THEORETICAL POPULATION BIOLOGY 20, 394–410 (1981)

Is the Between-Population Variance Negligible in the Total Variance of Heterozygosity? Case of a Finite Number of Loci Subject to the Infinite-Allele Model in Finite Monoecious Populations*

S. LESSARD†

Département de Génétique de Populations, Institut National d'Études Démographiques, Paris, France; and Département de Mathématique Appliquées, Université de Clermont, Aubiére, France

Received December 5, 1981

The decomposition of the variance of the average heterozygosity into variances between and within populations is studied in the general case of a finite number of loci. These loci are assumed randomly distributed over chromosome pairs having a non-interference recombination scheme, and independently subject to mutation according to the infinite-allele model. The equilibrium behavior of that decomposition is discussed in the monoecious mating case with regard to each parameter of the model: mutation rate per gene per generation (u), population size (N), number of loci (n), map length of chomosome pairs (L). It is shown that the proportion Q of the between-population variability in the total variance of the average heterozygosity is decreasing as either the mean heterozygosity ($\theta = 4Nu/(1 + 4Nu)$) or the mean number of mutations per gamete per generation ($v = nu$) is increasing. Moreover, even if Q is always smaller than $\frac{1}{3}$ for this model, it is not negligible unless θ is close to one or v is much larger than one for L long enough.

1. INTRODUCTION

The expected value of heterozygosity at a single locus maintained by selectively neutral mutations in a finite population has been known for a long time (Malécot, 1946; Kimura and Crow, 1964) while its variances between and within populations have been studied more recently (Nei and Roychoudhury, 1974; Li and Nei, 1975; Stewart, 1976; Géry 1978). On the other hand, assuming an infinite number of loci, Sved (1971) was concerned with the distribution of the length of homozygous chromosome segments, Franklin (1977) with the distribution of the proportion of genome which is

* Research supported in part by La Direction Générale de l'Enseignement Supérieur, Gouvernement du Québec.

† Present address: Department of Mathematics, Stanford University, Stanford, Calif. 94305.

0040-5809/81/060394-17$02.00/0

homozygous by descent in inbred individuals, and Weir *et al.* (1980) with the variation in inbreeding under several mating structures. Considering (as it is done in this article) a diploid monoecious population with random mating including random selfing, Avery and Hill (1979) showed that the between-population variance for the coefficient of inbreeding may be neglected asymptotically in the total variance. Incidentally, Jacquard (1975) had used this fact without being aware of it for the case dioecious with monogamy, and his method was later justified numerically and developed by Weir *et al.* (1980).

In this paper, we shall consider the average heterozygosity with respect to a finite number of loci. A general formulation of a decomposition of the variance of any quantitative genetical trait (Wright, 1951, 1952; Cockerham, 1969, 1973) and particular formulas for the average heterozygosity (Schnell, 1961; Sved, 1968; Weir *et al.*, 1980) will be recalled. Probabilistic arguments will be given to establish the decomposition of Wright and Cockerham, who had a more statistical point of view (Section 2 and Appendix A). Introducing mutation according to the infinite-allele model and considering a finite number of loci, the formulas of Weir, Avery, and Hill for the between-population variance and the total variance of the average heterozygosity will be deduced (Sections 3 and 4). Some probabilities will have to be determined. Methods of identity by descent of Malécot (1948) generalized by descent measures of Weir and Cockerham (1969, 1974), and Cockerham (1971) will be used introducing mutation as described by Serant (1974) (Appendix B). Then the proportion Q of the between-population variability in the total variance of the average heterozygosity will be studied at equilibrium. Analytical solutions will be given in the cases of loci randomly distributed over chromosome pairs of very short or very long map length, while numerical evaluations will be provided in the intermediate cases assuming a non-interference recombination scheme (Section 5). In the last section (Section 6), the behavior of Q will be discussed with regard to each parameter involved: mutation rate per gene per generation (u), population size (N), number of loci (n), map length of chromosome pairs (L).

2. Variances between and within Populations

Consider in a diploid population a quantitative genetical trait X determined at n loci. Chromosome (or gamete) types are described by the n-tuples

$$\xi = (\xi^{(1)},..., \xi^{(n)}), \tag{2.1}$$

where $\xi^{(k)}$ is one of the possible alleles at locus k ($k = 1,..., n$). A typical

genotype $\mathscr{G}$ composed of two chromosomes ξ and ζ as (2.1) is displayed in the form

$$\mathscr{G} = \xi/\zeta = (\xi^{(1)},..., \xi^{(n)})/(\zeta^{(1)},..., \zeta^{(n)}). \tag{2.2}$$

The genetical trait X is a variable which associates a quantity with each possible genotype. For the purposes of this paper, suppose this quantity is bounded by a constant. The genetical states of the entire population are given by

$$s = (\mathscr{G}_1,..., \mathscr{G}_N), \tag{2.3}$$

where $\mathscr{G}_i$ is the genotype of the individual i $(i = 1,..., N)$ and N is the population size $(N \geqslant 2)$. Let S be the space of all possible such states.

If we assume that the generations are discrete and that the individuals of each generation are obtained from the previous one as results of independent identical random trials (of mating, recombination, mutation, etc.), a transition probability on S can be well defined and then a Markov chain on S can be given for any initial distribution (see Karlin and Taylor, 1966; or any text on Markov chains). Let us denote by $h^{(g)}$ a typical genetical history of the entire population (i.e., a sequence of successive genetical states) up to the generation g inclusively, by $H^{(g)}$ the set of all possible $h^{(g)}$ and by $\mathscr{H}^{(g)}$ the σ-algebra of all parts of $H^{(g)}$.

Let $X_i^{(g)}$ be the trait X for the individual i in the generation g. The sequence of random variables $(X_1^{(g)},..., X_N^{(g)})$ is clearly exchangeable for any N. Then by De Finetti's theorem, it is also conditionally independent (Appendix A). Here it is evident by the preceding assumptions that it is conditionally independent with respect to the past. In other words, the sequence $(X_1^{(g)},..., X_N^{(g)})$ conditionally on $\mathscr{H}^{(g-1)}$ is independent identically distributed (cond. i.i.d.).

Let $E(X_1^{(g)} \mid \mathscr{H}^{(g-1)})$ be the conditional expectation of $X_1^{(g)}$ with respect to the σ-algebra $\mathscr{H}^{(g-1)}$. This is the random variable that, for each possible past history $h^{(g-1)}$, takes as a value the mathematical expectation of $X_1^{(g)}$ given $h^{(g-1)}$. By the law of large numbers for cond. i.i.d. sequences of random variables (see Appendix A), we get a helpful representation for $E(X_1^{(g)} \mid \mathscr{H}^{(g-1)})$, namely,

$$E(X_1^{(g)} \mid \mathscr{H}^{(g-1)}) = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} X_i^{(g)}. \tag{2.4}$$

One may consider this result as the standard law of large numbers applied for each possible past history $h^{(g-1)}$.

Now let us write the decomposition

$$X_1^{(g)} = [X_1^{(g)} - E(X_1^{(g)} \mid \mathscr{H}^{(g-1)})] + [E(X_1^{(g)} \mid \mathscr{H}^{(g-1)})]. \tag{2.5}$$

The basic properties of conditional expectation (see Loève, 1960; or any other text in probability theory) lead to a corresponding decomposition for the variances, namely,

$$V^{(g)} = V_w^{(g)} + V_b^{(g)}, \tag{2.6a}$$

where $V^{(g)}$, $V_w^{(g)}$ and $V_b^{(g)}$ are the respective variances for the variables in (2.5). This is the decomposition of Wright (1951, 1952). Moreover, by (A.3) in the Appendix, representation (2.4) enables us to go forth and find out the further relation

$$V_b^{(g)} = \mathrm{Cov}(X_1^{(g)}, X_2^{(g)}) \tag{2.6b}$$

(Cov for covariance). This property was pointed out by Cockerham (1969, 1973).

The interpretation of the variances involved in decomposition (2.6) is of great interest. Consider the island model of Wright (1951). More precisely, an infinite number of isolated populations, called "species" by agreement, originated initially from the same population and then evolved independently according to a model similar to that described above. Just before producing the generation g, it is expected that the histories $h^{(g-1)}$ will have been realized according to their chances of occurrence. Then let the next generation be infinite in each population. $V_w^{(g)}$ and $V_b^{(g)}$ should be the respective variances of the trait X within and between these populations. Here variance between populations means variance of the average values taken over the individuals in each population. So let us call $V_w^{(g)}$ the within-population variance and $V_b^{(g)}$ the between-population variance, both at generation g.

3. Heterozygosity

With each genotype $\mathscr{G}$ of the form (2.2), we can associate

$$\eta = (\eta^{(1)}, \ldots, \eta^{(n)}), \tag{3.1}$$

where $\eta^{(k)} = 1$ or 0 accordingly as $\mathscr{G}$ is heterozygous or homozygous at locus k, i.e., $\xi^{(k)}$ and $\zeta^{(k)}$ in representation (2.2) stand for different or identical alleles ($k = 1, \ldots, n$). Throughout the remainder of this article, we concentrate on the average heterozygosity; namely, we define

$$X = \frac{1}{n} \sum_{k=1}^{n} \eta^{(k)}. \tag{3.2}$$

In this particular case, the variance $V^{(g)}$ and the between-population variance $V_b^{(g)}$ in the decomposition (2.6) take the forms

$$V^{(g)} = \frac{1}{n^2}\left[\sum_{k=1}^{n}(\theta_k^{(g)} - \theta_k^{(g)2}) + 2\sum_{k=1}^{n-1}\sum_{l=k+1}^{n}(\Theta_{k,l}^{(g)} - \theta_k^{(g)}\theta_l^{(g)})\right], \quad (3.3a)$$

$$V_b^{(g)} = \frac{1}{n^2}\left[\sum_{k=1}^{n}(\delta_k^{(g)} - \theta_k^{(g)2}) + 2\sum_{k=1}^{n-1}\sum_{l=k+1}^{n}(\Delta_{k,l}^{(g)} - \theta_k^{(g)}\theta_l^{(g)})\right], \quad (3.3b)$$

where

$$\theta_k^{(g)} = E(\eta_A^{(k)}), \quad (3.4a)$$

$$\delta_k^{(g)} = E(\eta_A^{(k)}\eta_B^{(k)}), \quad (3.4b)$$

$$\Theta_{k,l}^{(g)} = E(\eta_A^{(k)}\eta_A^{(l)}), \quad (3.4c)$$

$$\Delta_{k,l}^{(g)} = E(\eta_A^{(k)}\eta_B^{(l)}) \quad (3.4d)$$

(E for expectation), with $k, l = 1,..., n$ $(k \neq l)$ and A, B two distinct individuals in generation g. Of course, the double summation in (3.3) disappears when $n = 1$.

4. Model

For the purpose of this paper, let the population be finite with constant size N and the individuals be diploid monoecious. Assume random mating including random selfing. In other words, at each generation, N diploid individuals independently provide an infinitely large identical number of gametes, and then die. The next generation is obtained by randomly pairing $2N$ gametes of the previous pool. Overproduced gametes die. Besides, assume that mutations independently occur with probability u per gene per generation and each one leads to a new allele at its locus in its generation of appearance. This is known as the infinite-allele model (Kimura and Crow, 1964).

Under the previous assumptions, it is proved in Appendix B that a unique stable equilibrium exists for measures (3.4). Moreover, the equilibrium values are reached independently of the initial conditions, are independent of the specific locus when only one locus is involved, and are dependent only on the recombination rate when two loci are involved. So let us denote the corresponding equilibrium values for those measures by θ, δ, $\Theta(r_{k,l})$, $\Delta(r_{k,l})$, where $r_{k,l}$ is the recombination rate between loci k and l. The expressions for these values are given in the Appendix. In the same way, let V, V_w, and V_b be the corresponding equilibrium variances for the variances in (2.6).

Now suppose that the n loci are randomly distributed over chromosome pairs of map length L (L is the expected number of crossovers). Let R be the recombination rate between any two of these loci. Then (3.3) leads to the following formulas at equilibrium:

$$\mathrm{Exp}(V) = \frac{\theta - \theta^2}{n} + \left(1 - \frac{1}{n}\right) \mathrm{Exp}[\Theta(R) - \theta^2], \tag{4.1a}$$

$$\mathrm{Exp}(V_b) = \frac{\delta - \theta^2}{n} + \left(1 - \frac{1}{n}\right) \mathrm{Exp}[\Delta(R) - \theta^2], \tag{4.1b}$$

where the notation Exp is used for the expectation taken over the distribution of the loci. Since the mean heterozygosity θ is independent of this distribution, it is clear by the properties of conditional expectation that $\mathrm{Exp}(V)$ and $\mathrm{Exp}(V_b)$ are actually the variances over the total probability space. Finally, decomposition (2.5) takes the form

$$\mathcal{V} = \mathcal{V}_w + \mathcal{V}_b, \tag{4.2}$$

at equilibrium, where $\mathcal{V}_w$ and $\mathcal{V}_b$ are respectively the total variances within and between populations of the total variance $\mathcal{V}$ of X. The expressions for $\mathcal{V}$ and $\mathcal{V}_b$ are given by (4.1a) and (4.1b), respectively.

At this point, the question of the distribution of R arises and a final assumption is included, namely, a non-interference recombination scheme (Haldane (1919)) or equivalently a Poisson-count crossover process (see, e.g., Karlin and Liberman, 1979). Then R is distributed as $(1 - e^{-2D})/2$, where D is the distance between two points chosen at random in $[0, L]$ and formulas (4.1) can be rewritten in a more explicit form, namely,

$$\mathcal{V} = \frac{\theta - \theta^2}{n} + \left(1 - \frac{1}{n}\right) \frac{2}{L^2} \int_0^L (L - y) \left[\Theta\left(\frac{1 - e^{-2y}}{2}\right) - \theta^2\right] dy, \tag{4.3a}$$

$$\mathcal{V}_b = \frac{\delta - \theta^2}{n} + \left(1 - \frac{1}{n}\right) \frac{2}{L^2} \int_0^L (L - y) \left[\Delta\left(\frac{1 - e^{-2y}}{2}\right) - \theta^2\right] dy. \tag{4.3b}$$

Weir *et al.* (1980), studying the variation in inbreeding ($u = 0$, $n = \infty$), pointed out these formulas for any generation g.

5. Results

Throughout the remainder of this paper, we will be exclusively concerned with the quotient

$$Q = \mathscr{V}_b/\mathscr{V} \tag{5.1}$$

and its equilibrium behavior according to each paramater involved. With this definition, Q is a measure of the relative weight of the between-population variability in the total variance of the average heterozygosity at equilibrium. Analytical determinations will be given in two limit cases while numerical evaluations will be provided for the others. It is understood throughout this section that the population size N is very large. Moreover, the results below neglect terms of small order with regard to the principal terms.

In the case $L = 0$ (case of completely linked loci), $R = 0$ and then (4.1) above and (B.6–B.9) in the Appendix lead to the result

$$Q = \frac{2(1-\theta)^2\,[3 + n\theta(2-\theta)]}{(2-\theta)(3-2\theta)(3-\theta)(1+n\theta)}, \tag{5.2}$$

where $\theta = 4Nu/(1 + 4Nu)$ is the mean heterozygosity.

On the other hand, in the case $L = \infty$ (case of completely unlinked loci), $R = 1$ and again (4.1) with (B.6–B.9) gives the second analytical result, namely,

$$Q = \frac{6(1-\theta)^2}{(2-\theta)(3-2\theta)[3 + 4(1-1/n)\,v]} + \frac{6(1-\theta)(1-1/n)\,v}{[3 + 4(1-1/n)\,v]\,N}, \tag{5.3}$$

where $v = nu$ is the mean number of mutations per gamete per generation.

In addition, using formulas (4.3) above and (B.5) in the Appendix, numerical evaluations of Q have been computed with some relevant values for the parameters n, N, u, L. These results are summarized in Figs. 1 and 2. Of course, these agree with the previous theoretical determinations when L is short enough or long enough.

Remarks. (1) In the case of one locus, Q is independent of L and given by

$$Q = \frac{V_b}{V} = \frac{2(1-\theta)^2}{(2-\theta)(3-2\theta)}. \tag{5.4}$$

This is in agreement with results already known (Stewart, 1976; Nei and Roychoudhury, 1974).

(2) In all cases, Q is bounded above by $\frac{1}{3}$.

(3) Q decreases as n increases. The decrease rate and the limit value depend on L. If L is very short, the limit

$$Q = \frac{2(1-\theta)^2}{(3-\theta)(3-2\theta)} \tag{5.5}$$

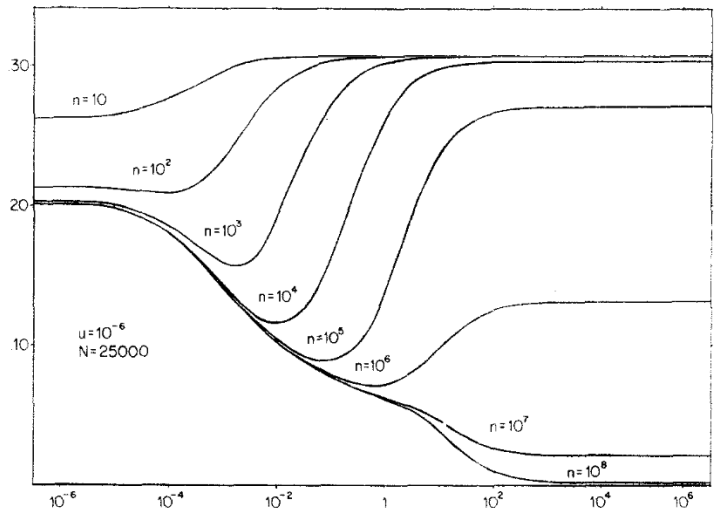

FIG. 1. Q as a function of L for different numbers of loci and a mean heterozygosity of 0.09.

is attained as soon as $n\theta$ is much larger than one, while if L is very long, the limit

$$Q = \frac{3(1-\theta)}{2N} \tag{5.6}$$

is reached only when nu is much larger than N. However, notice that in the latter case, Q is actually negligible as soon as nu is much larger than one.

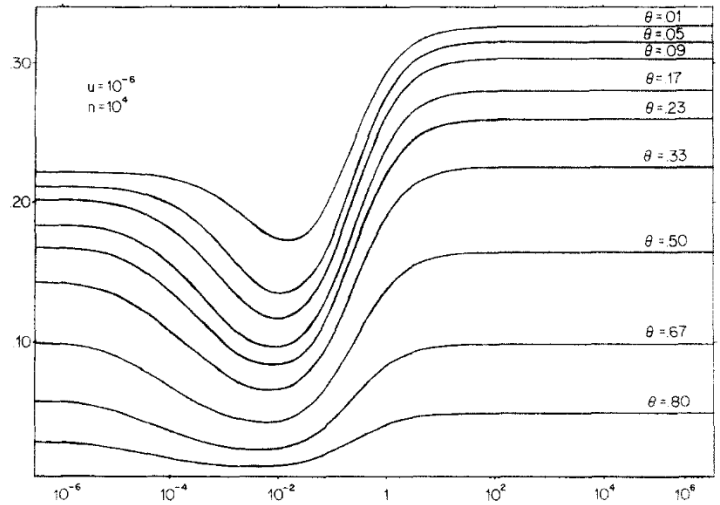

FIG. 2. Q as a function of L for different mean heterozygosities and a mean number of mutations per gamete per generation of 0.01.

(4) Q decreases as θ increases. The limit function is negligible as soon as $(1-\theta)$ is negligible with regard to one or equivalently Nu is much larger than one.

(5) In the case $L=\infty$, the first term of (5.3) gives a good approximation of Q which may be written in the form

$$Q=\frac{3}{(1+2Nu)(3+4Nu)(3-4u+4nu)}. \tag{5.7}$$

6. Discussion

It should be emphasized that (5.2) cannot be deduced from (5.4) by replacing u by nu. While in (5.4) the average heterozygosity can take only the values 0 or 1, in (5.2) it can take the values $0, 1/n,..., n/n$. Also (5.3) cannot be deduced from (5.4) by assuming independence between loci. Remind ourselves that loci are unlinked in individuals, not in populations. Our results reflect a dependence in all cases.

What about our basic assumptions? The island model without migration may describe a situation for different species. Moreover, as Zouros (1979) stated, the infinite-allele model seems acceptable for natural populations. However, the mutation rate u is not constant over all loci and a gamma distribution has been postulated (Nei, Fuerst, and Chakraborty 1976; Zouros 1979). Our assumption on u (as the others) is more convenient mathematically. Notice that if we let $u=0$, $n=\infty$, and $L=\infty$, then $(1-X)$ becomes the inbreeding coefficient and the variation of this trait was studied for several mating structures (Avery and Hill (1980)). Even though our assumptions may not always be appropriate, our conclusions might help in formulating some general principles.

It may be relevant to point out that formulas (3.3) with formulas (B.4) enables us to obtain the variances $\mathscr{V}$, $\mathscr{V}_w$, and $\mathscr{V}_b$ at any generation g, under any distribution of loci and any recombination scheme. However, computational restrictions have to be taken into account especially when the number of loci is large. The assumptions of loci randomly distributed and recombination scheme without interference were also made by Weir, Avery, and Hill (1980) (see also Avery and Hill (1977, 1979)). Different recombination schemes (a list is given in Karlin and Liberman (1979)) could be considered and the linkage effects on $\mathscr{V}$, $\mathscr{V}_w$, and $\mathscr{V}_b$ studied in each case. This will be the object of a subsequent study.

In the case of one locus, the between-population variance of heterozygosity at any generation g is given by

$$V_b^{(g)}=\delta_1^{(g)}-\theta_1^{(g)2}.$$

By methods based on gene frequencies, Gery (1978) obtained the exact expression for $V_b^{(g)}$, while Li and Nei (1975) found an approximative formula by differentiation. Computations using our exact determinations in (B.4) have shown that their approximations were quite accurate even for $g = 10$.

On the other hand, it may be of interest to compare our results with those of Wright (1951, 1952). So for a trait governed by additive genes with recurrent mutations, the proportion of the between-population variance in the total variance is given at steady state by

$$Q = \frac{2(1-\theta)}{2-\theta},$$

where θ is the mean heterozygosity. Notice that $Q > \frac{1}{2}$ for $\theta < \frac{2}{3}$.

For the average heterozygosity, we have shown that $Q < \frac{1}{3}$ in all cases, which implies that the between-population variability is less important than the within-population variability. Of course, this conclusion may be strongly dependent on our model. However, some data recently collected by Nevo (1978) for natural populations led to a conclusion in the same way, namely, that "levels of genetic variation may differ more within than between taxa (i.e., taxonomic groups of species)."

Besides it has been shown that Q decreases as either the mean heterozygosity ($\theta = 4Nu/(1+4Nu)$) or the mean number of mutations per gamete per generation ($v = nu$) increases (see Figs. 1 and 2). There is a symmetry between these two conditions, for an increase of θ is equivalent to an increase of Nu. If $1/N$ may be taken as a drift measure within populations, let us consider $1/n$ as a drift measure within individuals. As long as the average heterozygosity is concerned, it seems that any drift acts to emphasize differences between populations while mutation acts in the opposite direction. Here drift and mutation clearly appear as two opposite forces. Moreover, the between-population differences can actually be neglected if either θ is close to one (or equivalently Nu is much larger than one) or $v = nu$ is much larger than one (except in the cases of short L). While Q decreases rather uniformly with respect to L to a negligible value as N increases, the effect of an increase of n on Q is strongly dependent on L. If L is very short, the limit value is not a priori negligible and if L is very long, an increase of n is still less efficient than an increase of N (see (5.7)). Indeed, the number n of loci, even large, has to be taken into account.

APPENDIX A: EXCHANGEABILITY

A finite sequence $(X_1,..., X_N)$ of random variables is said to be exchangeable if the joint distribution of $(X_{\pi 1},..., X_{\pi N})$ is the same as that of

$(X_1,...,X_N)$ for every permutation π. Moreover, an infinite sequence $(X_1, X_2,...)$ of random variables is said to be exchangeable if $(X_1,...,X_N)$ is exchangeable for each $N \geqslant 2$.

A finite sequence $(X_1,...,X_N)$ of random variables is said to be conditionally independent identically distributed (cond. i.i.d.) if there is a σ-algebra $\mathscr{H}$ such that

$$P(X_1 < x_1,..., X_N < x_N|\mathscr{H}) = F(x_1) \cdots F(x_N), \tag{A.1}$$

where $F(x) = P(X_1 < x|\mathscr{H})$. An infinite sequence $(X_1, X_2,...)$ of random variables is said to be cond. i.i.d. if $(X_1,...,X_N)$ is cond. i.i.d. for each $N \geqslant 2$.

De Finetti's theorem states that an infinite sequence $(X_1, X_2,...)$ of random variables is exchangeable if and only if cond. i.i.d. Moreover, if $E|X_1| < \infty$, then

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} X_i = E(X_1|\mathscr{H}) \qquad \text{a.s.,} \tag{A.2}$$

where $\mathscr{H}$ is the σ-algebra on which the X_i are cond. i.i.d. This is the strong law of large numbers for exchangeable sequences (Loève, 1969; Kingman, 1978). Then a further result can be obtained if $E(X_1^2) < \infty$, namely,

$$\operatorname{Var}[E(X_1|\mathscr{H})] = \lim_{N\to\infty} \frac{1}{N^2} \left[\sum_{i=1}^{N} \operatorname{Var}(X_i) + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \operatorname{Cov}(X_i, X_j) \right] \tag{A.3a}$$

$$= \lim_{N\to\infty} \left[\frac{\operatorname{Var}(X_1)}{N} + \left(1 - \frac{1}{N}\right) \operatorname{Cov}(X_1, X_2) \right] \tag{A.3b}$$

$$= \operatorname{Cov}(X_1, X_2). \tag{A.3c}$$

APPENDIX B: Recurrence System

(a) *Recurrence Equations*

Let $(\xi_A^{(1)},..., \xi_A^{(n)})/(\zeta_A^{(1)},..., \zeta_A^{(n)})$ and $(\xi_B^{(1)},..., \xi_B^{(n)})/(\zeta_B^{(1)},..., \zeta_B^{(n)})$ be the genotypes of two distinct individuals A and B in generation g by agreement. Let us define

$$\eta_A^{(k)} = 1 \quad \text{if} \quad \xi_A^{(k)} \not\equiv \zeta_A^{(k)} \tag{B.1a}$$
$$= 0 \quad \text{otherwise,}$$

$$\eta_B^{(k)} = 1 \quad \text{if} \quad \xi_B^{(k)} \not\equiv \zeta_B^{(k)} \tag{B.1b}$$
$$= 0 \quad \text{otherwise,}$$

$$\eta_{AB}^{(k)} = 1 \quad \text{if} \quad \xi_A^{(k)} \not\equiv \xi_B^{(k)} \tag{B.1c}$$
$$= 1 \quad \text{otherwise,}$$

where $\not\equiv$ is a notation for non-identity between alleles. Then consider the measures

$$\theta_k^{(g)} = E(\eta_A^{(k)}), \tag{B.2a}$$

$$\gamma_k^{(g)} = E(\eta_A^{(k)}\eta_{AB}^{(k)}), \tag{B.2b}$$

$$\delta_k^{(g)} = E(\eta_A^{(k)}\eta_B^{(k)}), \tag{B.2c}$$

$$\Theta_{k,l}^{(g)} = E(\eta_A^{(k)}\eta_A^{(l)}), \tag{B.2d}$$

$$\Gamma_{k,l}^{(g)} = E(\eta_A^{(k)}\eta_{AB}^{(l)}), \tag{B.2e}$$

$$\Delta_{k,l}^{(g)} = E(\eta_A^{(k)}\eta_B^{(l)}), \tag{B.2f}$$

for $k, l = 1,..., n$ $(k \neq l)$. It is convenient to introduce the notation

$$\Psi_{k,l}^{(g)} = (\Theta_{k,l}^{(g)}, \Gamma_{k,l}^{(g)}, \Delta_{k,l}^{(g)})^{\mathrm{T}}, \tag{B.3}$$

where T is a notation for transpose. Then the recurrence relations

$$\theta_k^{(g)} = u(2-u) + (1-u)^2(1-\lambda)\,\theta_k^{(g-1)}, \tag{B.3a}$$

$$\begin{aligned}\gamma_k^{(g)} &= u + u^2(1-u) + 2u(1-u)^2(1-\lambda)\,\theta_k^{(g-1)} \\ &\quad + (1-u)^3\,[\lambda(1-\lambda)\,\theta_k^{(g-1)} + (1-\lambda)(1-2\lambda)\,\gamma_k^{(g-1)}],\end{aligned} \tag{B.3b}$$

$$\begin{aligned}\delta_k^{(g)} &= u^2(2-u)^2 + 2u(2-u)(1-u)^2(1-\lambda)\,\theta_k^{(g-1)} \\ &\quad + (1-u)^4\,[2\lambda^2(1-\lambda)\,\theta_k^{(g-1)} + 4\lambda(1-\lambda)(1-2\lambda)\,\gamma_k^{(g-1)} \\ &\quad + (1-\lambda)(1-2\lambda)(1-3\lambda)\,\delta_k^{(g-1)}],\end{aligned} \tag{B.3c}$$

$$\begin{aligned}\Psi_{k,l}^{(g)} &= [u^2(2-u)^2 + u(2-u)(1-u)^2(1-\lambda)(\theta_k^{(g-1)} + \theta_l^{(g-1)})]\,\mathbf{1} \\ &\quad + (1-u)^4\,M(r_{k,l})\,\Psi_{k,l}^{(g-1)},\end{aligned} \tag{B.3d}$$

can be deduced, where $\lambda = 1/2N$, $\mathbf{1} = (1, 1, 1)^{\mathrm{T}}$,

$$M(r) = \begin{bmatrix} r^2 + (1-2r)(1-\lambda) & 2r(1-r)(1-2\lambda) & r^2(1-2\lambda) \\ \lambda[(1-\lambda) - r(1-2\lambda)] & [(1-\lambda) - r(1-4\lambda)](1-2\lambda) & r(1-2\lambda)(1-3\lambda) \\ 2\lambda^2(1-\lambda) & 4\lambda(1-\lambda)(1-2\lambda) & (1-\lambda)(1-2\lambda)(1-3\lambda) \end{bmatrix} \tag{B.3e}$$

and $r_{k,l}$ is the recombination rate between loci k and l. Notice that $M(r)$ is a positive matrix of spectral radius less or equal to $(1-\lambda)$.

(b) *Deduction of* (B.3)

First of all, notice that every measure of (B.2) is an expectation of an indicator random variable and then a probability. It may also be useful to point out that chromosomes within the same generation are exchangeable by our assumptions. Probabilities will be written inside square brackets in the following.

For (B.3a). $\xi_A^{(k)} \not\equiv \zeta_A^{(k)}$ if at least one is a mutant $[2(2-u)]$ and otherwise $[(1-u)^2]$ if the parental genes are distinct $[(1-\lambda)]$ and different $[\theta_k^{(g-1)}]$.

For (B.3b). $\zeta_A^{(k)} \not\equiv \xi_A^{(k)} \not\equiv \xi_B^{(k)}$ if $\xi_A^{(k)}$ is a mutant $[u]$, or otherwise $[(1-u)]$ if $\zeta_A^{(k)}$ and $\xi_B^{(k)}$ are both mutants $[u^2]$ or only one is a mutant $[2u(1-u)]$ and the parents of both others are distinct and different $[(1-\lambda)\,\theta_k^{(g-1)}]$. If there are no mutants $[(1-u)^3]$, the event occurs if the parents of $\zeta_A^{(k)}$ and $\xi_B^{(k)}$ are the same one, distinct and different from that of $\xi_A^{(k)}$ $[\lambda(1-\lambda)\,\theta_k^{(g-1)}]$, or distinct ones, both distinct and different from that of $\xi_A^{(k)}$ $[(1-\lambda)(1-2\lambda)\,\gamma_k^{(g-1)}]$.

For (B.3c). $\xi_A^{(k)} \not\equiv \zeta_A^{(k)}$ and $\xi_B^{(k)} \not\equiv \zeta_B^{(k)}$ if there is at least one mutant in each pair $[u^2(1-u)^2]$ or in only one $[2u(2-u)(1-u)^2]$ and the parents for the other are distinct and different $[(1-\lambda)\,\theta_k^{(g-1)}]$. If there are no mutants $[(1-u)^4]$, the event occurs if the parents for each pair are distinct and different and this occurs with probability $2\lambda^2(1-\lambda)\,\theta_k^{(g-1)}$, $4\lambda(1-\lambda)\,\gamma_k^{(g-1)}$ or $4\lambda(1-\lambda)(1-2\lambda)\,\delta_k^{(g-1)}$ accordingly as the total number of parents is 4, 3 or 2, respectively.

For (B.3d). The expressions for $\Theta_{k,l}^{(g)}$, $\Gamma_{k,l}^{(g)}$, and $\Delta_{k,l}^{(g)}$ are obtained as previously up to the case of no mutants. In this last case $[(1-u)^4]$, the relation is expressed by a matrix which depends only on the recombination rate between loci k and l. We have denoted it by M. It has appeared previously in the literature (Weir and Cockerham, 1969, 1974; Serant, 1974).

(c) *Solution of* (B.3)

It is easy to find out the following determinations:

$$\theta_k^{(g)} = \theta_k^{(\infty)} + (\theta_k^{(0)} - \theta_k^{(\infty)})(1-u)^{2g}\,(1-\lambda)^g, \tag{B.4a}$$

$$\gamma_k^{(g)} = \gamma_k^{(\infty)} + C_k(1-u)^{2g}\,(1-\lambda)^g + (\gamma_k^{(0)} - \gamma_k^{(\infty)} - C_k)(1-u)^{3g}\,(1-\lambda)^g\,(1-2\lambda)^g, \tag{B.4b}$$

$$\delta_k^{(g)} = \delta_k^{(\infty)} + D_k(1-u)^{2g}\,(1-\lambda)^g + G_k(1-u)^{3g}\,(1-\lambda)^g\,(1-2\lambda)^g + (\delta_k^{(0)} - \delta_k^{(\infty)} - D_k - G_k)(1-u)^{4g}\,(1-\lambda)^g\,(1-2\lambda)^g\,(1-3\lambda)^g, \tag{B.4c}$$

$$\Psi_{k,l}^{(g)} = \Psi_{k,l}^{(\infty)} + (1-u)^{2g}(1-\lambda)^g \Phi_{k,l} + (1-u)^{4g} M(r_{k,l})^g [\Psi_{k,l}^{(0)} - \Psi_{k,l}^{(\infty)} - \Phi_{k,l}], \tag{B.4d}$$

where

$$\theta_k^{(\infty)} = \theta = \frac{u(2-u)}{1-(1-u)^2(1-\lambda)}, \tag{B.4e}$$

$$\gamma_k^{(\infty)} = \gamma = \frac{u(1+u-u^2) + [2u+\lambda(1-u)](1-u)^2(1-\lambda)\theta}{1-(1-u)^3(1-\lambda)(1-2\lambda)}, \tag{B.4f}$$

$$\delta_k^{(\infty)} = \delta = \frac{u^2(2-u)^2 + 2[u(2-u)+\lambda^2(1-u)^2](1-u)^2(1-\lambda)\theta + 4\lambda(1-u)^4(1-\lambda)(1-2\lambda)\gamma}{1-(1-u)^4(1-\lambda)(1-2\lambda)(1-3\lambda)}, \tag{B.4g}$$

$$C_k = \frac{[2u+\lambda(1-u)](\theta_k^{(0)} - \theta)}{1-(1-u)(1-2\lambda)}, \tag{B.4h}$$

$$D_k = \frac{4\lambda(1-u)^2(1-2\lambda)C_k + 2[u(2-u)+\lambda^2(1-u)^2](\theta_k^{(0)} - \theta)}{1-(1-u)^2(1-2\lambda)(1-3\lambda)}, \tag{B.4i}$$

$$G_k = \frac{4\lambda(1-u)(\gamma_k^{(0)} - \gamma - C_k)}{1-(1-u)(1-3\lambda)}, \tag{B.4j}$$

$$\Phi_{k,l} = u(2-u)(\theta_k^{(0)} + \theta_l^{(0)} - 2\theta)[I-(1-\lambda)^{-1}(1-u)^2 M(r_{k,l})]^{-1}\mathbf{1} \tag{B.4k}$$

$$\Psi_{k,l}^{(\infty)} = \Psi(r_{k,l}) = \theta^2[1-(1-u)^4(1-\lambda)^2][I-(1-u)^4 M(r_{k,l})]^{-1}\mathbf{1}. \tag{B.4l}$$

Of course, θ, γ, δ, and Ψ give the unique stable equilibrium measures. Let $\Psi(r) = (\Theta(r), \Gamma(r), \Delta(r))^{\mathrm{T}}$. Then the further determinations

$$\Theta(r) = (r^2 p_1(r) + r\lambda p_2(r) + \lambda^2 p_3(r))\,\theta^2/q(r), \tag{B.5a}$$

$$\Gamma(r) = (r^2 p_1(r) + r\lambda p_4(r) + \lambda^2 p_5(r))\,\theta^2/q(r), \tag{B.5b}$$

$$\Delta(r) = (r^2 p_1(r) + r\lambda p_4(r) + \lambda^2 p_6(r))\,\theta^2/q(r), \tag{B.5c}$$

with

$$q(r) = r^2 p_1(r) + r\lambda p_4(r) + \lambda^2 p_7(r), \tag{B.5d}$$

can be obtained, where $p_1(r),\ldots, p_7(r)$ are polynomials of r different from zero at $r=0$ given by

$$p_1(r) = (4+8\beta) - (2+4\beta)\,r, \tag{B.5e}$$

$$p_2(r) = (28+80\beta+48\beta^2) - (40+132\beta+92\beta^2)\,r + (13+48\beta+38\beta^2)\,r^2, \tag{B.5f}$$

$$p_3(r) = (36 + 144\beta + 176\beta^2 + 64\beta^3) - (136 + 672\beta + 920\beta^2 + 336\beta^3)\, r + (119 + 700\beta + 1094\beta^2 + 440\beta^3)\, r^2 - (30 + 212\beta + 376\beta^2 + 164\beta^3)\, r^3, \tag{B.5g}$$

$$p_4(r) = (26 + 76\beta + 48\beta^2) - (36 + 124\beta + 92\beta^2)\, r + (9 + 40\beta + 38\beta^2)\, r^2, \tag{B.5h}$$

$$p_5(r) = (24 + 112\beta + 160\beta^2 + 64\beta^3) - (99 + 560\beta + 850\beta^2 + 336\beta^3)\, r + (73 + 548\beta + 986\beta^2 + 440\beta^3)\, r^2 - (8 + 116\beta + 284\beta^2 + 164\beta^3)\, r^3, \tag{B.5i}$$

$$p_6(r) = (22 + 108\beta + 160\beta^2 + 64\beta^3) - (93 + 548\beta + 850\beta^2 + 336\beta^3)\, r + (65 + 532\beta + 986\beta^2 + 440\beta^3)\, r^2 - (4 + 108\beta + 284\beta^2 + 164\beta^3)\, r^3, \tag{B.5j}$$

$$p_7(r) = (18 + 108\beta + 160\beta^2 + 64\beta^3) - (83 + 548\beta + 850\beta^2 + 336\beta^3)\, r + (53 + 532\beta + 986\beta^2 + 440\beta^3)\, r^2 - (108\beta + 284\beta^2 + 164\beta^3)\, r^3, \tag{B.5k}$$

with the notation $\beta = 2Nu$. Terms of order $\lambda, \lambda^2, \ldots$, etc., have been neglected in p_3, p_5, p_6, and p_7.

(d) *Some Approximate Determinations*

It is useful to point out some approximations when N is large enough:

$$\theta = \frac{2\beta}{1 + 2\beta}, \tag{B.6}$$

$$\delta = \frac{1 + 4\beta + 2\beta^2}{3 + 5\beta + 2\beta^2}\,\theta, \tag{B.7}$$

$$\Theta(r) - \theta^2 = \left[\frac{1}{1 + 4\beta}\right]\theta^2 \quad \text{if} \quad r \ll \lambda \tag{B.8a}$$

$$= \left[\frac{5 + 10\beta + 4\beta^2}{12 + 48\beta + 52\beta^2 + 16\beta^3}\right]\theta^2 \quad \text{if} \quad r = \lambda \tag{B.8b}$$

$$= \left[\frac{1 - 2r + 2r^2}{2r - r^2}\right]\lambda\theta^2 \quad \text{if} \quad r \gg \lambda, \tag{B.8c}$$

$$\Delta(r) - \theta^2 = \left[\frac{2}{9 + 54\beta + 80\beta^2 + 32\beta^3}\right]\theta^2 \quad \text{if} \quad r \ll \lambda \tag{B.9a}$$

$$= \left[\frac{1}{12 + 48\beta + 52\beta^2 + 16\beta^3}\right]\theta^2 \qquad \text{if} \quad r = \lambda \qquad \text{(B.9b)}$$

$$= \left[\frac{1 - 2r + 2r^2}{(1 + 2\beta)\, r^2}\right]\lambda^2\theta^2 \qquad \text{if} \quad r \gg \lambda. \qquad \text{(B.9c)}$$

Equation (B.6) is classical (Malécot, 1946; Kimura, 1964), (B.7) can be deduced from Stewart (1976) or Nei and Roychoudhury (1974), and (B.8c) and (B.9c) can be found from Serant (1974).

Acknowledgments

I am very grateful to Professors A. Jacquard, S. Karlin, and B. S. Weir for their encouragement and their valuable comments.

References

Avery, P. J. and Hill, W. G. 1977. Variability in genetic parameters among small populations, *Genet. Res.* **29**, 193–213.

Avery, P. J. and Hill, W. G. 1979. Variance in quantitative traits due to linked dominant genes and variance in heterozygosity in small populations, *Genetics* **91**, 817–844.

Cockerham, C. C. 1969. Variance of gene frequencies, *Evolution* **23**, 72–84.

Cockerham, C. C. 1971. Higher order probability functions of identity of alleles by descent, *Genetics* **69**, 235–246.

Cockerham, C. C. 1973. Analyses of gene frequencies, *Genetics* **74**, 679–700.

Franklin, I. R. 1977. The distribution of the proportion of the genome which is homozygous by descent in inbred individuals, *Theor. Pop. Biol.* **11**, 60–80.

Géry, G. 1978. Coefficient de parenté et dispersion en population finie. Incidence de la mutation, *Ann. Génét. Sél. Anim.* **10**, 533–540.

Haldane, J. B. S. 1919. The combination of linkage values and the calculation of distance between the loci of linked factors, *J. Genet.* **8**, 299–309.

Jacquard, A. 1975. Inbreeding: One word, several meanings, *Theor. Pop. Biol.* **7**, 338–363.

Karlin, S. and Taylor, H. M. 1966. "A First Course in Stochastic Processes," Academic Press, New York.

Karlin, S. and Liberman, U. 1979. A natural class of multilocus recombination processes and related measures of crossover interference, *Adv. Appl. Prob.* **11**, 479–501.

Kimura, M. and Crow, J. F. 1964. The number of alleles that can be maintained in a finite population, *Genetics* **49**, 725–738.

Kingman, J. F. C. 1978. Uses of exchangeability, *Ann. Prob.* **6**, 183–197.

Li, W. H. and Nei, M. 1975. Drift variances of heterozygosity and genetic distance in transient states, *Genet. Res.* **25**, 229–248.

Loève, M. 1960. "Probability Theory," 2nd ed., Van Nostrand, Princeton, N. J.

Malécot, G. 1946. La consanguinité dans une population limitée, *C. R. Acad. Sci.* **222**, 841–843.

Malécot, G. 1948. "Les Mathématiques de l'Hérédité," Masson, Paris.

Nei, M., Fuerst, P. A., and Chakraborty, R. 1976. Testing the neutral mutation hypothesis by distribution of single locus heterozygosity, *Nature (London)* **262**, 491–493.

NEI, M. AND ROYCHOUDHURY, A. K. 1974. Sampling variances of heterozygosity and genetic distance, *Genetics* **76**, 379–390.

NEVO, E. 1978. Genetic variation in Natural populations: Patterns and theory, *Theor. Pop. Biol.* **13**, 121–177.

SCHNELL, F. W. 1961. Some general formulations of linkage effects in inbreeding, *Genetics* **46**, 947–957.

SERANT, D. 1974. Linkage and inbreeding coefficients in finite random mating population, *Theor. Pop. Biol.* **5**, 251–263.

STEWART, F. M. 1976. Variability in the amount of heterozygosity maintained by neutral mutations, *Theor. Pop. Biol.* **9**, 188–201.

SVED, J. A. 1968. The stability of linked systems of loci with a small population size, *Genetics* **59**, 543–563.

SVED, J. A. 1971. Linkage disequilibrium and homozygosity of chromosome segments in finite populations, *Theor. Pop. Biol.* **2**, 125–141.

WEIR, B. S., AVERY, P. J., AND HILL, W. G. 1980. Effect of mating structure on variation in inbreeding, *Theor. Pop. Biol.* **18**, 396–429.

WEIR, B. S. AND COCKERHAM, C. C. 1969. Group inbreeding with two linked loci, *Genetics* **63**, 711–743.

WEIR, B. S. AND COCKERHAM, C. C. 1974. Behavior of pairs of loci in finite monoecious populations, *Theor. Pop. Biol.* **6**, 323–354.

WRIGHT, S. 1951. The genetical structure of populations, *Ann. Eugen.* **15**, 323–354.

WRIGHT, S. 1952. The theoretical variance within and among subdivisions of a population that is in a steady state, *Genetics* **37**, 313–321.

ZOUROS, E. 1979. Mutation rates. Population sizes and amounts of electrophoretic variation of ensyme loci in natural populations, *Genetics* **92**, 623–646.