

THE DYNAMICS OF FINITE HAPLOID POPULATIONS WITH OVERLAPPING GENERATIONS. I. MOMENTS, FIXATION PROBABILITIES AND STATIONARY DISTRIBUTIONS¹

TED H. EMIGH

Department of Statistics, Iowa State University, Ames, Iowa 50011
Present address: Department of Statistics, University of Georgia,
Athens, Georgia 30602

Manuscript received November 15, 1977

Revised copy received July 31, 1978

ABSTRACT

Much of the work on finite populations with overlapping generations has been limited to deriving effective population numbers with the tacit assumption that the dynamics of the population will be similar to a population with nonoverlapping generations and the appropriate population number. In this paper, some exact and approximate results will be presented on the behavior of the first two moments of the gene frequencies. The probability of fixation of a neutral gene is found equal to the initial average reproductive value of the gene, and the means and covariances of the stable distribution with mutation in both directions are found by a simple extension of the values found by assuming nonoverlapping generations.

IN populations with overlapping generations, it is hoped to be able to summarize the complex mating, selection and mutation structure of the population to a few easily understood parameters. FELSENSTEIN (1971) calculated one such parameter, effective population number, for a haploid population with overlapping generations. In this and a subsequent paper, we explore the use of his effective number in describing the dynamics of such a population. To do this, we will introduce several other summary parameters, such as average reproductive value, which are useful in describing the population.

In this paper, we will define the model and present some results under the assumption of no selection. In a subsequent paper, the action of selection will be investigated through the use of the diffusion approximation.

THE MODEL

The model used in this discussion, which was first introduced by FELSENSTEIN (1971), is based on the assumption of a fixed life table and reproductive structure with independent births.

The model is clearly specified in FELSENSTEIN (1971), and only a review is given here. The number of haploid individuals in each age group ($i = 1, 2, \dots, k$)

¹ Journal Paper No. J-9009 of the Iowa Agriculture and Home Economics Experiment Station, Ames, Iowa. Project 1669. This work was supported by Public Health Service grant GM 13827 while the author was at Iowa State University.

is constant over time ($t = 0, 1, 2, \dots$) and is equal to $N(i)$. Newborn individuals are independently chosen from parents, with p_i equal to the probability that the parent of any particular newborn individual at time $t + 1$ came from age group i at time t . Individuals in age group $i + 1$ at time $t + 1$ are chosen without replacement from individuals in age group i at time t . Although we call age one the newborn age group, it is really the group at which the zygote is formed. We can consider this group as distinct individuals or as merely the zygote. If the latter is chosen, we may consider birth at age group two and thus allow for selection among zygotes.

FELSENSTEIN did not consider mutation or selection, but it is easy to incorporate this in the framework of his model. The age-specific mutation rates from A_1 to A_2 and A_2 to A_1 are μ_i and ν_i , respectively, for parents in age group i . The age-specific fertility selection coefficient of A_1 over A_2 is denoted r_i for parents in age group i , and the viability selection coefficient of A_1 over A_2 in surviving from age group $i - 1$ to i is denoted s_i . Let $N_1(i, t)$ be the number of individuals with allele A_1 in age group i at time t . Then, for the newborn age group,

$$N_1(1, t+1) | \mathbf{N}_1(t) \sim \text{Binomial} [N(1), P_r(t+1)], \tag{1}$$

independently of survival, where $\mathbf{N}_1(t) = (N_1(1, t), \dots, N_1(k, t))'$, $\mathbf{r} = (r_1, \dots, r_k)'$ and

$$P_r(t+1) = \sum_{i=1}^k p_i \left\{ \frac{(1 - \mu_i)(1 + r_i)N_1(i, t) + \nu_i[N(i) - N_1(i, t)]}{N(i) + r_i N_1(i, t)} \right\}$$

$$= v^* + \sum_i p_i (1 - \mu_i - \nu_i) \frac{(1 + r_i)N_1(i, t)}{N(i) + r_i N_1(i, t)}$$

with

$$v^* = \sum_i p_i \nu_i .$$

For survival,

$$N_1(i+1, t+1) | N_1(i, t) \sim \text{Noncentral hypergeometric} (N(i), N_1(i, t), N(i+1), s_{i+1}), \tag{2}$$

$$i=1, \dots, k-1,$$

$$\text{and } N_2(i+1, t+1) = N(i+1) - N_1(i+1, t+1), \tag{3}$$

independently of the other age groups and reproduction. The noncentral hypergeometric distribution is defined such that the probability of drawing gene A_1 if there are (a) A_1 genes and (b) A_2 genes left before that draw is a $(1 + s_{i+1}) / [a(1 + s_{i+1}) + b]$, instead of $a / [a + b]$ for the central hypergeometric distributions (see WALLENUS 1963, or JOHNSON and KOTZ 1969). FELSENSTEIN (1971; see also EMIGH 1976) has shown that, for this type of population, the effective population number is, ignoring selection,

$$N_e = \frac{N(1)T}{1 + \sum_i q_{i+1}^2 \left(\frac{1}{l_{i+1}} - \frac{1}{l_i} \right)}, \tag{4}$$

where $q_i = \sum_{j \geq i} p_j$ = probability of reproducing at or after age i , $l_i = N(i) / N(1)$ = probability of surviving to age i , and $T = \sum_i q_i = \sum_i i p_i$, the generation time. The

quantity q_i is related to FISHER's (1930) reproductive value as $v_i = q_i/l_i$. Letting $w_i = \frac{l_{i+1}}{l_i}$ and $d_i = 1 - w_i$, the effective population number can be written as

$$N_e = \frac{N(1)T}{1 + \sum_i l_i w_i d_i v_{i+1}^2} .$$

MOMENTS OF THE GENE FREQUENCIES

Let $X(i,t) = N_1(i,t)/N(i,t)$, the frequency of gene A_1 in age group i at time t , and define $\mathbf{X}(t) = [X(1,t), \dots, X(k,t)]'$. At the next time interval, $t+1$, the gene frequencies will change, governed by the probability distribution (1) and (2). The means and covariances are important characteristics of this change.

The means

Consider the newborn age class.

$$\begin{aligned} E[X(1,t+1) | \mathbf{N}_1(t)] &= P_r(t+1) \\ &= v^* + \sum_i p_i(1-\mu_i-v_i) \frac{(1+r_i)N_1(i,t)}{N(i)+r_iN_1(i,t)} . \end{aligned} \tag{6}$$

If $\{\mu_i\}$, $\{v_i\}$ and $\{r_i\}$ are $O\left(\frac{1}{N_e}\right)$, then (6) can be written as

$$\begin{aligned} E[X(1,t+1) | \mathbf{N}_1(t)] &= v^* + \sum_i p_i(1-\mu_i-v_i)X(i,t) \\ &\quad + \sum_i p_i r_i X(i,t) [1-X(i,t)] + O\left(\frac{1}{N_e^2}\right) . \end{aligned} \tag{7}$$

If there is no selection, then (6) can be written as

$$E[X(1,t+1) | \mathbf{N}_1(t)] = v^* + \sum_i p_i(1-\mu_i-v_i) X(i,t), \tag{8}$$

and with no selection or mutation,

$$E[X(1,t+1) | \mathbf{N}_1(t)] = \sum_i p_i X(i,t). \tag{9}$$

Consider age group $i+1$. If s_{i+1} is $O\left(\frac{1}{N_e}\right)$ and $N(i+1)$ is large, then the mean of the noncentral hypergeometric distribution given in equation (2) is approximately.

$$\begin{aligned} E[X(i+1, t+1) | N_1(i,t)] \\ &= (1+s_{i+1})X(i,t) - s_{i+1}X^2(i,t) + O\left(\frac{1}{N_e^2}\right) . \end{aligned} \tag{10}$$

This approximation is valid if either (1) $N(i+1)$ is small in comparison to $N(i)$, such as with some trees; or (2) $N(i+1)$ is close to $N(i)$, such as with humans. As $N(i)$ becomes very large, the noncentral hypergeometric distribution approximates the binomial. Caution must be exercised when a moderate drop in population number occurs during one year, although with very small s_{i+1} , even this is not serious.

If no selection is present, the mean is exactly

$$E[X(i+1,t+1) | N_1(i,t)] = X(i,t) . \tag{11}$$

If there is no selection present in the population, $r_i = s_i = 0$, all i , the means can be written in matrix form as

$$E[\mathbf{X}(t+1) | \mathbf{X}(t)] = \underline{v}^* + \mathbf{P}\mathbf{X}(t) ; \tag{12}$$

where

$$\underline{v}^* = (v^*, 0, 0, \dots, 0)',$$

and

$$\mathbf{P} = \begin{bmatrix} p_1(1-\mu_1-v_1) & p_2(1-\mu_2-v_2) & \dots & p_{k-1}(1-\mu_{k-1}-v_{k-1}) & p_k(1-\mu_k-v_k) \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}$$

This same type of matrix (without mutation) is described by HILL (1972), and is a special case of the type discussed by LESLIE (1945).

Hence, if $\mathbf{M}_t = E[\mathbf{X}(t)]$, then a recurrence relation can be written for the means when there is no selection as

$$\mathbf{M}_{t+1} = \underline{v}^* + \mathbf{P} \mathbf{M}_t \tag{13}$$

Variances, covariances and higher moments

Now, consider the variance-covariance matrix of the gene frequencies. If the selection coefficients are small [s_i, r_i are $O(1/N_e)$], then the variance will be approximately the same [$O(1/N_e^2)$] as if there was no selection. Let

$$\mathbf{V}(t+1) = E[(\mathbf{X}(t+1) - \mathbf{M}_{t+1})(\mathbf{X}(t+1) - \mathbf{M}_{t+1})'] .$$

The $(i,j)^{th}$ element of $\mathbf{V}(t+1)$ is

$$V_{ij}(t+1) = E[(X(i,t+1) - M_{i,t+1})(X(j,t+1) - M_{j,t+1})] ,$$

where $M_{i,t+1}$ is the i^{th} element of \mathbf{M}_{t+1} . By first conditioning on $\mathbf{X}(t)$, and then taking the expected value over $\mathbf{X}(t)$, we obtain,

$$V_{ij}(t+1) = E_{\mathbf{X}(t)} \{ \text{Cov}[X(i,t+1), X(j,t+1) | \mathbf{X}(t)] + \text{Cov}_{\mathbf{X}(t)} \{ E[X(i,t+1) | \mathbf{X}(t)], E[X(j,t+1) | \mathbf{X}(t)] \} \} ,$$

where $E_{\mathbf{X}(t)}$ is the expected value over $\mathbf{X}(t)$ and $\text{Cov}_{\mathbf{X}(t)}$ is the covariance over $\mathbf{X}(t)$.

Define $X'(1,t+1) = v^* + \sum_i p_i(1-\mu_i-v_i) X(i,t)$, and $X'(i,t+1) = X(i-1,t)$, $i \neq 1$.

Then,

$$E[X(i,t+1) | \mathbf{X}(t)] = X'(i,t).$$

Hence,

$$V_{ij}(t+1) = E_{\mathbf{X}(t)} \{ \text{Cov}[X(i,t+1), X(j,t+1) | \mathbf{X}(t)] + \text{Cov}_{\mathbf{X}(t)} [X'(i,t), X'(j,t)] \} \tag{14}$$

Consider the newborn age group ($i = j = 1$),

$$\text{Var} [X(1,t+1) | \mathbf{X}(t)] = \frac{X'(1,t)(1-X'(1,t))}{N(1)} . \tag{15}$$

The expected value of this quantity is

$$\frac{1}{N(1)} M_{1,t+1} - \frac{1}{N(1)} \{ v^{*2} + 2v^* \sum_i p_i(1-\mu_i-v_i) M_{i,t} + \sum_{ij} p_i p_j (1-\mu_i-v_i)(1-\mu_j-v_j) (V_{ij}(t) + M_{i,t} M_{j,t}) \}$$

Since $M_{1,t+1} = v^* + \sum_i p_i(1-\mu_i-\nu_i) M_{i,t}$, this can be rewritten as

$$\frac{1}{N(1)} M_{1,t+1} - \frac{1}{N(1)} M_{1,t+1}^2 - \frac{1}{N(1)} \sum_{ij} p_i p_j (1-\mu_i-\nu_i)(1-\mu_j-\nu_j) V_{ij}(t). \tag{16}$$

The variance of $X'(1,t)$ is easily calculated as

$$\sum_{ij} p_i p_j (1-\mu_i-\nu_i)(1-\mu_j-\nu_j) V_{ij}(t). \tag{17}$$

Combining (16) and (17) into (14),

$$V_{11}(t+1) = \left(1 - \frac{1}{N(1)}\right) \sum_{ij} p_i p_j (1-\mu_i-\nu_i)(1-\mu_j-\nu_j) V_{ij}(t) + \frac{1}{N(1)} M_{1,t+1}(1-M_{1,t+1}). \tag{18}$$

For $i = 1, j > 1$, since reproduction is independent of survival,

$$\text{Cov} \{X(1, t+1), X(j, t+1) \mid \mathbf{X}(t)\} = 0,$$

so

$$V_{1j}(t+1) = \sum_i p_i(1-\mu_i-\nu_i) V_{i,j-1}(t). \tag{19}$$

If $i \neq j, i, j > 1$

$$\text{Cov} [X(i, t+1), X(j, t+1) \mid \mathbf{X}(t)] = 0,$$

so

$$V_{ij}(t+1) = V_{i-1, j-1}(t). \tag{20}$$

Finally, if $i=j \neq 1$,

$$\text{Var} [X(i,t+1) \mid \mathbf{X}(t)] = X(i-1,t) [1-X(i-1,t)] \frac{N(i-1)-N(i)}{[N(i-1)-1]N(i)},$$

so

$$V_{ii}(t+1) = \frac{1 - \frac{1}{N(i)}}{1 - \frac{1}{N(i-1)}} V_{i-1,i-1}(t) + \frac{\frac{1}{N(i)} - \frac{1}{N(i-1)}}{1 - \frac{1}{N(i-1)}} M_{i-1,t}(1-M_{i-1,t}). \tag{21}$$

Using elementary properties of the binomial and hypergeometric distributions, it is easy to show that the third and fourth moments are $O(1/N_e^2)$ (see EMIGH 1976), although the skewness and kurtosis may not be negligible. In the next sections, these moments will be used to obtain elementary properties of the population.

PROBABILITY OF FIXATION OF A NEUTRAL GENE

Assume that there is no mutation or selection in the population. Since the population has finite size either A_1 or A_2 will eventually become fixed in the population if there exists a j such that $p_j, p_{j+1} > 0$. The condition that $p_j, p_{j+1} > 0$ assures that the population will not be cyclic. Populations where this condition may not hold are periodic cicadas (*Magicicada* sp.), with different populations emerging each year. It is then possible to have some populations fixed with gene A_1 and other populations fixed with gene A_2 . The above condition assures that there will be proper mixing across years.

For neutral genes, equation (12) becomes

$$E[\mathbf{X}(t+1) \mid \mathbf{X}(t)] = \mathbf{P} \mathbf{X}(t), \tag{22}$$

where

$$\mathbf{P} = \begin{bmatrix} p_1 & p_2 & \dots & p_{k-1} & p_k \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}$$

The matrix \mathbf{P} , is a special case of the type discussed by LESLIE (1945). The only positive real eigenvalue of this matrix is 1, and the other eigenvalues (complex and negative) have absolute values less than 1 so long as the process is not cyclic (see e.g., POLLAK and KEMPTHORNE 1970).

Therefore,

$$E[\mathbf{X}(t) \mid \mathbf{X}(0)] = \mathbf{P}^t \mathbf{X}(0). \tag{23}$$

As t increases, $\mathbf{X}(t) \mid \mathbf{X}(0)$ tends to a degenerate stable distribution with all the probability mass on two points, and

$$E[\mathbf{X}(t) \mid \mathbf{X}(0)] \rightarrow \lim_{t \rightarrow \infty} \mathbf{P}^t \mathbf{X}(0) = \mathbf{j} \mathbf{q}' \mathbf{X}(0), \tag{24}$$

where \mathbf{q} is such that $\mathbf{q}' = \mathbf{q}' \mathbf{P}$ and $\mathbf{q}' \mathbf{j} = 1$, where $\mathbf{j} = (1, \dots, 1)'$. Thus,

$$\lim_{t \rightarrow \infty} E[X(i,t) \mid \mathbf{X}(0)] = \mathbf{q}' \mathbf{X}(0) .$$

Denote the elements of \mathbf{q}' by $\frac{q_i}{T}$, $i = 1, \dots, k$, where the values for the q 's and T are to be determined. The equation $\mathbf{q}' = \mathbf{q}' \mathbf{P}$ may be written as

$$q_1 p_i + q_{i+1} = q_i, \quad i = 1, \dots, k-1, \tag{25}$$

and

$$q_1 p_k = q_k .$$

It is easy to verify that

$$q_i = \sum_{m \geq i} p_m$$

solves (25). The condition $\mathbf{q}' \mathbf{j} = 1$ implies that $T = \sum_i i p_i =$ Average age at reproduction, as follows:

$$1 = T^{-1} \sum_i q_i = T^{-1} \sum_i \sum_{m \geq i} p_m = T^{-1} \sum_i i p_i .$$

Therefore,

$$\begin{aligned} \lim_{t \rightarrow \infty} E[X(i,t) \mid \mathbf{X}(0)] &= x(0) \\ &= \frac{1}{T} \sum_i q_i X(i,0) . \end{aligned} \tag{26}$$

Since there is no mutation, if the population ever becomes entirely A_1 , or entirely A_2 , it will stay that way. Hence, the states $\mathbf{X}(t) = 0,0, \dots, 0)'$ and $\mathbf{X}(t) = (1, 1, \dots, 1)'$ are absorbing states. It is not difficult to see that all states communicate with at least one of the two absorbing states. Therefore, the population eventually will become entirely A_1 or entirely A_2 , so long as it is not cyclic. Hence,

$$\begin{aligned} \lim_{t \rightarrow \infty} E[X(i,t) \mid \mathbf{X}(0)] &= P[\text{Fixation of } A_1] \cdot 1 \\ &\quad + P[\text{Loss of } A_1] \cdot 0 \\ &= P[\text{Fixation of } A_1]. \end{aligned}$$

Therefore,

$$\begin{aligned}
 P[\text{Fixation} \mid \mathbf{X}(0)] &= x(0) \\
 &= \frac{1}{T} \sum_i q_i X(i,0).
 \end{aligned}
 \tag{27}$$

While the model used here assumes a constant number of individuals in the age groups, we need only the relationship (22) and that the gene will always go toward fixation or loss. In particular, the values for $N(i)$ do not appear. In a related model, EMIGH (1976) has shown that (27) holds for the case with $N(1)$ fixed and the p 's fixed. Survival is accomplished through the binomial distribution rather than the hypergeometric distribution.

It is possible to allow $N(1)$ to vary, either stochastically or deterministically, so long as $N(i,t) < N^*$, for some N^* . This guarantees that the gene will be either lost or fixed in the population. If the population were allowed to grow fast enough, it could be possible to have both genes present in the population for every t .

If the p 's also are considered random variables independent of t and $\mathbf{X}(t)$, then the p 's in (5) are replaced by their expected values and

$$x(0) = \frac{1}{E(T)} \sum_i E(q_i) X(i,0).$$

The joint distribution of the p 's does not matter, but the condition on the p 's then becomes a condition on the p 's in probability.

It is obvious that, as FISHER (1930) surmised, the reproductive value of a population is an important quantity in describing the population.

STABLE DISTRIBUTION WITH MUTATION AND NO SELECTION

Consider the population to have mutation but no selection. With mutation in both directions (*i.e.*, from A_1 to A_2 as well as from A_2 to A_1), one gene cannot be fixed in the population. The Markov chain describing the population can be seen to be an irreducible positive recurrent Markov chain, since, with mutation, any state of the population is possible and the population always will return to any given state after a sufficient time. Therefore, the population will reach a stable distribution and

$$\mathbf{M}_{t+1} = \mathbf{M}_t = \mathbf{M}.$$

Hence, from (12),

$$\mathbf{M} = \underline{v}^* + \mathbf{P} \mathbf{M},$$

or, so long as $\mu^* + \nu^* > 0$, then $(\mathbf{I} - \mathbf{P})^{-1}$ exists, and

$$\mathbf{M} = (\mathbf{I} - \mathbf{P})^{-1} \underline{v}^*.
 \tag{28}$$

Since \underline{v}^* is a column with only one non-zero element, this being in the first position, \mathbf{M} is the first column of $(\mathbf{I} - \mathbf{P})^{-1}$ multiplied by \underline{v}^* . Designate the first column of $(\mathbf{I} - \mathbf{P})^{-1}$ by \mathbf{F} , with the elements of \mathbf{F} denoted by f_i , $i = 1, \dots, k$, and the i^{th} row of $(\mathbf{I} - \mathbf{P})$ by \mathbf{G}_i . Then, since $(\mathbf{I} - \mathbf{P})(\mathbf{I} - \mathbf{P})^{-1} = \mathbf{I}$,

$$\mathbf{G}_1 \mathbf{F} = 1,$$

and

$$\mathbf{G}_i \mathbf{F} = 0, \quad i = 2, \dots, k.
 \tag{29}$$

For $i \neq 1$, $\mathbf{G}_i = (0, \dots, -1, 1, \dots, 0)$, with -1 in the $(i-1)^{th}$ position and 1 in the i^{th} position, with the other elements zero.

Therefore, for $i \neq 1$,

$$\mathbf{G}_i \mathbf{F} = -f_{i-1} + f_i = 0,$$

or

$$f_i = f, \text{ for all } i. \tag{30}$$

The i^{th} term of \mathbf{G}_1 is $-p_i(1 - \mu_i - \nu_i)$ if $i \neq 1$ and $1 - p_1(1 - \mu_1 - \nu_1)$ if $i = 1$.

Therefore,

$$\begin{aligned} 1 = \mathbf{G}_1 \mathbf{F} &= -\sum_i p_i(1 - \mu_i - \nu_i) f_i + f_1 \\ &= -f \sum_i p_i(1 - \mu_i - \nu_i) + f \\ &= f(\mu^* + \nu^*), \end{aligned} \tag{31}$$

where $\mu^* = \sum_i p_i \mu_i$ and $\nu^* = \sum_i p_i \nu_i$.

Hence,

$$f_i = 1/(\mu^* + \nu^*), \text{ for all } i,$$

or

$$\mathbf{M} = M \mathbf{j}, \tag{32}$$

where $M = \frac{\nu^*}{\mu^* + \nu^*}$ and \mathbf{j} is a column of ones.

That is, for every i ,

$$\lim_{t \rightarrow \infty} E[X(i,t)] = \frac{\nu^*}{\mu^* + \nu^*}. \tag{33}$$

Notice that this is independent of the initial gene frequencies.

To find the variance of the stable distribution, note that $\mathbf{V}(t+1) = \mathbf{V}(t) = \mathbf{V}$ and $\mathbf{M}_{t+1} = \mathbf{M}_t = \mathbf{M}$, where $\mathbf{M} = M \mathbf{j}$, with $M = \nu^*/(\mu^* + \nu^*)$, as above.

Replace \mathbf{M}_t with \mathbf{M} in equations (18) and (21). Then, at equilibrium,

$$V_{ii}(t+1) = (1 - 1/N(1)) \sum_{ij} p_i p_j (1 - \mu_i - \nu_i) (1 - \mu_j - \nu_j) V_{ij}(t) + \frac{1}{N(1)} M(1 - M) \tag{34}$$

and

$$V_{ii}(t+1) = \frac{1 - \frac{1}{N(i)}}{1 - \frac{1}{N(i-1)}} V_{i-1,i-1}(t) + \frac{\frac{1}{N(i)} - \frac{1}{N(i-1)}}{1 - \frac{1}{N(i-1)}} M(1 - M). \tag{35}$$

Now, consider equation (35). Replacing $V_{i-1,i-1}(t)$ by its value in terms of $V_{i-2,i-2}(t-1)$, and simplifying,

$$V_{ii}(t+1) = \frac{1 - \frac{1}{N(i)}}{1 - \frac{1}{N(i-2)}} V_{i-2,i-2}(t+1) + \frac{\frac{1}{N(i)} - \frac{1}{N(i-2)}}{1 - \frac{1}{N(i-2)}} M(1 - M).$$

By repeating this process,

$$V_{ii}(t+1) = \frac{1 - \frac{1}{N(i)}}{1 - \frac{1}{N(1)}} V_{11}(t-i+2) + \frac{\frac{1}{N(i)} - \frac{1}{N(1)}}{1 - \frac{1}{N(1)}} M(1 - M). \tag{36}$$

This is easily seen from noting that in sampling theory, a two-stage subsampling scheme is equivalent to a one-stage sampling scheme. Also, for $i < j$,

$$V_{ij}(t+1) = V_{1,j-i+1}(t-i+2), \tag{37}$$

from (20).

While equations (34), (35) and (37) can be solved directly at equilibrium, we will reduce the number of equations from k^2 to k by considering the covariances $V_{1j}, j=1, \dots, k$. First, consider equation (18),

$$\begin{aligned} V_{11}(t+1) &= \left(1 - \frac{1}{N(1)}\right) \sum_{ij} p_i p_j (1-\mu_i-\nu_i)(1-\mu_j-\nu_j) V_{ij}(t) \\ &\quad + \frac{1}{N(1)} M(1-M) \\ &= \left(1 - \frac{1}{N(1)}\right) \left\{ \sum_{i=1}^{k-1} p_i (1-\mu_i-\nu_i) \sum_j p_j (1-\mu_j-\nu_j) V_{ij}(t) \right. \\ &\quad \left. + p_k (1-\mu_k-\nu_k) \sum_{j=1}^{k-1} p_j (1-\mu_j-\nu_j) V_{kj}(t) \right. \\ &\quad \left. + p_k^2 (1-\mu_k-\nu_k)^2 V_{kk}(t) \right\} + \frac{1}{N(1)} M(1-M) . \end{aligned}$$

For simplicity, assume that $p_k = 0$. This will not exclude any type of age structure; we need only add another age group that does not participate in reproduction. For example, the case of nonoverlapping generations would have $p_1 = 1$ and $p_2 = 0$. Then

$$\begin{aligned} V_{11}(t+1) &= \left(1 - \frac{1}{N(1)}\right) \sum_{ij} p_i p_j (1-\mu_i-\nu_i)(1-\mu_j-\nu_j) V_{ij}(t) \\ &\quad + \frac{1}{N(1)} M(1-M) \end{aligned}$$

Using equation (19) and letting $V_{k,j}(t) = 0$, we obtain

$$V_{11}(t+1) = \left(1 - \frac{1}{N(1)}\right) \sum_i p_i (1-\mu_i-\nu_i) V_{1,i+1}(t+1) + \frac{1}{N(1)} M(1-M) \tag{38}$$

Since the population is in equilibrium, $V_{ij}(t+1) = V_{ij}(t) = \dots = V_{ij}$.

Therefore,

$$V_{11} = \left(1 - \frac{1}{N(1)}\right) \sum_i p_i (1-\mu_i-\nu_i) V_{1,i+1} + \frac{1}{N(1)} M(1-M), \tag{39}$$

$$V_{ij} = V_{1, |j-i| + 1}, i \neq j, i, j > 1 \tag{40}$$

$$V_{ii} = \frac{1-1/N(i)}{1-1/N(1)} V_{11} + \frac{1/N(i) - 1/N(1)}{1-1/N(1)} M(1-M), i > 1, \tag{41}$$

and

$$\begin{aligned} V_{1j} &= \sum_i p_i (1-\mu_i-\nu_i) V_{i,j-1} \\ &= \sum_{i < j-1} p_i (1-\mu_i-\nu_i) V_{1,j-1} + \sum_{i > j-1} p_i (1-\mu_i-\nu_i) V_{1,i-j+2} \\ &\quad + p_{j-1} (1-\mu_{j-1}-\nu_{j-1}) \frac{1 - \frac{1}{N(j-1)}}{1 - \frac{1}{N(1)}} V_{11} \end{aligned} \tag{42}$$

$$+ p_{j-1}(1-\mu_{j-1}-v_{j-1}) \frac{\frac{1}{N(j-1)} - \frac{1}{N(1)}}{1 - \frac{1}{N(1)}} M(1-M), \text{ for } j_i > 1 .$$

It is obvious that only $V_1 = (V_{11}, V_{12}, \dots, V_{1k})'$ is needed in finding the variances and covariances of the stable distribution. Using (40) and (41), it is possible to obtain the remaining elements of V .

Equations (39) and (42) can be written in matrix form as

$$V_1 = A V_1 + B, \tag{43}$$

where

$$B = M(1-M) \begin{bmatrix} \frac{1}{N(1)} \\ 0 \\ p_2(1-\mu_2-v_2) \frac{\frac{1}{N(2)} - \frac{1}{N(1)}}{1 - \frac{1}{N(1)}} \\ \vdots \\ p_{j-1}(1-\mu_{j-1}-v_{j-1}) \frac{\frac{1}{N(j-1)} - \frac{1}{N(1)}}{1 - \frac{1}{N(1)}} \\ \vdots \\ p_{k-1}(1-\mu_{k-1}-v_{k-1}) \frac{\frac{1}{N(k-1)} - \frac{1}{N(1)}}{1 - \frac{1}{N(1)}} \end{bmatrix}$$

and

If $I - A$ is nonsingular, then the solution is

$$V_1 = (I - A)^{-1} B . \tag{44}$$

APPROXIMATE COVARIANCES FOR THE STABLE DISTRIBUTION

While it is possible to solve (44) exactly for particular choices of $N(1), p_i, \mu_i$ and v_i , approximate analytic solutions for populations in terms of the N 's, p 's, μ 's, v 's, etc. are possible.

As a first approximation, assume that $V_{ij} = V_{i'j'}$, for all i, i', j, j' . This approximation is a reasonable first step, considering that this is essentially the assumption needed to obtain the effective number. Recalling that $\sum_i q_i = T$, consider the following weighted average of the V_{ij} 's.

$$\mathbf{A} = \begin{bmatrix}
 0 & \left(1 - \frac{1}{N(1)}\right) p_1 (1 - \mu_1 - v_1) & \dots & \left(1 - \frac{1}{N(1)}\right) p_{k-1} (1 - \mu_{k-1} - v_{k-1}) \\
 p_1 (1 - \mu_1 - v_1) & p_2 (1 - \mu_2 - v_2) & \dots & p_k (1 - \mu_k - v_k) \\
 \frac{1 - \frac{1}{N(2)}}{1 - \frac{1}{N(1)}} p_2 (1 - \mu_2 - v_2) & p_1 (1 - \mu_1 - v_1) & \dots & 0 \\
 \vdots & + p_3 (1 - \mu_3 - v_3) & \vdots & \vdots \\
 \frac{1 - \frac{1}{N(i-1)}}{1 - \frac{1}{N(1)}} p_{i-1} (1 - \mu_{i-1} - v_{i-1}) & p_{i-2} (1 - \mu_{i-2} - v_{i-2}) & \dots & 0 \\
 \vdots & + p_i (1 - \mu_i - v_i) & \vdots & \vdots \\
 \frac{1 - \frac{1}{N(k-1)}}{1 - \frac{1}{N(1)}} p_{k-1} (1 - \mu_{k-1} - v_{k-1}) & p_{k-2} (1 - \mu_{k-2} - v_{k-2}) & \dots & 0 \\
 \vdots & + p_k (1 - \mu_k - v_k) & \vdots & \vdots
 \end{bmatrix}$$

$$V(t+1) = \frac{1}{T^2} \sum_{ij} q_i q_j V_{ij}(t+1). \tag{45}$$

Let $\sum_{i,j>1}^{\neq}$ refer to the summation over i and j not equal and greater than one. From (18), (19), (20), and (21), the right hand side of (45) is

$$\begin{aligned} & \frac{1}{T^2} \{q_1^2 V_{11}(t+1) + q_1 \sum_{j>1} q_j V_{1j}(t+1) + q_1 \sum_{i>1} q_i V_{i1}(t+1) \\ & + \sum_{i,j>1}^{\neq} q_i q_j V_{ij}(t+1) + \sum_{i>1} q_i^2 V_{ii}(t+1)\} \\ & = \frac{1}{T^2} \{q_1^2 \left(1 - \frac{1}{N(1)}\right) \sum_{ij} p_i (1 - \mu_i - \nu_i) p_j (1 - \mu_j - \nu_j) V_{ij}(t) + q_1^2 \frac{1}{N(1)} M(1-M) \\ & + q_1 \sum_{j>1} q_j \sum_i p_i (1 - \mu_i - \nu_i) V_{i,j-1}(t) + q_1 \sum_{i>1} q_i \sum_j p_j (1 - \mu_j - \nu_j) V_{i-1,j}(t) \\ & + \sum_{i,j>1}^{\neq} q_i q_j V_{i-1,j-1}(t) + \sum_{i>1} q_i^2 \frac{1 - 1/N(i)}{1 - 1/N(i-1)} V_{i-1,i-1}(t) \\ & + \sum_{i>1} q_i^2 \frac{\frac{1}{N(i)} - \frac{1}{N(i-1)}}{1 - \frac{1}{N(i-1)}} M(1-M)\} . \end{aligned}$$

Define $q_{k+1} = 0$ and recall that $q_1 = 1$ and $q_j = \sum_{i \geq j} p_i$, so $q_{i+1} + p_i = q_i$. Replacing each $V_{ij}(t)$ by the approximated value V and simplifying, equation (45) can be written as

$$\begin{aligned} V &= \frac{1}{T^2} \left\{ [T^2 + (1 - \mu^* - \nu^*)^2 - 1 - \frac{1}{N(1)} (1 - \mu^* - \nu^*)^2 \right. \\ & \quad \left. - 2(T-1)(\mu^* + \nu^*) - \frac{1}{N(1)} \sum_i q_{i+1}^2 \left(\frac{1}{l_{i+1}} - \frac{1}{l_i} \right) \right] V \\ & \quad \left. + \frac{1}{N(1)} M(1-M) + \frac{1}{N(1)} \sum_i q_{i+1}^2 \left(\frac{1}{l_{i+1}} - \frac{1}{l_i} \right) M(1-M) \right\}. \end{aligned} \tag{46}$$

From (4), recalling that $\sum q_i = T$, the generation time,

$$\frac{1}{N(1)} \sum_i q_{i+1}^2 \left(\frac{1}{l_{i+1}} - \frac{1}{l_i} \right) = \frac{T}{N_e} - \frac{1}{N(1)}. \tag{47}$$

Hence, solving for V in (46),

$$V = \frac{M(1-M)}{\frac{N_e}{T} \left[\left(1 - \frac{1}{N(1)}\right) (1 - (1 - \mu^* - \nu^*)^2) + 2(T-1)(\mu^* + \nu^*) \right] + 1} \tag{48}$$

By ignoring terms of order μ^{*2} , ν^{*2} and $(\mu^* + \nu^*)/N(1)$, since $N(1)$ is large, this can be approximated by

$$V = \frac{M(1-M)}{1 + 2N_e(\mu^* + \nu^*)}. \tag{49}$$

This is the value found by WRIGHT (1931) for populations with nonoverlapping generations and is valid for large numbers in each of the age classes.

The approximation (48) appears to be quite good. This approximation can be improved somewhat by using (38) through (42) and iterating. The improved estimates are

$$V_{ii} = \left(1 - \frac{1}{N(i)}\right) (1 - \mu^* - \nu^*) V + \frac{1}{N(i)} M(1 - M), \quad (50)$$

and

$$V_{ij} = (1 - \mu^* - \nu^*) V + p_{|i-j|} (1 - \mu_{|i-j|} - \nu_{|i-j|}) \left(\frac{1}{N(|i-j|)} - \frac{1}{N(1)} \right) \\ \cdot [M(1 - M) - (1 - \mu^* - \nu^*) V], \quad i \neq j, \quad (51)$$

where V is given in (48).

A comparison of the exact and approximate covariances of V_{ij} , $j=1, \dots, k$ is presented in Table 1. The relative fertilities follow a similar pattern to that observed in females in the United States (EMIGH and POLLAK 1979) although the life table is different. The approximations are extremely good (within order $\mu^* + \nu^*$).

SUMMARY

The moments of the gene frequencies in a haploid model with overlapping generation are found through recurrence formulas that are easy to evaluate. In a population without mutation and selection, the unconditional mean reproductive value of the gene in the population, $E\left(\frac{1}{T} \sum_i q_i X(i, t)\right)$, is unchanged with time. This leads to the result that the probability of fixation of a neutral gene is equal to the relative initial reproductive value of the gene in the population $\left(\frac{1}{T} \sum_i q_i X(i, 0)\right)$.

With mutation present, but without selection, we are able to derive an exact mean and approximate covariances of the gene frequencies in the resulting stable distribution. The mean is exactly that found for populations with nonoverlapping generations, with $\mu^* = \sum_i p_i \mu_i$ and $\nu^* = \sum_i p_i \nu_i$, average mutation rates, replacing μ and ν . The correlations are seen to tend to unity, although not exactly (within order $\mu^* + \nu^*$). The common approximate variance is the same as for a population with nonoverlapping generations with μ^* , ν^* and N_e replacing μ , ν and N , respectively.

It seems reasonable, therefore, to believe that haploid populations with overlapping generations can be sufficiently described through the reproductive value, with the proper modifications to the parameters, in the same way as a population with nonoverlapping generations. This is investigated in a subsequent paper.

There has been some initial work done on diploid populations by HILL (1972), JOHNSON (1977a,b) and EMIGH and POLLAK (1979). It is expected that diploid populations will give results similar to this paper.

Portions of this paper appeared in my thesis (EMIGH 1976) under the direction of OSCAR KEMPTHORNE; I greatly appreciate his guidance. Helpful comments and suggestions by EDWARD POLLAK on the thesis and this paper also are appreciated. Comments on an earlier draft of this paper by JOSEPH FELSENSTEIN and an anonymous reviewer are appreciated.

TABLE 1

Comparison of true and approximated covariances for a population with overlapping generations

	l_i	f_i	μ_i	ν_i	V_{1i}		μ_i	ν_i	V_{1i}	
					Exact	Approx.			Exact	Approx.
1	1.0	0	—	—	0.249137	0.249140	—	—	0.245568	0.245571
2	1.0	0	—	—	0.249088	0.249094	—	—	0.245520	0.245526
3	0.9	0	—	—	0.249084	0.249094	—	—	0.245516	0.245526
4	0.9	1	10^{-5}	10^{-5}	0.249082	0.249094	10^{-5}	10^{-5}	0.245514	0.245526
5	0.8	3	10^{-5}	10^{-5}	0.249088	0.249095	10^{-5}	10^{-5}	0.245520	0.245527
6	0.7	3	10^{-5}	10^{-5}	0.249101	0.249098	10^{-5}	10^{-5}	0.245533	0.245530
7	0.6	2	10^{-5}	10^{-5}	0.249101	0.249100	10^{-5}	10^{-5}	0.245533	0.245531
8	0.5	1	10^{-5}	10^{-5}	0.249094	0.249099	10^{-5}	10^{-5}	0.245525	0.245531
9	0.5	1	10^{-5}	10^{-5}	0.249086	0.249097	10^{-5}	10^{-5}	0.245518	0.245529
10	0.5	0	—	—	0.249089	0.249097	—	—	0.245521	0.245529
$N(1)$			20				100			
M			0.5000				0.5000			
V			0.2490993				0.2455310			
N_e			91.16				455.78			
μ^*/ν^*			$10^{-5}/10^{-5}$				$10^{-5}/10^{-5}$			

	μ_i	ν_i	V_{1i}		μ_i	ν_i	V_{1i}	
			Exact	Approx.			Exact	Approx.
1	—	—	0.249719	0.249720	—	—	0.221074	0.221077
2	—	—	0.249703	0.249705	—	—	0.221009	0.221017
3	—	—	0.249702	0.249705	—	—	0.221003	0.221017
4	10^{-5}	10^{-5}	0.249701	0.249705	10^{-5}	2×10^{-5}	0.221000	0.221017
5	10^{-6}	10^{-6}	0.249703	0.249705	10^{-5}	2×10^{-5}	0.221009	0.221018
6	10^{-6}	10^{-6}	0.249707	0.249706	10^{-5}	2×10^{-5}	0.221026	0.221022
7	10^{-6}	10^{-6}	0.249707	0.249707	10^{-5}	2×10^{-5}	0.221026	0.221024
8	10^{-5}	10^{-5}	0.249705	0.249707	10^{-5}	2×10^{-5}	0.221016	0.221023
9	10^{-5}	10^{-5}	0.249702	0.249706	10^{-5}	2×10^{-5}	0.221006	0.221021
10	—	—	0.249703	0.249706	—	—	0.221009	0.221021
$N(1)$		20				20		
M		0.5000				0.3333		
V		0.2497066				0.2210235		
N_e		91.16				91.16		
μ^*/ν^*		$3.25 \times 10^{-6}/3.25 \times 10^{-6}$				$10^{-5}/2 \times 10^{-5}$		

LITERATURE CITED

EMIGH, T. H., 1976 The effects of finite population size on genetic populations with overlapping generations. Ph.D. Thesis, Iowa State University.

EMIGH, T. H. and E. POLLAK, 1979 Fixation probabilities and effective population numbers in diploid populations with overlapping generations. Theoret. Popul. Biol. **15**: (in press).

FELSENSTEIN, J., 1971 Inbreeding and variance effective numbers in populations with overlapping generations. Genetics **68**: 581-597.

FISHER, R. A., 1930 *The Genetical Theory of Natural Selection*. Oxford University Press, Oxford.

- HILL, W. G., 1972 Effective size of populations with overlapping generations. *Theoret. Popul. Biol.* **3**: 278-289.
- JOHNSON, D. L., 1977a Inbreeding in populations with overlapping generations. *Genetics* **87**: 581-591. —, 1977b Variance-covariance structure of group means with overlapping generations. pp. 851-858. In: *Proceedings of the International Conference on Quantitative Genetics*. Edited by E. POLLACK, O. KEMPTHORNE and T. B. BAILEY, JR. Iowa State University Press, Ames, Iowa.
- JOHNSON, N. L. and S. KOTZ, 1969 *Distributions in statistics: Discrete Distributions*. Houghton Mifflin Co., Boston.
- LESLIE, P. H., 1945 On the use of matrices in certain population mathematics. *Biometrika* **33**: 183-212.
- POLLAK, E. and O. KEMPTHORNE, 1970 Malthusian parameters in genetic populations I. Haploid and selfing models. *Theoret. Popul. Biol.* **1**: 315-345.
- WALLENIUS, K. T., 1963 Biased sampling: The noncentral hypergeometric probability distribution. Ph.D. Thesis, Stanford University.
- WRIGHT, S., 1931 Evolution in Mendelian populations. *Genetics* **16**: 97-159.

Corresponding editor: D. L. HARTL