Reprinted from THEORETICAL POPULATION BIOLOGY
All Rights Reserved by Academic Press, New York and London

Vol. 23, No. 1, February 1983
Printed in Belgium

The Probability that Related Individuals Share Some Section of Genome Identical by Descent

KEVIN P. DONNELLY[1]

Statistical Laboratory, Cambridge University, Cambridge, England

Received February 1981

A formal mathematical framework is presented for the study of linkage in man and the concept of chromosome pedigree is defined for both autosomes and X chromosomes. It is shown that, assuming no interference, all the crossover processes in the pedigree may be viewed jointly as a continuous-time Markov random walk on the vertices of a hypercube, the time parameter being map distance along the chromosome. The event that two individuals have a segment of chromosome in common, thus proving them to be related, corresponds to the random walk hitting a particular set of vertices. The probability of this happening is calculated for various types of relationship, making use of the symmetry of the situation to partition the vertices into a very much smaller number of orbits and render the computations manageable. The probability that an individual with n children passes on all his or her genes to them is also calculated in this way.

INTRODUCTION

This paper approaches a limited problem in the field of human pedigree reconstruction from the new viewpoint of complete genetic information being available for all 22 pairs of autosomes. It is hoped, however, that the methods developed may be of use in other genetic problems concerning the complete genome.

Edwards (1965) first suggested the possibility of using genetic information on living individuals in a small population to infer an unknown pedigree structure. Thompson (1974a, 1975, 1976) has done both theoretical and computational work on this problem, from the point of view of a finite number of independent loci, and indicates that significant practical progress can be made in estimating pairwise and multiple relationships. Although linked loci pose no theoretical problems, the computational complexity increases rapidly with the size of linkage groups.

[1] *Present address*: Forestry Commission, Northern Research Station, Roslin EH25 9SY, Scotland.

0040-5809/83 $3.00

The map of the human genome is being filled in increasingly rapidly in recent years (for a recent review of linkage analysis in man with further references see Conneally and Rivas (1980)), and there is the prospect of DNA sequencing becoming commonplace. It may therefore be timely to look tentatively toward the day when measurable informative loci are located densely throughout the genome, so that chromosomes are better represented by line segments, which are broken and respliced by crossovers, than as finite collections of loci. This is the approach we adopt in this paper. This approach is also the basis of Fisher's theory of junctions (Fisher, 1949, 1954, 1959; Bennett, 1953, 1954), and of the papers by Franklin (1977) and Stam (1980). Franklin's paper is closest to our own and gives further references. In this paper we assume as Franklin does that the process of crossovers is Poisson (no interference), although this is known not to be the case.

Parallels can also be seen between some of the concepts in a paper by Schnell (1961) and those in the present paper, although Schnell deals with a finite number of loci and makes no assumption about the crossover process.

The problem of pedigree reconstruction is complex, and we examine only a very small part of it. We calculate the probability that two individuals with a given relationship have some segment of chromosome in common, due to its descent from a common ancestor. We assume that in the event of this occurring the segment of chromosome contains a sufficient number of informative loci for its identity by descent in the two individuals to be conclusively established, thus proving the individuals to be related. This possibility was suggested as long ago as 1963 by Smith (1963).

In Section 1 we present a formal mathematical framework for the study of linkage. Although this is more formal than is strictly necessary for our immediate problem, it is hoped that it will provide a foundation for further studies. We define the useful concept of "chromosome pedigree" in which each nonfounder chromosome is obtained by a "crossover process" from two "parent chromosomes." If we label the parent chromosomes 0 and 1, then the crossover process may be considered as a continuous-time Markov random walk on these two states, the "time" parameter being distance along the chromosome. When we consider jointly all the crossover process in the chromosome pedigree, the states are vectors of zeros and ones, only a single coordinates of which changes when a crossover occurs somewhere in the pedigree, so that we have a continuous-time Markov random walk on the vertices of a hypercube. This representation of the gene flow in a pedigree as a random walk on the vertices of a hypercube forms the basis of this paper. Genetic events, such as a relationship between two individuals being detectable, correspond to the random walk hitting a particular subset of vertices.

In Section 2 we recall briefly some results on hitting times for continuous-time Markov processes. Our hypercubes have, however, such a large number

of vertices that the computation of results would be impossible were it not that we can exploit the symmetry of the situation to divide the vertices into sets of corresponding vertices, called "orbits." We then need not keep track of individual vertices, but only of the much smaller number of orbits. This is described in Section 3. In Section 4 various types of relationship are studied, obtaining the Q matrix and initial vector for the process of orbits, which can then be used to obtain the results presented in Section 5. Finally, in Section 6, we obtain an approximation to the exact results and use this to investigate robustness to some of the assumptions made.

The model used in this paper must be regarded as tentative pending further knowledge of DNA sequence structure and crossovers in man, although it is hoped that the methodology will be of use whatever the precise model turns out to be. A discussion of some of the assumptions made in the light of present knowledge, as well as suggestions for further developments, will be found in Donnelly (1982).

1. A Formal Framework for the Study of Linkage in Man

An Introductory Example

Consider two half-sibs with a father in common and let us restrict attention to the chromosome number 1's which they possess. In Fig. 1 we represent the paternal chromosomes C and D of the two children and the chromosomes from which they are derived, namely the father's maternal chromosome A and paternal chromosome B.

Since this diagram has the structure of a pedigree, except that A, B, C, and D are chromosomes rather than persons, we call it a chromosome pedigree. Taking the analogy further, we say that chromosome A is the mother chromosome of chromosome C and chromosome B is the father chromosome. A and B are the founder chromosomes of the pedigree.

The idea of chromosome pedigree also includes a notion of how the

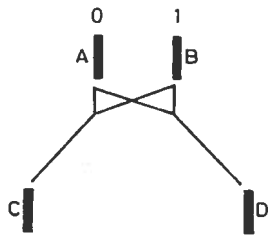

FIGURE 1

nonfounder chromosomes C and D are obtained by a random process from their parent chromosomes. At any point ("locus") along its length, chromosome C is equally likely to be a copy of A or a copy of B. We denote the former event by 0 and the latter by 1. As we move along the length of the chromosome, however, crossovers may occur at random causing us to switch from 0 to 1 or vice versa. The same thing is happening for chromosome D. At any locus we may write the joint state of C and D as one of the four possibilities {00, 01, 10, 11}, which correspond to the vertices of a square, as illustrated in Fig. 2. As we move in parallel along chromosomes C and D, the joint state may change for example from 00 to 10 if a crossover occurs in the formation of C, or from 00 to 01 if a crossover occurs in D. Since two crossovers are extremely unlikely to coincide, we cannot move directly from 00 to 11. We thus have a random walk on the vertices of a square, moving between adjacent vertices.

The event 00 corresponds to both C and D being at that point a copy of chromosome A, and the vertex 11 to them both being a copy of B; otherwise C and D are unequal. Thus the event that the two children share a section of chromosome identical by descent corresponds to the random walk hitting the hitting set {00, 11}.

We now place these ideas in a more formal and general mathematical context before going on to apply them. The central notion of chromosome pedigree applies to X chromosomes as well as autosomes, although we shall not study X chromosome pedigrees in this paper.

Chromosomes

There are 23 *types* of *chromosome*. Chromosomes of types 1 to 22 are called *autosomes*. Chromosomes of the 23rd type are called *X chromosomes*. (We ignore the existence of Y chromosomes since they are likely to have very few genes.)

Corresponding to each type of chromosome i is a positive real number l_i called the *length* or *map length* of chromosomes of type i. Although the map lengths of chromosomes appear to differ in males and females (Conneally and Rivas, 1980), we ignore this and assume a fixed map length for each chromosome type.

A chromosome of type i is a random function from $[0, l_i]$ to the set of nucleotides. (What nucleotides are will be of no importance to us in this paper.) The distribution of the random function is the same for all chromosomes of the same type.

Chromosome Pedigrees

A *pedigree structure* is a finite, simply ordered set $\mathscr{C}$, together with a subset $\mathscr{F}$, called *founders*, and functions m and f, called *mother* and *father* functions, from the set of nonfounders $\mathscr{C}\setminus\mathscr{F}$ into $\mathscr{C}$. The images of an

element C under m and f are written Cm and Cf, and must precede C in the ordering.

A *chromosome pedigree* is pedigree structure where the set consists of chromosomes of a given type i, whose joint distribution is given as follows:

(a) The founder chromosomes are independent random functions with the appropriate distribution for their type.

(b) For each nonfounder chromosome C there is a random function s_C (called a *switch*) from $[0, l_i]$ to $\{0, 1\}$. (This function indicates which parent chromosome is being copied.) The distribution of s_C is invariant under interchange of 0 and 1, so s_C is a two-point random walk on $\{0, 1\}$, the "time" parameter being distance along the chromosome. The switches are independent and identically distributed, so the ordered set $\{s_C\}_{C\in\mathscr{C}\setminus\mathscr{F}}$ is a random walk on a $\#(\mathscr{C}\setminus\mathscr{F})$-dimensional hypercube.

(c) Once we are given the founder chromosomes and the switches, the nonfounder chromosomes are obtained deterministically as

$$\begin{aligned} C(t) &= Cm(t) \quad \text{if} \quad s_C(t) = 0 \\ &= Cf(t) \quad \text{if} \quad s_C(t) = 1. \end{aligned}$$

Thus the *gene flow* for a chromosome pedigree is described by the random walk $\mathbf{s}$ on a hypercube.

(d) Assumptions (b) and (c) require that we have no translocations, deletions, mutations, and equal crossover rates in males and females. We now assume no interference by requiring the distribution of s_C to be in fact given as follows: $s_C(0) = 0$ or 1 with equal probability, and s_C changes from 0 to 1 or vice versa at the points of a Poisson process of rate 1. These points are called *crossovers*. $\mathbf{s}$ is thus a continuous time Markov process on the vertices of a hypercube, starting with the equilibrium distribution, which assigns equal probability to each vertex.

Person Pedigrees

A *person* consists of a *sex* (*male* or *female*) together with an ordered set of chromosomes. There are two chromosomes of each type, except that males have only one X chromosome. One of each type is called a *maternal chromosome* and one is called a *paternal chromosome*. The single X chromosome of a male is always a maternal chromosome. The chromosomes of a person P will be written P_1_m, P_1_p, P_X_m, etc., where, for example, P_1_m represents the maternal chromosome of type 1.

A *person pedigree* is a pedigree structure whose underlying set consists of persons, and which satisfies the conditions which we now list. For each nonfounder person P, the mother, Pm, of P, is female, and the father, Pf of P,

is male. Further, the joint distribution of all the chromosomes of the persons in the pedigree is given as follows:

(a) The paternal X chromosomes of nonfounder females P are the same as the father's X chromosome;

$$P\_X_p = Pf\_X_m.$$

(b) The autosomes of each type, and the X chromosomes, excluding paternal X chromosomes of nonfounder females, form 23 independent chromosome pedigrees with founder chromosomes defined to be the chromosomes of founder persons. The parent chromosomes of a maternal chromosome are the mother's two chromosomes, and the parent chromosomes of a paternal chromosome are the father's two chromosomes, except that paternal X chromosomes of nonfounder females are omitted since they come unchanged from the father. More formally, the mother and father functions for the chromosome pedigrees are defined as follows for autosomes of type i:

$$\begin{aligned}(P\_i_m)m &= Pm\_i_m,\\ (P\_i_m)f &= Pm\_i_p,\\ (P\_i_p)m &= Pf\_i_m,\\ (P\_i_p)f &= Pf\_i_p;\end{aligned}$$

and as follows for the maternal X chromosomes of nonfounders:

$$\begin{aligned}(P\_X_m)m &= Pm\_X_m\\ (P\_X_m)f &= Pm\_X_p \qquad \text{if} \quad Pm \text{ is a founder,}\\ &= Pmf\_X_m \qquad \text{if} \quad Pm \text{ is nonfounder.}\end{aligned}$$

Thus every person pedigree has 23 chromosome pedigrees embedded within it, with mother and father functions defined as above. The 22 autosomal chromosome pedigree structures are isomorphic.

Detectable Relationships between Chromosomes

Suppose we are dealing with a fixed chromosome pedigree $\mathscr{C}$, with founder set $\mathscr{F}$. The gene flow $\mathbf{s}$ is a continuous time Markov random walk on the hypercube

$$H = \{0, 1\}^{\mathscr{C}\setminus\mathscr{F}}.$$

The "*founder being copied from*" *function* F: $H \times \mathscr{C} \to \mathscr{F}$ is defined recursively by

$$F(\mathbf{s}(t), C) = C$$

if C is a founder, and

$$\begin{aligned}F(\mathbf{s}(t), C) &= F(\mathbf{s}(t), Cm) \qquad \text{if} \quad s_C(t) = 0\\ &= F(\mathbf{s}(t), Cf) \qquad \text{if} \quad s_C(t) = 1,\end{aligned}$$

if C is a nonfounder.

Thus $F(\mathbf{s}, C)$ specifies the founder chromosome of which chromosome C is a copy at each "time" $t \in |0, l_i|$, given the gene flow $\mathbf{s}$. Points t in $|0, l_i|$ are called *loci*.

Two chromosomes C and D are said to be *identical by descent at locus* t if $F(\mathbf{s}(t), C) = F(\mathbf{s}(t), D)$, that is, if, at locus t, they are both derived from the same founder chromosome. They will then almost surely be identical by descent in an interval about t.

Chromosomes C and D are said to be *detectably related* if they are equal on some interval $(t_1, t_2) \subseteq |0, l_i|$. We assume now that the information density (entropy density) of chromosomes is infinite everywhere on $|0, l_i|$. For practical purposes this means that any section of chromosome, down to the shortest in which we might be interested, contains sufficient polymorphic loci for it to be distinct in unrelated individuals. We are thus assuming that there are no fixed sequences of nucleotides of any significant length, that polymorphic loci can be detected everywhere along the chromosome, and that linkage disequilibrium does not rise with decreasing map distance in such a way as to make the information density finite. This assumption implies that two chromosomes are detectably related if and only if they are identical by descent at some locus t.

Thus C and D are detectably related if and only if by time l_i the random walk $\mathbf{s}$ has hit the hitting set $\mathscr{K}$ given by

$$\mathscr{K} = \mathscr{K}(C, D) = \{h \in H: F(h, C) = F(h, D)\}.$$

$\mathscr{K}$ is a subset of the vertices of the hypercube.

2. Hitting Time for a Continuous-Time Markov Process

For an introduction to the theory of Markov processes and a definition of the terms used in this section see, e.g., Parzen (1962) or Karlin and Taylor (1975).

We wish to calculate the probability that a Markov random walk on a

hypercube hits a given set of vertices $\mathscr{H}$ by time l. Thus we wish to calculate the hitting time distribution for the hitting set $\mathscr{H}$. Since we are not interested in what the process does after it first hits $\mathscr{H}$, we can make $\mathscr{H}$ an absorbing set and calculate, equivalently, the distribution of the absorption time.

The behavior of the Markov process is determined by its Q matrix, which we denote for the moment A, in order to save the symbol Q for a matrix of greater interest. The off-diagonal elements of A are the transition intensities, defined such that $a_{ji}\,\delta t$ is the probability of the process being in state j after a small time δt, starting from state i. The diagonal elements are the negative of the passage intensities, defined such that $-a_{ii}\,\delta t$ is the probability that if the process starts in state i, it has left it after time δt. The columns of A thus sum to 1. This is in fact the transpose of the usual Q-matrix definition, but we use it in order to be able to write the vectors of state probabilities as column vectors.

If we order the states so that the hitting set states (which are now absorbing states) come first, then we may write the Q matrix in the block form

$$A = \left(\begin{array}{c|c} P & R \\ \hline 0 & Q \end{array}\right).$$

The matrix Q, which describes the behavior of the process before it hits the hitting set, is all that need concern us, since the probability that the process has hit the hitting set may be deduced by summing the nonhitting state probabilities. Q is obtained by deleting from the full Q matrix the rows and columns corresponding to the hitting set. It is called a dishonest Q matrix because its columns may sum to less than 0; it describes a process which has some probability of leaving entirely the set of states under consideration.

If the initial vector of probabilities for the nonabsorbing states, were an eigenvector $\mathbf{v}$ of Q, corresponding to eigenvalue λ, then the vector of probabilities at time t would be $e^{\lambda t}\mathbf{v}$. λ is always negative. $\lambda\mathbf{v}$ is the rate at which probability is "leaking" into the absorbing states. The hitting time distribution would thus be exponential with parameter $-\lambda$.

Since in general the initial vector $\mathbf{u}$ may be expressed as a mixture of the eigenvectors $\mathbf{v}_i$ (which are normalized so that their elements sum to 1),

$$\mathbf{u} = \sum \alpha_i \mathbf{v}_i,$$

the hitting time distribution is a mixture of exponential distributions with parameters $-\lambda_i$, the weights being α_i.

The mean hitting time μ is the weighted average of the means $-\lambda_i^{-1}$ of the exponential distributions, and so is given by $\mu = -\sum \alpha_i \lambda_i^{-1}$. This may be

more easily calculated as $\mu = -\mathbf{1}^t Q^{-1}\mathbf{u}$, where $\mathbf{1}$ is a column vector of 1's and t denotes transpose, since

$$Q^{-1}\mathbf{u} = Q^{-1}\left(\sum \alpha_i \mathbf{v}_i\right)$$

$$= \sum \alpha_i \lambda_i^{-1} \mathbf{v}_i$$

and $\mathbf{1}^t\mathbf{v}_i = 1$ for each i, by the normalization of the $\mathbf{v}_i$'s.

3. Symmetries and Orbits

The method of analysis which we have just described cannot be used directly in our application due to computational difficulties. The number of states of the Markov process (vertices of the hypercube in our case) is so large that eigenvectors and eigenvalues of the Q matrix would take too long to compute. We now describe, and clarify with a simple example, how any symmetry present in the problem can be used to considerably reduce the computation involved. The symmetry induces a partition of the set of vertices into sets of like vertices, called "orbits." We need not keep track of individual vertices, but only of the orbit in which the process lies For an introduction to group theory and orbits see Green (1965) or Wielandt (1964), and for an example of their use in a genetic context see Thompson (1974b).

If G is a group of symmetries of the hypercube, the orbit O_α of a vertex α is defined to be the set of vertices onto which α is mapped by the group of symmetries:

$$O_\alpha = \{g(\alpha)\colon g \in G\}.$$

Belonging to the same orbit can easily be shown to be an equivalence relation, so that the set of orbits forms a partition of the set of vertices (Green, 1965, p. 53; Wielandt, 1964, p. 4). If at each time t we replace the vertex α on which our Markov process lies by the corresponding orbit O_α, we obtain another stochastic process whose state space is now the set of orbits instead of the set of vertices. This process of orbits may easily be shown to be a Markov process. If the group G leaves the hitting set $\mathscr{H}$ invariant, then $\mathscr{H}$ must be the union of a subset $\mathscr{H}'$ of the set of orbits, and the hitting time distribution may be found more easily by looking at the time the process of orbits takes to hit $\mathscr{H}'$. The maximum simplification is obtained when G is the stabiliser of $\mathscr{H}$, the group of symmetries which leaves $\mathscr{H}$ invariant. This also helps to solve the technical problem caused by the eigenvalues of Q not all being distinct.

A d-dimensional hypercube (considered as a graph of vertices and edges) has $2^d d!$ symmetries, these being given by changing 0 to 1 and vice versa in one of the 2^d subsets of the d coordinates, and then applying one of the $d!$ possible permutations to the ordering of the coordinates.

Example

Consider the continuous time random walk on the square, with state space $S = \{00, 01, 10, 11\}$ and the single-vertex hitting set $\mathcal{H} = \{00\}$. The group of symmetries G then consists of the identity, and the reflection r which interchanges 01 and 10 (Fig. 2). The orbits are $\{00\}$, $\{01, 10\}$, and $\{11\}$. The Q matrix for the process of orbits is

$$\begin{matrix} \mathcal{H} = \{00\} \\ \{01, 10\} \\ \{11\} \end{matrix} \begin{bmatrix} -2 & 1 & 0 \\ 2 & -2 & 2 \\ 0 & 1 & -2 \end{bmatrix},$$

and for the dishonest process with the hitting set $\mathcal{H}$ removed is

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

The initial vector for the dishonest process of orbits is

$$\mathbf{u} = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{4} \end{bmatrix},$$

assuming an initial uniform distribution on the four vertices of the square.

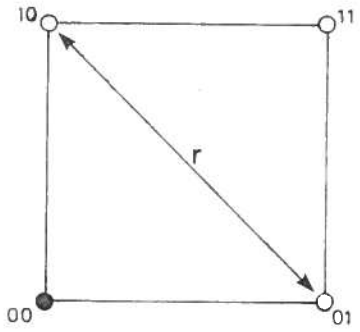

FIGURE 2

The mean hitting time is therefore

$$\begin{aligned} \mu &= -\mathbf{1}^t Q^{-1} \mathbf{u} \\ &= -\{1, 1\} \tfrac{1}{2} \begin{bmatrix} -2 & -2 \\ -1 & -2 \end{bmatrix} \begin{bmatrix} \frac{1}{2} \\ \frac{1}{4} \end{bmatrix} \\ &= 5/4, \end{aligned}$$

as may be verified by the usual simultaneous equations method.

4. Examples of Particular Relationships

We now look at the probability of two persons being detectably related through their autosomal chromosomes for some particular types of relationship. In this section we just derive the transition matrix and initial vector for the process of orbits. These are supplied to the computer program which produces the results given in Section 5.

Grandparent-Type Relationships

This is the type of relationship where person A is a grandparent, or more generally a direct ancestor, of person B. We can assume without loss of generality that the relationship is always through the maternal line. The chromosome pedigree is shown in Fig. 3, with the part of interest, containing the maternal chromosomes, boxed in.

We obtain a detectable relationship only when all d switches in the box point to the maternal chromosomes above. That is, the hitting set $\mathcal{H}$ consists of the single vertex $00 \cdots 0$ of a d-dimensional hypercube, where $d = 1$ for grandparents, 2 for great-grandparents, and so on. We could also include the trivial case $d = 0$, for parents, where a detectable relationship is certain.

The group of symmetries of the hypercube which leave $00 \cdots 0$ invariant consists of all permutations of the coordinates. The 2^d vertices partition into $d + 1$ orbits according to their distance from $00 \cdots 0$; that is, according to the number of coordinates equal to 1. The number of vertices in the ith orbit (vertices with i 1's) is given by the binomial coefficient $\binom{d}{i}$.

The Q matrix for the process of orbits is

$$\begin{bmatrix} -d & 1 & & & & \\ d & -d & 2 & & & \\ & d-1 & -d & \ddots & & \\ & & d-2 & \ddots & d-1 & \\ & & & \ddots & -d & d \\ & & & & 1 & -d \end{bmatrix}.$$

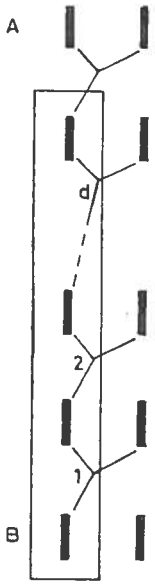

FIGURE 3

The dishonest Q-matrix for the nonabsorbing states is obtained by deleting the first row and column. Since all the vertices have equal initial probability 2^{-d}, the initial vector for the process of orbits is

$$\frac{1}{2^d}\begin{bmatrix}\binom{d}{0}\\ \binom{d}{1}\\ \vdots\\ \binom{d}{d}\end{bmatrix}.$$

Probability of All of a Person's Genes Being Passed On to His or Her Children

This is not a two-person problem but is included here because it is similar to the detectable relationship poblem for grandparent type relationships. For simplicity we consider only autosomal genes, as in our other examples. The autosomal chromosome pedigree for a person with d children is shown in Fig. 4.

In this problem it is the "not" event, the event that some part of the

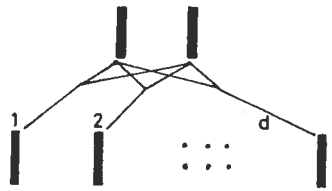

FIGURE 4

genome is not passed on, which corresponds to the random walk hitting a set of vertices of the hypercube. The event that some part of the person's paternal chromosome is not copied to any of the children corresponds to all of the d switches pointing simultaneously at some locus to the maternal chromosome, in other words, to hitting the vertex $000\cdots0$ of a d-dimensional hypercube. Similarly, the maternal chromosome not being copied corresponds to the diagonally opposite vertex $11\cdots1$. So the hitting set is the two-vertex set $\{00\cdots0,\ 11\cdots1\}$. This is shown in Fig. 5 for the case $d=3$.

The vertices of the hypercube partition into orbits according to how many 1's they contain, in a similar manner to the case of grandparent-type relationships, except that the orbits of vertices with 1 and $(d-1)$ 1's merge, as do the orbits of vertices with 2 and $(d-2)$ 1's, and so on. This merging of the orbits is caused by the fact that we now have an additional symmetry of the hypercube—the reflection which changes all coordinates which are 0 to 1, and vice versa.

If d is even, then there are $(d/2)+1$ orbits, and the Q matrix and initial vector for the process of orbits are, respectively,

$$\begin{bmatrix}
-d & 1 & & & & & \\
d & -d & 2 & & & & \\
 & d-1 & -d & \ddots & & & \\
 & & d-2 & \ddots & \frac{d}{2}-2 & & \\
 & & & \ddots & -d & \frac{d}{2}-1 & \\
 & & & & \frac{d}{2}+2 & -d & d \\
 & & & & & \frac{d}{2}+1 & -d
\end{bmatrix}$$

and

$$\frac{1}{2^{d-1}}\begin{bmatrix} \binom{d}{0} \\ \binom{d}{1} \\ \vdots \\ \binom{d}{\frac{d}{2}-1} \\ \frac{1}{2}\binom{d}{\frac{d}{2}} \end{bmatrix}.$$

If d is odd, then there are $(d+1)/2$ orbits, and the Q matrix and initial vector for the process of orbits are

$$\begin{bmatrix} -d & 1 & & & & & \\ d & -d & 2 & & & & \\ & d-1 & -d & \ddots & & & \\ & & d-2 & \ddots & \frac{d-3}{2} & & \\ & & & \ddots & -d & \frac{d-1}{2} & \\ & & & & \frac{d+3}{2} & -d & \frac{d+1}{2} \\ & & & & & \frac{d+1}{2} & -\frac{d+1}{2} \end{bmatrix}$$

and

$$\frac{1}{2^{d-1}}\begin{bmatrix} \binom{d}{0} \\ \binom{d}{1} \\ \vdots \\ \binom{d}{\frac{d-1}{2}} \end{bmatrix}.$$

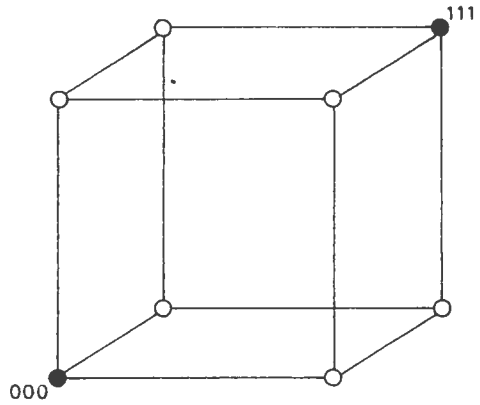

FIGURE 5

The hitting set is the first orbit. The Q matrix and initial vector for the dishonest process which makes this an external absorbing state are therefore obtained by removing the first row and column.

Half-Sib-Type Relationships

In this type of relationship person A, or some ancestor of A, is the half-sib of B, or of some ancestor of B, this being the only way in which A and B are related (Fig. 6). This way of extending the half-sib relationship to half-sib type relationships is similar to the way in which we shall extend the cousin relationship to cousin-type relationships, such as second cousins once removed, and the uncle relationship to uncle-type relationships, such as great-great uncles. The subsequent analysis is also similar. We analyze half-

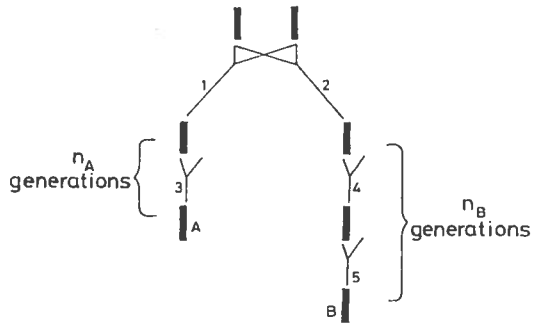

FIGURE 6

sib-type relationships first, although this is a less common type of relationship in most societies, because the hitting set and the number of orbits are smaller, making the analysis simpler.

For half-sibs, we can see from Fig. 6 that the hitting set is $\{00, 11\}$. For the case where A and B are the n_Ath and n_Bth generation descendants, respectively, of a pair of half-sibs, the only difference is that we have $n_A + n_B$ additional switches which must all point the right way, say to 0. (In Fig. 6, $n_A = 1$ and $n_B = 2$.) This means that in general the hitting set is the set $\{0000 \cdots 0,\ 1100 \cdots 0\}$, consisting of two vertices differing in precisely two coordinates. The dimension d of the hypercube is $n_A + n_B + 2$.

Note, incidentally, that two relationships with the same value of $n_A + n_B$ have exactly the same hitting set. This means that no amount of autosomal genetic information is of any help in distinguishing these two relationships. (Unless of course information is available on a third person related to both the individuals in question.) The same is true for relationships such as third cousins and second cousins twice removed. Moreover this does not depend on the process of crossovers being Poisson.

To find the orbits and the Q matrix for the process of orbits, consider first the half-sib relationship itself, with hitting set $\{00, 11\}$ on the two-dimensional hypercube as shown in Fig. 7. The group of symmetries is generated by the reflections about the diagonals and the orbits are the hitting set $\{00, 11\}$ itself, which we denote $A0$, and the set $\{01, 10\}$, which we denote $B0$. The Q matrix is

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

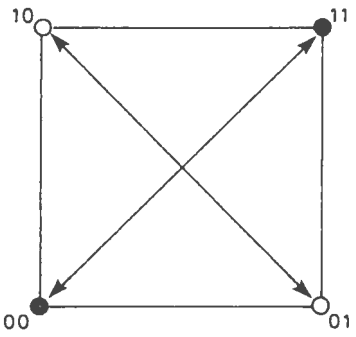

FIGURE 7

since there are two edges going from each vertex of $A0$ to $B0$ and from each vertex of $B0$ to $A0$, and the initial vector is

$$\mathbf{u} = \tfrac{1}{4}\begin{bmatrix} 2 \\ 2 \end{bmatrix},$$

reflecting the number of vertices in the orbits. It is convenient to write the Q matrix as $M - 2I$, where

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

Consider now the general relationship with hitting set $\{000 \cdots 0,\ 110 \cdots 0\}$ on the d-dimensional hypercube. The 2-face with the third and subsequent coordinates set to 0 is invariant under the group of symmetries, since it is the only 2-face containing the hitting set. The distances of vertices from this 2-face are also invariant.

The group of symmetries is thus generated by the symmetries which we already have for half-sibs, applied to the first two coordinates, together with the group of permutations of third and subsequent coordinates. The orbits may be written

$$A0,\quad B0,\quad A1,\quad B1,\ldots,A(d-2),\quad B(d-2),$$

where the A or B denotes the state of the first two coordinates, and the integer denotes the number of 1's among subsequent coordinates.

From each vertex of orbit Ai there are two edges leading to orbit Bi; there are i edges leading to orbit $A(i-1)$ (since a change of any of the i coordinates which are 1 among the third and subsequent coordinates gives a vertex of orbit $A(i-1)$); and there are $d-2-i$ edges leading to orbit $A(i+1)$. A similar statement holds for transitions from orbit Bi.

Thus the Q matrix has the block form

$$\begin{bmatrix} M-dI & I & & & \\ (d-2)I & M-dI & 2I & & \\ & (d-3)I & M-dI & \ddots & \\ & & (d-4)I & \ddots & \\ & & & \ddots & \end{bmatrix}$$

and the initial vector has the block form

$$\frac{1}{2^{d-2}}\begin{bmatrix} \binom{d-2}{0} & \mathbf{u} \\ \binom{d-2}{1} & \mathbf{u} \\ \vdots & \\ \binom{d-2}{d-2} & \mathbf{u} \end{bmatrix}.$$

M, I, and $\mathbf{u}$ are defined as for the half-sib relationship.

We see that the analysis for half-sib-type relationships is in some ways like a mixture of the half-sib relationship and grandparent-type relationships.

Cousin-Type Relationships

These are relationships of the type "sth cousins t times removed," where $s \geqslant 1$ and $t \geqslant 0$. For first cousins (A and B in Fig. 8), the simplest relationship of this type, the hitting set may be seen to consist of the following 16 vertices of a six-dimensional hypercube:

00*00*,
01*01*,
1*01*0,
1*11*1;

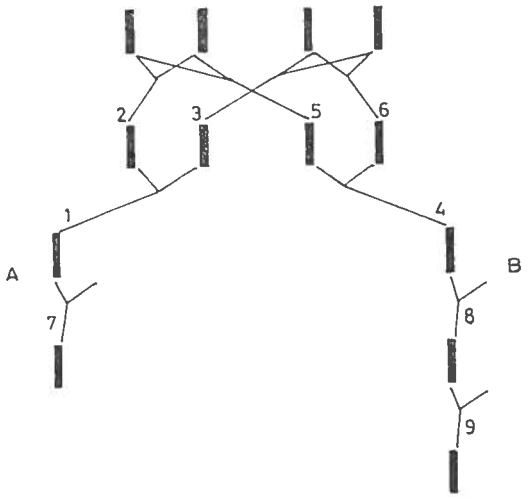

FIGURE 8

where each * can denote either 0 or 1. For sth cousins t times removed the hitting set consists of the same 16 vertices of a $(2s+t+4)$-dimensional hypercube, with the additional coordinates set to 0.

The determination of the orbits for the cousin relationship is more tedious than for previously considered relationships, but is even more important because the unsimplified problem has an unwieldy $2^6 \times 2^6$ Q matrix. We start off by looking for generators for the group G of symmetries which leaves the hitting set invariant. We denote by $v(i)$ the ith coordinate of a typical vertex v.

The first and fourth coordinates form a pair distinguishable from the others, so any element of G either transposes them or leaves them where they are. If we take the symmetry which does nothing other than transpose them as a generator, then we can confine our search for further generators to symmetries which leave them where they are.

No element of G can change $v(1)$ or $v(4)$ (i.e., change 0 to 1 and vice versa) without changing the other, because such a transformation would not leave the hitting set invariant. We take as generator the element of G which simultaneously changes $v(1)$, changes $v(4)$, transposes $v(2)$ and $v(3)$, and transposes $v(5)$ and $v(6)$, and confine our search for further generators to symmetries which do nothing to $v(1)$ and $v(4)$.

The second and fifth coordinates now form a distinguishable pair, as do the third and sixth. As before if we change one coordinate of the pair we must change both.

This argument leads to the following set of six generators for G:

(a) transpose $v(1)$ and $v(4)$,

(b) change $v(1)$, change $v(4)$, transpose $v(2)$ and $v(3)$, and transpose $v(5)$ and $v(6)$,

(c) transpose $v(2)$ and $v(5)$,

(d) change $v(2)$ and change $v(5)$,

(e) transpose $v(3)$ and $v(6)$,

(f) change $v(3)$ and change $v(6)$.

We know that this is a complete set of generators because after finding (f) we are left looking for symmetries which do nothing to any of the coordinates.

Having a set of generators available makes it easy to partition the vertices into orbits. To start with, the vertices divide up according to whether $v(1) = v(4)$, because this characteristic is not changed by any of the six generators. The vertices with $v(1) = v(4)$ split into four groups according to whether the following two statements are true:

(i) $[v(1)=v(4)=0$ and $v(2)=v(5)]$
or $[v(1)=v(4)=1$ and $v(3)=v(6)]$,

(ii) $[v(1)=v(4)=0$ and $v(3)=v(6)]$
or $[v(1)=v(4)=1$ and $v(2)=v(5)]$.

The vertices with $v(1)\neq v(4)$ split into three groups according to whether neither, one, or both of the following statements are true:

(iii) $v(2)=v(5)$,

(iv) $v(3)=v(6)$.

(Vertices with only (iii) true lie in the same orbit as vertices with only (iv) true, as may be seen by applying generators (a) and (b) above.)

It may be verified that no further splitting occurs, so that, choosing an arbitrary ordering, the orbits are as follows:

(1) $v(1)=v(4)$, (i) and (ii) true,
(2) $v(1)=v(4)$, (i) true, (ii) false,
(3) $v(1)=v(4)$, (i) false, (ii) true,
(4) $v(1)=v(4)$, (i) and (ii) false,
(5) $v(1)\neq v(4)$, (iii) and (iv) true,
(6) $v(1)\neq v(4)$, (iii) or (iv) true but not both,
(7) $v(1)\neq v(4)$, (iii) and (iv) false.

The hitting set consits of orbits (1) and (2).

Orbit (6) has 16 vertices; the rest have 8. Thus the initial vector for the process of orbits is

$$\mathbf{u}=\tfrac{1}{8}\begin{bmatrix}1\\1\\1\\1\\1\\2\\1\end{bmatrix}.$$

By choosing a vertex from each of the orbits and seeing which orbits it moves to when a single coordinate is changed, we find that the Q matrix for the process of orbits is $M-6I$, where I is the identity matrix, and

$$M=\begin{bmatrix}0&2&2&0&2&0&0\\2&0&0&2&0&1&0\\2&0&0&2&0&1&0\\0&2&2&0&0&0&2\\2&0&0&0&0&2&0\\0&2&2&0&4&0&4\\0&0&0&2&0&2&0\end{bmatrix}.$$

For the general case of sth cousins t times removed we find, in a similar manner to the analysis for half-sib-type relationships, that the Q matrix for the process of orbits has the block form

$$\begin{bmatrix}M-dI & I & & \\ (d-6)I & M-dI & 2I & \\ & (d-7)I & M-dI & \ddots \\ & & (d-8)I & \ddots \\ & & & \ddots\end{bmatrix},$$

where $d=2s+t+4$ is the dimension of the hypercube. The initial vector has the block form

$$\frac{1}{2^{d-6}}\begin{bmatrix}\binom{d-6}{0}\mathbf{u}\\ \binom{d-6}{1}\mathbf{u}\\ \vdots\\ \binom{d-6}{d-6}\mathbf{u}\end{bmatrix}.$$

The hitting set always consists of the first two orbits.

Uncle-Type Relationships

This type of relationship, illustrated in Fig. 9, includes the case where person A is the uncle, or great uncle, or great-great uncle, etc., of person B. Note that unlike previous two-person relationships which we have considered, it involves three chromosomes and does not reduce to a relationship between two chromosomes. However, the event that two such persons are detectably related still amounts to a random walk hitting a set of vertices on a hypercurbe.

For the uncle relationship, the simplest case, the hitting set consists of the following 16 vertices of a five-dimensional hypercube:

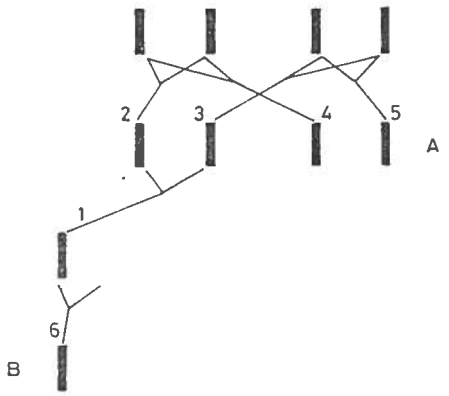

FIGURE 9

00*0*,
01*1*,
1*0*0,
1*1*1;

where each * may represent either 0 or 1.

This hitting set is similar to the hitting set for cousins, but the orbits are easier to determine. After a little thought it may be seen that the group of hypercube symmetries leaving the hitting set invariant is generated by the following symmetries ($v(i)$ represents the ith coordinate of a typical vertex v):

(a) change $v(1)$, transpose $v(2)$ and $v(3)$, and transpose $v(4)$ and $v(5)$,

(b) transpose $v(2)$ and $v(4)$,

(c) change $v(2)$ and change $v(4)$,

(d) transpose $v(3)$ and $v(5)$,

(e) change $v(3)$ and change $v(5)$.

It may then be seen that the vertices partition into four orbits according to whether or not each of the following two statements is true, these being the only characteristics which remain invariant under the above symmetries:

(i) $|v(1)=0$ and $v(2)=v(4)|$ or $|v(1)=1$ and $v(3)=v(5)|$,

(ii) $|v(1)=0$ and $v(3)=v(5)|$ or $|v(1)=1$ and $v(2)=v(4)|$.

Statement (i) in fact defines the hitting set.

The orbits are thus

$$1\begin{Bmatrix}0&00&00&1&00&00\\0&11&11&1&11&11\\0&01&01&1&10&10\\0&10&10&1&01&01\end{Bmatrix},\quad 2\begin{Bmatrix}0&00&01&1&00&10\\0&11&10&1&11&01\\0&01&00&1&10&00\\0&10&11&1&01&11\end{Bmatrix},$$

$$3\begin{Bmatrix}0&00&11&1&00&11\\0&11&00&1&11&00\\0&01&10&1&10&01\\0&10&01&1&01&10\end{Bmatrix},\quad 4\begin{Bmatrix}0&00&10&1&00&01\\0&11&01&1&11&10\\0&01&11&1&10&11\\0&10&00&1&01&00\end{Bmatrix},$$

orbits 1 and 2 being the hitting set.

The Q matrix for the process of orbits may be seen to be $M-5I$, where I is the identity matrix and

$$M=\begin{bmatrix}1&2&0&2\\2&0&2&1\\0&2&1&2\\2&1&2&0\end{bmatrix}.$$

The initial vector is

$$\mathbf{u}=\begin{bmatrix}\frac{1}{4}\\\frac{1}{4}\\\frac{1}{4}\\\frac{1}{2}\end{bmatrix}.$$

The analysis extends to general uncle-type relationships in exactly the same way as for cousin-type relationships and half-sib-type relationships. The Q matrix for the process of orbits is

$$\begin{bmatrix}M-dI & I & & \\ (d-5)I & M-dI & 2I & \\ & (d-6)I & M-dI & \ddots \\ & & (d-7)I & \ddots \\ & & & \ddots\end{bmatrix},$$

and the initial vector is

$$\frac{1}{2^{d-5}}\begin{bmatrix} \binom{d-5}{0} & \mathbf{u} \\ \binom{d-5}{1} & \mathbf{u} \\ \vdots & \\ \binom{d-5}{d-5} & \mathbf{u} \end{bmatrix},$$

where d, the dimension of the hypercube, is 6 for great uncles, 7 for great-great uncles, etc. The hitting set is always given by the first two orbits.

5. Results

Now that we have reduced the computational complexity of the problem by finding, for various relationships, the Q matrix and initial vector of the process of orbits, it is easy to calculate the probability of a detectable relationship. As previously described, we can express the hitting time distribution as a mixture of exponentials by removing the rows and columns of the hitting set from the Q matrix and then determining its eigenvalues and eigenvectors. This was done using the FORTRAN computer subroutines of the standard NAG package (Numerical Algorithms Group, 1978).

To determine the hitting probabilities we need to know the chromosome lengths. We take the total autosomal map length, $L = \sum_1^{22} l_i$, to be 33. In fact it appears to differ in males and females, being about 27.5 in males and 38.5 in females (Renwick, 1971, p. 87), but we ignore this. We take the relative chromosome lengths, as percentages of the total autosomal map length, to be those measured cytologically and given by Maynard-Smith *et al.* (1961), namely,

9.12 8.53 7.16 6.59 6.15 5.87 5.31 4.92 4.81 4.71 4.60

4.47 3.56 3.60 3.40 3.20 3.12 2.72 2.48 2.27 1.77 1.64

for chromosomes 1 to 22, respectively, and 5.84 for the X chromosome.

Using these figures we obtain the probabilities given in Table I. The probabilities, for the different types of relationships, are also plotted in Fig. 10. For comparability they are plotted against k, where $k = d - \log_2 \#(\mathscr{H})$. That is, k is $-\log_2(p)$, where p is the proportion of the vertices of the hypercube which lie in the hitting set, or the probability of a detectable relationship at any particular locus. The probabilities are in fact very similar for different relationships with the same value of k.

We see that there is a 82% probability of a detectable relationship between a person and his eighth generation ancestor (great-great-great-great-great-great grandparent), but only a 16% probability of a detectable relationship with a 12th generation ancestor. This means that someone descended from the Scottish poet Robert Burns (born 1759) probably carries some of his genes, but that someone unilineally descended from the English playwright William Shakespeare (born 1564) is unlikely to have any genes in common with him. (In fact, the time scale in the latter example would make it very difficult to rule out other relationships, so that the present analysis is really only saying that proof of descent from William Shakespeare does little to increase the probability that the claimant has genes in common with him.)

For cousins we see that a person has a 70% chance of sharing some genes with a fourth cousin, but only a 10% chance of sharing genes with a sixth cousin. To be 90% sure of passing on all his genes to his offspring, a person has to have at least 13 children; with only 7 children he has only a 3% chance of passing on all his genes.

All these statements refer only to ausosomal genes, are subject to the assumptions of the model, notably no interference and a total map length of 33 in both males and females, and assume that individuals are not inbred, or related in any way other than that referred to.

TABLE I

Probability of *No* Detectable Relationship
(or Probability That All Genes Are Passed on to Offspring)

	Type of "relationship"					
k	Cousin	Uncle	Offspring	Half-sib	$e^{-kl/2\lambda}$	Grandparent
4	0.0000	0.0000	0.0000	0.0000	0.0003	0.0001
5	0.0012	0.0023	0.0017	0.0029	0.0058	0.0056
6	0.0232	0.0308	0.0280	0.0349	0.0453	0.0495
7	0.1206	0.1387	0.1347	0.1484	0.1645	0.1776
8	0.3069	0.3290	0.3268	0.3410	0.3566	0.3743
9	0.5196	0.5380	0.5381	0.5483	0.5599	0.5753
10	0.6976	0.7099	0.7111	0.7170	0.7245	0.7350
11	0.8216	0.8287	0.8300	0.8310	0.8376	0.8438
12	0.8989	0.9028	0.9038	0.9053	0.9078	0.9112
13	0.9442	0.9462	0.9468	0.9476	0.9490	0.9507
14		0.9707	0.9711	0.9714	0.9722	0.9731
15		0.9842	0.9844	0.9846	0.9850	0.9854

Note. $k = n + 1$ for (great)^n grandparent ($(n + 2)$th generation ancestor).
$k = n + 1$ for (great)^n uncle (uncle n times removed).
$k = 2s + t$ for sth cousins t times removed.
$k = n + 1$ for half-sibs n times removed.
$k = n - 1$ for n offspring.

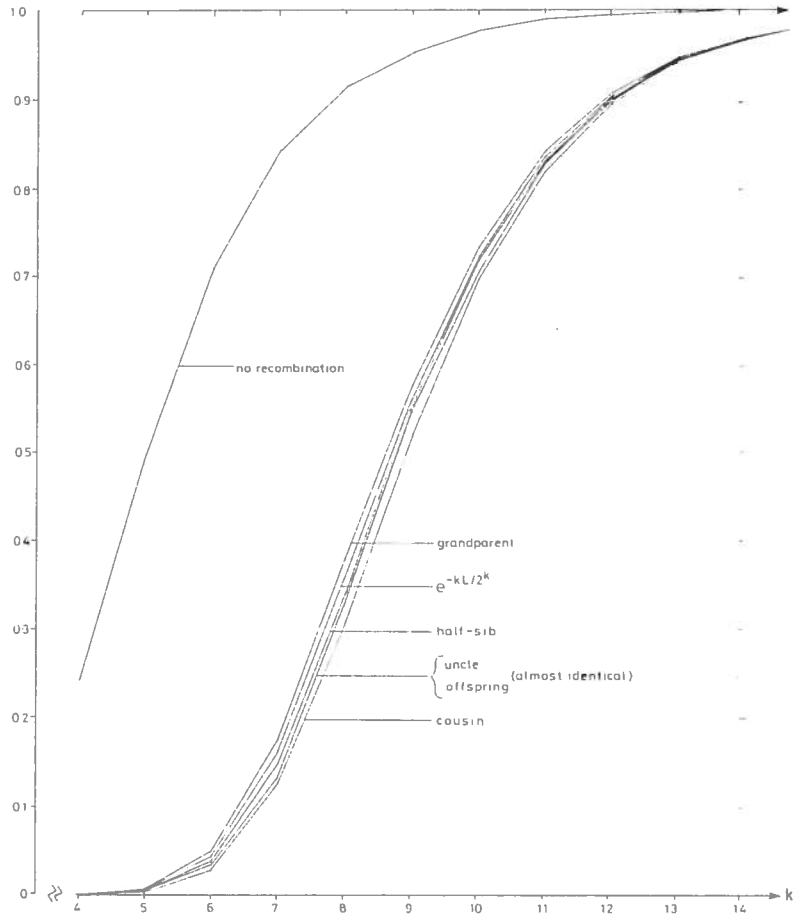

FIG. 10. Probability of no detectable relationship (or probability that all genes are passed on to offspring).

Figure 10 also shows what the probabilities would be if there were no crossing over, each of the 22 autosomes acting as a single independent locus, namely, $(1-2^{-k})^{22} \approx e^{-22/2^k}$. Crossing over can considerably increase the chances of a detectable relationship. For instance, in Smith's (1963) example of second cousins ($k=4$), the probability of no autosomally detectable relationship would be 0.25 if there were no crossing over. With crossing over the probability falls to 10^{-5}, so that Smith's intuition was in fact wrong in

his suggestion that recombination would not change the general order of magnitude.

6. Discussion

The exact results given in Table I could not have been obtained without the use of a computer. In this section, however, an empirical formula will be presented which gives a good approximation to the exact results and has a theoretical justification. This approximate formula indicates how the results would be changed by assuming a different total map length and slightly changed by assuming positive interference, and shows that the results are robust with respect to the relative chromosome map lengths and the exact nature of the positive interference. These predictions are verified by further exact calculations and by simulations.

Approximations

Also plotted in Fig. 9 is the function $e^{-kL/2^k}$. It can be seen that this gives a good approximation to the probabilities for all the types of relationship considered. The maximum error is less than 0.05 and occurs for cousin-type relationships.

This formula is mainly empirical, but it can be justified by the following argument. It can be shown (David Aldous, private communication) that a Markov random walk on the vertices of a hypercube has a memory which is short, of order d, compared to its mean return time, which is of order 2^d: the location of the random walk is little affected by where it was more than d steps ago. Thus returns to the hitting set occur approximately as a Poisson process.

The modification which must be made to this statement is that the random walk may, with small probability, return immediately, or almost immediately, to the same vertex of the hitting set, before it has time to "forget," and in the case of cousin, uncle or half-sib-type relationships, where the hitting set consists of a fairly compact set of vertices, the walk may move almost immediately to another vertex of the hitting set. Thus hits on the hitting set $\mathcal{H}$ come in widely spaced clusters, the mean number of hits in a cluster being say c, where c is slightly greater than 1. Since by symmetry the random walk spends a proportion $\#(\mathcal{H})/2^d = 2^{-k}$ of its time on the hitting set, the mean number of steps between clusters is $c2^k$.

The expected number of crossovers for chromosomes of type i is dl_i, so that the expected number of visits to vertices by the random walk is $dl_i + 1$. This is longer than the mean cluster length, but much shorter than the mean number of steps between clusters. The probability of a detectable relationship is the probability that a randomly located window of this width overlaps a

cluster. This probability is $(dl_i + 1)/(c2^k)$, or, slightly more accurately, $(dl_i + \gamma)/(c2^k)$, where γ is the mean length of a cluster, including any intervening nonhitting-set vertices. The probability of no detectable relationship for any autosome is therefore $\prod_i [1 - (dl_i + \gamma)/(c2^k)]$, or very nearly

$$\exp\left[-\frac{dL + 22\gamma}{c2^k}\right].$$

In the case of grandparent-type relationships and our offspring problem, we can derive the approximation $(d-2)/(d-3)$ for c, and taking γ to be approximately $2c - 1$ we can check the above formula, which proves to work very well. For all but small values of k it works better than the formula $e^{-kL/2^k}$, and for large values much better.

We can now see why the formula $e^{-kL/2^k}$ should be a good approximation for all types of relationship. It is because the effect of replacing $dL + 22\gamma$ by kL in the numerator is balanced by omitting c in the denominator. First the proportional effect, in any case fairly small, of the term 22γ on the numerator is close to the factor $(d-2)/(d-3)$ by which the mean cluster size c is changed by the possibility of an immediate return to the same vertex. Secondly, for cousin-type and uncle-type relationships, where the term dL is significantly greater than kL, the mean cluster size c is also increased because of the possibility of moving from one vertex of the hitting set to another. If all 16 hitting set vertices for these relationship types were gathered together on a single 4-face, we would have complete compensation, because removing the redundant four dimensions would get us back to the grandparent-type relationship with the same value of k. Since the hitting set is somewhat dispersed, however, the mean cluster size is lower and we do not have complete compensation, explaining why these types of relationship give lower probabilities in Table I. Similarly the "no hit" probabilities for our offspring problem are slightly lower than for half-sib-type relationships, because although both problems have a two vertex hitting set, it is more dispersed in the former case.

It would thus be unwise to rely on the approximate formula $e^{-kL/2^k}$ in the case of relationships with very widely dispersed hitting sets, as well as for other species such as *Drosophila* with much fewer chromosomes or a very different total map length.

Robustness

It might be expected in view of the above approximation that the effect of changing the relative chromosome lengths while keeping the total length L constant would be small. This was tested for cousin-, uncle-, and half-sib-type relationships by recomputing the probabilities, keeping $L = 33$ but

making all chromosomes the same length. The maximum change in the probabilities was less than 3×10^{-4}, which is indeed negligible.

A major assumption we have made is that the process of crossovers is Poisson, when in fact it is known not to be. It appears (Bailey, 1961) that in reality there is generally positive interference: the presence of a crossover makes the occurence of another crossover nearby less likely. In this case the arguments leading to the above approximation still hold good, on the whole. We still have widely spaced clusters of at most a very few hits, viewed through a randomly located window of expected width $dl_i + 1$. The main difference would appear to be that since a change of coordinate is unlikely to be immediately reversed, an immediate return to the same hitting set vertex is unlikely, and the mean cluster size c is correspondingly reduced slightly. We can therefore speculate that positive interference reduces the probabilities in Table I very slightly, this effect being independent of the precise form of interference.

This prediction was tested by simulating a renewal process of crossovers for each chromosome, with $\frac{1}{2}\Gamma(2)$ distributions ($\frac{1}{4}\chi^2(4)$ distributions) for the intervals between crossovers, and an equal $\frac{1}{4}\Gamma(1)$, $\frac{1}{4}\Gamma(2)$ mixture for the time to the first crossover for stationarity. From 10,000, 15,000, and 5000 simulations, respectively, of grandparent-type relationships with $k = 5$, $k = 7$, and $k = 9$, the "no detectable relationship" probabilities (± 1 SD) were estimated to be 0.0025 ± 0.0005, 0.1431 ± 0.0028, and 0.5400 ± 0.0070, respectively. These estimates agree with the values 0.0029, 0.1385, and 0.5363 obtained from the formula $\exp(-(kL + 22)/2^k)$, as we would expect if $c = 1$. Although the reduction from the "no interference" values in Table I is very small, it is proportionally greater for close relatives, making a detectable relationship even more certain. We can expect the reduction to be approximately the same for other types of relationship, although these were not simulated.

References

Bailey, N. T. J. 1961. "Introduction to the Theory of Genetic Linkage," Oxford Univ. (Clarendon) Press, Oxford.

Bennett, J. H. 1953. Junctions and inbreeding, *Genetica* **26**, 392–406.

Bennett, J. H. 1954. The distribution of heterogeneity upon inbreeding, *J. Roy. Statist. Soc.* **16**, 88–99.

Conneally, P. M., and Rivas, M. L. 1980. Linkage analysis in man, *Advan. Human Genet.* **10**, 209–266.

Donnelly, K. P. 1982. "Genetic Linkage, Detectable Relationship and Other Topics," Ph.D. thesis, University of Cambridge, England.

Edwards, A. W. F. 1967. Automatic construction of genealogies from phenotypic information (AUTOKIN), *Bull. Eur. Soc. Human Genet.* **1**, 42–43.

Fisher, R. A. 1949. "The Theory of Inbreeding," Oliver & Boyd, Edinburgh.

FISHER, R. A. 1954. A fuller theory of junctions. *Heredity* **8**, 187–197.
FISHER, R. A. 1959. An algebraically exact examination of junction formation and transmission in parent offspring breeding. *Heredity* **13**, 179–186.
FRANKLIN, I. R. 1977. The distribution of the proportion of the genome which is homozygous by descent in inbred individuals. *Theor. Pop. Biol.* **11**, 60–80.
GREEN, J. A. 1965. "Sets and Groups." Routledge & Kegan Paul, London.
KARLIN, S., AND TAYLOR, H. M. 1975. "A First Course in Stochastic Processes." Academic Press, New York.
MAYNARD-SMITH, S., PENROSE, L. A., AND SMITH, C. A. B. 1961. "Mathematical Tables for Research Workers in Human Genetics." Churchill, London.
NUMERICAL ALGORITHMS GROUP, 1978. NAG Fortran Library Manual, Mark 7, Numerical Algorithms Group, NAG Central Office, 7 Banbury Cross, Oxford, OX2 6NN, U.K.
PARZEN, E. 1962. "Stochastic Processes," Holden-Day, San Francisco.
RENWICK, J. H. 1971. The mapping of human chromosomes. *Ann. Rev. Genet.* **5**, 81–120.
SCHNELL, F. W. 1961. Some general formulations of linkage effects in inbreeding. *Genetics* **46**, 947–957.
SMITH, C. A. B. 1963. Discussion to J. Hajnal, "Concepts of random mating and the frequency of consanguineous marriages." *Proc. Roy. Soc. B* **159**, 176–177.
STAM, P. 1980. The distribution of the fraction of the genome identical by descent in finite random mating populations. *Genet. Res., Cambridge* **35**, 131–155.
THOMPSON, E. A. 1974a. "Mathematical Analysis of Human Evolution and Population Structure," Ph.D. Thesis, University of Cambridge.
THOMPSON, E. A. 1974b. Gene identities and multiple relationships. *Biometrics* **30**, 667–680.
THOMPSON, E. A. 1975. The estimation of pairwise relationships. *Ann. Hum. Genet.* **39**, 173–188.
THOMPSON, E. A. 1976. Inference of genealogical structure. *Soc. Sci. Inform.* **15**, 477–526.
WIELANDTH, H. 1964. "Finite Permutation Groups." Academic Press, New York.

Printed by the St. Catherine Press Ltd., Tempelhof 41, Bruges, Belgium