the Borel σ -algebra on \mathcal{X}_j and $\theta \in \Theta \subset \mathbb{R}^{q+p}$ an unknown parameter. We are interested in testing

$$\begin{aligned} H_0 &: \theta \in \Theta_0, \\ H_1 &: \theta \notin \Theta_0 \end{aligned} \tag{1}$$

for some subset Θ_0 of the parameter space Θ based on subsample sizes n_j and total sample size $n = n_1 + \dots + n_J$. To this end, define the log likelihood function

$$l_n(\theta) = \sum_{j=1}^J \sum_{i=1}^{n_j} \log f_j(x_{ji}; \theta), \tag{2}$$

where x_{ji} is the observed value of X_{ji} , and the log likelihood ratio statistic

$$\lambda_n = 2 \left(\sup_{\theta \in \Theta} l_n(\theta) - \sup_{\theta \in \Theta_0} l_n(\theta) \right). \tag{3}$$

The null hypothesis H_0 is rejected when λ_n exceeds a given threshold. For standard scenarios, it is well known that λ_n converges in distribution to a $\chi^2(p)$ -distribution, where p is the difference in dimensions between Θ and Θ_0 . This requires certain smoothness assumptions on l_n and that the true parameter $\theta_0 \in \Theta_0$ is an inner point of the parameter space, see e.g. Cramér (1946), Chernoff (1954) and Serfling (1980).

When θ_0 is a boundary point of Θ , standard asymptotic fails. The asymptotic distribution of λ_n then depends on the exact form of Θ in local neighborhoods of θ_0 . If local neighborhood are cone shaped, the asymptotic distribution of λ_n is a finite mixture of χ^2 -distributions of various degrees of freedoms, see for instance Chernoff (1954), Self and Liang (1987), Shapiro (1988), Sen and Silvapulle (2002, 2005) and references therein. Such mixtures are frequently referred to as having a $\bar{\chi}^2$ distribution.

Another complication arises when θ is non-identifiable under the null hypothesis, i.e.

$$f_j(\cdot; \theta) = f_{j0}(\cdot) \text{ for all } \theta \in \Theta_0 \tag{4}$$

and some null densities f_{j0} , $j = 1, \dots, J$. Assuming only the first q components of θ are non-identifiable in (4), we write $\theta = (t, \xi)$, where $t \in \mathbb{T} = [T_{11}, T_{12}] \times \dots \times [T_{q1}, T_{q2}] \subset \mathbb{R}^q$ and $\xi \in \Xi(t) = \{\xi; (t, \xi) \in \Theta\} \subset \mathbb{R}^p$. In

particular, $\Theta = \mathbb{T} \times \Xi$ when $\Xi(t) \equiv \Xi$ for all t . We formulate the hypothesis testing problem (1) as

$$\begin{aligned} H_0 &: \xi = \xi_0, \\ H_1 &: \xi \neq \xi_0, \end{aligned} \tag{5}$$

for some $\xi_0 \in \cap_{t \in \mathbb{T}} \Xi(t)$, so that $\Theta_0 = \mathbb{T} \times \{\xi_0\}$. In view of (4), we thus require that the null hypothesis is simple.

Davies (1977) approximates the asymptotic distribution of λ_n when $q = p = 1$ by the supremum of a Gaussian process, which is truncated or not depending on whether ξ_0 is a boundary point of Ξ . The former case happens when testing the number of components in a mixture distribution with unknown mean, see e.g. Bickel and Chernoff (1993), Lemdani and Pons (1999), Chen and Chen (2001), Garel (2001), Delmas (2003), Liu and Shao (2004) and Azaïs et al. (2004)

The purpose of this paper is to consider the asymptotic distribution of λ_n for a general class of models, where θ_0 is both non-identifiable and at the boundary of the parameter space for general q and p . We prove that under H_0 the asymptotic distribution of λ_n is the supremum of a so called $\bar{\chi}^2$ -process Y over \mathbb{T} , i.e.

$$\lambda_n \xrightarrow{\mathcal{L}} \lambda = \sup_{t \in \mathbb{T}} Y(t). \tag{6}$$

The process Y is constructed so that the marginal distribution of $Y(t)$ is $\bar{\chi}^2$ for each $t \in \mathbb{T}$.

The asymptotic dimensionality of the testing problem can be summarized by the triplet

$$(q, p, r),$$

where q and p are the number of unidentifiable and identifiable parameters under H_0 , as described above, and r is the number of boundary restrictions imposed on the identifiable parameters. It turns out that $Y(t)$ for each $t \in \mathbb{T}$ is a mixture of $r + 1$ distinct χ^2 -distributions with $p - r, \dots, p$ degrees of freedom.

If $\alpha_n(c) = P_{\theta_0}(\lambda_n \geq c)$ is the significance level of the LR-test with threshold $c \in \mathbb{R}$ and $\alpha(c) = P_{\theta_0}(\lambda \geq c)$ the corresponding asymptotic quantity, (6) implies that

$$\alpha_n(c) \rightarrow \alpha(c) \text{ as } n \rightarrow \infty$$

at all continuity points c of the distribution function F_λ of λ .

The paper is organized as follows: In Section 2 we define the $\bar{\chi}^2$ -process Y and outline the main ideas behind (6), the proof of which is given, in two versions, in Sections 3-4. In Sections 5-6 we illustrate the theory with

LR-tests for i) the number of components of a normal mixture and ii) the presence of a disease gene along a chromosome using genetic linkage analysis. In the latter case we derive the limit distribution of MOD scores (Risch, 1984, Clerget-Darpoux et al., 1986) and MLS scores (Risch, 1990). A discussion of the results is given in Section 7 and proofs and technical results are gathered in the appendix.

Throughout the paper, we let, for any positive integer s , $|\cdot|$ denote the Euclidean norm in \mathbb{R}^s . $\|\cdot\|$ the supremum norm of functions $\mathbb{T} \rightarrow \mathbb{R}^s$, and $\|A\|_1 = \sum_{kl} |A_{kl}|$ the L_1 norm of $s \times s$ matrices $A = (A_{kl})$.

2 Quadratic Approximation and $\bar{\chi}^2$ -Process

We will assume that each $f_j(x; \theta)$ is smooth in ξ but not necessarily in t . Define the $1 \times p$ vector-valued score function

$$\psi_j(x; \theta) = \frac{\partial \log f_j(x; \theta)}{\partial \xi},$$

put $\psi_j(x; t) = \psi_j(x; t, \xi_0)$ and introduce the $p \times p$ -matrix

$$I_j(t, t') = E_{\theta_0} \left(\psi_j(X_{j1}; t)^T \psi_j(X_{j1}; t') \right), \quad (7)$$

where ψ^T is the conjugate of ψ and $\theta_0 \in \Theta_0$. Notice that (7) is well defined, since, in view of (4), we need not specify θ_0 . We assume that $I_j(t, t')$ exists for all $t, t' \in \mathbb{T}$. When $t = t'$, $I_j(t) = I_j(t, t)$ is the Fisher information matrix of the j^{th} subsample for ξ at $\theta = (t, \xi_0)$. For the whole sample, define

$$I_n(t, t') = \sum_{j=1}^J (n_j/n) I_j(t, t'). \quad (8)$$

As $n \rightarrow \infty$, we assume that the subsample sizes converge to fixed proportions

$$n_j/n \rightarrow \pi_j, \quad (9)$$

where π_1, \dots, π_J are non-negative numbers summing to one. It follows from (9) that $I_n(t, t') \rightarrow I(t, t')$, where

$$I(t, t') = \sum_{j=1}^J \pi_j I_j(t, t'). \quad (10)$$

Asymptotically, the Fisher information of ξ per observation for the combined sample is $I(t) = I(t, t)$ at (t, ξ_0) . Assuming that $I(t)$ is non-singular for all

$t \in \mathbb{T}$ (this assumption will be relaxed in Section 4), let $A(t)$ be a square root of $I(t)$, i.e. any $p \times p$ matrix satisfying

$$I(t) = A(t)^T A(t). \quad (11)$$

It will be convenient to reparametrize ξ and put

$$\begin{aligned} \tilde{l}_n(t, \delta) &= l_n(t, \xi_0 + n^{-1/2} \delta A(t)^{-T}) \\ \Delta_n(t) &= n^{1/2} (\Xi(t) - \xi_0) A(t)^T, \end{aligned} \quad (12)$$

where $A(t)^{-T}$ is the conjugate of $A(t)^{-1}$. Then

$$\lambda_n = \sup_{t \in \mathbb{T}} Y_n(t), \quad (13)$$

where

$$Y_n(t) = \sup_{\xi \in \Xi(t)} 2(l_n(t, \xi) - l(t, \xi_0)) = \sup_{\delta \in \Delta_n(t)} 2(\tilde{l}_n(t, \delta) - \tilde{l}_n(t, 0)). \quad (14)$$

We view $Y_n = \{Y_n(t); t \in \mathbb{T}\}$ as a random variable on $(D(\mathbb{T}), \mathcal{B}(D(\mathbb{T})))$. $D(\mathbb{T})$ is the space of functions $\mathbb{T} \rightarrow \mathbb{R}$ defined as the closure, with respect to supremum norm on \mathbb{T} , of all simple functions that are constant on sub-cubes $[a_1, b_1) \times \dots \times [a_q, b_q)$. This is the natural generalization of cadlag functions when $q = 1$ in that each $y \in D(\mathbb{T})$ has limits at t along the $2^q - 1$ quadrants 'below' t and is continuous at t along the quadrant 'above' t for each $t \in \mathbb{T}$. We endow $D(\mathbb{T})$ with the Skorohod topology (Bickel and Wichura, 1971) in the definition of $\mathcal{B}(D(\mathbb{T}))$.

In Section 3, the major step in showing (6) is to establish

$$Y_n \xrightarrow{\mathcal{L}} Y \quad (15)$$

on $(D(\mathbb{T}), \mathcal{B}(D(\mathbb{T})))$, where $Y = \{Y(t); t \in \mathbb{T}\}$ is a $\bar{\chi}^2$ -process, to be defined below. Then (6) follows by a simple application of the Continuous Mapping Theorem (Billingsley, 1968, Theorem 5.1).

In order to define Y , we first need an asymptotic approximation of Y_n . Let

$$Z_n(t) = \left. \frac{\partial \tilde{l}_n(t, \delta)}{\partial \delta} \right|_{\delta=0} = n^{-1/2} \sum_{j=1}^J \sum_{i=1}^{n_j} \psi_j(X_{ji}; t) A(t)^{-1} \quad (16)$$

be the score vector with respect to δ at $(t, 0)$. We assume that $Z_n = \{Z_n(t); t \in \mathbb{T}\}$ is a random element of $(D^p(\mathbb{T}), \mathcal{B}(D^p(\mathbb{T})))$. Here $D^p(\mathbb{T})$ is

the space of functions $\mathbb{T} \rightarrow \mathbb{R}^p$ that have limits from below and are continuous from above, defined as when $p = 1$. Likewise, the Skorohod topology of $D^p(\mathbb{T})$ is defined analogously as when $p = 1$. The covariance function of Z_n is

$$\rho_n(t, t') = \text{Cov}(Z_n(t), Z_n(t')) = A(t)^{-T} I_n(t, t') A(t')^{-1}. \quad (17)$$

A Taylor expansion of \tilde{l}_n with respect to δ at $\delta = 0$ yields

$$\tilde{l}_n(t, \delta) = \tilde{l}_n(t, 0) + Z_n(t) \delta^T - \frac{1}{2} |\delta|^2 + \tilde{R}_n(t, \delta), \quad (18)$$

for $t \in \mathbb{T}$ and $\delta \in \Delta_n(t)$, with $\tilde{R}_n(t, \delta)$ a remainder term that is small as $n \rightarrow \infty$. For each t , (18) is the usual asymptotic expansion of the log likelihood used to derive the asymptotic distribution of the LR-test, see e.g. Serfling (1980). Hence, (18) can be viewed as a process analogue when considering several t simultaneously. Inserting (18) into (14) we obtain

$$Y_n(t) = \sup_{\delta \in \Delta_n(t)} \left(2Z_n(t) \delta^T - |\delta|^2 + 2\tilde{R}_n(t, \delta) \right) \quad (19)$$

for all $t \in \mathbb{T}$.

Now (19) gives some ideas on how to define the limit process Y . We need to I) show that \tilde{R}_n is asymptotically negligible, II) replace Z_n by a limit process Z and III) replace $\Delta_n(t)$ by a limit region $\Delta(t)$.

To deal with II), we assume that

$$Z_n \xrightarrow{\mathcal{L}} Z \text{ as } n \rightarrow \infty \quad (20)$$

on $(D^p(\mathbb{T}), \mathcal{B}(D^p(\mathbb{T})))$, where $Z = \{Z(t), t \in \mathbb{T}\}$ is a Gaussian process with covariance function

$$\rho(t, t') = A(t)^{-T} I(t, t') A(t')^{-1}, \quad (21)$$

and $A(t)$ satisfies (11). Notice that $\rho(t, t')$ is the limit of (17) because of (9). Since $\rho(t, t) = \text{Id}$, the identity matrix of order p , the marginal distribution of $Z(t)$ is p -dimensional standard normal for all $t \in \mathbb{T}$.

To deal with III), we assume that $\Xi(t)$ is shaped as a cone bounded by hyperplanes locally around ξ_0 . In more detail, let $V(t)$ be a $r \times p$ matrix with linearly independent row vectors $v_1(t), \dots, v_r(t)$, $0 \leq r \leq p$. Define

$$C(V(t)) = \{\xi \in \mathbb{R}^p; v_i(t) \xi^T \geq 0, i = 1, \dots, r\}$$

as a cone in \mathbb{R}^p bounded by the r hyperplanes through the origin that are perpendicular to $v_1(t), \dots, v_r(t)$. Let d be the distance metric between sets, defined in the appendix, and assume that

$$\lim_{\tau \rightarrow 0^+} \sup_{t \in \mathbb{T}} d\left(\tau^{-1}(\Xi - \{\xi_0\}), C(V(t))\right) = 0. \quad (22)$$

Define

$$\Delta(t) = C(V(t))A(t)^T = C(U(t)), \quad (23)$$

where $U(t) = V(t)A(t)^{-1}$. It is proved in the appendix that (12) and (22), under some regularity conditions, lead to

$$\lim_{n \rightarrow \infty} \sup_{t \in \mathbb{T}} d(\Delta_n(t), \Delta(t)) = 0. \quad (24)$$

Going back to (19), we define $Y(t)$ by replacing $\tilde{R}_n(t, \cdot)$, $Z_n(t)$ and $\Delta_n(t)$ by 0, $Z(t)$ and $\Delta(t)$. This yields

$$\begin{aligned} Y(t) &= \sup_{\delta \in \Delta(t)} \left(2Z(t)\delta^T - |\delta|^2\right) \\ &= |Z(t)|^2 - |Z(t) - P_{\Delta(t)}Z(t)|^2 \\ &= |P_{\Delta(t)}Z(t)|^2, \end{aligned} \quad (25)$$

where $P_{\Delta}z = \arg \min_{y \in \Delta} |z - y|$ is the projection of $z \in \mathbb{R}^p$ onto $\Delta \subset \mathbb{R}^p$. The last equality of (25) follows since $\Delta(t)$ is a cone in \mathbb{R}^p , see e.g. Self and Liang (1987), Shapiro (1988) and Sen and Silvapulle (2002, 2005). In words, $Y(t)$ is defined as the squared Euclidean norm of the projection of $Z(t)$ onto a (possible t -varying) cone in \mathbb{R}^p bounded by r hyperplanes.

Since $\Delta(t)$ is a cone bounded by hyperplanes for each t , it follows that the marginal distribution of the $\bar{\chi}^2$ -process Y is a mixture of χ^2 -variables,

$$Y(t) \in \sum_{i=0}^p w_i(t) \chi^2(i), \quad (26)$$

for some non-negative weights $\{w_i(t)\}_{i=0}^p$ summing to one. In more detail $w_i(t)$ is the probability that the projection $P_{\Delta(t)}Z(t)$ belongs to a face of $\Delta(t)$ with dimension i , see Sen and Silvapulle (2002) and references therein.

When $r = 0$ we have $\Delta(t) = \mathbb{R}^p$, so that $w_p(t) = 1$ in (26), and Y is a pure $\chi^2(p)$ -process. When $r = 1$, $\Delta(t)$ is a half-plane and $Y(t)$ is a 0.5 : 0.5-mixture of χ^2 -distributions with $p-1$ and p degrees of freedom, i.e. $w_{p-1}(t) = w_p(t) = 0.5$. When $r = 2$, $Y(t)$ is a mixture of three types of χ^2 -distributions, with weights $w_{p-2}(t) = 0.5 - w_p(t)$, $w_{p-1} = 0.5$ and

$$w_p(t) = \cos^{-1}(-u_1(t)u_2(t)^T / (|u_1(t)||u_2(t)|)) / 2\pi, \quad (27)$$

where $u_1(t)$ and $u_2(t)$ are the two row vectors of $U(t)$. In case $r = p = 2$, we can rewrite (27) in the more explicit form

$$w_2(t) = \cos^{-1}(-B_{12}(t)/\sqrt{B_{11}(t)B_{22}(t)})/2\pi, \quad (28)$$

where $B(t) = (B_{kl}(t))_{k,l=1}^2 = V(t)QI(t)Q^T V(t)^T$ and

$$Q = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

This is a generalization of Self and Liang (1987), who derived (28) when $V(t) = \text{Id}_2$, the identity matrix of order 2.

3 Asymptotic Distribution - Non-Singular Case

We will assume the following regularity conditions:

- i. Let $\hat{\xi}_n(t) = \arg \max_{\xi \in \Xi(t)} l_n(t, \xi)$. The estimators $\{\hat{\xi}_n(t); t \in \mathbb{T}\}$ are uniformly consistent, i.e.

$$\lim_{n \rightarrow \infty} P_{\theta_0} \left(\sup_{t \in \mathbb{T}} |\hat{\xi}_n(t)| \geq \varepsilon \right) = 0$$

for any $\varepsilon > 0$.

- ii. The functions $t \rightarrow \log f_j(x; t, \xi_0)$, $t \rightarrow \log \bar{f}_j(x; t) = \log \sup_{\xi \in \Xi(t)} f_j(x; t, \xi)$ belong to $D(\mathbb{T})$ and $t \rightarrow \psi_j(x; t)$ belongs to $D^p(\mathbb{T})$ for all $j = 1, \dots, J$ and $x \in \mathcal{X}'_j \subset \mathcal{X}_j$, where $P_{\theta_0}(\mathcal{X}'_j) = 1$.
- iii. For all $t \in \mathbb{T}$ and $j = 1, \dots, J$, it holds that $E_{\theta_0} \psi_j(X_{j1}; t) = 0$. Put $\zeta_j(x; t) = (\partial^2 \log f_j(x; t, \xi) / \partial \xi_k \partial \xi_l)_{k,l=1}^p \Big|_{\xi=\xi_0}$. Then $E_{\theta_0}(\zeta_j(X_{j1}; t)) = -I_j(t)$ and

$$\sup_{t \in \mathbb{T}} \|\hat{I}_n(t) - I(t)\|_1 = o_p(1) \text{ under } H_0, \quad (29)$$

with $\hat{I}_n(t) = -\sum_{j=1}^J \sum_{i=1}^{n_j} \zeta_j(X_{ji}; t)/n$.

- iv. $(t, t') \rightarrow I(t, t')$ is continuous on $\mathbb{T} \times \mathbb{T}$ and $t \rightarrow A(t)$ is continuous on \mathbb{T} with respect to the topology introduced by $\|\cdot\|_1$. The function ρ in (21) is the covariance function of a well defined Gaussian process $Z \in D^p(\mathbb{T})$, which is continuous with probability one.
- v. Let $\kappa_{\min}(t)$ be the smallest eigenvalue of $I(t)$. Then

$$\kappa_{\min} = \inf_{t \in \mathbb{T}} \kappa_{\min}(t) > 0.$$

vi. For some $\varepsilon > 0$ and each $j = 1, \dots, J$, there exists a $\mathcal{B}(\mathcal{X}_j)$ -measurable function $M_j : \mathcal{X}_j \rightarrow \mathbb{R}$ with $E_{\theta_0} M_j(X) < \infty$ such that

$$\max_{k; |k|=3} \sup_{\theta \in \Theta_\varepsilon} \left| \frac{\partial^3 f_j(x; \theta)}{\partial \xi^k} \right| \leq M_j(x),$$

where $k = (k_1, \dots, k_p)$, each k_i is a non-negative integer, $|k| = \sum_{i=1}^p k_i$, $\partial \xi^k = \partial \xi_1^{k_1} \dots \partial \xi_p^{k_p}$, $\Theta_\varepsilon = \{(t, \xi); t \in \mathbb{T}, \xi \in \Xi(t) \text{ and } |\xi - \xi_0| \leq \varepsilon\}$.

vii. The matrix-valued function $t \rightarrow V(t)$ is continuous on \mathbb{T} .

Theorem 1 (Weak convergence of LR-process.) *Consider the log likelihood function (2), assume the non-identifiability condition (4) holds, as well as (9), (20), (22) and i-vii. Then (15) and (6) follow. \square*

Remark 1 The consistency condition i can be removed if we weaken the interpretation of $\hat{\xi}_n(t)$ to be a local maximum of $l_n(t, \cdot)$ such that the sequence $\{\hat{\xi}_n(t); t \in \mathbb{T}\}$ is uniformly consistent with respect to t as $n \rightarrow \infty$. It is implicit from the proof of Theorem 1 that such a sequence exists. \square

Remark 2 Conditions ii and iv guarantee that $Z_n \in D^p(\mathbb{T})$ and ii that $Y_n \in D(\mathbb{T})$. A sufficient set of conditions for $\log \bar{f}_j(x; \cdot) \in D(\mathbb{T})$ to hold is that $\Xi(t) \equiv \Xi$ is compact, that $t \rightarrow \log f_j(x; t, \xi) \in D(\mathbb{T})$ for each $\xi \in \Xi$ and $\sup_{\xi \in \Xi, t \in \mathbb{T}} |\psi_j(x; t, \xi)| < \infty$. Then there exists a dense sequence ξ_1, ξ_2, \dots in Ξ , so that

$$\log \bar{f}_j(x; t) = \lim_{m \rightarrow \infty} \max_{1 \leq k \leq m} \log f_j(x; t, \xi_k),$$

is the limit (in m) of a Cauchy sequence with respect to supremum norm on $D(\mathbb{T})$. Completeness of $D(\mathbb{T})$ with respect to the weaker Skorohood metric proves $\log \bar{f}_j(x; \cdot) \in D(\mathbb{T})$. \square

Remark 3 By Prohorov's Theorem (Billingsley, 1968, Chapter 1.6), in order to prove (20) we need to establish convergence of finite-dimensional distributions of $\{Z_n\}$ and tightness. Convergence of finite-dimensional distributions follows immediately from (9) and the Lindeberg-Feller Central Limit Theorem. A sufficient condition for tightness is given by Theorem 15.6 of Billingsley (1968) when $q = p = 1$ and Theorem 3 of Bickel and Wichura (1971) for general q and $p = 1$. The latter result translates to $p > 1$ as well. \square

Remark 4 For (29), we use the bound

$$\|\hat{I}_n(t) - I(t)\|_1 \leq \sum_{j=1}^J (n_j/n) |\hat{I}_{n_j}(t) - I_j(t)| + \sum_{j=1}^J |n_j/n - \pi_j| |I_j(t)| \quad (30)$$

where $\hat{I}_{n_j}(t) = -\sum_{i=1}^{n_j} \zeta(X_{ji}; t, \xi_0)/n_j$. The second term of (30) tends to zero uniformly for $t \in \mathbb{T}$ according to (9) and $|I_j(t)|$ is uniformly bounded in t because of v. To handle the first term, we need show that $\sup_{t \in \mathbb{T}} |\hat{I}_{n_j,kl}(t) - I_{j,kl}(t)| = o_p(1)$ each for each $j = 1, \dots, J$ and $1 \leq k, l \leq p$. This is a uniform law of large numbers result, see e.g. Bickel and Millar (1992) for regularity conditions. \square

4 Asymptotic Distribution - Singular Case

We will now generalize Theorem 1 and allow for singular Fisher information. To this end, define

$$\mathbb{T}_s = \{t \in \mathbb{T}; I(t) \text{ is singular}\}. \quad (31)$$

It turns out that (15) typically fails when $\mathbb{T}_s \neq \emptyset$, although the weaker requirement (6) does not. In order to derive (6), we will assume that \mathbb{T}_s contains at most a finite number of points. The major technical difficulty is that the reparametrization (12) is only valid for $t \in \mathbb{T}_{ns} = \mathbb{T} \setminus \mathbb{T}_s$, the set of non-singular points. Still, it is more convenient to work with the transformed parameter δ instead of ξ when $\mathbb{T}_s \neq \emptyset$. Define

$$\tilde{f}_j(x; t, \delta) = f_j(x; t, \xi_0 + \delta A(t)^{-T}) \quad (32)$$

for all $t \in \mathbb{T}_{ns}$, so that

$$\tilde{l}_n(t, \delta) = \sum_{j=1}^J \sum_{i=1}^{n_j} \log \tilde{f}_j(X_{ji}; t, n^{-1/2} \delta).$$

Then i-vii are changed to:

i'. Let $\hat{\delta}_n(t) = \arg \max_{\delta \in \Delta_n(t)} \tilde{l}_n(t, \delta)$. Then

$$\lim_{n \rightarrow \infty} P_{\theta_0} \left(\sup_{t \in \mathbb{T}_{ns}} |\hat{\delta}_n(t)| \geq \sqrt{n} \varepsilon \right) = 0$$

for each $\varepsilon > 0$.

- ii'. For each $j = 1, \dots, J$ the following holds: The functions $t \rightarrow \log f_j(x; t, \xi_0)$, $t \rightarrow \log \tilde{f}_j(x; t)$ belong to $D(\mathbb{T})$ for all $x \in \mathcal{X}'_j \subset \mathcal{X}_j$, where $P_{\theta_0}(\mathcal{X}'_j) = 1$. The function $\tilde{\psi}_j : \mathbb{T}_{ns} \rightarrow \mathbb{R}^p$, defined by $\tilde{\psi}_j(x; t) = \psi_j(x; t)A(t)^{-1}$, can be uniquely extended to \mathbb{T} so that $\tilde{\psi}_j \in D^p(\mathbb{T})$ for all $x \in \mathcal{X}'_j$.
- iii'. For all $t \in \mathbb{T}_{ns}$ and $j = 1, \dots, J$, it holds that $E_{\theta_0}\psi_j(X_{j1}; t) = 0$. Put $\tilde{\zeta}_j(x; t) = \left(\partial^2 \log \tilde{f}_j(x; t, \delta) / \partial \delta_k \partial \delta_l \right)_{k,l=1}^p \Big|_{\delta=0}$. Then $E_{\theta_0}(\tilde{\zeta}_j(X_{j1}; t)) = -A(t)^{-T} I_j(t) A(t)^{-1}$, and

$$\sup_{t \in \mathbb{T}_{ns}} \|\widehat{\text{Id}}_n(t) - \text{Id}\|_1 = o_p(1) \text{ under } H_0,$$

$$\text{with } \widehat{\text{Id}}_n(t) = -\sum_{j=1}^J \sum_{i=1}^{n_j} \tilde{\zeta}_j(X_{ji}; t) / n.$$

- iv'. The conditions on $I(\cdot, \cdot)$ and $A(\cdot)$ in iv hold. In addition, the function $\rho : \mathbb{T}_{ns} \times \mathbb{T}_{ns} \rightarrow \mathbb{R}$ in (21) can be extended to a continuous function $\mathbb{T} \times \mathbb{T} \rightarrow \mathbb{R}$ which is the covariance function of a Gaussian process $Z \in D^p(\mathbb{T})$ that is continuous with probability one.
- v'. The number of singularity points \mathbb{T}_s is positive but finite.
- vi'. For some $\varepsilon > 0$ and each $j = 1, \dots, J$, there exists a $\mathcal{B}(\mathcal{X}_j)$ -measurable function $\tilde{M}_j : \mathcal{X}_j \rightarrow \mathbb{R}$ with $E_{\theta_0}\tilde{M}_j(X) < \infty$ such that

$$\max_{k:|k|=3} \sup_{(t,\delta) \in \Upsilon_\varepsilon} \left| \frac{\partial^3 \tilde{f}_j(x; t, \delta)}{\partial \delta^k} \right| \leq \tilde{M}_j(x).$$

where $k = (k_1, \dots, k_p)$ and $\Upsilon_\varepsilon = \{(t, \delta), t \in \mathbb{T}_{ns}, |\delta| \leq \varepsilon \text{ and } \delta \in \Delta_1(t)\}$.

- vii'. $\tau^{-1}(\Xi(t) - \{\xi_0\})$ is a non-decreasing function of $\tau > 0$ for each $t \in \mathbb{T}_{ns}$, $t \rightarrow V(t)$ is continuous on \mathbb{T}_{ns} and

$$\lim_{\tau \rightarrow 0^+} \sup_{t \in \mathbb{T}_\varepsilon} d(\tau^{-1}(\Xi(t) - \{\xi_0\}), C(V(t))) = 0$$

for each $\varepsilon > 0$, where $\mathbb{T}_\varepsilon = \{t \in \mathbb{T}; \inf_{t' \in \mathbb{T}_s} |t - t'| \geq \varepsilon\}$. $\Delta(t)$, defined by (23) for $t \in \mathbb{T}_{ns}$, can be uniquely extended to a continuous function on \mathbb{T} .

Theorem 2 Consider the log likelihood function (2), assume the non-identifiability condition (4) holds, as well as (9), (20) and i'-vii'. Then (6) follows.

Remark 5 Condition ii' guarantees that $Y_n \in D(\mathbb{T})$ and $Z_n \in D^p(\mathbb{T})$. Since iv' implies that $Z \in D_p(\mathbb{T})$ is well defined and continuous with probability one, it follows from vii' that $Y \in D(\mathbb{T})$ is also continuous with probability one. \square

Remark 6 A sufficient condition in vii' for $\tau \rightarrow \tau^{-1}(\Xi(t) - \{\xi_0\})$ to be non-decreasing is that $\Xi(t)$ is convex. It guarantees that $\Delta_n(t)$ is a non-decreasing sequence of sets, converging to $\Delta(t)$ as $n \rightarrow \infty$ for all $t \notin \mathbb{T}_{ns}$. This will simplify the proof of Theorem 2. \square

5 Examples

In the examples to follow, we write $\Delta(t) = \Delta$, $A(t) = A$ and $V(t) = V$ whenever these quantities are independent of $t \in \mathbb{T}$.

Example 1 (Identifiability under H_0 .) When $\Theta_0 = \{\theta_0\}$, we put $q = 0$, $\mathbb{T} = \{0\}$, $\xi_0 = \theta_0$ and $\lambda = Y(0) = Y$. The 'process' Y is then a $\bar{\chi}^2$ -distributed random variable. \square

Example 2 (Polar coordinates.) Assume $J = 1$ and that $f(x; \eta)$ is parametrized by a two-dimensional Cartesian parameter vector $\eta = (\eta_1, \eta_2)$. We wish to test $H_0 : \eta = \eta_0$ against $H_1 : \eta \neq \eta_0$, and assume that η_0 is an inner point of the parameter space. This is a special case of Example 1 with $r = 0$, so that $\lambda_n \xrightarrow{\mathcal{L}} \chi^2(2)$ as $n \rightarrow \infty$. However, by switching to polar coordinates we may formulate this as $q = p = 1$, using

$$\eta = \eta_0 + \xi(\cos(t), \sin(t)),$$

for $t \in \mathbb{T} = [0, 2\pi]$, $\xi \in \Xi = [0, \infty)$ and $\xi_0 = \{0\}$. Let \bar{I} be the 2×2 Fisher information matrix for η at $\eta = \eta_0$ and $\bar{l}(\eta)$ the log likelihood function for η . Then $I(t) = u(t)\bar{I}u(t)^T$, with $u(t) = (\cos(t), \sin(t))$, so that $\mathbb{T}_s = \emptyset$ if \bar{I} is non-singular. Let $W_n = \partial \bar{l}_n(\eta) / \partial \eta \Big|_{\eta=\eta_0} (n\bar{I})^{-1/2}$ be the score vector. Then $Z_n(t) = u(g(t))W_n^T$ where $g : [0, 2\pi] \rightarrow [0, 2\pi]$ is a transformed angle, defined by $u(g(t)) = u(t)\bar{I}^{1/2} / \sqrt{I(t)}$. Letting $n \rightarrow \infty$, $W_n \xrightarrow{\mathcal{L}} W = (W_I, W_{II}) \in N_2(0, \text{Id})$, where Id is the identity matrix of order 2. Hence the limiting distribution of Z_n is the process

$$Z(t) = u(g(t))W^T,$$

a time transformed sinusoidal Gaussian random process with covariance function

$$\rho(t, t') = \cos(g(t) - g(t')).$$

We have $r = 1$ and $\Delta = [0, \infty)$, so that

$$Y(t) = \max(Z(t), 0)^2 = \max(W_I \cos(g(t)) + W_{II} \sin(g(t)), 0)^2.$$

Since g is a bijection on $[0, 2\pi]$,

$$\lambda = \sup_{0 \leq t \leq 2\pi} Y(t) = W_I^2 + W_{II}^2 \in \chi^2(2)$$

in accordance with standard LR theory. \square

In the previous example, non-identifiability under H_0 was a consequence of using polar coordinates. It could simply be removed by switching to Cartesian coordinates. We now give three examples with Gaussian mixtures, $J = 1$ and $\mathcal{X} = \mathbb{R}$, where no such trick is available.

Example 3 (Mixture distributions with $q = p = 1$.) Assume

$$X_i \in (1 - \xi)N(0, 1) + \xi N(t, 1) \tag{33}$$

with $\mathbb{T} = [-T, T]$ for some $T > 0$ and $\Xi = [0, 1]$. If $\xi_0 = \{0\}$, it follows that $r = p = 1$, $\Delta = [0, \infty)$, the score vector is $\psi(x; t) = \exp(tx - t^2/2) - 1$, $I(t, t') = \exp(tt') - 1$, $\mathbb{T}_s = \{0\}$ and

$$Z_n(t) = \frac{1}{\sqrt{n(\exp(t^2) - 1)}} \sum_{i=1}^n (\exp(tX_i - t^2/2) - 1)$$

if $t \neq 0$ and $Z_n(0) = \sum_{i=1}^n X_i / \sqrt{n}$. The covariance function of Z_n and Z is

$$\rho(t, t') = (\exp(tt') - 1) / \sqrt{(\exp(t^2) - 1)(\exp((t')^2) - 1)}$$

and $Y(t) = \max(Z(t), 0)^2$, as in the previous example. Weak convergence $\lambda_n \xrightarrow{\mathcal{L}} \lambda$ has been established by Chen and Chen (2001, Theorem 2). \square

Example 4 (Mixture distributions with $q = 2, p = 1$.) We generalize Example 3 so that the mean *and* variance of the second component in (33) are unknown, i.e.

$$X_i \in (1 - \xi)N(0, 1) + \xi N(t_1, t_2),$$

with $t = (t_1, t_2)$ and $\mathbb{T} = [-T_1, T_1] \times [T_{21}, T_{22}]$ for some $T_1 > 0$ and $0 < T_{21} < T_{22} < 2$. In this case $q = 2$, $p = r = 1$, $\Delta = [0, \infty)$, $\mathbb{T}_s = \{(0, 1)\}$ and

$$\psi(x; t) = \frac{1}{\sqrt{t_2}} \exp\left(\frac{1}{2}\left(1 - \frac{1}{t_2}\right)x^2 + \frac{t_1}{t_2}x - \frac{t_1^2}{2t_2}\right).$$

Hence

$$\lambda = \sup_{t \in \mathbb{T}} \max(0, Z(t))^2,$$

where the covariance function ρ of Z is deduced from (7) and (21) on $\mathbb{T}_{ns} \times \mathbb{T}_{ns}$ and then extended to all of $\mathbb{T} \times \mathbb{T}$. Notice the restriction $t_2 < 2$, which is needed in order for $I(t)$ to be finite. \square

Example 5 (Mixture distributions with $q = 1, p = 2$.) A second generalization of Example 3 is

$$X_i \in (1 - \xi_2)N(\xi_1, 1) + \xi_2 N(t, 1),$$

with $\mathbb{T} = [-T, T]$, $\xi = (\xi_1, \xi_2)$ and $\Xi = [-T, T] \times [0, 0.5]$. The restriction $\xi_2 \leq 0.5$ is imposed for identifiability, since $\theta = (t, \xi_1, \xi_2)$ and $\theta = (\xi_1, t, 1 - \xi_2)$ give the same model. The null hypotheses is defined by $\xi_0 = (0, 0)$. The score vector is

$$\begin{aligned} \psi(x; t) &= (x, \exp(tx - t^2/2) - 1), \\ I(t, t') &= \begin{pmatrix} 1 & t' \\ t & e^{tt'} - 1 \end{pmatrix} \end{aligned} \quad (34)$$

and $\mathbb{T}_s = \{0\}$. We use a Cholesky decomposition of $I(t)$, i.e. choose an upper triangular $A(t)$ in (11), defined through its inverse

$$A(t)^{-1} = \begin{pmatrix} 1 & -t/\sigma_t \\ 0 & 1/\sigma_t \end{pmatrix}, \quad (35)$$

with $\sigma_t^2 = e^{t^2} - 1 - t^2$. From (16) it follows that $Z_n(t) = (Z_{n1}(t), Z_{n2}(t))$, where

$$\begin{aligned} Z_{n1}(t) &= \sum_{i=1}^n X_i / \sqrt{n}, \\ Z_{n2}(t) &= \sum_{i=1}^n (\exp(tX_i - t^2/2) - 1 - tX_i) / (\sqrt{n}\sigma_t), \quad t \neq 0, \\ Z_{n2}(0) &= \sum_{i=1}^n (X_i^2 - 1) / \sqrt{2n}. \end{aligned}$$

The covariance function of Z_n and Z is found by inserting (34) and (35) into (21), giving

$$\rho(t, t') = \begin{pmatrix} 1 & 0 \\ 0 & \sigma_{tt'}/(\sigma_t\sigma_{t'}) \end{pmatrix},$$

where $\sigma_{tt'} = e^{tt'} - 1 - tt'$. Hence $Z(t) = (Z_1, Z_2(t))$, where $Z_1 \in N(0, 1)$ and Z_2 is a real-valued Gaussian process, independent of Z_1 , with covariance function $\sigma_{tt'}/(\sigma_t\sigma_{t'})$.

We have $r = 1$, $V = (01)$ and $C(V)$ is the upper half plane. Hence $U(t) = (01)A(t)^{-1} = (01)/\sigma_t$. This yields $\Delta = \{(\delta_1, \delta_2); \delta_2 \geq 0\}$, the upper half plane for all $t \in \mathbb{T}$. It follows that $Y(t) = |P_\Delta Z(t)|^2 = Z_1^2 + \max(Z_2(t)^2, 0)$ and hence

$$\lambda^2 = Z_1^2 + \sup_{-T \leq t \leq T} \max(Z_2(t)^2, 0).$$

Weak convergence $\lambda_n \xrightarrow{\mathcal{L}} \lambda$ has been proved by Chen and Chen (2002, Theorem 3). \square

6 Linkage Analysis and MOD Scores

It has been noticed that some test statistics arising in genetic linkage analysis can be formulated as the supremum of a one-dimensional Gaussian or $\chi^2(1)$ -process along an interval, see e.g. Lander and Botstein (1989) and Feingold et al. (1993). Asymptotically, such processes correspond to the supremum of a $\bar{\chi}^2$ -process when either $(q, p, r) = (1, 1, 1)$ and $\Delta(t) = [0, \infty)$, or $(q, p, r) = (1, 1, 0)$. Dupuis and Siegmund (1995) define another test statistic, which asymptotically is the the supremum of a $\chi^2(2)$ -process along an interval (i.e. $(q, p, r) = (1, 2, 0)$).

In this section, we show that LR-tests in linkage analysis, usually referred to as MOD scores, fit into our framework. We start with a brief summary of genetic linkage analysis, see Sham (1998) for more details. The objective is to test whether a disease susceptibility gene τ is present along a chromosome $[0, T]$ of (genetic map) length T , formally written as

$$\begin{aligned} H_0 : \tau &\notin [0, T], \\ H_1 : \tau &\in [0, T]. \end{aligned} \tag{36}$$

At our disposal we have n families. For each family, disease related quantities, so called phenotypes, and DNA marker data, is registered. To begin with, we assume that all n families have the same pedigree structure \mathcal{P} with

no inbreeding, such that each $k \in \mathcal{P}$ has both or none of his/her parents in \mathcal{P} . In the former case k is a nonfounder and in the latter case k is a founder. The phenotype vector $\Phi = (\Phi_k, k \in \mathcal{P})$ is also identical for all n families. During meiosis, i.e. production of germ cells, a child receives genetic material from a parent, inherited from the grandfather or grandmother. There are two meioses (one from each parent) of each nonfounder, giving a total of m meioses in the pedigree, where $m/2$ is the number of nonfounders. Transmission of genetic material at position (locus) s for a given pedigree can be described by means of a binary inheritance vector $v(s) = (v_1(s), \dots, v_m(s))$, where $v_l(s)$ is zero or one depending on whether grandpaternal or grandmaternal DNA was transmitted at s during the l^{th} meiosis. Assuming complete marker data, we register DNA transmission

$$x = \{v(s); 0 \leq s \leq T\}$$

along the whole chromosome. Let \mathcal{X} denote the set of all such x with a most a finite number of discontinuities. A discontinuity of $v_l(\cdot)$ at s corresponds to a crossover, i.e. switching between grandpaternal to grandmaternal DNA transmission at s .

We assume Mendelian inheritance and Haldane's map function, which implies that $\{v_l(s); 0 \leq s \leq T\}_{l=1}^m$ are independent and stationary Markov processes on $\{0, 1\}$. The intensities of jumping from 0 to 1 and from 1 to 0 are both 1 (measuring distance in Morgans).

Under H_1 , we assume that Φ gives some information about $v(\tau)$ and that x and Φ are conditionally independent given $v(\tau)$. The conditional distribution of the inheritance process given phenotypes is then

$$f(x; \theta) = f(x; \theta, \mathcal{P}, \Phi) = P_\theta(x|\Phi) = P_{\mathbf{p}, \gamma}(v(\tau)|\Phi)P(x|v(\tau)), \quad (37)$$

where $\theta = (\tau, \mathbf{p}, \gamma)$, \mathbf{p} contains frequencies of all possible alleles (= expressions of the gene) at τ , and γ contains penetrance parameters. Notice that $P(x|v(\tau)) = \exp(-mT)$ is a constant that can be dropped. In the sequel, we refer to (\mathcal{P}, Φ) as the *pedigree type*.

Using Bayes' rule, we find that

$$P_{\mathbf{p}, \gamma}(v(\tau)|\Phi) = \frac{2^{-m}P_{\mathbf{p}, \gamma}(\Phi|v(\tau))}{P_{\mathbf{p}, \gamma}(\Phi)} \propto P_{\mathbf{p}, \gamma}(\Phi|v(\tau)), \quad (38)$$

where the proportionality constant involves \mathbf{p} and γ , but not $v(\tau)$.

Let F be the number of founders in the pedigree and $N = F + m/2$ the total number of individuals. To expand (38) further, let $G = (G_1, \dots, G_N)$ denote

the set of genotypes of all pedigree members at τ , where $G_k = (a_{2k-1}, a_{2k})$ contains two alleles of k at locus τ , one inherited from the father (a_{2k-1}) and one inherited from the mother (a_{2k}). We consider a biallelic disease gene, i.e. $a_k \in \{0, 1\}$, with $P(1) = p$ and $P(0) = q = 1 - p$. Let $a = (a_1, \dots, a_{2F})$ contain the $2F$ founder alleles (assuming the F founders are numbered as $1, \dots, F$). Since $G = G(a, v)$ is obtained by spreading founder alleles according to $v = v(\tau)$ to all nonfounders and Φ depends on v only through G , we find, by conditioning on a , that

$$\begin{aligned} P_{p,\gamma}(\Phi|v) &= \sum_a P_p(a) P_\gamma(\Phi|G(a, v)) \\ &= \sum_a p^{|a|} q^{2F-|a|} P_\gamma(\Phi|G(a, v)), \end{aligned} \quad (39)$$

where $|a| = \sum_1^{2f} a_k$ is the number of 1-alleles among the $2f$ founder alleles.

In principle, our framework includes the possibility of discrete or continuous phenotypes, as well as polygenic and shared environmental effects. For simplicity however, we will most of the time assume a monogenic disease with binary phenotypes, conditionally independent given disease genotypes, i.e.

$$P_\gamma(\Phi|G) = \prod_{k=1}^N P_\gamma(\Phi_k|G_k), \quad (40)$$

where $\Phi_k \in \{0, 1, ?\}$, 1='affected', 0='unaffected' and ?='unknown'. Writing $\gamma = (\gamma_0, \gamma_1, \gamma_2)$, where γ_j is the probability of that an individual with j 1-alleles is affected, we can write each term of (40) as

$$P_\gamma(\Phi_k|G_k) = \gamma_{|G_k|}^{\{\Phi_k=1\}} (1 - \gamma_{|G_k|})^{\{\Phi_k=0\}}, \quad (41)$$

where $|G_k| = a_{2k-1} + a_{2k}$.

With P and Γ the set of allowable values p and γ we have

$$\Theta = [0, T] \times P \times \Gamma. \quad (42)$$

Since, for our genetic model, p is a probability and γ is a vector of three probabilities, we put $P = [0, 1]$ and $\Gamma = [0, 1]^3$. In order to define Θ_0 , we notice that the null density $f_0(x) = 2^{-m} \exp(-mT)$ is obtained by replacing $P_{p,\gamma}(v(\tau)|\Phi)$ with the uniform distribution 2^{-m} in (37). Hence we put

$$\begin{aligned} \Theta_0 &= \{\theta \in \Theta; P_{p,\gamma}(v(\tau)|\Phi) = 2^{-m} \text{ for all } v(\tau) \in \{0, 1\}^m\} \\ &= [0, T] \times \cup_{p \in P} (\{p\} \times \Gamma_0(p)). \end{aligned} \quad (43)$$

Given p , $\Gamma_0(p)$ is the set of penetrance vectors for which the disease gene cannot be detected. If no more than one $k \in \mathcal{P}$ has known phenotype, there

is not enough information to detect the disease gene for any genetic model (p, γ) . This leads to a degenerate testing problem, since $\Gamma_0(p) = \Gamma$ for all $p \in P$ and consequently $\Theta_0 = \Theta$. To avoid such problems, we assume that the pedigree type (\mathcal{P}, Φ) is such that

$$\Gamma_0(p) = \{\gamma; \gamma_0 = \gamma_1 = \gamma_2\} \text{ for all } p \in P. \quad (44)$$

In general, $\Gamma_0(p)$ is always at least as large as (44), since it is never possible to detect the disease gene when all three penetrance probabilities are equal. However, for some pedigree types (e.g. unilineal affected pairs, see below) $\Gamma_0(p)$ may be strictly larger than (44).

Motivated by (44), it is convenient to reparametrize the penetrance parameters. To this end, for each fixed p , introduce the scalar product $(x, y) = q^2 x_0 y_0 + 2qp x_1 y_1 + p^2 x_2 y_2$ on \mathbb{R}^3 , define the three orthonormal unit vectors $e_0 = (1, 1, 1)$, $e_1 = (-2p, q - p, 2q)/\sqrt{2pq}$ and $e_2 = (q^{-1} - 1, -1, p^{-1} - 1)$. Then write

$$\gamma = K e_0 + \varepsilon_1 e_1 + \varepsilon_2 e_2. \quad (45)$$

One notices that $K = E(\Phi_k)$ is the prevalence of the disease and the genetic variance $\text{Var}(E(\Phi_k|G_k)) = \varepsilon_1^2 + \varepsilon_2^2$ is split into additive and dominance components ε_1^2 and ε_2^2 .

Formulas (43) and (45) suggest $q = 3$, with

$$t = (\tau, p, K). \quad (46)$$

However, $\xi = (\varepsilon_1, \varepsilon_2)$ yields a singular estimation problem¹ $I(t) = 0$ for all $t \in \mathbb{T} = [0, T] \times [0, 1]^2$. To alleviate this, put

$$\xi = (\xi_1, \xi_2) = (\varepsilon_1^2, \varepsilon_2^2), \text{ where } \varepsilon_1, \varepsilon_2 \geq 0, \quad (47)$$

and

$$\Xi(t) = \{\xi \in [0, \infty) \times [0, \infty); \gamma(\xi; K, p) \in \Gamma\}, \quad (48)$$

where $\gamma(\xi; K, p) = K e_0 + \sqrt{\xi_1} e_1 + \sqrt{\xi_2} e_2$. Hence $p = r = 2$, and the null parameter corresponds to $\xi_0 = (0, 0)$ according to (43)-(44).

Notice that the restriction $\varepsilon_1, \varepsilon_2 \geq 0$ reduces Θ . It is made just for convenience and does not affect the asymptotic distribution of λ_n . The reason is that all penetrance vectors $K e_0 \pm \varepsilon_1 e_1 \pm \varepsilon_2 e_2$ essentially give the same likelihood when ε_1 and ε_2 are small (i.e. θ close to Θ_0), see McPeck (1999) and Hössjer (2003, 2005a) for details.

¹This is in contrast to Section 4, where $I(t)$ was singular only at a finite number of points.

It can be deduced from Hössjer (2005a), that the likelihood score function is

$$\psi(x; t) = S(v(\tau); K), \quad (49)$$

where $S = (S_1, S_2)$ has components

$$\begin{aligned} S_1(v) &= \sum_{1 \leq k < l \leq N} \omega_{kl} \text{IBD}_{kl} / 2 - C_1, \\ S_2(v) &= \sum_{1 \leq k < l \leq N} \omega_{kl} 1_{\{\text{IBD}_{kl}=2\}} - C_2, \end{aligned} \quad (50)$$

$\text{IBD}_{kl}(v)$ is the number of founder alleles shared identical-by-descent by k and l and C_i is a centering constant assuring that $E_0(S_i(v)) = 2^{-m} \sum_v S_i(v) = 0$. The weight ω_{kl} for each pair (k, l) of individuals is defined as

$$\omega_{kl} = \frac{\partial^2 P(\Phi|G) / \partial \gamma_{|G_k|} \partial \gamma_{|G_l|} \Big|_{\gamma=(K,K,K)}}{P(\Phi)},$$

In particular, for binary phenotypes without polygenic and shared environmental effects, it follows from (40) and (41) that

$$\omega_{kl} = \begin{cases} K^{-2}, & \text{if } \Phi_k = \Phi_l = 1, \\ -K^{-1}(1-K)^{-1}, & \text{if } \Phi_k = 1, \Phi_l = 0 \text{ or } \Phi_k = 0, \Phi_l = 1, \\ (1-K)^{-2}, & \text{if } \Phi_k = \Phi_l = 0, \\ 0, & \text{otherwise.} \end{cases} \quad (51)$$

Inserting (49) into (7) and (10), we obtain

$$\begin{aligned} I(t, t') &= E_0 \left(S(v(\tau); K)^T S(v(\tau'); K') \right) \\ &= \sum_{v, v'} S(v; K)^T S(v'; K') P(v(\tau) = v, v(\tau') = v'), \end{aligned} \quad (52)$$

with t as in (46), $t' = (\tau', p', K')$ and the sum ranges of all $v, v' \in \{0, 1\}^m$.

Suppose now that $I(t)$ is non-singular for all t . This typically requires that (44) holds. In addition, we require that $\epsilon \leq p, K \leq 1 - \epsilon$ for some $\epsilon > 0$. This is in order to make the transformation $\gamma \rightarrow (K, \xi)$ and the weights ω_{kl} in (51) well defined. Then Z is also well defined with covariance function ρ obtained from (11), (21) and (52). We notice in particular that $\rho(t, t')$ does not depend on p and p' . From (22), (47) and (48) follows $V = \text{Id}$, the identity matrix of order 2, so that $\Delta(t) = C(A(t)^{-1})$ does not depend on p either. Hence from the definition of Y in (25) we deduce² $Y(t) = Y(\tau, K)$, so Y can be viewed as a process on $[0, T] \times [\epsilon, 1 - \epsilon]$ with marginal distribution

$$Y(\tau, K) \in (0.5 - w_2(\tau, K))\chi^2(0) + 0.5\chi^2(1) + w_2(\tau, K)\chi^2(2), \quad (53)$$

²When 1 is the disease allele and 0 the normal allele, one may restrict $\Xi(t)$ in (48) by putting $\Gamma = \{\gamma; \gamma_0 \leq \gamma_1 \leq \gamma_2\}$. It may be shown that this results in a region $\Delta(t)$ depending on p , so that $Y(t) = Y(\tau, K)$ no longer holds.

where $w_2(\tau, K) = \cos^{-1} \left(I_{12}(t) / \sqrt{I_{11}(t)I_{22}(t)} \right) / 2\pi$ is deduced from (28). The asymptotic LR-statistic is

$$\lambda = \sup_{\tau, K} Y(\tau, K), \quad (54)$$

where the supremum ranges over $\tau, K \in [0, T] \times [\epsilon, 1 - \epsilon]$.

Since $I(t)$ is non-singular for all $t \in \mathbb{T}$, we would like to apply Theorem 1. A technical difficulty, due to the reparametrization $\varepsilon \rightarrow \xi$, is that $\log f(x; t, \xi)$ is not twice differentiable with respect to ξ , so that iii and vi are violated. They are used for deriving an asymptotically quadratic expansion of the log likelihood in the proof of Theorem 1. However, it is possible to show that this expansion still holds:

Theorem 3 (Quadratic log likelihood expansion for linkage.) *The logarithm of the linkage likelihood function (2), with $J = 1$ and $f_1(x; \theta) = f(x; \theta)$ as in (37), admits a quadratic expansion (A.5), with Z_n defined with score function (49), $I(t) = I(t, t)$ as in (52), $A(t)$ given by (11) and R_n asymptotically negligible.*

Example 6 (MOD scores with only affecteds.) The covariance function $\rho(t, t')$, as well as the cone $\Delta(t)$ simplifies considerably when there are no unaffecteds in Φ , since then

$$S(v) = K^{-2}(S_{\text{pairs}}(v) - E_0(S_{\text{pairs}}), S_{\text{g-prs}}(v) - E_0(S_{\text{g-prs}})), \quad (55)$$

where $S_{\text{pairs}} = \sum_{k < l} \text{IBD}_{kl} / 2$ (Whittemore and Halpern, 1994), $S_{\text{g-prs}} = \sum_{k < l} 1_{\{\text{IBD}_{kl}=2\}}$ (McPeck, 1999) and the sums ranges over all pairs (k, l) of affecteds. Using that $\{v(s); 0 \leq s \leq T\}$ is a stationary Markov process and that K^{-2} only enters as a multiplicative constant in (55), we obtain

$$\rho(t, t') = \rho(\tau' - \tau) \quad (56)$$

from (52) and (21). That is, $Z(t) = Z(\tau)$ can be viewed as a stationary Gaussian process on $[0, T]$. It is shown in the appendix that

$$\rho(s) = \sum_{l=1}^m \kappa_l \exp(-2l|s|), \quad (57)$$

where $\kappa_1, \dots, \kappa_m$ are 2×2 -matrices. This corresponds to a stationary diffusion process on $[0, T]$.

In view of (52) and (55), we may choose a square root $A(t) = K^{-2}B$ of $I(t)$ such that B is a constant, independent of t . This implies $\Delta(t) = \Delta = C(B^{-1})$ and $Y(t) = Y(\tau)$. Hence we can view Y as a stationary process along $[0, T]$ with marginal distribution (53) and $\chi^2(2)$ -weight

$$w_2 = \cos^{-1} \left(I_{12}(t) / \sqrt{I_{11}(t)I_{22}(t)} \right) / 2\pi. \quad (58)$$

The asymptotic distribution of the LR-statistic is

$$\lambda = \sup_{0 \leq \tau \leq T} Y(\tau). \quad (59)$$

Figure 1 shows a number of pedigree types with affecteds. The associated $I(t)$, w_2 , Δ and κ_l are given in Tables 1-2. In particular, for affected sibling pedigrees (pedigree types 1-4), the two components of Z are independent Ornstein-Uhlenbeck processes. \square

An affected unilineal pair (k, l) is a pedigree type such that k and l are both affected, all other pedigree members have unknown phenotype and, moreover, (k, l) is unilineal, i.e. can share no more than one allele IBD. Examples are affected first cousins and uncle-nephew pairs, but not affected sib pairs. For affected unilineal pairs, it turns out that $I(t)$ singular. In this case another parametrization $\theta = (t, \xi)$ is needed, with $\mathbb{T} \subset \mathbb{R}^4$ and $\Xi(t) \subset \mathbb{R}$. It turns out that the limit process Z satisfies $Z(t) = Z(\tau)$, so that (59) holds, with $p = r = 1$, $\Delta(t) = [0, \infty)$ and $Y(\tau)$ a 0.5:0.5 mixture of $\chi^2(0)$ and $\chi^2(1)$. Formally, this corresponds to $w_2 = 0$.

Example 7 (Affected sib pairs and MLS scores.) It is interesting to note that for affected sib pairs (pedigree type 1 in Figure 1), there is another more compact parametrization $\theta = (t, \xi)$ that leads to an equivalent formulation of the MOD score, referred to as the MLS score. The reason is that p and the three penetrance parameters are confounded, and can be replaced by two parameters. Put

$$t = \tau, \quad \xi = (z_0, z_1), \quad (60)$$

where z_i is the probability that the ASP shares i alleles IBD at τ . The condensed vector ξ in (60) can be written as an explicit function of K , ε_1^2 and ε_2^2 , see Suarez et al. (1978). We get

$$f(x; \theta) = P_\xi(\text{IBD}(t) | \Phi) P(v(t) | \text{IBD}(t)) P(x | v(t)), \quad (61)$$

where $\text{IBD}(t)$ is the number of alleles shared IBD by the affected sib pair at t ,

$$P_{\xi}(\text{IBD}|\Phi) = z_0^{\{\text{IBD}=0\}} z_1^{\{\text{IBD}=1\}} (1 - z_0 - z_1)^{\{\text{IBD}=2\}},$$

and $P(v(t)|\text{IBD}(t)) = 1/8$ if $\text{IBD}(t) = 1$ and $1/4$ if $\text{IBD}(t) = 0$ or 2 . Moreover, $\xi_0 = (0.25, 0.5)$ and

$$\Xi = \{\xi; z_0 \geq 0, z_1 \leq 0.5 \text{ and } 2z_0 \leq z_1\} \quad (62)$$

is the possible triangle (Holmans, 1993). The parametrization (60) is more convenient than (47) in the sense that Theorem 1 can be applied directly. On the other hand, (60) cannot be extended to more general pedigree structures.

To verify that the MLS score calculation of $\{\kappa_l\}_{l=1}^m$ and Δ agree with the MOD score calculation for sib pairs in Tables 1-2, we first differentiate the logarithm of (61) at ξ_0 and find that

$$\begin{aligned} \psi(x; t) &= \left(4 \cdot 1_{\{\text{IBD}(t)=0\}} - 4 \cdot 1_{\{\text{IBD}(t)=2\}}, 2 \cdot 1_{\{\text{IBD}(t)=0\}} - 4 \cdot 1_{\{\text{IBD}(t)=2\}} \right) \\ &= \begin{pmatrix} -8 & 4 \\ 0 & -8 \end{pmatrix} (S_{\text{pairs}}(v(t)) - E_0(S_{\text{pairs}}), S_{\text{g-prs}}(v(t)) - E_0(S_{\text{g-prs}})) \\ &=: S(v(t)). \end{aligned} \quad (63)$$

Hence, applying (52), with S as in (63), we obtain

$$I(t) = \begin{pmatrix} 8 & 4 \\ 4 & 6 \end{pmatrix}. \quad (64)$$

If A is chosen as the upper triangular square root of $I(t)$, it follows that the coefficients κ_l agree with those in Table 2 for affected sib pairs. From (22) and (62) we find that

$$V = \begin{pmatrix} -2 & 1 \\ 0 & -1 \end{pmatrix}. \quad (65)$$

Inserting (64) and (65) into (23) and (28), and normalizing the rows of U (where $\Delta = C(U)$) to have unit length, we arrive at

$$U = \begin{pmatrix} -0.5774 & 0.8165 \\ 0 & -1.0000 \end{pmatrix}, \quad w_2 = 0.0980,$$

which is equivalent to the affected sib pair entry of Table 1. \square

Throughout the section we have considered a single type of pedigree and phenotype vector. More generally, we may assume $J > 1$ family types $(\mathcal{P}_1, \Phi_1), \dots, (\mathcal{P}_J, \Phi_J)$, with n_j families of type j and

$$f_j(x; \theta) = f(x; \theta, \mathcal{P}_j, \Phi_j).$$

Example 8 (MOD score with affected sib pairs and first cousins.)

Put $J = 2$ with asymptotic proportions $\pi_1 = \pi$ of affected sib pairs and $\pi_2 = 1 - \pi$ of affected first cousins. Then

$$I(t) = K^{-4} \begin{pmatrix} 0.125\pi + 0.0469(1 - \pi) & 0.125\pi \\ 0.125\pi & 0.1875\pi \end{pmatrix}.$$

The $\chi^2(2)$ -weight (58) of the marginal distribution of Y is plotted as a function of π in Figure 2. It is surprising that $w_2 \rightarrow 0.25$ as $\pi \rightarrow 0+$, whereas $\pi = 0$ corresponds to the degenerate case described above with $w_2 = 0$. This discontinuity of the limit distribution at $\pi = 0$ suggests that very large sample sizes are needed when π is positive but small in order for $\mathcal{L}(\lambda_n)$ to be close to its asymptotic distribution (59). \square

7 Discussion

In this paper, we have provided a general framework for LR-tests when part of the parameter vector under the null hypothesis is non-identifiable and another part is at the boundary of the parameter space. Under appropriate regularity conditions, the limit distribution of twice the log likelihood ratio statistic was found to equal the supremum of a $\bar{\chi}^2$ -process. The theory was illustrated with several examples (all having $p = 1$ or 2), including testing of normal mixtures and tests for presence of a disease gene in linkage analysis.

Examples with $p > 2$ are mixtures of normals with more than two components and genetic linkage analysis for models involving two disease genes, where $\tau = (\tau_1, \tau_2)$ contains the locations of the two genes, $\mathbf{p} = (p_1, p_2)$ the two disease allele frequencies and γ contains nine penetrance parameters (assuming a binary phenotype).

We have assumed a simple null hypothesis (4). A possible extension is to assume $\theta = (t, \xi, \eta) \in \mathbb{R}^{q+p+s}$, with

$$\begin{aligned} H_0 : & \xi = \xi_0, \\ H_1 : & \xi \neq \xi_0 \end{aligned}$$

as before, η an s -dimensional nuisance parameter and the non-identifiability condition (4) changed to

$$f_j(t, \xi_0, \eta) = f_j(t', \xi_0, \eta) \text{ for all } t, t' \in \mathbb{T}, \eta$$

and $j = 1, \dots, J$. In this case Θ_0 is s -dimensional. We conjecture that our framework can be extended naturally if η is an inner point and ξ_0 a boundary point of Θ_0 . What changes is mainly the definition of $I(t, t')$ in (10), which affects the covariance function (21) of $Z \in D^p(\mathbb{T})$ as well as the form (23) of the cone $\Delta(t)$. The modified $I(t, t')$ will involve the covariance structure of the score functions with respect to ξ and η . A more complicated extension is when both η and ξ are boundary points of Θ_0 , see Self and Liang (1987) for examples when $q = 0$.

Hypothesis testing for change-point problems is very similar to our setting. When $J = 1$, the basic difference is that a triangular array of random variables $\{X_{ni}; i = 1, \dots, n, n = 1, 2, 3, \dots\}$ is used with $\theta = (t, \xi)$, $\mathbb{T} = [0, T]$, hypothesis test (5) and model

$$f_{X_{ni}}(x) = \begin{cases} f(x; \xi_0), & i \leq [tn/T], \\ f(x; \xi), & i > [tn/T], \end{cases} \quad (66)$$

see e.g. Siegmund (1985). Hence t is the change-point and ξ_0 and ξ the parameter of data before and after the change-point. Non-identifiability (4) follows immediately from (66) and ξ_0 may or may not be on the boundary of Ξ . The dimension p of Ξ may be 1, 2, 3 or larger and hence a broad class of $\bar{\chi}^2$ -process are of interest for change-point problems. Typically, the limit process Z is not differentiable but of diffusion type, as in Section 6.

Appendix

Distance metric between sets. Given any two sets $\Delta, \Delta' \subset \mathbb{R}^p$ and $K > 0$, define

$$d_K(\Delta, \Delta') = \max \left(\sup_{z \in \Delta \cap B(0, K)} |z - P_{\Delta'} z|, \sup_{z \in \Delta' \cap B(0, K)} |z - P_{\Delta} z| \right)$$

and

$$d(\Delta, \Delta') = \sum_{K=1}^{\infty} 2^{-K} \min(d_K(\Delta, \Delta'), 1). \quad (\text{A.1})$$

Then d defines a metric between sets and (22) is equivalent to Chernoff regularity (Chernoff, 1954). This is equivalent to $C(V)$ being equal to the

ordinary and derivable tangent cone of Ξ at ξ_0 , see Geyer (1994), Theorem 2.1 and Shapiro (2000) for definitions.

Lemma 1 *Given v and (22), (24) follows. Moreover, for any $K > 0$,*

$$\lim_{n \rightarrow \infty} \sup_{t \in \mathbb{T}, |z| \leq K} \left| |z - P_{\Delta_n(t)} z|^2 - |z - P_{\Delta(t)} z|^2 \right| = 0. \quad (\text{A.2})$$

Proof. To prove (24), it suffices to establish

$$\lim_{n \rightarrow \infty} \sup_{t \in \mathbb{T}} d_K(\Delta_n(t), \Delta(t)) = 0, \quad (\text{A.3})$$

for any $K > 0$. Suppose $z' \in \Delta_n(t)$ and $z \in \Delta(t)$, with $|z'|, |z| \leq K$. Let $C(t) = C(V(t))$ and $C_\tau(t) = (\Xi - \xi_0)/\tau$. Then $z' = y'A(t)^T$ and $z = yA(t)^T$, with $y' \in C_{1/\sqrt{n}}(t)$, $y \in C(t)$ and $|y|, |y'| \leq K/\sqrt{\kappa_{\min}(t)}$. Moreover, if $\kappa_{\max}(t)$ is the largest eigenvalue of $I(t)$,

$$|z' - z|^2 = (y' - y)I(t)(y' - y) \leq \kappa_{\max}(t)|y' - y|^2.$$

Since $t \rightarrow I(t)$ continuous (cf. iv) it follows that $t \rightarrow \kappa_{\max}(t)$ is continuous as well and so $\kappa_{\max} = \sup_{t \in \mathbb{T}} \kappa_{\max}(t) < \infty$ by the compactness of \mathbb{T} . Hence

$$\sup_{t \in \mathbb{T}} d_K(\Delta_n(t), \Delta(t)) \leq \sqrt{\kappa_{\max}} \sup_{t \in \mathbb{T}} d_{K/\sqrt{\kappa_{\min}}}(C_{1/\sqrt{n}}(t), C(t)),$$

which in conjunction with (22) implies (A.3).

To prove (A.2), introduce $d_{nK} = \sup_{t \in \mathbb{T}} d_K(\Delta_n(t), \Delta(t))$. Then, given any $z \in B(0, K)$ and $t \in \mathbb{T}$ we have $|z - P_{\Delta_n(t)} z| \leq K$ and $|z - P_{\Delta(t)} z| \leq K$ since $0 \in \Delta_n(t)$ and $0 \in \Delta(t)$. Hence

$$\left| |z - P_{\Delta_n(t)} z|^2 - |z - P_{\Delta(t)} z|^2 \right| \leq 2K \left| |z - P_{\Delta_n(t)} z| - |z - P_{\Delta(t)} z| \right| \leq 2K d_{nK}, \quad (\text{A.4})$$

where in the last step we used the definitions of $P_{\Delta_n(t)} z$ and $P_{\Delta(t)} z$. Since $\lim_{n \rightarrow \infty} d_{nK} = 0$ according to (24) and the upper bound in (A.4) does not involve z or t , (A.2) follows. \square

Proof of Theorem 1. The quadratic expansion of l_n is

$$\begin{aligned} l_n(t, \xi) &= l_n(t, \xi_0) + \sqrt{n} Z_n(t) A(t) (\xi - \xi_0)^T \\ &\quad - \frac{n}{2} (\xi - \xi_0) I(t) (\xi - \xi_0)^T + R_n(t, \xi), \end{aligned} \quad (\text{A.5})$$

where $R_n(t, \xi) = \tilde{R}_n(t, \sqrt{n}(\xi - \xi_0)A(t)^T)$. It follows from Taylor expansion of $l_n(t, \cdot)$, iii and vi that

$$|R_n(t, \xi)| \leq \frac{n}{2} |(\xi - \xi_0)(\hat{I}_n(t) - I(t))(\xi - \xi_0)^T| + KnM_n |\xi - \xi_0|^3, \quad (\text{A.6})$$

with $M_n = \sum_{j=1}^J \sum_{i=1}^{n_j} M_j(X_{ji})/n$, for some constant $K > 0$ and $(t, \xi) \in \Theta_\varepsilon$ when $\varepsilon > 0$ is small enough. Let $M = \sum_{j=1}^J \pi_j E_{\theta_0}(M_j(X_{j1}))$. Applying (29), (9), iii, vi and the Law of Large Numbers we get

$$|R_n(t, \xi)| \leq o_p(n)|\xi - \xi_0|^2 + Kn(M + o_p(1))|\xi - \xi_0|^3 \quad (\text{A.7})$$

uniformly for all $(t, \xi) \in \Theta_\varepsilon$. Using (20) and the Continuous Mapping Theorem, we find that

$$\|Z_n\| \xrightarrow{\mathcal{L}} \|Z\| \quad (\text{A.8})$$

as $n \rightarrow \infty$, where $\|Z_n\| = \sup_{t \in \mathbb{T}} |Z_n(t)|$ and $\|Z\| = \sup_{t \in \mathbb{T}} |Z(t)|$. Hence, if $2KM\varepsilon < \kappa_{\min}/4$, it follows from (A.5), (A.7), (A.8) and v that

$$l_n(t, \xi) - l_n(t, \xi_0) \leq 2\sqrt{n\kappa_{\max}}|Z||\xi - \xi_0| - \frac{n\kappa_{\min}}{4}|\xi - \xi_0|^2$$

uniformly for $(t, \xi) \in \Theta_\varepsilon$, with probability tending to one. This and i implies

$$\sup_{t \in \mathbb{T}} |\hat{\xi}_n(t) - \xi_0| = O_p(n^{-1/2}). \quad (\text{A.9})$$

Fix $K > 0$ and define

$$\Upsilon_{K,n} = \{(t, \delta), t \in \mathbb{T}, \delta \in \Delta_n(t) \text{ and } |\delta| \leq K\}. \quad (\text{A.10})$$

Put also $\hat{\delta}_n(t) = \sqrt{n}(\hat{\xi}_n(t) - \xi_0)A(t)^T$ and $\|\hat{\delta}_n\| = \sup_{t \in \mathbb{T}} |\hat{\delta}_n(t)|$. Then (A.9) implies that $P(|\hat{\delta}_n| \leq K)$ can be made arbitrarily close to one (uniformly in n) provided K is chosen large enough.

Define $\tilde{Y}_n = \{\tilde{Y}_n(t); t \in \mathbb{T}\}$, where $\tilde{Y}_n(t)$ is obtained by ignoring the remainder term in (19), i.e.

$$\begin{aligned} \tilde{Y}_n(t) &= \sup_{\delta \in \Delta_n(t)} (2Z_n(t)\delta^T - |\delta|^2) \\ &= |Z_n(t)|^2 - |Z_n(t) - P_{\Delta_n(t)}Z_n(t)|^2. \end{aligned} \quad (\text{A.11})$$

Let $\tilde{\delta}_n(t)$ be the argmax of (A.11). It is easy to see that $\|Z_n\| \leq K/2$ implies $\|\tilde{\delta}_n\| = \sup_{t \in \mathbb{T}} |\tilde{\delta}_n(t)| \leq K$. Assume $\|\hat{\delta}_n\|, \|\tilde{\delta}_n\| \leq K$. (According to (A.8) and the discussion below (A.10), this event has an arbitrarily small probability, uniformly in n , provided K is chosen large enough.). It follows that

$$\|\tilde{Y}_n - Y_n\| = \sup_{t \in \mathbb{T}} |\tilde{Y}_n(t) - Y_n(t)| \leq 2 \sup_{(t, \delta) \in \Upsilon_{K,n}} |\tilde{R}_n(t, \delta)| = o_p(1), \quad (\text{A.12})$$

where the last identity follows from (A.7), v and $\tilde{R}_n(t, \delta) = R_n(t, \xi_0 + n^{-1/2}\delta A(t)^{-T})$. Define $\check{Y}_n(t)$ by replacing $\Delta_n(t)$ with $\Delta(t)$ in (A.11) i.e.

$$\check{Y}_n(t) = |Z_n(t)|^2 - |Z_n(t) - P_{\Delta(t)}Z_n(t)|^2 = |P_{\Delta(t)}Z_n(t)|^2.$$

In the last equality, we used that $\Delta(t)$ is a cone. It follows from (A.8) and (A.2) that

$$\|\check{Y}_n - \tilde{Y}_n\| = o_p(1). \quad (\text{A.13})$$

Define the mapping $g : D^p(\mathbb{T}) \rightarrow D(\mathbb{T})$ through $g(Z)(t) = |P_{\Delta(t)}Z(t)|^2$ if $Z = \{Z(t); t \in \mathbb{T}\}$. The fact that $g(Z) \in D(\mathbb{T})$ follows from continuity of $A(t)$ and $V(t)$ (see iv and vii). For any $z_1, z_2 \in \mathbb{R}^p$ and a convex cone Δ bounded by hyperplanes we have

$$|P_{\Delta}z_2 - P_{\Delta}z_1| \leq |z_2 - z_1|. \quad (\text{A.14})$$

Since $P_{\Delta}0 = 0$, this implies $|P_{\Delta}z| \leq |z|$. Hence

$$\begin{aligned} \|g(Z_2) - g(Z_1)\| &= \sup_{t \in \mathbb{T}} \left| |P_{\Delta(t)}Z_2(t)|^2 - |P_{\Delta(t)}Z_1(t)|^2 \right| \\ &\leq \sup_{t \in \mathbb{T}} (|Z_1(t)| + |Z_2(t)|) \left| |P_{\Delta(t)}Z_2(t)| - |P_{\Delta(t)}Z_1(t)| \right| \\ &\leq (\|Z_1\| + \|Z_2\|) \|Z_2 - Z_1\|, \end{aligned}$$

so that g is continuous with respect to supremum norm. A slight modification of the argument proves that g is continuous with respect to the Skorohod topology as well. Since $\check{Y}_n = g(Z_n)$ and $Y = g(Z)$, (20) and the Continuous Mapping Theorem imply

$$\check{Y}_n \xrightarrow{\mathcal{L}} Y \text{ as } n \rightarrow \infty. \quad (\text{A.15})$$

Equations (A.12), (A.13) and (A.15) and Slutsky's Theorem prove (15). Another application of the Continuous Mapping Theorem (with $g : D(\mathbb{T}) \rightarrow \mathbb{R}$, $g(y) = \sup_{t \in \mathbb{T}} y(t)$, $\lambda_n = g(Y_n)$ and $\lambda = g(Y)$) proves (6). \square

Proof of Theorem 2. Since \mathbb{T}_s is finite and $Y_n \in D(\mathbb{T})$ according to Remark 5, it follows that

$$\lambda_n = \sup_{t \in \mathbb{T}_{ns}} Y_n(t). \quad (\text{A.16})$$

As in Theorem 1, the main idea of the proof is to use a locally quadratic expansion $Y_n(t)$. In view of (A.16), it suffices to consider such an expansion for $t \in \mathbb{T}_{ns}$. Hence we can use (18), which is only defined for $t \in \mathbb{T}_{ns}$. Define $\tilde{M}_n = \sum_{j=1}^J \sum_{i=1}^{n_j} \tilde{M}_j(X_{ji})/n$. Then, by iii' and vi', the remainder term in (18) satisfies

$$|\tilde{R}_n(t, \delta)| \leq \frac{1}{2} |\delta(\widehat{\text{Id}}_n(t) - \text{Id})\delta^T| + Kn^{-1/2} \tilde{M}_n |\delta|^3 \quad (\text{A.17})$$

for some constant $K > 0$, uniformly on

$$\Upsilon_{\varepsilon\sqrt{n}, n} = \{(t, \delta); t \in \mathbb{T}_{ns}, \delta \in \Delta_n(t) \text{ and } |\delta| \leq \varepsilon\sqrt{n}\},$$

provided $\varepsilon > 0$ is small enough. Let $\tilde{M} = \sum_{j=1}^J \pi_j E_{\theta_0}(\tilde{M}_j(X_{j1}))$. Then (A.17), (9), iii', vi' and the Law of Large Numbers imply

$$|\tilde{R}_n(t, \delta)| \leq o_p(1)|\delta|^2 + Kn^{-1/2}(\tilde{M} + o_p(1))|\delta|^3$$

uniformly for all $(t, \delta) \in \Upsilon_{\varepsilon\sqrt{n}, n}$. As in the proof of Theorem 1 we use (A.8) to conclude that

$$\tilde{l}_n(t, \delta) - \tilde{l}_n(t, 0) \leq 2|Z||\delta| - \frac{1}{4}|\delta|^2$$

uniformly for all $(t, \varepsilon) \in \Upsilon_{\varepsilon\sqrt{n}, n}$, with probability tending to one. Together with i' this proves

$$\sup_{t \in \mathbb{T}_{ns}} |\hat{\delta}_n(t)| = O_p(1).$$

Then we argue as in the proof of Theorem 1 to deduce (A.12), which in turn implies

$$\tilde{\lambda}_n - \lambda_n = o_p(1), \quad (\text{A.18})$$

where $\tilde{\lambda}_n = \sup_{t \in \mathbb{T}_{ns}} \tilde{Y}_n(t)$. We will prove below that

$$\check{\lambda}_n - \tilde{\lambda}_n = o_p(1), \quad (\text{A.19})$$

where $\check{\lambda}_n = \sup_{t \in \mathbb{T}} \check{Y}_n(t)$. Since $Z_n, Z \in D^p(\mathbb{T})$ according to Remark 5, we can use (20) and the Continuous Mapping Theorem to establish (A.15), as in the proof of Theorem 1. Another application of the Continuous Mapping Theorem implies

$$\check{\lambda}_n \xrightarrow{\mathcal{L}} \lambda. \quad (\text{A.20})$$

Combining (A.18), (A.19), (A.20) and Slutsky's Lemma we obtain (6).

Hence it remains to prove (A.19). Assume without loss of generality that $\mathbb{T}_s = \{0\}$ and that 0 is an inner point of \mathbb{T} . Put $\tilde{\lambda}_{\varepsilon n} = \sup_{t \in \mathbb{T}_\varepsilon} \tilde{Y}_n(t)$ and $\check{\lambda}_{\varepsilon n} = \sup_{t \in \mathbb{T}_\varepsilon} \check{Y}_n(t)$, with \mathbb{T}_ε as in vii' and ε small enough for the ball around 0 with radius ε to be a subset of \mathbb{T} . It follows from vii' that $\Delta_n(t) \nearrow \Delta(t)$ is an increasing function of n (see also Remark 6). Hence

$$\begin{aligned} 0 &\leq \tilde{\lambda}_{\varepsilon n} \leq \tilde{\lambda}_n \leq \check{\lambda}_n, \\ 0 &\leq \check{\lambda}_{\varepsilon n} \leq \check{\lambda}_n \leq \tilde{\lambda}_n. \end{aligned} \quad (\text{A.21})$$

It follows from iv that $t \rightarrow I(t)$ is continuous and hence that $\min_{t \in \mathbb{T}_\varepsilon} \kappa_{\min}(t) > 0$. This and vii' implies (cf. (A.2) and its proof)

$$\lim_{n \rightarrow \infty} \sup_{t \in \mathbb{T}_\varepsilon, |z| \leq K} \left| |z - P_{\Delta_n(t)} z|^2 - |z - P_{\Delta(t)} z|^2 \right| = 0 \quad (\text{A.22})$$

for all $\varepsilon, K > 0$. Proceeding as in the proof of (A.13) in Theorem 1, (A.22) leads to

$$\tilde{\lambda}_{\varepsilon n} - \check{\lambda}_{\varepsilon n} = o_p(1). \quad (\text{A.23})$$

By (20) and the Continuous Mapping Theorem, it follows that

$$\check{\lambda}_{\varepsilon n} - \check{\lambda}_n \xrightarrow{\mathcal{L}} \lambda_\varepsilon - \lambda \quad (\text{A.24})$$

as $n \rightarrow \infty$, where $\lambda_\varepsilon = \sup_{t \in \mathbb{T}_\varepsilon} Y(t)$. The modulus of the right-hand side of (A.24) is upper bounded by

$$\begin{aligned} |\lambda_\varepsilon - \lambda| &\leq \sup_{|t|, |t'| \leq \varepsilon} |Y(t') - Y(t)| \\ &\leq 2 \sup_{|t| \leq \varepsilon} |Y(t) - Y(0)| \\ &\leq 2 \sup_{|t| \leq \varepsilon} \left| |P_{\Delta(t)}(Z(t))|^2 - |P_{\Delta(t)}Z(0)|^2 \right| \\ &\quad + 2 \sup_{|t| \leq \varepsilon} \left| |P_{\Delta(t)}(Z(0))|^2 - |P_{\Delta(0)}Z(0)|^2 \right| \\ &\leq 4\|Z\|\omega_Z(\varepsilon) + 4\|Z\| \sup_{|t| \leq \varepsilon} d_{Z(0)}(\Delta(t), \Delta(0)), \end{aligned} \quad (\text{A.25})$$

where $\omega_Z(\varepsilon) = \sup_{|t| \leq \varepsilon} |Z(t) - Z(0)|$. In the last inequality of (A.25) we used (A.14) for the first term and the same estimate as in the last step of (A.4) for the second term. Combining (A.21)-(A.25), we find that

$$\begin{aligned} \overline{\lim}_{n \rightarrow \infty} P(|\check{\lambda}_n - \tilde{\lambda}_n| > \varepsilon) &\leq \overline{\lim}_{n \rightarrow \infty} P(|\check{\lambda}_n - \tilde{\lambda}_{\varepsilon n}| > \varepsilon) \\ &\leq \overline{\lim}_{n \rightarrow \infty} P(|\check{\lambda}_n - \tilde{\lambda}_{\varepsilon n}| > \varepsilon/2) + \overline{\lim}_{n \rightarrow \infty} P(|\tilde{\lambda}_{\varepsilon n} - \check{\lambda}_{\varepsilon n}| > \varepsilon/2) \\ &\leq P(|\lambda_\varepsilon - \lambda| > \varepsilon/4) + 0 \\ &\leq P(\|Z\| \geq K) + P(\omega_Z(\varepsilon) > \varepsilon/(32K)) \\ &\quad + P\left(\sup_{|t| \leq \varepsilon} d_K(\Delta(t), \Delta(0)) > \varepsilon/(32K)\right). \end{aligned} \quad (\text{A.26})$$

According to iv' and vii', given any $\varepsilon > 0$, the right-hand side of (A.26) can be made arbitrarily small by first choosing K large and then ε small. This proves (A.19) and hence also Theorem 2. \square

Proof of Theorem 3. With a slight abuse of notation, let $l_n(t, \varepsilon)$ denote the log likelihood with parametrization t as in (46) and $\varepsilon = (\varepsilon_1, \varepsilon_2)$. Following Rotnitzky et al. (2000), we perform a Taylor expansion of l_n up to order 4,

$$l_n(t, \varepsilon) = l_n(t; 0) + \sum_{k; 1 \leq |k| \leq 4} l_n^{(k)}(t, 0) \frac{\varepsilon^k}{k!} + R_n(t; \varepsilon), \quad (\text{A.27})$$

where $0 = (0, 0)$, $l_n^{(k)}(t, \varepsilon) = \partial^k l_n(t, \varepsilon) / \partial \varepsilon^k$, $k = (k_1, k_2)$ and $\varepsilon^k = \varepsilon_1^{k_1} \varepsilon_2^{k_2}$. Then

$$l_n^{(k)}(t, 0) = \sum_{i=1}^n \psi_k(X_i; t),$$

where

$$\psi_k(x; t) = \frac{\partial^k \log f(x; t, \varepsilon)}{\partial \varepsilon^k} = \frac{\partial^k \log P_{p, \gamma}(v(\tau) | \Phi)}{\partial \varepsilon^k}. \quad (\text{A.28})$$

In the last step of (A.28) we used (37), with γ as in (45). Introduce the abbreviated notation $P(v) = P_{p, \gamma}(v | \Phi)|_{\varepsilon=0} = 2^{-m}$ and $P^{(k)}(v) = \partial^k P_{p, \gamma}(v | \Phi) / \partial \varepsilon^k|_{\varepsilon=0}$. Using results from Hössjer (2005a), we find, by repeated differentiation w.r.t. ε , that

$$\begin{aligned} \psi_{10}(x; t) &= 0, \\ \psi_{01}(x; t) &= 0, \\ \psi_{20}(x; t) &= P^{(20)}(v(\tau)) / P(v(\tau)) = 2S_1(v(\tau)), \\ \psi_{11}(x; t) &= 0, \\ \psi_{02}(x; t) &= P^{(02)}(v(\tau)) / P(v(\tau)) = 2S_2(v(\tau)), \\ \psi_{40}(x; t) &= -3I_{11}(t, t) + r_{40}(x; t), \\ \psi_{22}(x; t) &= -I_{12}(t, t) + r_{22}(x; t), \\ \psi_{04}(x; t) &= -3I_{22}(t, t) + r_{04}(x; t), \end{aligned}$$

with $S = (S_1, S_2)$ defined in (49), $I(t, t) = E_0(S^T(v)S(v))$ and $E(r_k(X; t)) = 0$ for $k = (4, 0)$, $(2, 2)$ and $(0, 4)$. Moreover, $E(\psi_k(X; t)) = 0$ when $|k| = 3$ and $k = (3, 1)$, $(1, 3)$. Substituting $\xi = (\xi_1, \xi_2) = (\varepsilon_1^2, \varepsilon_2^2)$ back into (A.27), we find that

$$l_n(t, \xi) = l_n(t, 0) + \left(\sum_{i=1}^n \psi(X_i; t) \right) \xi^T - \frac{n}{2} \xi I(t) \xi^T + R_n(t, \xi), \quad (\text{A.29})$$

where $\psi(x; t) = (\psi_{10}(x; t), \psi_{01}(x; t)) / 2$ agrees with (49),

$$\begin{aligned} R_n(t, \xi) &= \sum_{k \in I} l^{(k)}(t, 0) \frac{\varepsilon^k}{k!} + \sum_{k \in II} \frac{\varepsilon^k}{k!} \sum_{i=1}^n r_k(X_i; t) + R_n(t, \varepsilon) \\ &= o_p(|\xi|^{3/2} n^{1/2}) + R_n(t, \varepsilon), \end{aligned} \quad (\text{A.30})$$

$I = \{k; |k| = 3 \text{ or } k = (3, 1), (1, 3)\}$ and $II = \{(4, 0), (2, 2), (0, 4)\}$. From (16) and the fact that $\xi_0 = (0, 0)$ we deduce that (A.5) and (A.29) are identical.

In (A.30) we notice that $o_p(|\xi|^{3/2} n^{1/2})$ is small in comparison to $\sqrt{n} Z_n(t) A(t) \xi^T$ and $R_n(t, \varepsilon)$ can be made small uniformly w.r.t. all (t, ε) in a small neighborhood of $\mathbb{T} \times \{(0, 0)\}$ by imposing a regularity condition for all $l_n^{(k)}(t, \varepsilon)$ with $|k| = 5$, analogous to Condition vi of Section 3. This implies that $R_n(t, \xi)$ is asymptotically negligible.

□

Derivation of (57). We generalize Theorem 2 of Hössjer (2005b) from one- to two-dimensional score functions. We introduce the space \mathcal{A} of mappings

$\{0, 1\}^m \rightarrow \mathbb{R}$, so that both components of $S = (S_1, S_2)$ are elements of \mathcal{A} . Endow \mathcal{A} with the scalar product

$$\langle S_1, S_2 \rangle = 2^{-m} \sum_w S_1(w) S_2(w),$$

Given any $w \in \{0, 1\}^m$, let $S_w(u) = (-1)^{w \cdot u}$, where $w \cdot u = \sum_{l=1}^m w_l u_l$ is the vector dot product of w and u . Then $\{S_w\}$ is an orthonormal basis on \mathcal{A} and S can be expanded as

$$S = \sum_w R_S(w) S_w,$$

where $R_S(w) = (R_{S_1}(w), R_{S_2}(w))$ and $R_{S_k}(w) = \langle S_k, S_w \rangle$. Notice that $R_S(0) = 0$ since both S_1 and S_2 are standardized to have mean zero, under the null hypothesis of no linkage, i.e. $E_0(S_k) = \langle S_k, S_0 \rangle = 0$. It follows from (52) and the proof of Theorem 2 in Hössjer (2005b) that

$$I(t, t') = \sum_{w \neq 0} R_S^T(w) R_S(w) \exp(-2|w| |\tau' - \tau|), \quad (\text{A.31})$$

where $|w| = \sum_{l=1}^m w_l$ is the number of one-components of w . Combining (56) and (A.31) we arrive at (57), with

$$\begin{aligned} \kappa_l &= A(t)^{-T} \sum_{w; |w|=l} R_S^T(w) R_S(w) A(t)^{-1} \\ &= B^{-T} \sum_{w; |w|=l} R_{S_0}^T(w) R_{S_0}(w) B^{-1}. \end{aligned}$$

where $S_0 = K^2 S = (S_{\text{pairs}} - E_0(S_{\text{pairs}}), S_{\text{g-prs}} - E_0(S_{\text{g-prs}}))$ and $B = K^2 A(t)$ are both independent of t . \square

Acknowledgment

This work was supported by the Swedish Research Council, contract number 621-2005-2810. The author wishes to thank Annica Dominicus and Tobias Ryden for providing valuable references.

References

- Azaïs, J.-M., Gassiat, É. and Mercadier, C. (2004). Asymptotic distribution and power of the likelihood ratio test for mixtures: bounded and unbounded cases. Manuscript.
- Bickel, P. and Chernoff, H. (1993). Asymptotic distribution of the likelihood ratio statistic in a prototypical non regular problem. In Ghosh, J.K., Mitra,

- S.K., Parthasarathy, K.R., Prakasa Rao, B.L.S. (eds.) *Statistics and Probability: A Raghu Raj Bahadur Festschrift*. Wiley Eastern, New Dehli, pp. 83-96.
- Bickel, P. and Millar, P. (1992). Uniform convergence of probability measures on classes of functions. *Stat. Sinica* **2**, 1-15.
- Bickel, P. and Wichura, M. (1971). Convergence criteria for multiparameter stochastic processes and some applications. *Ann. Math. Statist.* **42**(5), 1656-1670.
- Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley, New York.
- Chen, H. and Chen, J. (2001). Large sample distribution of the likelihood ratio test for normal mixtures. *Stat. Prob. Letters* **52**, 125-133.
- Chernoff, H. (1954). On the distribution of the likelihood ratio. *Annals of Mathematical Statistics* **25**, 573-578.
- Clerget-Darpoux, F., Bonaïti-Pellié, C. and Hochez, J. (1986). Effects of misspecifying genetic parameters in lod score analysis. *Biometrics*, **42**, 393-399.
- Cramér, H. (1946). *Mathematical Methods of Statistics*, Princeton University Press, Princeton, N.J.
- Davies, R.B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **64**(2), 247-254.
- Delmas, C. (2003). On likelihood ratio tests in Gaussian mixture models. *Sankhya* **65**(3), 513-531.
- Dupuis, J., Brown, P.O. and Siegmund, D. (1995). Statistical methods for linkage analysis of complex traits from high-resolution maps of identity by descent. *Genetics* **140**, 843-856.
- Feingold, E., Brown, P.O. and Siegmund, D. (1993). Gaussian models for genetic linkage analysis using complete high-resolution maps of identity by descent. *Am. J. Hum. Genet.*, **53**, 234-251.
- Garel, B. (2001). Likelihood ratio test for univariate Gaussian mixture. *J. Stat. Plan. Inf.* **95**, 325-350.
- Geyer, C.J. (1994). On the asymptotic of constrained M -estimation. *Ann. Statist.* **22**(4), 1993-2010.
- Holmans, P. (1993). Asymptotic properties of affected-sib-pair linkage analysis. *Am. J. Hum. Genet.* **52**, 362-374.
- Hössjer, O. (2003). Determining inheritance distributions via stochastic penetrances. *J. Amer. Statist. Ass.* **98**, 1035-1051.

- Hössjer, O. (2005a). Conditional likelihood score functions in linkage analysis. *Biostatistics* **6**, 313-332.
- Hössjer, O. (2005b). Spectral decomposition of score functions in linkage analysis. *Bernoulli* **11**(6), 1093-1113.
- Lander, E.S. and Botstein, D. (1989). Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**, 185-199.
- Lemdani, M. and Pons, O. (1999). Likelihood ratio tests in contamination models. *Bernoulli* **5**(4), 705-719.
- Liu, X. and Shao, Y. (2004). Asymptotics for the likelihood ratio test in a two-component normal mixture model. *J. Stat. Plan. Inf.* **123**, 61-81.
- McPeck, S. (1999). Optimal allele-sharing statistics for genetic mapping using affected relatives. *Gen. Epidemiol.* **16**, 225-249.
- Risch, N. (1984). Segregation analysis incorporating genetic markers. I. Single-locus models with an application to type I diabetes. *Am. J. Hum. Genet.*, **36**, 363-386.
- Risch, N. (1990). Linkage strategies for genetically complex traits III. The effect of marker polymorphism and analysis of affected relative pairs. *Am. J. Hum. Genet.*, **46**, 242-253.
- Rotnitzky, A., Cox, D.R., Bottai, M. and Robins, J. (2000). Likelihood-based inference with singular information matrix. *Bernoulli* **6**, 243-284.
- Self, S. and Liang K-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J. Amer. Statist. Assoc.* **82**, 605-610.
- Sen, P.K. and Silvapulle, M.J. (2002). An appraisal of statistical inference under inequality constraints. *J. Stat. Plan. Inf.* **107**, 3-43.
- Sen, P.K. and Silvapulle, M.J. (2005). *Constrained Statistical Inference. Inequality, Order and Shape Restrictions*. Wiley, Hoboken, New Jersey.
- Serfling, R. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.
- Sham, P. (1998). *Statistics in Human Genetics*, Arnold Applications of Statistics, London.
- Shapiro, A. (1988). Towards a unified theory of inequality constrained testing in multivariate analysis. *Int. Statist. Rev.* **56**(1), 49-62.
- Shapiro, A. (2000). On the asymptotics of constrained local M -estimators. *Ann. Statist.* **28**(3), 948-960.
- Suarez, B.K., Rice, J. and Reich, T. (1978). The generalized sib pair IBD distribution: its use in detection of linkage. *Ann. Hum. Genet.*, **44**, 87-94.

Whittemore A.S. and Halpern, J. (1994). A class of tests for linkage using affected pedigree members. *Biometrics*, **50**, 118-127.

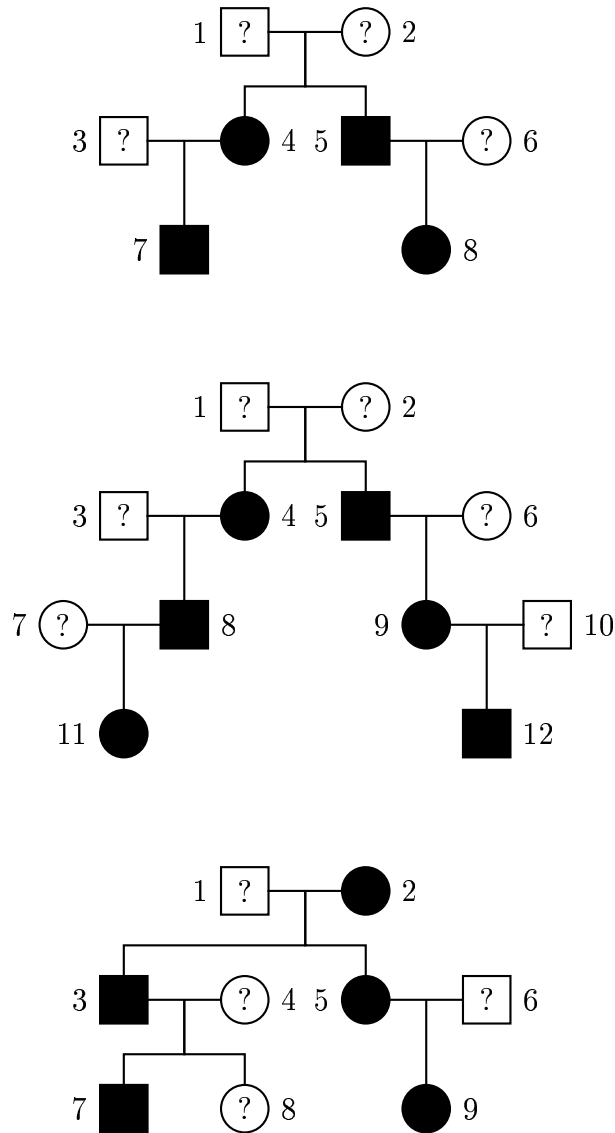


Figure 1: Pedigree types (\mathcal{P}_j, Φ_j) used in Tables 1-2. For $j = 1, 2, 3, 4$, \mathcal{P}_j consists of two parents with unknown phenotypes and $k + 1$ affected offspring. (In particular, (\mathcal{P}_1, Φ_1) is an affected sib pair.) \mathcal{P}_5 (upper), \mathcal{P}_6 (middle) and \mathcal{P}_7 (lower) are shown above with individual numbers. Males and females correspond to squares and circles, affected individuals have black and unaffected ones have white symbols. Individuals with unknown phenotypes have question marks.

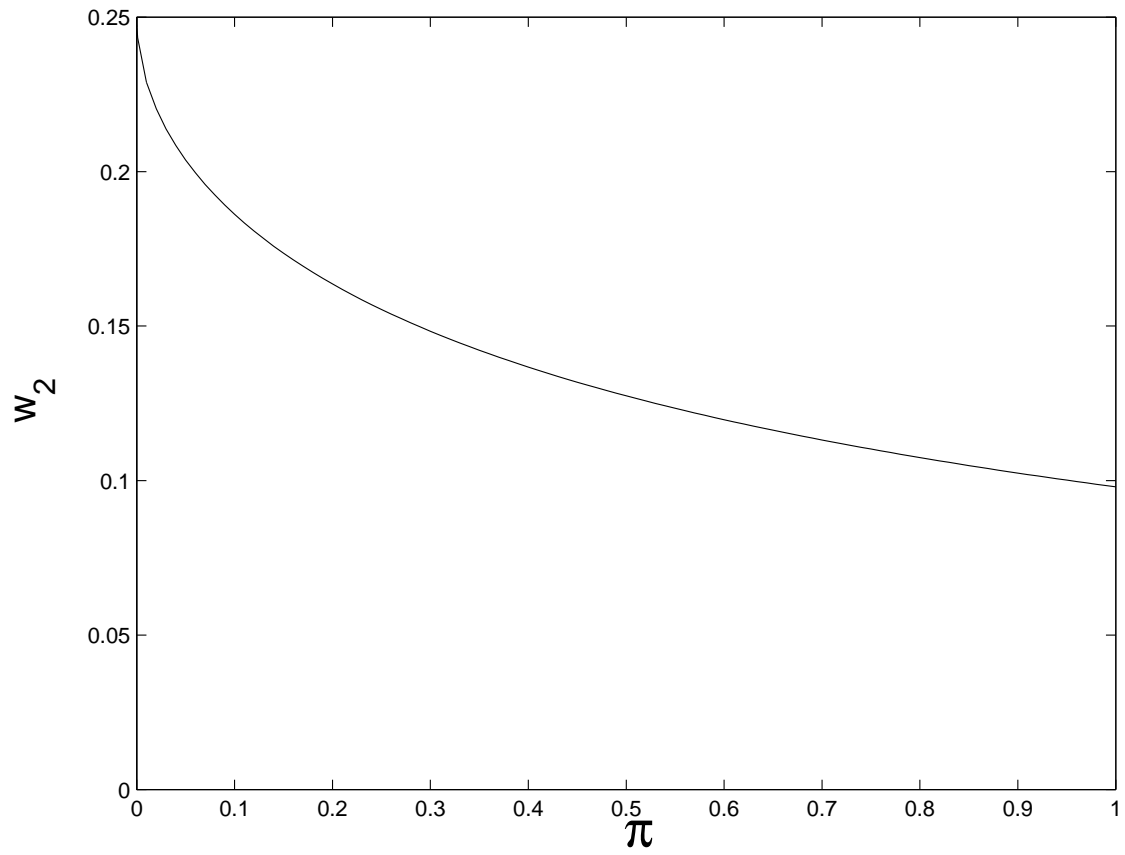


Figure 2: Plot of $\chi^2(2)$ -weight w_2 as function of the proportion π of affected sib pairs for a mixture of affected sib and first cousin pairs.

Pedigree type	$I(t)$	U	w_2
(\mathcal{P}_1, Φ_1)	$K^{-4} \cdot \begin{pmatrix} 0.1250 & 0.1250 \\ 0.1250 & 0.1875 \end{pmatrix}$	$\begin{pmatrix} 0.5774 & -0.8165 \\ 0 & 1.0000 \end{pmatrix}$	0.0980
(\mathcal{P}_2, Φ_2)	$K^{-4} \cdot \begin{pmatrix} 0.3750 & 0.3750 \\ 0.3750 & 0.5625 \end{pmatrix}$	$\begin{pmatrix} 0.5774 & -0.8165 \\ 0 & 1.0000 \end{pmatrix}$	0.0980
(\mathcal{P}_3, Φ_3)	$K^{-4} \cdot \begin{pmatrix} 0.7500 & 0.7500 \\ 0.7500 & 1.1250 \end{pmatrix}$	$\begin{pmatrix} 0.5774 & -0.8165 \\ 0 & 1.0000 \end{pmatrix}$	0.0980
(\mathcal{P}_4, Φ_4)	$K^{-4} \cdot \begin{pmatrix} 1.2500 & 1.2500 \\ 1.2500 & 1.8750 \end{pmatrix}$	$\begin{pmatrix} 0.5774 & -0.8165 \\ 0 & 1.0000 \end{pmatrix}$	0.0980
(\mathcal{P}_5, Φ_5)	$K^{-4} \cdot \begin{pmatrix} 0.7969 & 0.2813 \\ 0.2813 & 0.1875 \end{pmatrix}$	$\begin{pmatrix} 0.6860 & -0.7276 \\ 0 & 1.0000 \end{pmatrix}$	0.1203
(\mathcal{P}_6, Φ_6)	$K^{-4} \cdot \begin{pmatrix} 2.2959 & 0.3828 \\ 0.3828 & 0.1875 \end{pmatrix}$	$\begin{pmatrix} 0.8121 & -0.5835 \\ 0 & 1.0000 \end{pmatrix}$	0.1508
(\mathcal{P}_7, Φ_7)	$K^{-4} \cdot \begin{pmatrix} 0.9219 & 0.2813 \\ 0.2813 & 0.1875 \end{pmatrix}$	$\begin{pmatrix} 0.7365 & -0.6765 \\ 0 & 1.0000 \end{pmatrix}$	0.1318

Table 1: Values of Fisher information matrix $I(t)$, boundary region $\Delta = C(U)$ and $\chi^2(2)$ -weight w_2 (see (53)) for the pedigree types of Figure 1. The upper triangular square root $A(t)$ of $I(t)$ is used for calculating U .

j	κ_1	κ_2	κ_3
1-4	$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$
5	$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0.8137 & 0.1976 \\ 0.1976 & 0.0821 \end{pmatrix}$	$\begin{pmatrix} 0.1765 & -0.1872 \\ -0.1872 & 0.1985 \end{pmatrix}$
6	$\begin{pmatrix} 0.1914 & -0.1375 \\ -0.1375 & 0.0988 \end{pmatrix}$	$\begin{pmatrix} 0.5455 & 0.3265 \\ 0.3265 & 0.2601 \end{pmatrix}$	$\begin{pmatrix} 0.2161 & -0.1552 \\ -0.1552 & 0.1115 \end{pmatrix}$
7	$\begin{pmatrix} 0.1356 & -0.1246 \\ -0.1246 & 0.1144 \end{pmatrix}$	$\begin{pmatrix} 0.7034 & 0.2725 \\ 0.2725 & 0.1352 \end{pmatrix}$	$\begin{pmatrix} 0.1525 & -0.1401 \\ -0.1401 & 0.1287 \end{pmatrix}$
j	κ_4	κ_5	κ_6
1-4	$\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$
5	$\begin{pmatrix} 0.0098 & -0.0104 \\ -0.0104 & 0.7194 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$
6	$\begin{pmatrix} 0.0425 & -0.0306 \\ -0.0306 & 0.5273 \end{pmatrix}$	$\begin{pmatrix} 0.0043 & -0.0031 \\ -0.0031 & 0.0022 \end{pmatrix}$	$\begin{pmatrix} 0.0002 & -0.0002 \\ -0.0002 & 0.0001 \end{pmatrix}$
7	$\begin{pmatrix} 0.0085 & -0.0078 \\ -0.0078 & 0.6217 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$

Table 2: Matrices $\kappa_l = (\kappa_{l,uv})_{u,v=1}^2$ in (57) for the pedigree types (\mathcal{P}_j, Φ_j) of Figure 1. The upper triangular square root $A(t)$ of $I(t)$ is used.