The Annals of Probability
1973, Vol. 1, No. 5, 758-771

AN INDEX OF GENEALOGICAL RELATEDNESS DERIVED FROM A GENETIC MODEL

BY STUART P. LLOYD AND COLIN L. MALLOWS

Bell Laboratories, Murray Hill

The coefficient de parenté $\sigma(x, y)$ of Malécot, which measures the genealogical relatedness of individuals x and y, is derived as a correlation coefficient of certain genotype scores. An explicit formula is given for $\sigma(x, y)$ involving path sums in the ancestral charts of x and y. The expected value of $\sigma(x, y)$ is worked out for certain random population models.

1. Introduction. The notion of a genealogical distance function, measuring the degree of unrelatedness of two individuals in a population, is due to Mycielski and Ulam. These authors define various such distance functions in [7], and compute the expected values in a certain random population model. Kahane and Marr in [2] extend the considerations of [7] to a somewhat more general model.

One of the Mycielski–Ulam distances is the l_1 metric on a certain vector space; the corresponding l_p metric for $1 \leqq p \leqq \infty$ gives a whole family of genealogical distance functions, called here $d_p(x, y)$. In [4] it is shown that the inner product associated with $d_2(x, y)$ serves to define an index of genealogical relatedness $r(x, y)$ which measures the degree of cousinship of individuals x and y, and the expected value of $r(x, y)$ is obtained for certain random population models, including the Mycielski–Ulam model. It turns out that common ancestors in the remote past count for too much in $r(x, y)$; one has $r(x, y) \equiv +\infty$ w.p. 1 in the random models when the past is infinite. A discount or mutation rate $1 - \xi > 0$ can be introduced which makes $r(x, y; \xi)$ finite, with divergence at the natural value $\xi \to 1$.

In [6] an index of genealogical relatedness $\rho(x, y)$ is defined as the probability that x and y have the same makeup with respect to a pseudogene pair A, a obeying a simplified version of the Mendelian inheritance law. Associated with $\rho(x, y)$ there is a genealogical distance $d_\rho(x, y) = 1 - \rho(x, y)$. In this treatment one has $\rho(x, y) \equiv 1$ and $d_\rho(x, y) \equiv 0$ w.p. 1 for populations with an infinite past unless a mutation rate $1 - \xi > 0$ is introduced. The difficulties are again connected with the influence of the remote past, but in a more subtle way.

The *coefficient de parenté* of Malécot [5] is the first general formation of an index of genealogical relatedness based on Mendelian genetics, although the notion appears in many places in the earlier literature. (A substantial bibliography is given in [3].) In the present paper we derive the *coefficient de parenté* $\sigma(x, y)$ as a correlation coefficient, define an associated genealogical distance function $d_\sigma(x, y)$, and work out the expected values of these for the random population

---

Received May 8, 1972; revised January 8, 1973.

AMS 1970 *subject classifications*. Primary 92A10; Secondary 60J99.

Key words and phrases. Genealogical index, Malécot coefficient.

models of [4]. The difficulties with the remote past are not removed; we find $\sigma(x, y) \equiv 1$ and $d_\sigma(x, y) \equiv 0$ w.p. 1 for populations with infinite past unless a mutation rate $1 - \xi > 0$ is introduced.

2. The population. We follow the notation of [4]. The population $\mathscr{P}$ under consideration, a subset of some designated sexually reproducing species, is assumed to have the properties:

(i) $\mathscr{P}$ is finite.

(ii) each member of $\mathscr{P}$ is the issue of two other distinct individuals, his father and his mother, not necessarily members of $\mathscr{P}$;

(iii) if $x \in \mathscr{P}$ then all descendants of x up to the present are members of $\mathscr{P}$.

Let $\mathscr{B}$ denote the set $\mathscr{B} = \{0, 1\}$, and let $\mathscr{S}$ be the set $\mathscr{S} = \{\emptyset, 0, 1, 00, 01, \cdots\}$ of all finite sequences of members of $\mathscr{B}$, including the empty sequence $\emptyset$. The length of $s \in \mathscr{S}$ is denoted by $\lambda(s)$, the concatenation of s, $t \in \mathscr{S}$ is denoted by st, of length $\lambda(st) = \lambda(s) + \lambda(t)$, and $s \lesssim t$ means $t = s\sigma$ for some $\sigma \in \mathscr{S}$. For each fixed $x \in \mathscr{P}$ the ancestral designating function x_s, $s \in \mathscr{S}$, is determined by: $x_\emptyset = x$, $x_{t0} =$ father of x_t, $x_{t1} =$ mother of x_t, $t \in \mathscr{S}$. With $\mathscr{P}_0 = \{x_0 : x \in \mathscr{P}\}$ and $\mathscr{P}_1 = \{x_1 : x \in \mathscr{P}\}$ the parental sets of members of $\mathscr{P}$, the set of progenitors of $\mathscr{P}$ is $\partial\mathscr{P} = (\mathscr{P}_0 \cup \mathscr{P}_1) - \mathscr{P}$; we put $\mathscr{P}^+ = \mathscr{P} \cup \partial\mathscr{P}$. For $x \in \mathscr{P}^+$, $\mathscr{A}_x = \{x_s : s \in \mathscr{S}$ and $x_s \in \mathscr{P}^+\}$ denotes the set of members of $\mathscr{P}^+$ who are ancestors of x, plus x himself. No $x \in \mathscr{P}^+$ is his own proper ancestor, whence $x \in \mathscr{A}_y$ and $y \in \mathscr{A}_x$ implies $y = x$. The ancestral chart $\mathscr{C}_x$ is the graph $\mathscr{C}_x = \{(s, x_s) : s \in \mathscr{S}\}$ of the function x.

3. The genetic model. The genetics we use is the elementary model of [1] Section V. 5. We single out for attention a particular chromosome pair $C = \{C^{(0)}, C^{(1)}\}$ in the species containing $\mathscr{P}$, and on $C^{(0)}$, $C^{(1)}$ paired locations $L^{(0)}$, $L^{(1)}$. We assume that for any given individual, $L^{(0)}$ is occupied by one of two genes A, a and that $L^{(1)}$ is also occupied by one of A, a. For $\iota \in \mathscr{B}$ the indicator $h^{(\iota)}(x)$ will specify the gene on location $L^{(\iota)}$ in the typical cells of $x \in \mathscr{P}$ according to: $h^{(\iota)}(x) = +1$ if A appears at $L^{(\iota)}$ and $h^{(\iota)}(x) = -1$ if a appears at $L^{(\iota)}$. We make the assumptions:

(i) a randomly chosen gamete (germ cell) of any $x \in \mathscr{P}$ contains either $C^{(0)}$ or else $C^{(1)}$ with probability $\frac{1}{2}$ each;

(ii) C is not the sex determining pair;

(iii) genes A, a are neutral as far as gamete viability, survival value, mating preference, or other phenotypic expression are concerned.

The simple Mendelian inheritance law then yields the distribution

$$
\begin{aligned}
(1) \qquad h^{(\iota)}(x) &= h^{(0)}(x_\iota) \qquad \text{with probability } \tfrac{1}{2} \\
&= h^{(1)}(x_\iota) \qquad \tfrac{1}{2}, \qquad \iota \in \mathscr{B},\ x \in \mathscr{P},
\end{aligned}
$$

for the origin of the genes of x. (The assumptions (i)—(iii) are necessary because

(1) is an a postiori distribution, given that x is a surviving member of $\mathscr{P}$, of sex not specified.)

There are various equivalent ways of introducing a mutation parameter; the following is the most convenient for our purposes. With each $x \in \mathscr{P}$ we associate a pair $m^{(0)}(x)$, $m^{(1)}(x)$ of ± 1-valued random variables; the probability distribution of the m's will be specified presently. In place of (1) we assume

$$
\begin{aligned}
h^{(\iota)}(x) &= h^{(0)}(x_\iota) && \text{with probability } \tfrac{1}{2}\xi \\
&= h^{(1)}(x_\iota) && \tfrac{1}{2}\xi\,, \qquad \iota \in \mathscr{B},\ x \in \mathscr{P}, \\
&= m^{(\iota)}(x) && 1-\xi
\end{aligned}
\tag{2}
$$

with $0 \leqq \xi \leqq 1$ the mutation parameter, the same for all x; when $\xi = 1$ we have (1).

Our measure of relatedness is based on the following idea. Suppose that genes A, a are assigned at random to the members of the progenitor set $\partial\mathscr{P}$, the assignments being independent. If $x, y \in \mathscr{P}$ have no ancestors in $\partial\mathscr{P}$ in common then the genotypes of x and y will be statistically independent, whereas common ancestors in $\partial\mathscr{P}$ will introduce dependence, and the more dependence the closer x and y are related. Our relatedness index measures the correlation of the genotypes of x and y, in a sense to be made precise.

Explicitly, a probability structure is determined by the following assumptions:

(i) the various choices (2) for $x \in \mathscr{P}$ are statistically independent;

(ii) $\mathscr{U} = \{h^{(\iota)}(x) : x \in \partial\mathscr{P}, \iota \in \mathscr{B}\} \cup \{m^{(\iota)}(x) : x \in \mathscr{P}, \iota \in \mathscr{B}\}$ is a set of mutually independent random variables, independent of the random choices (2);

(iii) the members u of $\mathscr{U}$ are identically distributed with distribution $P\{u = +1\} = p$, $P\{u = -1\} = q = 1 - p$ for some $0 \leqq p \leqq 1$.

Observe that (2) is suspended for $x \in \partial\mathscr{P}$, e.g., if both $z \in \partial\mathscr{P}$ and $z_0 \in \partial\mathscr{P}$ we still require $h^{(0)}(z)$, $h^{(1)}(z)$, $h^{(0)}(z_0)$, $h^{(1)}(z_0)$ to be mutually independent.

The gene frequencies p, q of A, a are stable under (2); that is, $P\{h^{(\iota)}(x) = +1\} = p$, $P\{h^{(\iota)}(x) = -1\} = q$ for all $x \in \mathscr{P}$, $\iota \in \mathscr{B}$. (Cf. [1] pages 135–136.) The following proof of this introduces an idea to be used later. Let $\mathscr{U} = \{u_\gamma : \gamma \in \Gamma\}$ be an indexing of $\mathscr{U}$. By induction from (2), $h^{(\iota)}(x)$ is obtained by selecting a member of $\mathscr{U}$ according to some probability distribution $P\{h^{(\iota)}(x) = u_\gamma\} = p(\gamma \mid x, \iota)$, $\gamma \in \Gamma$, this random selection being independent of the values of the members of $\mathscr{U}$. Although it is easy enough to give the explicit form for $p(\gamma \mid x, \iota)$, we need only the property $\sum_\gamma p(\gamma \mid x, \iota) = 1$. We find

$$
\begin{aligned}
E\{h^{(\iota)}(x)\} &= \textstyle\sum_\gamma E\{u_\gamma\} p(\gamma \mid x, \iota) \\
&= \textstyle\sum_\gamma (p - q) p(\gamma \mid x, \iota) \\
&= p - q\,, \qquad\qquad x \in \mathscr{P},\ \iota \in \mathscr{B}.
\end{aligned}
$$

The proof is concluded by the remark that the distribution of a ± 1-valued random variable u is determined by $E\{u\}$ according to $P\{u = \pm 1\} = \frac{1}{2}[1 \pm E\{u\}]$.

For each $x \in \mathscr{P}^+$ the quantity $h(x) = \frac{1}{2}[h^{(0)}(x) + h^{(1)}(x)]$ scores the genotype of x according to $h(x) = 1, 0, -1$ for x being AA, Aa, aa, respectively. We define the genealogical distance $d_\sigma^2(x, y) = E\{[h(x) - h(y)]^2\}$. The expectation $E\{\ \}$ is over the probability structure designated above, so that $d_\sigma(x, y)$ depends on the parameters ξ, p. We will write $d_\sigma(x, y; \xi)$ when the ξ dependence is of interest; the p dependence will require no notational observance. With an inner product $R(x, y)$ defined by $R(x, y) = E\{h(x)h(y)\}$, $x, y \in \mathscr{P}^+$, d_σ and R satisfy the familiar $d_\sigma^2(x, y) = R(x, x) + R(y, y) - 2R(x, y)$.

THEOREM 1. *$d_\sigma(x, y)$ is a metric on $\mathscr{P}^+$ provided* $0 < p < 1$, $0 \leqq \xi < 1$.

PROOF. The properties $d_\sigma(x, y) \geqq 0$, $d_\sigma(x, y) = d_\sigma(y, x)$ and $d_\sigma(x, y) \leqq d_\sigma(x, z) + d_\sigma(z, y)$ derive from the inner product structure in the usual way. It remains to show that under the hypotheses, $d_\sigma(x, y) \neq 0$ if $x \neq y$. The condition $0 < p < 1$ is necessary, of course; if $p = 1$ or 0 then the $h^{(\cdot)}(\cdot)$ are either all $+1$ or all -1, respectively, with probability 1. The selections (2) being statistically independent for $x \neq y$, the probability is at least $(1 - \xi)^4$ that the four $h^{(\cdot)}(\cdot)$ for x and y constitute four distinct members $u_{\gamma_1}, u_{\gamma_2}, u_{\gamma_3}, u_{\gamma_4}$ of $\mathscr{U}$, e.g., each $h^{(\cdot)}(\cdot)$ equal to the appropriate $m^{(\cdot)}(\cdot)$ when $x, y \in \mathscr{P}$. There follows, for $x, y \in \mathscr{P}^+$, $x \neq y$,

$$\begin{aligned} d_\sigma^2(x, y; \xi) &\geqq \frac{(1-\xi)^4}{4} E\{[u_{\gamma_1} + u_{\gamma_2} - u_{\gamma_3} - u_{\gamma_4}]^2\} \\ &= 4pq(1 - \xi)^4 > 0\,, \qquad 0 \leqq \xi < 1, 0 < p < 1\,. \ \square \end{aligned}$$

The covariance of the random variables $h(x)$, $h(y)$ is $E\{h(x)h(y)\} - E\{h(x)\}E\{h(y)\} = R(x, y) - (p - q)^2$, since $E\{h(x)\} = E\{h(y)\} = p - q$. We show now that this covariance is always nonnegative; a multiple of it will be our genealogical relatedness index $\sigma(x, y)$.

The pair $h^{(\iota)}(x)$, $h^{(\theta)}(y)$ for any fixed $x, y \in \mathscr{P}^+$ and $\iota, \theta \in \mathscr{B}$ is obtained by selecting a pair from $\mathscr{U}$ according to some probability distribution $P\{h^{(\iota)}(x) = u_{\gamma'}$ and $h^{(\theta)}(y) = u_{\gamma''}\} = p(\gamma', \gamma'' \mid x, \iota; y, \theta)$, $\gamma', \gamma'' \in \Gamma$; this distribution is determined by $\mathscr{C}_x$ and $\mathscr{C}_y$. The random selections (2) involved are independent of the values of the members of $\mathscr{U}$, and it follows that

$$\begin{aligned} E\{h^{(\iota)}(x)h^{(\theta)}(y)\} &= \textstyle\sum_{\gamma'} \sum_{\gamma''} E\{u_{\gamma'} u_{\gamma''}\} p(\gamma', \gamma'' \mid x, \iota; y, \theta) \\ &= \textstyle\sum_{\gamma} p(\gamma, \gamma \mid x, \iota; y, \theta) \\ &\qquad + (p - q)^2 \textstyle\sum_{\gamma' \neq \gamma''} \sum p(\gamma' \gamma'' \mid x, \iota; y, \theta) \\ &= (p - q)^2 + 4pq \textstyle\sum_{\gamma} p(\gamma, \gamma \mid x, \iota; y, \theta)\,, \end{aligned} \tag{3}$$

where we have used $\sum_{\gamma'} \sum_{\gamma''} p(\gamma', \gamma'' \mid x, \iota; y, \theta) = 1$. Since $R(x, y) = (\frac{1}{4}) \sum_\iota \sum_\theta E\{h^{(\iota)}(x)h^{(\theta)}(y)\}$, if we define $\sigma(x, y)$ by

$$\sigma(x, y) = \tfrac{1}{4} \textstyle\sum_\iota \sum_\theta \sum_\gamma p(\gamma, \gamma \mid x, \iota; y, \theta) \tag{4}$$

then we have

$$R(x, y) = (p - q)^2 + 4pq\sigma(x, y)\,, \qquad x, y \in \mathscr{P}^+\,,$$

and further

$$d_\sigma^2(x, y) = 4pq[\sigma(x, x) + \sigma(y, y) - 2\sigma(x, y)]\,, \qquad x, y \in \mathscr{P}^+\,.$$

The quantity $\sigma(x, y)$ is independent of p, and is just $R(x, y)$ for the value $p = \frac{1}{2}$ of the parameter p. It is obvious from the definition that $0 \leqq \sigma(x, y) \leqq 1$ always. The quantity $\sigma(x, y)$ is our index of genealogical relatedness in the Mendelian genetic model.

For the case $x = y$ we see from (4) that $\sigma(x, x) = \frac{1}{2} + \frac{1}{2}\sigma(x)$ with $\sigma(x)$ defined by $\sigma(x) = \sum_\gamma p(\gamma, \gamma \mid x, 0; x, 1)$. The quantity $0 \leqq \sigma(x) \leqq 1$ measures the inbreeding in the ancestry of x. It also parameterizes the distribution of the genotype indicator $h(x)$; from

$$\begin{aligned} P\{h(x) = \pm 1\} &= \tfrac{1}{4}E\{[1 \pm h^{(0)}(x)][1 \pm h^{(1)}(x)]\} \\ &= \tfrac{1}{2}[R(x, x) \pm (p - q)]\,, \\ P\{h(x) = 0\} &= 1 - E\{h^2(x)\} \\ &= 1 - R(x, x)\,, \end{aligned}$$

and $R(x, x) = p^2 + q^2 + 2pq\sigma(x)$, we find

$$\begin{aligned} P\{h(x) = 1\} &= p^2 + pq\sigma(x) \\ P\{h(x) = 0\} &= 2pq[1 - \sigma(x)] \\ P\{h(x) = -1\} &= q^2 + pq\sigma(x)\,. \end{aligned}$$

Another interpretation of $\sigma(x, y)$ is the following. Let $h^{(\cdot)}(x)$, $h^{(\cdot)}(y)$ specify the genes at L on two independently chosen gametes, one from x and one from y. Then $\sigma(x, y)$ is seen to be $P\{h^{(\cdot)}(x) = u_\gamma = h^{(\cdot)}(y) \text{ for some } \gamma\}$. This is the probability that the genes $h^{(\cdot)}(x)$, $h^{(\cdot)}(y)$ are identical by descent in the terminology of Malécot, and our $\sigma(x, y)$ is just the *coefficient de parenté* of Malécot [5]. Moreover, $\sigma(x)$ is the *coefficient de consanguenite* of [5].

4. An explicit formula for $\sigma(x, y)$. We next derive some recurrence properties satisfied by $R(x, y)$. First, the symmetry $R(x, y) = R(y, x)$ is obvious from the definition $R(x, y) = E\{h(x)h(y)\}$; taking $p = \frac{1}{2}$ gives the symmetry

$$\sigma(x, y) = \sigma(y, x)\,, \qquad x, y \in \mathscr{P}^+\,. \tag{5}$$

If $x \notin \mathscr{A}_y$ the choices (2) for $x \in \mathscr{P}$ are conditionally independent given the values of $h^{(0)}(y)$, $h^{(1)}(y)$, and moreover, $m^{(\iota)}(x)$ is independent of $h(y)$, $\iota \in B$, since $p(\gamma, \gamma \mid x, \iota; y, \theta) = 0$ if $u_\gamma = m^{(\iota)}(x)$. From

$$\begin{aligned} E\{h^{(\iota)}(x)h(y)\} &= \frac{\xi}{2}[E\{h^{(0)}(x_\iota)h(y)\} + E\{h^{(1)}(x_\iota)h(y)\}] + (1 - \xi)E\{m^{(\iota)}(x)h(y)\} \\ &= \xi E\{h(x_\iota)h(y)\} + (1 - \xi)(p - q)^2\,, \qquad x \notin \mathscr{A}_y,\ x \in \mathscr{P},\ y \in \mathscr{P}^+\,, \end{aligned}$$

there follows

$$R(x, y) = \frac{\xi}{2}\,R(x_0, y) + \frac{\xi}{2}\,R(x_1, y) + (p - q)^2(1 - \xi)\,,$$
$$x \notin \mathscr{A}_y,\ x \in \mathscr{P},\ y \in \mathscr{P}^+\,,$$

and taking $p = \frac{1}{2}$,

$$(6) \qquad \sigma(x, y) = \frac{\xi}{2}\sigma(x_0, y) + \frac{\xi}{2}\sigma(x_1, y)\,, \qquad x \notin \mathscr{A}_y,\ x \in \mathscr{P},\ y \in \mathscr{P}^+\,.$$

When $y = x \in \mathscr{P}$ we have by (2),

$$\begin{aligned} 2R(x, x) - 1 &= E\{h^{(0)}(x)h^{(1)}(x)\} \\ &= \left(\frac{\xi}{2}\right)^2 \textstyle\sum_\iota \sum_\theta E\{h^{(\iota)}(x_0)h^{(\theta)}(x_1)\} \\ &\quad + \xi(1-\xi)[E\{h(x_0)m^{(1)}(x)\} + E\{m^{(0)}(x)h(x_1)\}] \\ &\quad + (1-\xi)^2 E\{m^{(0)}(x)m^{(1)}(x)\} \\ &= \xi^2 R(x_0, x_1) + (1-\xi^2)(p-q)^2\,, \qquad x \in \mathscr{P}; \end{aligned}$$

with $p = \frac{1}{2}$ we find

$$(7) \qquad \sigma(x, x) = \tfrac{1}{2} + \frac{\xi^2}{2}\sigma(x_0, x_1)\,, \qquad x \in \mathscr{P}.$$

By the basic assumption of the model, $R(x, y) = (p - q)^2 + 2pq\delta_{xy}$ for x, $y \in \partial\mathscr{P}$, whence

$$(8) \qquad \sigma(x, y) = \tfrac{1}{2}\delta_{xy}\,, \qquad x, y \in \partial\mathscr{P},$$

where δ_{xy} is the Kronecker symbol.

THEOREM 2. *The properties* (5)—(8) *determine* $\sigma(x, y)$ *uniquely for* x, $y \in \mathscr{P}^+$.

PROOF. Let $\tau(x, y) = \sigma'(x, y) - \sigma''(x, y)$ be the difference of two solutions. Then, for all x, $y \in \mathscr{P}^+$,

(i) $\tau(x, y) = \tau(y, x)$,
(ii) $\tau(x, y) = (\xi/2)\tau(x_0, y) + (\xi/2)\tau(x_1, y)$ if $x \notin \mathscr{A}_y$, $x \in \mathscr{P}$,
(iii) $\tau(x, x) = (\xi^2/2)\tau(x_0, x_1)$ if $x \in \mathscr{P}$,
(iv) $\tau(x, y) = 0$ if x, $y \in \partial\mathscr{P}$.

Arguing as in [4], we apply (i)—(iii) in some order to obtain

$$\tau(x, y) = \sum_{z', z'' \in \partial\mathscr{P}}\sum \tau(z', z'')K(z', z''; x, y)\,, \qquad x, y \in \mathscr{P}^+\,,$$

where the K's are combinatorial coefficients depending possibly on the algorithm used for the application of (i)—(iii). (This last point was not discussed in [4], but the argument for [4] is similar.) From (iv) it then follows that $\sigma'(x, y) = \sigma''(x, y)$, $x, y \in \mathscr{P}^+$. □

We will now exhibit an expression $\sigma(x, y)$ which satisfies (5)—(8). For $x \in \mathscr{P}^+$ and $s \in \mathscr{S}$ the path $\pi_s(x)$ is defined to be the chain of successive ancestors of x along s; that is, $\pi_s(x)$ is the sequence $\pi_s(x) = [x_{s_j} : j = 0, 1, \cdots, \lambda(s)]$ with $s_j \in \mathscr{S}$ the unique $s_j \precsim s$ such that $\lambda(s_j) = j$. The edges of $\pi_s(x)$ are the ordered pairs $[x_{s_j}, x_{s_{j+1}}]$ for $j = 0, 1, \cdots, \lambda(s) - 1$. For $x, y \in \mathscr{P}^+$ and $s, t \in \mathscr{S}$ we write $\pi_s(x) \perp \pi_t(y)$ if the pair of paths $\pi_s(x)$, $\pi_t(y)$ has the properties:

(i) $\pi_s(x)$ and $\pi_t(y)$ have no edge in common;
(ii) neither $\pi_s(x)$ nor $\pi_t(y)$ has an edge lying entirely in $\partial\mathscr{P}$.

That is, common nodes are permitted ($x_{s_j} = y_{t_k}$ for some $j \leqq \lambda(s)$, $k \leqq \lambda(t)$) but not successive common nodes.

THEOREM 3. *The relatedness index $\sigma(x, y)$ is given explicitly by*

$$\sigma(x, y) = \tfrac{1}{2} \textstyle\sum_s \sum_t \left\{ \left(\dfrac{\xi}{2} \right)^{\lambda(s)+\lambda(t)} : x_s = y_t \in \mathscr{P}^+ \text{ and } \pi_s(x) \perp \pi_t(y) \right\}, \qquad x, y \in \mathscr{P}^+ . \tag{9}$$

PROOF. We will show that the expression (9) has the properties (5)—(8). The symmetry $\sigma(x, y) = \sigma(y, x)$ of (9) is obvious, and it is soon apparent that (9) has the values $\sigma(x, y) = \frac{1}{2}\delta_{xy}$ for $x, y \in \partial\mathscr{P}$, which is property (8).

The ancestral chart $\mathscr{C}_x$ of $x \in \mathscr{P}^+$ may be regarded as the disjoint union $\mathscr{C}_x = \{x\} \cup \mathscr{C}_{x_0} \cup \mathscr{C}_{x_1}$ in the sense that either $x_s = x$ for $s = \emptyset$ or else $x_s = x_{0s'} = (x_0)_{s'}$ biuniquely for $s' \in \mathscr{S}$ or else $x_s = x_{1s''} = (x_1)_{s''}$ biuniquely for $s'' \in \mathscr{S}$. With $x \in \mathscr{P}$, $y \in \mathscr{P}^+$ such that $x \notin \mathscr{A}_y$, suppose $s, t \in \mathscr{S}$ are such that $x_s = y_t$. The assumption $x \notin \mathscr{A}_y$ entails $s \neq \emptyset$, so that $s = \iota s'$ for unique $\iota \in \mathscr{B}$, $s' \in \mathscr{S}$. If $\pi_s(x) \perp \pi_t(y)$ then $\pi_{s'}(x_\iota) \perp \pi_t(y)$ certainly. On the other hand, for any $\iota \in \mathscr{B}$ and $s', t \in \mathscr{S}$ such that $x_{\iota s'} = y_t$ and $\pi_{s'}(x_\iota) \perp \pi_t(y)$, the paths $\pi_s(x)$, $\pi_t(y)$ for $s = \iota s'$ will also satisfy $\pi_s(x) \perp \pi_t(y)$, since the added edge $[x, x_\iota]$ is not in $\pi_t(y)$, from $x \notin \mathscr{A}_y \supset \pi_t(y)$, nor in $\partial\mathscr{P}$, from $x \in \mathscr{P}$. Thus (9) becomes

$$\begin{aligned} \sigma(x, y) &= \tfrac{1}{2} \textstyle\sum_\iota \sum_{s'} \sum_t \left\{ \left(\dfrac{\xi}{2} \right)^{1+\lambda(s')+\lambda(t)} : x_{\iota s'} = y_t \in \mathscr{P}^+ \text{ and } \pi_{s'}(x_\iota) \perp \pi_t(y) \right\} \\ &= \frac{\xi}{2}\sigma(x_0, y) + \frac{\xi}{2}\sigma(x_1, y), \qquad x \notin \mathscr{A}_y,\ x \in \mathscr{P},\ y \in \mathscr{P}^+ , \end{aligned}$$

showing that (9) has property (6).

When $x = y \in \mathscr{P}^+$ we have $x_s = y_t \in \mathscr{P}^+$ and $\pi_s(x) \perp \pi_t(y)$ iff either $s = t = \emptyset$ or else $x \in \mathscr{P}$ and $x_{\iota s'} = x_{\theta t'} \in \mathscr{P}^+$ and $\pi_{s'}(x_\iota) \perp \pi_{t'}(x_\theta)$ with $\theta = 1 - \iota$ and $s = \iota s'$, $t = \theta t'$ for unique $\iota \in \mathscr{B}$ and $s', t' \in \mathscr{S}$. From this one verifies that (9) has property (7).

By the uniqueness result of Theorem 2, expression (9) is an explicit representation for the relatedness index $\rho(x, y)$ defined in (4). □

We next obtain an expression for $\sigma(x, y)$ corresponding to [4] Equation (3). Let us say that paths $\pi_s(x)$, $\pi_t(y)$ satisfy $\pi_s(x) \underline{\perp} \pi_t(y)$ provided:

(i) $x_{s_j} \neq y_{t_k}$ for all $0 \leqq j < \lambda(s)$ and $0 \leqq k < \lambda(t)$;

(ii) neither $\pi_s(x)$ nor $\pi_t(y)$ has an edge lying entirely in $\partial\mathscr{P}$.

Thus a pair $\pi_s(x)$, $\pi_t(y)$ for which $x_s = y_t \in \mathscr{P}$ is minimal in the sense of [4] provided $\pi_s(x) \underline{\perp} \pi_t(y)$.

THEOREM 4. *In terms of minimal path pairs, $\sigma(x, y)$ is*

$$\sigma(x, y) = \textstyle\sum_s \sum_t \left\{ \left(\dfrac{\xi}{2} \right)^{\lambda(s)+\lambda(t)} \sigma(z, z) : x_s = y_t (= z) \in \mathscr{P}^+ \text{ and } \pi_s(x) \underline{\perp} \pi_t(y) \right\}, \qquad x, y \in \mathscr{P}^+ . \tag{10}$$

PROOF. If $x, y \in \mathscr{P}^+$ and $s, t \in \mathscr{S}$ are such that $x_s = y_t \in \mathscr{P}^+$ and $\pi_s(x) \perp \pi_t(y)$ then there exist uniquely $z \in \mathscr{P}^+$, $s', t', \sigma, \tau \in \mathscr{S}$ such that $s = s'\sigma$, $t = t'\tau$, $x_{s'} = y_{t'} = z \in \mathscr{P}^+$, $\pi_{s'}(x) \perp \pi_{t'}(y)$, and further, $z_\sigma = z_\tau \in \mathscr{P}^+$ and $\pi_\sigma(z) \perp \pi_\tau(z)$. Conversely, if $x, y \in \mathscr{P}^+$ and $s', t' \in \mathscr{S}$ are such that $x_{s'} = y_{t'} \in \mathscr{P}^+$ and $\pi_{s'}(x) \perp \pi_{t'}(y)$, and if $z = x_{s'} = y_{t'}$ and $\sigma, \tau \in \mathscr{S}$ are such that $z_\sigma = z_\tau \in \mathscr{P}^+$ and $\pi_\sigma(z) \perp \pi_\tau(z)$, then the composites $s = s'\sigma$, $t = t'\tau$ satisfy $x_s = y_t \in \mathscr{P}^+$ and $\pi_s(x) \perp \pi_t(y)$. The theorem follows. □

The relatedness index $\rho(x, y)$ of [6] being

$$\rho(x, y) = \textstyle\sum_s \sum_t \left\{\left(\frac{\xi}{2}\right)^{\lambda(s)+\lambda(t)} : x_s = x_t \in \mathscr{P}^+ \text{ and } \pi_s(x) \perp \pi_y(t)\right\}, \qquad x, y \in \mathscr{P}^+,$$

we see that $\frac{1}{2}\rho(x, y) \leqq \sigma(x, y) \leqq \rho(x, y)$, $x, y \in \mathscr{P}^+$, since the coefficients $\sigma(z, z)$ in (10) satisfy $\frac{1}{2} \leqq \sigma(z, z) \leqq 1$, from (7).

5. The random families. We describe now the probabilistic population models treated in [4], for which $E\{\sigma(x, y)\}$ can be obtained explicitly. The population is stratified into generations $\mathscr{P}^{(0)}, \mathscr{P}^{(1)}, \cdots$ into the past, with the parents of $z \in \mathscr{P}^{(m)}$ being members of $\mathscr{P}^{(m+1)}$. The $\mathscr{P}$ of the preceding sections will be $\mathscr{P}^{(0)} \cup \cdots \cup \mathscr{P}^{(M)}$ for some large M, but only the limit as $M \to \infty$ will be given for quantities involving M. In a few places we use without mention the uniform convergence of the series for $\sigma(x, y; \xi)$ on $0 \leqq \xi \leqq 1$.

Each $\mathscr{P}^{(m)}$ consists of $2c$ individuals, named $1, \cdots, 2c$, paired into couples $\mathscr{K}_i^{(m)} = \{i, c + i\}$, $1 \leqq i \leqq c$, with $c > 1$ independent of m. Each $\mathscr{K}_i^{(m+1)}$ has as issue a son and a daughter in $\mathscr{P}^{(m)}$. For $1 \leqq i \leqq c$ let the infant names of the two children of $\mathscr{K}_i^{(m+1)}$ be $\{i, c + i\}$. The adult names of the members of $\mathscr{P}^{(m)}$ are obtained by applying a permutation operator $\pi^{(m)}$ to the list $1, \cdots, 2c$ of infant names. The random family structure is determined by:

(i) for each m the permutation $\pi^{(m)}$ is a random variable uniformly distributed on a set of admissible permutations, to be designated;

(ii) the $\pi^{(m)}$ for various m are mutually independent random variables, independent also of the basic random variables of the genetic part of the model.

The models we treat are:

Model $\tau = 2$: An admissible permutation is the identity on $(1, \cdots, c)$ and nonrencontre on $(c + 1, \cdots, 2c)$. The interpretation is that the couples are surnamed $1, \cdots, c$. The son i of $\mathscr{K}_i^{(m+1)}$ = "Mr. and Mrs. i" of $\mathscr{P}^{(m+1)}$ becomes "Mr. i" of $\mathscr{K}_i^{(m)}$, and the daughter $c + i$ = "Miss i" becomes $c + v_i$ = "Mrs. v_i" of $\mathscr{K}_{v_i}^{(m)}$, with $(v_1, \cdots, v_c)$ a nonrencontre permutation of $(1, \cdots, c)$. That is, no $\mathscr{K}_i^{(m)}$ consists of siblings.

Model $\tau = 0$: The same without the taboo; each permutation which is the identity on $(1, \cdots, c)$ is admissible.

Model $\tau = 1$: Each permutation of $(1, \cdots, 2c)$ is admissible. The determination of sex and surnames in this model involves random renumbering of the couples, which we will not need to discuss in detail, see [4] Section 4.

The probability $P\{\ \}$ and expectation $E\{\ \}$ of the preceding sections are to be treated as conditionals for given $\pi^{(0)}, \cdots, \pi^{(M-1)}$: in the present section $P\{\ \}$ and $E\{\ \}$ involve averaging also over the $\pi^{(m)}$'s. We denote by $E\{\ \mid \mathscr{F}_m\}$ the conditional expectation for given $\pi^{(0)}, \cdots, \pi^{(m-1)}$.

Consider now the random variable $\sigma(x, y)$ for $x \in \mathscr{P}^{(k)}$, $y \in \mathscr{P}^{(l)}$, with $0 \leqq k \leqq l$. The rest of the present section is devoted to obtaining $E\{\sigma(x, y)\}$ explicitly in the various cases. Let us first examine (10). For $z \in \mathscr{P}^{(m)}$ the random variable $\sigma(z, z)$ depends only on $\pi^{(m)}, \pi^{(m+1)}, \cdots$, while the event $\{x_s = y_t = z$ and $\pi_s(x) \perp \pi_t(y)\}$ is $\mathscr{F}_m$ measurable. Let random variable $M_m(x, y)$ be the total number of minimal path pairs to some (unspecified) member of $\mathscr{P}^{(m)}$. Using $E\{\sigma(z, z) \mid \mathscr{F}_m\} = E\{\sigma(z, z)\}$ for $z \in \mathscr{P}^{(m)}$ and the fact (to be verified presently) that $E\{\sigma(z, z)\}$ is the same for all $z \in \mathscr{P}$, we obtain from (10):

$$(11) \qquad E\{\sigma(x, y)\} = E\{\sigma(z, z)\} \textstyle\sum_{m=l}^{\infty} \left(\dfrac{\xi}{2}\right)^{2m-k-l} E\{M_m(x, y)\}$$

$$= E\{\sigma(z, z)\}E\{\rho(x, y)\}\,, \qquad x, y \in \mathscr{P}.$$

Now, the genealogical index $r(x, y)$ of [4] satisfies a similar identity

$$(12) \qquad E\{r(x, y)\} = E\{r(z, z)\}E\{\rho(x, y)\}$$

for these random families. Combining (11) and (12), we obtain the remarkable result

$$(13) \qquad E\{\sigma(x, y)\} = \left[\frac{E\{\sigma(z, z)\}}{E\{r(z, z)\}}\right] E\{r(x, y)\}\,, \qquad x, y \in \mathscr{P}.$$

Since $E\{r(x, y)\}$ is given explicitly in [4] for all cases, we need only to calculate $E\{\sigma(z, z)\}$ to obtain $E\{\sigma(x, y)\}$ in all cases.

Let us classify pairs (z', z'') in each $\mathscr{P}^{(m)}$ according to

$z'\langle 1\rangle z''$ if $z' = z''$
$z'\langle 2\rangle z''$ if $z' \neq z''$ belong to the same couple
$z'\langle 3\rangle z''$ if $z' \neq z''$ are of the same sex
$z'\langle 4\rangle z''$ otherwise .

For a path pair $\pi_s(x)$, $\pi_t(y)$ we write $\pi_s(x)\langle\mu\rangle\pi_t(y)$ for $\mu = 1, 2, 3, 4$ provided $\pi_s(x) \perp \pi_t(y)$ and $x_s\langle\mu\rangle y_t$ are satisfied. Consider a path pair $\pi_s(x) \perp \pi_t(y)$ such that $x_s = z' \in \mathscr{P}^{(m)}$, $y_t = z'' \in \mathscr{P}^{(m)}$. There are four extensions to a path pair $\pi_{s\iota}(x)$, $\pi_{t\theta}(y)$ with $\iota, \theta \in \mathscr{B}$ and $x_{s\iota} = w' \in \mathscr{P}^{(m+1)}$, $x_{t\theta} = w'' \in \mathscr{P}^{(m+1)}$. The number of extensions which satisfy $\pi_{s\iota}(x)\langle\mu\rangle\pi_{t\theta}(y)$ depends on the relatedness of z', z'' according to:

		$w'\langle 1\rangle w''$	$w'\langle 2\rangle w''$	$w'\langle 3\rangle w''$	$w'\langle 4\rangle w''$
	$z' = z''$	0	2	0	0
(14)	$z' \neq z''$ sibs	2	2	0	0
	$z' \neq z''$ not sibs	0	0	2	2 .

Whether $z', z'' \in \mathscr{P}^{(m)}$ are sibs or not depends on $\pi^{(m)}$, as well as the class of z', z''; we find easily

$$\begin{array}{llcccc} & & z'\langle 1\rangle z'' & z'\langle 2\rangle z'' & z'\langle 3\rangle z'' & z'\langle 4\rangle z'' \\ (15) & P\{z' \neq z'' \text{ sibs} \mid \mathscr{F}_m\} & 0 & \dfrac{2-\tau}{2c-\tau} & 0 & \dfrac{2}{2c-\tau}\,. \end{array}$$

For $\mu = 1, 2, 3, 4$ let $N_\mu{}^{(m)}(x, y)$ denote the random total number of $\langle\mu\rangle$-path-pairs to $\mathscr{P}^{(m)}$:

$$N_\mu{}^{(m)}(x, y) = \textstyle\sum_s \sum_t \{1 : x_s \in \mathscr{P}^{(m)} \text{ and } y_t \in \mathscr{P}^{(m)} \text{ and } \pi_s(x)\langle\mu\rangle\pi_t(y)\}\,, \qquad m \geqq l\,,$$

and let $||N^{(m)}(x, y)||$ denote the column vector with components $||N^{(m)}(x, y)||_\mu = N_\mu{}^{(m)}(x, y)$, $1 \leqq \mu \leqq 4$. With $||A||$ the matrix

$$A = \left(\frac{2}{2c-\tau}\right)\left|\begin{array}{cccc} 0 & 2-\tau & 0 & 2 \\ 2c-\tau & 2-\tau & 0 & 2 \\ 0 & 2c-2 & 2c-\tau & 2c-\tau-2 \\ 0 & 2c-2 & 2c-\tau & 2c-\tau-2 \end{array}\right|$$

there holds $E\{||N^{(m+1)}(x, y)|| \mid \mathscr{F}_m\} = ||AE\{N^{(m)}(x, y)\}||$, $m \geqq l$, using (14) and (15). Taking expectations, we see that $E\{||N^{(m)}(x, y)||\} = ||A^{m-l}E\{N^{(l)}(x, y)\}||$, so that

$$\begin{aligned} E\{\sigma(x, y)\} &= \tfrac{1}{2} \textstyle\sum_{m=l}^{\infty} \left(\dfrac{\xi}{2}\right)^{2m-k-l} E\{N_1{}^{(m)}(x, y)\} \\ &= \tfrac{1}{2} \textstyle\sum_{m=l}^{\infty} \left(\dfrac{\xi}{2}\right)^{2m-l-k} ||A^{m-l}E\{N^{(l)}(x, y)\}||_1 \\ &= \tfrac{1}{2} \left(\dfrac{\xi}{2}\right)^{l-k} ||[1 - (\xi^2/4)A]^{-1}E\{N^{(l)}(x, y)\}||_1\,. \end{aligned}$$

According to (13), we need to work this out only for $y = x$, in which case $N_\nu{}^{(k)}(x, x) = \delta_{1\nu}$ and $E\{\sigma(x, x)\} = \frac{1}{2}||[1 - (\xi^2/4)A]^{-1}||_{1,1}$. We find

$$\det ||1 - (\xi^2/4)A|| = \frac{(8c-4\tau) - (8c-6\tau)\xi^2 - \tau\xi^4 + (1-\tau)\xi^6}{8c-4\tau}$$

and altogether

$$(16) \qquad E\{\sigma(x, x)\} = \frac{(4c-2\tau) - (4c-3\tau)\xi^2 + (1-\tau)\xi^4}{(8c-4\tau) - (8c-6\tau)\xi^2 - \tau\xi^4 + (1-\tau)\xi^6}$$

$$\frac{E\{\sigma(x, x)\}}{E\{r(x, x)\}} = \frac{4(1-\xi^2)[2c-\tau-(c-\tau)\xi^2]}{8c-4\tau-(8c-6\tau)\xi^2-\tau\xi^4+(1-\tau)\xi^6}\,.$$

For $\tau = 1$ the roots h of $(8c - 4\tau)h^3 - (8c - 6\tau)h^2 - \tau h + 1 - \tau$ are

$$h_1 = \frac{(4c-3) + [4(2c-1)^2+1]^{\frac{1}{2}}}{4(2c-1)} = 1 - \frac{1}{8c-3} + O\left(\frac{1}{c^4}\right)$$

$$h_2 = \frac{(4c-3) - [4(2c-1)^2+1]^{\frac{1}{2}}}{4(2c-1)} = -\frac{1}{8c-5} + O\left(\frac{1}{c^4}\right)$$

$$h_3 = 0\,.$$

For $\tau = 0, 2$ the roots are

$$h_1 = 1 - \frac{1}{8c - \tau - 2} + O\left(\frac{1}{c^3}\right)$$

$$h_2, h_3 = \frac{1 - 2\tau}{16c} + O\left(\frac{1}{c^2}\right) \pm \left[\frac{\tau - 1}{8c} + O\left(\frac{1}{c^2}\right)\right]^{\frac{1}{2}},$$

the last two being complex when $\tau = 0$. In all cases we find

$$E\{N_1^{(m)}(x, y)\} \sim (\text{const.}) \times 2^{2m} \times h_1^m$$
$$= (\text{const.}) \times \left[4 - \frac{1}{2c} + O\left(\frac{1}{c^2}\right)\right]^m$$

for the asymptotic behavior of the number of $\perp$-path-pairs to $\mathscr{P}^{(m)}$. Observe that this exponential increase in the number of $\perp$-path-pairs is just barely less that the discount factor 4^{-m} in the series for $\sigma(x, y; 1)$.

6. Further considerations in the random families. Let us return to the identity (13). From [4], $E\{r(x, y)\}$ has for each τ the form

$$E\{r(x, y; \xi)\} = \frac{N(x, y; \xi)}{D_r(\xi)}, \qquad 0 \leqq \xi < 1, \tag{17}$$

with $N(x, y; \xi)$ a polynomial in ξ and $D_r(\xi)$ the polynomial $D_r(\xi) = 4(1 - \xi^2)[2c - \tau - (c - \tau)\xi^2]$; the τ dependence need not be displayed. If we define $D_\rho(\xi) = N(x, x; \xi) = (4c - 2\tau) - (4c - 3\tau)\xi^2 + (1 - \tau)\xi^4$ then we have, as is shown in [4],

$$E\{\rho(x, y; \xi)\} = \frac{N(x, y; \xi)}{D_\rho(\xi)}, \qquad 0 \leqq \xi \leqq 1, \tag{18}$$

for the same $N(x, y; \xi)$. Comparing with (16), we see that $E\{\sigma(x, x; \xi)\} = D_\sigma(\xi)/D_\rho(\xi)$ with $D_\sigma(\xi) = (8c - 4\tau) - (8c - 6\tau)\xi^2 - \tau\xi^4 + (1 - \tau)\xi^6$, whence, from (13).

$$E\{\sigma(x, y; \xi)\} = \frac{N(x, y; \xi)}{D_\sigma(\xi)}, \qquad 0 \leqq \xi \leqq 1, \tag{19}$$

again with the same $N(x, y; \xi)$.

Since $N(x, y; 1) \equiv 1$ and $D_\sigma(1) = 1$, there follows $E\{\sigma(x, y; 1)\} = 1$ for all $x, y \in \mathscr{P}$. But $\sigma(x, y)$ is a random variable with the property $0 \leqq \sigma(x, y) \leqq 1$, so we have further $\sigma(x, y; 1) = 1$ with probability 1 for all $x, y \in \mathscr{P}$. This failure to measure relatedness can be explained as follows. The vector process $[h^{(\iota)}(x) : x \in \mathscr{P}^{(m)}, \iota \in \mathscr{B}]$ is Markov in the forward time sense $m = M, M - 1, \cdots$ with stationary transition probabilities. Of the 2^{4c} states, two are absorbing when $\xi = 1$ and accessible from all the other states viz., all $h^{(\cdot)}(\cdot) = +1$ and all $h^{(\cdot)}(\cdot) = -1$. With probability 1, then, the system will be eventually in one of the absorbing states. (An explicit calculation appears in [1] page 380.) This is to say,

$$\lim_{m \to -\infty} P\{\sigma(x, y) = 1 \text{ for all } x, y \in \mathscr{P}^{(m)} \mid \text{state at } M\} = 1 .$$

The process being stationary, we have also

$$\lim_{M\to+\infty} P\{\sigma(x, y) = 1 \text{ for all } x, y \in \mathscr{P}^{(m)} \mid \text{state at } M\} = 1$$

for any $m \geqq 0$, implying $\sigma(x, y) = 1$ for all $x, y \in \mathscr{P}$. The same argument applies to $\rho(x, y)$ of [6].

Added in proof. Indeed, an "Adam effect" is apparent: as $M - m \to \infty$ all $4c$ genes in generation m will w.p. 1 eventually be replicates of a single one of the $8c$ progenitor genes, treating these as distinct.

From [4], the genealogical distance $d_2(x, y)$ defined there has for the random families the form

$$E\{[d_2(x, y;\, \xi)]^2\} = \frac{8c(1-\xi)\Delta(x, y;\, \xi)}{D_r(\xi)}, \qquad 0 \leqq \xi < 1,$$

for a certain polynomial $\Delta(x, y;\, \xi)$; the expression $\Delta(x, y;\, 1) = E\{[d_2(x, y;\, 1)]^2\}$ is given explicitly in [4] Section 5. Since $[d_2(x, y)]^2 = 2[r(x, x) + r(y, y) - 2r(x, y)]$, it follows from (17) that

$$\begin{aligned} E\{[d_2(x, y;\, \xi)]^2\} &= \frac{4[N(x, x;\, \xi) - N(x, y;\, \xi)]}{D_r(\xi)} \\ &= \left[\frac{4D_\rho(\xi)}{D_r(\xi)}\right] E\{d_\rho(x, y;\, \xi)\}, \end{aligned}$$

whence

$$E\{d_\rho(x, y;\, \xi)\} = \frac{2c(1-\xi)\Delta(x, y;\, \xi)}{D_\rho(\xi)}.$$

This and $\Delta(x, y;\, \xi) \neq 0$ for $x \neq y$ shows that $\Delta(x, y;\, \xi)$ is a metric. In [4] Section 6 a genealogical distance function $\tilde{d}(x, y)$ was defined by $\tilde{d}(x, y) = \lim_{\xi\to 1} d_\rho(x, y;\, \xi)/[2(1-\xi)]$, so that $E\{\tilde{d}(x, y)\} = c\Delta(x, y;\, 1)$, a result which appears in [4]. For the Mendelian distance $d_\sigma(x, y)$ we have similarly

$$\begin{aligned} E\{[d_\sigma(x, y;\, \xi)]^2\} &= 2E\{\sigma(x, x;\, \xi) - \sigma(x, y;\, \xi)\} \\ &= \frac{2}{D_\sigma(\xi)}[N(x, x;\, \xi) - N(x, y;\, \xi)] \\ &= \left[\frac{2D_\rho(\xi)}{D_\sigma(\xi)}\right] E\{d_\rho(x, y;\, \xi)\} \\ &= \frac{4c(1-\xi)\Delta(x, y;\, \xi)}{D_\sigma(\xi)}, \qquad 0 \leqq \xi \leqq 1, \end{aligned}$$

whence $E\{[d_\sigma(x, y;\, 1)]^2\} = 0$ and

$$\lim_{\xi\to 1} \frac{E\{[d_\sigma(x, y;\, \xi)]^2\}}{1-\xi} = 4c\Delta(x, y;\, 1).$$

7. **Some generalizations.** Instead of the pair A, a, suppose that there are n genes $A_1, \cdots, A_n$ which can occupy locus L, and that these have been assigned score values $v_1, \cdots, v_n$, assumed for convenience to be distinct. For each individual x we may define scoring indicators $h^{(0)}(x)$, $h^{(1)}(x)$ according to: $h^{(\epsilon)}(x) = v_i$

if A_i appears at locus $L^{(\iota)}$ in the cells of x, for $\iota \in \mathscr{B}$; we define also the average score $h(x) = \frac{1}{2}[h^{(0)}(x) + h^{(1)}(x)]$. The inheritance law still being (2), assume as before that the $\mathscr{U} = \{u_\gamma : \gamma \in \Gamma\}$ are independent identically distributed random variables, the distribution of $u \in \mathscr{U}$ being $P\{u = v_i\} = p_i$, $1 \leqq i \leqq n$, for some choice of the p's. Denote the mean and variance of each $u \in \mathscr{U}$ by $M = \sum_1^n p_i v_i$ and $S^2 = \sum_1^n p_i v_i^2 - M^2$, respectively. In place of (3) there holds

$$\begin{aligned} E\{h^{(\iota)}(x)h^{(\theta)}(y)\} &= \textstyle\sum_{\gamma'} \sum_{\gamma''} E\{u_{\gamma'} u_{\gamma''}\} p(\gamma', \gamma'' \mid x, \iota; y, \theta) \\ &= (\textstyle\sum_{i=1}^n p_i v_i^2) \sum_\gamma p(\gamma, \gamma \mid x, \iota; y, \theta) \\ &\quad + (\textstyle\sum_{i=1}^n p_i v_i)^2 \sum\limits_{\gamma' \neq \gamma''}\sum p(\gamma' \gamma'' \mid x, \iota; y, \theta) \\ &= M^2 + S^2 \textstyle\sum_\gamma p(\gamma, \gamma \mid x, \iota; y, \theta) . \end{aligned}$$

With $R(x, y) = E\{h(x)h(y)\}$ as before, we find

$$R(x, y) = M^2 + S^2\sigma(x, y) , \qquad x, y \in \mathscr{P}^+ ,$$

with $\sigma(x, y)$ the same quantity previously defined in (4).

Even more generally, suppose that the $h^{(\cdot)}(\bullet)$ take values in a linear space with inner product $\langle\ ,\ \rangle$, still with inheritance law (2) and the members of $\mathscr{U}$ independent and identically distributed. The only requirement is that $E\{\langle u_{\gamma'}, u_{\gamma''}\rangle\}$, $\gamma', \gamma'' \in \Gamma$ be well defined, necessarily of the form $E\{\langle u_{\gamma'}, u_{\gamma''}\rangle\} = M^{(2)} + S^2\delta_{\gamma'\gamma''}$ for real constants $M^{(2)}$ and $S^2 \geqq 0$. With $R(x, y)$ now $R(x, y) = E\{\langle h(x), h(y)\rangle\}$, $x, y \in \mathscr{P}^+$, a calculation similar to the above shows $R(x, y) = M^{(2)} + S^2\sigma(x, y)$, again for the same $\sigma(x, y)$.

Thus $\sigma(x, y)$ of (4) emerges as a coefficient of covariance in any L_2 setup for the $h^{(\cdot)}(\bullet)$. Suppose the basic random variables $u \in \mathscr{U}$ are real valued, with the property $E\{|u|^\alpha\} < \infty$ for some $1 \leqq \alpha \leqq \infty$. Then a genealogical distance function $d_\sigma^{(\alpha)}(x, y)$ is determined according to $d_\sigma^{(\alpha)}(x, y) = [E\{|h(x) - h(y)|^\alpha\}]^{1/\alpha}$, $x, y \in \mathscr{P}^+$; as we have seen, the case $\alpha = 2$ is distinguished by the existence of an associated relatedness index.

There is a curious relation among the distances $d_\sigma^{(\alpha)}(x, y)$ when the $h^{(\cdot)}(\bullet)$ are ± 1-valued, as originally specified. With each of $h(x)$, $h(y)$ taking values $1, 0, -1$, there holds $E\{|h(x) - h(y)|^\alpha\} = 2^\alpha A(x, y) + B(x, y)$ with $A(x, y) = P\{$one of x, y is AA, the other $aa\}$ and $B(x, y) = P\{$exactly one of x, y is $Aa\}$. The functions $A(x, y)$, $B(x, y)$ are determined if we know $d_\sigma^{(\alpha)}(x, y)$ for two distinct values of $1 \leqq \alpha \leqq \infty$; e.g., using $\alpha = 1, 2$ we obtain, for all $1 \leqq \alpha < \infty$,

$$[d_\sigma^{(\alpha)}(x, y)]^\alpha = 2^{\alpha-1}\{[d_\sigma(x, y)]^2 - d_\sigma^{(1)}(x, y)\} + \{2d_\sigma^{(1)}(x, y) - [d_\sigma(x, y)]^2\} , \qquad x, y \in \mathscr{P}^+ .$$

There is a discontinuity in $d_\sigma^{(\infty)}(x, y)$ at $\xi = 1$ for certain x, y provided $0 < p < 1$ and $\mathscr{P}$ is finite:

$$\begin{aligned} d_\sigma^{(\infty)}(x, y; \xi) &= 2(1 - \delta_{xy}) && \text{if } 0 \leqq \xi < 1 , \\ d_\sigma^{(\infty)}(x, y; 1) &= 1 && \text{if one of } x, y \text{ is parent of the other ,} \\ &= 2(1 - \delta_{xy}) && \text{otherwise .} \end{aligned}$$

8. Addendum. Reference [3] came to our attention after the above was written. The *coefficient of kinship* of Karlin ([3] page 85 *et. seq.*) involves the joint distribution of all four of $h^{(0)}(x)$, $h^{(1)}(x)$, $h^{(0)}(y)$, $h^{(1)}(y)$, and does not satisfy a recurrence scheme as simple as (5)—(8).

REFERENCES

[1] FELLER, WILLIAM (1968). *An Introduction to Probability Theory and its Applications* **1**, 3rd ed. Wiley, New York.

[2] KAHANE, JOSEPH and MARR, ROBERT (1972). *On a class of stochastic pairing processes and the Mycielski-Ulam notions of genealogical distance. J. Combinatorial Theory*, **13** 383-400.

[3] KARLIN, SAMUEL (1969). *Equilibrium Behavior of Population Genetic Models with Non-random Mating.* Gordon and Breach, New York.

[4] LLOYD, S. P. (1974). *An index of genealogical relatedness, Advances in Appl. Probability.*

[5] MALÉCOT, G. (1948). *Les Mathématiques de L'Hérédite.* Masson et Cic, Paris.

[6] MALLOWS, C. L. (1974). *Another definition of genealogical distance. J. Appl. Probability.*

[7] MYCIELSKI, JAN and ULAM, S. M. (1969). *On the pairing process and the notion of genealogical distance. J. Combinatorial Theory* **6** 227-234.

BELL LABORATORIES
600 MOUNTAIN AVENUE
MURRAY HILL, NEW JERSEY 07974