HIGHER ORDER PROBABILITY FUNCTIONS OF IDENTITY OF ALLELES BY DESCENT[1]

C. CLARK COCKERHAM

Department of Statistics, North Carolina State University, Raleigh, N. C.

Received June 7, 1971

HARRIS (1964) provided a parameterization of nine probabilities for the fifteen states of identity of allelic genes of inbred relatives, and with these developed the joint frequencies and covariances of the relatives. He also showed how to find the probabilities when one relative is not an ancestor of the other. At about the same time (1965) GILLOIS organized various probability functions for the states of identity and, although not involving explicit expressions of the joint frequencies, found the covariances of inbred relatives, which he later (1966) extended to multiple alleles. DENNISTON (1967) produced a system of path counting for arriving at the probabilities, and extended his system to include two linked loci.

In this work all two-, three- and four-gene probability functions for a single locus are interrelated for all situations, one, two, three, or four individuals in which the gene states of identity can arise. The probability functions in conjunction with gene frequencies provide the joint genotypic frequencies. An algorithm is developed for the systematic computation of all the probability functions from pedigrees. For regular systems of mating the algorithm need be applied only once between successive generations and is illustrated for full sib mating and a finite monoecious population.

As mentioned before, the probability functions were used to find the covariances of inbred relatives. They can be used to compute exact probabilities of fixation and the time to fixation in finite populations, but which requires an extension of the set of functions to the order of the number of genes in the population. Most important, however, they provide models for the analyses of gene frequencies, models which clearly define parameters that are estimable and hypotheses that are testable. These models may be applied to the analyses of frequencies of individuals and groupings of individuals for which the pedigree relationships are known or, more important, for which the relationships are unknown—the usual situation in natural populations.

STATES OF IDENTITY BY DESCENT

The inbreeding coefficient, F, and coancestry, θ, are probability statements about pairs of alleles being identical by descent; F for the pair of alleles of an

---

[1] Paper No. 3175 of the Journal Series of the North Carolina State University Agricultural Experiment Station, Raleigh, N. C. This investigation was supported in part by Public Health Service Grant GM 11546 from the Division of General Medical Sciences.

individual and θ for a random pair of alleles, one from each of two individuals. Moving to the next step of three genes, for any three alleles, a, b and c, there are five states of identity by descent. These states and corresponding probabilities are

Prob	Genes Identical
γ_{abc}	$a \equiv b \equiv c$
γ_{ab}	$a \equiv b$
γ_{ac}	$a \equiv c$
γ_{bc}	$b \equiv c$
γ_{0}	None

These probabilities sum to one so that one must reckon with only four which can be further reduced to one by bringing in the probabilities for gene-pair states. The probabilities for pairs of genes being identical by descent can be found by adding the probabilities over the appropriate states:

$$\theta_{ab} = P(a \equiv b) = \gamma_{abc} + \gamma_{ab}$$
$$\theta_{ac} = P(a \equiv c) = \gamma_{abc} + \gamma_{ac}$$
$$\theta_{bc} = P(b \equiv c) = \gamma_{abc} + \gamma_{bc} \ .$$

Now, if a, b, and c are random alleles from individuals X, Y, and Z, respectively, then

$$\gamma_{XYZ} = \gamma_{abc}, \ \theta_{XY} = \theta_{ab}, \ \theta_{XZ} = \theta_{ac}, \ \theta_{YZ} = \theta_{bc},$$

and the three gene-pair probabilities are just the three coancestries for the three pairings of the individuals. Therefore, γ_{XYZ} is required in addition to the coancestries to find the probabilities for all of the three-gene states:

$$\gamma_{ab} = \theta_{XY} - \gamma_{XYZ}, \ \gamma_{ac} = \theta_{XZ} - \gamma_{XYZ}, \ \gamma_{bc} = \theta_{YZ} - \gamma_{XYZ}$$
$$\gamma_{0} = 1 - \theta_{XY} - \theta_{XZ} - \theta_{YZ} + 2\gamma_{XYZ}.$$

If the two genes a and b are contained by X, and the other, c, is a random allele from Y, then

$$\gamma_{\ddot{X}Y} = \gamma_{abc}, \quad F_{\ddot{X}} = \theta_{ab}, \quad 2\theta_{XY} = \theta_{ac} + \theta_{bc}.$$

In this case no distinction is necessary between θ_{ac} and θ_{bc}. In any application, the average, which is just the coancestry between X and Y, suffices. The other gene-pair state is for the genes of X, and the probability is the inbreeding coefficient. Note that $\gamma_{\ddot{X}Y}$ is a probability of three genes being identical by descent, and the notation indicates that the probability is for the two genes of X and a one from Y.

Turning now to four-gene states of identity, there are 15 states (GILLOIS 1965). These and their corresponding probabilities for alleles a, b, c, and d are

Prob	Genes Identical	Prob	Genes Identical
δ_{abcd}	$a \equiv b \equiv c \equiv d$	δ_{ab}	$a \equiv b$
δ_{abc}	$a \equiv b \equiv c$	δ_{ac}	$a \equiv c$
δ_{abd}	$a \equiv b \equiv d$	δ_{ad}	$a \equiv d$
δ_{acd}	$a \equiv c \equiv d$	δ_{bc}	$b \equiv c$
δ_{bcd}	$b \equiv c \equiv d$	δ_{bd}	$b \equiv d$
$\delta_{ab \cdot cd}$	$a \equiv b, c \equiv d$	δ_{cd}	$c \equiv d$
$\delta_{ac \cdot bd}$	$a \equiv c, b \equiv d$	δ_{0}	None
$\delta_{ad \cdot bc}$	$a \equiv d, b \equiv c$		

These fall into the categories of four-gene, three-gene, two-gene-pair, gene-pair and none, but all are for four-gene states identity by descent .Summarizing these in terms of three-gene and gene-pair probabilities,

$$\begin{array}{ll}
\gamma_{abc} = \delta_{abcd} + \delta_{abc}\ , & \theta_{ab} = \delta_{abcd} + \delta_{abc} + \delta_{abd} + \delta_{ab\cdot cd} + \delta_{ab} \\
\gamma_{abd} = \delta_{abcd} + \delta_{abd}\ , & \theta_{ac} = \delta_{abcd} + \delta_{abc} + \delta_{acd} + \delta_{ac\cdot bd} + \delta_{ac} \\
\gamma_{acd} = \delta_{abcd} + \delta_{acd}\ , & \theta_{ad} = \delta_{abcd} + \delta_{abd} + \delta_{acd} + \delta_{ad\cdot bc} + \delta_{ad} \\
\gamma_{bcd} = \delta_{abcd} + \delta_{bcd}\ , & \theta_{bc} = \delta_{abcd} + \delta_{abc} + \delta_{bcd} + \delta_{ad\cdot bc} + \delta_{bc} \\
 & \theta_{bd} = \delta_{abcd} + \delta_{abd} + \delta_{bcd} + \delta_{ac\cdot bd} + \delta_{bd} \\
 & \theta_{cd} = \delta_{abcd} + \delta_{acd} + \delta_{bcd} + \delta_{ab\cdot cd} + \delta_{cd}\ ,
\end{array}$$

leaves four four-gene probabilities that must be retained to account for all fifteen probabilities. These are δ_{abcd}, $\delta_{ab\cdot cd}$, $\delta_{ac\cdot bd}$, and $\delta_{ad\cdot bc}$. The latter three are easier to handle in many applications as two-gene-pair probabilities, $\Delta_{ab\cdot cd} = \delta_{abcd} + \delta_{ab\cdot cd}$, $\Delta_{ac\cdot bd} = \delta_{abcd} + \delta_{ac\cdot bd}$, and $\Delta_{ad\cdot bc} = \delta_{abcd} + \delta_{ad\cdot bc}$.

Now, if the genes a, b, c, d are random alleles from the individuals X, Y, Z, W, respectively, the fourteen probabilities may be summarized as δ_{XYZW}, $\Delta_{XY\cdot ZW}$, $\Delta_{XZ\cdot YW}$, $\Delta_{XW\cdot YZ}$, γ_{XYW}, γ_{XYZ}, γ_{XZW}, γ_{YZW}, θ_{XY}, θ_{XZ}, θ_{XW}, θ_{YZ}, θ_{YW}, θ_{ZW}, each of which may be identified with the appropriate gene-pair, three-gene, two-gene-pair, and four-gene probability functions of the genes.

If the genes a and b are contained by X and c and d are random alleles from Y and Z, respectively, the set of probabilities is as follows:

$$\begin{array}{llllll}
\delta_{\ddot{X}YZ} & = \delta_{abcd}\ , & \Delta_{\ddot{X}\cdot YZ} & = \Delta_{ab\cdot cd}\ , & 2\Delta_{\ddot{X}+YZ} & = \Delta_{ac\cdot bd} + \Delta_{ad\cdot bc} \\
\gamma_{\ddot{X}Y} & = \gamma_{abc}\ , & \gamma_{\ddot{X}Z} & = \gamma_{abd}\ , & 2\gamma_{XYZ} & = \gamma_{acd} + \gamma_{bcd} \\
F_{\ddot{X}} & = \theta_{ab}\ , & 2\theta_{XY} & = \theta_{ac} + \theta_{bc}, & 2\theta_{XZ} & = \theta_{ad} + \theta_{bd}\ .
\end{array}$$

The number of distinct probabilities is reduced to nine because four of them are averages of two. It should be stressed that $\delta_{\ddot{X}YZ}$ implies a four-gene all identity probability, and in this case it is for the two genes of X and a random gene from each of Y and Z. Similarly, Δ is for a two-gene-pair state, $\Delta_{\ddot{X}\cdot YZ}$ for the simultaneous identity of the two genes of X and of the two random genes, one from Y and one from Z, and $\Delta_{\ddot{X}+YZ}$ for the average of the two probabilities for the other two pairings of the two genes of X and a random one from each of Y and Z.

There remains to consider the case of a and b as alleles contained by X and c and d by Y. The probabilities become

$$\begin{array}{llllll}
\delta_{\ddot{X}\ddot{Y}} & = \delta_{abcd}\ , & \Delta_{\ddot{X}\cdot\ddot{Y}} & = \Delta_{ab\cdot cd}\ , & 2\Delta_{\ddot{X}+\ddot{Y}} & = \Delta_{ac\cdot bd} + \Delta_{ad\cdot bc}\ , \\
2\gamma_{\ddot{X}Y} & = \gamma_{abc} + \gamma_{abd}\ , & 2\gamma_{X\ddot{Y}} & = \gamma_{acd} + \gamma_{bcd}\ , & 4\theta_{XY} & = \theta_{ac} + \theta_{ad} + \theta_{bc} + \theta_{bd}\ , \\
F_{\ddot{X}} & = \theta_{ab}\ , & F_{\ddot{Y}} & = \theta_{cd}\ . & &
\end{array}$$

We have now related gene-pair, two-gene-pair, three-gene and four-gene probabilities for all situations, two individuals, three individuals and four individuals, in which four-gene states can arise.

The utility of the notation will be seen in pedigree manipulations and computations to be considered later. For expressing the frequencies of pairs of individuals, the notation for the nine distinct probabilities for the four-gene states is clarified.

$$
\begin{array}{lll}
\delta_{\ddot{X}\ddot{Y}} & = \delta_{\ddot{X}\ddot{Y}} & = \delta_{abcd} \\
2\delta_{\ddot{X}Y} & = 2\gamma_{\ddot{X}Y} - 2\delta_{\ddot{X}\ddot{Y}} & = \delta_{abc} + \delta_{abd} \\
2\delta_{X\ddot{Y}} & = 2\gamma_{X\ddot{Y}} - 2\delta_{\ddot{X}\ddot{Y}} & = \delta_{acd} + \delta_{bcd} \\
\delta_{\ddot{X}\cdot\ddot{Y}} & = \Delta_{\ddot{X}\cdot\ddot{Y}} - \delta_{\ddot{X}\ddot{Y}} & = \delta_{ab\cdot cd} \\
2\delta_{\ddot{X}+\ddot{Y}} & = 2\Delta_{\ddot{X}+\ddot{Y}} - 2\delta_{\ddot{X}\ddot{Y}} & = \delta_{ac\cdot bd} + \delta_{ad\cdot bc} \\
\delta_{\ddot{X}} & = F_{\ddot{X}} - \Delta_{\ddot{X}\cdot\ddot{Y}} - 2\gamma_{\ddot{X}Y} + 2\delta_{\ddot{X}\ddot{Y}} & = \delta_{ab} \\
\delta_{\ddot{Y}} & = F_{\ddot{Y}} - \Delta_{\ddot{X}\cdot\ddot{Y}} - 2\gamma_{X\ddot{Y}} + 2\delta_{\ddot{X}\ddot{Y}} & = \delta_{cd} \\
4\delta_{XY} & = 4\theta_{XY} - 4\Delta_{\ddot{X}+\ddot{Y}} - 4\gamma_{\ddot{X}Y} - 4\gamma_{X\ddot{Y}} + 8\delta_{\ddot{X}\ddot{Y}} & = \delta_{ac} + \delta_{ad} + \delta_{bc} + \delta_{bd} \\
\delta_{\overline{XY}} & = 1 - F_{\ddot{X}} - F_{\ddot{Y}} - 4\theta_{XY} + \Delta_{\ddot{X}\cdot\ddot{Y}} + 2\Delta_{\ddot{X}+\ddot{Y}} + 4\gamma_{\ddot{X}Y} + 4\gamma_{X\ddot{Y}} - 6\delta_{\ddot{X}\ddot{Y}} & = \delta_0
\end{array}
\tag{1}
$$

Illustrated also are the manipulations of the gene-pair, two-gene-pair and three- and four-gene probabilities to produce the probabilities for the four-gene states.

JOINT FREQUENCIES OF RELATIVES

In developing the joint frequencies it is assumed that identity by descent is independent of the genes in the population. Then, the total array of pairs of relatives may be summarized as did HARRIS (1964), where p_i is the frequency of allele a_i,

$$
\begin{array}{llll}
\delta_{\ddot{X}\ddot{Y}} & \sum_i p_i (\mathrm{a}_i\mathrm{a}_i)(\mathrm{a}_i\mathrm{a}_i) & +2\delta_{\ddot{X}Y} & \sum_{ij} p_i p_j (\mathrm{a}_i\mathrm{a}_i)(\mathrm{a}_i\mathrm{a}_j) \\
+2\delta_{X\ddot{Y}} & \sum_{ij} p_i p_j (\mathrm{a}_i\mathrm{a}_j)(\mathrm{a}_i\mathrm{a}_i) & +2\delta_{\ddot{X}+\ddot{Y}} & \sum_{ij} p_i p_j (\mathrm{a}_i\mathrm{a}_j)(\mathrm{a}_i\mathrm{a}_j) \\
+\delta_{\ddot{X}\cdot\ddot{Y}} & \sum_{ij} p_i p_j (\mathrm{a}_i\mathrm{a}_i)(\mathrm{a}_j\mathrm{a}_j) & +\delta_{\ddot{X}} & \sum_{ijk} p_i p_j p_k (\mathrm{a}_i\mathrm{a}_i)(\mathrm{a}_j\mathrm{a}_k) \\
+\delta_{\ddot{Y}} & \sum_{ijk} p_i p_j p_k (\mathrm{a}_i\mathrm{a}_j)(\mathrm{a}_k\mathrm{a}_k) & +4\delta_{XY} & \sum_{ijk} p_i p_j p_k (\mathrm{a}_i\mathrm{a}_j)(\mathrm{a}_i\mathrm{a}_k) \\
 & & +\delta_{\overline{XY}} & \sum_{ijkl} p_i p_j p_k p_l (\mathrm{a}_i\mathrm{a}_j)(\mathrm{a}_k\mathrm{a}_l) \; ,
\end{array}
\tag{2}
$$

where the pairs of genotypes are for (X)(Y). A bit more explanation than HARRIS gave may be helpful. If all four genes are identical by descent then they will all be a_i with frequency p_i (first term). If only three are identical by descent then the three will be a_i with frequency p_i and the other gene will be a_j (including $j = i$) with frequency p_j, and so on to the case where none are identical by descent and the genotypes occur together as in a randomly mating population (last term).

To obtain the frequency for a specific pair of genotypes one must collect

TABLE 1

Joint genotypic frequencies

X	Y	X	Y	X	Y	X	Y
a_ia_i	a_ia_i	a_ia_i	a_ia_j	a_ia_i	a_ja_j	a_ia_j	a_ia_j
$\delta_{\ddot{X}\ddot{Y}}$	p_i	$\delta_{\ddot{X}Y}$	$2p_ip_j$	$\delta_{\ddot{X}\cdot\ddot{Y}}$	p_ip_j	$\delta_{\ddot{X}+\ddot{Y}}$	$4p_ip_j$
$+2\delta_{\ddot{X}Y}$	p_i^2	$+\delta_{\ddot{X}}$	$2p_i^2p_j$	$+\delta_{\ddot{X}}$	$p_ip_j^2$	$+\delta_{XY}$	$4p_i^2p_j$
$+2\delta_{\ddot{Y}X}$	p_i^2	$+2\delta_{XY}$	$2p_i^2p_j$	$+\delta_{\ddot{Y}}$	$p_i^2p_j$	$+\delta_{XY}$	$4p_ip_j^2$
$+2\delta_{\ddot{X}+\ddot{Y}}$	p_i^2	$+\delta_{\overline{XY}}$	$2p_i^3p_j$	$+\delta_{\overline{XY}}$	$p_i^2p_j^2$	$+\delta_{\overline{XY}}$	$4p_i^2p_j^2$
$+\delta_{\ddot{X}\cdot\ddot{Y}}$	p_i^2						
$+\delta_{\ddot{X}}$	p_i^3	X	Y	X	Y	X	Y
$+\delta_{\ddot{Y}}$	p_i^3	a_ia_i	a_ja_k	a_ia_j	a_ia_k	a_ia_j	a_ka_ℓ
$+4\delta_{XY}$	p_i^3	$\delta_{\ddot{X}}$	$2p_ip_jp_k$	δ_{XY}	$4p_ip_jp_k$	$\delta_{\overline{XY}}$	$4p_ip_jp_kp_\ell$
$+\delta_{\overline{XY}}$	p_i^4	$+\delta_{\overline{XY}}$	$2p_i^2p_jp_k$	$+\delta_{\overline{XY}}$	$4p_i^2p_jp_k$		

together the appropriate frequencies from the various terms in (2). The seven possible types of pairs of genotypes and their frequencies are displayed in Table 1. From these a joint frequency table can be written for any number of alleles by appropriate interchanges of X and Y.

GENERAL PEDIGREE METHOD OF DETERMINING THE PROBABILITIES

While it is simple to find the probabilities for some relatives, the arguments become tedious for complex pedigrees and a systematic approach is required. The handling of F and θ is well known. To this list we want to add the two-gene-pair probability, Δ, the three-gene probability, γ, and the four-gene probability, δ. The expansion of these probabilities to probabilities of states of identity of genes back in the pedigree will be indicated by expanding the subscripts according to the origin of the genes. For example, if the two parents of X are M and N, then we replace X by (MN). If the probability involves the two genes within X, then the replacement means a random gene from M and a random gene from N. If the probability involves a random gene from X, then replacement implies a random gene from M half of the time and a random gene from N half of the time. For two individuals, neither of which is an ancestor of the other, the expansions are continued until a common ancestor is indicated, and special consideration is given to these probabilities. When one is an ancestor of the other, the descendant is always expanded back to the ancestor.

Consider first the three-gene probabilities, γ's. The probability $\gamma_{\ddot{X}Y}$ is for the two genes of X and a random one from Y. Bringing in the parents of X,

$$\gamma_{\ddot{X}Y} = \gamma_{(MN)Y} = \gamma_{MNY} \ ,$$

and it is just the three-gene probability for a random gene from each of M, N, and Y. Bringing in the parents of Y first,

$$\gamma_{\ddot{X}Y} = \gamma_{\ddot{X}(ZW)} = 1/2\,(\gamma_{\ddot{X}Z} + \gamma_{\ddot{X}W}) \ ,$$

averages the corresponding function over the parents. Further, bringing in the parents of X leads to

$$\gamma_{\ddot{X}Y} = 1/2\,[\gamma_{(MN)Z} + \gamma_{(MN)W}] = 1/2\,(\gamma_{MNZ} + \gamma_{MNW}) \ .$$

The same result is obtained by the other route,

$$\gamma_{\ddot{X}Y} = \gamma_{MNY} = \gamma_{MN(ZW)} = 1/2\,(\gamma_{MNZ} + \gamma_{MNW}) \ ,$$

and the process is continued by bringing in the parents of each until genes from the same individual become involved. Expansions must always proceed to bring in common individuals. If Y is a parent of X then expand on X, $\gamma_{\ddot{X}Y} = \gamma_{(MY)Y} = \gamma_{MYY}$ and not Y, but if X is a parent of Y, the expansion must be on Y.

Special probability arguments are required when genes of the same individual become involved,

$$\gamma_{XXY} = 1/2\,(\theta_{XY} + \gamma_{\ddot{X}Y}), \ \gamma_{XXX} = 1/4\,(1 + 3F_{\ddot{X}}), \ \gamma_{\ddot{X}X} = F_{\ddot{X}} \ .$$

The case of γ_{XXY} is for two random genes (gametes) from X and one from Y. The two random genes from X are the same parental gene ½ the time, and are identical by descent, and are identical to the one from Y with probability θ_{XY}. They are different parental genes ½ the time but are identical and identical to the one from Y with probability $\gamma_{\ddot{X}Y}$. In the case of three random gametes from the same individual, γ_{XXX}, they are all the same parental gene ¼ the time and thus are identical by descent, and the remainder of the time, ¾, they are all identical when the two parental genes are identical, $\gamma_{\ddot{X}X} = F_{\ddot{X}}$. Only these three special functions are required in the evaluation of all other functions involving common ancestors.

Since δ is always for four genes, random gametes of parents are introduced in parentheses until four random gametes are involved, after which parentheses play the role of introducing more parental gametes to be averaged over. Expansion must always be from the descendant to an ancestor. When one relative is not an ancestor of the other the expansion may proceed in terms of the origins of genes of both relatives,

$$\delta_{\ddot{X}\ddot{Y}} = \delta_{\ddot{X}(ZW)} = \delta_{(MN)(ZW)} = \delta_{MNZW} \ ,$$

and bringing in the parents of M,

$$\delta_{MNZW} = \delta_{(UV)NZW} = 1/2\,(\delta_{UNZW} + \delta_{VNZW}) \ .$$

For an individual and its gametes (offspring if the gametes are united),

$$\delta_{\ddot{X}(XY)} = \gamma_{\ddot{X}Y} , \qquad \delta_{\ddot{X}(XX)} = F_{\ddot{X}} = \delta_{\ddot{X}\ddot{X}} .$$

Other types of relatives are indicated by parenthetical punctuation.

$$\delta_{(XX)(XX)} = \delta_{XXXX} = \tfrac{1}{8}(1 + 7F_{\ddot{X}})$$

$$\delta_{(XX)(XY)} = \delta_{XXXY} = \tfrac{1}{4}(\theta_{XY} + 3\gamma_{\ddot{X}Y})$$

$$\delta_{(XX)(YY)} = \delta_{(XY)(XY)} = \delta_{XXYY} = \tfrac{1}{4}(\theta_{XY} + \gamma_{\ddot{X}Y} + \gamma_{\ddot{Y}X} + \delta_{\ddot{X}\ddot{Y}})$$

$$\delta_{(XY)(XZ)} = \delta_{(XX)(YZ)} = \delta_{XXYZ} = \tfrac{1}{2}(\gamma_{XYZ} + \delta_{\ddot{X}YZ}) .$$

Similar procedures are applied to the two-gene-pair states. If one relative is not the ancestor of the other,

$$\Delta_{\ddot{X}\cdot\ddot{Y}} = \Delta_{\ddot{X}\cdot(ZW)} = \Delta_{(MN)\cdot(ZW)} = \Delta_{MN\cdot ZW}$$

and it is the simultaneous identity of each separated gene pair either in individuals or in pairs of random gametes that is considered. The other two-gene-pair probability,

$$\Delta_{\ddot{X}+\ddot{Y}} = \Delta_{(MN)+(ZW)} = \Delta_{MN+ZW} = \tfrac{1}{2}(\Delta_{MZ\cdot NW} + \Delta_{MW\cdot NZ}) ,$$

is the average of the two probabilities for the other two possible pairings. Bringing in the parents of M, the expansions proceed in the usual manner,

$$\Delta_{(UV)N\cdot ZW} = \tfrac{1}{2}(\Delta_{UN\cdot ZW} + \Delta_{VN\cdot ZW}) ,$$
$$\Delta_{(UV)N+ZW} = \tfrac{1}{2}(\Delta_{UN+ZW} + \Delta_{VN+ZW}) .$$

The descendant is expanded back to the ancestor, and the probabilities for an individual and its random gametes (or offspring) are

$$\Delta_{\ddot{X}\cdot XY} = \gamma_{\ddot{X}Y} , \quad \Delta_{\ddot{X}+XY} = \tfrac{1}{2}(\theta_{XY} + \gamma_{\ddot{X}Y}) , \quad \Delta_{\ddot{X}\cdot XX} = F_{\ddot{X}} , \quad \Delta_{\ddot{X}+XX} = \tfrac{1}{4}(1 + 3F_{\ddot{X}}) .$$

Further manipulation shows that

$$\Delta_{XX\cdot XX} = \Delta_{XX+XX} = \Delta_{\ddot{X}+XX} , \qquad \Delta_{XX\cdot XY} = \Delta_{XX+XY} = \Delta_{\ddot{X}+XY} .$$

Other special cases are

$$\Delta_{XX\cdot YY} = \tfrac{1}{4}(1 + F_{\ddot{X}} + F_{\ddot{Y}} + \Delta_{\ddot{X}\cdot\ddot{Y}}) , \qquad \Delta_{XY\cdot XY} = \tfrac{1}{4}(\theta_{XY} + \gamma_{\ddot{X}Y} + \gamma_{\ddot{Y}X} + \Delta_{\ddot{X}+\ddot{Y}}) ,$$

$$\Delta_{XX\cdot YZ} = \tfrac{1}{2}(\Delta_{\ddot{X}\cdot YZ} + \theta_{YZ}) , \qquad \Delta_{XY\cdot XZ} = \tfrac{1}{2}(\Delta_{\ddot{X}+YZ} + \gamma_{XYZ}) .$$

From these the other two-gene-pair probabilities are found,

$$\Delta_{XX+YY} = \Delta_{XY\cdot XY} , \qquad \Delta_{XY+XY} = \tfrac{1}{2}(\Delta_{XX\cdot YY} + \Delta_{XY\cdot XY}) ,$$
$$\Delta_{XX+YZ} = \Delta_{XY\cdot XZ} , \qquad \Delta_{XY+XZ} = \tfrac{1}{2}(\Delta_{XX\cdot YZ} + \Delta_{XY\cdot XZ}) .$$

We now have a systematic method of calculating all of the probabilities for a specific pedigree or for some system of mating.

SYSTEMS OF MATING

When the system of mating is the same over time, it is not required to apply the pedigree method over more than one generation. One simply finds the expressions, transition matrix, for the probabilities in one generation in terms of the probabilities of the previous generation. Then the probabilities for any generation can be found by powering the matrix and multiplying by the vector of initial conditions.

As an example consider full sib mating. The results of the previous pedigree method may be applied directly, (t indexes the generations)

$$\delta_{\ddot{X}\ddot{Y}t+1} = \delta_{(XY)(XY)t} = \tfrac{1}{4}(\theta_{XYt} + \gamma_{\ddot{X}Yt} + \gamma_{\ddot{Y}Xt} + \delta_{\ddot{X}\ddot{Y}t}) \ .$$

Because of the symmetry between full sibs, $\gamma_{\ddot{X}Y} = \gamma_{\ddot{Y}X}$ and $F_{\ddot{X}} = F_{\ddot{Y}}$. Let $\Delta = \Delta_{\ddot{X}\cdot\ddot{Y}}$ and $\Delta' = \Delta_{\ddot{X}+\ddot{Y}}$ so that the X,Y subscripts can be dropped. Then,

$$\begin{aligned}
\delta_{t+1} &= \tfrac{1}{4}(\theta_t + 2\gamma_t + \delta_t) & \gamma_{t+1} &= \tfrac{1}{2}(\gamma_t + \theta_t) \\
\Delta'_{t+1} &= \tfrac{1}{8}(1 + 2\gamma_t + \Delta'_t + \Delta_t + 2F_t + 2\theta_t) & \Delta_{t+1} &= \tfrac{1}{4}(2\gamma_t + \Delta'_t + \theta_t) \quad (3) \\
\theta_{t+1} &= \tfrac{1}{4}(1 + F_t + 2\theta_t) & F_{t+1} &= \theta_t
\end{aligned}$$

the other expressions being found in a similar manner. It is easiest to work with the complements of these, analogous to the panmictic coefficient, 1−F. Let

$$g'_t = (1-\delta_t, 1-\gamma_t, 1-\Delta'_t, 1-\Delta_t, 1-F_t, 1-\theta_t) \ .$$

Then, $g_{t+1} = Ag_t$, where

$$A = \begin{bmatrix}
\tfrac{1}{4} & \tfrac{1}{2} & 0 & 0 & 0 & \tfrac{1}{4} \\
0 & \tfrac{1}{2} & 0 & 0 & 0 & \tfrac{1}{2} \\
0 & \tfrac{1}{4} & \tfrac{1}{8} & \tfrac{1}{8} & \tfrac{1}{4} & \tfrac{1}{8} \\
0 & \tfrac{1}{2} & \tfrac{1}{4} & 0 & 0 & \tfrac{1}{4} \\
0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & \tfrac{1}{4} & \tfrac{1}{2}
\end{bmatrix}$$

is the transition matrix, and $g_t = A^t g_0$. Of course, 1−F and 1−θ may be treated separately from the others, as they have been in the past, since their transition does not depend on the other probabilities.

Alternatively, one may make substitutions (1) into both sides of (3) to arrive at the relationships between the probabilities for distinct four-gene states of identity of the two generations. Let

$$h'_t = (\delta_{\ddot{X}\ddot{Y}t}, 4\delta_{\ddot{X}Yt}, 2\delta_{\ddot{X}+\ddot{Y}t}, \delta_{\ddot{X}\cdot\ddot{Y}t}, 2\delta_{\ddot{X}t}, 4\delta_{XYt}, \delta_{\overline{XYt}}) \ ,$$

then $h_t = B^t h_0$, where

$$B = \begin{bmatrix}
1 & \tfrac{1}{4} & \tfrac{1}{8} & 0 & 0 & \tfrac{1}{16} & 0 \\
0 & \tfrac{1}{2} & \tfrac{1}{2} & 0 & 0 & \tfrac{1}{4} & 0 \\
0 & \tfrac{1}{4} & \tfrac{1}{4} & 1 & \tfrac{1}{2} & \tfrac{3}{16} & \tfrac{1}{4} \\
0 & 0 & \tfrac{1}{8} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \tfrac{1}{8} & 0 \\
0 & 0 & 0 & 0 & \tfrac{1}{2} & \tfrac{3}{8} & \tfrac{1}{2} \\
0 & 0 & 0 & 0 & 0 & 0 & \tfrac{1}{4}
\end{bmatrix}$$

is the generation matrix given by FISHER (1949, Table 5) relating the frequencies of mating types in adjacent generations for sib mating. FISHER produced the latent roots and a solution for B^t.

Of more interest is a finite randomly mating population. For simplicity let it be monoecious, and the sex of the gamete be random with respect to any parent (the combined sampling plan of WEIR and COCKERHAM 1969). In order to argue the gametic probability functions back through the parents to gametes in the previous generation one must take into account the ways the gametes relate to the parents. For a random pair of gametes they may be both from the same parent or one from each of two parents with probabilities P^2 and P^{11}, respectively, $P^2 + P^{11} = 1$. For three gametes the probabilities are P^3, P^{21} and P^{111}, and for four gametes are P^4, P^{31}, P^{22}, P^{211}, and P^{1111}, where the notation indicates the ways of obtaining three and four gametes from one, two, three, and four parents. (These P's are considered in detail in WEIR and COCKERHAM 1969).

Since genes in random sets of gametes bear the same relationship to each other regardless of how they are constituted within or between individuals, certain probability functions such as F and θ will be the same. We denote these probability functions in a manner corresponding to the P's. For example F^2_{t+1} is for a pair of gametes from the same parent, and consequently,

$$F^2_{t+1} = \frac{1+F_t}{2} \ ,$$
$$F^{11}_{t+1} = \theta_t = F_t \ ,$$

F^{11} being for two gametes from separate parents. Then, F_{t+1} is just the average,

$$F_{t+1} = P^2F^2_{t+1} + P^{11}F^{11}_{t+1} = \frac{P^2}{2} + \frac{(P^2+2P^{11})}{2} F_t$$

or

$$(1-F_{t+1}) = \frac{P^2+2P^{11}}{2} (1-F_t) \ .$$

The same procedure is followed for the other functions.

$$\gamma^3_{t+1} = \frac{1+3F_t}{4}, \ \gamma^{21}_{t+1} = \frac{F_t+\gamma_t}{2}; \ \gamma^{111}_{t+1} = \gamma_t$$

$$1-\gamma_{t+1} = \frac{P^{21}+2P^{111}}{2} (1-\gamma_t) + \frac{3P^3+2P^{21}}{4} (1-F_t)$$

$$\delta^4_{t+1} = \frac{1+7F_t}{8}, \ \delta^{31}_{t+1} = \frac{F_t+3\gamma_t}{4}, \ \delta^{22}_{t+1} = \frac{F_t+2\gamma_t+\delta_t}{4}$$

$$\delta^{211}_{t+1} = \frac{\gamma_t+\delta_t}{2}, \ \delta^{1111}_{t+1} = \delta_t$$

$$1-\delta_{t+1} = \frac{P^{22}+2P^{211}+4P^{1111}}{4} (1-\delta_t) + \frac{3P^{31}+2P^{22}+2P^{211}}{4} (1-\gamma_t)$$
$$+ \frac{7P^4+2P^{31}+2P^{22}}{8} (1-F_t) \ .$$

Only one Δ is required since the probability is the same for a random set of two pairs of gametes in any arrangement. The three arrangements are argued separately for P^{211} and P^{22} and averaged.

$$\Delta^4_{t+1} = \frac{1+3F_t}{4}, \quad \Delta^{31}_{t+1} = \frac{F_t+\gamma_t}{2}, \quad \Delta^{211}_{t+1} = \frac{F_t+2\gamma_t+3\Delta_t}{6}$$

$$\Delta^{22}_{t+1} = \frac{1+4F_t+4\gamma_t+3\Delta_t}{12}, \quad \Delta^{1111}_{t+1} = \Delta_t$$

$$1-\Delta_{t+1} = \frac{P^{22}+2P^{211}+4P^{1111}}{4}(1-\Delta_t) + \frac{3P^{31}+2P^{22}+2P^{211}}{6}(1-\gamma_t) + \frac{9P^4+6P^{31}+4P^{22}+2P^{211}}{12}(1-F_t) \, .$$

With these results, we may set up the transition matrix C for the probability functions

$$k'_t = (1-\Delta_t, 1-\delta_t, 1-\gamma_t, 1-F_t)$$

and

$$k_t = C^t k_0 \, .$$

The values of the P's depend on the distribution of the number of offspring per parent and are discussed in WEIR and COCKERHAM (1969). With equal chance of each parent producing any gamete,

$$C = \begin{bmatrix} \frac{(2N-1)(2N-2)(2N-3)}{(2N)^3} & 0 & \frac{4(2N-1)(2N-2)}{(2N)^3} & \frac{2(4N^2-1)}{(2N)^3} \\ 0 & \frac{(2N-1)(2N-2)(2N-3)}{(2N)^3} & \frac{6(2N-1)(2N-2)}{(2N)^3} & \frac{7(2N-1)}{(2N)^3} \\ 0 & 0 & \frac{(2N-1)(2N-2)}{(2N)^2} & \frac{3(2N-1)}{(2N)^2} \\ 0 & 0 & 0 & \frac{2N-1}{2N} \end{bmatrix}$$

where N is the number of individuals. The roots of the matrix are the diagonal terms. The asymptotic behavior of all elements of k is determined by the largest root, $(2N-1)/2N$, the one corresponding to $1-F$.

COMMENTS

To arrive at the joint frequencies of relatives as was done in Table 1, it is assumed that the founder individuals are not inbred and not related and with alleles associated at random, i.e., HARDY-WEINBERG proportions. For example, it may be verified for full sib mating that letting $\delta_{\overline{XY}} = 1$ for the initial mate pairs and applying (1) gives the same results as applying the transition (B) directly to the frequencies of initial mate pairs in HARDY-WEINBERG proportions and mated at random.

The parameterization provides a complete description of populations, sib mating, parent offspring or random mating, in which only four genes are involved. More genes require an extension of the probability functions to accommodate all possible states of identity. The number of states increases at an alarming rate with N. If, however, one is primarily interested in evaluating rates and average times to fixation or loss of genes then only $2N-1$ probability functions are required. KEMPTHORNE (1967) gave the transition matrix C extended to accommodate the probability functions $1-\delta^i$, $i=2, 3 \ldots 2N$ for random sampling of genes.

The probability of being fixed at generation t is δ_t^{2N}, and $\delta_t^{2N}-\delta_{t-1}^{2N}$ is the probability of becoming fixed at time t. The average time to fixation for a neutral gene that is being fixed is $\bar{t}_{2N} = \sum_{t=1}^{\infty} (\delta_t^{2N}-\delta_{t-1}^{2N})t$. Now consider a general transition matrix of dimensions K × K corresponding to C for the vector containing $1-\delta^{2N}$. There is a solution of the form

$$1-\delta_t^{2N} = \sum_{i=1}^{K} a_i \lambda_i^t$$

where λ_i's are roots of the matrix and the a_i's are determined by the initial conditions, and also $\sum_{i=1}^{K} a_i = 1$. The probability of becoming fixed at time t is found to be

$$\delta_t^{2N}-\delta_{t-1}^{2N} = \sum_{i=1}^{K} a_i(1-\lambda_i)\lambda_i^{t-1}$$

and the average time to fixation is

$$\bar{t}_{2N} = \sum_{t=1}^{\infty} \sum_{i=1}^{K} a_i(1-\lambda_i)\lambda_i^{t-1}t = \sum_{i=1}^{K} \frac{a_i}{1-\lambda_i}$$

For matrix C the roots are just the diagonal terms which would generally be the case for the monoecious system. For $2N=2$, $K=1$, $a_1 = 1$, $\lambda_1 = 1/2$, and $\bar{t}_2 = 2$, the usual result for self-fertilization. For $2N = 3$, not biologically reasonable, $K=2$, $a_1 = 3/2$, $\lambda_1 = 2/3$, $a_2 = -1/2$, $\lambda_2 = 2/9$, and $\bar{t}_3 = 27/7$. For $2N = 4$, $K = 3$, $a_1 = 25/14$, $\lambda_1 = 3/4$, $a_2 = -1$, $\lambda_2 = 3/8$, $a_3 = 3/14$, $\lambda_3 = 3/32$, and $\bar{t}_4 = 5866/1015$. The function $t_{2N}^* = \bar{t}_{2N}/(2N-1)$ is of interest.

$$t_2^* = 2, \quad t_3^* = 1.9286, \quad t_4^* = 1.9265,$$

because it provides the rate of loss of neutral genes, and which appears to be close to one gene every two generations. Thus the results of KIMURA and OHTA (1969) appear to approximate well those for small populations with the substitution of $2N-1$ for $2N$, which is of no consequence if N is large. Other facets, such as the effects of small and large variation in the number of offspring per parent, need a more thorough investigation.

The four-gene parameters provide the frequencies of pairs of disomic relatives for any system of mating. It was this description that permitted HARRIS (1964) to formulate generally the covariances among relatives for a single locus, and the extension to multiple loci when genotypes of different loci are not correlated (linkage and identity equilibrium). However, loci which do not recombine freely are not in identity equilibrium during intermediate stages of inbreeding, although the founder population is in linkage equilibrium (WEIR and COCKERHAM 1969). Consequently, his results require some modification for linked genes.

Probably most important, the four-gene parameters provide a model for estimation and testing of hypotheses with gene frequency data. The individuals need not necessarily be relatives but other groupings or pairings such as mates for which the parameters are unknown. One procedure of analysis is to use analysis of variance (quadratic) techniques for estimation and testing of hypotheses involving two-gene parameters, as exemplified by COCKERHAM (1969) for groupings of individuals. There is a logical extension to components of higher moments, cubic and quartic techniques, for three- and four-gene parameters. While test statistics are not available for testing these higher order functions directly, the functions do provide unbiased estimators of classes of frequencies to be used as expected values for χ^2 testing, which is important unless the data are extensive.

Alternatively, one may use maximum likelihood estimation procedures. In any case the model delimits all possible associations of the genes whatever the causes for these associations.

SUMMARY

All two-, three- and four-gene and two-gene-pair probability functions for a single locus are interrelated for all situations, one, two, three, and four individuals, in which the gene states of identity can arise. Systematic procedures are developed for computing the probabilities from pedigrees. In the case of regular systems of mating, these procedures need be applied for only one generation, and are illustrated for sib mating and a randomly mating finite monoecious population. The method produces exact probabilities of fixation and average times to fixation. Extensions to an arbitrary size of monoecious population are suggested. The four-gene probability functions provide joint genotypic frequencies for any type of relatives. They also provide a model for the analysis of frequency data for pairs of individuals, such as mate pairs, for which the relationships are unknown.

LITERATURE CITED

COCKERHAM, C. CLARK, 1969 Variance of gene frequencies. Evolution **23**: 72–84.

DENNISTON, C., 1967 Probability and genetic relationship. Unpublished Ph.D., thesis, Library, University of Wisconsin, Madison, Wisconsin.

FISHER, R. A., 1949 *The Theory of Inbreeding.* Oliver and Boyd Ltd., Edinburgh.

GILLOIS, M., 1965 Relation d'identité en génétique. I. Postulats et axiomes mendéliens. II. Corrélation génétique dans le cas de dominance. Ann. Inst. Henri Poincaré. Section B. **II**: 1–94. ——, 1966 Note sur la variance et la covariance génotypiques entre apparentés. Ann. Inst. Henri Poincaré. Section B. **II**: 349–352.

HARRIS, D. L., 1964 Genotypic covariances between inbred relatives. Genetics **50**: 1319–1348.

KEMPTHORNE, O., 1967 The concept of identity of genes by descent. Proc. of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Vol. IV: 333–348.

KIMURA, M. and T. OHTA, 1969 The average number of generations until fixation of a mutant gene in a finite population. Genetics **61**: 763–771.

WEIR, B. S. and C. CLARK COCKERHAM, 1969 Pedigree mating with two linked loci. Genetics **61**: 923–940.