

University of California, Berkeley
U.C. Berkeley Division of Biostatistics Working Paper Series

Year 2003

Paper 128

IBD Configuration Transition Matrices and
Linkage Score Tests for Unilineal Relative
Pairs

Sandrine Dudoit*

*Division of Biostatistics, School of Public Health, University of California, Berkeley, sandrine@stat.berkeley.edu

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.

<http://biostats.bepress.com/ucbbiostat/paper128>

Copyright ©2003 by the author.

IBD Configuration Transition Matrices and Linkage Score Tests for Unilineal Relative Pairs

Sandrine Dudoit

Abstract

Properties of transition matrices between IBD configurations are derived for four general classes of unilineal relative pairs obtained from the grand-parent/ grand-child, half-sib, avuncular, and cousin relationships. In this setting, IBD configurations are defined as orbits of groups acting on a set of inheritance vectors. Properties of the transition matrix between IBD configurations at two linked loci are derived by relating its infinitesimal generator to the adjacency matrix of a quotient graph. The second largest eigenvalue of the infinitesimal generator and its multiplicity are key in determining the form of the transition matrix and of likelihood-based linkage tests such as score tests.

1 Introduction

Genetic mapping is concerned with identifying genes that predispose to given phenotypes in humans and model organisms. The general approach to genetic mapping involves identifying loci at which genotypes are associated with phenotypes. Genotype-phenotype associations are viewed as suggesting linkage, or “closeness”, of a genetic locus to a gene influencing the phenotype. Since the seminal work of Haseman and Elston (1972) on the linkage analysis of quantitative traits, identity by descent (IBD) genotypes have been widely used in genetic mapping studies. Reviews of IBD-based linkage analysis methods are given in Elston and Cordell (2001) and Shih and Whittemore (2001). Likelihood-based methods for studying associations between phenotypes and IBD configurations at marker loci typically involve computing transition probabilities between IBD configurations at two linked loci. Properties of IBD configuration transition matrices are therefore important in terms of understanding statistical properties of linkage test statistics and, more generally, for studying patterns of transmission of DNA in families.

Focusing on sibships, Dudoit (1999), Dudoit and Speed (1999), and Dudoit and Speed (2000) proposed a unified likelihood-based approach for the genetic mapping of complex human traits, qualitative and quantitative, using IBD data from small pedigrees. The unified approach considers the likelihood of IBD data conditional on phenotypes, and tests for linkage between a marker locus and a gene influencing the trait using a score test. For a given pedigree type, the form of the linkage score statistic is determined by the second largest eigenvalue and corresponding eigenvector(s) of the transition matrix for IBD configurations. The simulation studies of Dudoit and Speed (2000), Goldstein et al. (2000), and Goldstein et al. (2001) demonstrated that the linkage score test for quantitative traits had good power and robustness properties compared to alternative genetic mapping methods based on IBD data from unilineal relative pairs and sib-pairs.

The present article derives theoretical properties of transition matrices between IBD configurations at two linked loci for four general classes of unilineal relative pairs, obtained from the grand-parent/grand-child, half-sib, avuncular (uncle/nephew), and cousin relationships. In this setting, IBD configurations are defined as orbits of groups acting on a set of inheritance vectors (Donnelly, 1983). Section 2 describes general properties of transition matrices for IBD configurations. We show that the transition matrix satisfies a semi-group property (Proposition 1, p. 5) and derive a spectral

representation of the matrix in terms of the eigenvalues and eigenvectors of its infinitesimal generator (Proposition 2, p. 7). Properties of the eigenvalues of the infinitesimal generator are obtained by relating it to the adjacency matrix of a quotient graph (Proposition 4, p. 9). The second largest eigenvalue of the infinitesimal generator and its multiplicity are key in determining the form of the transition matrix. Sections 3 – 6 describe specific properties of the transition matrices for the four general classes of unilineal relative pairs. Based on these properties, Section 7 derives linkage score statistics, for qualitative and quantitative traits, for the different types of relative pairs. Table 1 summarizes the main properties of the transition matrices and score statistics for the four classes of unilineal relative pairs.

2 General properties of transition matrices for IBD configurations

2.1 Inheritance vectors and IBD configurations

DNA at the same locus on two homologous chromosomes is said to be *identical by descent* (IBD) if it originated from the same ancestral chromosome. Identity by descent (IBD) patterns within a pedigree may be summarized at any locus by *inheritance vectors* which indicate the outcome of meioses giving rise to the non-founders. For a pedigree with n non-founders, Kruglyak et al. (1996) and Lander and Green (1987) define the inheritance vector at a particular locus to be a binary $2n$ -vector whose coordinates describe the outcome of the $2n$ paternal and maternal meioses giving rise to the n non-founders. The $(2i - 1)$ st coordinate is 0 or 1 according to whether grand-paternal or grand-maternal DNA was transmitted in the paternal meiosis giving rise to the i th non-founder, $i = 1, \dots, n$. The $(2i)$ th coordinate contains the same information for the maternal meiosis. Thus, for a pedigree with f founders, the inheritance vector completely specifies which of the $2f$ founder DNA variants are inherited by each non-founder at the locus of interest.

In many cases, only a subset of meioses are relevant. For instance, when considering IBD for unilineal relative pairs, the only relevant meioses are those intervening between the two relatives and their common ancestor(s) (see Donnelly (1983) and Sections 3 – 6 below for specific definitions).

Note that inheritance vectors as defined above summarize IBD *within* a pedigree. In other words, inheritance vectors follow the transmission of the

founders' DNA within the pedigree, ignoring any IBD already present among these founders. Thus, in general, we allow the possibility of inbreeding in the population of interest and the founders could in principle be related.

For d meioses, let \mathcal{X} denote the set of all 2^d inheritance vectors. This set is highly redundant, as subsets of inheritance vectors correspond to the same amount of IBD sharing among relatives. Inheritance vectors may be partitioned into a smaller number of *IBD configurations*. For instance, for sib-pairs, it is common practice to collapse the $2^4 = 16$ inheritance vectors into three IBD configurations corresponding to the number $j = 0, 1, 2$ of chromosomes sharing DNA IBD at a given locus.

In this article, unless otherwise indicated, we consider a general definition of IBD configurations as *orbits of groups* acting on the set of inheritance vectors \mathcal{X} . The group definitions for IBD configurations depend on the pedigree type. Dudoit and Speed (1999) consider IBD configurations for sibships of arbitrary size. In particular, it is shown that the usual 0, 1, 2 IBD configurations for sib-pairs correspond to the orbits of $S_2 \times D_4$, the direct product of the symmetric group S_2 on 2 letters and the dihedral group D_4 of the square. In Sections 3 – 6, we consider the group definitions of Donnelly (1983) for the four general classes of unilineal relative pairs. Working with IBD configurations rather than inheritance vectors serves at least two purposes: (i) practical convenience achieved from data reduction and (ii) mathematical symmetry for the transition matrices between IBD configurations, as shown below.

2.2 Transition matrices for inheritance vectors

Because of crossovers, the IBD configuration of a given pedigree varies along a chromosome. In this section, we summarize general properties of transition matrices between IBD configurations at two linked loci. Proofs of the general results in Propositions 1 – 4 are given in Dudoit (1999) and Dudoit and Speed (1999). Consider two loci \mathcal{L}_1 and \mathcal{L}_2 with recombination fraction $\theta \in [0, 1/2]$, and denote inheritance vectors (for d meioses) 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 \mathcal{L}_1 and \mathcal{L}_2 in the corresponding meiosis. Let $\Delta(x, y)$ denote the number of coordinates at which the inheritance vectors x and y differ (Hamming distance), i.e., the number of recombination events between the two loci. Recombination events are assumed to be independent across meioses and to occur with probability

θ in each meiosis. Thus, the probability $r_{xy}(\theta)$ of a transition from x to y is the probability of $\Delta(x, y)$ recombination events in the d meioses under consideration

$$r_{xy}(\theta) = \theta^{\Delta(x,y)}(1 - \theta)^{d-\Delta(x,y)}.$$

The *transition matrix* $R(\theta) = (r_{xy}(\theta))$ between inheritance vectors at loci with recombination fraction θ may be expressed as the Kronecker power of 2×2 transition matrices corresponding to transitions in each of the d coordinates

$$R(\theta) = \left[\begin{array}{cc} 1 - \theta & \theta \\ \theta & 1 - \theta \end{array} \right]^{\otimes d}. \quad (1)$$

In the extreme case of fully linked loci ($\theta = 0$), the transition matrix simplifies to $R(0) = I_{2^d}$, the $2^d \times 2^d$ identity matrix. In the other extreme of unlinked loci ($\theta = 1/2$), all the entries of $R(1/2)$ are equal to 2^{-d} . Thus, the inheritance vectors at two unlinked loci are independent. This corresponds to *Mendel's Second Law*.

Note that we are assuming equal male and female recombination fractions; otherwise, we would have a 2×2 transition matrix for paternal meioses and a 2×2 transition matrix for maternal meioses. If θ_m and θ_f denote respectively the male and female recombination fractions between \mathcal{L}_1 and \mathcal{L}_2 , then the transition matrix is

$$R(\theta_m, \theta_f) = \left[\begin{array}{cc} 1 - \theta_m & \theta_m \\ \theta_m & 1 - \theta_m \end{array} \right]^{\otimes d_m} \otimes \left[\begin{array}{cc} 1 - \theta_f & \theta_f \\ \theta_f & 1 - \theta_f \end{array} \right]^{\otimes d_f}, \quad (2)$$

where d_m and d_f denote, respectively, the number of paternal and maternal meioses among the d meioses.

2.3 Transition matrices for IBD configurations

The transition matrix $R(\theta)$ for inheritance vectors has the same form for any pedigree structure, however, the transition matrix for IBD configurations depends on the type of pedigree under consideration and the type of group action defining these IBD configurations. Suppose that for a given pedigree structure, IBD configurations are defined as orbits of a group G acting on the set of inheritance vectors \mathcal{X} . Denote the m IBD configurations by \mathcal{C}_j , $j = 1, \dots, m$. The transition matrix $T(\theta) = (t_{ij}(\theta))$ between IBD configurations

at loci with recombination fraction θ is the $m \times m$ matrix with entries

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

Note that these entries are polynomials in θ of degree d , the number of meioses.

Dudoit (1999) and Dudoit and Speed (1999) showed that these conditional probabilities simplify to

$$\begin{aligned} t_{ij}(\theta) &= \sum_{y \in \mathcal{C}_j} \theta^{\Delta(x,y)} (1 - \theta)^{d - \Delta(x,y)}, \quad \text{where } x \text{ is any } x \in \mathcal{C}_i \quad (3) \\ &= \frac{|\mathcal{C}_j|}{|\mathcal{C}_i|} \sum_{x \in \mathcal{C}_i} \theta^{\Delta(x,y)} (1 - \theta)^{d - \Delta(x,y)}, \quad \text{where } y \text{ is any } y \in \mathcal{C}_j. \end{aligned}$$

In the extreme case of fully linked loci ($\theta = 0$), the transition matrix simplifies to $T(0) = I_m$, the $m \times m$ identity matrix. In the other extreme of unlinked loci ($\theta = 1/2$), $t_{ij}(1/2) = |\mathcal{C}_j|/2^d$.

The next two propositions relate the transition matrix $T(\theta)$ to the adjacency matrix of a quotient graph whose eigenvalues and their multiplicities are key in determining properties of $T(\theta)$ and its derivatives.

Proposition 1 Semi-group property for $T(\theta)$.

Let $T(\theta)$ denote the transition matrix between IBD configurations $\{\mathcal{C}_j : j = 1, \dots, m\}$ defined as orbits of a group G acting on the set \mathcal{X} of inheritance vectors. Define a binary operation $*$ as $\theta_1 * \theta_2 = \theta_1(1 - \theta_2) + \theta_2(1 - \theta_1)$. Then, $T(\theta)$ satisfies the semi-group property

$$T(\theta_1 * \theta_2) = T(\theta_1)T(\theta_2). \quad (4)$$

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

$$T(\theta) = e^{d(\theta)Q}, \quad (5)$$

where $d(\theta) = -\ln(1 - 2\theta)/2$ is the inverse of the Haldane map function and $Q = T'(0)$ is the infinitesimal generator. The infinitesimal generator Q has the form

$$Q = T'(0) = B - dI, \quad (6)$$

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

$$b_{ij} = \sum_{y \in \mathcal{C}_j} I(\Delta(x, y) = 1), \quad \text{for any } x \in \mathcal{C}_i, \quad (7)$$

$\Delta(x, y)$ denotes Hamming distance, i.e., the number of coordinates at which the inheritance vectors x and y differ, and $I(\cdot)$ denotes the indicator function. The transition matrix $T(\theta)$ has stationary distribution

$$\alpha = (\alpha_1, \dots, \alpha_m) = \frac{1}{2^d} (|\mathcal{C}_1|, \dots, |\mathcal{C}_m|) \quad (8)$$

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

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

Hence, for one meiosis, 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 $d(\theta) = -\ln(1 - 2\theta)/2$, the inverse of the Haldane map function. Jointly, the d crossover processes are independent and identically distributed (i.i.d.) and so embeddable in a continuous-time random walk on the vertices of the hypercube $\{0, 1\}^d$ (Donnelly, 1983). The random walk model for the crossover process is widely used and is referred to in the genetics literature as the *no interference model*. Under the no interference model, the crossover process on individual meiotic products is a Poisson process with intensity 1. The Haldane map function $M(d) = (1 - e^{-2d})/2$ relates recombination fractions to genetic map distances d (measured in Morgans, M) under the no interference model (Speed, 1996). Note that if we have three ordered loci, and θ_1 and θ_2 are the recombination fractions between the first and second loci and second and third loci, respectively, then $\theta_1 * \theta_2$ is the recombination fraction between the first and third loci under the assumption that recombination events in disjoint intervals are independent, i.e., no crossover interference. Also note that we do not need to assume no crossover interference to derive the semi-group property in Proposition 1. 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 map distance along a chromosome. From condition (15) p. 63 in Rosenblatt (1974), it follows that the IBD configurations also form a continuous-time Markov

chain.

In order to compute linkage score statistics as in Dudoit and Speed (1999), one needs derivatives of the transition matrix $T(\theta)$ at $\theta = 1/2$. These may be computed by direct differentiation of the polynomial entries of $T(\theta)$, but one gains more knowledge on the transition matrix and on the form of the score statistic by using the following spectral decomposition of $T(\theta)$.

Proposition 2 Spectral decomposition for $T(\theta)$.

Let $T(\theta)$ denote the transition matrix between IBD configurations $\{\mathcal{C}_j : j = 1, \dots, m\}$ defined as orbits of a group G acting on the set \mathcal{X} of inheritance vectors. Then $T(\theta)$ may be written as

$$T(\theta) = \sum_h e^{\lambda_h d(\theta)} P_h = \sum_h (1 - 2\theta)^{-\lambda_h/2} P_h, \tag{9}$$

where λ_h are the distinct 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 $\sum_h P_h = I$. The matrix 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 ij th entry of P_h is $\alpha_j v_{ih} v_{jh}$, where v_{ih} is the i th 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_α .

Thus, eigenvalues of Q and their multiplicities give us information regarding the derivatives of the transition matrix $T(\theta)$ and hence the form of the score statistic in θ . We relate the infinitesimal generator Q to the adjacency matrix of a quotient graph in order to derive properties of its eigenvalues and corresponding eigenvectors. Consider the graph \mathcal{X} with vertex set the set of all inheritance vectors of length d and adjacency matrix $A(\mathcal{X}) = A$, with (x, y) -entry

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

\mathcal{X} is the graph defined by the first associates in the *Hamming scheme* $H(d, 2)$ (van Lint and Wilson (1992), Chapter 30). Consider any group G acting on the set \mathcal{X} of inheritance vectors. The matrix B , defined in Proposition 1, is the *adjacency matrix of the quotient graph* \mathcal{X}/G , which is the multi-digraph with the orbits of G as its vertices and with b_{ij} arcs going from \mathcal{C}_i to \mathcal{C}_j . Recall that $Q = B - dI$, consequently, one may work with B to derive the

eigenvalues of Q ; the eigenvalues of Q are simply the eigenvalues of B minus the number of meioses d .

In order to derive the eigenvalues of B , we rely on the following general properties of the eigenvalues of the adjacency matrix A . To describe the eigenvectors of A it is convenient to code the inheritance vectors $x = (x_1, x_2, \dots, x_d)$ as in a 2^d factorial experiment, where $x_i = 1$ when factor i is absent and 0 when it is present. The eigenvectors of A have the following patterns.

Proposition 3 Eigenvalues and eigenvectors of adjacency matrix A .

The eigenvalues of A belong to the set $\{d - 2i \binom{d}{i} : i = 0, \dots, d\}$, where $\binom{d}{i}$ is the multiplicity of the eigenvalue $d - 2i$. The eigenvector corresponding to the eigenvalue $\lambda = d$ is the grand mean term, $V_0 = (1, 1, \dots, 1)^T$.

The eigenvectors corresponding to the eigenvalue $\lambda = d - 2$ are the d main effect terms, V_1, V_2, \dots, V_d , where

$$V_i(x) = I(x_i = 1) - I(x_i = 0). \tag{10}$$

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

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

In general, the eigenvectors corresponding to the eigenvalue $\lambda = d - 2i$, $i = 0, \dots, d$, are the $\binom{d}{i}$ i -factor interactions, V_{j_1, j_2, \dots, j_i} , $1 \leq j_1 < j_2 < \dots < j_i \leq d$, where

$$V_{j_1, j_2, \dots, j_i}(x) = V_{j_1}(x)V_{j_2}(x) \dots V_{j_i}(x). \tag{11}$$

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

In addition, we make use of the following general facts concerning adjacency matrices of quotient graphs (Godsil (1993), Chapter 5).

Lemma 1 (based on Lemma 2.2 in Godsil (1993))

The eigenvalues of B are a subset of the eigenvalues of A .

Lemma 2 (based on Lemma 2.2 in Godsil (1993))

Let C denote the characteristic matrix for the partition $\{\mathcal{C}_j : j = 1, \dots, m\}$; C is a $2^d \times m$ matrix, with ij th entry 1 or 0 according as the i th vertex of \mathcal{X} is contained in the orbit \mathcal{C}_j or not. If v is an eigenvector of B , then $V = Cv$ is an eigenvector of A which is constant over the orbits of G , with entry $V(x) = v_i$ for $x \in \mathcal{C}_i$.

Lemma 3 (Proof in Dudoit and Speed (1999))

If V is an eigenvector of A which is constant over the orbits of G , with $V(x) = v_i \forall x \in \mathcal{C}_i$, then the vector v , with i th entry v_i , is an eigenvector of B .

Proposition 4 Eigenvalues of the infinitesimal generator Q .

The eigenvalues of the infinitesimal generator Q , for IBD configurations defined as orbits of a group G acting on the set \mathcal{X} of inheritance vectors, belong to the set $\{-2i \binom{d}{i} : i = 0, \dots, d\}$, where $\binom{d}{i}$ is the largest possible multiplicity of the eigenvalue $-2i$.

Since the matrix $R(\theta)$ has the same form for any type of pedigree, Propositions 1 – 4 apply to arbitrary pedigrees and IBD configurations. The i th derivative, $i = 0, \dots, d$, of $T(\theta)$ is therefore given by

$$T^{(i)}(\theta) = \sum_h \left\{ \prod_{j=0}^{i-1} (\lambda_h + 2j) \right\} (1 - 2\theta)^{-(\lambda_h + 2i)/2} P_h. \quad (12)$$

Note that the projection matrix for the largest eigenvalue of Q , $\lambda_1 = 0$, is $P_1 = T(1/2)$, the matrix whose rows are equal to the stationary distribution α . The first non-zero derivative of $T(\theta)$ at $\theta = 1/2$ and its rank are determined by the *second largest eigenvalue* of Q , λ_2 , and its multiplicity. If $\lambda_2 = -2\kappa$, the first non-zero derivative is the κ th derivative

$$U = T^{(\kappa)}\left(\frac{1}{2}\right) = (-2)^\kappa \kappa! P_2, \quad (13)$$

where P_2 is the projection matrix for $\lambda_2 = -2\kappa$, with rank the multiplicity of λ_2 . Furthermore, the rate of convergence to $T(1/2)$ as $\theta \rightarrow 1/2$ is determined by the second largest eigenvalue

$$T(\theta) = T\left(\frac{1}{2}\right) + (1 - 2\theta)^\kappa P_2 + o((1 - 2\theta)^\kappa). \quad (14)$$

The smaller the second largest eigenvalue, the faster the convergence to $T(1/2)$. Dudoit and Speed (1999) showed that for sibships of any size k , the second largest eigenvalue of the infinitesimal generator is $\lambda_2 = -4$. However, this is not the case for all types of relatives. Below, we consider various types of unilineal relative pairs for which the second largest eigenvalue is -2 .

In the next four sections, the four main types of unilineal relative pairs are studied in detail. Unilineal relative pairs share DNA IBD on either 0 or 1 chromosome at any given locus, and are of four general types, derived from the grand-parent/grand-child, half-sib, avuncular, and cousin relationships. In general, the usual 0/1 IBD configurations are not orbits of groups acting on the set of inheritance vectors and the 2×2 transition matrix $T(\theta)$ does not satisfy the semi-property of Proposition 1. To get around this problem, we follow the work of Donnelly (1983) and define *augmented IBD configurations* which are orbits of groups acting on the set of inheritance vectors for relevant meioses. Propositions 1 – 4 can then be applied to obtain the infinitesimal generator Q for the augmented IBD configurations, its second largest eigenvalue and corresponding eigenvectors. In turn, these are used to derive linkage score statistics for the usual 0/1 IBD configurations.

3 Grand-parent-type relationship

3.1 Augmented IBD configurations

Suppose person \mathcal{A} is a direct ancestor of person \mathcal{B} , d generations removed, and assume without loss of generality that the relationship is always through the paternal line. For example, when $d = 0$, \mathcal{A} is the father of \mathcal{B} , and when $d = 1$, \mathcal{A} is the paternal grand-father of \mathcal{B} .

When studying IBD between \mathcal{A} and \mathcal{B} there are only d relevant meioses, namely, the paternal meioses giving rise to individuals $\mathcal{A}_2, \mathcal{A}_3, \dots, \mathcal{A}_d, \mathcal{A}_{d+1} = \mathcal{B}$, where \mathcal{A}_i is the i th descendant of \mathcal{A} . The case when $d = 0$ is trivial, as a child and his father always share DNA IBD on 1 chromosome. For $d > 0$, the relevant inheritance vector is $x = (x_1, \dots, x_d)$, where $x_i = 0$ if grand-paternal DNA was transmitted in the paternal meiosis giving rise to \mathcal{A}_{i+1} , 1 otherwise. The $d = 2$ case (great-grand-parent/great-grand-child) is depicted in Figure 1. Individuals \mathcal{A} and \mathcal{B} share DNA IBD on 1 chromosome if the inheritance vector is $x_0 = (0, 0, \dots, 0)$, otherwise, they share DNA IBD

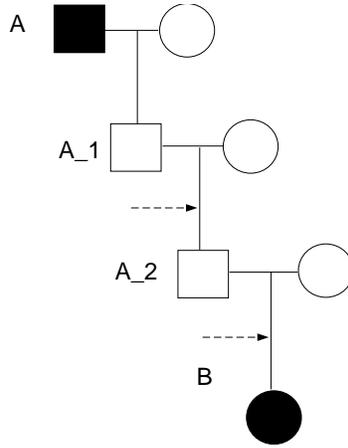


Figure 1: Grand-parent-type relationship, $d = 2$. \mathcal{A} is the great-grand-father of \mathcal{B} . The relevant paternal meioses are indicated by arrows.

on 0 chromosome.

Following Donnelly (1983), IBD configurations are defined to be the orbits of the symmetric group S_d on d letters acting on the set \mathcal{X} of all 2^d inheritance vectors. There are $d+1$ orbits, where the i th orbit is the set of all inheritance vectors with i coordinates equal to 1, that is, for $i = 0, \dots, d$,

$$\mathcal{C}_i = \{x \in \mathcal{X} : \Delta(x, x_0) = i\} \quad \text{and} \quad |\mathcal{C}_i| = \binom{d}{i},$$

where $\Delta(x, x_0)$ is the Hamming distance, i.e., the number of coordinates at which x and x_0 differ.

Proposition 5 Transition matrix for grand-parent-type relationship.

The $(d+1) \times (d+1)$ transition matrix for the augmented IBD configurations has the form $T(\theta) = \sum_h (1 - 2\theta)^{-\lambda_h/2} P_h$, where λ_h are the real eigenvalues of the infinitesimal generator $Q = T'(0)$ and P_h are the corresponding projection matrices. The ij th entry, $i, j = 0, \dots, d$, of the infinitesimal generator is

$$q_{ij} = \begin{cases} -d, & \text{if } i = j, \\ d - i, & \text{if } j = i + 1, \\ i, & \text{if } j = i - 1, \\ 0, & \text{otherwise,} \end{cases}$$

that is, Q has the tridiagonal form

$$Q = \begin{bmatrix} -d & d & 0 & 0 & \dots & 0 & 0 & 0 \\ 1 & -d & d-1 & 0 & \dots & 0 & 0 & 0 \\ 0 & 2 & -d & d-2 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & d-1 & -d & 1 \\ 0 & 0 & 0 & 0 & \dots & 0 & d & -d \end{bmatrix}.$$

The stationary distribution α has entries

$$\alpha_i = \frac{\binom{d}{i}}{2^d}, \quad i = 0, \dots, d.$$

The second largest eigenvalue of Q is $\lambda_2 = -2$, with multiplicity one, and the corresponding eigenvector is v with entries

$$v_i = \frac{2i - d}{\sqrt{d}},$$

and with unit norm with respect to the inner product \langle, \rangle_α . Consequently, the first derivative U of the transition matrix $T(\theta)$ at $\theta = 1/2$ has rank 1 and entries

$$u_{ij} = -2\alpha_j v_i v_j = -2 \frac{\binom{d}{j}}{2^d} \frac{(2i - d)(2j - d)}{d}.$$

Proof. The proof relies on Propositions 1 – 4 and follows closely the methods used in Appendix C of Dudoit and Speed (1999). Thus, only the main steps are presented.

From Proposition 1, the entries of Q are given by $q_{ij} = \sum_{y \in \mathcal{C}_j} I(\Delta(x, y) = 1) - dI(i = j)$, where $x \in \mathcal{C}_i$. From Proposition 3 and Lemma 1, the eigenvalues of $B = Q + dI$ belong to the set $\left\{ (d - 2i) \binom{d}{i} : i = 0, \dots, d \right\}$. Furthermore, from Proposition 3, the d eigenvectors of the adjacency matrix A corresponding to the eigenvalue $d - 2$ are given by $V_i(x) = 2I(x_i = 1) - 1$, $i = 1, \dots, d$. Let $V = \sum_{i=1}^d V_i$. Then

$$V(x) = 2 \sum_{i=1}^d I(x_i = 1) - d = 2j - d, \quad \text{if } x \in \mathcal{C}_j,$$

and V is an eigenvector of A which is constant over the IBD configurations. Thus, by Lemma 3, V yields an eigenvector of B corresponding to the eigenvalue $d - 2$. It remains to show that $d - 2$ has multiplicity one for B , that is, V is the only eigenvector of A for $\lambda = d - 2$ that is constant over the orbits. We may follow a similar argument as in the proof of Proposition 4 in Dudoit and Speed (1999). The orthogonal complement of V in the eigenspace of A for $\lambda = d - 2$ is spanned by the following d vectors

$$W_i = V_i - \frac{\langle V_i, V \rangle}{|V|^2} V = V_i - \frac{|V_i|^2}{|V|^2} V = V_i - \frac{1}{d} V, \quad 1 \leq i \leq d.$$

For any orbit \mathcal{C}

$$\sum_{x \in \mathcal{C}} W_i(x) = \sum_{x \in \mathcal{C}} V_i(x) - \frac{1}{d} \sum_{i=1}^d \sum_{x \in \mathcal{C}} V_i(x) = 0,$$

where we use the fact that $\forall i, j = 1, \dots, d, \sum_{x \in \mathcal{C}} V_i(x) = \sum_{x \in \mathcal{C}} V_j(x)$. Hence, no eigenvector in the orthogonal complement of V in the eigenspace of A for $\lambda = d - 2$ is constant over the orbits of S_d . Consequently, by Lemma 2, $d - 2$ is an eigenvalue of B with multiplicity 1. \square

3.2 0/1 IBD configurations

For grand-parent-type relationships, the $d + 1$ augmented IBD configurations are collapsed into the usual two configurations, $\tilde{\mathcal{C}}_1 = \mathcal{C}_0$ and $\tilde{\mathcal{C}}_0 = \cup_{i=1}^d \mathcal{C}_i$, corresponding to sharing 1 and 0 IBD, respectively. The stationary distribution for these two configurations (in the order 0, 1) is

$$\tilde{\alpha} = \left(1 - \frac{1}{2^d}, \frac{1}{2^d} \right).$$

The first derivative of the collapsed transition matrix at $\theta = 0$ is

$$\tilde{Q} = d \begin{bmatrix} -\frac{1}{2^{d-1}} & \frac{1}{2^{d-1}} \\ 1 & -1 \end{bmatrix},$$

and at $\theta = 1/2$ is

$$\tilde{U} = \frac{d}{2^{d-1}} \begin{bmatrix} -\frac{1}{2^{d-1}} & \frac{1}{2^{d-1}} \\ 1 & -1 \end{bmatrix}.$$

In general, the collapsed transition matrix $\tilde{T}(\theta)$ for the usual 0/1 IBD configurations does not satisfy the semi-group property $\tilde{T}(\theta_1 * \theta_2) = \tilde{T}(\theta_1)\tilde{T}(\theta_2)$ of Proposition 1. The semi-group property holds for grand-parent/grand-child pairs ($d = 1$, Section 3.3), but not for the great-grand-parent/great-grand-child case ($d = 2$, Section 3.4).

3.3 Grand-parent/grand-child pair

If \mathcal{A} is the grand-father of \mathcal{B} , $d = 1$ and there are only two augmented IBD configurations, the usual configurations corresponding to sharing DNA IBD on 0 and 1 chromosome. The infinitesimal generator is

$$Q = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix},$$

with eigenvalues $\lambda_1 = 0$ and $\lambda_2 = -2$. The stationary distribution is

$$\alpha = \left(\frac{1}{2}, \frac{1}{2}\right),$$

and the transition matrix, satisfying the semi-group property, is

$$T(\theta) = \begin{bmatrix} 1 - \theta & \theta \\ \theta & 1 - \theta \end{bmatrix}.$$

3.4 Great-grand-parent/great-grand-child pair

If \mathcal{A} is the great-grand-father of \mathcal{B} , $d = 2$ and there are three augmented IBD configurations, $\mathcal{C}_0 = \{(0, 0)\}$, $\mathcal{C}_1 = \{(0, 1), (1, 0)\}$, and $\mathcal{C}_2 = \{(1, 1)\}$. The infinitesimal generator is

$$Q = \begin{bmatrix} -2 & 2 & 0 \\ 1 & -2 & 1 \\ 0 & 2 & -2 \end{bmatrix},$$

with eigenvalues $\lambda_1 = 0$, $\lambda_2 = -2$, and $\lambda_3 = -4$ and corresponding eigenvectors $(1, 1, 1)$, $(-\sqrt{2}, 0, \sqrt{2})$, and $(1, -1, 1)$, with unit norm with respect to the inner product \langle, \rangle_α . The stationary distribution is

$$\alpha = \left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right),$$

and the transition matrix is

$$\begin{aligned}
 T(\theta) &= T\left(\frac{1}{2}\right) + (1 - 2\theta)P_2 + (1 - 2\theta)^2P_3 \\
 &= \frac{1}{4} \begin{bmatrix} 1 & 2 & 1 \\ 1 & 2 & 1 \\ 1 & 2 & 1 \end{bmatrix} + (1 - 2\theta) \begin{bmatrix} -\sqrt{2} \\ 0 \\ \sqrt{2} \end{bmatrix} \begin{bmatrix} -\frac{\sqrt{2}}{4} & 0 & \frac{\sqrt{2}}{4} \end{bmatrix} \\
 &\quad + (1 - 2\theta)^2 \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \begin{bmatrix} \frac{1}{4} & -\frac{1}{2} & \frac{1}{4} \end{bmatrix} \\
 &= \begin{bmatrix} 1 - 2\theta + \theta^2 & 2\theta - 2\theta^2 & \theta^2 \\ \theta - \theta^2 & 1 - 2\theta + 2\theta^2 & \theta - \theta^2 \\ \theta^2 & 2\theta - 2\theta^2 & 1 - 2\theta + \theta^2 \end{bmatrix}.
 \end{aligned}$$

The collapsed transition matrix for the 0/1 IBD configurations is

$$\tilde{T}(\theta) = \begin{bmatrix} \frac{1}{3}(3 - 2\theta + \theta^2) & \frac{1}{3}(2\theta - \theta^2) \\ 2\theta - \theta^2 & 1 - 2\theta + \theta^2 \end{bmatrix},$$

with stationary distribution

$$\tilde{\alpha} = \left(\frac{3}{4}, \frac{1}{4}\right),$$

and matrix of first derivatives at $\theta = 1/2$

$$\tilde{U} = \begin{bmatrix} -\frac{1}{3} & \frac{1}{3} \\ 1 & -1 \end{bmatrix}.$$

Note that the collapsed 2×2 transition matrix $\tilde{T}(\theta)$ does not satisfy the semi-group property $\tilde{T}(\theta_1 * \theta_2) = \tilde{T}(\theta_1)\tilde{T}(\theta_2)$.

4 Half-sib-type relationship

4.1 Augmented IBD configurations

In the half-sib-type relationship, person \mathcal{A} , or some ancestor of \mathcal{A} , is the half-sib of \mathcal{B} , or of some ancestor of \mathcal{B} . In other words, \mathcal{A} and \mathcal{B} are the n_A th and n_B th generation descendants, respectively, of a pair of half-sibs. To simplify notation, assume without loss of generality that the relationship is through

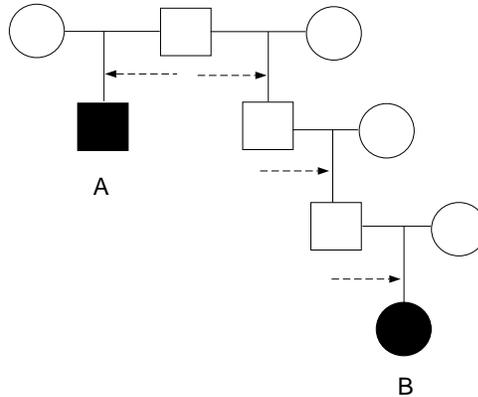


Figure 2: Half-sib-type relationship, $d = 4$ ($n_A = 0, n_B = 2$). The relevant meioses are indicated by arrows.

the paternal line. Then, there are $d = 2 + n_A + n_B$ relevant meioses, and the corresponding inheritance vector is $x = (x_1, \dots, x_d)$, where x_1 and x_2 refer to the meioses giving rise to the half-sibs, x_3, \dots, x_{n_A+2} refer to the meioses giving rise to \mathcal{A} and his ancestors and x_{n_A+3}, \dots, x_d refer to the meioses giving rise to \mathcal{B} and her ancestors. The first two entries of the inheritance vector are specific to the half-sib relationship, while the remaining entries are like in the grand-parent-type relationship. Individuals \mathcal{A} and \mathcal{B} share DNA IBD on 1 chromosome if $x = (0, 0, 0, \dots, 0)$ or $(1, 1, 0, \dots, 0)$, otherwise they share DNA IBD on 0 chromosome.

Following Donnelly (1983), there are $2(d - 1)$ IBD configurations, which correspond to the orbits of the cyclic group of order 2, C_2 , acting on the first 2 coordinates of the inheritance vectors and of the symmetric group on $d - 2$ letters, S_{d-2} , acting on the remaining $d - 2$ coordinates. The IBD configurations are labeled \mathcal{C}_{A_i} and \mathcal{C}_{B_i} , $i = 0, \dots, d - 2$, where \mathcal{C}_{A_i} refers to the set of inheritance vectors with first two coordinates $(0, 0)$ or $(1, 1)$, and i 1's among the last $d - 2$ coordinates. Likewise, for \mathcal{C}_{B_i} , the first two coordinates are $(0, 1)$ or $(1, 0)$.

$$|\mathcal{C}_{A_i}| = |\mathcal{C}_{B_i}| = 2 \binom{d-2}{i}.$$

Proposition 6 Transition matrix for half-sib-type relationship.

The $2(d - 1) \times 2(d - 1)$ transition matrix for the augmented IBD config-

urations has the form $T(\theta) = \sum_h (1 - 2\theta)^{-\lambda_h/2} P_h$, where λ_h are the real eigenvalues of the infinitesimal generator $Q = T'(0)$ and P_h are the corresponding projection matrices. When the IBD configurations are ordered as $\mathcal{C}_{A_0}, \mathcal{C}_{B_0}, \mathcal{C}_{A_1}, \mathcal{C}_{B_1}, \dots, \mathcal{C}_{A_{(d-2)}}, \mathcal{C}_{B_{(d-2)}}$, the infinitesimal generator has the following block-tridiagonal form

$$Q = \begin{bmatrix} B_2 - dI & (d-2)I & & & \dots & & & & & \\ & I & B_2 - dI & (d-3)I & & & & & & \\ & & 2I & B_2 - dI & (d-4)I & \dots & & & & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \\ & & & & & \dots & (d-3)I & B_2 - dI & I & \\ & & & & & \dots & & (d-2)I & B_2 - dI & \end{bmatrix},$$

where B_2 is the adjacency matrix for the $d = 2$ quotient graph, that is, the infinitesimal generator Q for the $d = 2$ case plus twice the identity matrix

$$B_2 = \begin{bmatrix} 0 & 2 \\ 2 & 0 \end{bmatrix}.$$

The stationary distribution α has entries

$$\alpha_{A_i} = \alpha_{B_i} = \frac{2 \binom{d-2}{i}}{2^d}, \quad i = 0, \dots, d-2.$$

For half-sibs ($d = 2$), the 2 eigenvalues of Q are $\lambda_1 = 0$ and $\lambda_2 = -4$, and the first derivative of the transition matrix vanishes at $\theta = 1/2$. However, for more distant half-sibs ($d > 2$), the second largest eigenvalue of Q is $\lambda_2 = -2$, with multiplicity one, and the corresponding eigenvector is v with entries

$$v_{A_i} = v_{B_i} = \frac{2i - (d-2)}{\sqrt{d-2}},$$

and with unit norm with respect to the inner product \langle, \rangle_α . Consequently, for $d > 2$, the first derivative U of the transition matrix $T(\theta)$ at $\theta = 1/2$ has rank 1 and entries

$$u_{A_i, A_j} = u_{A_i, B_j} = u_{B_i, B_j} = u_{B_i, A_j} = -2 \frac{\binom{d-2}{j} (2i - (d-2))(2j - (d-2))}{2^{d-1} (d-2)}.$$

Proof. Again, the proof relies on Propositions 1 – 4 and follows closely the methods used in Appendix C of Dudoit and Speed (1999). We only sketch the main steps.

When $d = 2$,

$$Q = \begin{bmatrix} -2 & 2 \\ 2 & -2 \end{bmatrix},$$

with eigenvalues $\lambda_1 = 0$ and $\lambda_2 = -4$.

Consider now the case when $d > 2$. From Proposition 3 and Lemma 1, the eigenvalues of $B = Q + dI$ belong to the set $\left\{ (d - 2i) \binom{d}{i} : i = 0, \dots, d \right\}$. Furthermore, from Proposition 3, the d eigenvectors of the adjacency matrix A corresponding to the eigenvalue $d - 2$ are given by $V_i(x) = 2I(x_i = 1) - 1$, $i = 1, \dots, d$. Let $V = \sum_{i=3}^d V_i$. Then

$$V(x) = 2 \sum_{i=3}^d I(x_i = 1) - (d - 2) = 2j - (d - 2), \quad \text{if } x \in \mathcal{C}_{A_j} \text{ or } \mathcal{C}_{B_j},$$

and V is an eigenvector of A which is constant over the IBD configurations. Thus, by Lemma 3, V yields an eigenvector of B corresponding to the eigenvalue $d - 2$. It remains to show that $d - 2$ has multiplicity one for B . We may follow a similar argument as in the proof of Proposition 5, using the facts that for any orbit \mathcal{C} and $\forall i, j \geq 3$

$$\sum_{x \in \mathcal{C}} V_i(x) = \sum_{x \in \mathcal{C}} V_j(x),$$

and for any \mathcal{C}

$$\sum_{x \in \mathcal{C}} V_1(x) = \sum_{x \in \mathcal{C}} V_2(x) = 0.$$

□

4.2 0/1 IBD configurations

For distant half-sibs with $d > 2$ (see Section 4.3 for the special case $d = 2$), the $2(d - 1)$ augmented IBD configurations are collapsed into the usual two configurations, $\tilde{\mathcal{C}}_1 = \mathcal{C}_{A0}$ and $\tilde{\mathcal{C}}_0 = \mathcal{C}_{B0} \cup \bigcup_{i=1}^{d-2} (\mathcal{C}_{A_i} \cup \mathcal{C}_{B_i})$, corresponding to sharing 1 and 0 IBD, respectively. The stationary distribution for these two configurations (in the order 0, 1) is

$$\tilde{\alpha} = \left(1 - \frac{1}{2^{d-1}}, \frac{1}{2^{d-1}} \right).$$

The first derivative of the collapsed transition matrix at $\theta = 0$ is

$$\tilde{Q} = d \begin{bmatrix} -\frac{1}{2^{d-1}-1} & \frac{1}{2^{d-1}-1} \\ 1 & -1 \end{bmatrix},$$

and at $\theta = 1/2$ is

$$\tilde{U} = \frac{d-2}{2^{d-2}} \begin{bmatrix} -\frac{1}{2^{d-1}-1} & \frac{1}{2^{d-1}-1} \\ 1 & -1 \end{bmatrix}.$$

4.3 Half-sibs

If \mathcal{A} is the half-sib of \mathcal{B} , $d = 2$ and there are only two IBD configurations, the usual configurations corresponding to sharing DNA IBD on 0 and 1 chromosome. The infinitesimal generator is

$$Q = \begin{bmatrix} -2 & 2 \\ 2 & -2 \end{bmatrix},$$

with eigenvalues $\lambda_1 = 0$ and $\lambda_2 = -4$. The stationary distribution is

$$\alpha = \left(\frac{1}{2}, \frac{1}{2} \right),$$

and the transition matrix is

$$T(\theta) = \begin{bmatrix} \psi & 1 - \psi \\ 1 - \psi & \psi \end{bmatrix},$$

where $\psi = \theta^2 + (1 - \theta)^2$. The first derivative at $\theta = 1/2$ vanishes and the second derivative is given by

$$U = T''\left(\frac{1}{2}\right) = \begin{bmatrix} 4 & -4 \\ -4 & 4 \end{bmatrix}.$$

4.4 Half-sibs once removed

If \mathcal{A} is the half-sib once removed of \mathcal{B} , $d = 3$ and there are four IBD configurations, $\mathcal{C}_{A0} = \{(0, 0, 0), (1, 1, 0)\}$, $\mathcal{C}_{B0} = \{(0, 1, 0), (1, 0, 0)\}$, $\mathcal{C}_{A1} = \{(0, 0, 1), (1, 1, 1)\}$, and $\mathcal{C}_{B1} = \{(0, 1, 1), (1, 0, 1)\}$. The infinitesimal generator is

$$Q = \begin{bmatrix} -3 & 2 & 1 & 0 \\ 2 & -3 & 0 & 1 \\ 1 & 0 & -3 & 2 \\ 0 & 1 & 2 & -3 \end{bmatrix},$$

with eigenvalues $\lambda_1 = 0$, $\lambda_2 = -2$, $\lambda_3 = -4$, and $\lambda_4 = -6$, and corresponding eigenvectors $(1, 1, 1, 1)$, $(-1, -1, 1, 1)$, $(-1, 1, -1, 1)$, and $(-1, 1, 1, -1)$, with unit norm with respect to the inner product \langle, \rangle_α . The stationary distribution is

$$\alpha = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4} \right),$$

and the transition matrix is

$$\begin{aligned} T(\theta) &= T\left(\frac{1}{2}\right) + (1 - 2\theta)P_2 + (1 - 2\theta)^2P_3 + (1 - 2\theta)^3P_4 \\ &= \frac{1}{4} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} + \frac{1}{4}(1 - 2\theta) \begin{bmatrix} 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 \\ -1 & -1 & 1 & 1 \end{bmatrix} \\ &\quad + \frac{1}{4}(1 - 2\theta)^2 \begin{bmatrix} 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \end{bmatrix} + \frac{1}{4}(1 - 2\theta)^3 \begin{bmatrix} 1 & -1 & -1 & 1 \\ -1 & 1 & 1 & -1 \\ -1 & 1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}. \end{aligned}$$

The matrix of first derivatives at $\theta = 1/2$ is

$$U = \frac{-1}{2} \begin{bmatrix} 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 \\ -1 & -1 & 1 & 1 \end{bmatrix}.$$

The collapsed transition matrix for the 0/1 IBD configurations is

$$\tilde{T}(\theta) = \begin{bmatrix} \frac{1}{3}(3 - 3\theta + 4\theta^2 - 2\theta^3) & \frac{1}{3}(3\theta - 4\theta^2 + 2\theta^3) \\ 3\theta - 4\theta^2 + 2\theta^3 & 1 - 3\theta + 4\theta^2 - 2\theta^3 \end{bmatrix},$$

with stationary distribution

$$\tilde{\alpha} = \left(\frac{3}{4}, \frac{1}{4} \right),$$

and matrix of first derivatives at $\theta = 1/2$

$$\tilde{U} = \begin{bmatrix} -\frac{1}{6} & \frac{1}{6} \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix}.$$

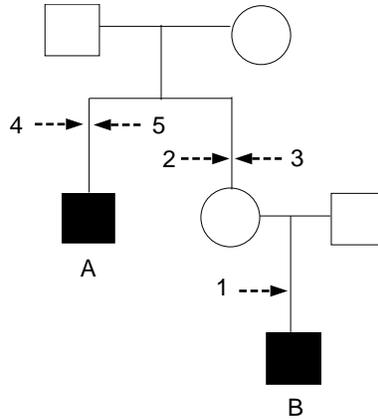


Figure 3: Uncle/nephew, $d = 5$. The relevant meioses are indicated by arrows and numbered.

5 Avuncular relationship

5.1 Uncle/nephew pair

In the special case when \mathcal{A} is the uncle of \mathcal{B} , there are $d = 5$ relevant meioses, namely the 2 meioses giving rise to \mathcal{A} (meioses 4 and 5 in Figure 3), the 2 meioses giving rise to the mother of \mathcal{B} (meioses 2 and 3) and the maternal meiosis giving rise to \mathcal{B} (meiosis 1). The relevant inheritance vector is $x = (x_1, \dots, x_5)$, where $x_i = 0$ if grand-paternal DNA was transmitted in the i th meiosis, 1 otherwise. Individuals \mathcal{A} and \mathcal{B} share DNA IBD on 1 chromosome if the inheritance vector is one of the following 16 inheritance vectors $(0, 0, *, 0, *)$, $(0, 1, *, 1, *)$, $(1, *, 0, *, 0)$, or $(1, *, 1, *, 1)$, where $*$ is either 0 or 1. Otherwise they share DNA IBD on 0 chromosome.

As in Donnelly (1983), consider the following four IBD configurations

$$\mathcal{C}_1 = \begin{Bmatrix} 00000 & 10000 \\ 01111 & 11111 \\ 00101 & 11010 \\ 01010 & 10101 \end{Bmatrix}, \quad \mathcal{C}_2 = \begin{Bmatrix} 00001 & 10010 \\ 01110 & 11101 \\ 00100 & 11000 \\ 01011 & 10111 \end{Bmatrix},$$

$$\mathcal{C}_3 = \begin{Bmatrix} 00011 & 10011 \\ 01100 & 11100 \\ 00110 & 11001 \\ 01001 & 10110 \end{Bmatrix}, \quad \mathcal{C}_4 = \begin{Bmatrix} 00010 & 10001 \\ 01101 & 11110 \\ 00111 & 11011 \\ 01000 & 10100 \end{Bmatrix}.$$

The first two IBD configurations correspond to sharing DNA IBD on 1 chromosome. The infinitesimal generator is

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

and the stationary distribution is

$$\alpha = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4} \right).$$

The infinitesimal generator has eigenvalues $\lambda_1 = 0$, $\lambda_2 = -4$, $\lambda_3 = -6$, and $\lambda_4 = -8$. The eigenvector corresponding to $\lambda_2 = -4$ is $v = \sqrt{2}(1, 0, -1, 0)$. Thus, the first derivative of the transition matrix vanishes at $\theta = 1/2$ and the second derivative is

$$U = (8\alpha_j v_i v_j) = \begin{bmatrix} 4 & 0 & -4 & 0 \\ 0 & 0 & 0 & 0 \\ -4 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

The 4 augmented IBD configurations are collapsed into the usual two configurations, $\tilde{\mathcal{C}}_1 = \mathcal{C}_1 \cup \mathcal{C}_2$ and $\tilde{\mathcal{C}}_0 = \mathcal{C}_3 \cup \mathcal{C}_4$, corresponding to sharing 1 and 0 IBD, respectively. The transition matrix for these two configurations (in the order 0,1) is

$$\tilde{T}(\theta) = \begin{bmatrix} \psi(1-\theta) + \frac{1}{2}\theta & 1 - \psi(1-\theta) - \frac{1}{2}\theta \\ 1 - \psi(1-\theta) - \frac{1}{2}\theta & \psi(1-\theta) + \frac{1}{2}\theta \end{bmatrix}.$$

This matrix is given in Campbell and Elston (1971) and Thompson (1986). Note that the collapsed transition matrix $\tilde{T}(\theta)$ does not satisfy the semi-group property $\tilde{T}(\theta_1 * \theta_2) = \tilde{T}(\theta_1)\tilde{T}(\theta_2)$. The stationary distribution is $\tilde{\alpha} = (1/2, 1/2)$, the first derivative of the collapsed transition matrix at $\theta = 0$ is

$$\tilde{Q} = \frac{5}{2} \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix},$$

and the second derivative of the collapsed transition matrix at $\theta = 1/2$ is

$$\tilde{U} = \begin{bmatrix} 2 & -2 \\ -2 & 2 \end{bmatrix}.$$

The stationary distribution α also has a block form, with ij th entry

$$\alpha_{ji} = \frac{\binom{d-5}{i}}{2^{d-3}}, \quad i = 0, \dots, d-5, \quad j = 1, 2, 3, 4.$$

For $d > 5$, the second largest eigenvalue of Q is $\lambda_2 = -2$, with multiplicity one, and the corresponding eigenvector is v with entries

$$v_{ji} = \frac{2i - (d-5)}{\sqrt{d-5}},$$

and with unit norm with respect to the inner product \langle, \rangle_α . Consequently, the first derivative U of the transition matrix $T(\theta)$ at $\theta = 1/2$ has rank 1 and entries

$$u_{ji,lk} = -2\alpha_{lk}v_{ji}v_{lk} = -2\frac{\binom{d-5}{k}}{2^{d-3}} \frac{(2i - (d-5))(2k - (d-5))}{d-5}.$$

The proof is similar to that for the half-sib-type relationship and considers the first 5 coordinates of the inheritance vectors separately from the last $d-5$.

5.3 0/1 IBD configurations

Individuals \mathcal{A} and \mathcal{B} share DNA IBD on 1 chromosome if $x \in \tilde{\mathcal{C}}_1 = \mathcal{C}_{10} \cup \mathcal{C}_{20}$, otherwise they share DNA IBD on 0 chromosome. The stationary distribution for these two configurations (in the order 0, 1) is

$$\tilde{\alpha} = \left(1 - \frac{1}{2^{d-4}}, \frac{1}{2^{d-4}}\right).$$

The first derivative of the collapsed transition matrix at $\theta = 0$ is

$$\tilde{Q} = (d-5/2) \begin{bmatrix} -\frac{1}{2^{d-4}-1} & \frac{1}{2^{d-4}-1} \\ 1 & -1 \end{bmatrix},$$

and at $\theta = 1/2$ is

$$\tilde{U} = \frac{d-5}{2^{d-5}} \begin{bmatrix} -\frac{1}{2^{d-4}-1} & \frac{1}{2^{d-4}-1} \\ 1 & -1 \end{bmatrix}.$$

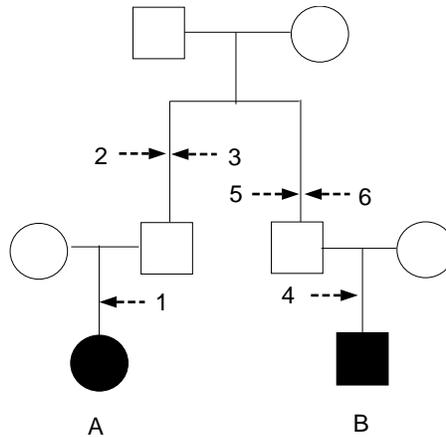


Figure 4: First-cousins, $d = 6$. The relevant meioses are indicated by arrows.

6 Cousin-type relationship

6.1 First-cousin pair

In the special case when \mathcal{A} and \mathcal{B} are first-cousins, there are $d = 6$ relevant meioses, namely the two paternal meioses giving rise to \mathcal{A} and \mathcal{B} , and the four meioses giving rise to the fathers of \mathcal{A} and \mathcal{B} (Figure 4). The relevant inheritance vector is $x = (x_1, \dots, x_6)$, where $x_i = 0$ if grand-paternal DNA was transmitted in the i th meiosis, 1 otherwise. Individuals \mathcal{A} and \mathcal{B} share DNA IBD on 1 chromosome if the inheritance vector is one of the following 16 inheritance vectors $(0, 0, *, 0, 0, *)$, $(0, 1, *, 0, 1, *)$, $(1, *, 0, 1, *, 0)$, or $(1, *, 1, 1, *, 1)$, where $*$ is either 0 or 1. Otherwise they share DNA IBD on 0 chromosome. We consider the 7 IBD configurations defined by Donnelly (1983) p. 52–53. The first two IBD configurations correspond to sharing DNA IBD on 1 chromosome.

The infinitesimal generator is

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

and the stationary distribution is

$$\alpha = \frac{1}{8}(1, 1, 1, 1, 1, 2, 1).$$

The infinitesimal generator has eigenvalues $\lambda_1 = 0$, $\lambda_2 = -4$ (multiplicity 2), $\lambda_3 = -6$, $\lambda_4 = -8$ (multiplicity 2), $\lambda_5 = -12$. The two eigenvectors corresponding to $\lambda_2 = -4$ are $v = (3, 1, 1, -1, 1, -1, -3)/\sqrt{3}$ and $\tilde{v} = \sqrt{2/3}(0, -1, -1, -2, 2, 1, 0)$, and $\langle v, \tilde{v} \rangle_\alpha = 0$. Thus, the first derivative of the transition matrix vanishes at $\theta = 1/2$ and the second derivative is

$$U = (8\alpha_j(v_i v_j + \tilde{v}_i \tilde{v}_j)) = \begin{bmatrix} 3 & 1 & 1 & -1 & 1 & -2 & -3 \\ 1 & 1 & 1 & 1 & -1 & -2 & -1 \\ 1 & 1 & 1 & 1 & -1 & -2 & -1 \\ -1 & 1 & 1 & 3 & -3 & -2 & 1 \\ 1 & -1 & -1 & -3 & 3 & 2 & -1 \\ -1 & -1 & -1 & -1 & 1 & 2 & 1 \\ -3 & -1 & -1 & 1 & -1 & 2 & 3 \end{bmatrix}.$$

The 7 augmented IBD configurations are collapsed into the usual two configurations, $\tilde{\mathcal{C}}_1 = \mathcal{C}_1 \cup \mathcal{C}_2$ and $\mathcal{C}_0 = \mathcal{C}_3 \cup \dots \cup \mathcal{C}_7$, corresponding to sharing 1 and 0 IBD, respectively. The transition matrix for these two configurations (in the order 0,1) is

$$\tilde{T}(\theta) = \begin{bmatrix} \frac{1}{3}(2 + \frac{1}{2}\theta^2 + \psi(1 - \theta)^2) & \frac{1}{3}(1 - \frac{1}{2}\theta^2 - \psi(1 - \theta)^2) \\ 1 - \frac{1}{2}\theta^2 - \psi(1 - \theta)^2 & \frac{1}{2}\theta^2 + \psi(1 - \theta)^2 \end{bmatrix}.$$

This matrix is given in Campbell and Elston (1971) and Thompson (1986). Note that the collapsed transition matrix does not satisfy the semi-group property $\tilde{T}(\theta_1 * \theta_2) = \tilde{T}(\theta_1)\tilde{T}(\theta_2)$. The stationary distribution is $\tilde{\alpha} = (3/4, 1/4)$, the first derivative of the collapsed transition matrix at $\theta = 0$ is

$$\tilde{Q} = 4 \begin{bmatrix} -\frac{1}{3} & \frac{1}{3} \\ 1 & -1 \end{bmatrix},$$

and the second derivative of the collapsed transition matrix at $\theta = 1/2$ is

$$\tilde{U} = \begin{bmatrix} 1 & -1 \\ -3 & 3 \end{bmatrix}.$$

The stationary distribution α also has a block form, with ij th entry

$$\alpha_{ji} = \frac{\binom{d-6}{i}}{2^{d-3}}, \quad j \neq 6, \quad \text{and} \quad \alpha_{6i} = \frac{\binom{d-6}{i}}{2^{d-4}}, \quad i = 0, \dots, d-6.$$

For $d > 6$, the second largest eigenvalue of Q is $\lambda_2 = -2$, with multiplicity one, and the corresponding eigenvector is v with entries

$$v_{ji} = \frac{2i - (d-6)}{\sqrt{d-6}},$$

and with unit norm with respect to the inner product \langle, \rangle_α . Consequently, the first derivative U of the transition matrix $T(\theta)$ at $\theta = 1/2$ has rank 1 and entries

$$u_{ji,lk} = -2\alpha_{lk}v_{ji}v_{lk}.$$

The proof is similar to that for the half-sib-type relationship and considers the first 6 coordinates of the inheritance vectors separately from the last $d-6$.

6.3 0/1 IBD configurations

Individuals \mathcal{A} and \mathcal{B} share DNA IBD on 1 chromosome if $x \in \tilde{\mathcal{C}}_1 = \mathcal{C}_{10} \cup \mathcal{C}_{20}$, otherwise they share DNA IBD on 0 chromosome. The stationary distribution for these two configurations (in the order 0, 1) is

$$\tilde{\alpha} = \left(1 - \frac{1}{2^{d-4}}, \frac{1}{2^{d-4}}\right).$$

The first derivative of the collapsed transition matrix at $\theta = 0$ is

$$\tilde{Q} = (d-2) \begin{bmatrix} -\frac{1}{2^{d-4}-1} & \frac{1}{2^{d-4}-1} \\ 1 & -1 \end{bmatrix},$$

and at $\theta = 1/2$ is

$$\tilde{U} = \frac{d-6}{2^{d-5}} \begin{bmatrix} -\frac{1}{2^{d-4}-1} & \frac{1}{2^{d-4}-1} \\ 1 & -1 \end{bmatrix}.$$

7 Linkage score statistics for unilineal relative pairs

7.1 Conditional IBD probabilities at a marker

Genetic mapping involves identifying loci at which genotypes are associated with phenotypes. Genotype-phenotype associations are viewed as suggesting linkage, or “closeness”, of a genetic locus to a gene influencing the phenotype. In general, phenotypes and IBD configurations of related individuals at loci *linked* to genes influencing the phenotype (trait loci) are *dependent*, while phenotypes and IBD configurations at loci *unlinked* to genes are *independent*.

Consider a marker locus \mathcal{M} linked to a gene \mathcal{D} , and let θ denote the recombination fraction between these two loci. The gene \mathcal{D} could be one of several genes, unlinked to each other, influencing a quantitative or a qualitative phenotype. Let $\pi_j = \pi_j(\phi_1, \phi_2; \nu)$ and $\rho_j = \rho_j(\phi_1, \phi_2; \theta, \nu)$ denote the conditional probabilities that a relative pair with phenotypes (ϕ_1, ϕ_2) shares DNA IBD on $j = 0, 1$ chromosome at the gene \mathcal{D} and at the marker locus \mathcal{M} , respectively. Here, ν denotes parameters of the genetic model for the trait, such as genotype frequencies at the gene(s) and penetrances (i.e., conditional distribution of phenotypes given genotypes at the gene(s)). Then,

$$[\rho_0, \rho_1] = [\pi_0, \pi_1] T(\theta), \quad (15)$$

where, for unilineal relative pairs, $T(\theta)$ is a 2×2 transition matrix

$$T(\theta) = \begin{bmatrix} t_{00}(\theta) & t_{01}(\theta) \\ t_{10}(\theta) & t_{11}(\theta) \end{bmatrix},$$

with entries the probabilities of sharing DNA IBD (or not) at one locus, given sharing (or not) at another locus separated by a recombination fraction θ . Thus, the IBD probabilities at a marker locus \mathcal{M} linked to a gene \mathcal{D} have two components: one component $T(\theta)$ involving the recombination fraction θ between the two loci, and the other involving the conditional IBD probabilities π_j at the gene \mathcal{D} . The later depend on typically unknown and numerous parameters ν of the genetic model for the trait. Thompson (1997) refers to these two components as the *scale* of the genetic distance of interest and the *specificity* of the gene, respectively. Both the scale and specificity components affect the strength of the association between phenotypes and IBD configurations at the marker locus \mathcal{M} , and hence the power to detect

linkage. The strength of the association at the gene \mathcal{D} depends on the mode of inheritance, the genotype frequencies at the gene(s), and the phenotypes of the pair of relatives. As θ increases from 0 to $1/2$, this association is attenuated by recombination between the gene and the marker locus. The rate of convergence, as $\theta \rightarrow 1/2$, of the IBD probabilities ρ to the stationary distribution α is determined by the second largest eigenvalue of the infinitesimal generator Q (cf. equation (14) in Section 2).

7.2 General form of linkage score statistic

We are interested in testing the null hypothesis of no linkage between the marker \mathcal{M} and the gene \mathcal{D} , $H_0 : \theta = 1/2$. Suppose we have phenotype and IBD data for n unilineal relative pairs of a given type (e.g., n cousin pairs), in the form of phenotype pairs (ϕ_{1i}, ϕ_{2i}) and indicators N_{ji} for the number $j = 0, 1$ of chromosomes sharing DNA IBD at the marker locus \mathcal{M} for the i th relative pair, $i = 1, \dots, n$. Let $\pi_{ji} = \pi_j(\phi_{1i}, \phi_{2i}; \nu)$ and $\rho_{ji} = \rho_j(\phi_{1i}, \phi_{2i}; \theta, \nu)$ denote the conditional probabilities that a relative pair with phenotypes (ϕ_{1i}, ϕ_{2i}) shares DNA IBD on $j = 0, 1$ chromosome at the gene \mathcal{D} and at the marker locus \mathcal{M} , respectively.

Dudoit and Speed (2000) show that, under the following two general sampling assumptions, the different relative pairs make multiplicative contributions to the likelihood function for the IBD data conditional on phenotypes.

Assumption S1. For a given relative pair, phenotypes are conditionally independent of any phenotype and marker genotype data from other pairs, given the genotypes at the gene(s) for this pair.

Assumption S2. For a given relative pair, genotypes at the gene(s) are independent of any phenotype and marker genotype data from other pairs.

Assumption S1 rules out the influence on phenotypes of environmental factors shared by groups of families. Accommodating *between family* environmental factors would require conditioning on the environment from which the families are sampled. Note that Assumption S1 does not however rule out *within family* environmental factors. Assumption S2 rules out related pairs.

Under Assumptions S1 and S2, the log-likelihood of the IBD data *conditional* on phenotypes is

$$l(\theta, \nu) = \sum_{i=1}^n \sum_{j=0}^1 N_{ji} \ln \rho_{ji}. \quad (16)$$

Score tests are locally most powerful tests, based on the first non-zero derivative in the Taylor series expansion of the log-likelihood about the null parameter $\theta = 1/2$ (Cox and Hinkley, 1974; Rao, 1973). Derivatives of the log-likelihood are based on derivatives of the transition matrix $T(\theta)$ via the IBD sharing probabilities ρ .

Proposition 9 Linkage score statistic for unilineal relative pairs.

When the first non-zero derivative of the transition matrix $T(\theta)$ at $\theta = 1/2$ is its κ th derivative, $U = (u_{ij}) = T^\kappa(1/2)$, the score statistic for the null hypothesis $H_0 : \theta = 1/2$ of no linkage is the κ th derivative of the log-likelihood evaluated at $\theta = 1/2$ and is given by

$$ST(\nu) = \frac{u_{11}}{\alpha_1} \sum_{i=1}^n \left(\pi_{1i} - \frac{\alpha_1}{\alpha_0} \pi_{0i} \right) \left(N_{1i} - \frac{\alpha_1}{\alpha_0} N_{0i} \right), \quad (17)$$

where $\alpha = (\alpha_0, \alpha_1)$, the vector of unconditional probabilities of sharing DNA IBD on 0 and 1 chromosome, is the stationary distribution of $T(\theta)$. The null hypothesis of no linkage is rejected for large values of $ST(\nu)$ when κ is even and for small values of $ST(\nu)$ when κ is odd.

Proof. Using the fact that $T(\theta)$ is a 2×2 stochastic matrix with stationary distribution $\alpha = (\alpha_0, \alpha_1)$, the unconditional probabilities of sharing DNA IBD on 0 and 1 chromosome, we obtain the following identities satisfied by the first derivative $Q = T'(0)$ and the first non-zero derivative $U = T^\kappa(1/2)$ of $T(\theta)$ evaluated at $\theta = 1/2$.

$$q_{01} = -q_{00}, \quad q_{10} = -q_{11}, \quad \alpha_1 q_{11} = \alpha_0 q_{00},$$

and

$$u_{01} = -u_{00}, \quad u_{10} = -u_{11}, \quad \alpha_1 u_{11} = \alpha_0 u_{00}.$$

Since the first $\kappa - 1$ derivatives of $T(\theta)$ vanish at $\theta = 1/2$, then so do the first $\kappa - 1$ derivatives of the IBD probabilities ρ and of the log-likelihood $l(\theta, \nu)$ in

equation (16). Thus, the linkage score statistic is based on the κ th derivative of the log-likelihood and given by

$$\begin{aligned}
 ST(\nu) &= \left. \frac{\partial^\kappa l(\theta, \nu)}{\partial \theta^\kappa} \right|_{\theta=\frac{1}{2}} \\
 &= \sum_{i=1}^n \sum_{j=0}^1 N_{ji} \left. \frac{\partial^\kappa \rho_{ji}}{\partial \theta^\kappa} \frac{1}{\rho_{ji}} \right|_{\theta=\frac{1}{2}} \\
 &= \sum_{i=1}^n \left(N_{0i} \pi_{0i} \frac{u_{00}}{\alpha_0} + N_{0i} \pi_{1i} \frac{u_{10}}{\alpha_0} + N_{1i} \pi_{0i} \frac{u_{01}}{\alpha_1} + N_{1i} \pi_{1i} \frac{u_{11}}{\alpha_1} \right) \\
 &= \sum_{i=1}^n \left(u_{11} \pi_{1i} - u_{00} \pi_{0i} \right) \left(\frac{N_{1i}}{\alpha_1} - \frac{N_{0i}}{\alpha_0} \right) \\
 &= \frac{u_{11}}{\alpha_1} \sum_{i=1}^n \left(\pi_{1i} - \frac{\alpha_1}{\alpha_0} \pi_{0i} \right) \left(N_{1i} - \frac{\alpha_1}{\alpha_0} N_{0i} \right).
 \end{aligned}$$

The direction of rejection is determined by considering the Taylor series expansion of the log-likelihood about $\theta = 1/2$. \square

It is important to note that Proposition 9 holds for qualitative and quantitative phenotypes and whether or not IBD configurations are defined as orbits of groups acting on the set of inheritance vectors, i.e., whether or not $T(\theta)$ satisfies the semi-group property of Proposition 1.

The linkage score statistic consists of two components: a component $(N_{1i} - \alpha_1 N_{0i} / \alpha_0)$ involving the IBD indicators N_{ji} and a similar component, $(\pi_{1i} - \alpha_1 \pi_{0i} / \alpha_0)$, involving the IBD sharing probabilities π_{ji} at the gene. For complex traits, the genetic model for the trait is typically unknown and hence so are the weights $(\pi_{1i} - \alpha_1 \pi_{0i} / \alpha_0)$ in the score statistic. The simulation studies of Dudoit and Speed (2000), Goldstein et al. (2000), and Goldstein et al. (2001) for sib-pairs, half-sibs, avuncular pairs, and grand-parent/grand-child pairs showed that the weights are primarily driven by phenotypes, rather than by parameters ν of the genetic model, and that the linkage score test is generally robust to misspecification of the genetic model.

For IBD configurations defined as orbits of groups acting on the set of inheritance vectors \mathcal{X} (e.g., augmented IBD configurations in Sections 3 – 6), the results of Section 2 show that the first non-zero derivative of $T(\theta)$ at $\theta = 1/2$ and its rank are determined by the second largest eigenvalue of

the infinitesimal generator Q , λ_2 , and its multiplicity. If this second largest eigenvalue is $\lambda_2 = -2\kappa$, then the first non-zero derivative of $T(\theta)$ is its κ th derivative. Note that if the κ th derivative of the transition matrix for augmented IBD configurations vanishes, then so does the κ th derivative of the transition matrix for collapsed 0/1 IBD configurations. So, if $\lambda_2 = -2\kappa$ for the infinitesimal generator of the augmented IBD configurations, then the score statistic for both the augmented IBD configurations and collapsed 0/1 IBD configurations is based on the κ th derivative of the log-likelihood.

Sections 3 – 6 demonstrate that the infinitesimal generators for different types of relative pairs have different second largest eigenvalues, and thus linkage score tests based on different types of relative pairs have different power properties at the null. In particular, linkage score tests for pairs with $\lambda_2 = -2$ (i.e., grand-parent/grand-child, distant half-sibs with $d > 2$, distant avuncular pairs with $d > 5$, distant cousin pairs with $d > 6$) are infinitely more powerful near the null than score tests for pairs with $\lambda_2 = -4$ (i.e., half-sibs with $d = 2$, avuncular pairs with $d = 5$, cousin pairs with $d = 6$). For the latter, the slope of the power function is zero at the null. The linkage score statistic based on IBD data from different types of relative pairs with the same second largest eigenvalue (e.g., avuncular and cousin pairs) is simply the sum of the score statistics for each type of relative pairs. In Section 7.4, below, we consider the different types of relative pairs of Sections 3 – 6 and use properties of their infinitesimal generator Q to derive the specific form of the linkage score statistic for each type of pair. Results are summarized in Table 1.

Under the null hypothesis of no linkage between the marker locus \mathcal{M} and the gene \mathcal{D} , the first two moments of $ST(\nu)$ are

$$E_0[ST(\nu)|\phi] = 0,$$

and

$$\begin{aligned} Var_0[ST(\nu)|\phi] &= \left(\frac{u_{11}}{\alpha_1}\right)^2 \sum_{i=1}^n \left(\pi_{1i} - \frac{\alpha_1}{\alpha_0} \pi_{0i}\right)^2 \left(\alpha_0 \alpha_1 \left(1 + \frac{\alpha_1}{\alpha_0}\right)^2\right) \\ &= \frac{u_{11}^2}{\alpha_0 \alpha_1} \sum_{i=1}^n \left(\pi_{1i} - \frac{\alpha_1}{\alpha_0} \pi_{0i}\right)^2. \end{aligned}$$

Thus, the standardized linkage score statistic (up to the sign of u_{11}) is

$$SST(\nu) = \sqrt{\frac{\alpha_0}{\alpha_1}} \frac{\sum_{i=1}^n (\pi_{1i} - \frac{\alpha_1}{\alpha_0} \pi_{0i}) (N_{1i} - \frac{\alpha_1}{\alpha_0} N_{0i})}{\sqrt{\sum_{i=1}^n (\pi_{1i} - \frac{\alpha_1}{\alpha_0} \pi_{0i})^2}}. \quad (18)$$

So far, we have assumed full IBD information for the marker locus \mathcal{M} . In practice, this is usually not the case, as founders may not be available for typing or they may not exhibit sufficient polymorphism to establish IBD unambiguously. Several statistical methods are available to infer IBD status from marker genotype data. Common approaches rely on hidden Markov models to estimate the *inheritance distribution*, that is, the conditional distribution of IBD configurations given observed multipoint marker genotypes (Abecasis et al., 2002; Kruglyak and Lander, 1995a; Kruglyak et al., 1996; Lander and Green, 1987). An incomplete data linkage statistic may be defined as

$$\tilde{S}T(\nu) = E_0[ST(\nu)|M, \phi] = \frac{u_{11}}{\alpha_1} \sum_{i=1}^n \left(\pi_{1i} - \frac{\alpha_1}{\alpha_0} \pi_{0i} \right) \left(r_{1i} - \frac{\alpha_1}{\alpha_0} r_{0i} \right), \quad (19)$$

where M_i denotes the multipoint marker genotype data for relative pair i and the IBD indicators N_{ji} have been replaced by their expected values given the multipoint marker genotypes, $r_{ji} = pr(\text{Relative pair has IBD configuration } \mathcal{C}_j \text{ at } \mathcal{M} | M_i)$ (see Dudoit (1999) and Dudoit and Speed (2000) for greater details).

7.3 Auto-correlation function for linkage score statistic

Dense maps of genetic markers are currently available and, in practice, linkage tests are performed simultaneously at a large number of chromosomal loci (e.g., every 10 centiMorgans (cM) or more in a genome scan). Assessments of the statistical significance of mapping results should therefore be adjusted for multiples testing. Let $SST(s; \nu)$ denote the standardized linkage score statistic at a locus \mathcal{L}_s for a given type of unilineal relative pair. Feingold et al. (1993), followed by Kruglyak and Lander (1995b) and Lander and Kruglyak (1995), used extreme value theory for the distribution of Ornstein-Uhlenbeck processes to derive an approximation for the genome-wide probability that a linkage statistic exceeds a particular threshold. One of the key ingredients in this approximation is the auto-correlation function for the linkage statistics.

Proposition 10 Auto-correlation function of linkage score statistic for unilineal relative pairs - general case.

Consider loci \mathcal{L}_t and \mathcal{L}_{t+s} located $|s|$ Morgans (M) apart on a given chromosome. Under the null hypothesis of no genes on the chromosome, the auto-correlation function for linkage score statistics computed at loci \mathcal{L}_t and \mathcal{L}_{t+s} is given by

$$C(t, t + s) = \frac{t_{11}(\theta_{t,t+s}) - \alpha_1}{\alpha_0}, \quad (20)$$

where $\theta_{t,t+s}$ denotes the recombination fraction between loci \mathcal{L}_t and \mathcal{L}_{t+s} . For a stationary crossover process, with map function $M(s)$ satisfying $M(0) = 0$, $M'(0) = 1$, and $\theta_{t,t+s} = M(|s|)$, then

$$C(s) \equiv C(t, t + s) = \frac{t_{11}(M(|s|)) - \alpha_1}{\alpha_0}. \quad (21)$$

The first derivative of the auto-correlation function at $s = 0$, taken as the limit from the right, is given by

$$C'(0) = \frac{M'(0)t'_{11}(M(0))}{\alpha_0} = \frac{q_{11}}{\alpha_0}. \quad (22)$$

Proof. Let $N_{ji}(t)$ denote the IBD indicators at locus \mathcal{L}_t , $i = 1, \dots, n$, $j = 0, 1$. Then,

$$\begin{aligned} C(t, t + s) &= E_0[SST(t; \nu)SST(t + s; \nu) | \phi] \\ &= \frac{\alpha_0/\alpha_1}{\sum_{i=1}^n (\pi_{1i} - \frac{\alpha_1}{\alpha_0}\pi_{0i})^2} \\ &\quad \times \sum_{i,i'=1}^n (\pi_{1i} - \frac{\alpha_1}{\alpha_0}\pi_{0i})(\pi_{1i'} - \frac{\alpha_1}{\alpha_0}\pi_{0i'}) E_0 \left[\left(N_{1i}(t) - \frac{\alpha_1}{\alpha_0}N_{0i}(t) \right) \left(N_{1i'}(t + s) - \frac{\alpha_1}{\alpha_0}N_{0i'}(t + s) \right) | \phi \right] \\ &= \frac{\alpha_0/\alpha_1}{\sum_{i=1}^n (\pi_{1i} - \frac{\alpha_1}{\alpha_0}\pi_{0i})^2} \sum_{i=1}^n (\pi_{1i} - \frac{\alpha_1}{\alpha_0}\pi_{0i})^2 \\ &\quad \times \left(\alpha_1 t_{11}(\theta_{t,t+s}) - \frac{\alpha_1}{\alpha_0} \alpha_1 t_{10}(\theta_{t,t+s}) - \frac{\alpha_1}{\alpha_0} \alpha_0 t_{01}(\theta_{t,t+s}) + \left(\frac{\alpha_1}{\alpha_0} \right)^2 \alpha_0 t_{00}(\theta_{t,t+s}) \right) \\ &= \frac{\alpha_0}{\alpha_1} \left(\alpha_1 t_{11}(\theta_{t,t+s}) - \frac{\alpha_1^2}{\alpha_0} (1 - t_{11}(\theta_{t,t+s})) - \alpha_1 (1 - t_{00}(\theta_{t,t+s})) + \frac{\alpha_1^2}{\alpha_0} t_{00}(\theta_{t,t+s}) \right) \\ &= t_{00}(\theta_{t,t+s}) + t_{11}(\theta_{t,t+s}) - 1 \\ &= \frac{t_{11}(\theta_{t,t+s}) - \alpha_1}{\alpha_0}, \end{aligned}$$

where we use the facts that IBD data from different relative pairs are conditionally independent given phenotypes, $\alpha_0 t_{01}(\theta_{t,t+s}) + \alpha_1 t_{11}(\theta_{t,t+s}) = \alpha_1$, and $t_{00}(\theta_{t,t+s}) = 1 - t_{01}(\theta_{t,t+s})$. \square

Note that the above results hold even when $T(\theta)$ does not satisfy the semi-group property. Next, we derive further properties of the auto-correlation function $C(s)$ by assuming that the 0/1 IBD configurations are defined as orbits of groups acting on the set of inheritance vectors.

Proposition 11 Auto-correlation function of linkage score statistic for unilineal relative pairs - semi-group.

Suppose the 0/1 IBD configurations are defined as orbits of groups acting on the set \mathcal{X} of inheritance vectors. Consider loci \mathcal{L}_t and \mathcal{L}_{t+s} located $|s|$ Morgans (M) apart on a given chromosome. Under the null hypothesis of no genes on the chromosome, the auto-correlation function for linkage score statistics computed at loci \mathcal{L}_t and \mathcal{L}_{t+s} is given by

$$C(t, t + s) = (1 - 2\theta_{t,t+s})^{-\lambda_2/2} = e^{(\lambda_2 d(\theta_{t,t+s}))}, \tag{23}$$

where λ_2 denotes the second largest eigenvalue of the infinitesimal generator Q (Propositions 1 – 4), $\theta_{t,t+s}$ denotes the recombination fraction between loci \mathcal{L}_t and \mathcal{L}_{t+s} , and $d(\theta) = -\ln(1 - 2\theta)/2$ is the inverse of the Haldane map function. In particular, under the no interference model, $\theta_{t,t+s} = M(|s|) = (1 - e^{-2|s|})/2$,

$$C(s) = e^{\lambda_2 |s|}, \tag{24}$$

and the first derivative of the auto-correlation function at $s = 0$, taken as the limit from the right, is given by

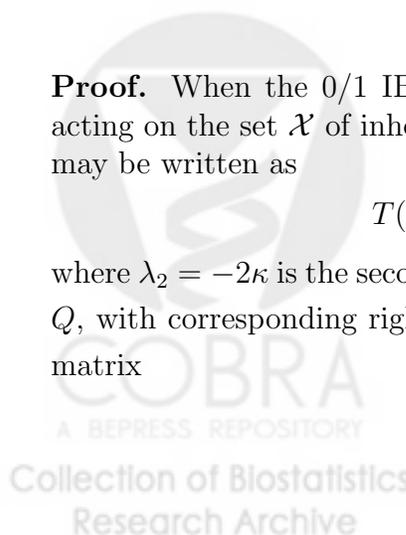
$$C'(0) = \lambda_2. \tag{25}$$

Proof. When the 0/1 IBD configurations are defined as orbits of groups acting on the set \mathcal{X} of inheritance vectors, then, by Propositions 1 – 4, $T(\theta)$ may be written as

$$T(\theta) = T\left(\frac{1}{2}\right) + (1 - 2\theta)^\kappa P_2,$$

where $\lambda_2 = -2\kappa$ is the second largest eigenvalue of the infinitesimal generator Q , with corresponding right eigenvector $v = \left(\sqrt{\frac{\alpha_1}{\alpha_0}}, -\sqrt{\frac{\alpha_0}{\alpha_1}}\right)$ and projection matrix

$$P_2 = \begin{bmatrix} \alpha_1 & -\alpha_1 \\ -\alpha_0 & \alpha_0 \end{bmatrix}.$$



Thus

$$t_{11}(\theta) = \alpha_1 + \alpha_0(1 - 2\theta)^\kappa$$

and

$$C(t, t + s) = (1 - 2\theta_{t,t+s})^\kappa.$$

□

Note that as the second largest eigenvalue λ_2 decreases, correlations between score statistics at linked loci are attenuated. The auto-correlation function is not affected by the multiplicity of the second largest eigenvalue.

In order to compute genome-wide significance levels, we may approximate the asymptotic distribution of the process $SST(s; \nu)$ by the distribution of a standard Ornstein-Uhlenbeck process with auto-correlation function $e^{(C'(0)|s|)} = e^{(q_{11}|s|/\alpha_0)}$ (Aldous, 1989; Feingold et al., 1993; Kruglyak and Lander, 1995b; Lander and Kruglyak, 1995; Leadbetter et al., 1983). The error in this approximation for the auto-correlation function is $o(|s|)$ as $s \rightarrow 0$. Under the null hypothesis of no gene influencing the phenotype, the genome-wide probability that $SST(s; \nu)$ exceeds the threshold T is approximated by

$$\begin{aligned} & pr_0(SST(s; \nu) \text{ exceeds } T \text{ somewhere in the genome} | \phi) \\ & \approx pr_0(SST(s; \nu) \geq T | \phi) \left(C - C'(0)GT^2 \right), \end{aligned} \quad (26)$$

where, for the human genome, $C = 23$ chromosome pairs and $G = \text{total genome length} \approx 33 \text{ M}$.

Expression (26) is based on three approximations: (i) an asymptotic approximation for the number n of relative pairs; (ii) an approximation of the crossover process by a Poisson process (no interference model); and (iii) a Markov approximation for the IBD configuration process. Although $T(\theta)$ does not in general satisfy the semi-group property, Feingold (1993) found this third approximation to be adequate for avuncular and cousin pairs. Relative pairs with skewed IBD distribution, i.e., small α_1 , require a larger sample size in order to apply this asymptotic result.

7.4 Linkage score statistic for unilineal relative pairs

7.4.1 Linkage score statistic for grand-parent-type relationship

From Proposition 5, the second largest eigenvalue of the infinitesimal generator Q for the augmented IBD configurations is $\lambda_2 = -2$. Thus, the linkage

score statistic is based on the *first derivative* of the log-likelihood. From Proposition 9 and derivations in Section 3.2, the score statistic for general grand-parent/grand-child pairs is given by

$$ST(\nu) = -2d \sum_i \left(\pi_{1i} - \frac{1}{2^d - 1} \pi_{0i} \right) \left(N_{1i} - \frac{1}{2^d - 1} N_{0i} \right).$$

Note that for this type of relationship, the null hypothesis of no linkage is rejected for *small* values of the score statistic.

In the special case of grand-parent/grand-child pairs ($d = 1$), the score statistic is

$$ST(\nu) = -2 \sum_i (\pi_{1i} - \pi_{0i})(N_{1i} - N_{0i})$$

and for great-grand-parent/great-grand-child pairs ($d = 2$) it is

$$ST(\nu) = -4 \sum_i (\pi_{1i} - \pi_{0i}/3)(N_{1i} - N_{0i}/3).$$

7.4.2 Linkage score statistic for half-sib-type relationship

From Proposition 6, one needs to distinguish between two main cases, half-sibs ($d = 2$) and distant half-sibs ($d > 2$).

Half-sibs ($d = 2$). For half-sibs, the second largest eigenvalue of the infinitesimal generator Q for the IBD configurations is $\lambda_2 = -4$. Thus, from results in Proposition 9 and Section 4.3, the linkage score statistic is based on the *second derivative* of the log-likelihood and is given by

$$ST(\nu) = 8 \sum_i (\pi_{1i} - \pi_{0i})(N_{1i} - N_{0i}).$$

In this case, the null hypothesis of no linkage is rejected for *large* values of the score statistic.

Distant half-sibs ($d > 2$). For more distant half-sibs ($d > 2$), the second largest eigenvalue of the infinitesimal generator Q for the augmented IBD configurations is $\lambda_2 = -2$. Thus, unlike the half-sib case with $d = 2$, results in Proposition 9 and Section 4.2 imply that the linkage

score statistic is based on the *first derivative* of the log-likelihood and is given by

$$ST(\nu) = -2(d-2) \sum_i \left(\pi_{1i} - \frac{1}{2^{d-1}-1} \pi_{0i} \right) \left(N_{1i} - \frac{1}{2^{d-1}-1} N_{0i} \right).$$

Note that in this case, the null hypothesis of no linkage is rejected for *small* values of the score statistic.

7.4.3 Linkage score statistic for avuncular relationship

From Proposition 7, one needs to distinguish between two main cases, avuncular pairs ($d = 5$) and distant avuncular pairs ($d > 5$).

Avuncular pairs ($d = 5$). For the simplest avuncular relationship ($d = 5$), the results of Section 5.1 show that the second largest eigenvalue of the infinitesimal generator Q for the augmented IBD configurations is $\lambda_2 = -4$. Thus, from results in Proposition 9 and Section 5.1, the linkage score statistic is based on the *second derivative* of the log-likelihood and is given by

$$ST(\nu) = 4 \sum_i (\pi_{1i} - \pi_{0i})(N_{1i} - N_{0i}).$$

In this case, the null hypothesis of no linkage is rejected for *large* values of the score statistic.

Distant avuncular pairs ($d > 5$). For more distant avuncular pairs ($d > 5$), the second largest eigenvalue of the infinitesimal generator Q for the augmented IBD configurations is $\lambda_2 = -2$. Thus, unlike the avuncular case with $d = 5$, results in Proposition 9 and Section 5.3 imply that the linkage score statistic is based on the *first derivative* of the log-likelihood and is given by

$$ST(\nu) = -2(d-5) \sum_i \left(\pi_{1i} - \frac{1}{2^{d-4}-1} \pi_{0i} \right) \left(N_{1i} - \frac{1}{2^{d-4}-1} N_{0i} \right).$$

Note that in this case, the null hypothesis of no linkage is rejected for *small* values of the score statistic.

In the simple avuncular case ($d = 5$), the transition matrix $T(\theta)$ for the usual 0/1 IBD configurations does not satisfy the semi-group property. Also, the auto-correlation function for the linkage score statistic does not have the exponential form of Proposition 11. Under the no interference model, the auto-correlation function is, for $s > 0$,

$$C(s) = \frac{1}{\alpha_0} \left(t_{11} \left(\frac{1}{2} (1 - e^{-2s}) \right) - \alpha_1 \right) = \frac{1}{2} e^{-4s} (1 + e^{-2s}).$$

7.4.4 Linkage score statistic for cousin-type relationship

From Proposition 8, one needs to distinguish between two main cases, first-cousins ($d = 6$) and distant cousin pairs ($d > 6$).

First-cousins ($d = 6$). For the simplest cousin relationship ($d = 6$), the results of Section 6.1 show that the second largest eigenvalue of the infinitesimal generator Q for the augmented IBD configurations is $\lambda_2 = -4$. Thus, from results in Proposition 9 and Section 6.1, the linkage score statistic is based on the *second derivative* of the log-likelihood and is given by

$$ST(\nu) = 12 \sum_i \left(\pi_{1i} - \frac{1}{3} \pi_{0i} \right) \left(N_{1i} - \frac{1}{3} N_{0i} \right).$$

In this case, the null hypothesis of no linkage is rejected for *large* values of the score statistic.

Distant cousin pairs ($d > 6$). For more distant cousin pairs ($d > 6$), the second largest eigenvalue of the infinitesimal generator Q for the augmented IBD configurations is $\lambda_2 = -2$. Thus, unlike the first-cousins case with $d = 6$, results in Proposition 9 and Section 6.3 imply that the linkage score statistic is based on the *first derivative* of the log-likelihood and is given by

$$ST(\nu) = -2(d - 6) \sum_i \left(\pi_{1i} - \frac{1}{2^{d-4} - 1} \pi_{0i} \right) \left(N_{1i} - \frac{1}{2^{d-4} - 1} N_{0i} \right).$$

Note that in this case, the null hypothesis of no linkage is rejected for *small* values of the score statistic.

8 Conclusions

In this article, we have derived theoretical properties of transition matrices for IBD configurations in four general classes of unilineal relative pairs obtained from the grand-parent/grand-child, half-sib, avuncular, and cousin relationships. The sibship case was studied in detail in Dudoit and Speed (1999) and Dudoit and Speed (2000). Instead of working directly with the usual 0/1 IBD configurations, general properties of the transitions matrices were obtained by considering augmented IBD configurations defined as orbits of groups acting on the set of inheritance vectors. In this setting, the transition matrix satisfies a semi-group property (Proposition 1, p. 5) and one can derive a spectral representation of the matrix in terms of the eigenvalues and eigenvectors of its infinitesimal generator Q (Proposition 2, p. 7). Properties of the eigenvalues of the infinitesimal generator were obtained by relating it to the adjacency matrix of a quotient graph (Proposition 4, p. 9). The second largest eigenvalue of the infinitesimal generator Q and its multiplicity are key in determining the form of the transition matrix and of likelihood-based linkage tests such as score tests. In general, if the second largest eigenvalue of Q is $\lambda_2 = -2\kappa$, then the transition matrix has the form

$$T(\theta) = T\left(\frac{1}{2}\right) + (1 - 2\theta)^\kappa P_2 + o((1 - 2\theta)^\kappa),$$

where $T(1/2)$ is the matrix with rows equal to the stationary distribution, α , and P_2 is the projection matrix for λ_2 , with rank the multiplicity of λ_2 . As shown in Proposition 9, the linkage score test is based on the κ th derivative of the log-likelihood evaluated at the null $\theta = 1/2$. The second largest eigenvalue not only determines the form of the linkage score statistic, but also the auto-correlation function between score statistics computed at different locations in the genome (Section 7.3). This property can be used to derive approximations for genome-wide significance levels of linkage score tests.

Table 1 summarizes the main properties of the transition matrices and linkage score statistics for the four different types of unilineal relative pairs. Relative pairs fall into two main categories, according to the second largest eigenvalue λ_2 of the infinitesimal generator Q for the augmented IBD configurations.

$\lambda_2 = -2$. The second largest eigenvalue of the infinitesimal generator Q is $\lambda_2 = -2$ for grand-parent/grand-child pairs (any d), distant half-sib

pairs ($d > 2$), distant avuncular pairs ($d > 5$), and distant cousin pairs ($d > 6$). In this case, the rate of convergence of the transition matrix $T(\theta)$ to the stationary distribution is $O(1 - 2\theta)$ as $\theta \rightarrow 1/2$ and the linkage score test is based on the *first derivative* of the log-likelihood. Linkage score tests for relative pairs with $\lambda_2 = -2$ are infinitely more powerful for local alternatives than score tests for pairs with $\lambda_2 = -4$.

$\lambda_2 = -4$. The second largest eigenvalue of the infinitesimal generator is $\lambda_2 = -4$ for half-sibs ($d = 2$), avuncular pairs ($d = 5$), and first-cousins ($d = 6$). In this case, the rate of convergence of the transition matrix $T(\theta)$ to the stationary distribution is $O(1 - 2\theta)^2$ as $\theta \rightarrow 1/2$ and the linkage score test is based on the *second derivative* of the log-likelihood. Linkage score tests based on these types of relative pairs are less efficient for local alternatives, the first derivative of the power function at the null being zero.

Linkage score tests as described in Section 7 provide a unified likelihood-based framework for the genetic mapping of complex human traits, qualitative and quantitative, using IBD data from small pedigrees. A more detailed discussion of such tests is given in Dudoit and Speed (1999) in the context of sibships. Dudoit and Speed (2000), Goldstein et al. (2000), and Goldstein et al. (2001) performed extensive simulation studies to investigate power and robustness properties of linkage score statistics. These studies showed that the linkage score test for quantitative traits had good power and robustness properties compared to alternative genetic mapping methods based on IBD data from unilineal relative pairs and sib-pairs.

The present article demonstrated that properties of IBD configuration transition matrices are important in terms of understanding the behavior of linkage test statistics. These matrices are relevant more broadly for studying patterns of transmission of DNA in families. Note that the general properties described in Section 2 hold for any type of pedigree, so long as IBD configurations are defined as orbits of groups acting on the set of inheritance vectors. The linkage score test approach may be extended to other types of relative pairs (e.g., double first-cousins) and small pedigrees (e.g., two sibs and a cousin) by defining suitable IBD configurations and deriving the transition matrices for these IBD configurations.

References

- G. R. Abecasis, S. S. Cherny, W. O. Cookson, and L. R. Cardon. Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. *Nature Genetics*, 30:97–101, 2002.
- D. Aldous. *Probability approximations via the Poisson clumping heuristic*. Springer-Verlag, New York, 1989.
- M. A. Campbell and R. C. Elston. Relatives of probands: models for preliminary genetic analysis. *Ann. Hum. Genet., Lond.*, 35:225–236, 1971.
- D. R. Cox and D. V. Hinkley. *Theoretical Statistics*. Chapman and Hall, 1974.
- K. P. Donnelly. The probability that related individuals share some section of genome identical by descent. *Theor. Pop. Biol.*, 23:34–63, 1983.
- S. Dudoit. *Linkage analysis of complex human traits using identity by descent data*. Ph.D. thesis, Department of Statistics, University of California, Berkeley, 1999.
- S. Dudoit and T. P. Speed. A score test for linkage using identity by descent data from sibships. *Annals of Statistics*, 27(3):943–986, 1999.
- S. Dudoit and T. P. Speed. A score test for the linkage analysis of qualitative and quantitative traits based on identity by descent data on sib-pairs. *Biostatistics*, 1(1):1–26, 2000.
- R. C. Elston and H. J. Cordell. Overview of model-free methods for linkage analysis. *Advances in Genetics*, 42:135–150, 2001.
- E. Feingold. Markov processes for modeling and analyzing a new genetic mapping method. *J. Appl. Probability*, 30:766–779, 1993.
- E. Feingold, P. O. Brown, and D. Siegmund. Gaussian models for genetic linkage analysis using complete high-resolution maps of identity by descent. *Am. J. Hum. Genet.*, 53:234–251, 1993.
- C. D. Godsil. *Algebraic Combinatorics*. Chapman & Hall mathematics. Chapman & Hall, New York, 1993.

- D. R. Goldstein, S. Dudoit, and T. P. Speed. Power of a score test for quantitative trait linkage analysis of relative pairs. *Genetic Epidemiology*, 19(Suppl. 1):S85–S91, 2000.
- D. R. Goldstein, S. Dudoit, and T. P. Speed. Power and robustness of a score test for linkage analysis of quantitative traits using identity by descent data on sib pairs. *Genetic Epidemiology*, 20(4):415–431, 2001.
- J. K. Haseman and R. C. Elston. The investigation of linkage between a quantitative trait and a marker locus. *Behavior Genetics*, 2:3–19, 1972.
- L. Kruglyak, M. J. Daly, M. P. Reeve-Daly, and E. S. Lander. Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am. J. Hum. Genet.*, 58:1347–1363, 1996.
- L. Kruglyak and E. S. Lander. Complete multipoint sib-pair analysis of qualitative and quantitative traits. *Am. J. Hum. Genet.*, 57:439–454, 1995a.
- L. Kruglyak and E. S. Lander. Nonparametric approach for mapping quantitative trait loci. *Genetics*, 139:1421–1428, 1995b.
- E. S. Lander and P. Green. Construction of multilocus genetic maps in humans. *Proc. Natl. Acad. Sci.*, 84:2363–2367, 1987.
- E. S. Lander and L. Kruglyak. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature Genetics*, 11: 241–247, 1995.
- M. R. Leadbetter, G. Lindgren, and H. Rootzen. *Extremes and related properties of random sequences and processes*. Springer series in statistics. Springer-Verlag, New York, 1983.
- C. R. Rao. *Linear Statistical Inference and Its Applications*. John Wiley & Sons, 2nd edition, 1973.
- M. Rosenblatt. *Random Processes*. Number 17 in Graduate texts in mathematics. Springer-Verlag, New York, 2nd edition, 1974.
- M. C. Shih and A. S. Whittemore. Allele-sharing among affected relatives: non-parametric methods for identifying genes. *Statistical Methods in Medical Research*, 10(1):27–55, 2001.

- T. P. Speed. What is a genetic map function? In T. P. Speed and M. S. Waterman, editors, *Genetic Mapping and DNA Sequencing*, volume 81 of *IMA Volumes in Mathematics and its Applications*. Springer-Verlag, New York, 1996.
- E. A. Thompson. *Pedigree Analysis in Human Genetics*. The Johns Hopkins series in contemporary medicine and public health. Johns Hopkins University Press, Baltimore, 1986.
- E. A. Thompson. Conditional gene identity in affected individuals. In I-H. Pawlowitzki, J. H. Edwards, and E. A. Thompson, editors, *Genetic Mapping of Disease Genes*, chapter 10, pages 137–146. Academic Press Inc., San Diego, London, 1997.
- J. H. van Lint and R. M. Wilson. *A Course in Combinatorics*. Cambridge University Press, Cambridge, New York, 1992.



Table 1: *Summary of properties of transition matrices and linkage score statistics for unilineal relative pairs.* d is the number of relevant meioses; the column “semi-group” indicates whether or not the 2×2 transition matrix $T(\theta)$ for the usual 0/1 IBD configurations satisfies the semi-group property of Proposition 1; $\lambda_2 = -2\kappa$ is the second largest eigenvalue of the infinitesimal generator Q for the augmented IBD configurations (subscript indicates its multiplicity); for the usual 0/1 IBD configurations, $\alpha = (\alpha_0, \alpha_1)$ are the unconditional (null) IBD probabilities; $U = (u_{ij})_{\{i,j=0,1\}} = T^\kappa(1/2)$, $Q = (q_{ij})_{\{i,j=0,1\}} = T'(0)$; $C(s)$ is the auto-correlation function for score statistics computed at loci s Morgans apart.

Type of pair	d	Semi-group	λ_2	α_1	q_{11}	u_{11}	$C'(0)$
Grand-parent/grand-child	1	YES	-2_1	$\frac{1}{2}$	-1	-1	-2
	> 1	NO	-2_1	$\frac{1}{2^d}$	$-d$	$-\frac{d}{2^{d-1}}$	$-\frac{d2^d}{2^{d-1}}$
Half-sibs	2	YES	-4_1	$\frac{1}{2}$	-2	4	-4
	> 2	NO	-2_1	$\frac{1}{2^{d-1}}$	$-d$	$-\frac{d-2}{2^{d-2}}$	$-\frac{d2^{d-1}}{2^{d-1}-1}$
Avuncular	5	NO	-4_1	$\frac{1}{2}$	$-\frac{5}{2}$	2	-5
	> 5	NO	-2_1	$\frac{1}{2^{d-4}}$	$\frac{5-2d}{2}$	$-\frac{d-5}{2^{d-5}}$	$\frac{(5-2d)2^{d-5}}{2^{d-4}-1}$
Cousins	6	NO	-4_2	$\frac{1}{4}$	-4	3	$-\frac{16}{3}$
	> 6	NO	-2_1	$\frac{1}{2^{d-4}}$	$2-d$	$-\frac{d-6}{2^{d-5}}$	$\frac{(2-d)2^{d-4}}{2^{d-4}-1}$
			$\alpha_1 u_{11} = \alpha_0 u_{00}, u_{01} = -u_{00}, u_{10} = -u_{11}$		$\alpha_1 q_{11} = \alpha_0 q_{00}, q_{01} = -q_{00}, q_{10} = -q_{11}$		
			$ST(\nu) = \frac{u_{11}}{\alpha_1} \sum_i (\pi_{1i} - \frac{\alpha_1}{\alpha_0} \pi_{0i})(N_{1i} - \frac{\alpha_1}{\alpha_0} N_{0i})$		$C'(0) = \frac{q_{11}}{\alpha_0}$		