LIMITING DISTRIBUTION UNDER ASSORTATIVE MATING*

G. L. GHAI

Statistical Laboratory, Iowa State University, Ames, Iowa 50010

Manuscript received October 9, 1972
Revised copy received July 9, 1973

ABSTRACT

A multi-locus model for complete positive assortative mating is discussed. For a two-locus model, if the gene frequencies for the two loci are different, as they are likely to be, it is shown that in equilibrium the population is not composed of only two homozygous types, as is usually thought. The limiting distribution will have three homozygous genotypes depending upon the initial gene frequencies. If there are m-loci such that gene frequencies at all loci are different, there will be $(m+1)$ such homozygous genotypes present in the equilibrium population, one in each phenotypic group.

ASSORTATIVE mating is a form of nonrandom mating where mating is based on the phenotypic properties of mates. This tendency to mate assortatively is known to occur in certain bird, mammal, and insect populations. In human populations where homogamy prevails, conventional barriers greatly restrict the choice of a mate. Hence human parents tend to resemble one another in heritable characters, e.g., height, intelligence, etc.

The simplest model of complete positive assortative mating for a single locus with complete dominance was discussed initially by JENNINGS (1916). These results were later extended by WENTWORTH and REMICK (1916) who gave a more general formula. Positive assortative mating leads to complete homozygosity of a population, though very slowly. In the case of a single locus with alleles A and a and complete dominance, the frequency of Aa individuals after n generations of positive assortative mating is given by (e.g., LI 1955, p. 234)

$$H_n = \frac{2p_A H_0}{2p_A + nH_0}$$

where H_0 is the initial heterozygosity, and p_A is the gene frequency of A which is invariant over time. If there is no dominance the results are similar to those under selfing. In the limit as $n \to \infty$, $H_n \to 0$, and AA and aa individuals occur with frequency p_A and p_a, respectively.

Consider now a population segregating with respect to m loci, each with two alleles. We assume that there is no dominance at either locus, and that the effects of different loci are equal and additive. Let us define $\{f^{(n)}_{ijk\ldots}\}$, $n = 0,1,2,\ldots,$ to be the probability distribution of genotypes in generation n, where $i = 0,1,2,$ de-

* Journal Paper No. J-7399 of the Iowa Agriculture and Home Economics Experiment Station, Project 1669. Partial support provided by National Institute of Health Grant No. GM 13827.

pending on the number of A genes in the genotype for the A-a locus and, likewise, for $j,k, \ldots$ etc. For example, for a two-locus model, $f_{12}^{(n)}$ will be the probability of genotype $AaBB$ ($P[AaBB]$) in generation n. The initial distribution corresponds to $n=0$. The frequencies in the limit as $n \to \infty$ are defined by $\{f_{ijk \ldots}\}$. With m loci, there are $(2m+1)$ phenotypic groups, and the individuals within each group mate at random.

First, we give a complete analysis of the two-locus case. We then consider the cases of 3 and $m(>3)$ loci. Some of the recurrence relations are obtained under the assumption of no linkage. But such an assumption is not necessary and is merely introduced for simplification. Linkage will lead to the same limiting distributions, though the rate of approach will be affected.

TWO-LOCUS MODEL

Consider the case of complete positive assortative mating with respect to a character which depends upon two factors, (A,a) and (B,b). WRIGHT (1921) discussed this model with equal gene frequencies of one-half of each allele. He concluded that assortative mating based on the resemblance leads to a composition of the population very different from that reached under inbreeding. With perfect assortative mating, a two-factor population is converted ultimately into only two extreme types $AABB$ and $aabb$. With inbreeding, all the four homozygous types tend equally toward fixation.

Obviously, WRIGHT's conclusion was based on the assumption of equal gene frequencies. But this result is often quoted without mentioning this assumption (e.g., LI 1955, p. 237; CROW and FELSENSTEIN, 1968) and without realizing that this result may not be valid in a more general situation. For example, LI states that ". . . for a metric character dependent on two pairs of genes with additive and equal effects (Ch. 8, Ex. 5) complete assortative mating within each of the five phenotypes would ultimately lead to a population consisting of the two extreme types. . . ." That this result is not true in a general situation, and in particular in the case of Ex. 5 (Ch. 8) of LI, can be easily seen if one recalls that with this kind of mating system gene frequencies are invariant over time. However, as we shall show, such a result will still hold if $p_B = p_A$, not necessarily one-half.

TABLE 1

Distribution of phenotypes

Group	Phenotype	Value	Probability
1	*AABB*	4	$f_{22}^{(n)}$
2	*AABb, AaBB*	3	$f_{21}^{(n)} + f_{12}^{(n)}$
3	*AAbb, AaBb, aaBB*	2	$f_{20}^{(n)} + f_{11}^{(n)} + f_{02}^{(n)}$
4	*Aabb, aaBb*	1	$f_{10}^{(n)} + f_{01}^{(n)}$
5	*aabb*	0	$f_{00}^{(n)}$

There are five phenotypic groups on the basis of which assortative mating is to be made. Their distribution is given in Table 1. Within each group, individuals mate at random. Group 1 individuals have progeny of their own type; group 2 produces individuals with frequencies proportional to

$$[(1/2)\{f_{12}^{(n)}+f_{21}^{(n)}\}AB+(1/2)f_{12}^{(n)}aB+(1/2)f_{21}^{(n)}Ab]^2$$

and so on. Thus, one generation of assortative mating yields, for example, in the case of no linkage,

$$\begin{aligned}
f_{22}^{(n+1)} &= f_{22}^{(n)}+(1/4)\{f_{21}^{(n)}+f_{12}^{(n)}\}+(1/16)\frac{\{f_{11}^{(n)}\}^2}{f_{11}^{(n)}+f_{20}^{(n)}+f_{02}^{(n)}} \\
f_{21}^{(n+1)} &= (1/2)f_{21}^{(n)}+(1/2)f_{11}^{(n)}\frac{\{f_{20}^{(n)}+(1/4)f_{11}^{(n)}\}}{f_{11}^{(n)}+f_{20}^{(n)}+f_{02}^{(n)}} \qquad (1) \\
f_{20}^{(n+1)} &= \frac{(1/4)\{f_{21}^{(n)}\}^2}{f_{21}^{(n)}+f_{12}^{(n)}}+\frac{\{f_{20}^{(n)}+(1/4)f_{11}^{(n)}\}^2}{f_{11}^{(n)}+f_{20}^{(n)}+f_{02}^{(n)}}+\frac{(1/4)\{f_{10}^{(n)}\}^2}{f_{10}^{(n)}+f_{01}^{(n)}}.
\end{aligned}$$

We can write similarly the probabilities of other genotypes. We can easily show that $p_A^{(n+1)}=p_A^{(n)}$, $p_B^{(n+1)}=p_B^{(n)}$; i.e., the gene frequencies remain unchanged.

We note that $\{f_{22}^{(n)}\}_n$ is a monotonic nondecreasing bounded sequence and therefore will converge. Let $f_{22}^{(n)}\to f_{22}$ as $n\to\infty$. Then from the first equation in (1) we get, as $n\to\infty$,

$$f_{11}^{(n)}\to 0,\quad f_{21}^{(n)}\to 0,\quad f_{12}^{(n)}\to 0\,.$$

We can also show in a similar manner that as $n\to\infty$, $f_{10}^{(n)}\to 0$ and $f_{01}^{(n)}\to 0$. Hence, with this system of mating, all heterozygosity will be lost ultimately.

If the two loci are linked with the recombination fraction $\rho(0\le\rho\le 1/2)$, then the recurrence relations will be altered. For example, the first equation in (1) will become

$$f_{22}^{(n+1)}=f_{22}^{(n)}+(1/4)\{f_{21}^{(n)}+f_{12}^{(n)}\}+\frac{\left\{\frac{1-\rho}{2}f_{11c}^{(n)}+\frac{\rho}{2}f_{11r}^{(n)}\right\}^2}{f_{11}^{(n)}+f_{20}^{(n)}+f_{02}^{(n)}}$$

where $f_{11c}=P[AB/ab]$
$f_{11r}=P[Ab/aB]$
and $f_{11}=f_{11c}+f_{11r}$.

Since the limit of $f_{22}^{(n)}$ exists, we get, as before, as $n\to\infty$,

$$f_{21}^{(n)}\to 0\quad f_{12}^{(n)}\to 0$$
$$f_{11c}^{(n)}\to 0\quad f_{11r}^{(n)}\to 0 \Rightarrow f_{11}^{(n)}\to 0\,.$$

We can also show similarly that $f_{10}^{(n)}\to 0$ and $f_{01}^{(n)}\to 0$. Hence, again all heterozygosity will be lost ultimately.

Thus we get the same results in the limit whether the two loci are linked or segregate independently. This is also true when more than two loci are involved.

In the limiting case, let f_{22}, f_{20}, f_{02}, and f_{00} be the frequencies of the homozygous genotypes *AABB*, *AAbb*, *aaBB*, and *aabb* respectively, such that $f_{22} + f_{20} + f_{02} + f_{00} = 1$. There will be three phenotypic classes—(*AABB*), (*AAbb*, *aaBB*), and (*aabb*)—with the corresponding frequencies of f_{22}, $f_{20} + f_{02}$, and f_{00}. This yields in the next generation,

$$\begin{aligned} f'_{22} &= f_{22} & f'_{00} &= f_{00} \\ f'_{20} &= f^2_{20}/(f_{20} + f_{02}) & f'_{02} &= f^2_{02}/(f_{20} + f_{02}) \\ P'[Ab/aB] &= 2f_{20}f_{02}/(f_{20} + f_{02}) \,. \end{aligned} \tag{2}$$

Since the population is in equilibrium, we have

$$f_{20}f_{02} = 0 \,. \tag{3}$$

Also, we have

$$\begin{aligned} f_{22} + f_{20} &= p_A & f_{22} + f_{02} &= p_B \\ f_{02} + f_{00} &= p_a & f_{20} + f_{00} &= p_b \end{aligned} \tag{4}$$

which do not change with time. From Equations (4) we get, if $p_A \neq p_B$,

$$f_{20} - f_{02} = p_A - p_B \neq 0 \,. \tag{5}$$

Equations (3) and (5) imply that either $f_{20} = 0$ or $f_{02} = 0$ but not both zero because $f_{20} - f_{02} \neq 0$. This shows that, at equilibrium, only three homozygous types can be present in the population. In addition, it also shows that the population is not composed of only two homozygous genotypes as is usually thought, because f_{20} and f_{02} cannot both be zero simultaneously.

Now consider Equations (4) again. If $p_A > p_B$, then Equation (3) yields that $f_{02} = 0$ because $f_{20} > f_{02}$. Similarly, if $p_A < p_B$, then $f_{20} = 0$. We can determine the frequencies of other genotypes by using Equations (4). For example, if $p_A > p_B$, then $f_{02} = 0$, $f_{22} = p_B$, $f_{00} = p_a$, and $f_{20} = p_A - p_B$. Therefore, the distribution of phenotypes or genotypes in an equilibrium population will be independent of the initial zygotic distribution but will depend upon the initial gene frequencies. Thus, if $p_A \neq p_B$, this mating system will result in three homozygous genotypes in a population in equilibrium instead of the two extreme types with the frequencies as given in Table 2.

TABLE 2

Genotypic distribution in equilibrium

Genotype	Frequency	
	$p_A < p_B$	$p_A > p_B$
AABB	$f_{22} = p_A$	$f_{22} = p_B$
AAbb	$f_{20} = 0$	$f_{20} = p_A - p_B$
aaBB	$f_{02} = p_B - p_A$	$f_{02} = 0$
aabb	$f_{00} = p_b$	$f_{00} = p_a$

It can be seen that if $p_A = p_B$ there will be only two extreme types with frequencies p_A and p_a.

THREE-LOCUS MODEL

It can be shown, similar to a two-locus model, that gene frequencies remain unchanged and that the population eventually becomes homozygous. Thus in the limit there will be four phenotypic groups ($AABBCC$), ($AABBcc$, $AAbbCC$, $aaBBCC$), ($AAbbcc$, $aaBBcc$, $aabbCC$), and ($aabbcc$) with the corresponding frequencies of f_{222}, $(f_{220} + f_{202} + f_{022})$, $(f_{200} + f_{020} + f_{002})$, and f_{000} adding up to unity. This yields

$$\begin{aligned} f_{222} + f_{220} + f_{202} + f_{200} &= p_A \\ f_{222} + f_{220} + f_{022} + f_{020} &= p_B \\ f_{222} + f_{202} + f_{022} + f_{002} &= p_C \quad . \end{aligned} \tag{6}$$

Also, group 1 individuals have progeny of their own type, group 2 individuals have an offspring array proportional to

$$(f_{220}\,ABc + f_{202}\,AbC + f_{022}\,aBC)^2$$

and so on. Since the population is in equilibrium

$$f_{220}f_{202} = 0, \quad f_{220}f_{022} = 0, \quad f_{202}f_{022} = 0 \tag{7}$$

because there is no heterozygosity in equilibrium.

Similarly,

$$f_{200}f_{020} = 0, \quad f_{200}f_{002} = 0, \quad f_{020}f_{002} = 0 \quad . \tag{8}$$

We have to determine the f's such that Equations (6)–(8) are satisfied. There are 16 possible solutions of Equations (7) and (8). If the gene frequencies are all different, i.e., $p_A \neq p_B \neq p_C$, there is a unique solution for the whole system. Without loss of generality, let us take $p_A > p_B > p_C$. From all possible solutions of Equations (7) and (8), the only solution which also satisfies relations (6) and the inequalities among the gene frequencies, is

$$\begin{aligned} f_{222} &= p_C \\ f_{220} &= p_B - p_C \\ f_{200} &= p_A - p_B \\ f_{000} &= 1 - p_A \quad . \end{aligned} \tag{9}$$

The frequencies of other genotypes are zero. Any other solution will be inadmissible. The solution of the system for other similar inequalities among the gene frequencies can be easily obtained by permuting the subscripts properly.

The other possible relationships among the gene frequencies are of the type where some of the gene frequencies are equal, e.g.; $p_A = p_B \neq p_C$. To be more specific, let us consider $p_A = p_B > p_C$. Consider first the case where $p_A = p_B + \varepsilon$, $\varepsilon > 0$, which yields $p_A = p_B + \varepsilon > p_B > p_C$. The solution in this case is given by

Equations (9) for all $\varepsilon > 0$. In the limit as $\varepsilon \to 0$, $p_A \to p_B$ and consequently we have the distribution in equilibrium

$$f_{222} = p_C, \quad f_{220} = p_A - p_C, \quad f_{000} = 1 - p_A \ . \tag{10}$$

We can similarly obtain the solution when $p_A = p_B = p_C$. This shows that the limiting distribution in all cases is determined from the general solution (9) where the inequalities ($>$) among the gene frequencies are replaced by $\geq$. These results also show that the marginal distribution, in equilibrium, for a pair of loci is the same as for a two-locus model.

m-LOCUS MODEL

Following the same approach as for one-, two-, and three-locus models, we can easily extend the analysis to multiple loci, say m. In equilibrium, there is no heterozygosity in the population and thus there will be $(m+1)$ phenotypic groups of homozygous genotypes. Since the population is in equilibrium and there is no heterozygosity, each phenotypic group can have at the most one genotype depending upon the relations among the gene frequencies. Without loss of generality, let $p_A \geq p_B \geq \ldots \geq p_L \geq p_M$. From the form of earlier solutions for $m = 1,2,3$, it can be deduced that the limiting distribution in this case is given by

$$\begin{aligned} f_{22\ldots22} &= p_M \\ f_{22\ldots20} &= p_L - p_M \\ &\ldots \qquad\qquad \ldots \\ f_{20\ldots00} &= p_A - p_B \\ f_{00\ldots00} &= 1 - p_A \ . \end{aligned} \tag{11}$$

The frequencies of the other genotypes are zero.

The author wishes to express his appreciation to PROFESSOR OSCAR KEMPTHORNE for many helpful suggestions and comments. Thanks are also due to DR. W. G. HILL for going through the manuscript and for suggesting improvements in the presentation. Thanks are also due to a referee for some useful suggestions.

LITERATURE CITED

CROW, J. F. and J. FELSENSTEIN, 1968 The effect of assortative mating on the genetic composition of a population. Eugen. Quart. **15**: 85–97.

JENNINGS, H. S., 1916 The numerical results of diverse systems of breeding. Genetics **1**: 53–89.

LI, C. C., 1955 *Population Genetics*. Univ. Chicago Press, Chicago, Illinois.

WENTWORTH, E. N. and B. L. REMICK, 1916 Some breeding properties of the generalized Mendelian population. Genetics **1**: 608–616.

WRIGHT, S., 1921 Systems of mating: III. Assortative mating based on somatic resemblance. Genetics **6**: 144–161.

Corresponding editor: T. PROUT