Behavior Genetics, Vol. 2, No. 1, 1972

(2)

The Investigation of Linkage Between a Quantitative Trait and a Marker Locus

J. K. Haseman[1] and R. C. Elston[2]

Received 16 Feb. 1970—Final 24 March 1970

Procedures are given, using sib pairs, for estimating linkage between a known m*-allele locus and a hypothesized two-allele locus that governs a quantitative trait. Random mating and linkage equilibrium are assumed. Also given are parametric and nonparametric methods for detecting linkage when the trait in question is governed by several two-allele loci, provided there is no epistasis.*

INTRODUCTION

It is not an easy matter in man to demonstrate unequivocally that a genetic component is involved in the determination of a behavioral trait. Twin studies can provide a certain amount of evidence, but since the methods of analysis are based upon assumptions that cannot be verified, such studies are often liable to overestimate the genetic component (Haseman and Elston, 1970). Classical segregation analysis can be tried for qualitative traits but cannot be used for quantitative traits. One method of demonstrating the existence of genetic control for a quantitative trait would be to determine a linkage relationship between that trait and a marker locus (i.e., a genetically defined polymorphic system), for it is difficult to see how environmental influences can simulate the effects of genetic linkage. In view of the increasing number of marker loci that are becoming available in man, this technique

---

This investigation was supported by a U.S. Public Health Service Training Grant (2 T01 GM00038), Research Career Development Award (1-K3-GM-31,732), and a research grant (GM HD 16697) from the National Institute of General Medical Sciences.

[1]National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina.

[2]Department of Biostatistics and the Genetics Curriculum, University of North Carolina, Chapel Hill, North Carolina.

will undoubtedly become of greater usefulness in human behavior–genetic studies. It may well be a useful technique for animal studies also, but the particular method we consider here is proposed specifically for studies in man, in which it is not possible to set up matings according to an experimental plan.

So far as the quantitative trait is concerned, we restrict our attention to data gathered on sib pairs. This eliminates the large biases, due to secular or age effects, that could occur if an attempt were made to utilize data on two or more generations simultaneously. The problem of detecting linkage between a quantitative trait and a marker locus from sib pair data was first considered by Penrose (1938). In the present paper, however, we allow for the incorporation of data on the sibs' parents with regard to the marker locus, for the phenotypic classification of an individual with respect to a marker locus does not usually depend upon the date or age at which it is determined. In addition, we consider the problem of estimating the recombination fraction between a major gene for the trait in question and the marker locus; we allow for multiple allelism at the marker locus but, in view of the numerical difficulties that would be involved in practice, not at the trait locus.

Letting x_{1j} and x_{2j} be the observed trait values for the first and second sibs, respectively, in the jth sib pair, we assume the general model

$$\left.\begin{aligned} x_{1j} &= \mu + g_{1j} + e_{1j} \\ x_{2j} &= \mu + g_{2j} + e_{2j} \end{aligned}\right\} \tag{1}$$

where μ is the overall mean and g_{ij} and e_{ij} are the genetic and environmental effects, respectively. However, it will be convenient, to begin with, to assume that only one locus determines g_{ij}; the generalization to the case where many loci are involved will be shown later to be a simple step if there is no epistasis. Suppose that just two alleles are involved, B and b, with gene frequencies p and q, respectively. Then we can define the genotypic values

$$\left.\begin{aligned} g_{ij} &= a \text{ for a } BB \text{ individual} \\ &= d \text{ for a } Bb \text{ individual} \\ &= -a \text{ for a } bb \text{ individual} \end{aligned}\right\} \tag{2}$$

The genetic variance at this locus, σ_g^2, is composed of an additive component, σ_a^2, and a dominance component, σ_d^2, and it is well known (see, for instance, Li, 1955) that under random mating these are given by

$$\sigma_a^2 = 2pq[a-d(p-q)]^2 \tag{3}$$

and

$$\sigma_d^2 = 4p^2q^2d^2 \tag{4}$$

We shall let $e_j = e_{1j} - e_{2j}$, and for convenience denote $E(e_j^2)$ by σ_e^2. Thus σ_e^2 is a function of the environmental variance, the environmental covariance

between sibs, and any order effect. Random mating will be assumed throughout.

In Section 1, we derive the expectation of the squared sib pair differences and show how this expectation, conditional on the proportion of genes the sibs have identical by descent (i.b.d.) at the trait locus, is a function of σ_e^2, σ_a^2, and σ_d^2. It follows as a direct consequence of this result that a simple regression procedure could be used to estimate σ_g^2 if this proportion were known for each sib pair. In Section 2, a procedure is given for estimating, for each sib pair, the proportion of genes i.b.d. at the marker locus. Then, in Section 3, it is shown that if this estimate replaces the proportion of genes i.b.d. at the trait locus in the regression procedure of Section 1, linkage can be detected. In Section 4, nonparametric methods for detecting linkage, based on the same general principle, are discussed. Finally, in Section 5, a maximum likelihood procedure is derived that under certain conditions permits estimation of both the genetic effect of a major trait locus and the recombination fraction between it and a marker locus.

One assumption that will be made is that of linkage equilibrium. The consequences of linkage disequilibrium will of course depend upon the cause of that disequilibrium, whether it is selection, assortative mating, or simply that equilibrium has not yet been reached. These possibilities will be investigated at a future date, with a view to obtaining the appropriate analysis in each case.

1. CONDITIONAL EXPECTATION OF THE SQUARED PAIR DIFFERENCES

Let $Y_j = (x_{1j} - x_{2j})^2$ be the squared pair difference for sib pair j. Then, for fixed e_j, Y_j can take on one of seven values depending upon the genotypes of the first and second sibs. These values, obtained from (1) and (2), are shown

Table I. Conditional Distribution of Y_j

Sib pair	Y_j	Conditional probability $\pi_j = 0$	$\pi_j = \frac{1}{2}$	$\pi_j = 1$
BB–BB	e_j^2	p^4	p^3	p^2
bb–bb	e_j^2	q^4	q^3	q^2
Bb–Bb	e_j^2	$4p^2q^2$	pq	$2pq$
BB–Bb	$(a-d+e_j)^2$	$2p^3q$	p^2q	0
Bb–BB	$(-a+d+e_j)^2$	$2p^3q$	p^2q	0
Bb–bb	$(a+d+e_j)^2$	$2pq^3$	pq^2	0
bb–Bb	$(-a-d+e_j)^2$	$2pq^3$	pq^2	0
BB–bb	$(2a+e_j)^2$	p^2q^2	0	0
bb–BB	$(-2a+e_j)^2$	p^2q^2	0	0

这个结果是等式等于无关。 See C.C. Li's ITO method

in the second column of Table I. For example, for the sib pair *BB–Bb* we have from (1) and (2)

$$x_{1j} = \mu + a + e_{1j}, \quad x_{2j} = \mu + d + e_{2j}$$

and hence

$$Y_j = (a + e_{1j} - d - e_{2j})^2 = (a - d + e_j)^2$$

In order to obtain the expectation of Y_j conditional on the proportion of genes i.b.d., we need to know the distribution of Y_j conditional on this proportion. Any particular sib pair must have zero, one, or two genes i.b.d. at the trait locus, and so the proportion of genes i.b.d. must be 0, $\frac{1}{2}$, or 1. Let the proportion be π_j for the jth sib pair. Then the conditional distribution of the sib pairs given π_j is shown in Table I.

When $\pi_j = 0$, the sibs are "unrelated" at the trait locus, and so the distribution of sib pairs is simply the same as the well-known distribution of matings in a random mating population. When $\pi_j = 1$, both sibs have the same genotype, and in that case the probability of the sib pair is simply the probability in the population of one of them. To obtain the distribution for $\pi_j = \frac{1}{2}$, we argue as follows: It is obvious that the pair *BB–bb* is impossible if the sibs are to have one gene i.b.d. The probability of the pair *BB–BB*, given one gene i.b.d., is simply the probability of three *B* genes occurring together, or p^3. For the case *BB–Bb*, the *B* gene in the second sib must be i.b.d. with a *B* gene in the first sib, and so the desired probability is simply that of *BB* and *b*, or p^2q. In the sib pair *Bb–Bb*, either *B* is the gene i.b.d., in which case the sib pair has probability pq^2, or *b* is, in which case the probability is p^2q; adding these two probabilities, we obtain pq. All the other probabilities are obtained analogously.

We can now use Table I to calculate the expected value of Y_j conditional on π_j. We have

$$E(Y_j|\pi_j = 1) = E\{e_j^2[p^2 + q^2 + 2pq]\} = E(e_j^2) = \sigma_e^2 \tag{5}$$

$$\begin{aligned} E(Y_j|\pi_j = \tfrac{1}{2}) &= E\{e_j^2[p^3 + q^3 + pq] + [(a-d+e_j)^2 + (-a+d+e_j)^2]p^2q \\ &\quad + [(a+d+e_j)^2 + (-a-d+e_j)^2]pq^2\} \\ &= \sigma_e^2 + (a^2+d^2)[2p^2q + 2pq^2] + 4ad[pq^2 - p^2q] \\ &= \sigma_e^2 + 2pq[a^2 + d^2 - 2ad(p-q)] \\ &= \sigma_e^2 + 2pq[a-(p-q)d]^2 + 2pqd^2[1-(p-q)^2] \\ &= \sigma_e^2 + \sigma_a^2 + 2pqd^2[4pq] \qquad \text{using (3)} \\ &= \sigma_e^2 + \sigma_a^2 + 2\sigma_d^2 \qquad \text{using (4)} \end{aligned} \tag{6}$$

and similarly it can be shown that

$$E(Y_j|\pi_j = 0) = \sigma_e^2 + 2\sigma_a^2 + 2\sigma_d^2 \tag{7}$$

It is clear from (5)–(7) that if there is no dominance ($d = 0$, or equivalently $\sigma_d^2 = 0$), we can write

$$E(Y_j|\pi_j) = (\sigma_e^2+2\sigma_g^2)-2\sigma_g^2\pi_j, \quad \pi_j = 0,\tfrac{1}{2},1 \tag{8}$$

This can be written in the form

$$E(Y_j|\pi_j) = \alpha+\beta\pi_j \tag{9}$$

where $\alpha = \sigma_e^2+2\sigma_g^2$ and $\beta = -2\sigma_g^2$. Thus if π_j were known and we fitted the simple linear regression model (9), then $-\frac{1}{2}\hat{\beta}$ would be an unbiased estimator of σ_g^2, where $\hat{\beta}$ is the usual least squares estimator of β. This same result will hold asymptotically even when dominance is present. It is shown in Appendix I that in this more general case

$$E(Y_j|\pi_j) = (\sigma_e^2+2\sigma_g^2)-\pi_j\{2\sigma_g^2+2n_1(n_2-n_0)\sigma_d^2/[4n_0n_2+n_0n_1+n_1n_2]\} \tag{10}$$

where n_i ($i = 0, 1, 2$) is the number of sib pairs in the sample that have the proportion $i/2$ genes i.b.d. at the trait gene locus. As the sample size increases, n_2 and n_0 tend to equality, and so the term in σ_d^2 vanishes asymptotically.

2. ESTIMATING π_j FOR A MARKER LOCUS

Let f_{ji} be the probability that the jth sib pair should have i genes i.b.d. at a marker locus, conditional on I_m, the information available on the sib pair and parental phenotypes at the marker locus. Then the estimator we shall use for π_j, the proportion of genes i.b.d. at the marker locus, is

$$\hat{\pi}_j = f_{j2}+\tfrac{1}{2}f_{j1} \tag{11}$$

As defined by (11), $\hat{\pi}_j$ is the Bayes estimate of π_j when a squared error loss function is used. It also has the property of having maximum possible correlation with π_j, when π_j is considered as a random variable taking on values 0, $\frac{1}{2}$, and 1. For a proof of both of these results, see Haseman (1970).

When sib and parental genotypes are both known, $\hat{\pi}_j$ can easily be calculated for every conceivable mating and sib pair. For the general multiallele system, there are seven mating types, and similarly seven sib pair types. We are here using the term *mating type*, and analogously the term *sib pair type*, in the same broad sense as does Kempthorne (1957). Thus $A_1A_1 \times A_1A_1$ and $A_2A_2 \times A_2A_2$ are two genotypically different matings of the same mating type, since they both involve identical homozygotes. All the possibilities, together with their probabilities and the appropriate values of $\hat{\pi}_j$, are given in Table II, in which it is assumed that the gene frequency of the allele A_i is p_i. Each number in parentheses in Table II is the number of genotypically different sib pairs of the indicated type that could result from the indicated mating. For example, a type IV mating may produce a type V sib pair that is either A_iA_j-A_iA_j or A_iA_k-A_iA_k. These are both equally likely type V

Table II. $\hat{\pi}_j$ When Both Parental and Sib Genotypes Are Known

Mating type	Sib pair type	Probability	f_{j0}	f_{j1}	f_{j2}	$\hat{\pi}_j$
I: $A_iA_i \times A_iA_i$	I: A_iA_i–A_iA_i	p_i^4	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2}$
II: $A_iA_i \times A_jA_j$	V: A_iA_j–A_iA_j	$2p_i^2p_j^2$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2}$
III: $A_iA_i \times A_iA_j$	I: A_iA_i–A_iA_i	$p_i^3p_j$	0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{3}{4}$
	III: A_iA_i–A_iA_j	$2p_i^3p_j$	$\frac{1}{2}$	$\frac{1}{2}$	0	$\frac{1}{4}$
	V: A_iA_j–A_iA_j	$p_i^3p_j$	0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{3}{4}$
IV: $A_iA_i \times A_jA_k$	V: (2)	$p_i^2p_jp_k$	0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{3}{4}$
	VI: A_iA_j–A_iA_k	$2p_i^2p_jp_k$	$\frac{1}{2}$	$\frac{1}{2}$	0	$\frac{1}{4}$
V: $A_iA_j \times A_iA_j$	I: (2)	$p_i^2p_j^2/4$	0	0	1	1
	II: A_iA_i–A_jA_j	$p_i^2p_j^2/2$	1	0	0	0
	III: (2)	$p_i^2p_j^2$	0	1	0	$\frac{1}{2}$
	V: A_iA_j–A_iA_j	$p_i^2p_j^2$	$\frac{1}{2}$	0	$\frac{1}{2}$	$\frac{1}{2}$
VI: $A_iA_j \times A_iA_k$	I: A_iA_i–A_iA_i	$p_i^2p_jp_k/2$	0	0	1	1
	III: (2)	$p_i^2p_jp_k$	0	1	0	$\frac{1}{2}$
	IV: A_iA_i–A_jA_k	$p_i^2p_jp_k$	1	0	0	0
	V: (3)	$p_i^2p_jp_k/2$	0	0	1	1
	VI: A_iA_j–A_iA_k	$p_i^2p_jp_k$	1	0	0	0
	VI: A_iA_j–A_jA_k	$p_i^2p_jp_k$	0	1	0	$\frac{1}{2}$
	VI: A_iA_k–A_jA_k					
VII: $A_iA_j \times A_kA_l$	V: (4)	$p_ip_jp_kp_l/2$	0	0	1	1
	VI: (4)	$p_ip_jp_kp_l$	0	1	0	$\frac{1}{2}$
	VII: (2)	$p_ip_jp_kp_l$	1	0	0	0

sib pairs, with identical f_{ji} values. Table II also gives the probability of each combination of mating and sib pair type; these probabilities are easily verifiable.

When some of the genotypes are unknown, the calculation of f_{ji} becomes more difficult. An algorithm that can be used for this purpose will now be given. The procedure is very general and can be used when no information is available as to the parental genotypes, or when there is dominance (i.e., we know phenotypes but not the genotypes).

Let P_{1p} and P_{2p} denote the *phenosets* (Cotterman, 1969) for the two parents; i.e., P_{1p} is the set of all genotypes that could give rise to the phenotype of one parent, and P_{2p} is the analogous set for the other parent. If there is no information as to the parental phenotypes, then the phenosets will consist of all possible genotypes. Let P_p denote the set consisting of all possible ordered pairs of genotypes resulting when an element of P_{1p} is paired with an element of P_{2p}. Thus, if there are N_1 genotypes in P_{1p} and N_2 genotypes in P_{2p}, then there are N_1N_2 elements in P_p. Similarly, let P_{1s} and P_{2s} denote the phenosets for the two sibs and P_s the set of all possible ordered pairs of genotypes from P_{1s} and P_{2s}. Let v and w be elements of P_p and P_s, respectively. Then

$$f_{ji} = \sum_{v \epsilon P_p} \sum_{w \epsilon P_s} P\{v \text{ and}$$

The numerator of t
and that π_j should e
joint probabilities fo

Each term in th
since it is the produc
of the correspondin
belong to P_{1p}, P_{2p}, i
belong to P_p and P_s.
result in the observe
that the observed sil
grammed in general
no parental informa
in Table III.

3. DERI

In Section 1, we
locus, π_j, is known f
given by (9) will re
$\sigma_d^2 = 0$. This estimat
is present. In the la
portion of genes i.b.
the regression analy
$\hat{\pi}_j$ for π_j in the regr
of a two-allele mar
mation,

where

and c is the recomb

We now disting
marker loci for sib p
and we denote by $\hat{\pi}$
equilibrium betwee

es Are Known

f_{j0}	f_{j1}	f_{j2}	$\hat{\pi}_j$
$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2}$
$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2}$
0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{3}{4}$
$\frac{1}{2}$	$\frac{1}{2}$	0	$\frac{1}{4}$
0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{3}{4}$
0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{3}{4}$
$\frac{1}{2}$	$\frac{1}{2}$	0	$\frac{1}{4}$
0	0	1	1
1	0	0	0
0	1	0	$\frac{1}{2}$
$\frac{1}{2}$	0	$\frac{1}{2}$	$\frac{1}{2}$
0	0	1	1
0	1	0	$\frac{1}{2}$
1	0	0	0
0	0	1	1
1	0	0	0
0	1	0	$\frac{1}{2}$
0	0	1	1
0	1	0	$\frac{1}{2}$
1	0	0	0

he probability of each
obabilities are easily

ulation of f_{ji} becomes
purpose will now be
hen no information is
is dominance (i.e., we

n, 1969) for the two
give rise to the pheno-
the other parent. If
s, then the phenosets
e set consisting of all
an element of P_{1p} is
genotypes in P_{1p} and
in P_p. Similarly, let
, the set of all possible
and w be elements of

$$f_{ji} = \sum_{v\epsilon P_p} \sum_{w\epsilon P_s} P\left\{v \text{ and } w \text{ and } \pi_j = \frac{i}{2}\right\} \Big/ \sum_{h=0}^{2} \sum_{v\epsilon P_p} \sum_{w\epsilon P_s} P\left\{v \text{ and } w \text{ and } \pi_j = \frac{h}{2}\right\},$$

$$i = 0,1,2 \qquad (12)$$

The numerator of this expression is the joint probability of observing I_m and that π_j should equal $i/2$; the denominator is the sum of the three such joint probabilities for $i = 0$, 1, and 2.

Each term in the summations in (12) can be obtained from Table II, since it is the product of one of the probabilities in the third column and one of the corresponding f_{ji}. It is necessary only to specify the genotypes that belong to P_{1p}, P_{2p}, P_{1s}, and P_{2s}, or equivalently the *pairs* of genotypes that belong to P_p and P_s. The elements of P_p are the possible matings that could result in the observed sib pair; the elements of P_s are the sib pair genotypes that the observed sib pair could assume. The calculation can easily be programmed in general for a computer. The special case of no dominance and no parental information, which permits an easy algebraic solution, is given in Table III.

3. DERIVING THE EXPECTED VALUE OF THE REGRESSION COEFFICIENT

In Section 1, we showed that if the proportion of genes i.b.d. at the trait locus, π_j, is known for each sib pair, then the simple linear regression model given by (9) will result in $-\beta/2$ being an unbiased estimate of σ_g^2 when $\sigma_d^2 = 0$. This estimate is also asymptotically unbiased even when dominance is present. In the last section, we have derived an estimate, $\hat{\pi}_j$, of the proportion of genes i.b.d. at a marker locus. In this section, we investigate how the regression analysis of Section 1 is affected if we substitute our estimate $\hat{\pi}_j$ for π_j in the regression equation. We shall show that for the special case of a two-allele marker locus, no dominance, and complete parental information,

$$E(Y_j|\hat{\pi}_j) = \alpha + \beta\hat{\pi}_j \qquad (13)$$

where

$$\beta = -2(1-2c)^2\sigma_g^2 \qquad (14)$$

and c is the recombination fraction between the trait and marker loci.

We now distinguish between the proportion of genes i.b.d. at the trait and marker loci for sib pair j, denoting these proportions π_{jt} and π_{jm}, respectively; and we denote by $\hat{\pi}_{jm}$ the estimate of π_{jm} given by (11). We assume linkage equilibrium between the marker and trait loci, so that for fixed π_{jt}, Y_j and

Table III. $\hat{\pi}_J$ When There Is No Dominance and the Parental Genotypes Are Unknown

Sib pair type	Probability	f_{J0}	f_{J1}	f_{J2}	$\hat{\pi}_J$
I: A_IA_I–A_IA_I	$p_I^2(1+p_I)^2/4$	$\frac{p^2_I}{(1+p_I)^2}$	$\frac{2p_I}{(1+p_I)^2}$	$\frac{1}{(1+p_I)^2}$	$\frac{1}{(1+p_I)}$
II: A_IA_I–A_JA_J	$p_I^2p_J^2/2$	1	0	0	0
III: A_IA_I–A_IA_J	$p_I^2p_J(1+p_I)$	$\frac{p_I}{1+p_I}$	$\frac{1}{1+p_I}$	0	$\frac{1}{2(1+p_I)}$
IV: A_IA_I–A_JA_K	$p_I^2p_Jp_K$	1	0	0	0
V: A_IA_J–A_IA_J	$\frac{1}{2}p_Ip_J(1+p_I+p_J+2p_Ip_J)$	$\frac{2p_Ip_J}{(1+p_I+p_J+2p_Ip_J)}$	$\frac{p_I+p_J}{(1+p_I+p_J+2p_Ip_J)}$	$\frac{1}{(1+p_I+p_J+2p_Ip_J)}$	$\frac{2+p_I+p_J}{2(1+p_I+p_J+2p_Ip_J)}$
VI: A_IA_J–A_IA_K	$p_Ip_Jp_K(1+2p_I)$	$\frac{2p_I}{1+2p_I}$	$\frac{1}{1+2p_I}$	0	$\frac{1}{2(1+2p_I)}$
VII: A_IA_J–A_KA_I	$2p_Ip_Jp_Kp_I$	1	0	0	0

T

π_{Jt}	
0	
$\frac{1}{2}$	Ψ(
1	(
Total	

$\hat{\pi}_{jm}$ are independent
It follows that

$E(Y_j|$

where the summatio

$E(Y_j|\pi_{jt})$ is giv
derived in Appendix
The joint distributi
marker gene, no d
derived easily from
if and only if there
with probability p^2q
been AA–AA or aa-
Similarly, the remai
and the joint probal

Table V. Joint Distrib
D

	0
$\hat{\pi}_{Jm}$	
0	$\frac{1}{2}p^2q^2$
$\frac{1}{4}$	p^3q+pq^3
$\frac{1}{2}$	$\frac{1}{4}(p^4+4p^2q^2+$
$\frac{3}{4}$	0
1	0
Total	$\frac{1}{4}$

Table IV. Joint Distribution of π_{jm} and π_{jt}

	π_{jm}			
	0	½	1	Total
π_{jt}				
0	$\Psi^2/4$	$\Psi(1-\Psi)/2$	$(1-\Psi)^2/4$	¼
½	$\Psi(1-\Psi)/2$	$(1-2\Psi+2\Psi^2)/2$	$\Psi(1-\Psi)/2$	½
1	$(1-\Psi)^2/4$	$\Psi(1-\Psi)/2$	$\Psi^2/4$	¼
Total	¼	½	¼	1

$\hat{\pi}_{jm}$ are independent, and also for fixed π_{jm}, π_{jt} and $\hat{\pi}_{jm}$ are independent. It follows that

$$E(Y_j|\hat{\pi}_{jm}) = \sum_{\pi_{jt}} E\{Y_j|\pi_{jt}\}P\{\pi_{jt}|\hat{\pi}_{jm}\}$$
$$= \sum_{\pi_{jt}}\sum_{\pi_{jm}} E(Y_j|\pi_{jt})P\{\pi_{jt}|\pi_{jm}\}P\{\pi_{jm}|\hat{\pi}_{jm}\} \qquad (15)$$

where the summations are over the three values that π_{jt} and π_{jm} can assume.

$E(Y_j|\pi_{jt})$ is given by (5)–(7). The joint distribution of π_{jt} and π_{jm} is derived in Appendix II and is given in Table IV, in which $\Psi = c^2 + (1-c)^2$. The joint distribution of π_{jm} and $\hat{\pi}_{jm}$, for the special case of a two-allele marker gene, no dominance, and complete parental information can be derived easily from Table II. For example, we see from Table II that $\hat{\pi}_{jm} = 0$ if and only if there is an $Aa \times Aa$ mating and an AA–aa sib pair, an event with probability $p^2q^2/2$. Naturally, $\pi_{jm} = 0$ for this sib pair. If the sibs had been AA–AA or aa–aa (each with probability $p^2q^2/4$), then $\hat{\pi}_{jm} = \pi_{jm} = 1$. Similarly, the remaining matings and resulting sib pairs can be examined, and the joint probability is found to be as given in Table V.

Table V. Joint Distribution of $\hat{\pi}_{jm}$ and π_{jm} for a Two-Allele Marker Locus with No Dominance and Complete Parental Information

	π_{jm}			
	0	½	1	Total
$\hat{\pi}_{jm}$				
0	$\frac{1}{2}p^2q^2$	0	0	$\frac{1}{2}p^2q^2$
¼	p^3q+pq^3	p^3q+pq^3	0	$2(p^3q+pq^3)$
½	$\frac{1}{4}(p^4+4p^2q^2+q^4)$	$\frac{1}{2}(p^4+6p^2q^2+q^4)$	$\frac{1}{4}(p^4+4p^2q^2+q^4)$	$(p^4+5p^2q^2+q^4)$
¾	0	p^3q+pq^3	p^3q+pq^3	$2(p^3q+pq^3)$
1	0	0	$\frac{1}{2}p^2q^2$	$\frac{1}{2}p^2q^2$
Total	¼	½	¼	1

From Tables IV and V and from (5)–(7) and (15), we have

$$\begin{aligned}E(Y_{jt}|\hat{\pi}_{jm}=0\} &= \sigma_e^2[(1-\Psi)^2(1)+\Psi(1-\Psi)(0)+\Psi^2(0)]\\ &\quad+[\sigma_e^2+\sigma_g^2][2\Psi(1-\Psi)(1)+(1-2\Psi+2\Psi^2)(0)+2\Psi(1-\Psi)(0)]\\ &\quad+[\sigma_e^2+2\sigma_g^2][\Psi^2(1)+\Psi(1-\Psi)(0)+(1-\Psi)^2(0)]\\ &= \sigma_e^2+\sigma_g^2[2\Psi-2\Psi^2+2\Psi^2]=\sigma_e^2+2\Psi\sigma_g^2\end{aligned} \tag{16}$$

Also,

$$\begin{aligned}E(Y_{jt}|\hat{\pi}_{jm}=\tfrac{1}{4}) &= \sigma_e^2[(1-\Psi)^2(\tfrac{1}{2})+\Psi(1-\Psi)\tfrac{1}{2}+\Psi^2(0)]\\ &\quad+[\sigma_e^2+\sigma_g^2][2\Psi(1-\Psi)(\tfrac{1}{2})+(1-2\Psi+2\Psi^2)\tfrac{1}{2}+2\Psi(1-\Psi)0]\\ &\quad+[\sigma_e^2+2\sigma_g^2][\Psi^2(\tfrac{1}{2})+\Psi(1-\Psi)\tfrac{1}{2}+(1-\Psi)^2(0)]\\ &= \sigma_e^2+(\tfrac{1}{2}+\Psi)\sigma_g^2\end{aligned} \tag{17}$$

Similarly, it can be shown that

$$E(Y_{jt}|\hat{\pi}_{jm}=\tfrac{1}{2})=\sigma_e^2+\sigma_g^2 \tag{18}$$

$$E(Y_{jt}|\hat{\pi}_{jm}=\tfrac{3}{4})=\sigma_e^2+(\tfrac{3}{2}-\Psi)\sigma_g^2 \tag{19}$$

$$E(Y_{jt}|\hat{\pi}_{jm}=1)=\sigma_e^2+2(1-\Psi)\sigma_g^2 \tag{20}$$

Thus, from (16)–(20), we see that

$$\begin{aligned}E(Y_{jt}|\hat{\pi}_{jm}) &= [\sigma_e^2+2\Psi\sigma_g^2]+2(1-2\Psi)\sigma_g^2\,\hat{\pi}_{jm}\\ &= [\sigma_e^2+2(1-2c+2c^2)\sigma_g^2]-2(1-2c)^2\sigma_g^2\,\hat{\pi}_{jm}\end{aligned} \tag{21}$$

This result is shown by Haseman (1970) to hold also in the case of a multiallele marker locus, provided there is no dominance at the trait locus. When there is dominance at the trait locus (13), (14) and (21) are approximate rather than exact, but the bias is small for large samples. Thus the regression analysis described in Section 1 can be used with $\hat{\pi}_{jm}$ replacing π_{jr}, and the hypothesis that $\beta = 0$ can be tested approximately by comparing the calculated $\hat{\beta}$ to its estimated standard error; a significantly large $|\hat{\beta}|$ indicates that $c \neq \frac{1}{2}$, and so linkage is present. Note, however, that it is only possible to detect linkage, not to estimate c, using this regression procedure.

Finally, suppose that there are K trait loci, each linked to the marker locus; then (14) will hold for each trait locus separately, and if the trait loci are mutually unlinked and there is no epistasis

$$E(\hat{\beta}) = -2\sum_{i=1}^{K}(1-2c_i)^2\sigma_i^2 \tag{22}$$

where σ_i^2 is the contribution to the total genetic variance of the ith trait locus and c_i is the recombination fraction between it and the marker locus. (The equality is exact if there is no dominance.)

An even stronger result holds if linkage equilibrium among the trait loci is assumed. At linkage equilibrium, the genetic effects at two loci are

independent, whicl
traits are linked, a
(i.e., there is no epi
is a linkage relatio
loci.

4. DETECTING

If there is a majo
definite (inverse) ass
π_j, the proportion c
if there is no major t
be independent. H
Kendall's τ and Spe
analysis, π_j is replac

First, $\hat{\pi}_j$ and $|x$
tied scores being assi
are then calculated b
(1956) and Kendall
Siegel, 1956), and a
either a relatively larg
or that there is a sma

This test proced
$\hat{\pi}_j$ and the sib pair
assumptions for the
being a nonparametr
relatively large sampl

5. MAXIMUM

One disadvantage
with the recombinati
detected it can not be
likelihood (ML) techn

We assume that
given by (2), located a
We denote by π_{jm} an
trait loci, respectively,
and marker loci and a
More precisely, we assu
as a mixture of up to
given in Table I, we s

we have

$'^2)(0)+2\Psi(1-\Psi)(0)]$
$\Psi)^2(0)]$
(16)

$'^2)\frac{1}{2}+2\Psi(1-\Psi)0]$
$')^2(0)]$
(17)

(18)

(19)

(20)

$)^2\sigma_g^2\hat{\pi}_{jm}$ (21)

so in the case of a
e at the trait locus.
d (21) are approxi-
samples. Thus the
with $\hat{\pi}_{jm}$ replacing
ately by comparing
ificantly large $|\hat{\beta}|$
ever, that it is only
ression procedure.
ked to the marker
nd if the trait loci

(22)

the ith trait locus
arker locus. (The

among the trait
ts at two loci are

independent, which implies that (22) will hold at equilibrium even if the traits are linked, as long as the effects at the different loci are additive (i.e., there is no epistasis). Thus a significantly large $|\hat{\beta}|$ indicates that there is a linkage relationship between the marker locus and one or more trait loci.

4. DETECTING LINKAGE BY NONPARAMETRIC METHODS

If there is a major trait gene located near a marker, then there should be a definite (inverse) association between the sib pair difference $|x_{1j}-x_{2j}|$ and π_j, the proportion of genes i.b.d. at the marker locus. On the other hand, if there is no major trait gene near the marker, then $|x_{ij}-x_{2j}|$ and π_j should be independent. Hence standard rank correlation procedures, such as Kendall's τ and Spearman's ρ, can be used as a test of such linkage. In the analysis, π_j is replaced by its estimate $\hat{\pi}_j$, as defined by (11).

First, $\hat{\pi}_j$ and $|x_{1j}-x_{2j}|$ are separately ranked in order of magnitude, tied scores being assigned the average of the tied ranks. The rank correlations are then calculated by the standard formulas given, for example, by Siegel (1956) and Kendall (1955). Tables of critical values are available (e.g., Siegel, 1956), and a significant negative correlation implies that there is either a relatively large genetic effect at a moderate distance from the marker or that there is a smaller genetic effect close to the marker.

This test procedure is easy to apply, requiring only the calculation of $\hat{\pi}_j$ and the sib pair differences. Furthermore, it requires no distributional assumptions for the trait of interest. The primary disadvantage is that, being a nonparametric test with π_j estimated by $\hat{\pi}_j$, it is likely to require relatively large samples in order to detect anything but fairly close linkage.

5. MAXIMUM LIKELIHOOD ESTIMATION OF LINKAGE

One disadvantage of the methods discussed above is that σ_g^2 is confounded with the recombination fraction c, and hence although linkage can be *detected* it can not be *estimated*. In this section, we show how maximum likelihood (ML) techniques can be used to overcome this difficulty.

We assume that there is a two-allele trait locus, with genetic effects given by (2), located at a linkage distance c from a multiallele marker locus. We denote by π_{jm} and π_{jt} the proportion of genes i.b.d. at marker and trait loci, respectively, for sib pair j. We assume linkage equilibrium for trait and marker loci and also that sib pair differences are normally distributed. More precisely, we assume that, for each value of π_{jt}, $x_{1j}-x_{2j}$ is distributed as a mixture of up to seven normal distributions. From the values of Y_j given in Table I, we see immediately that if $E(e_j) = 0$ the means of these

seven distributions are 0, $a-d$, $a+d$, $-a+d$, $-a-d$, $2a$, and $-2a$, depending upon the sib pair genotypes. We shall assume $E(e_j) = 0$, i.e., that the data have been corrected for any sib order effect, and so the variance of each distribution is σ_e^2.

Without loss of generality, we can reduce the number of distributions from seven to four by considering only the absolute pair differences; for example, no distinction is made between a *BB–Bb* and a *Bb–BB* sib pair. This is reasonable, since the order of the sibs' scores is unimportant if we correct for age. Thus henceforth we consider only the absolute differences $D_j = |x_{1j} - x_{2j}|$.

The distribution of D_j, conditional on the sib pair genotypes at the trait locus, is simply twice the conditional distribution of $x_{1j} - x_{2j}$, but limited to the range $x_{1j} - x_{2j} \geqslant 0$. Thus we can write it as

$$f_1 = f(D_j|\text{sibs } BB\text{–}BB, bb\text{–}bb, \text{or } Bb\text{–}Bb) = (1/\sigma_e)(2/\pi)^{\frac{1}{2}} \exp(-D_j^2/2\sigma_e^2), \quad D_j \geqslant 0 \tag{23}$$
$$= 0, \quad \text{otherwise}$$

$$f_2 = f(D_j|\text{sibs } BB\text{–}Bb \text{ or } Bb\text{–}BB) = (1/\sigma_e)(2/\pi)^{\frac{1}{2}} \exp[-(D_j - a + d)^2/2\sigma_e^2], \quad D_j \geqslant 0 \tag{24}$$
$$= 0, \quad \text{otherwise}$$

$$f_3 = f(D_j|\text{sibs } Bb\text{–}bb \text{ or } bb\text{–}Bb) = (1/\sigma_e)(2/\pi)^{\frac{1}{2}} \exp[-(D_j - a - d)^2/2\sigma_e^2], \quad D_j \geqslant 0 \tag{25}$$
$$= 0, \quad \text{otherwise}$$

$$f_4 = f(D_j|\text{sibs } BB\text{–}bb \text{ or } bb\text{–}BB) = (1/\sigma_e)(2/\pi)^{\frac{1}{2}} \exp[-(D_j - 2a)^2/2\sigma_e^2], \quad D_j \geqslant 0 \tag{26}$$
$$= 0, \quad \text{otherwise}$$

If we knew the sib pair genotypes at locus B for all sib pairs, the likelihood function could be easily constructed. Instead, we have information on the sib pair phenotypes at the *marker* locus and possibly, in addition, the phenotypes of one or both parents at this locus. We now show how this information at the marker locus, I_m, can be used to obtain the likelihood function for an observed sib pair.

The likelihood function for sib pair j may be written

$$L = f(D_j|I_m) = f(D_j \text{ and } I_m)/P\{I_m\}$$
$$= \sum_{\pi_{jt}} f(D_j \text{ and } I_m|\pi_{jt}) P\{\pi_{jt}\}/P\{I_m\} = \sum_{\pi_{jt}} f(D_j|\pi_{jt}) P\{I_m|\pi_{jt}\} P\{\pi_{jt}\}/P\{I_m\}$$

(because of linkage equilibrium)

$$= \sum_{\pi_{jt}} f(D_j|\pi_{jt})$$
$$= \sum_{\pi_{jt}} \sum_{\pi_{jm}} f(D_j|\pi$$

$$= \sum_{\pi_{jt}} \sum_{\pi_{jm}} f(D_j|\pi$$
$$= \sum_{h=0}^{2} \sum_{k=0}^{2} f(l$$

Apart from a
Table IV, and $P\{$
in simple cases, or
use of (12).

We now find $f($
buted as a mixture

$$m_{hi}$$

we have

$$f(D_j|\pi_{jt} = h$$

The coefficients m_{hi}
VI. Thus

$$m_{01} = \Pr(\text{sibs}$$

$$m_{02} = \Pr(\text{sibs}$$

Thus $f(D_j|\pi_{jt}$
VI, and hence all ele
five parameters in t
frequencies at the m
estimated.) After M
the estimated addit
stituting the parame

h
0
1
2

$$= \sum_{\pi_{jt}} f(D_j|\pi_{jt}) \left[\sum_{\pi_{jm}} P\{I_m \text{ and } \pi_{jt}|\pi_{jm}\} P\{\pi_{jm}\}/P\{I_m\} \right]$$

$$= \sum_{\pi_{jt}} \sum_{\pi_{jm}} f(D_j|\pi_{jt}) P(I_m|\pi_{jm}) P(\pi_{jt}|\pi_{jm}) P(\pi_{jm})/P(I_m)$$

(because of linkage equilibrium)

$$= \sum_{\pi_{jt}} \sum_{\pi_{jm}} f(D_j|\pi_{jt}) P(\pi_{jt}|\pi_{jm}) P(\pi_{jm}|I_m)$$

$$= \sum_{h=0}^{2} \sum_{k=0}^{2} f(D_j|\pi_{jt} = h/2) P\{\pi_{jt} = h/2|\pi_{jm} = k/2\} \Pr(\pi_{jm} = k/2|I_m) \quad (27)$$

Apart from a factor of 2 or 4, $P\{\pi_{jt} = h/2|\pi_{jm} = k/2\}$ is given in Table IV, and $P\{\pi_{jm} = k/2|I_m\}$ can be obtained from Tables II or III in simple cases, or found numerically in more complex situations by the use of (12).

We now find $f(D_j|\pi_{jt} = h/2)$. Note that D_j, conditional on π_{jt}, is distributed as a mixture of the four distributions given by (23)–(26), i.e., letting

$$m_{hi} = P(D_j \text{ has density function } f_i|\pi_{jt} = h/2)$$

$$(h = 0, 1, 2 \text{ and } i = 1, 2, 3, 4)$$

we have

$$f(D_j|\pi_{jt} = h/2) = m_{h1}f_1 + m_{h2}f_2 + m_{h3}f_3 + m_{h4}f_4, \quad h = 0, 1, 2 \quad (28)$$

The coefficients m_{hi} can be calculated from Table I and are given in Table VI. Thus

$$m_{01} = \Pr(\text{sibs } BB\text{–}BB, bb\text{–}bb, \text{ or } Bb\text{–}Bb|\pi_{jt} = 0) = p^4 + q^4 + 4p^2q^2$$

$$m_{02} = \Pr(\text{sibs } BB\text{–}Bb \text{ or } Bb\text{–}BB|\pi_{jt} = 0) = 4p^3q, \quad \text{etc.}$$

Thus $f(D_j|\pi_{jt} = h/2)$ can be obtained using (23)–(26), (28), and Table VI, and hence all elements in the likelihood (27) can be calculated. There are five parameters in the likelihood function: c, p, σ_e^2, a, and d. (If the gene frequencies at the marker locus are unknown, they too can be appropriately estimated.) After ML estimates of these five parameters have been obtained, the estimated additive and dominance variance can be calculated by substituting the parameter estimates for the true values in (3) and (4).

Table VI. Values of the Coefficient m_{hi}

h \ i	1	2	3	4
0	$p^4+4p^2q^2+q^4$	$4p^3q$	$4pq^3$	$2p^2q^2$
1	$1-2pq$	$2p^2q$	$2pq^2$	0
2	1	0	0	0

Because of the complexity of the likelihood function, computer methods must be used in order to find the ML estimates. Note, however, that little information is needed beyond that already supplied for the computer calculation of $\hat{\pi}_{jm}$. The only additional information required in order to evaluate the likelihood function (27) are Tables IV and VI and the density functions (23)–(26). In fact, note that these additional quantities, $P(\pi_{jt}|\pi_{jm})$ and $f(D_j|\pi_{jt})$, do not in any way depend upon what is observed at the marker locus and hence are constant for all sib pairs. Thus the only probabilities in the likelihood function that vary from sib pair to sib pair are those given by (12). Once the likelihood has been programmed, various methods are available for calculating the ML estimates; the simplest is to search the likelihood surface directly, as explained elsewhere (Kaplan and Elston, 1972).

Finally, it might be noted that if we make the simplifying assumption that $\sigma_d^2 = 0$, then the number of distributions involved is reduced from four to three, and the number of parameters to be estimated is reduced from five to four. The ML procedure described above can be modified accordingly to permit estimation of c, p, σ_e^2, and a.

APPENDIX A

Consider the simple linear regression model (9) in which we regress the squared pair differences Y_j on π_j, which is assumed to be known. Suppose that of the n sib pairs used in the analysis, n_i ($i = 0, 1, 2$) have $i/2$ of their genes i.b.d. at the trait locus. Then in matrix notation we can write

$$E(\mathbf{y}) = \mathbf{A}\gamma$$

where $\mathbf{y}$ is a column vector whose n elements are Y_j; $\mathbf{A}$ is a $n \times 2$ matrix whose first n_2 rows are (1,1), whose next n_1 rows are $(1,\frac{1}{2})$, and whose last n_0 rows are (1,0); and γ is a column vector whose two elements are α and β. It is well known (e.g., Graybill, 1961) that $\hat{\gamma}$, the least squares estimator of γ, can be obtained by solving the system of equations

$$\mathbf{A}'\mathbf{A}\hat{\gamma} = \mathbf{A}'\mathbf{y}$$

so that

$$\mathbf{A}'\mathbf{A}E(\hat{\gamma}) = E(\mathbf{A}'\mathbf{y})$$

In this particular case, we have

$$\mathbf{A}'\mathbf{A} = \begin{bmatrix} n & n_2 + \frac{n_1}{2} \\ n_2 + \frac{n_1}{2} & n_2 + \frac{n_1}{4} \end{bmatrix}$$

$$E(\hat{\gamma}) = E\begin{bmatrix}\hat{\alpha}\\ \hat{\beta}\end{bmatrix},$$

$$E(\mathbf{A}'\mathbf{y}) = \begin{pmatrix} n\sigma_e^2+(n_1+2n_0)\sigma_a^2+2(n_1+n_0)\sigma_d^2 \\ (n_2+n_1/2)\sigma_e^2+(n_1/2)\sigma_a^2+n_1\sigma_d^2 \end{pmatrix} =$$

$$\begin{pmatrix} n\sigma_e^2+(n_1+2n_0)\sigma_g^2+n_1\sigma_d^2 \\ (n_2+n_1/2)\sigma_e^2+(n_1/2)\sigma_g^2+(n_1/2)\sigma_d^2 \end{pmatrix}$$

Thus to find $E(\hat{\beta})$, we solve

$$\begin{cases} nE(\hat{\alpha})+(n_2+n_1/2)E(\hat{\beta}) = n\sigma_e^2+(n_1+2n_0)\sigma_g^2+n_1\sigma_d^2 \\ (n_2+n_1/2)E(\hat{\alpha})+(n_2+n_1/4)E(\hat{\beta}) = (n_2+n_1/2)\sigma_e^2+(n_1/2)(\sigma_g^2+\sigma_d^2) \end{cases}$$

Eliminating $E(\hat{\alpha})$, we see that

$$E(\hat{\beta})\{(n_2+n_1/2)^2-(n_2+n_1/4)n\} = \sigma_g^2[(n_1+2n_0)(n_2+n_1/2)-n_1n/2] + \sigma_d^2\{n_1(n_2+n_1/2)-n_1n/2\}$$

which reduces to

$$-[E(\hat{\beta})/4][4n_2n_0+n_1n_2+n_1n_0] = = (\sigma_g^2/2)[4n_2n_0+n_1n_2+n_1n_0]+[-n_1n_0/2+n_1n_2/2]\sigma_d^2$$

or

$$E(\hat{\beta}) = -2\sigma_g^2-2n_1(n_2-n_0)\sigma_d^2/(4n_2n_0+n_1n_2+n_1n_0)$$

APPENDIX B

Consider a general mating that at two loci A and B is

$$\frac{A_1B_1}{A_2B_2} \times \frac{A_3B_3}{A_4B_4}$$

Let c be the recombination fraction (assumed the same for both sexes) between these two loci. Then the gametic frequencies are

Parent I		Parent II	
Gamete	frequency	Gamete	frequency
A_1B_1	$(1-c)/2$	A_3B_3	$(1-c)/2$
A_2B_2	$(1-c)/2$	A_4B_4	$(1-c)/2$
A_1B_2	$c/2$	A_3B_4	$c/2$
A_2B_1	$c/2$	A_4B_3	$c/2$

Suppose that two sibs result from the above mating and we wish to find $P\{\pi_{jm} = \pi_{jt} = 1\}$ in these sibs, where π_{jm} and π_{jt} are the proportions of genes these sibs have i.b.d. at the A and B loci, respectively. This probability can be found by summing the squares of all 16 zygote frequencies formed when a gamete frequency from parent I is multiplied by a gamete frequency from parent II. For example, one way that $\pi_{jm} = \pi_{jt} = 1$ is for both sibs

to be A_1B_1/A_3B_3, and this probability is $[(1-c)/2]^2[(1-c)/2]^2 = (1-c)^4/16$. The remaining 15 probabilities are calculated similarly, and so

$$\begin{aligned} P\{\pi_{jm} = \pi_{jt} = 1\} &= 4(c^4/16)+8[c^2(1-c)^2/16]+4[(1-c)^4/16] \\ &= [c^4+2c^2(1-c)^2+(1-c)^4]/4 \\ &= [c^2+(1-c)^2]^2/4 = \Psi^2/4 \end{aligned}$$

where

$$\Psi = c^2+(1-c)^2$$

By symmetry we have

$$P\{\pi_{jm} = \pi_{jt} = 0\} = P\{\pi_{jm} = \pi_{jt} = 1\} = \Psi^2/4$$

which can also be established by summing the appropriate cross-product frequencies.

We now find $P\{\pi_{jm} = 1 \text{ and } \pi_{jt} = 0\}$. Note that $\pi_{jm} = 1$ and $\pi_{jt} = 0$ if, for example, the first sib is A_1B_1/A_3B_3, the second A_1B_2/A_3B_4. The probability of this sib pair is

$$[(1-c)/2][(1-c)/2][c/2][c/2] = c^2(1-c)^2/16$$

There are 15 other sib pairs that could result in $\pi_{jm} = 1$ and $\pi_{jt} = 0$, and all 15 have the same probability $c^2(1-c)^2/16$. Hence

$$\begin{aligned} P\{\pi_{jm} = 1 \text{ and } \pi_{jt} = 0\} &= 16[c^2(1-c)^2/16] \\ &= c^2(1-c)^2 \\ &= (1-\Psi)^2/4 \end{aligned}$$

By symmetry we have

$$\Pr(\pi_{jm} = 0 \text{ and } \pi_{jt} = 1) = \Pr(\pi_{jm} = 1 \text{ and } \pi_{jt} = 0) = (1-\Psi)^2/4$$

Since we know that the marginal distribution of π_{jm} (and π_{jt}) is given by

$$P\{\pi_{jm}\} = \begin{cases} 1/4 & \text{if } \pi_{jm} = 0 \\ 1/2 & \text{if } \pi_{jm} = \frac{1}{2} \\ 1/4 & \text{if } \pi_{jm} = 1 \end{cases}$$

the remaining probabilities in the joint distribution of π_{jm} and π_{jt} can be obtained by subtraction. For example,

$$\begin{aligned} P\{\pi_{jm} = 1 \text{ and } \pi_{jt} = 1/2\} &= P\{\pi_{jm} = 1\} - P\{\pi_{jm} = 1 \text{ and } \pi_{jt} = 0\} \\ &\quad - P\{\pi_{jm} = 1 \text{ and } \pi_{jt} = 1\} \\ &= \tfrac{1}{4} - \Psi^2/4 - (1-\Psi)^2/4 = \tfrac{1}{4}(1-\Psi^2-1+2\Psi-\Psi^2) \\ &= \Psi(1-\Psi)/2 \end{aligned}$$

The other probabilities in Table IV can be obtained likewise.

Let S be # of genes two sibs shared i.e ibd then
S has distribution $\begin{matrix} 2 & 1 & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{matrix} \Rightarrow \therefore ES = 1$ *

Linkage Between Quanti

Cotterman, C. W. (196
Computer Applicati
Graybill, F. A. (1961).
York.
Haseman, J. K. (1970).
data. Unpublished F
Haseman, J. K., and Els
data. *Behav. Genet.*
Kaplan, E. B., and Elsto
Estimation (MAXL
Carolina.
Kempthorne, O. (1957).
Kendall, M. G. (1955). *R*
Li, C. C. (1955). *Populati*
Penrose, L. S. (1938).
133–138.
Siegel, S. (1956). *Nonpa*
New York.

Suppose mating

sib

A_1 A_3

sib 2

A_1A_3	2
A_1A_4	1
A_2A_3	1
A_2A_4	0

P{ Two sibs sh

P{ Two sibs sha

P{ Two sibs sh

c

← continuous

$c)/2]^2 = (1-c)^4/16.$
and so
$[(1-c)^4/16]$

$\Psi^2/4$

priate cross-product

$\tau_{jm} = 1$ and $\pi_{jt} = 0$
id A_1B_2/A_3B_4. The

$^2/16$

and $\pi_{jt} = 0$, and all

$/16]$

$) = (1-\Psi)^2/4$

$_m$ (and π_{jt}) is given by

f π_{jm} and π_{jt} can be

$\pi_{jt} = 0\}$

$-\Psi^2-1+2\Psi-\Psi^2)$

kewise.

b d then

$= 1$

REFERENCES

Cotterman, C. W. (1969). Factor-union phenotype systems. In Morton, N. E. (ed.), *Computer Applications in Genetics*, University of Hawaii Press, Honolulu, pp. 1–19.

Graybill, F. A. (1961). *An Introduction to Linear Statistical Models*, McGraw-Hill, New York.

Haseman, J. K. (1970). The genetic analysis of quantitative traits using twin and sib data. Unpublished Ph.D. thesis, University of North Carolina.

Haseman, J. K., and Elston, R. C. (1970). The estimation of genetic variance from twin data. *Behav. Genet.* **1**: 11–20.

Kaplan, E. B., and Elston, R. C. (1972). *A Subroutine Package for Maximum Likelihood Estimation (MAXLIK)*. Institute of Statistics Mimeo Series, University of North Carolina.

Kempthorne, O. (1957). *An Introduction to Genetic Statistics*, John Wiley, New York.

Kendall, M. G. (1955). *Rank Correlation Methods*, Charles Griffin, London.

Li, C. C. (1955). *Population Genetics*, University of Chicago Press, Chicago.

Penrose, L. S. (1938). Genetic linkage in graded human characters. *Ann. Eugen.* **6**: 133–138.

Siegel, S. (1956). *Nonparametric Statistics for the Behavioral Sciences*, McGraw Hill, New York.

Suppose mating are $A_1A_2 \times A_3A_4$

sib1

sib2 \ sib1	A_1A_3	A_1A_4	A_2A_3	A_2A_4
A_1A_3	2	1	1	0
A_1A_4	1	2	0	1
A_2A_3	1	0	2	1
A_2A_4	0	1	1	2

number of genes shared by two sibs

$P\{\text{Two sibs shared 2 genes} / A_1A_2 \times A_3A_4\} = \frac{\#\text{ of 2's}}{16} = \frac{4}{16} = \frac{1}{4}$

$P\{\text{Two sibs share 1 gene} / A_1A_2 \times A_3A_4\} = \frac{\#\text{ of 1's}}{16} = \frac{8}{16} = \frac{1}{2}$

$P\{\text{Two sibs share 0 gene} / A_1A_2 \times A_3A_4\} = \frac{\#\text{ of 0's}}{16} = \frac{4}{16} = \frac{1}{4}$

← continuous