JOURNAL OF COMPUTATIONAL BIOLOGY
Volume 8, Number 2, 2001
Mary Ann Liebert, Inc.
Pp. 191–200

On Relationship Inference Using Gamete Identity by Descent Data

HONGYU ZHAO[1] and FENG LIANG[2]

ABSTRACT

Related individuals are identical by descent (IBD) at a genetic locus if they share the same DNA material from a common ancestor. Continuous gamete IBD data consist of the lengths of (in order) IBD and non-IBD regions along the genomes for gametes segregating from two related individuals and can be used to distinguish different relationships. Under the assumption that the crossovers follow a Poisson process, we show that the exact calculation of the likelihood of a particular relationship for a given gamete IBD datum is tractable. Greatgrandparent–greatgrandchild and cousin relationships are used as examples to illustrate our methods.

Key words: identity by descent, gamete, Markov process, relationship inference, genomic data.

INTRODUCTION

RELATED INDIVIDUALS ARE IDENTICAL BY DESCENT (IBD) at a genetic locus if they share the same DNA material from a common ancestor. The IBD data provides essential information to map disease genes (Risch, 1990) and to infer relationships among individuals (Boehnke and Cox, 1997) on the basis of genetic marker data. With the rapid advances of the Human Genome Project, it has become possible to obtain IBD data at very densely spaced markers, and new techniques, such as genomic mismatch scanning (Nelson *et al.*, 1993), will potentially yield continuous IBD data.

In this article, we consider continuous gamete IBD data consisting of the lengths of (in order) IBD and non-IBD regions along the genomes for gametes segregating from two related individuals. Based on continuous gamete IBD data, Browning (1998) proposed a Monte Carlo method to estimate the likelihood of a particular relationship. The likelihood ratio test statistic estimated from simulations can be used to distinguish between relationships. Although the proposed simulation methods are easy to implement, the computing time needed in such simulations can be considerable. For example, likelihood estimation on 100,000 Morgans of data took several days of computing time on a DEC Alpha Unix workstation (Browning, 1998). In the Methods section of this article, we present a closed-form formula for the exact calculation of the likelihood for gamete IBD data for a given relationship. Compared to the simulation method, this formula can be used to calculate the likelihood both accurately and efficiently. We then

[1]Department of Epidemiology and Public Health, Yale University School of Medicine, New Haven, CT 06520-8034.
[2]Department of Statistics, Yale University, New Haven, CT 06520-8290.

discuss in detail in the Applications section the calculations involved to distinguish greatgrandparent–greatgrandchild and cousin relationships through continuous gamete IBD data.

METHODS

Assumptions and notation

During meiosis, homologous chromosomes pair and duplicate. Crossovers then occur between nonsister chromatids, with each resulting gamete usually consisting of both paternal and maternal chromosomal materials. At a given locus x along the genome in a gamete, we define the state $v(x)$ at x as $v(x) = 0$ if the DNA has a maternal origin and as $v(x) = 1$ if the DNA has a paternal origin. Throughout this article, we assume that the crossovers on the four-strand bundle during meiosis follow a Poisson process and that there is no chromatid interference. Therefore, if we follow the maternal and paternal origin along a chromosome, the stochastic process $v(x)$ along a gamete for a chromosome is a time-homogeneous two-state Markov process with the following intensity matrix:

$$\begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}.$$

The intensity is 1 because the distance along the gametes is measured in Morgans. The use of the Poisson process model assumes no crossover interference during meiosis. Although crossover interference has been observed in many organisms (Zhao *et al.*, 1995), the Poisson process can serve as both a first approximation and as a starting model for general models incorporating crossover interference (Lange *et al.*, 1997).

Two gametes generated from two related individuals are separated by a certain number of meioses. For example, the gametes produced from a mother and her child are separated by three meioses (see discussion on the mother–child relationship in the following subsection). If two gametes are separated by k meioses,

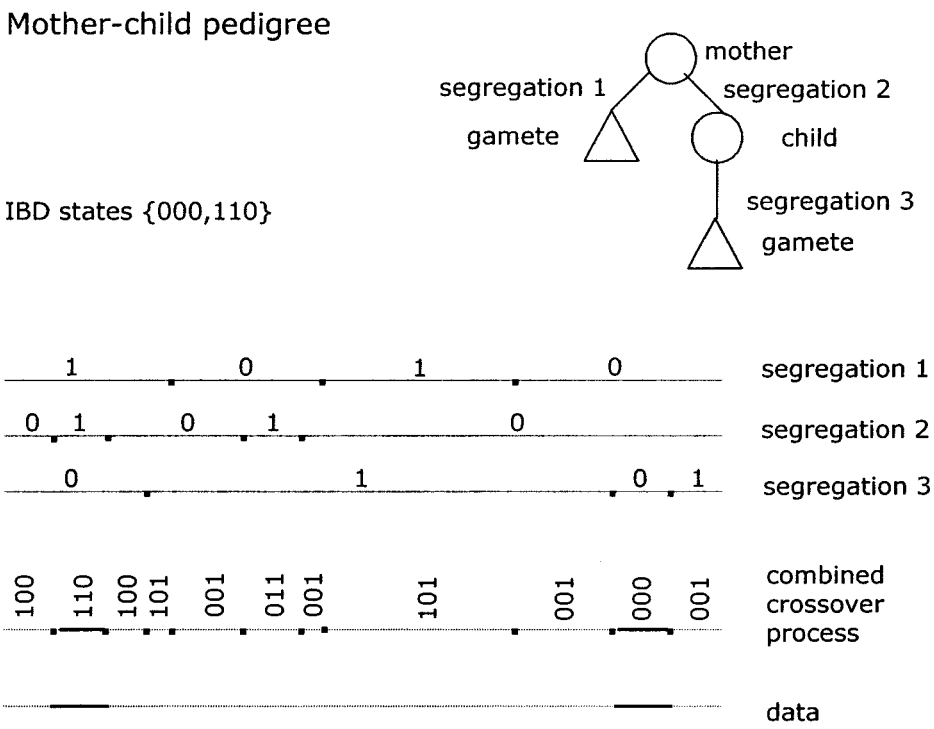

FIG. 1. The IBD process for a mother–child relationship. This graph is based on Fig. 1 in Browning (1998). A realization of the combined crossover process along a chromosome is shown with crossovers represented by points. The observed data consist of IBD (solid line) and non-IBD (dotted line) regions.

we consider all of the gametes generated from these meioses together. At a genetic locus x, we define the joint states of all k gametes by a vector $\mathbf{v}(x) = \{v_1(x), v_2(x), \ldots, v_k(x)\}$ of length k, where $v_i(x)$ describes the gamete from the ith meiosis: $v_i(x) = 0$ if the ith gamete has a maternal origin, and $v_i(x) = 1$ if the ith gamete has a paternal origin at x. There are a total of 2^k possible joint states. The crossover processes during these meioses are independent as we move along the chromosome, so the joint states $\mathbf{v}(x)$ form a 2^k-state Markov process on the hypercube $H = \{0, 1\}^k$. However, for continuous gamete IBD data, not all of these 2^k states are observable. In fact, we can only observe whether the two gametes are IBD or not IBD at a locus x. Therefore, we need to relate the observed IBD and non-IBD data to the unobservable joint states data $\mathbf{v}(x)$ to calculate the likelihood of continuous gamete IBD data for a relationship.

In this section, we first discuss the exact calculation of the likelihood for the mother–child relationship. We then present the general closed-form expression for the likelihood of any relationship. The calculations are then simplified using results obtained by Donnelly (1983), who demonstrated that the 2^k joint states for two gametes separated by k meiosis can be reduced from the symmetry among these states.

Mother–child relationship

The simplest relationship for two individuals is the mother–child relationship. The corresponding gametes are separated by three meioses. This relationship is plotted in Fig. 1 following Browning (1998). There are eight joint states, with joint states (000) and (110) corresponding to IBD and the other six joint states corresponding to non-IBD. We reorder the eight states such that the first two states are the two IBD states, (000) and (110), followed by the six non-IBD states, (001), (010), (011), (100), (101), and (111). Denote the initial distribution of the joint states by $\boldsymbol{\pi} = \{\pi_i\} = \{1/8, 1/8, \ldots, 1/8\}$. The intensity matrix for the *unobservable* joint states $\mathbf{v}(x)$ is:

$$\mathbf{T} = \begin{pmatrix} -3 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & -3 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & -3 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & -3 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & -3 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & -3 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & -3 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & -3 \end{pmatrix}.$$

We partition the intensity matrix $\mathbf{T}$ into four submatrices

$$\mathbf{T} = \begin{pmatrix} \mathbf{T}_{00} & \mathbf{T}_{01} \\ \mathbf{T}_{10} & \mathbf{T}_{11} \end{pmatrix}, \tag{1}$$

where

$$\mathbf{T}_{00} = \begin{pmatrix} -3 & 0 \\ 0 & -3 \end{pmatrix}, \qquad \mathbf{T}_{01} = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix},$$

$$\mathbf{T}_{10} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 0 \\ 1 & 1 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \mathbf{T}_{11} = \begin{pmatrix} -3 & 0 & 1 & 0 & 1 & 0 \\ 0 & -3 & 1 & 0 & 0 & 0 \\ 1 & 1 & -3 & 0 & 0 & 1 \\ 0 & 0 & 0 & -3 & 1 & 0 \\ 1 & 0 & 0 & 1 & -3 & 1 \\ 0 & 0 & 1 & 0 & 1 & -3 \end{pmatrix}.$$

The submatrix $\mathbf{T}_{00}$ represents transition among the two IBD states, the submatrix $\mathbf{T}_{01}$ represents transitions from the two IBD states to the six non-IBD states, the submatrix $\mathbf{T}_{10}$ represents transitions from the six non-IBD states to the two IBD states, and the submatrix $\mathbf{T}_{11}$ represents transitions among the six non-IBD states. The initial distribution $\boldsymbol{\pi}$ is also partitioned into two components $\{\boldsymbol{\pi}_0, \boldsymbol{\pi}_1\}$, where $\boldsymbol{\pi}_0 = \{1/8, 1/8\}$ and $\boldsymbol{\pi}_1 = \{1/8, 1/8, 1/8, 1/8, 1/8, 1/8\}$. We number these eight states from 1 to 8 and denote the state at x by $s(x)$.

Define

$$\mathbf{P}_{00}(d) = e^{\mathbf{T}_{00}d},$$

and the (i, j)th entry in this matrix is

$$P\{s(d) = j \text{ and IBD from 0 to } d \mid s(0) = i\}.$$

Similarly, the (i, j)th entry in $\mathbf{P}_{11}(d) = e^{\mathbf{T}_{11}d}$ is

$$P\{s(d) = j \text{ and non-IBD from 0 to } d \mid s(0) = i\}.$$

It is straightforward to calculate the likelihood if the two gametes are all IBD or all non-IBD. If the observed data is $D = \{\text{two gametes of length } d \text{ are all IBD}\}$,

$$P\{D\} = \boldsymbol{\pi}_0 e^{\mathbf{T}_{00}d}\mathbf{1}_0,$$

where $\mathbf{1}_0 = (1, 1)^T$. This is so because

$$\begin{aligned}
&P\{\text{all IBD}\}\\
&= \sum_{i=1}^{2}\sum_{j=1}^{2} P\{s(0) = i, s(d) = j, \text{ and IBD from 0 to } d\}\\
&= \sum_{i=1}^{2}\sum_{j=1}^{2} P\{s(d) = j \text{ and IBD from 0 to } d \mid s(0) = i\}\\
&\quad \times P\{\text{initial state is } s(0)\}\\
&= \boldsymbol{\pi}_0 e^{\mathbf{T}_{00}d}\mathbf{1}_0.
\end{aligned}$$

Similarly, if the observed data is $D = \{\text{two gametes of length } d \text{ share nothing IBD}\}$,

$$P\{D\} = \boldsymbol{\pi}_1 e^{\mathbf{T}_{11}d}\mathbf{1}_1,$$

where $\mathbf{1}_1 = (1, 1, 1, 1, 1, 1)^T$.

Now consider the case of there being both IBD and non-IBD regions. If the two gametes are IBD from 0 to u and non-IBD from u to d, the likelihood for D is

$$f\{D\} = \boldsymbol{\pi}_0 e^{\mathbf{T}_{00}u}\mathbf{T}_{01}e^{\mathbf{T}_{11}(d-u)}\mathbf{1}_1. \tag{2}$$

This is so because

$$\begin{aligned}
&f\{D\} \times \delta\\
&= P\{\text{IBD from 0 to } u, \text{ changes from IBD to non-IBD on } (u, u+\delta),\\
&\quad \text{non-IBD from } u+\delta \text{ to } d\}\\
&= \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=3}^{8}\sum_{l=3}^{8} P\{\text{IBD from 0 to } u, s(u) = j, s(u+\delta) = k,\\
&\quad \text{non-IBD from } s(u+\delta) \text{ to } s(d), s(d) \text{ is } l \mid s(0) \text{ is } i\}\\
&\quad \times P\{s(0) = i\}\\
&= \boldsymbol{\pi}_0 e^{\mathbf{T}_{00}u}\mathbf{T}_{01}e^{\mathbf{T}_{11}(d-u)}\mathbf{1}_1 \times \delta + o(\delta).
\end{aligned}$$

General results

Having discussed the above cases for the mother–child relationship, we now present the general result for the likelihood for an arbitrary relationship separated by k meioses. There are a total of 2^k joint states. We order these states so that the first a states are the IBD states and the last $b = 2^k - a$ states are the non-IBD states.

Let $\mathbf{T}$ denote the intensity matrix and partition it as in (1). Then the $a \times a$ submatrix $\mathbf{T}_{00}$ represents transitions among the IBD states, the $a \times b$ submatrix $\mathbf{T}_{01}$ represents transitions from the IBD states to the non-IBD states, the $b \times a$ submatrix $\mathbf{T}_{10}$ represents transitions from the non-IBD states to the IBD states, and the $b \times b$ submatrix $\mathbf{T}_{11}$ represents transitions among the non-IBD states. The initial distribution $\boldsymbol{\pi}$ for all 2^k states is similarly partitioned into $\boldsymbol{\pi} = \{\boldsymbol{\pi}_0, \boldsymbol{\pi}_1\}$.

Denote the continuous gamete IBD data by $D = (u_l, i_l, l = 1, \ldots, m)$, where m is the number of IBD and non-IBD segments, u_l is the length of the lth IBD or non-IBD segment, $i_l = 0$ if the lth segment is IBD, and $i_l = 1$ if the lth segment is non-IBD. The likelihood for the observed data D on an arbitrary relationship is

$$f(D) = \boldsymbol{\pi}_{i_1} \left\{ \prod_{l=1}^{m-1} [\exp\{\mathbf{T}_{i_l i_l} u_l\} \mathbf{T}_{i_l i_{l+1}}] \right\} \exp\{\mathbf{T}_{i_m i_m} u_m\} \mathbf{1}_{i_m}, \tag{3}$$

where $\mathbf{1}_0$ is a vector of 1's of length a and $\mathbf{1}_1$ is a vector of 1's of length b. This result follows from the same reasoning that leads to the likelihood expression in (2).

Symmetry among the joint states

Although the expression in (3) can be applied to derive the exact likelihood for an arbitrary relationship, large matrices are involved for relatives separated by many segregations. For example, a greatgrandparent and greatgrandchild are separated by five meioses and two first cousins are separated by eight meioses. Therefore, the intensity matrix has size $32 \times 32 (2^5)$ and $256 \times 256 (2^8)$ if all possible 2^k joint states $\mathbf{v}(x)$ are treated separately. Although the complete model used to describe all the meioses $\mathbf{v}(x)$ is a Markov process on the hypercube vertices, we can only observe two states at a given locus for two individuals, IBD or non-IBD. Making use of the symmetry among the 2^k states, Donnelly (1983) showed that, when only IBD and non-IBD states are of interest, the 2^k states can be reduced to sets called "orbits," and only the orbit in which the Markov process lies is needed to characterize the IBD process. Donnelly (1983) derived the orbits for several types of relationships. The use of these orbits instead of the original joint states can reduce the dimension of our model, and the reduction is considerable for some relationships. As an application of his results, Donnelly (1983) calculated the probability that two related individuals share some section of genome IBD. We note here that Donnelly (1983) assumed that IBD can be detected on the diploid data between two related individuals, where we consider IBD between two gametes generated from these two individuals. Therefore, for each relationship, two further meioses are involved for gamete data. For the mother–child relationship, only one meiosis is needed in Donnelly's setting, whereas three meioses are needed here for gamete IBD data.

If, at each locus, we replace the joint state $\mathbf{v}(x)$ by its corresponding orbit $o(x)$, the process of orbits also follows a Markov process. In the case of a mother–child relationship, the eight joint states belong to four orbits, $A0 = \{000, 110\}$, $B0 = \{010, 100\}$, $A1 = \{001, 111\}$, and $B1 = \{011, 101\}$. Because gamete IBD data are analyzed here, the mother–child relationship is actually the half-sib-type relationship studied by Donnelly. Orbit $A0$ includes the two IBD states, while orbits $B0$, $A1$, and $B1$ each consists of two of the remaining six joint states. The intensity matrix for the orbits has the following form:

$$\begin{pmatrix} -3 & 2 & 1 & 0 \\ 2 & -3 & 0 & 1 \\ 1 & 0 & -3 & 2 \\ 0 & 1 & 2 & -3 \end{pmatrix}.$$

The initial distribution for the four orbits is $\{1/4, 1/4, 1/4, 1/4\}$. Therefore, instead of working on the eight vertices of the hypercube $\{0, 1\}^3$ for the mother–child relationship to calculate the exact likelihood,

we use symmetry among the states to allow us to work on the reduced four-state Markov process. For an arbitrary relationship, we can still use the general result in (3) to calculate the likelihood by replacing the original joint states by the orbits in the formula.

APPLICATIONS

In this section, we apply the formula described in the previous section to distinguish the two specific relationships studied by Browning (1998): the greatgrandparent–greatgrandchild relationship and the first cousin relationship.

Greatgrandparent–greatgrandchild relationship

When gamete IBD data are studied, the greatgrandparent–greatgrandchild relationship corresponds to the half-sib-type relationship in Donnelly (1983). The two gametes are separated by five meioses (Fig. 2). Among the $2^5 = 32$ possible joint states on $\{0, 1\}^5$, there are eight orbits. Donnelly (1983) gave a detailed derivation of these eight orbits. The first orbit consists of the two IBD states, (00000) and (11000), denoted by $A0$. The seven other orbits correspond to non-IBD states. The second orbit consists of two non-IBD states, (01000) and (10000), denoted by $B0$. Three orbits $A1$, $A2$, and $A3$ differ from orbit $A0$ at the last three components, with Ai having a total number of i 1's in the last three components. For example, orbit $A1$ consists of six joint states (00001), (11001), (00010), (11010), (00100), and (11100). Similarly, three orbits $B1$, $B2$, and $B3$ differ from orbit $B0$ at the last three components, with Bi having a total number of i 1's in the last three components. Define

$$\mathbf{M} = \begin{pmatrix} 0 & 2 \\ 2 & 0 \end{pmatrix}, \quad \mathbf{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \text{and} \quad \mathbf{0} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Then the 8×8 intensity matrix among the eight orbits has the following form:

$$\mathbf{T} = \begin{pmatrix} \mathbf{M} - 5\mathbf{I} & 3\mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{I} & \mathbf{M} - 5\mathbf{I} & 2\mathbf{I} & \mathbf{0} \\ \mathbf{0} & 2\mathbf{I} & \mathbf{M} - 5\mathbf{I} & \mathbf{I} \\ \mathbf{0} & \mathbf{0} & 3\mathbf{I} & \mathbf{M} - 5\mathbf{I} \end{pmatrix}. \tag{4}$$

The initial distribution on the eight orbits is $\pi = \{1/8, 1/8, 1/8, 1/8, 1/8, 1/8, 1/8, 1/8\}$.

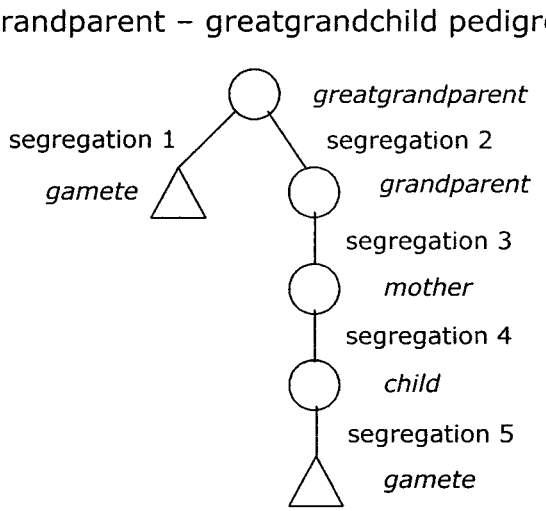

FIG. 2. A greatgrandparent–greatgrandchild relationship. The five meioses that separate the gametes from these two individuals are numbered from 1 to 5. All transmissions are assumed to be through females in this pedigree.

First cousin relationship

When gamete IBD data are studied, the two gametes from two first cousins are separated by eight meioses (Fig. 3). Among the $2^8 = 256$ possible joint states $\mathbf{v}(x) = \{v_1(x), \ldots, v_8(x)\}$ on $\{0, 1\}^8$, we find 21 orbits following the arguments in Donnelly (1983). It is easy to see that only when $v_7(x) = v_8(x) = 0$ can the two gametes be IBD at x. Furthermore, the first six components have to be one of the following forms to make two gametes IBD at x: 00*00*, 01*01*, 1*01*0, and 1*11*1, where * can be either 0 or 1. When the last two components are 0, $v_7(x) = v_8(x) = 0$, there are seven orbits. Before describing these seven orbits, we define the following four conditions:

(i) $[v_1(x) = v_4(x) = 0$ and $v_2(x) = v_5(x)]$ or $[v_1(x) = v_4(x) = 1$ and $v_3(x) = v_6(x)]$,
(ii) $[v_1(x) = v_4(x) = 0$ and $v_3(x) = v_6(x)]$ or $[v_1(x) = v_4(x) = 1$ and $v_2(x) = v_5(x)]$,
(iii) $v_2(x) = v_5(x)$,
(iv) $v_3(x) = v_6(x)$.

Orbit $A0$ consists of those states with both conditions (i) and (ii) being true: (00000000), (00100100), (01001000), (01101100), (10010000), (11011000), (10110100), and (11111100). Orbit $B0$ consists of those states with condition (i) being true but condition (ii) being false: (00000100), (00100000), (01001100), (01101000), (10011000), (11010000), (10111100), and (11110100). The 16 joint states on these two orbits are the *only* IBD states. The other five orbits when $v_7(x) = v_8(x) = 0$ are: orbit $C0$ consisting of states with (i) being false but (ii) being true; orbit $D0$ consisting of states with $v(1) = v(4)$ but both (i) and (ii) being false; orbit $E0$ consisting of states with $v(1) \neq v(4)$ and both (iii) and (iv) being true; orbit $F0$ consisting of states with $v(1) \neq v(4)$, (iii) or (iv) begin true but not both, and orbit $G0$ consisting of states with $v(1) \neq v(4)$ and both (iii) and (iv) being false.

With orbit $A0$ defined above, the states in orbit $A1$ differ from the states in orbit $A0$ in that one of the last two components is 1; i.e., $v_7(x) = 1$ or $v_8(x) = 1$ but not both, and the states in orbit $A2$ differ from the states in orbit $A0$ in that both last two components are 1, $v_7(x) = v_8(x) = 1$. Orbits $B1$ to $G1$ and orbits $B2$ to $G2$ can be similarly defined. These are the 21 orbits for gamete IBD data from two first cousins. Therefore, instead of dealing with the original 256 states, we need to focus on only the 21 orbits in our likelihood calculations.

Define

$$\mathbf{M} = \begin{pmatrix} 0 & 2 & 2 & 0 & 2 & 0 & 0 \\ 2 & 0 & 0 & 2 & 0 & 2 & 0 \\ 2 & 0 & 0 & 2 & 0 & 2 & 0 \\ 0 & 2 & 2 & 0 & 0 & 0 & 2 \\ 2 & 0 & 0 & 0 & 0 & 4 & 0 \\ 0 & 1 & 1 & 0 & 2 & 0 & 2 \\ 0 & 0 & 0 & 2 & 0 & 4 & 0 \end{pmatrix},$$

$\mathbf{I}$ as the identity matrix, and $\mathbf{0}$ as the zero matrix. The intensity matrix for the transitions among the 21 orbits is:

$$\begin{pmatrix} \mathbf{M} - 8\mathbf{I} & 2\mathbf{I} & \mathbf{0} \\ \mathbf{I} & \mathbf{M} - 8\mathbf{I} & \mathbf{I} \\ \mathbf{0} & 2\mathbf{I} & \mathbf{M} - 8\mathbf{I} \end{pmatrix}. \tag{5}$$

The initial distribution for these 21 orbits is:

$$\{\boldsymbol{\alpha}/4, \boldsymbol{\alpha}/2, \boldsymbol{\alpha}/4\}, \quad \text{where } \boldsymbol{\alpha} = \{1/8, 1/8, 1/8, 1/8, 1/8, 2/8, 1/8\}.$$

Calculation of $e^{\mathbf{T}u}$

In our analysis, we used the eigenvalue method to calculate $e^{\mathbf{T}}$. For both relationships, we can decompose the matrices into the form $T = P\ \mathrm{diag}\{\lambda_1, \ldots, \lambda_n\}P^{-1}$. Then $e^{\mathbf{T}u} = P\ \mathrm{diag}\{e^{\lambda_1 u}, \ldots, e^{\lambda_n u}\}P^{-1}$.

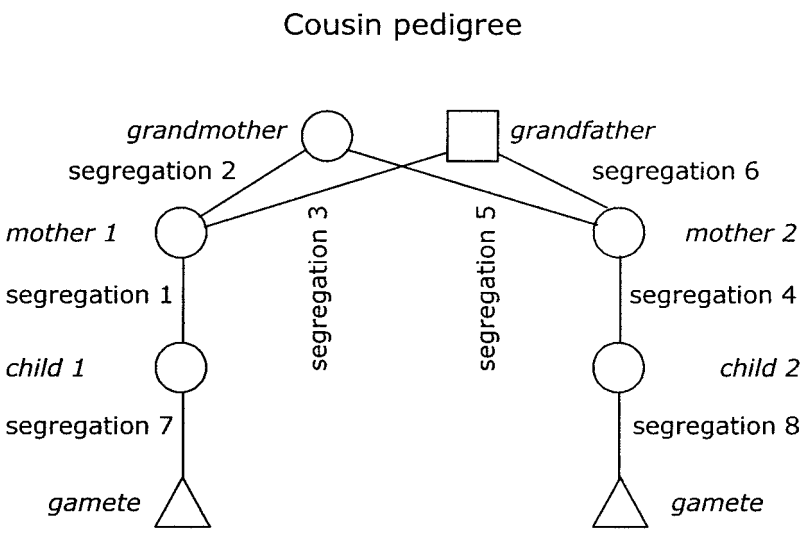

FIG. 3. A first cousin relationship. The eight meioses that separate the gametes from these two individuals are numbered from 1 to 8. The transmissions in the two parents and the two children are assumed to be through females in this pedigree.

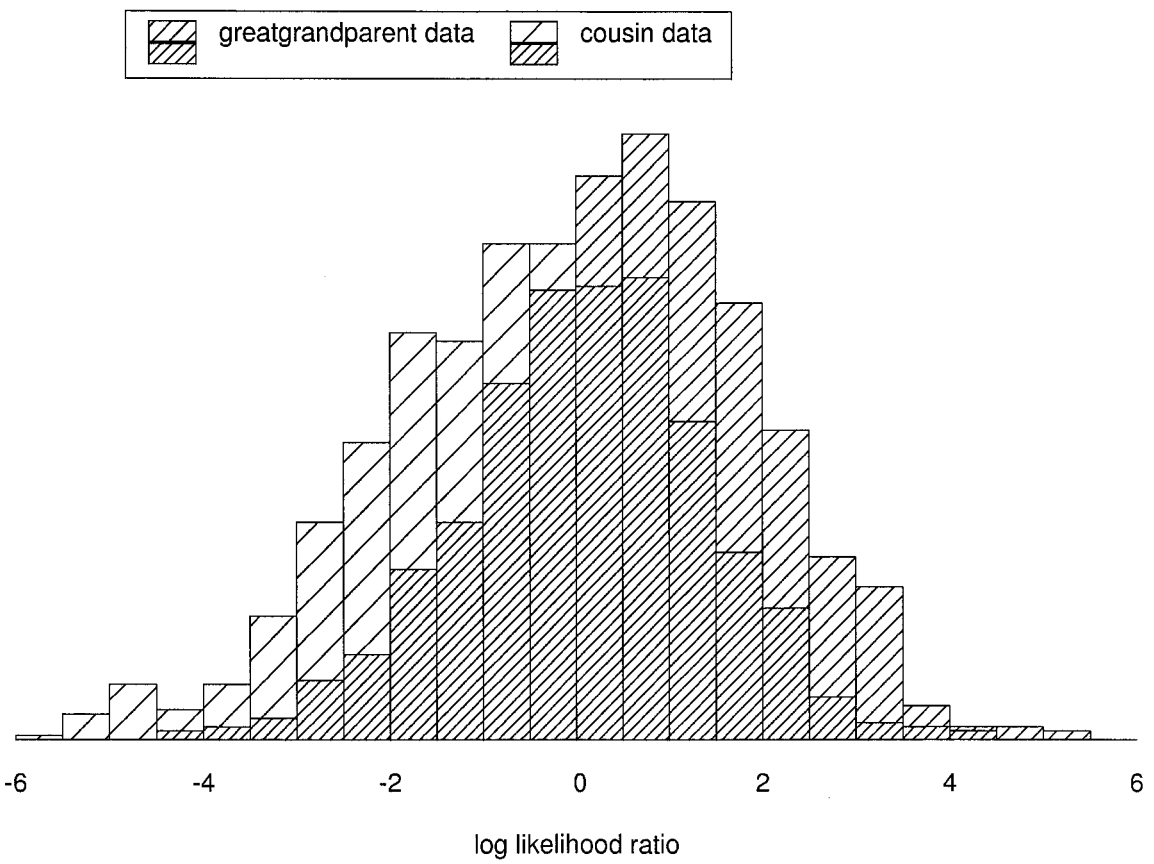

FIG. 4. Histograms of calculated likelihood ratio test statistics for the null hypothesis of the greatgrandparent-greatgrandchild relationship versus the alternative hypothesis of the first cousin relationship. One histogram is for the data generated from the greatgrandparent–greatgrandchild relationship, and the other histogram is for the data generated from the first cousin relationship. For each relationship, there are 1,000 data sets generated, with each data set consisting of 30 chromosomes of length 1 Morgan each.

Simulations

Similar to the simulation examples in Browning (1998), 1000 sets of continuous gamete IBD data from both greatgrandparent–greatgrandchild and first cousin relationships are generated. Each data set consists of gamete IBD data on 30 chromosomes of length 1 Morgan. The likelihoods are evaluated using the general result in (3) and the intensity matrix on the orbits in (4) for the greatgrandparent–greatgrandchild relationship and (5) for the first cousin relationship. The likelihood ratio test statistics are then calculated for the null hypothesis of the greatgrandparent–greatgrandchild relationship versus the alternative hypothesis of the first cousin relationship. The histograms are given in Fig. 4. This figure is very similar to the histograms presented in Browning (1998) through Monte Carlo estimates of the likelihoods. However, the exact method is both accurate and computationally much more efficient than the Monte Carlo simulation method. This is particularly advantageous if a large number of gamete IBD data are examined, as in the power study of using gamete IBD data to distinguish between different relationships.

DISCUSSION

In this article, a general formula is introduced to calculate the exact likelihood for any relationship using continuous gamete IBD data. We also show how the computation can be simplified by reducing the original joint states to a set of orbits through symmetry among the many joint states. Compared to the Monte Carlo method proposed by Browning (1998), our method is accurate, computationally efficient, and feasible for most relationships of common interest. For example, the total computation time for all of the cases considered in our simulations was only 63 seconds on a Pentium III 600 MHz desktop with 64MB memory.

With some extensions, our approach can be used to calculate the likelihood for diploid continuous IBD data studied by Donnelly (1983). Compared to discrete marker data (Boehnke and Cox, 1997), continuous IBD data provide more information for distinguishing different relationships and serve as an upper bound on information in discrete marker data.

We have assumed that the crossover process is a Poisson process, which implies no crossover interference. However, crossover interference has been observed in many organisms and stochastic models incorporating crossover interference, e.g., the chi-square model (Zhao *et al.*, 1995) and the Poisson-skip model (Lange *et al.*, 1997), have been found to provide better fit to data from a variety of organisms. We are currently investigating how to obtain the exact likelihoods when crossover interference models are assumed and will report our findings in future work.

ACKNOWLEDGMENTS

This work was supported in part by research grant FY98-0752 from the March of Dimes Birth Defects Foundation and by NIH grants HD36834 and GM59507.

REFERENCES

Boehnke, M., and Cox, N.J. 1997. Accurate inference of relationships in sib-pair linkage studies. *Am. J. Human Genetics* 61, 423–429.

Browning, S. 1998. Relationship information contained in gamete identity by descent data. *J. Comp. Biol.* 5, 323–334.

Donnelly, K.P. 1983. The probability that related individuals share some section of genome identical by descent. *Theoretical Population Biology* 23, 34–63.

Lange, K., Zhao, H., and Speed, T.P. 1997. The Poisson-skip model of crossing-over. *Annals of Applied Probability* 7, 299–313.

Risch, N. 1990. Linkage strategies for genetically complex traits. *I. Multilocus* models. *Am. J. Human Genetics* 46, 222–228.

Nelson, S.F., McCusker, J.H., Sander, M.A., Kee, Y., Modrich, P., and Brown, P.O. 1993. Genomic mismatch scanning: A new approach to genetic linkage mapping. *Nature Genetics* 4, 11–18.

Zhao, H., Speed, T.P., and McPeck, M.S. 1995. Statistical analysis of crossover interference using the chi-square model. *Genetics* 139, 1045–1056.

Address correspondence to:
Hongyu Zhao
Department of Epidemiology and Public Health
Yale University School of Medicine
New Haven, CT 06520-8034

E-mail: hongyu.zhao@yale.edu