The Annals of Statistics
1999, Vol. 27, No. 3, 943–986

A SCORE TEST FOR LINKAGE USING IDENTITY BY DESCENT DATA FROM SIBSHIPS

BY SANDRINE DUDOIT[1] AND TERENCE P. SPEED[2]

University of California, Berkeley

We consider score tests of the null hypothesis H_0: $\theta = \frac{1}{2}$ against the alternative hypothesis H_1: $0 \le \theta < \frac{1}{2}$, based upon counts multinomially distributed with parameters n and $\rho(\theta, \pi)_{1\times m} = \pi_{1\times m} T(\theta)_{m\times m}$, where $T(\theta)$ is a transition matrix with $T(0) = I$, the identity matrix, and $T(\frac{1}{2}) = (1, \ldots, 1)^T(\alpha_1, \ldots, \alpha_m)$. This type of testing problem arises in human genetics when testing the null hypothesis of no linkage between a marker and a disease susceptibility gene, using identity by descent data from families with affected members. In important cases in this genetic context, the score test is independent of the nuisance parameter π and based on a widely used test statistic in linkage analysis. The proof of this result involves embedding the states of the multinomial distribution into a continuous-time Markov chain with infinitesimal generator Q. The second largest eigenvalue of Q and its multiplicity are key in determining the form of the score statistic. We relate Q to the adjacency matrix of a quotient graph in order to derive its eigenvalues and eigenvectors.

1. Introduction. This paper concerns a class of score tests which arise naturally in human genetics. However, their essence can be described quite efficiently without any of the genetic background, and we now do so. Let $\alpha = (\alpha_1, \ldots, \alpha_m)$ and $\pi = (\pi_1, \ldots, \pi_m)$ be two multinomial distributions, viewed as points in a simplex, and let $\{T(\theta)\colon 0 \le \theta \le \frac{1}{2}\}$ be a one-parameter family of transition matrices such that $T(0) = I$, the identity matrix, and $T(\frac{1}{2}) = \mathbf{1}^T\alpha$, where $\mathbf{1} = (1, \ldots, 1)$. These objects allow us to define the curve $\mathscr{C}_\pi(\theta)$ of distributions $\rho(\theta, \pi) = \pi T(\theta)$, $0 \le \theta \le \frac{1}{2}$, connecting $\pi = \rho(0, \pi)$ to $\alpha = \rho(\frac{1}{2}, \pi)$. Our interest is a score test for the null hypothesis H_0: $\theta = \frac{1}{2}$ against the alternative H_1: $0 \le \theta < \frac{1}{2}$, that is, for testing H_0: $\rho = \alpha$ against alternatives along the curve $\mathscr{C}_\pi(\theta)$, based upon counts $N = (N_1, \ldots, N_m)$ multinomially distributed with parameters $n = \sum_i N_i$ and $\rho(\theta, \pi)$. The associated log-likelihood is $l(\theta, \pi) = \sum_i N_i \ln(\rho_i(\theta, \pi))$, and the score test in question should be based on $l'(\frac{1}{2}, \pi) = \sum_i N_i \rho_i'(\frac{1}{2}, \pi)/\alpha_i$, where $'$ denotes differentiation in θ. It turns out in our genetic context that $l'(\frac{1}{2}, \pi) \equiv 0$, and so we consider the second derivative, obtaining $l''(\frac{1}{2}, \pi) = \sum_i N_i \rho_i''(\frac{1}{2}, \pi)/\alpha_i = \sum_i N_i(\sum_j \pi_j u_{ji})/\alpha_i$, where $U = (u_{ij}) = T''(\frac{1}{2})$. Now, we would normally need

Received August 1998; revised March 1999.

[1]Supported in part by a PMMB Burroughs-Wellcome Fellowship.

[2]Supported in part by NIH Grant R01 HG01093.

AMS 1991 *subject classifications*. Primary 62F03; secondary 92D30, 60J20, 15A18, 05C20, 05C30.

Key words and phrases. Score test, Markov chain, infinitesimal generator, quotient graph, adjacency matrix, eigenvalues, orbits, Pólya's theory, linkage analysis.

to deal with the nuisance parameter π in this score test. This study was motivated by the observation that in some important cases in linkage analysis, U has rank 1, that is, $u_{ij} = a_i b_j$, for suitable vectors (a_i) and (b_i). In such cases, $l''(\frac{1}{2}, \pi) = (\sum_j a_j \pi_j)(\sum_i b_i N_i/\alpha_i)$, and the score test is independent of the nuisance parameter π. Moreover, for our genetic problem, the score test is based on a widely used nonparametric statistic in linkage analysis, S_{pairs} [Whittemore and Halpern (1994a), Kruglyak, Daly, Reeve-Daly and Lander (1996)]. We thought it would be of interest to understand the origins of this property and to learn just how far it extended.

In Section 2, we present the genetic problem which motivated our study, the linkage analysis of disease susceptibility genes using identity by descent (IBD) data from sets of siblings (sibships). This involves describing how IBD patterns in pedigrees (i.e., collections of related individuals) may be summarized by inheritance vectors which correspond to the vertices of a hypercube. The inheritance vectors along a chromosome are embeddable in a continuous-time random walk on the vertices of the hypercube, with time parameter the genetic distance along the chromosome. For our purpose, the inheritance vectors may be partitioned into so-called IBD configurations, which are orbits of groups acting on the set of inheritance vectors. In Section 3, we derive a semigroup property for the IBD configuration transition matrix $T(\theta)$ and present a spectral decomposition of $T(\theta)$ in terms of the eigenvalues and eigenvectors of its infinitesimal generator Q. The second largest eigenvalue of Q and its multiplicity are key in determining the form of the linkage score statistic. In order to derive the eigenvalues and eigenvectors of the infinitesimal generator, we relate it to the adjacency matrix of a quotient graph. Finally, in Section 4, we derive score statistics for testing linkage in sibships and illustrate the results with sib-pairs and sib-trios in Section 5. Remarkably, in an affected-only analysis, where siblings not affected by the disease are ignored, the score test for sibships of a given size does not depend on the nuisance parameter π and is based on a well-known statistic in linkage analysis, S_{pairs} [Whittemore and Halpern (1994a), Kruglyak, Daly, Reeve-Daly and Lander (1996)].

2. Testing linkage using identity by descent data. Genetic mapping is based upon the phenomenon of *crossing-over*, which is the exchange of corresponding DNA between chromosomes from the same pair during gamete (egg/sperm) formation. The human genome is distributed along 23 pairs of chromosomes, 22 autosomal pairs and the sex chromosome pair (XX for females and XY for males). Each pair consists of a paternally inherited chromosome and a maternally inherited chromosome. As a result of crossovers, chromosomes passed from parent to offspring are combinations of the two grandparental chromosomes (see Figures 1 and 2). In general, the DNA variants (*alleles*) passed from parent to offspring at two nearby chromosomal locations (*loci*) have the same grand-parental origin (e.g., at both loci, the maternally inherited alleles are from the maternal grandfather). This is sometimes called *cosegregation*, as *segregation* is the process leading to the choice of one of a parent's two variants (maternal or paternal) at any given locus for transmis-

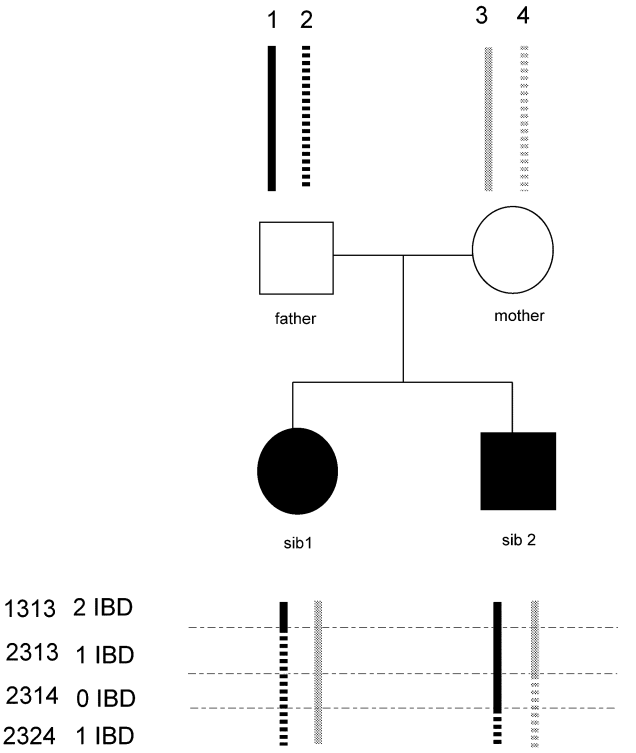

FIG. 1. *Segregation products for a sibship of size* 2 *and a single chromosome pair. Male and female individuals are represented by squares and circles, respectively, and colored symbols indicate affectedness by the disease under study. The paternal and maternal chromosome pairs are labeled by* $(1, 2)$ *and* $(3, 4)$*, respectively. The inheritance vectors and IBD configurations of the sib-pair are indicated on the left.*

sion to a child. Exceptions to cosegregation occur due to crossovers; then, the variants passed on to the child have different grand-parental origins at the two loci and the chromosome is said to be *recombinant* (e.g., for the maternally inherited chromosome, the variant from the maternal grandfather was inherited at one locus and that from the maternal grandmother was inherited at the other locus). The frequency with which this occurs is the *recombination fraction* between the two loci, conventionally denoted by θ. [Recombination fracctions are assumed to be constant across conditions (e.g., age and temperature) and individuals. Under general models for crossovers, the recombination fraction between two loci belongs to the interval $[0, 1/2]$.] In general, two loci are said to be *linked* if their recombination fraction is less than $\frac{1}{2}$, and *unlinked* if it is $\frac{1}{2}$. Thus, unlinked loci may be widely separated on the same chromosome, or on different chromosomes. Loci are said to be *tightly linked* if the recombination fraction θ is close to 0, for example, $\theta < 0.05$ [see Ott (1991), McPeek (1996) and Speed (1996) for an introduction to linkage analysis].

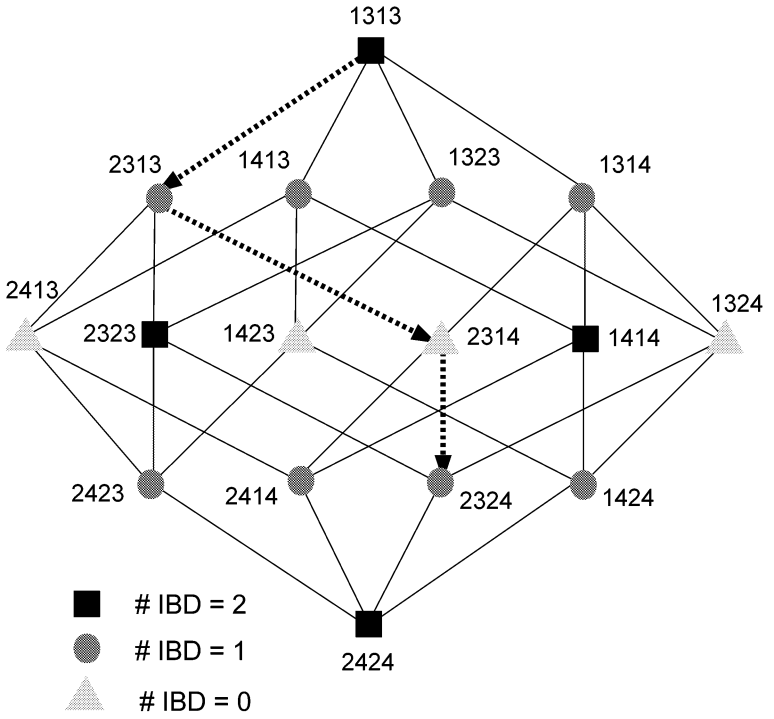

FIG. 2. *Four-dimensional hypercube whose vertices correspond to the* 16 *possible inheritance vectors for a sib-pair and whose edges represent permissible transitions. The arrows indicate the transitions for the segregation products represented in Figure* 1.

When mapping disease susceptibility (DS) genes, we are interested in testing whether genetic markers with known location are linked or not to DS genes, that is, in testing a null hypothesis of the form H_0: $\theta = \frac{1}{2}$, where θ is the recombination fraction between a genetic marker and a putative DS gene. This could be done by studying the cosegregation of variants of the DS genes with those of other mapped genes or markers. (By now, there are hundreds of well-mapped markers along each human, mouse and many other chromosomes.) Frequent cosegregation of a DS locus with a mapped marker would imply a small recombination fraction between the two loci, and hence an accurate placement of the DS locus. However, for most diseases of interest, we do not in general know, and are unable to determine, the alleles present at the DS loci prior to their being mapped. Indeed, much of the interest in mapping DS loci is to determine the variants segregating in populations. Thus, a direct approach to mapping DS loci is generally not available. Many ingenious methods have been developed by geneticists to circumvent this problem, and this paper concerns one such which studies marker identity by descent in sibships with affected members. DNA at the same locus on two chromosomes from the same pair is said to be *identical by descent* (*IBD*) if it originated from

the same ancestral chromosome. This is by contrast to *identity by state* (*IBS*), where the same DNA variant in two individuals may have entered the family under study through different ancestors and hence may not be IBD. Linkage analysis methods based on IBD data seek to exploit the association between the *sharing of DNA identical by descent at loci linked to DS loci* and *disease status* (*phenotype*) in families with affected individuals. At loci unlinked to DS loci, IBD sharing is independent of phenotype. For example, for sibships, this association arises from the fact that full sibs get all their genes from the same source, their parents. Consequently, if susceptibility to the disease under study has a genetic component, the disease status (affected or not) of the sibs should be associated with their IBD status (identical or not) at the DS loci [see Tables 3 and 4 in Dudoit and Speed (1999) for a simple example of this association in sib-pairs]. This approach is effective to the extent that disease susceptibility is affected by genes rather than, say, a shared environment, or other "random" factors. Determining IBD status at or near a DS locus is usually feasible, where determining the gene variant is not, because one can readily determine IBD status at so-called marker loci, one of which may be tightly linked to the DS locus [cf. Kruglyak and Lander (1995) and Kruglyak, Daly, Reeve-Daly and Lander (1996) for a treatment of incomplete IBD data]. IBD-based methods for detecting linkage to DS loci will thus be successful if (and only if) (1) there is a noticeable association between phenotype and IBD status of relatives at the DS loci, and (2) this association is strong enough to remain detectable when IBD status at an (unknown) DS locus is replaced by observed IBD status at a marker locus. Recombination between a DS locus and a marker locus will attenuate the association between phenotype and IBD status. If we have a dense enough set of marker loci, problem (2) would appear to be solved, but in truth there is always a close connection between the magnitude of the association in (1) and the density of the marker set necessary for its detection. These issues were addressed by Thompson (1997), who refers to the two components (1) and (2) as the *specificity* of the DS loci and the *scale* of the genetic distance, respectively. We refer the reader to Whittemore and Halpern (1994a, b), Whittemore (1996), Kruglyak, Daly, Reeve-Daly and Lander (1996), Kong and Cox (1997), Feingold and Siegmund (1997), Teng and Siegmund (1997), Dudoit (1999), and Dudoit and Speed (1999), McPeek (1999) for recent discussions of linkage analysis using IBD data.

The IBD pattern within a pedigree may be summarized at any chromosomal locus by the inheritance vector. Consider a sibship of $k \geq 2$ sibs and suppose we wish to identify the parental origin of the DNA inherited by each sib at a particular autosomal locus, $\mathscr{L}$ say (for loci on sex chromosomes, males and females need to be treated differently). Arbitrarily label the paternal chromosomes containing the locus of interest by $(1, 2)$, and similarly label the maternal chromosomes by $(3, 4)$. The *inheritance vector* of the sibship at the locus $\mathscr{L}$ is the $2k$-vector $x = (x_1, x_2, \ldots, x_{2k-1}, x_{2k})$, indicating the outcome of each of the $2k$ segregations giving rise to the sibship. More precisely, for $i = 1, \ldots, k$, x_{2i-1} is the label of the paternal chromosome from which sib i inherited DNA at $\mathscr{L}$, 1 or 2, and x_{2i} is the label of the maternal chromosome

from which sib i inherited DNA at $\mathscr{L}$, 3 or 4 (see Figures 1 and 2). Denote by $\mathscr{X}$ the set of all 2^{2k} inheritance vectors.

Consider now two loci, $\mathscr{L}_1$ and $\mathscr{L}_2$, separated by a recombination fraction θ, and denote the inheritance vectors at the two loci by x and y, respectively. If these two inheritance vectors differ at a particular entry, this indicates the occurrence of a recombination between $\mathscr{L}_1$ and $\mathscr{L}_2$ in the corresponding segregation. The chance of a recombination between the two loci is the recombination fraction θ, taken to be constant across conditions and individuals. the transition matrix $R(\theta)$ between inheritance vectors at loci separated by a recombination fraction θ has entries

$$r_{xy}(\theta) = \theta^{\Delta(x,y)}(1-\theta)^{2k-\Delta(x,y)}, \tag{2.1}$$

where $\Delta(x, y)$ is the number of coordinates at which the inheritance vectors x and y differ, that is, the number of recombination events between the two loci. The matrix $R(\theta)$ may be expressed as the Kronecker power of 2×2 transition matrices corresponding to transitions in each of the $2k$ coordinates,

$$R(\theta) = \begin{bmatrix} 1-\theta & \theta \\ \theta & 1-\theta \end{bmatrix}^{\otimes 2k}. \tag{2.2}$$

The notion of inheritance vector generalizes to arbitrary pedigrees, where pedigree members are separated into founders (individuals whose parents are not in the pedigree) and nonfounders (individuals whose parents are in the pedigree). In the case of sibships, the parents are founders and the sibs are nonfounders. For a pedigree with k nonfounders, the inheritance vector at a particular locus is defined to be a $2k$-vector with coordinates describing the outcome of the paternal and maternal segregations giving rise to the k nonfounders [Lander and Green (1987); Kruglyak and Lander (1995) and Kruglyak, Daly, Reeve-Daly and Lander (1996)]. The $(2i-1)$st coordinate is 0 or 1 according to whether the grand-paternal or grand-maternal DNA was transmitted in the paternal segregation giving rise to the ith nonfounder. The $(2i)$th coordinate contains the same information for the maternal segregation. For the purpose of this paper, we prefer the definition introduced earlier for sibship inheritance vectors (with labels $1, 2, 3, 4$ for parental chromosomes) to the more common definition with binary labels, since this facilitates the presentation in Section 3 of group action on inheritance vectors. It is easy to show that for general pedigrees the transition matrix $R(\theta)$ between inheritance vectors also has the form given in (2.2). Although this paper is primarily concerned with IBD data from sibships, the general setup and some of the results presented here (Propositions 1, 2 and 3) apply to arbitrary pedigree types, and we discuss generalizations where appropriate.

For the purpose of linkage analysis of disease genes, certain inheritance vectors are equivalent to each other, in that they have the same probability of arising at DS genes in pedigrees with given phenotypes. Although not needed for an understanding of this paper, a discussion of these probabilities for sibships and the genetic model under which they are calculated may be found in

Dudoit and Speed (1999) and Dudoit (1999). For an arbitrary pedigree type, we define *IBD configurations* as any partitioning of the set of inheritance vectors. For example, for k affected sibs, Ethier and Hodge (1985) partition the 2^{2k} inheritance vectors into a much smaller number of equivalence classes as follows. Two inheritance vectors belong to the same IBD configuration if one may be obtained from the other by applying any combination of the following four operations: (1) interchange the paternal labels 1 and 2, (2) interchange the maternal labels 3 and 4, (3) interchange the parental origin of the DNA by interchanging 1 and 3 and 2 and 4, and (4) permute the sibs. With this definition, the 16 possible inheritance vectors for a sib-pair are partitioned into three IBD configurations, corresponding to the number of chromosomes sharing DNA IBD at the locus of interest (cf. Table 1). In general, inheritance vectors may be partitioned in various ways for different purposes, and we address this question in greater detail for sibships in Section 3.

For a pedigree with given phenotypes, the conditional probability vector of IBD configurations at a genetic marker $\mathscr{M}$ linked to a DS locus $\mathscr{D}$ at recombination fraction θ is given by

$$\rho(\theta, \pi)_{1\times m} = \pi_{1\times m} T(\theta)_{m\times m},$$

where π is the conditional probability vector of IBD configurations at the DS locus (possibly one of several unlinked DS loci), m is the number of IBD configurations and $T(\theta)$ is the transition matrix between IBD configurations θ apart. In general, π depends on unknown and numerous genetic parameters, such as penetrances and genotype frequencies. In this paper, we consider a general genetic model with multiple genes unlinked to each other, arbitrary penetrances, and no population genetic assumptions such as random mating or Hardy–Weinberg equilibrium. Under the null hypothesis of no linkage between the marker and the DS locus, the IBD distribution at the marker is

$$\alpha = \rho\left(\frac{1}{2}, \pi\right) = \frac{1}{2^{2k}}\left(|\mathscr{C}_1|, \ldots, |\mathscr{C}_m|\right),$$

where $|\mathscr{C}_i|$ is the number of inheritance vectors in $\mathscr{C}_i$, the ith IBD configuration.

Thus, the IBD probabilities at the marker have two separate components: one component involving the recombination fraction θ between the marker and the DS locus (*scale*), the other depending on the mode of inheritance of

TABLE 1
Sib-pair IBD configurations

Orbits of $S_2 \times D_4$	**Orbits of $S_2 \times (C_2 \times C_2)$**	**Inheritance vectors**
0 IBD	0 IBD	(1, 3, 2, 4), (1, 4, 2, 3), (2, 3, 1, 4), (2, 4, 1, 3)
1 IBD	1 paternal IBD	(1, 3, 1, 4), (1, 4, 1, 3), (2, 3, 2, 4), (2, 4, 2, 3)
	1 maternal IBD	(1, 3, 2, 3), (1, 4, 2, 4), (2, 3, 1, 3), (2, 4, 1, 4)
2 IBD	2 IBD	(1, 3, 1, 3), (1, 4, 1, 4), (2, 3, 2, 3), (2, 4, 2, 4)

the disease (*specificity*). Our score test in the recombination fraction θ focuses on the scale component and seems to achieve some robustness with respect to the specificity (π). Examples of the transition matrix $T(\theta)$ are given in Section 5 for sib-pairs and sib-trios.

Suppose we collect data on n sibships of given size and phenotypes and wish to test the null hypothesis of no linkage between a genetic marker and a DS locus. Denote by N_i the number of sibships with IBD configuration i, $i = 1, \ldots, m$, at the genetic marker. Under certain sampling assumptions [Dudoit (1990)], $(N_1, \ldots, N_m)$ have a Multinomial$(n, \rho(\theta, \pi))$ distribution. There is no uniformly most powerful test of H_0: $\theta = \frac{1}{2}$, however, the *score test* is *locally most powerful*. Moreover, for affected-only sibships of a given size, the score statistic does not involve the nuisance parameter π and reduces to a widely used statistic in linkage analysis, S_{pairs}, which is obtained by forming all possible pairs of sibs and averaging the proportions of chromosomes on which they share DNA IBD at the marker [Whittemore and Halpern (1994a) and Kruglyak, Daly, Reeve-Daly and Lander (1996)]. This result is a corollary to Theorem 2 in Section 4.

COROLLARY 1. *For affected sib-k-tuples, using the IBD configurations of Ethier and Hodge, the score test of* H_0: $\theta = \frac{1}{2}$ *is based on* S_{pairs}, *regardless of the model for disease susceptibility, that is, regardless of* π. *For one affected sib-k-tuple,*

$$S_{\text{pairs}} = \frac{\sum_{i<j} S_{ij}}{k(k-1)},$$

where S_{ij} *is the number of chromosomes on which the* ij*th sib-pair shares DNA IBD. Under the null hypothesis of no linkage, the* S_{ij}*'s are pairwise independent* Binomial$(2, \frac{1}{2})$ *random variables, and thus*

$$E_0[S_{\text{pairs}}] = \frac{1}{2}, \qquad \text{Var}_0[S_{\text{pairs}}] = \frac{1}{4k(k-1)}.$$

For a collection of affected sib-k-tuples, S_{pairs} *is summed over all sibships.*

Thus, for affected-only sibships of a *given* size, S_{pairs} is locally most powerful (in θ), and may be calculated easily by considering each sib-pair one at a time and without the need for assigning sibships to IBD configurations. This finding extends the work of Knapp, Seuchter and Baur (1994a) to sibships of any size and to general genetic models with multiple unlinked DS loci and no population genetic assumptions such as random mating or Hardy–Weinberg equilibrium. Unfortunately, this simple property does not hold with all types of sibships, and we consider examples where it fails in Sections 4 and 5. Furthermore, we show that the linkage score statistic combining IBD data from sibships of *different* sizes assigns different weights to the various sibship types and these weights depend on the genetic model.

The remainder of this paper is concerned with the proof of Corollary 1 and with deriving score statistics for general sibships, with any number of affected

and unaffected sibs, and distinguishing the parental origin of the DNA. In general, the form of the score statistic is based on properties of the transition matrix $T(\theta)$, which in turn are determined by the pedigree type and the choice of IBD configurations. Thus, we first describe how sibship inheritance patterns may be summarized by IBD configurations which are orbits of groups acting on the set of inheritance vectors.

3. Transition matrix for IBD configurations.

3.1. *Sibship IBD configurations.* Let $a = (1, 3)$, $b = (1, 4)$, $c = (2, 3)$ and $d = (2, 4)$ denote all four possible segregation outcomes at a particular locus for a given sib. Then we may think of the set of inheritance vectors $\mathscr{X}$ as the set of mappings x: $\{1, \ldots, k\} \to \{a, b, c, d\}$. In this setting, the IBD configurations are *orbits* of groups acting on $\mathscr{X}$, where the groups are determined by the type of operations allowed within IBD configurations [cf. Fraleigh (1989), Section 3.2 for an introduction to group action]. Let

$\alpha = (ac)(bd)$	interchange labels 1 and 2 of paternal chromosomes,
$\beta = (ab)(cd)$	interchange labels 3 and 4 of maternal chromosomes,
$\gamma = (bc)$	interchange parental origin of DNA.

The group of permutations generated by α, β and γ is actually the *dihedral group*, D_4 (α and γ are sufficient to generate D_4), and the group generated by α and β is the *Klein four-group*, $C_2 \times C_2$. Figure 3 displays a square with vertices a, b, c, and d. Permutations α and β correspond to mirror images in the perpendicular bisectors of the sides, and permutations γ corresponds to a diagonal flip. The diherdral group D_4 is the group of symmetries of the square. The IBD configurations of Ethier and Hodge (1985) for affected-only sibships are the orbits of the direct product $S_k \times D_4$, of the *symmetric group* S_k on k letters and the dihedral group D_4, acting on $\mathscr{X}$. In some situations (e.g., parental

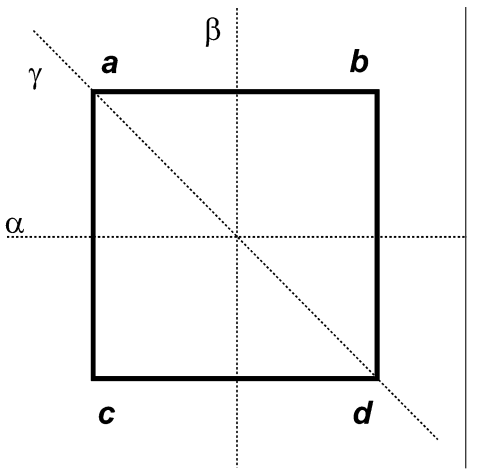

FIG. 3. *Permutations α, β and γ of the vertices of the square.*

imprinting, when disease susceptibility is different for maternally and paternally inherited disease alleles), it may be appropriate to distinguish between sharing of maternal and paternal DNA and exclude the group operation γ. For example, for a sib-pair, it may be appropriate to distinguish between the two inheritance vectors $(1, 3, 1, 4)$ and $(1, 3, 2, 3)$; for $(1, 3, 1, 4)$ the sibs share DNA IBD on the paternal chromosome, whereas for $(1, 3, 2, 3)$ the sibs share DNA IBD on the maternal chromosome (cf. Table 1). For sibships with both affected and unaffected individuals, similar configurations may be defined, but the sibs are permuted only among affecteds or unaffecteds. The different types of group action are listed in Table 2.

For a group $G \times H$ acting on the set $\mathscr{X}$, let m denote the number of orbits and $\mathscr{C}_i$ denote the ith orbit. In general, we may use the *Pólya theory of counting* [van Lint and Wilson (1992) and deBruijn (1964)] to find the number of orbits of groups acting on mappings, and hence determine the number of IBD configurations of each type (see Appendix, Section A). Ethier and Hodge derived the number of IBD configurations of affected sib-k-tuples, as well as the number of inheritance vectors in each IBD configuration, without reference to the group $S_k \times D_4$. Instead, they based their calculations on labels for the equivalence classes which are triples of integers [cf. pages 264, 265 in Ethier and Hodge (1985) and Appendix, Section A].

3.2. *Properties of transition matrix.* In this section, we derive properties of transition matrices between IBD configurations. While Theorems 1 and 2 are specific to the sibship IBD configurations defined in the previous section, Propositions 1 and 2 hold more generally for any type of pedigree, with IBD configurations defined as orbits of groups acting on the set of inheritance vectors. The transition matrix $T(\theta)$ between IBD configurations at loci separated by a recombination fraction θ is the $m \times m$ matrix with entries

$$t_{ij}(\theta) = \frac{1}{|\mathscr{C}_i|} \sum_{x \in \mathscr{C}_i} \sum_{y \in \mathscr{C}_j} r_{xy}(\theta) = \frac{1}{|\mathscr{C}_i|} \sum_{x \in \mathscr{C}_i} \sum_{y \in \mathscr{C}_j} \theta^{\Delta(x,y)} (1-\theta)^{2k - \Delta(x,y)}.$$

However, given any two inheritance vectors in $\mathscr{C}_i$, the probability of a transition to $\mathscr{C}_j$ is the same, that is,

$$\sum_{y \in \mathscr{C}_j} r_{xy}(\theta) = \sum_{y \in \mathscr{C}_j} r_{\tilde{x}y}(\theta) \quad \text{for any } x,\ \tilde{x} \in \mathscr{C}_i. \tag{3.1}$$

TABLE 2
Sibship IBD configurations

# affected, # unaffected	Distinguish maternal from paternal sharing	Group $G \times H$
$k, 0$	NO	$S_k \times D_4$
	YES	$S_k \times (C_2 \times C_2)$
$h, k-h$	NO	$(S_h \times S_{k-h}) \times D_4$
	YES	$(S_h \times S_{k-h}) \times (C_2 \times C_2)$

This result follows by observing that if $\sim$ denotes the operation applied to x to obtain $\tilde{x}$, then $\Delta(\tilde{x}, y) = \Delta(\tilde{\tilde{x}}, \tilde{y}) = \Delta(x, \tilde{y})$ and $\tilde{y} \in \mathscr{C}_j$. Consequently,

$$
\begin{aligned}
t_{ij}(\theta) &= \sum_{y \in \mathscr{C}_j} \theta^{\Delta(x,y)}(1-\theta)^{2k-\Delta(x,y)} \quad \text{where } x \text{ is any } x \in \mathscr{C}_i \\
&= \frac{|\mathscr{C}_j|}{|\mathscr{C}_i|} \sum_{x \in \mathscr{C}_i} \theta^{\Delta(x,y)}(1-\theta)^{2k-\Delta(x,y)} \quad \text{where } y \text{ is any } y \in \mathscr{C}_j.
\end{aligned} \tag{3.2}
$$

The next two propositions relate the transition matrix $T(\theta)$ to the adjacency matrix of a quotient graph, whose eigenvalues are key in determining the form of the score statistic (see Appendix, Section B for proofs).

PROPOSITION 1. *Let $T(\theta)$ denote the transition matrix between IBD configurations which are orbits of groups acting on the set of inheritance vectors $\mathscr{X}$, and let $\theta_1 * \theta_2 = \theta_1(1-\theta_2) + \theta_2(1-\theta_1)$. Then, $T(\theta)$ satisfies the semigroup property*

$$T(\theta_1 * \theta_2) = T(\theta_1)T(\theta_2).$$

Thus, $T(\theta)$ may be written as

$$T(\theta) = e^{d(\theta)Q},$$

where $d(\theta) = -\frac{1}{2}\ln(1-2\theta)$ is the inverse of the Haldane map function and Q is the infinitesimal generator. The infinitesimal generator is given by

$$Q = B - 2kI,$$

where B is the $m \times m$ matrix with entries

$$b_{ij} = \sum_{y \in \mathscr{C}_j} I(\Delta(x, y) = 1) \quad \text{for any } x \in \mathscr{C}_i$$

$\Delta(x, y)$ is the number of coordinates at which x and y differ and $I(\ )$ denotes the indicator function. The stationary distribution of $T(\theta)$ is

$$\alpha = (\alpha_1, \ldots, \alpha_m) = \frac{1}{2^{2k}}\left(|\mathscr{C}_1|, \ldots, |\mathscr{C}_m|\right)$$

and $T(\theta)$ is reversible, that is,

$$\alpha_i t_{ij}(\theta) = \alpha_j t_{ji}(\theta).$$

Hence, for one segregation, the crossover process is embeddable in a continuous-time random walk on $\{0, 1\}$, where 0 and 1 denote, respectively, the transmission of paternal and maternal DNA to one's child, and the time parameter is $d(\theta) = -\frac{1}{2}\ln(1-2\theta)$. Jointly, the crossover processes are i.i.d. and hence embeddable in a continuous-time random walk on the vertices of the hypercube $\{0, 1\}^{2k}$ (cf. Donnelly (1983) and Figures 1 and 2). The

random walk model for the crossover process is widely used and is referred to in the genetics literature as the *no interference model*. The *Haldane map function* relates the recombination fraction to the map distance under the no interference model. The *map distance* between two loci is the expected number of crossover events, that is, of changes of the grand-parental origin of the DNA, occurring on a segregation product in the chromosomal interval between the two loci. Under the no interference model, the crossover process on individual segregation products is a Poisson process with intensity 1 [cf. Ott (1991), McPeek (1996) and Speed (1996) for an introduction to map functions and a more detailed discussion of crossover processes]. Note that, if we have three ordered loci, and θ_1 and θ_2 are the recombination fractions between the first and second and second and third locus, respectively, then $\theta_1 * \theta_2$ is the recombination fraction between the first and third locus, under the assumption that recombination events in disjoint intervals are independent, that is, there is no crossover interference. Also, note that we did not need to assume no crossover interference to derive the semigroup property. If, however, we do assume no crossover interference, then the inheritance vectors along a chromosome form a continuous-time Markov chain with time parameter the genetic distance along a chromosome. From (3.1) and condition (15), page 63 in Rosenblatt (1974), it follows that the IBD configurations also form a continuous-time Markov chain.

In order to compute score statistics, we need derivatives of the transition matrix at $\theta = \frac{1}{2}$. These may be computed by differentiating (3.2); however, we gain more knowledge on the transition matrix $T(\theta)$ and on the form of the score statistic by using the following spectral decomposition of $T(\theta)$.

PROPOSITION 2. *Let $T(\theta)$ denote the transition matrix between IBD configurations which are orbits of groups acting on the set of inheritance vectors $\mathscr{X}$. Then $T(\theta)$ may be written as*

$$T(\theta) = \sum_h \exp(\lambda_h d(\theta))\, P_h = \sum_h (1-2\theta)^{-\lambda_h/2}\, P_h, \tag{3.3}$$

where λ_h are the m real eigenvalues of the infinitesimal generator Q, and P_h are projection matrices satisfying $P_h^2 = P_h = P_h^*$, $P_h P_l = 0$, $h \neq l$, and $\sum_h P_h = I$. Here P_h^* is the adjoint of P_h with respect to the inner product $\langle x, y\rangle_\alpha = \sum_i \alpha_i x_i y_i$. The ijth entry of P_h is $\alpha_j v_{ih} v_{jh}$, where v_{ih} is the ith entry of the right eigenvector $\mathbf{v}_h$ of Q corresponding to λ_h, and the eigenvectors $\mathbf{v}_h$ are orthonormal with respect to the inner product $\langle\ ,\ \rangle_\alpha$. In particular, the first two derivatives of the transition matrix with respect to θ are

$$T'(\theta) = \sum_h \lambda_h (1-2\theta)^{-(\lambda_h+2)/2} P_h \tag{3.4}$$

and

$$T''(\theta) = \sum_h \lambda_h(\lambda_h+2)(1-2\theta)^{-(\lambda_h+4)/2} P_h. \tag{3.5}$$

Thus, eigenvalues of Q and their multiplicity give us information regarding the derivatives of the transition matrix $T(\theta)$ and hence, the form of the score statistic. In particular, powers of θ in $T(\theta)$ are determined by the eigenvalues of Q, and the first nonzero derivative of $T(\theta)$ at $\theta = \frac{1}{2}$ and its rank are determined by the second largest eigenvalue of Q and its multiplicity. We will relate Q to the adjacency matrix of a quotient graph in order to derive its eigenvalues. Consider the graph $\mathscr{X}$ with vertex set the set of all inheritance vectors of length $2k$ and adjacency matrix $A(\mathscr{X}) = A$ with (x, y)-entry,

$$a_{xy} = \begin{cases} 1, & \text{if } \Delta(x, y) = 1, \\ 0, & \text{otherwise.} \end{cases}$$

Here $\mathscr{X}$ is the graph defined by the first associates in the Hamming scheme $H(2k, 2)$ [Chapter 30 in van Lint and Wilson (1992)]. Consider any of the four groups $G \times H$ described in Table 2. The matrix B, defined in Proposition 1, is the *adjacency matrix of the quotient graph* $\mathscr{X}/(G \times H)$, which is the multidigraph with the orbits of $G \times H$ as its vertices and with b_{ij} arcs going from $\mathscr{C}_i$ to $\mathscr{C}_j$. Recall that $Q = B - 2kI$; consequently, we may work with B to derive the eigenvalues of Q. The following theorem relies on general facts concerning eigenvectors and eigenvalues of adjacency matrices of quotient graphs, as well as specific facts regarding the behavior of eigenvectors of A on the orbits of the four groups $G \times H$ described in Section 3.1 (see Appendix, Section C for proof).

THEOREM 1 (Eigenvalues of infinitesimal generator Q for sibship IBD configurations). *The largest eigenvalue of Q is* 0, *with multiplicity one*, *and the second largest eigenvalue of Q is* -4, *with multiplicity depending on the group $G \times H$ defining the IBD configurations*.

(a) $S_k \times D_4$: -4 *has multiplicity one*;
(b) $S_k \times (C_2 \times C_2)$: -4 *has multiplicity two*;
and for $k \geq 3$
(c) $(S_h \times S_{k-h}) \times D_4$: -4 *has multiplicity two if* $h = 1$ *or* $h = k - 1$, *and three if* $2 \leq h \leq k - 2$;
(d) $(S_h \times S_{k-h}) \times (C_2 \times C_2)$: -4 *has multiplicity four if* $h = 1$ *or* $h = k - 1$, *and six if* $2 \leq h \leq k - 2$.

Furthermore, *all other eigenvalues of Q belong to the set* $\{-2i_{\binom{2k}{i}}: i = 3, \ldots, 2k\}$, *where the subscript* $\binom{2k}{i}$ *is the largest possible multiplicity of the eigenvalue* $-2i$. *Thus*, *from* (3.4) *and* (3.5),

$$T'\left(\tfrac{1}{2}\right) = 0 \tag{3.6}$$

and

$$U = T''\left(\tfrac{1}{2}\right) = 8P_{-4}, \tag{3.7}$$

where P_{-4} is the projection matrix for the second largest eigenvalue, -4, *with rank the multiplicity of* -4. *In general*, *the ijth entry of P_{-4} is* $\alpha_j \sum v_i v_j$,

where the v's are the right orthonormal (with respect to the inner product $\langle\ ,\ \rangle_\alpha$*) eigenvectors of* Q *corresponding to* -4*, and the sum is over all such eigenvectors.*

Note that we may also show that $T'(\frac{1}{2}) = 0$ by simple algebra, but this approach does not yield any particular insight into other general properties of $T(\theta)$.

The projection matrix for the largest eigenvalue of Q, $\lambda_1 = 0$, is $T(\frac{1}{2})$, the matrix whose rows are equal to the stationary distribution α. From Proposition 2 and Theorem 1, transition matrices for sibship IBD configurations have the form

$$T(\theta) = T\left(\tfrac{1}{2}\right) + (1-2\theta)^2 P_{-4} + o((1-2\theta)^2),$$

and the rate of convergence of $T(\theta)$ to $T(\frac{1}{2})$ as $\theta \to \frac{1}{2}$ is $O((1-2\theta)^2)$. Under the no interference model, the rate of convergence in terms of the map distance $d = -\ln(1-2\theta)/2$ is $O(e^{-4d})$.

More generally, since the matrix $R(\theta)$ has the same form for any type of pedigree, Propositions 1, 2 and 3 apply to arbitrary pedigrees, as long as the IBD configurations are defined as orbits of groups. Thus, for $2k$ segregations, the transition matrix for IBD configurations has the general form

$$T(\theta) = \sum_h (1-2\theta)^{-\lambda_h/2} P_h,$$

where the eigenvalues λ_h of Q belong to the set $\{-2i\binom{2k}{i}: i = 0, \ldots, 2k\}$. The ith derivative, $i = 0, \ldots, 2k$, of $T(\theta)$ is given by

$$T^{(i)}(\theta) = \sum_h \left\{ \prod_{j=0}^{i-1} (\lambda_h + 2j) \right\} (1-2\theta)^{-(\lambda_h+2i)/2} P_h.$$

The first nonzero derivative of $T(\theta)$ at $\theta = \frac{1}{2}$ and its rank are determined by the second largest eigenvalue of Q, λ_2, and its multiplicity. If $\lambda_2 = -2i$, the first nonzero derivative is the ith derivative,

$$T^{(i)}\left(\tfrac{1}{2}\right) = (-2)^i i! P_{-2i},$$

and the rank of this ith derivative is the multiplicity of $\lambda_2 = -2i$. Furthermore, the rate of convergence to $T(\frac{1}{2})$ is determined by the second largest eigenvalue,

$$T(\theta) = T\left(\tfrac{1}{2}\right) + (1-2\theta)^{-\lambda_2/2} P_{\lambda_2} + o((1-2\theta)^{-\lambda_2/2}).$$

The smaller the second largest eigenvalue, the faster the convergence to $T(\frac{1}{2})$. We proved that for sibships, the second largest eigenvalue is -4, but this is not the case for all types of relatives. In turns out that for grandparent/grandchild pairs, who can share DNA IBD on either 0 or 1 chromosome, the second largest eigenvalue is -2 [Dudoit (1999)]. It is noteworthy that for some types of relative pairs (e.g., cousin pairs), the usual IBD configurations (0 or 1 IBD) are not orbits of groups and the transition matrix for these usual configura-

tions does not satisfy the semigroup property [Donnelly (1983) and Dudoit (1999)].

In the next section, we will explore the implications of Theorem 1 on score tests of the null hypothesis of no linkage between a marker and a gene using IBD data from sibships.

4. Linkage score test for sibships. Suppose we have data on n sibships of a given type (e.g., affected sib-k-tuples with orbits of $S_k \times D_4$), in the form of multinomial counts N_i, $i = 1, \ldots, m$, for the number of sibships with IBD configuration i at a marker $\mathscr{M}$. We wish to test the null hypothesis of no linkage between the marker $\mathscr{M}$ and a DS locus $\mathscr{D}$, which could be one of several DS loci unlinked to each other; that is, we wish to test

$$\mathrm{H}_0\colon\ \theta = \tfrac{1}{2}\ \text{(no linkage)} \quad \text{versus} \quad \mathrm{H}_1\colon\ 0 \le \theta < \tfrac{1}{2}\ \text{(linkage)},$$

where θ denotes the recombination fraction between $\mathscr{M}$ and $\mathscr{D}$. Note that $\theta = \frac{1}{2}$ and $\pi = \alpha$ are not identifiable.

The log-likelihood of the IBD data, conditional on the phenotypes, is

$$l(\theta, \pi) = \sum_i N_i \ln(\rho_i(\theta, \pi)),$$

where

$$\rho(\theta, \pi)_{1\times m} = \pi_{1\times m} T(\theta)_{m\times m}.$$

The *score test* is based on the first nonzero derivative in the Taylor series expansion of the log-likelihood about $\theta = \frac{1}{2}$. In our problem, the first derivative vanishes, so we turn to the second derivative of the log-likelihood with respect to θ, which yields a test that maximizes the second derivative of the power function at the null. We find the score statistic for the given sibship type to be

$$S = \left.\frac{\partial^2 l(\theta, \pi)}{\partial \theta^2}\right|_{\theta=1/2} = \sum_{i=1}^{m} N_i \left.\frac{\partial^2 \rho_i(\theta, \pi)/\partial\theta^2}{\rho_i(\theta, \pi)}\right|_{\theta=1/2} = \sum_{i=1}^{m} N_i \frac{\sum_{j=1}^{m} \pi_j u_{ji}}{\alpha_i},$$

where $U = T''(\frac{1}{2}) = 8P_{-4} = (8\alpha_j \sum v_i v_j)$, the v's are right orthonormal (with respect to $\langle\ ,\ \rangle_\alpha$) eigenvectors of Q corresponding to the eigenvalue -4, and the sum in U is over all such eigenvectors. The null hypothesis of no linkage is rejected for large values of the score statistic S. We show next that for affected sib-k-tuples, with the orbits of $S_k \times D_4$, the second largest eigenvalue has multiplicity one, and as a result, the score statistic for affected-only sibships of a given size is independent of the nuisance parameter π.

4.1. *Affected sib-k-tuples, orbits of $S_k \times D_4$.* A very widely used statistic in linkage analysis is S_{pairs} [cf. Kruglyak, Daly, Reeve-Daly and Lander (1996), S_P of Whittemore and Halpern (1994a), and *PAIRS* and *WP* of Suarez and van Eerdewegh (1984)]. For a sibship of size k, S_{pairs} is obtained by forming all

possible pairs of sibs and averaging the proportions of chromosomes on which they share DNA IBD at the locus of interest; that is,

$$S_{\text{pairs}} = \frac{\sum_{i<j} S_{ij}}{k(k-1)},$$

where S_{ij} is the number of chromosomes on which the ijth sib-pair shares DNA IBD. The corollary in Section 2 results from the following theorem.

THEOREM 2 (Affected sib-k-tuple score statistic, orbits of $S_k \times D_4$). *Suppose we have IBD data on n affected sib-k-tuples, with IBD configurations defined as the orbits of $S_k \times D_4$. Then the score test of the null hypothesis of no linkage between a marker and a DS locus, H_0: $\theta = \frac{1}{2}$, is based on S_{pairs}. The contribution of n affected sib-k-tuples to the overall score statistic is*

$$\begin{aligned} S &= 8\Big(\sum_{j=1}^{m} v_j \pi_j\Big)\Big(\sum_{i=1}^{m} v_i N_i\Big) \\ &= 2^{2k-2}\Big(\sum_{j=1}^{m} u_{j1}\pi_j\Big)(2S_{\text{pairs}} - n) \\ &= 8\sqrt{k(k-1)}\Big(\sum_{j=1}^{m} v_j \pi_j\Big)(2S_{\text{pairs}} - n), \end{aligned} \tag{4.1}$$

where u_{i1} is the ith entry of the first column of $U = T''(\frac{1}{2})$, v_i is the ith entry of the right eigenvector of Q corresponding to the eigenvalue -4 and $u_{i1} = 2^{5-2k}\sqrt{k(k-1)}v_i$. For n sibships, S_{pairs} is summed over all sibships.

The proof of Theorem 2 may be found in the Appendix, Section D, and relies on Theorem 1 and the following identity. For a sibship with inheritance vector x,

$$\begin{aligned} S_{\text{pairs}} &= \frac{\sum_{i=1}^{4} a_i(x)(a_i(x)-1)}{2k(k-1)} \\ &= \frac{a_1(x)^2 + a_2(x)^2 + a_3(x)^2 + a_4(x)^2 - 2k}{2k(k-1)}, \end{aligned} \tag{4.2}$$

where $a_i(x)$ is the number of i labels in the inheritance vector x of the sibship, $i = 1, 2, 3, 4$, and $a_1(x) + a_2(x) + a_3(x) + a_4(x) = 2k$. Without loss of generality, we let the first IBD configuration be the one for which all sibs inherited the same maternal and paternal DNA, that is, with representative inheritance vector $(1, 3, 1, 3, \ldots, 1, 3)$ and label $(0, 0, 0)$ in the notation of Ethier and Hodge (1985). The entries of the first column of U are easily computed, as seen in the proof.

Thus, for affected sib-k-tuples and without distinguishing between sharing of maternal and paternal DNA, the score test is based on S_{pairs}, regardless

of the genetic model, and may be calculated easily by considering each sib-pair one at a time and without the need for assigning sibships to IBD configurations. For commonly studied genetic models, the affected sib-pair "possible triangle" constraints hold [Dudoit and Speed (1999)], and as a result, $\sum_j v_j \pi_j \geq 0$. This follows by noting that $n \sum_j v_j \pi_j$ is the expected value of $\sum_j v_j N_j = \sqrt{k(k-1)}(2S_{\text{pairs}} - n)$ when $\theta = 0$. Also, for each sibship, S_{pairs} is an average of sib-pair statistics, each with expected value $\pi_2 + \pi_1/2 \geq 1/2$, where in this expression π_i, $i = 0, 1, 2$, is the probability that an affected sib-pair shares DNA IBD on i chromosomes at the DS locus. For IBD data from sibships of a *given* size, it is thus appropriate to reject the null hypothesis of no linkage for large values of S_{pairs}. However, the score statistic for combining IBD data from sibships of *different* sizes involves weights which do depend on the genetic model ($\sum_i v_i \pi_i$).

4.2. *Affected sib-k-tuples, orbits of* $S_k \times (C_2 \times C_2)$. For affected sib-$k$-tuples and distinguishing between sharing of maternal and paternal DNA, the second largest eigenvalue of the infinitesimal generator Q, -4, has multiplicity two (Theorem 1). Hence, the second derivative of the transition matrix at $\theta = \frac{1}{2}$ has rank 2 and entries

$$u_{ij} = 8\alpha_j(v_i v_j + \tilde{v}_i \tilde{v}_j),$$

where $v = (v_1, \ldots, v_m)^T$ and $\tilde{v} = (\tilde{v}_1, \ldots, \tilde{v}_m)^T$ are the right orthonormal (with respect to the inner product $\langle\ ,\ \rangle_\alpha$) eigenvectors of Q corresponding to the second largest eigenvalue. These eigenvectors are based on V_e and V_o, respectively (see Appendix, Section C). The score statistic is given by

$$\begin{aligned} S &= \sum_{i=1}^{m} N_i \frac{\sum_{j=1}^{m} \pi_j 8\alpha_i(v_i v_j + \tilde{v}_i \tilde{v}_j)}{\alpha_i} \\ &= 8\Bigg(\sum_{j=1}^{m} v_j \pi_j\Bigg)\Bigg(\sum_{i=1}^{m} v_i N_i\Bigg) + 8\Bigg(\sum_{j=1}^{m} \tilde{v}_j \pi_j\Bigg)\Bigg(\sum_{i=1}^{m} \tilde{v}_i N_i\Bigg). \end{aligned}$$

Thus, in general, the score test depends on the parameters of the genetic model through π. However, in some situations (e.g., no imprinting), this score statistic reduces to S_{pairs}.

4.3. *Discordant sib-k-tuples.* For sibships of size at least 3, with both affected and unaffected individuals [orbits of $(S_h \times S_{k-h}) \times D_4$ or of $(S_h \times S_{k-h}) \times (C_2 \times C_2)$], the second largest eigenvalue of the infinitesimal generator has multiplicity at least two (Theorem 1). Hence, in general, the score statistic depends on the genetic model and is a sum of terms of the form $(\sum_{j=1}^{m} v_j \pi_j)(\sum_{i=1}^{m} v_i N_i)$, where the v's are the right orthonormal eigenvectors of Q corresponding to the second largest eigenvalue.

In the next section, we consider the examples of sib-pairs and sib-trios and present the transition matrix $T(\theta)$, the infinitesimal generator Q and the score statistic for these sibship types.

5. Examples.

5.1. *Sib-pairs, orbits of* $S_2 \times D_4$. For sib-pairs with either 0, 1 or 2 affected individuals and without distinguishing between sharing of maternal and paternal DNA, there are three distinct IBD configurations, labeled 0, 1, 2, according to the number of chromosomes sharing DNA IBD at the locus of interest. The transition matrix is

$$T(\theta) = \begin{bmatrix} \psi^2 & 2\psi\bar{\psi} & \bar{\psi}^2 \\ \psi\bar{\psi} & \psi^2 + \bar{\psi}^2 & \psi\bar{\psi} \\ \bar{\psi}^2 & 2\psi\bar{\psi} & \psi^2 \end{bmatrix},$$

where $\psi = \theta^2 + (1-\theta)^2$ and $\bar{\psi} = 1 - \psi$. Figure 4 is a barycentric representation of curves $\mathscr{C}_\pi(\theta) = \{\rho(\theta, \pi) = \pi T(\theta)\colon 0 \le \theta \le \frac{1}{2}\}$ for the sib-pair transition matrix $T(\theta)$ and for $\pi = (\pi_0, \pi_1, \pi_2)$ on the boundaries of the simplex. The infinitesimal generator is

$$Q = \begin{bmatrix} -4 & 4 & 0 \\ 2 & -4 & 2 \\ 0 & 4 & -4 \end{bmatrix}.$$

Q has eigenvalues $\lambda = 0, -4$ and -8. The left and right eigenvectors of Q corresponding to $\lambda_2 = -4$ are $(1/2\sqrt{2}, 0, -1/2\sqrt{2})$ and $(\sqrt{2}, 0, -\sqrt{2})$, respectively

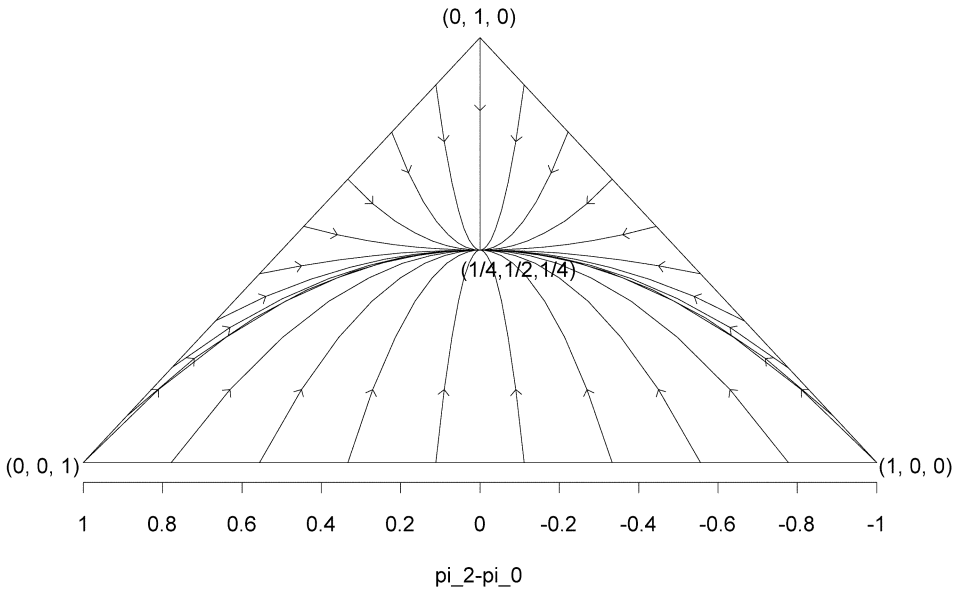

FIG. 4. *$S_2 \times D_4$-Barycentric representation of curves $\mathscr{C}_\pi(\theta)$, $0 \le \theta \le \frac{1}{2}$, for π on boundaries of simplex.*

(right eigenvector has unit norm with respect to the inner product $\langle\ ,\ \rangle_\alpha$). Hence

$$U = 8P_{-4} = 8\begin{bmatrix} \sqrt{2} \\ 0 \\ -\sqrt{2} \end{bmatrix}\left[\frac{1}{2\sqrt{2}}, 0, -\frac{1}{2\sqrt{2}}\right] = \begin{bmatrix} 4 & 0 & -4 \\ 0 & 0 & 0 \\ -4 & 0 & 4 \end{bmatrix}.$$

If we let N_i denote the number of affected sib-pairs sharing DNA IBD on i chromosomes at the marker, $i = 0, 1, 2$, then the score statistic for affected sib-pairs is

$$16(\pi_2 - \pi_0)(N_2 - N_0),$$

similarly for discordant and unaffected sib-pairs. Note that $N_2 - N_0$ may be rewritten as $N_2 - (n - N_1 - N_2) = 2(N_2 + \frac{1}{2}N_1) - n$. Thus, the score test is based on $S_{\text{pairs}} = N_2 + \frac{1}{2}N_1$, also known as the *mean IBD statistic*. These findings extend the work of Knapp, Seuchter and Baur (1994a), who proved the local optimality of the mean IBD statistic for affected sib-pairs under a single DS locus model with random mating and Hardy–Weinberg equilibrium. The score statistic for combining IBD data from affected, discordant and unaffected sib-pairs is a linear combination of the mean IBD statistics for each type of sib-pair, with weights $\pi_2 - \pi_0$ depending on the genetic model.

Note that this setup also applies to testing linkage to quantitative trait loci. In this case, each sib-pair may have different continuous phenotypes, and the score statistic is given by

$$16\sum_i(\pi_{2i} - \pi_{0i})(N_{2i} - N_{0i}),$$

where $N_{ji} = 1$ if the ith sib-pair shares DNA IBD on $j = 0, 1, 2$ chromosomes at the marker and $N_{ji} = 0$ otherwise, and π_{ji} is the conditional probability that the ith sib-pair shares DNA IBD on j chromosomes at the trait locus given the phenotypes of the sibs [Dudoit (1999)].

5.2. *Sib-pairs, orbits of* $S_2 \times (C_2 \times C_2)$. For sib-pairs with any number of affecteds and distinguishing between sharing of maternal and paternal DNA, there are four distinct IBD configurations, conveniently labeled by the pair (i, j), $i, j = 0, 1$, where i and j denote the number of paternally and maternally inherited chromosomes sharing DNA IBD, respectively. The transition matrix is

$$T(\theta) = \begin{bmatrix} \psi^2 & \psi\bar{\psi} & \psi\bar{\psi} & \bar{\psi}^2 \\ \psi\bar{\psi} & \psi^2 & \bar{\psi}^2 & \psi\bar{\psi} \\ \psi\bar{\psi} & \bar{\psi}^2 & \psi^2 & \psi\bar{\psi} \\ \bar{\psi}^2 & \psi\bar{\psi} & \psi\bar{\psi} & \psi^2 \end{bmatrix}$$

and the infinitesimal generator is

$$Q = \begin{bmatrix} -4 & 2 & 2 & 0 \\ 2 & -4 & 0 & 2 \\ 2 & 0 & -4 & 2 \\ 0 & 2 & 2 & -4 \end{bmatrix}.$$

Q has eigenvalues $\lambda = 0, -4, -4$ and -8. The two orthonormal right eigenvectors corresponding to $\lambda_2 = -4$ are $(\sqrt{2}, 0, 0, -\sqrt{2})$ and $(0, \sqrt{2}, -\sqrt{2}, 0)$, hence

$$\begin{aligned} U = 8P_{-4} &= 8\begin{bmatrix} \sqrt{2} \\ 0 \\ 0 \\ -\sqrt{2} \end{bmatrix}\left[\frac{1}{2\sqrt{2}}, 0, 0, -\frac{1}{2\sqrt{2}}\right] + 8\begin{bmatrix} 0 \\ \sqrt{2} \\ -\sqrt{2} \\ 0 \end{bmatrix}\left[0, \frac{1}{2\sqrt{2}}, -\frac{1}{2\sqrt{2}}, 0\right] \\ &= \begin{bmatrix} 4 & 0 & 0 & -4 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -4 & 0 & 0 & 4 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 4 & -4 & 0 \\ 0 & -4 & 4 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \end{aligned}$$

Let N_{ij} denote the number of affected sib-pairs sharing DNA IBD on i paternal and j maternal chromosome at the marker, $i, j = 0, 1$. The score statistic is given by

$$16(\pi_{11} - \pi_{00})(N_{11} - N_{00}) + 16(\pi_{10} - \pi_{01})(N_{10} - N_{01}),$$

and in general depends on the genetic model, similarly for discordant and unaffected sib-pairs. When $\pi_{10} = \pi_{01}$, the score test is based on $N_{11} - N_{00}$, that is, $N_2 - N_0$ in the more usual notation.

5.3. *Affected sib-trios, orbits of $S_3 \times D_4$.* For affected sib-trios (ASTs), there are four IBD configurations, with representative inheritance vectors and labels [defined as in Ethier and Hodge (1985)] listed in Table 3.

TABLE 3
IBD configurations for affected sib-trios

IBD configuration i	**Representative inheritance vector**	**Label**	$\lvert\mathcal{O}_i\rvert$
1	(1, 3, 1, 3, 1, 3)	(0, 0, 0)	4
2	(1, 3, 1, 3, 1, 4)	(0, 0, 1)	24
3	(1, 3, 1, 4, 2, 3)	(0, 1, 1)	24
4	(1, 3, 1, 3, 2, 4)	(1, 1, 1)	12

The transition matrix $T(\theta)$ is given by

$$\begin{bmatrix} (1-3\theta+3\theta^2)^2 & 6\theta\bar{\theta}(1-3\theta+3\theta^2) & 6\theta^2\bar{\theta}^2 & 3\theta^2\bar{\theta}^2 \\ \theta\bar{\theta}(1-3\theta+3\theta^2) & 1-4\theta+10\theta^2-12\theta^3+6\theta^4 & 2\theta\bar{\theta}(1-\theta+\theta^2) & \theta\bar{\theta}(1-\theta+\theta^2) \\ \theta^2\bar{\theta}^2 & 2\theta\bar{\theta}(1-\theta+\theta^2) & 1-4\theta+10\theta^2-12\theta^3+6\theta^4 & \theta\bar{\theta}(2-5\theta+5\theta^2) \\ \theta^2\bar{\theta}^2 & 2\theta\bar{\theta}(1-\theta+\theta^2) & 2\theta\bar{\theta}(2-5\theta+5\theta^2) & 1-6\theta+17\theta^2-22\theta^3+11\theta^4 \end{bmatrix}$$

and the infinitesimal generator is

$$Q = \begin{bmatrix} -6 & 6 & 0 & 0 \\ 1 & -4 & 2 & 1 \\ 0 & 2 & -4 & 2 \\ 0 & 2 & 4 & -6 \end{bmatrix}.$$

Q has eigenvalues $\lambda = 0, -4, -8, -8$, and the left and right eigenvectors corresponding to $\lambda_2 = -4$ are $\frac{1}{16}\sqrt{\frac{2}{3}}(3, 6, -6, -3)$ and $\sqrt{\frac{2}{3}}(3, 1, -1, -1)$, respectively. Hence

$$U = 8P_{-4} = 8\sqrt{\frac{2}{3}}\begin{bmatrix} 3 \\ 1 \\ -1 \\ -1 \end{bmatrix}\frac{1}{16}\sqrt{\frac{2}{3}}[3, 6, -6, -3] = \begin{bmatrix} 3 & 6 & -6 & -3 \\ 1 & 2 & -2 & -1 \\ -1 & -2 & 2 & 1 \\ -1 & -2 & 2 & 1 \end{bmatrix}.$$

Let N_i denote the number of ASTs with IBD configuration i at the marker, $i = 1, 2, 3, 4$. Then the score statistic for testing linkage is

$$S = \tfrac{16}{3}(3\pi_1 + \pi_2 - \pi_3 - \pi_4)(3N_1 + N_2 - N_3 - N_4).$$

Note that $3N_1 + N_2 - N_3 - N_4$ may be rewritten as $2(3N_1 + 2N_2 + N_3 + N_4) - 3n = 6S_{\text{pairs}} - 3n$.

5.4. *Discordant sib-trios, orbits of* $(S_1 \times S_2) \times D_4$. For discordant sib-trios (DSTs), where the first sib is the "odd" sib (i.e., the only affected sib or the only unaffected sib), there are seven IBD configurations, with representative inheritance vectors listed in Table 4.

TABLE 4
IBD configurations for discordant sib-trios

IBD configuration i	**Representative inheritance vector**	$\lvert\mathscr{O}_i\rvert$
1	(1, 3, 1, 3, 1, 3)	4
2	(1, 3, 1, 3, 1, 4)	16
3	(1, 3, 1, 3, 2, 4)	8
4	(1, 3, 1, 4, 2, 3)	8
5	(1, 3, 1, 4, 2, 4)	16
6	(1, 3, 1, 4, 1, 4)	8
7	(1, 3, 2, 4, 2, 4)	4

The infinitesimal generator is

$$Q = \begin{bmatrix} -6 & 4 & 0 & 0 & 0 & 2 & 0 \\ 1 & -5 & 1 & 1 & 1 & 1 & 0 \\ 0 & 2 & -6 & 2 & 2 & 0 & 0 \\ 0 & 2 & 2 & -6 & 2 & 0 & 0 \\ 0 & 1 & 1 & 1 & -5 & 1 & 1 \\ 1 & 2 & 0 & 0 & 2 & -6 & 1 \\ 0 & 0 & 0 & 0 & 4 & 2 & -6 \end{bmatrix}.$$

Q has eigenvalues $\lambda = 0, -4, -4, -8, -8, -8, -8$, and the two orthonormal right eigenvectors corresponding to $\lambda_2 = -4$ are $v = \sqrt{\frac{2}{3}}(-1, -1, -1, -1, 1, 1, 3)$ and $\tilde{v} = (1/\sqrt{3})(4, 1, -2, -2, -1, 2, 0)$. Hence

$$U = 8P_{-4} = \begin{bmatrix} 3 & 4 & -2 & -2 & -4 & 2 & -1 \\ 1 & 2 & 0 & 0 & -2 & 0 & -1 \\ -1 & 0 & 2 & 2 & 0 & -2 & -1 \\ -1 & 0 & 2 & 2 & 0 & -2 & -1 \\ -1 & -2 & 0 & 0 & 2 & 0 & 1 \\ 1 & 0 & -2 & -2 & 0 & 2 & 1 \\ -1 & -4 & -2 & -2 & 4 & 2 & 3 \end{bmatrix}.$$

Denote the ith column of U by u_i, then $u_1 + u_7 = -u_3 = -u_4 = u_6$ and $u_1 - u_7 = u_2 = -u_5$. Let N_i denote the number of DSTs with IBD configuration i at the marker, $i = 1, \ldots, 7$. Then, the score statistic for testing linkage is

$$\begin{aligned} S &= 8\Big(\sum_{j=1}^{7} v_j \pi_j\Big)\Big(\sum_{i=1}^{7} v_i N_i\Big) + 8\Big(\sum_{j=1}^{7} \tilde{v}_j \pi_j\Big)\Big(\sum_{i=1}^{7} \tilde{v}_i N_i\Big) \\ &= \frac{8}{3}\Big(2(-\pi_1 - \pi_2 - \pi_3 - \pi_4 + \pi_5 + \pi_6 + 3\pi_7) \\ &\qquad \times (-N_1 - N_2 - N_3 - N_4 + N_5 + N_6 + 3N_7) \\ &\qquad + (4\pi_1 + \pi_2 - 2\pi_3 - 2\pi_4 - \pi_5 + 2\pi_6) \\ &\qquad \times (4N_1 + N_2 - 2N_3 - 2N_4 - N_5 + 2N_6)\Big). \end{aligned}$$

6. Discussion. In this paper, we have derived score statistics for testing the null hypothesis of no linkage between a marker and a disease gene using identity by descent (IBD) data from sibships. We considered IBD configurations which are orbits of groups acting on the set of inheritance vectors, and proved that the transition matrix between IBD configurations satisfies a semigroup property (Proposition 1). For general pedigree types, we derived the

following spectral decomposition for the IBD configuration transition matrix $T(\theta)$:

$$T(\theta) = \sum_h (1 - 2\theta)^{-\lambda_h/2} P_h,$$

where λ_h and P_h are the eigenvalues and projection matrices of the infinitesimal generator Q, respectively, and λ_h are negative even integers (Propositions 2 and 3). By relating Q to the adjacency matrix of a quotient graph, we derived properties of its eigenvalues and eigenvectors. In general, the second largest eigenvalue of Q and its multiplicity determine the form of the score statistic for a given pedigree type. If the second largest eigenvalue of Q is $-2i$, the score test is based on the ith derivative of the log-likelihood, and if it has multiplicity one, the score statistic is independent of the genetic model (i.e., of the nuisance parameter π). For affected-only sibships of a given size, the second largest eigenvalue is -4 and has multiplicity one. As a result, the score test is based on the second derivative of the log-likelihood and is independent of the genetic model. Furthermore, the score statistic for affected-only sibships of a given size reduces to a well-known nonparametric statistic in linkage analysis, S_{pairs}.

Testing the null hypothesis of no linkage using IBD data, as we do here, is an instance of a general class of testing problems in which the null hypothesis is that a Markov chain has reached its stationary distribution [Diaconis (1988)]. The second largest eigenvalue of the infinitesimal generator not only determines the rate of convergence to the stationary distribution, but also plays an important role in hypothesis testing, as illustrated by our study.

Since the mid-1970's, linkage statistics have been the subject of numerous studies in the genetics literature. Earlier research focussed on IBD data from sib-pairs, and the sib-pair case still remains popular for theoretical work on linkage analysis and in practice for genetic studies. Several test statistics have been examined to test the null hypothesis of no linkage between a marker and a DS locus, based on the numbers (N_0, N_1, N_2) of affected sib-pairs sharing DNA IBD on 0, 1 and 2 chromosomes, respectively. These include the mean IBD statistic, that is, $S_{\text{pairs}} = N_2 + \frac{1}{2}N_1$ [Blackwelder and Elston (1985), Schaid and Nick (1990), Knapp, Seuchter and Baur (1994a, b)], N_2 [Day and Simons (1976), Blackwelder and Elston (1985) and Schaid and Nick (1990)], $N_2 + \frac{1}{4}N_1$ [Feingold and Siegmund (1997)], likelihood ratio statistics and χ^2 goodness-of-fit statistics, either unrestricted or restricted to the "possible triangle" [Risch (1990a, b), Holmans (1993), Faraway (1993), Feingold, Brown and Siegmund (1993), Holmans and Clayton (1995), Knapp, Seuchter and Baur (1994c) and Kruglyak and Lander (1995)]. Knapp, Seuchter and Baur (1994a) first pointed out the relationship of S_{pairs} to likelihood-based tests, in the special case of affected sib-pairs and genetic models with a single disease locus with random mating and Hardy–Weinberg equilibrium. In the past five years, nonparametric allele-sharing statistics for small pedigrees have been actively studied. Current methods for analyzing IBD data from small pedigrees rely on "scoring functions" which are defined as functions of the IBD configurations of affected pedigree members. [Whittemore and Halpern

(1994a), Kruglyak, Daly, Reeve-Daly and Lander (1996), Kong and Cox (1997), Teng and Siegmund (1997) and McPeek (1999)]. A few scoring functions have been suggested on empirical grounds, such as S_{pairs} and S_{all} [Whittemore and Halpern (1994a)], but the optimality and relationship of these scoring functions to likelihood analysis deserve further study. In addition, while it is clear that different pedigree types differ in their linkage information, IBD data from different pedigrees are typically combined by assigning equal weights to the standardized allele-sharing statistics of the various pedigrees. Different standardization and weighing schemes have been suggested, but the issue of combining IBD data from different pedigree types has not yet been thoroughly examined in the literature and remains an open problem. Whittemore (1996) proposed a likelihood-based approach to linkage analysis using multipoint marker data from general pedigrees. In the context of a genome scan, Whittemore considered the null hypothesis of no DS gene linked to the candidate locus against the alternative hypothesis of a DS gene *at* the candidate locus ($\theta = 0$). The null and alternative hypotheses were expressed in terms of a specificity parameter β, which is a function of genetic parameters, such as penetrances and allele frequencies. Whittemore proposed testing the null hypothesis of no DS gene linked to the candidate locus using a score test in β, that is, in the *specificity* parameter. Like our score test in the recombination fraction θ, this is a likelihood-based test which simultaneously treats the IBD data from all pedigrees. However, this type of score test depends intrinsically on the parameterization of the genetic model. This is a potentially serious shortcoming for complex diseases, for which there typically is no established knowledge regarding the number of DS loci, the penetrances, or the frequency of genotypes in the study population. Similarly, McPeek (1999) derived allele-sharing statistics for different types of alternative genetic models by focusing on specificity parameters.

It is important to note that in genome scans, where marker density is high, one expects to be testing near $\theta = 0$ rather than near $\theta = \frac{1}{2}$. If one has some knowledge concerning the mode of inheritance of the disease (e.g., rare recessive disease), then making use of this information in deriving allele-sharing statistics as in McPeek (1999) is desirable. However, the main obstacle in obtaining optimal statistics for an alternative hypothesis where $\theta = 0$ is precisely their dependence on the genetic model, which is usually unknown. By contrast, our score test in the recombination fraction θ focuses on the *scale* component and seems to achieve some robustness with respect to specificity. Although our score statistic in θ is locally most powerful near $\theta = \frac{1}{2}$, S_{pairs} was also found to be optimal for some classes of genetic models when $\theta = 0$ by McPeek (1999). A recent study by Davis and Weeks (1997) on affected sib-pairs found S_{pairs} to perform well for a variety of two-locus genetic models. Our preliminary work on applying the sib-pair score test in θ to the linkage analysis of quantitative traits also demonstrates the power of this type of score test for nonlocal alternatives (Dudoit (1999)). However, when the alternative hypothesis is such that $\theta = 0$, the optimal statistic depends on the genetic model, and there are inevitably classes of models for which S_{pairs} is outperformed by

other statistics. For example, Feingold and Siegmund (1997) reported that for sib-pairs N_2 performed better than S_{pairs} for rare heterogeneous traits. For larger sibships, the performance of S_{pairs} needs to be further investigated.

As mentioned earlier, an important issue is the combination of IBD data from different pedigree types. The score test in θ presented in this paper produces weights for combining IBD data from different types of sibships ($\sum_j v_j \pi_j$), but these weights are linear combinations of the IBD probabilities, π, which depend on the usually unknown genetic model. Our work on quantitative traits revealed that the sib-pair score test was robust to the choice of model for computing the weights $\pi_2 - \pi_0$. For sib-pairs sampled randomly with respect to their phenotypes, the score test was found to be similar in power to the widely used Haseman–Elston test, and for sib-pairs selected with extreme concordant or discordant phenotypes, the score test was found to be far more powerful than the Haseman–Elston test [cf. Haseman and Elston (1972), Risch and Zhang (1995, 1996), Dudoit (1999)].

We believe that score statistics in θ as described in this paper have the potential to be useful in practice for linkage analysis using IBD data from sibships and relative pairs, when the mode of inheritance is unknown. We have derived score statistics in θ for IBD data from other types of relative pairs (e.g., cousin pairs), using the IBD configurations considered by Donnelly (1983) and the general framework of Propositions 1 and 2 [Dudoit (1999)]. These score statistics may readily be extended to accommodate incomplete IBD data by replacing the IBD counts by their expected value given multipoint marker data [cf. inheritance distribution of Kruglyak and Lander (1995) and Kruglyak, Daly, Reeve-Daly and Lander (1996)]. We were very encouraged by our preliminary studies with sib-pairs and will further explore the issue of combining IBD data from different sibship types and relative pairs.

APPENDIX

A. Pólya theory of counting. Let A and B be finite sets, $|A| = n$, and let G and H be finite groups, G acting on A and H on B. By Theorem 35.3 in van Lint and Wilson (1992), the number of orbits of $G \times H$ acting on B^A, the set of mappings from A to B, is given by

$$\frac{1}{|H|} \sum_{\tau \in H} Z_G(m_1(\tau), \ldots, m_n(\tau)),$$

where

$$m_i(\tau) := \sum_{j|i} j z_j(\tau), \qquad i = 1, \ldots, n,$$

$$z_j(\tau) := \text{number of cycles of } \tau \text{ having length } j, \qquad j = 1, \ldots, |B|.$$

For a group G acting on a set of n elements, the *cycle index* Z_G is a polynomial in n letters, $X_1, \ldots, X_n$, defined by

$$Z_G(X_1, \ldots, X_n) := \frac{1}{|G|} \sum_{\sigma \in G} X_1^{z_1(\sigma)} \cdots X_n^{z_n(\sigma)}.$$

The cycle index for the symmetric group on n letters is

$$Z_{S_n}(X_1, \ldots, X_n) = \sum_{(1^{k_1} \cdots n^{k_n})} \frac{1}{1^{k_1} \cdots n^{k_n} k_1! \cdots k_n!} X_1^{k_1} \cdots X_n^{k_n},$$

where $(1^{k_1} \cdots n^{k_n})$ denotes a partition of n with k_i parts of size i, $i = 1, \ldots, n$. $Z_{S_n}(X_1, \ldots, X_n)$ is also the coefficient of z^n in the expansion of $\exp(\sum_{i=1}^{\infty}(z^i/i)X_i)$ [cf. deBruijn (1964), page 147]; this is the formula which is most appropriate for our problem. In our problem, we wish to determine the number of orbits of $G \times H$ acting on the set of mappings $\{a, b, c, d\}^{\{1,\ldots,k\}}$ (i.e., the set of inheritance vectors), where $H = D_4$ or $C_2 \times C_2$ and $G = S_k$ or $S_h \times S_{k-h}$. Table 5 lists for each permutation in D_4 the number of cycles of length j, $j = 1, 2, 3, 4$. The m_i's of Table 6 are calculated using Table 5, and the fact that for $\tau \in D_4$ and $i \geq 0$,

$$\begin{aligned} m_4(\tau) &= m_{4i+4}(\tau), \\ m_1(\tau) &= m_{4i+1}(\tau) = m_{4i+3}(\tau), \\ m_2(\tau) &= m_{4i+2}(\tau). \end{aligned}$$

Table 7 lists for each permutation in D_4 the cycle index $Z_{S_k}(m_1(\tau), \ldots, m_k(\tau))$. When $H = C_2 \times C_2$, only the first and third rows of Table 7 are used. When $G = S_h \times S_{k-h}$, we note that the cycle index polynomial of the direct product of two groups is simply the product of the cycle indices of the two groups and Table 7 may be used again. The number of IBD configurations for the four groups are listed below.

$S_k \times D_4$:

$$m = \begin{cases} (k+1)(k+3)(k+5)/48, & k \text{ odd}, \\ (k+2)(k^2+7k+18)/48, & k \text{ even and } k/2 \text{ odd}, \\ (k+4)(k^2+5k+12)/48, & k \text{ even and } k/2 \text{ even}, \end{cases}$$

which agrees with equation (5) of Ethier and Hodge (1985).

TABLE 5

Cycles of D_4. $z_j(\tau)$ denotes the number of cycles of τ having length j *

Permutation τ	$z_1(\tau)$	$z_2(\tau)$	$z_3(\tau)$	$z_4(\tau)$
ι	4	0	0	0
$\rho_1 = (cadb)$	0	0	0	1
$\rho_2 = (ad)(bc)$	0	2	0	0
$\rho_3 = (bdac)$	0	0	0	1
$\mu_1 = (ab)(cd)$	0	2	0	0
$\mu_2 = (ac)(bd)$	0	2	0	0
$\delta_1 = (bc)$	2	1	0	0
$\delta_2 = (ad)$	2	1	0	0

*The elements of D_4 are listed according to the notation of Fraleigh [(1989), page 70].

TABLE 6
$m_i(\tau)$ *for* $\tau \in D_4$

Permutation τ	$m_{4i+1}(\tau) = m_{4i+3}(\tau)$	$m_{4i+2}(\tau)$	$m_{4i+4}(\tau)$
ι	4	4	4
ρ_1, ρ_3	0	0	4
ρ_2, μ_1, μ_2	0	4	4
δ_1, δ_2	2	4	4

$S_k \times (C_2 \times C_2)$:

$$m = \begin{cases} (k+3)(k+2)(k+1)/24, & k \text{ odd}, \\ (k+2)(k^2+4k+12)/24, & k \text{ even}. \end{cases}$$

$(S_h \times S_{k-h}) \times D_4$:

$$m = \frac{1}{8}\Bigg[\binom{h+3}{3}\binom{k-h+3}{3} + 2I(4|h)I(4|k-h) + \frac{3}{4}I(2|h)I(2|k-h)(h+2)(k-h+2) + \frac{1}{8}(I(2|h)+(h+1)(h+3))(I(2|k-h)+(k-h+1)(k-h+3))\Bigg],$$

where $I(j|i) = 1$ if j divides i and 0 otherwise.

$(S_h \times S_{k-h}) \times (C_2 \times C_2)$:

$$m = \frac{1}{4}\Bigg[\binom{h+3}{3}\binom{k-h+3}{3} + \frac{3}{4}I(2|h)I(2|k-h)(h+2)(k-h+2)\Bigg].$$

For any inheritance vector x, let pat denote the less frequent of the paternal labels 1 and 2, and similarly let mat denote the less frequent of the maternal labels 3 and 4. The number of paternal labels pat in the inheritance vector x is denoted by $|pat|$, similarly for the maternal labels. The number of sibs with a pair of labels (pat, mat) is denoted by $|(pat, mat)|$. Ethier and Hodge (1985) define the label of the particular inheritance vector x to be the triple (l_1, l_2, l_3) where

$$l_1 = |(pat, mat)|, \qquad l_2 = \min(|pat|, |mat|), \qquad l_3 = \max(|pat|, |mat|).$$

TABLE 7
$Z_{S_k}(m_1(\tau), \ldots, m_k(\tau))$ *for* $\tau \in D_4$. $I(\ )$ *is the indicator function*

Permutation τ	$\exp\left(\sum_{i=1}^{\infty} m_i(\tau)\frac{z^i}{i}\right)$	$Z_{S_k}(m_1(\tau), \ldots, m_k(\tau))$
ι	$(1-z)^{-4}$	$\binom{k+3}{3}$
ρ_1, ρ_3	$(1-z^4)^{-1}$	$I(4\|k)$
ρ_2, μ_1, μ_2	$(1-z^2)^{-2}$	$I(2\|k)(k+2)/2$
δ_1, δ_2	$(1-z)^{-2}(1-z^2)^{-1}$	$\frac{1}{4}(I(2\|k)+(k+1)(k+3))$

For example, if $x = (1, 3, 1, 4, 2, 3)$, the less frequent of the paternal labels is 2, thus $pat = 2$ and $|pat| = 1$. Similarly, $mat = 4$ and $|mat| = 1$. $(mat, pat) = (2, 4)$ and the number of sibs with pair of labels $(2, 4)$ is 0. Thus $l_1 = 0$, $l_2 = 1$ and $l_3 = 1$. In ambiguous cases such as $|1| = |2| = k/2$ or $|3| = |4| = k/2$, Ethier and Hodge suggest making a choice that results in $l_1 \geq l_2/2$. Then the triple satisfies

$$0 \leq l_1 \leq l_2 \leq l_3 \leq k/2, \quad \text{and} \quad l_1 \geq l_2/2 \text{ if } l_3 = k/2.$$

We can modify the labeling of Ethier and Hodge for the orbits of $S_k \times (C_2 \times C_2)$ and let

$$l_1 = |(pat, mat)|, \qquad l_2 = |pat|, \qquad l_3 = |mat|.$$

B. Transition matrix for IBD configurations.

B.1. *Proof of Proposition* 1. Let $\bar{\theta} = 1 - \theta$. We first prove the semigroup property for the transition matrix $R(\theta)$ of inheritance vectors. Let $\Delta = \Delta(x, y)$, then

$$\begin{aligned} r_{xy}(\theta_1 * \theta_2) &= (\theta_1 * \theta_2)^{\Delta}(1 - (\theta_1 * \theta_2))^{2k-\Delta} \\ &= (\theta_1\bar{\theta}_2 + \bar{\theta}_1\theta_2)^{\Delta}(\theta_1\theta_2 + \bar{\theta}_1\bar{\theta}_2)^{2k-\Delta} \\ &= \sum_{i=0}^{\Delta} \sum_{j=0}^{2k-\Delta} \binom{\Delta}{i} \binom{2k-\Delta}{j} \theta_1^{i+j} \bar{\theta}_1^{\,2k-(i+j)} \theta_2^{\Delta-i+j} \bar{\theta}_2^{\,2k-(\Delta-i+j)}. \end{aligned}$$

Also,

$$\sum_z r_{xz}(\theta_1) r_{zy}(\theta_2) = \sum_z \theta_1^{\Delta(x,z)} \bar{\theta}_1^{\,2k-\Delta(x,z)} \theta_2^{\Delta(y,z)} \bar{\theta}_2^{\,2k-\Delta(y,z)}.$$

Now, for $i = 0, \ldots, \Delta$, $j = 0, \ldots, 2k - \Delta$, divide the set of all 2^{2k} inheritance vectors into groups of $\binom{\Delta}{i}\binom{2k-\Delta}{j}$ inheritance vectors z, such that z differs from x at i of the Δ positions at which x and y differ and z differs from x at j of the $2k - \Delta$ positions at which x and y agree. Then, $\Delta(x, z) = i + j$, $\Delta(y, z) = (\Delta - i) + j$, and

$$\begin{aligned} &\sum_z \theta_1^{\Delta(x,z)} \bar{\theta}_1^{\,2k-\Delta(x,z)} \theta_2^{\Delta(y,z)} \bar{\theta}_2^{\,2k-\Delta(y,z)} \\ &\quad = \sum_{i=0}^{\Delta} \sum_{j=0}^{2k-\Delta} \binom{\Delta}{i} \binom{2k-\Delta}{j} \theta_1^{i+j} \bar{\theta}_1^{\,2k-(i+j)} \theta_2^{\Delta-i+j} \bar{\theta}_2^{\,2k-(\Delta-i+j)}. \end{aligned}$$

Therefore,

$$r_{xy}(\theta_1 * \theta_2) = \sum_z r_{xz}(\theta_1) r_{zy}(\theta_2).$$

Consider now the transition matrix for IBD configurations. From (3.2),

$$\begin{aligned} t_{ij}(\theta_1 * \theta_2) &= \sum_{y \in \mathscr{C}_j} r_{xy}(\theta_1 * \theta_2) \quad (\text{where } x \text{ is any } x \in \mathscr{C}_i) \\ &= \sum_{y \in \mathscr{C}_j} \sum_{z} r_{xz}(\theta_1) r_{zy}(\theta_2) = \sum_{y \in \mathscr{C}_j} \sum_{l} \sum_{z \in \mathscr{C}_l} r_{xz}(\theta_1) r_{zy}(\theta_2) \\ &= \sum_{l} \sum_{z \in \mathscr{C}_l} r_{xz}(\theta_1) \sum_{y \in \mathscr{C}_j} r_{zy}(\theta_2) \\ &= \sum_{l} t_{lj}(\theta_2) \sum_{z \in \mathscr{C}_l} r_{xz}(\theta_1) \\ &= \sum_{l} t_{il}(\theta_1) t_{lj}(\theta_2). \end{aligned}$$

Hence, $T(\theta)$ satisfies the semigroup property $T(\theta_1 * \theta_2) = T(\theta_1)T(\theta_2)$. Now $T(\theta)$ is differentiable and for $\theta \neq \frac{1}{2}$,

$$\begin{aligned} \frac{T(\theta + h(1-2\theta)) - T(\theta)}{h(1-2\theta)} &= \frac{T(\theta * h) - T(\theta)}{h(1-2\theta)} \\ &= \left(\frac{T(\theta)}{1-2\theta}\right)\left(\frac{T(h) - I}{h}\right) = \left(\frac{T(h) - I}{h}\right)\left(\frac{T(\theta)}{1-2\theta}\right). \end{aligned}$$

Thus $T'(\theta)$, the matrix of first derivatives of the transition probabilities, is given by

$$T'(\theta) = \lim_{h \to 0} \frac{T(\theta + h) - T(\theta)}{h} = \lim_{h \to 0} \frac{T(\theta + h(1-2\theta)) - T(\theta)}{h(1-2\theta)},$$

that is,

$$T'(\theta) = \frac{T(\theta)}{1-2\theta}\, T'(0) = T'(0)\, \frac{T(\theta)}{1-2\theta},$$

and hence

$$T(\theta) = e^{d(\theta)Q},$$

where $d(\theta) = -\frac{1}{2}\ln(1-2\theta)$ is the inverse of the Haldane map function and $Q = T'(0)$ is the infinitesimal generator. Q has entries,

$$q_{ij} = \sum_{y \in \mathscr{C}_j} (-2k\, I(\Delta(x, y) = 0) + I(\Delta(x, y) = 1)) = \sum_{y \in \mathscr{C}_j} I(\Delta(x, y) = 1) - 2k\delta_{ij},$$

where x is any inheritance vector in $\mathscr{C}_i$ and $\delta_{ij} = 1$ if $i = j$ and 0 otherwise. Then Q may be written as $Q = B - 2kI$, where B is the $m \times m$ matrix with entries

$$b_{ij} = \sum_{y \in \mathscr{C}_j} I(\Delta(x, y) = 1) \quad \text{for any } x \in \mathscr{C}_i.$$

$T(\theta)$ satisfies

$$|\mathscr{C}_i| t_{ij}(\theta) = |\mathscr{C}_j| t_{ji}(\theta);$$

hence, the stationary distribution of $T(\theta)$ is

$$\alpha = (\alpha_1, \ldots, \alpha_m) = \frac{1}{2^{2k}} \big(|\mathscr{C}_1|, \ldots, |\mathscr{C}_m|\big),$$

since

$$\sum_i \alpha_i t_{ij}(\theta) = \sum_i \alpha_j t_{ji}(\theta) = \alpha_j. \qquad \square$$

B.2. *Proof of Proposition* 2. Q satisfies the reversibility condition $\alpha_i q_{ij} = \alpha_j q_{ji}$, hence Q is self-adjoint with respect to the real inner product $\langle x, y\rangle_\alpha = \sum_i \alpha_i x_i y_i$ on $\mathbb{R}^m$. Hence, from the principal axis theorem [cf. Jacob (1990), page 288], Q has an orthonormal basis of eigenvectors with only real eigenvalues, λ_h, $h = 1, \ldots, m$ (not necessarily distinct). Denote the hth right eigenvector by $\mathbf{v}_h$ and its ith entry by v_{ih}. Then $\langle \mathbf{v}_h, \mathbf{v}_l\rangle_\alpha = \sum_i \alpha_i v_{ih} v_{il} = \delta_{hl}$. Since Q is reversible, the row vector $\mathbf{w}_h$ with ith entry $w_{hi} = \alpha_i v_{ih}$ is the left eigenvector of Q corresponding to the hth eigenvalue. Hence Q may be written as

$$Q = \sum_h \lambda_h P_h,$$

where

$$(P_h)_{ij} = v_{ih} w_{hj} = \alpha_j v_{ih} v_{jh},$$

that is,

$$P_h = \mathbf{v}_h \mathbf{w}_h.$$

The projection matrices satisfy $P_h^2 = P_h = P_h^*$, $P_h P_l = 0$, $h \neq l$ and $\sum_h P_h = I$, where P_h^* is the adjoint of P_h with respect to $\langle\ ,\ \rangle_\alpha$. It follows that

$$T(\theta) = \sum_h \exp(\lambda_h d(\theta))\, P_h = \sum_h (1 - 2\theta)^{-\lambda_h/2}\, P_h. \qquad \square$$

C. Adjacency matrix of quotient graph $\mathscr{X}/(\boldsymbol{G} \times \boldsymbol{H})$. Consider the graph $\mathscr{X}$ with vertex set the set of all inheritance vectors of length $2k$ and adjacency matrix $A(\mathscr{X}) = A$ with (x, y)-entry,

$$a_{xy} = \begin{cases} 1, & \text{if } \Delta(x, y) = 1, \\ 0, & \text{otherwise.} \end{cases}$$

To describe the eigenvectors of A it is convenient to code the inheritance vectors $x = (x_1, x_2, \ldots, x_{2k})$ as in a 2^{2k} factorial experiment, where $x_{2i-1} = 1$ when factor $2i - 1$ is absent and 2 when it is present, and $x_{2i} = 3$ when factor $2i$ is absent and 4 when it is present. The eigenvectors of A have the following patterns.

PROPOSITION 3 (Eigenvalues and eigenvectors of adjacency matrix A). *The eigenvector corresponding to the eigenvalue $\lambda = 2k$ is the grand mean term $V_0 = (1, 1, \ldots, 1)^T$. The eigenvectors corresponding to the eigenvalue $\lambda = 2k-2$ are the $2k$ main effect terms, $V_1, V_2, \ldots, V_{2k}$, where*

$$V_{2i-1}(x) = I(x_{2i-1} = 2) - I(x_{2i-1} = 1),$$
$$V_{2i}(x) = I(x_{2i} = 4) - I(x_{2i} = 3).$$

The eigenvectors corresponding to the eigenvalue $\lambda = 2k-4$ are the $\binom{2k}{2}$ 2-factor interactions, V_{ij}, $1 \le i < j \le 2k$, where

$$V_{ij}(x) = V_i(x)V_j(x).$$

In general, the eigenvectors corresponding to the eigenvalue $\lambda = 2(k-i)$, $i = 0, \ldots, 2k$, are the $\binom{2k}{i}$ i-factor interactions, $V_{j_1, j_2, \ldots, j_i}$, $1 \le j_1 < j_2 < \cdots < j_i \le 2k$, where

$$V_{j_1, j_2, \ldots, j_i}(x) = V_{j_1}(x)V_{j_2}(x)\cdots V_{j_i}(x).$$

Let H denote the matrix with rows the 2^{2k} eigenvectors of A described above. Then, H is an Hadamard matrix; that is, its entries are 1 and -1 and $HH^T = 2^{2k}I$.

PROOF (Partial). We need not distinguish the parental origin of the DNA, hence, for simplicity, denote 1's and 3's by 0's and 2's and 4's by 1's. Then

$$V_i(x) = I(x_i = 1) - I(x_i = 0) = 2\,I(x_i = 1) - 1.$$

$\lambda = 2k$: the rows of A sum to $2k$ hence $\lambda = 2k$ is an eigenvalue of A with eigenvector V_0.

$\lambda = 2k - 2$:

$$\begin{aligned}\sum_y a_{xy}V_i(y) &= \sum_y I(\Delta(x, y) = 1)(2\,I(y_i = 1) - 1)\\ &= 2\sum_y I(\Delta(x, y) = 1, y_i = 1) - 2k\\ &= 2(I(x_i = 1)(2k-1) + I(x_i = 0)) - 2k\\ &= 2((2k-2)\,I(x_i = 1) + 1) - 2k\\ &= (2k-2)(2\,I(x_i = 1) - 1) = (2k-2)V_i(x).\end{aligned}$$

Hence $\lambda = 2k-2$ is an eigenvalue of A with eigenvectors V_i, $i = 1, \ldots, 2k$. It is easy to show that $\langle V_i, V_j\rangle = 2^{2k}\delta_{ij}$.

$\lambda = 2k - 4$:

$$\begin{aligned}
&\sum_y a_{xy} V_i(y) V_j(y) \\
&\quad = \sum_y I(\Delta(x, y) = 1)(2\, I(y_i = 1) - 1)(2\, I(y_j = 1) - 1) \\
&\quad = \sum_y I(\Delta(x, y) = 1)(4\, I(y_i = 1,\ y_j = 1) - 2\, I(y_i = 1) - 2\, I(y_j = 1) + 1) \\
&\quad = 4(I(x_i = 1,\ x_j = 1)(2k - 2) + I(x_i = 1,\ x_j = 0) + I(x_i = 0,\ x_j = 1)) \\
&\qquad - 2(I(x_i = 1)(2k - 1) + I(x_i = 0)) \\
&\qquad - 2(I(x_j = 1)(2k - 1) + I(x_j = 0)) + 2k \\
&\quad = 4I(x_i = 1,\ x_j = 1)(2k - 2) \\
&\qquad + 4(I(x_i = 1) - I(x_i = 1,\ x_j = 1)) \\
&\qquad + 4(I(x_j = 1) - I(x_i = 1,\ x_j = 1)) \\
&\qquad - 2((2k - 2)I(x_i = 1) + 1) - 2((2k - 2)I(x_j = 1) + 1) + 2k \\
&\quad = (2k - 4)(4I(x_i = 1,\ x_j = 1) - 2I(x_i = 1) - 2I(x_j = 1) + 1) \\
&\quad = (2k - 4)V_i(x)V_j(x).
\end{aligned}$$

Hence $\lambda = 2k - 4$ is an eigenvalue of A with eigenvectors V_{ij}. □

In order to prove Theorem 1, we rely on the following general facts concerning quotient graphs [Chapter 5 in Godsil (1993)]. Consider a group $G \times H$ acting on the vertices of $\mathscr{X}$, as described in Table 2. Then, by the same argument as that leading to (3.1), the orbits of $G \times H$, $\mathscr{C}_i$, $i = 1, \ldots, m$, form an equitable partition of the vertex set of $\mathscr{X}$. The matrix B defined in Proposition 1 is the *adjacency matrix of the quotient graph* $\mathscr{X}/(G \times H)$, which is the multidigraph with the orbits of $G \times H$ as its vertices and with b_{ij} arcs going from $\mathscr{C}_i$ to $\mathscr{C}_j$. Let C denote the *characteristic matrix* of the partition $(\mathscr{C}_i)$; C is a $2^{2k} \times m$ matrix, with ijth entry 1 or 0 according as the ith vertex of $\mathscr{X}$ is contained in the orbit $\mathscr{C}_j$ or not.

FACT 1 [Based on Lemma 2.2 in Godsil (1993)]. The eigenvalues of B are a subset of the eigenvalues of A.

FACT 2 [Based on Lemma 2.2 in Godsil (1993)]. If v is an eigenvector of B, then Cv is an eigenvector of A which is constant over the orbits of $G \times H$, with entry v_i on $\mathscr{C}_i$.

FACT 3. If V is an eigenvector of A which is constant over the orbits of $G \times H$, with $V(x) = v_i\ \forall\ x \in \mathscr{C}_i$, then the vector v, with ith entry v_i, is an eigenvector of B.

PROOF. For any $x \in \mathscr{C}_i$

$$\lambda v_i = \lambda V(x) = \sum_y a_{xy} V(y) = \sum_j v_j \sum_{y \in \mathscr{C}_j} a_{xy} = \sum_j v_j b_{ij}. \qquad \square$$

The proof of Theorem 1 also relies on the following specific properties of the eigenvectors of A on the orbits of $G \times H$.

C.1. *Quotient graph* $\mathscr{X}/(S_k \times D_4)$.

FACT 4. The $2k$ eigenvectors of A corresponding to the eigenvalue $2k-2$ sum to 0 over the orbits of $S_k \times D_4$, that is, $\forall\ i = 1, \ldots, 2k$, and any orbit $\mathscr{C}$,

$$\sum_{x \in \mathscr{C}} V_i(x) = 0.$$

PROOF. Let $\iota \in S_k$ denote the identity permutation and as before, let $\alpha = (ac)(bd)$ denote the permutation of D_4 which corresponds to interchanging the paternal labels 1 and 2. Let $\tilde{x} = (\iota, \alpha)(x)$ denote the inheritance vector obtained from x by interchanging the paternal labels. Then, for $1 \le i \le k$,

$$\begin{aligned} V_{2i-1}(x) &= (I(x_{2i-1} = 2) - I(x_{2i-1} = 1)) \\ &= (I(\tilde{x}_{2i-1} = 1) - I(\tilde{x}_{2i-1} = 2)) = -V_{2i-1}(\tilde{x}), \end{aligned}$$

and since applying (ι, α) to the elements of $\mathscr{C}$ results in a permutation of the inheritance vectors in $\mathscr{C}$, then

$$\sum_{x \in \mathscr{C}} V_{2i-1}(x) = -\sum_{x \in \mathscr{C}} V_{2i-1}(\tilde{x}) = -\sum_{x \in \mathscr{C}} V_{2i-1}(x).$$

Consequently,

$$\sum_{x \in \mathscr{C}} V_{2i-1}(x) = 0.$$

The proof for V_{2i} is similar, but uses the permutation β instead of α. $\square$

FACT 5. The k^2 eigenvectors of A corresponding to the eigenvalue $2k-4$ and involving "odd" and "even" factors sum to 0 over the orbits of $S_k \times D_4$, that is, $\forall\ i, j = 1, \ldots, k$, and any orbit $\mathscr{C}$,

$$\sum_{x \in \mathscr{C}} V_{2i-1}(x) V_{2j}(x) = 0.$$

PROOF. Here again, let $\tilde{x} = (\iota, \alpha)(x)$. Then

$$V_{2i-1}(x) V_{2j}(x) = (-V_{2i-1}(\tilde{x})) V_{2j}(\tilde{x})$$

and

$$\sum_{x \in \mathscr{C}} V_{2i-1}(x) V_{2j}(x) = -\sum_{x \in \mathscr{C}} V_{2i-1}(\tilde{x}) V_{2j}(\tilde{x}) = -\sum_{x \in \mathscr{C}} V_{2i-1}(x) V_{2j}(x).$$

Hence

$$\sum_{x \in \mathscr{C}} V_{2i-1}(x) V_{2j}(x) = 0. \qquad \square$$

FACT 6. Let

$$V(x) = \sum_{(i,\, j)} \left\{ V_{2i-1,\, 2j-1}(x) + V_{2i,\, 2j}(x) \right\},$$

where the sum is over all $\binom{k}{2}$ unordered pairs (i, j) of distinct integers ranging from 1 to k. Then V is an eigenvector of A corresponding to the eigenvalue $2k - 4$. Furthermore, V is constant over the orbits of $S_k \times D_4$, that is, for any orbit $\mathscr{C}$,

$$V(x) = V(\tilde{x}) \quad \text{whenever } x, \tilde{x} \in \mathscr{C}.$$

PROOF. Members of the same orbit are obtained by a combination of any of the following three operations: a permutation $\sigma \in S_k$ of the sibs, and permutations α and γ of the pairs of labels of all sibs simultaneously. We consider a particular configuration $\mathscr{C}$ and the effect of each operation separately on $x \in \mathscr{C}$.

$\tilde{x} = (\iota, \alpha)(x)$, where ι is the identity in S_k and $\alpha = (ac)(bd)$: for each pair (i, j),

$$V_{2i-1}(\tilde{x}) V_{2j-1}(\tilde{x}) + V_{2i}(\tilde{x}) V_{2j}(\tilde{x}) = (-V_{2i-1}(x))(-V_{2j-1}(x)) + V_{2i}(x) V_{2j}(x),$$

hence $V(\tilde{x}) = V(x)$.

$\tilde{x} = (\iota, \gamma)(x)$, where ι is the identity in S_k and $\gamma = (bc)$: for each $1 \le i \le k$,

$$\begin{aligned} I(\tilde{x}_{2i-1} = 1) &= I(\tilde{x}_{2i-1} = 1,\ \tilde{x}_{2i} = 3) + I(\tilde{x}_{2i-1} = 1,\ \tilde{x}_{2i} = 4) \\ &= I(x_{2i-1} = 1,\ x_{2i} = 3) + I(x_{2i-1} = 2,\ x_{2i} = 3) \\ &= I(x_{2i} = 3), \end{aligned}$$

and similarly

$$I(\tilde{x}_{2i-1} = 2) = I(x_{2i} = 4).$$

Hence, for $1 \le i \le k$,

$$V_{2i-1}(\tilde{x}) = V_{2i}(x), \tag{C.1}$$

and consequently $V(x) = V(\tilde{x})$.

$\tilde{x} = (\sigma, \iota)(x)$, where $\sigma \in S_k$ and ι is the identity in D_4: For $1 \le i \le k$, $\tilde{x}_{2i-1} = x_{2\sigma^{-1}(i)-1}$ and $\tilde{x}_{2i} = x_{2\sigma^{-1}(i)}$, thus

$$\begin{aligned} V(\tilde{x}) &= \sum_{(i,\, j)} \left\{ V_{2\sigma^{-1}(i)-1}(x) V_{2\sigma^{-1}(j)-1}(x) + V_{2\sigma^{-1}(i)}(x) V_{2\sigma^{-1}(j)}(x) \right\} \\ &= \sum_{(i,\, j)} \left\{ V_{2i-1}(x) V_{2j-1}(x) + V_{2i}(x) V_{2j}(x) \right\} = V(x). \end{aligned}$$

In particular, for $k > 1$,

$$V(1,3,1,3,\ldots,1,3) = \sum_{(i,j)} (-1)(-1) + (-1)(-1) = 2\binom{k}{2} = k(k-1) \neq 0.$$

Hence, since V is a linear combination of eigenvectors of A which is nonzero, then V is an eigenvector of A corresponding to the eigenvalue $2k-4$. Furthermore, V is constant on the orbits of $S_k \times D_4$. □

FACT 7. The $k(k-1)$ 2-factor eigenvectors $\{V_{2i-1,2j-1}, V_{2i,2j}: 1 \leq i < j \leq k\}$ have the same sums over the orbits of $S_k \times D_4$, that is, for any orbit $\mathscr{C}$ and $1 \leq i_1 < j_1 \leq k$, $1 \leq i_2 < j_2 \leq k$,

$$\sum_{x\in\mathscr{C}} V_{2i_2-1,2j_2-1}(x) = \sum_{x\in\mathscr{C}} V_{2i_1-1,2j_1-1}(x) = \sum_{x\in\mathscr{C}} V_{2i_1,2j_1}(x) = \sum_{x\in\mathscr{C}} V_{2i_2,2j_2}(x).$$

PROOF. Let $x \in \mathscr{C}$, then $\tilde{x} = (\iota, \gamma)(x) \in \mathscr{C}$, and by (C.1), for each $1 \leq i < j \leq k$,

$$\sum_{x\in\mathscr{C}} V_{2i-1}(x)V_{2j-1}(x) = \sum_{x\in\mathscr{C}} V_{2i}(\tilde{x})V_{2j}(\tilde{x}) = \sum_{x\in\mathscr{C}} V_{2i}(x)V_{2j}(x).$$

Also, consider any permutation $\sigma \in S_k$, then $\tilde{x} = (\sigma, \iota)(x) \in \mathscr{C}$ and

$$\sum_{x\in\mathscr{C}} V_{2i}(x)V_{2j}(x) = \sum_{x\in\mathscr{C}} V_{2i}(\tilde{x})V_{2j}(\tilde{x}) = \sum_{x\in\mathscr{C}} V_{2\sigma^{-1}(i)}(x)V_{2\sigma^{-1}(j)}(x).$$

Similarly for $V_{2i-1,2j-1}$.

PROPOSITON 4 [Eigenvalues of adjacency matrix B of quotient graph $\mathscr{X}/(S_k \times D_4)$]. *The two largest eigenvalues of Bare $2k$ and $2k-4$, and have multiplicity one. All other eigenvalues of B are strictly less than $2k-4$ and belong to the set $\{2(k-i)_{\binom{2k}{i}}: i = 3,\ldots,2k\}$, where $\binom{2k}{i}$ is the largest possible multiplicity of the eigenvalue $2(k-i)$. The eigenvector v corresponding to $2k-4$ may be obtained from*

$$V(x) = \sum_{(i,j)} \{V_{2i-1,2j-1}(x) + V_{2i,2j}(x)\},$$

by letting

$$v_i = V(x) \quad \textit{where } x \textit{ is any } x \in \mathscr{C}_i.$$

PROOF. From Proposition 3 and Fact 1, the eigenvalues of B belong to the set $\{2(k-i)_{\binom{2k}{i}}: i = 0,\ldots,2k\}$.

$\lambda = 2k$: the rows of B sum to $2k$, hence $2k$ is an eigenvalue of B with corresponding eigenvector $\mathbf{1} = (1,1,\ldots,1)^T$.

$\lambda = 2k-2$: from Fact 4, eigenvectors of A corresponding to the eigenvalue $2k-2$ sum to 0 over the orbits of $S_k \times D_4$, hence no eigenvector of A can

be constant and nonzero over the orbits. Hence, from Fact 2, $2k-2$ is not an eigenvalue of B.

$\lambda = 2k-4$: we have shown with Fact 6 that V is an eigenvector of A, corresponding to the eigenvalue $2k-4$, which is constant over the orbits. Hence, by Fact 3, V yields an eigenvector of B. It remains to show that B has no other eigenvector, that is, V is the only eigenvector of A which is constant over the orbits. The orthogonal complement of V in the eigenspace of A for $\lambda = 2k-4$ is spanned by the following $2k^2-k$ vectors:

$$\begin{aligned}
W_{2i-1,2j} &= V_{2i-1,2j} - \frac{\langle V_{2i-1,2j}, V\rangle}{|V|^2}\, V \\
&= V_{2i-1,2j}, \qquad 1 \le i, j \le k, \\
W_{2i-1,2j-1} &= V_{2i-1,2j-1} - \frac{\langle V_{2i-1,2j-1}, V\rangle}{|V|^2}\, V \\
&= V_{2i-1,2j-1} - \frac{1}{k(k-1)} V, \qquad 1 \le i < j \le k, \\
W_{2i,2j} &= V_{2i,2j} - \frac{\langle V_{2i,2j}, V\rangle}{|V|^2}\, V \\
&= V_{2i,2j} - \frac{1}{k(k-1)} V, \qquad 1 \le i < j \le k.
\end{aligned}$$

By Fact 5, for any orbit $\mathscr{C}$,

$$\sum_{x\in\mathscr{C}} W_{2i-1,2j}(x) = 0.$$

Also, by Fact 7,

$$\begin{aligned}
&\sum_{x\in\mathscr{C}} W_{2i-1,2j-1}(x) \\
&\quad = \sum_{x\in\mathscr{C}} V_{2i-1,2j-1}(x) - \frac{1}{k(k-1)} \sum_{(i,j)} \sum_{x\in\mathscr{C}} \{V_{2i-1,2j-1}(x) + V_{2i,2j}(x)\} = 0,
\end{aligned}$$

and similarly

$$\sum_{x\in\mathscr{C}} W_{2i,2j}(x) = 0.$$

Hence, no eigenvector in the orthogonal complement of V in the eigenspace of A for $\lambda = 2k-4$ is constant over the orbits of $S_k \times D_4$. Consequently, by Fact 2, $2k-4$ is an eigenvalue of B with multiplicity 1. □

C.2. *Quotient graph* $\mathscr{X}/(S_k \times (C_2 \times C_2))$. Facts 4 and 5 also apply to the orbits of $S_k \times (C_2 \times C_2)$. Facts 6 and 7 may be modified as follows.

FACT 8. Let

$$V_o(x) = \sum_{(i,\,j)} V_{2i-1,\,2j-1}(x)$$

and

$$V_e(x) = \sum_{(i,\,j)} V_{2i,\,2j}(x),$$

where the sums are over all $\binom{k}{2}$ unordered pairs (i, j) of distinct integers ranging from 1 to k. Then, V_e and V_o are two eigenvectors of A corresponding to the eigenvalue $2k-4$. Furthermore, V_e and V_o are constant over the orbits of $S_k \times (C_2 \times C_2)$.

FACT 9. The $k(k-1)/2$ 2-factor eigenvectors $\{V_{2i-1,2j-1}\colon 1 \le i < j \le k\}$ have the same sums over the orbits of $S_k \times (C_2 \times C_2)$, that is, for any orbit $\mathscr{O}$ and $1 \le i_1 < j_1 \le k$, $1 \le i_2 < j_2 \le k$,

$$\sum_{x \in \mathscr{O}} V_{2i_1-1,\,2j_1-1}(x) = \sum_{x \in \mathscr{O}} V_{2i_2-1,\,2j_2-1}(x).$$

Similarly for the $k(k-1)/2$ 2-factor eigenvectors $\{V_{2i,\,2j}\colon 1 \le i < j \le k\}$.

PROPOSITION 5 [Eigenvalues of adjacency matrix B of quotient graph $\mathscr{X}/(S_k \times (C_2 \times C_2))$]. *The largest eigenvalue of B is $2k$, with multiplicity one, and the second largest eigenvalue is $2k-4$, with multiplicity two. All other eigenvalues of B are strictly less than $2k-4$ and belong to the set $\{2(k-i)_{\binom{2k}{i}}\colon i = 3, \ldots, 2k\}$. The eigenvectors corresponding to $2k-4$ may be obtained from V_e and V_o.*

PROOF. From Proposition 3 and Fact 1, the eigenvalues of B belong to the set $\{2(k-i)_{\binom{2k}{i}}\colon i = 0, \ldots, 2k\}$.

$\lambda = 2k$: the rows of B sum to $2k$, hence $2k$ is an eigenvalue of B with corresponding eigenvector $\mathbf{1} = (1, 1, \ldots, 1)^T$.

$\lambda = 2k-2$: from Fact 4, eigenvectors of A corresponding to the eigenvalue $2k-2$ sum to 0 over the orbits of $S_k \times (C_2 \times C_2)$, hence no eigenvector of A can be constant and nonzero over the orbits. Hence, from Fact 2, $2k-2$ is not an eigenvalue of B.

$\lambda = 2k-4$: from Fact 8, V_o and V_e are eigenvectors of A, corresponding to the eigenvalue $2k-4$, which are constant over the orbits. Hence, by Fact 3, V_e and V_o yield two eigenvectors of B. It remains to show that B has only two eigenvectors, that is, V_e and V_o are the only eigenvectors of A which are constant over the orbits. The orthogonal complement of $\mathrm{Span}\{V_o, V_e\}$ in the eigenspace of A for $\lambda = 2k-4$ is spanned by the following $2k^2 - k$ vectors:

$$\begin{aligned} W_{2i-1,\,2j} &= V_{2i-1,\,2j} - \frac{\langle V_{2i-1,\,2j},\ V_e\rangle}{|V_e|^2}\ V_e - \frac{\langle V_{2i-1,\,2j},\ V_o\rangle}{|V_o|^2}\ V_o \\ &= V_{2i-1,\,2j}, \qquad 1 \le i,\ j \le k, \end{aligned}$$

$$W_{2i-1,2j-1} = V_{2i-1,2j-1} - \frac{\langle V_{2i-1,2j-1},\, V_e\rangle}{|V_e|^2}\, V_e - \frac{\langle V_{2i-1,2j-1},\, V_o\rangle}{|V_o|^2}\, V_o$$
$$= V_{2i-1,2j-1} - \frac{2}{k(k-1)} V_o, \qquad 1 \le i < j \le k,$$
$$W_{2i,2j} = V_{2i,2j} - \frac{\langle V_{2i,2j},\, V_e\rangle}{|V_e|^2}\, V_e - \frac{\langle V_{2i,2j},\, V_o\rangle}{|V_o|^2}\, V_o$$
$$= V_{2i,2j} - \frac{2}{k(k-1)} V_e, \qquad 1 \le i < j \le k.$$

By Fact 5, for any orbit $\mathscr{C}$,

$$\sum_{x\in\mathscr{C}} W_{2i-1,2j}(x) = 0.$$

Also, by Fact 9,

$$\sum_{x\in\mathscr{C}} W_{2i-1,2j-1}(x) = \sum_{x\in\mathscr{C}} V_{2i-1,2j-1}(x) - \frac{2}{k(k-1)} \sum_{(i,j)} \sum_{x\in\mathscr{C}} V_{2i-1,2j-1}(x) = 0,$$

and similarly,

$$\sum_{x\in\mathscr{C}} W_{2i,2j}(x) = 0.$$

Hence, no eigenvector in the orthogonal complement of $\mathrm{Span}\{V_o, V_e\}$ in the eigenspace of A for $\lambda = 2k-4$ is constant over the orbits of $S_k \times (C_2 \times C_2)$. Consequently, by Fact 2, $2k-4$ is an eigenvalue of B with multiplicity 2. □

C.3. *Quotient graph* $\mathscr{X}/((S_h \times S_{k-h}) \times D_4)$. Facts 4 and 5 also apply to the orbits of $\mathscr{X}/((S_h \times S_{k-h}) \times D_4)$. The proof for sibships with both affected and unaffected sibs is similar to that for affected-only sibships, but involves new combinations of eigenvectors. Without loss of generality, order the sibs such that the first h are affected and the last $k-h$ unaffected. For $k \ge 3$, define

$$V^a(x) = \sum_{1\le i<j\le h} \big\{V_{2i-1,2j-1}(x) + V_{2i,2j}(x)\big\}, \qquad h \ge 2,$$
$$V^u(x) = \sum_{h+1\le i<j\le k} \big\{V_{2i-1,2j-1}(x) + V_{2i,2j}(x)\big\}, \qquad h \le k-2,$$
$$V^{au}(x) = \sum_{1\le i\le h,\, h+1\le j\le k} \big\{V_{2i-1,2j-1}(x) + V_{2i,2j}(x)\big\}.$$

Facts 6 and 7 are then modified as follows.

FACT 10. For $k \ge 3$, V^a $(h \ge 2)$, V^u $(h \le k-2)$ and V^{au} are eigenvectors of A corresponding to the eigenvalue $2k-4$. Furthermore, these are constant over the orbits of $\mathscr{X}/((S_h \times S_{k-h}) \times D_4)$.

FACT 11. For any orbit $\mathscr{C}$ of $\mathscr{X}/((S_h \times S_{k-h}) \times D_4)$ and $1 \le i_1 < j_1 \le h$, $1 \le i_2 < j_2 \le h$,

$$\sum_{x \in \mathscr{C}} V_{2i_2-1,\,2j_2-1}(x) = \sum_{x \in \mathscr{C}} V_{2i_1-1,\,2j_1-1}(x) = \sum_{x \in \mathscr{C}} V_{2i_1,\,2j_1}(x) = \sum_{x \in \mathscr{C}} V_{2i_2,\,2j_2}(x).$$

Similarly for $h+1 \le i_1 < j_1 \le k$, $h+1 \le i_2 < j_2 \le k$ and $1 \le i_1, i_2 \le h$, $h+1 \le j_1,\ j_2 \le k$.

PROPOSITION 6 [Eigenvalues of adjacency matrix B of quotient graph $\mathscr{X}/((S_h \times S_{k-h}) \times D_4)$]. *The largest eigenvalue of B is $2k$, with multiplicity one, and the second largest eigenvalue is $2k-4$, with multiplicity three if $2 \le h \le k-2$ and two otherwise. All other eigenvalues of B are strictly less than $2k-4$ and belong to the set $\{2(k-i)_{\binom{2k}{i}}\colon i = 3, \ldots, 2k\}$. The eigenvectors corresponding to $2k-4$ may be obtained from V^a, V^u and V^{au}.*

PROOF. From Proposition 3 and Fact 1, the eigenvalues of B belong to the set $\{2(k-i)_{\binom{2k}{i}}\colon i = 0, \ldots, 2k\}$. We give the proof for $\lambda = 2k-4$; for the other eigenvalues, the proof is as for Propositions 4 and 5. From Fact 10, V^a $(h \ge 2)$, V^u $(h \le k-2)$ and V^{au} are eigenvectors of A, corresponding to the eigenvalue $2k-4$, which are constant over the orbits. Hence, by Fact 3, they yield eigenvectors of B. It remains to show that these are the only eigenvectors of B, that is, V^a, V^u and V^{au} are the only eigenvectors of A which are constant over the orbits. The orthogonal complement of $\mathrm{Span}\{V^a, V^u, V^{au}\}$ in the eigenspace of A for $\lambda = 2k-4$ is spanned by the $2k^2 - k$ vectors $W_{i,j}$, $1 \le i,\ j \le 2k$, $i \ne j$, defined as follows

$$W_{i,j} = V_{i,j} - \frac{\langle V_{i,j},\, V^a\rangle}{|V^a|^2}\, V^a - \frac{\langle V_{i,j},\, V^u\rangle}{|V^u|^2}\, V^u - \frac{\langle V_{i,j}, V^{au}\rangle}{|V^{au}|^2}\, V^{au}.$$

The $W_{i,j}$'s simplify to

$$\begin{aligned}
W_{2i-1,\,2j} &= V_{2i-1,\,2j}, \qquad i, j = 1, \ldots, k,\\
W_{2i-1,\,2j-1} &= V_{2i-1,\,2j-1} - \frac{1}{h(h-1)} V^a, \qquad 1 \le i < j \le h,\\
&= V_{2i-1,\,2j-1} - \frac{1}{(k-h)(k-h-1)} V^u, \qquad h+1 \le i < j \le k,\\
&= V_{2i-1,2j-1} - \frac{1}{2h(k-h)} V^{au}, \qquad 1 \le i \le h,\ h+1 \le j \le k,\\
W_{2i,\,2j} &= V_{2i,\,2j} - \frac{1}{h(h-1)} V^a, \qquad 1 \le i < j \le h,\\
&= V_{2i,2j} - \frac{1}{(k-h)(k-h-1)} V^u, \qquad h+1 \le i < j \le k,\\
&= V_{2i,2j} - \frac{1}{2h(k-h)} V^{au}, \qquad 1 \le i \le h,\ h+1 \le j \le k.
\end{aligned}$$

By Fact 5, for any orbit $\mathscr{O}$,

$$\sum_{x\in\mathscr{O}} W_{2i-1,\,2j}(x) = 0.$$

Also, by Fact 11,

$$\sum_{x\in\mathscr{O}} W_{2i-1,\,2j-1}(x) = 0$$

and

$$\sum_{x\in\mathscr{O}} W_{2i,\,2j}(x) = 0.$$

Hence, no eigenvector in the orthogonal complement of $\mathrm{Span}\{V^a, V^u, V^{au}\}$ in the eigenspace of A for $\lambda = 2k-4$ is constant over the orbits of $\mathscr{X}/((S_h \times S_{k-h}) \times D_4)$. Consequently, by Fact 2, $2k-4$ is an eigenvalue of B with multiplicity three if $2 \le h \le k-2$ and two otherwise. □

C.4. *Quotient graph* $\mathscr{X}/((S_h \times S_{k-h}) \times (C_2 \times C_2))$. For $(S_h \times S_{k-h}) \times (C_2 \times C_2)$ we again separate "even" and "odd" eigenvectors and consider six new combinations of eigenvectors,

$$V_e^a(x) = \sum_{1\le i<j\le h} V_{2i,\,2j}(x), \qquad h \ge 2,$$

$$V_o^a(x) = \sum_{1\le i<j\le h} V_{2i-1,\,2j-1}(x), \qquad h \ge 2,$$

$$V_e^u(x) = \sum_{h+1\le i<j\le k} V_{2i,\,2j}(x), \qquad h \le k-2,$$

$$V_o^u(x) = \sum_{h+1\le i<j\le k} V_{2i-1,\,2j-1}(x), \qquad h \le k-2,$$

$$V_e^{au}(x) = \sum_{1\le i\le h,\ h+1\le j\le k} V_{2i,\,2j}(x),$$

$$V_o^{au}(x) = \sum_{1\le i\le h,\ h+1\le j\le k} V_{2i-1,\,2j-1}(x).$$

Facts 6 and 7 may then be suitably modified.

D. Score statistic: Proof of Theorem 2. From Theorem 1, -4 is an eigenvalue of the infinitesimal generator Q with multiplicity 1. Hence, the second derivative of the transition matrix at $\theta = \frac{1}{2}$ has rank 1 and entries

$$u_{ij} = 8\alpha_j v_i v_j,$$

where $v = (v_1, \ldots, v_m)^T$ is the right eigenvector of Q with unit norm with respect to the inner product $\langle \cdot, \cdot \rangle_\alpha$. The score statistic is given by

$$S = \sum_{i=1}^{m} N_i \frac{\sum_{j=1}^{m} \pi_j 8\alpha_i v_i v_j}{\alpha_i}$$
$$= 8\Big(\sum_{j=1}^{m} v_j \pi_j \Big)\Big(\sum_{i=1}^{m} v_i N_i \Big).$$

It is convenient to express the score statistic in terms of the first column of U, $8\alpha_1 v_1 v$. Without loss of generality, we let the first IBD configuration be the one for which all sibs inherited the same maternal and paternal DNA, that is, with representative inheritance vector $(1, 3, 1, 3, \ldots, 1, 3)$ and label $(0, 0, 0)$ in the notation of Ethier and Hodge (1985):

$$S = \frac{8}{8(8\alpha_1 v_1^2)\alpha_1}\Big(\sum_{j=1}^{m} u_{j1} \pi_j \Big)\Big(\sum_{i=1}^{m} u_{i1} N_i \Big)$$
$$= \frac{2^{2k}}{u_{11}|\mathscr{C}_1|}\Big(\sum_{j=1}^{m} u_{j1} \pi_j \Big)\Big(\sum_{i=1}^{m} u_{i1} N_i \Big).$$

By differentiating (3.2) we find that

$$u_{ij} = 2^{4-2k} \sum_{y \in \mathscr{C}_j} ((\Delta(x, y) - k)^2 - k/2), \text{ where } x \text{ is any inheritance vector in } \mathscr{C}_i$$

and in particular, $u_{11} = 2^{5-2k} k(k-1)$. The contribution of an affected sib-k-tuple with inheritance vector $x \in \mathscr{C}_i$ to the score statistic is based on

$$\begin{aligned}
u_{i1} &= 2^{4-2k} \sum_{y \in \mathscr{C}_1} ((\Delta(x, y) - k)^2 - k/2) \\
&= 2^{4-2k}((a_2(x) + a_4(x) - k)^2 + (a_2(x) + a_3(x) - k)^2 \\
&\quad + (a_1(x) + a_4(x) - k)^2 + (a_1(x) + a_3(x) - k)^2 - 2k) \\
&= 2^{4-2k}(2(a_1(x)^2 + a_2(x)^2 + a_3(x)^2 + a_4(x)^2) \\
&\quad + 2(a_2(x)a_4(x) + a_2(x)a_3(x) + a_1(x)a_4(x) + a_1(x)a_3(x)) \\
&\quad - 2k(2a_1(x) + 2a_2(x) + 2a_3(x) + 2a_4(x)) + 4k^2 - 2k) \\
&= 2^{5-2k}(a_1(x)^2 + a_2(x)^2 + a_3(x)^2 + a_4(x)^2 \\
&\quad + (a_1(x) + a_2(x))(a_3(x) + a_4(x)) - 4k^2 + 2k^2 - k) \\
&= 2^{5-2k}(a_1(x)^2 + a_2(x)^2 + a_3(x)^2 + a_4(x)^2 + k^2 - 2k^2 - k) \\
&= 2^{5-2k}(a_1(x)^2 + a_2(x)^2 + a_3(x)^2 + a_4(x)^2 - k(k+1)).
\end{aligned}$$

Hence, from (4.2),

$$u_{i1} = 2^{5-2k}k(k-1)(2S_{\text{pairs}} - 1).$$

Now, $u_{i1} = 8\alpha_1 v_1 v_i$, $u_{11} = 2^{5-2k}k(k-1)$ and $\alpha_1 = 2^{2-2k}$. Thus

$$v_1 = \sqrt{k(k-1)} \quad \text{and} \quad v_i = \sqrt{k(k-1)}(2S_{\text{pairs}} - 1).$$

Hence

$$S = 2^{2k-2}\Big(\sum_j u_{j1}\pi_j\Big)(2S_{\text{pairs}} - n) = 8\sqrt{k(k-1)}\Big(\sum_j v_j\pi_j\Big)(2S_{\text{pairs}} - n),$$

where S_{pairs} is summed over all sibships with k affected sibs. □

Acknowledgments. We are grateful to Steve Evans for many helpful discussions on Markov chains and to Cheryl Praeger, Alice Niemeyer and Nick Wormald for graph-related ideas. We also thank the referees for their valuable comments and suggestions on a previous version of this paper.

REFERENCES

BLACKWELDER, W. C. and ELSTON, R. C. (1985). A comparison of sib-pair linkage tests for disease susceptibility loci. *Genet. Epidemiol.* **2** 85–97.

DAVIS, S. and WEEKS, D. E. (1997). Comparison of nonparametric statistics for detection of linkage in nuclear families: single-marker evaluation. *Amer. J. Hum. Genet.* **61** 1431–1444.

DAY, N. E. and SIMONS, M. J. (1976). Disease-susceptibility genes—their identification by multiple case family studies. *Tissue Antigens* **8** 109–119.

DEBRUIJN, N. G. (1964). Pólya's theory of counting. In *Applied Combinatorial Mathematics* (E. F. Beckenbach, ed.) 144–184. Wiley, New York.

DIACONIS, P. (1988). *Group Representations in Probability and Statistics*. IMS, Hayward, CA.

DONNELLY, K. P. (1983). The probability that related individuals share some section of genome identical by descent. *Theoret. Population Biol.* **23** 34–63.

DUDOIT, S. (1999). Linkage analysis of complex human traits using identity by descent data. Ph.D. dissertation, Univ. California, Berkeley.

DUDOIT, S. and SPEED, T. P. (1999). Triangle constraints for sib-pair identity by descent probabilities under a general multilocus model for disease susceptibility. In *Statistics in Genetics* (M. E. Halloran and S. Geisser, eds.). Springer, New York.

ETHIER, S. N. and HODGE, S. E. (1985). Identity-by-descent analysis of sibship configurations. *Amer. J. Med. Genet.* **22** 263–272.

FARAWAY, J. J. (1993). Improved sib-pair linkage test for disease susceptibility loci. *Genet. Epidemiol.* **10** 225–233.

FEINGOLD, E., BROWN, P. O. and SIEGMUND, D. (1993). Gaussian models for genetic linkage analysis using complete high-resolution maps of identity by descent. *Amer. J. Hum. Genet.* **53** 234–251.

FEINGOLD, E. and SIEGMUND, D. (1997). Strategies for mapping heterogeneous recessive traits by allele sharing methods. *Amer. J. Hum. Genet.* **60** 965–978.

FRALEIGH, J. B. (1989). *A First Course in Abstract Algebra*, 4th ed. Addison-Wesley, Reading, MA.

GODSIL, C. D. (1993). *Algebraic Combinatorics*. Chapman and Hall, New York.

HASEMAN, J. K. and ELSTON, R. C. (1972). The investigation of linkage between a quantitative trait and a marker locus. *Behavior Genetics* **2** 3–19.

HOLMANS, P. (1993). Asymptotic properties of affected sib-pair linkage analysis. *Amer. J. Hum. Genet.* **52** 362–374.

HOLMANS, P. and CLAYTON, D. (1995). Efficiency of typing unaffected relatives in an affected-sib-pair linkage study with single-locus and multiple tightly linked markers. *Amer. J. Hum. Genet.* **57** 1221–1232.

JACOB, B. (1990). *Linear Algebra*. Freeman, New York.

KNAPP, M., SEUCHTER, S. A. and BAUR, M. P. (1994a). Linkage analysis in nuclear families I: optimality criteria for affected sib-pair tests. *Hum. Hered.* **44** 37–43.

KNAPP, M., SEUCHTER, S. A. and BAUR, M. P. (1994b). Linkage analysis in nuclear families II: relationship between affected sib-pair tests and lod score analysis. *Hum. Hered.* **44** 44–51.

KNAPP, M., SEUCHTER, S. A. and BAUR, M. P. (1994c). Two-locus disease models with two marker loci: the power of affected-sib-pair tests. *Amer. J. Hum. Genet.* **55** 1030–1041.

KONG, A. and COX, N. J. (1997). Allele-sharing models: lod scores and accurate linkage tests. *Amer. J. Hum. Genet.* **61** 1179–1188.

KRUGLYAK, L., DALY, M. J., REEVE-DALY, M. P. and LANDER, E. S. (1996). Parametric and nonparametric linkage analysis: a unified multipoint approach. *Amer. J. Hum. Genet.* **58** 1347–1363.

KRUGLYAK, L. and LANDER, E. S. (1995). Complete multipoint sib-pair analysis of qualitative and quantitative traits. *Amer. J. Hum. Genet.* **57** 439–454.

LANDER, E. S. and GREEN, P. (1987). Construction of multilocus genetic maps in humans. *Proc. Nat. Acad. Sci. U.S.A.* **84** 2363–2367.

MCPEEK, M. S. (1996). An introduction to recombination and linkage analysis. In *Genetic Mapping and DNA Sequencing* (T. P. Speed and M. S. Waterman, eds.) 1–14. Springer, New York.

MCPEEK, M. S. (1999). Optimal allele-sharing statistics for genetic mapping using affected relatives. *Genet. Epidemiol.* **16** 225–249.

OTT, J. (1991). *Analysis of Human Genetic Linkage*, rev. eds. Johns Hopkins Univ. Press, Baltimore.

RISCH, N. (1990a). Linkage strategies for genetically complex traits II. The power of affected relative pairs. *Amer. J. Hum. Genet.* **46** 229–241.

RISCH, N. (1990b). Linkage strategies for genetically complex traits III. The effect of marker polymorphism on analysis of affected relative pairs. *Amer. J. Hum. Genet.* **46** 242–253.

RISCH, N. and ZHANG, H. (1995). Extreme discordant sib pairs for mapping quantitative trait loci in humans. *Science* **268** 1584–1589.

RISCH, N. and ZHANG, H. (1996). Mapping quantitative trait loci with extreme discordant sib pairs: sampling considerations. *Amer. J. Hum. Genet.* **58** 836–843.

ROSENBLATT, M. (1974). *Random Processes. Graduate Texts in Mathematics* **17**, 2nd ed. Springer, New York.

SCHAID, D. J. and NICK, T. G. (1990). Sib-pair linkage tests for disease suscpetibility loci: common tests vs. the asymptotically most powerful test. *Genet. Epidemiol.* **7** 359–370.

SPEED, T. P. (1996). What is a genetic map function? In *Genetic Mapping and DNA Sequencing* (T. P. Speed and M. S. Waterman, eds.) 65–88. Springer, New York.

SUAREZ, B. K. and VAN EERDEWEGH, P. (1984). A comparison of three affected-sib-pair scoring methods to detect HLA-linked disease susceptibility genes. *Amer. J. Med. Genet.* **18** 135–146.

TENG, J. and SIEGMUND, D. O. (1997). Combining information within and between pedigrees for mapping complex traits. *Amer. J. Hum. Genet.* **60** 979–992.

THOMPSON, E. A. (1997). Conditional gene identity in affected individuals. In *Genetic Mapping of Disease Genes* (I.-H. Pawlowitzki, J. H. Edwards and E. A. Thompson, eds.) 137–146. Academic Press, San Diego.

VAN LINT, J. H. and WILSON, R. M. (1992). *A Course in Combinatorics*. Cambridge Univ. Press.
WHITTEMORE, A. S. (1996). Genome scanning for linkage: an overview. *Amer. J. Hum. Genet.* **59** 704–716.
WHITTEMORE, A. S. and HALPERN, J. (1994a). A class of tests for linkage using affected pedigree members. *Biometrics* **50** 118–127.
WHITTEMORE, A. S. and HALPERN, J. (1994b). Probability of gene identity by descent: computation and applications. *Biometrics* **50** 109–117.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA, BERKELEY
BERKELEY, CALIFORNIA 94720-3860
E-MAIL: sandrine@stat.berkeley.edu
terry@stat.berkeley.edu