Hardy-Weinberg Law: Asymptotic Approach to a Generalized Form

Abstract. *The equilibrium frequencies of a generalized Hardy-Weinberg law are approached at a geometric rate under assortative mating, irrespective of the initial genotypic frequencies. The asymptotic form is similar to that of Wright, and the pattern of assortative mating is based on deviations from the mean genotypic value.*

This report presents a model of assortative mating with respect to a single locus with two alleles a and A and genotypes aa, aA, and AA. It was suggested by a model developed by Fisher (*1*) for quantitative traits, which is exposited by Malécot (*2*) and Kempthorne (*3*).

Fisher considered mating according to phenotype and obtained the frequency of matings of individuals with scores a, b to be $M'N'$ exp $(\mu ab/V)$, where M' and N' are the frequency of mating individuals with scores a and b, respectively, V is the variance of the trait, and μ is the correlation between mates. He then approximated the exponential by $(1 + \mu ab/V)$, which suggested the mating function (Eq. 3) given below.

It will be shown that whereas Fisher used approximations to obtain his result, the analogous model for a single locus is exact. First the outline of the model in equilibrium is stated and then it is proved that the equilibrium is stable in the sense that, providing the correlation between mates is held constant, the population moves toward such an equilibrium.

First assume that the parental distribution of genotypes aa, aA, and AA is given by the respective frequencies

$$\begin{aligned} f_0 &= q^2 + \lambda pq \\ f_1 &= 2(1 - \lambda)pq \\ f_2 &= p^2 + \lambda pq \end{aligned} \qquad (1)$$

This distribution is a parametrization which can always be made when there are two alleles with frequencies q and p. The constant λ in Eq. 1 is referred to as the constant of homogamy after Malécot (*4*) and plays the same role as Wright's (*5*) fixation index F, except that λ is used here only in respect of a population in equilibrium.

If the genotypes are given scores 0, 1, and 2, respectively, that is, equal to the number of A genes carried, the mean and variance of distribution in Eq. 1 are $m = 2p$ and $V = 2(1 + \lambda)pq$, respectively, and the deviations of scores from the mean are

$$\begin{aligned} d_0 &= 0 - 2p = -2p \\ d_1 &= 1 - 2p = q - p \\ d_2 &= 2 - 2p = 2q \end{aligned} \qquad (2)$$

respectively.

It can be proved, by calculating the frequencies of genotypes under the usual Mendelian assumptions, that the parental distribution (Eq. 1) is duplicated by that of the offspring if the frequencies of mating pairs, f_{xy}, are given by

$$f_{xy} = f_x f_y(1 + \mu d_x d_y/V) \qquad (x = 0, 1, 2; y = 0, 1, 2) \qquad (3)$$

provided that μ is related to λ by the function

$$\mu = 2\lambda/(1 + \lambda) \qquad (4)$$

In order for the frequencies in Eq. 3 to be nonnegative, μ must be restricted to the ranges

$$[1 - 1/p^{1/2}, 1], (q \leqq \tfrac{1}{2})$$

and

$$[1 - 1/q^{1/2}, 1], (q > \tfrac{1}{2})$$

If the genes a and A are given scores 0 and 1, respectively, consistent with additive contributions to genotypes, the correlation between uniting gametes in Eq. 1 is equal to λ, and also the correlation between mates in Eq. 3 is equal to μ. The relation in Eq. 4 is thus a basic relation between correlation coefficients, which is better known when rewritten as

$$\lambda = \mu/(2 - \mu) \qquad (5)$$

a function which arises, for example, in connection with mixed self-fertilization and random mating, a point which receives comment at the end of this report.

Suppose now that the population has frequencies

$$\begin{aligned} aa&: f_0 = q^2 + \theta_0 pq \\ aA&: f_1 = 2pq - 2\theta_0 pq \\ AA&: f_2 = p^2 + \theta_0 pq \end{aligned} \qquad (6)$$

The gene frequencies, $q = f_0 + \frac{1}{2} f_1$ and $p = 1 - q = \frac{1}{2} f_1 + f_2$, respectively, remain constant under assortative mating of the type given below, which is of the same form as that given in Eq. 3. The genotypic scores are kept the same as before, so that the deviations from the mean are as given in Eq. 2, but the variance of Eq. 6, V, is given by

$$V = 4(p^2 + \theta_0 pq) + 2(1 - \theta_0)pq - 4p^2 = 2(1 + \theta_0)pq \qquad (7)$$

Let the frequency of matings, f_{xy}, be

$$f_{xy} = f_x f_y(1 + k\, d_x d_y/V) \qquad (8)$$

Then the frequency of genotype aa among offspring is

$$\begin{aligned} &f_0^2(1 + k\, d_0^2/V) + f_0 f_1(1 + k\, d_0 d_1/V) + \tfrac{1}{4} f_1^2(1 + k\, d_1^2/V) \\ &= (f_0 + \tfrac{1}{2} f_1)^2 + (f_0 d_0 + \tfrac{1}{2} f_1 d_1)^2\, k/V \\ &= q^2 + p^2 q^2 (1 + \theta_0)^2\, k/V \\ &= q^2 + \tfrac{1}{2} k(1 + \theta_0)pq \end{aligned} \qquad (9)$$

since $f_0 d_0 + \frac{1}{2} f_1 d_1 = -pq(1 + \theta_0)$.

Note that if $\frac{1}{2} k(1 + \theta_0) = \theta_0$, the frequency of aa is unchanged.

The correlation between mates in Eq. 8 is

$$\Sigma\Sigma\, f_x f_y(1 + k\, d_x d_y/V) d_x d_y/V = k$$

a result which is immediate because $\{d_x/V^{\frac{1}{2}}\}$ is a vector of standardized scores with respect to Eq. 6.

If it is assumed that the correlation of mates with regard to score is constant over generations and equal to μ, then, by virtue of Eq. 9, the population in successive generations has the genotypic arrays

$$(q^2 + \theta_t pq)aa + (2pq - 2\theta_t pq)aA + (p^2 + \theta_t pq)AA$$

with the coefficients θ_t obeying the simple linear recurrence equation

$$\theta_{t+1} = \tfrac{1}{2}\mu(1 + \theta_t) \qquad (10)$$

Equation 10 has the general solution

$$\theta_t = \mu/(2 - \mu) + (\mu/2)^t \{\theta_0 - \mu/(2 - \mu)\}$$

and in the limit θ becomes $\mu/(2 - \mu)$, which is the function of μ given in Eq. 5. Hence, with the mating system above, the population tends to the form

$$(q^2 + \lambda pq)aa + (2pq - 2\lambda pq)aA + (p^2 + \lambda pq)AA$$

where λ is linked to μ by Eq. 5. The solution of recurrence equation (Eq. 10) shows that the rate of approach to equilibrium is rapid, since it involves the factor $\mu/2$, whose absolute value is less than or equal to 0.5.

There is some similarity between this

Table 1. Heterozygous mating frequencies.

Mating		Assortative mating	Partial self-fertilization
Female	Male		
aA	aa	$(1 - \lambda^2)/8$	$(1 - \lambda)^2/8$
aA	aA	$(1 - \lambda)^2/4$	$(1 - \lambda^2)/4$
aA	AA	$(1 - \lambda^2)/8$	$(1 - \lambda)^2/8$

model and partial self-fertilization, as described in Crow and Kimura (*6*, pp. 92–94, 144, 152). There is a geometric approach to equilibrium, and at equilibrium, F, the fixation index of Wright, is related to h, the proportion of self-fertilization, by $F = h/(2 - h)$. In partial self-fertilization a fraction h of mates have unit correlation and the remainder zero, so that overall the correlation coefficient is h. In the model given here the correlation between mates is μ, and hence the relation between h and F is the same as that between μ and λ. Wright's F corresponds to λ, and h to μ.

However, there are two differences between the models. First, in Wright's model of self-fertilization F is essentially positive, whereas in the model given here λ can be negative. Second, and more important, the equilibrium mating frequencies are different in the two models. For example, if $q = \frac{1}{2}$, the genotype frequencies are $f_0 = f_2 = \frac{1}{4}(1 + \lambda)$, $f_1 = \frac{1}{2}(1 - \lambda)$, and the mating frequencies involving heterozygotes are as shown in Table 1.

The list of papers which have been written on assortative mating in relation to a single locus is too extensive to be given here, but a number of the more important ones are cited by Crow and Kimura (*6*, pp. 143, 161). Lewontin *et al.* (*7*) draw distinctions between assortative mating, inbreeding, and selective mating, which are relevant to the previous two paragraphs of this report. The equilibrium system given by Eq. 3 was introduced by Stark [see (*8*), where it is discussed in greater detail].

A. E. STARK

School of Community Medicine,
University of New South Wales,
Post Office Box 1, Kensington,
New South Wales, Australia 2033

References and Notes

1. R. A. Fisher, *Trans. R. Soc. Edinburgh* **52**, 399 (1918).
2. G. Malécot, thesis, University of Paris (1939).
3. O. Kempthorne, *An Introduction to Genetic Statistics* (Wiley, New York, 1957), pp. 493–498.
4. G. Malécot, *Colloq. Int. C.N.R.S.* **13**, 121 (1949).
5. S. Wright, *Evolution and the Genetics of Populations* (Univ. of Chicago Press, Chicago, 1969), vol. 2, p. 173.
6. J. F. Crow and M. Kimura, *An Introduction to Population Genetics Theory* (Harper & Row, New York, 1970).
7. R. Lewontin, D. Kirk, J. Crow, *Eugen. Q.* **15**, 141 (1968).
8. A. E. Stark, *Nature (London)* **259**, 44 (1976); *Ann. Hum. Genet.* **39**, 455 (1976).
9. I thank Prof. G. Malécot, head of the department (Mathématiques Appliquées, Université Claude Bernard, Lyon 1) where this work was completed, for his advice and hospitality, and two referees: the first pointed out the similarity of my model to that of partial self-fertilization, the second provided comments on the work of Fisher and a proof of the approach to equilibrium which I used to make a substantial improvement to my original manuscript.

9 December 1975; revised 24 May 1976

The Effect of Stimulus Sequence on the Waveform of the Cortical Event-Related Potential

Abstract. *The waveform of the cortical event-related potential is extremely sensitive to variations in the sequence of stimuli preceding the eliciting event. The waveform changes were manifested primarily in the amplitudes of the negative component of the potential that peaked at 200 milliseconds, the positive component that peaked at 300 milliseconds, and the slow-wave components. A quantitative model was developed relating the waveform changes to changes in event expectancy. Expectancy is assumed to depend on a decaying memory for events within the prior sequence, the specific structure of the sequence, and the global probability of event occurrence. For stimuli relevant to the task, the less expected the stimulus the larger the amplitudes of late components of the event-related potentials.*

The cortical event-related potential (ERP) associated with the rare outcomes of Bernoulli trials that are relevant to a task is different from that associated with the frequent outcomes (*1–3*). If, for example, a subject is instructed to count the occurrences of a few low-pitched tones embedded in a series of high-pitched tones, the low-pitched tones elicit ERP's characterized by a large negative component peaking at 200 msec (N200), a large positive component peaking at 300 msec (P300), and large slow-wave (SW) components (*3*). The distinction between the types of ERP waveforms made it possible for us to develop a discriminant function in which ERP's to individual tones were classified as "rare" or "frequent" (*4*). When a "correct" classification was defined as the categorization of an ERP elicited by a rare tone as "rare" and by a frequent tone as "frequent," the discriminant function formed from the ERP data of one group of subjects correctly classified 81 percent of all ERP's from a group of new subjects. Although the classification technique was successful, the reasons for misclassifying 19 percent of the trials remained unclear. An analysis of the waveforms associated with the misclassifications suggested that the erroneous classifications reflected systematic trial-to-trial variations in the ERP waveforms. Some of the "rare" events seemed to elicit a "frequent" waveform, and vice versa (*5*). Since the underlying assumption of research involving ERP's is that the ERP's elicited by all occurrences of a particular type of event are identical, such trial-to-trial variations in the ERP required further examination.

Inspection of the trial-to-trial waveform measures suggested that the variations might have been due to short-term sequential dependencies. Remington (*6*) and others (*7–10*) have demonstrated sequential dependencies in choice tasks, in which reaction time (RT) on any given trial is sensitive to the specific sequence of preceding events. These sequential effects have been attributed by some to trial-to-trial changes in the subject's expectancies (*9, 10*). Since expectancy has been implicated as one determinant of the ERP waveform, we attempted to determine whether the waveform of the ERP exhibits similar sequential dependencies.

Seven subjects listened to series of regularly presented tone bursts. On each trial a high-pitched and a low-pitched tone were equally likely to occur (*11*). The subject was instructed to count the high-pitched tones silently and to report the count after each block of 200 trials. In a second condition, the probabilities of the high- and low-pitched tones were changed to .3 and .7, respectively (*12*). Each subject was tested on 800 to 1600 trials, depending on the condition.

During testing, the subject was comfortably seated in a reclining chair in a well-lighted experimental room. The electroencephalogram (EEG) was recorded with Burden Ag-AgCl electrodes from F_z, C_z, and P_z (according to the 10-20 system) which were referred to a linked mastoid electrode, with a wrist ground. The band pass of the amplifier system was set for a time constant of 0.8 second and an upper half-amplitude frequency of 35 hertz. Additional electrodes (Beckman) were situated above and on the outer margin of the right eye to record eye movement and blink potentials. On each trial, a 768-msec epoch of the EEG, beginning 100 msec before the stimulus onset, was digitized from each of the recording channels (at a rate of one sample every 3 msec) and stored on digital magnetic tape. The EEG epochs contaminated with eye movement or blink artifacts were excluded from the subsequent waveform analysis. All trials entered into the tabulations of sequences.

Remington's terminology (*6*) will be followed. An "A" represents whichever stimulus event occurred on trial *N* (a first-order sequence). For the second-order sequences there were two possible