On the Calculation of the Effective Population Size by a Method Based on the Theory of Branching Processes[1]

E. Pollak
Iowa State University, Ames, USA

SUMMARY

The effective population size N_e of a large population with discrete generations can be calculated from a T-type branching process, in which each type has at least one copy of an allele A in its genotype. With neutral alleles, the branching process is positively regular with a first moment matrix M that has a dominant eigenvalue 1 and corresponding left and right eigenvectors $p' = (p_1, \ldots, p_T)$ and $v = (v_1, \ldots, v_T)'$ which satisfy the equations $\Sigma_i p_i v_i = 1$ and $\Sigma_i p_i = 1$. In this paper it is shown that if Y_{ij} is the number of offspring of type j of a parent of type i, then $\Sigma_i p_i \text{Var}(\Sigma_j Y_{ij} v_j)$ is proportional to $K / 2N_e$, where K is the number of copies of a gene among fertilized eggs. Examples of the calculation of N_e in this way are then given for random mating dioecious populations and populations that reproduce partly by selfing or full-sib mating. A generalization of the theory that applies to age-structured random mating populations is also discussed. Finally, it is shown that N_e / K is proportional to an approximate expression that has recently been obtained for the probability of the long- term survival of A when it is initially present in one individual with a single copy of this allele in its genotype.

Key words : Effective population size, Branching process, Survival probability.

1. Introduction

One of the ways to define the effective size N_e of a population of constant size and unchanging demographic structure is in terms of the variance of the shift in the frequency of an allele. Thus, if there is no selection, q and q_1 are the frequencies of an allele A in generations 0 and 1, and $\Delta q = q_1 - q$, we have

$$\text{Var}(\Delta q) = \frac{q(1-q)}{2N_e}$$

Then N_e is what Crow [2] called the variance effective population number. This concept was developed more fully in later papers by Crow and Morton [4],

Kimura and Crow [16], Crow and Denniston [3], Latter [17], Hill ([11], [12]) and Ethier and Nagylaki [6].

In this paper, the approach that will be used in calculating N_e will be to take some well known results from the theory of multitype branching processes and adapt them to approximate properties of a stochastic process that governs the change in numbers of an initially rare allele A in a large finite population. After a statement of these results in Section 2, it will be shown in Section 3 that the reciprocal of N_e can be approximated by a multiple of an expression on the right sides of equations (5) and (6). The first of these approximate equations applies if there is complete random mating and the second if reproduction is partly by a regular system of inbreeding. Use will then be made of this theory to derive N_e for a dioecious random mating population, a partly self-fertilizing population, and a population consisting of N/2 permanent couples that reproduce partly by full-sib mating. When there is a dioecious population, results will be obtained for both autosomal and sex-linked loci. Generalizations applicable to random mating age-structured populations will also be discussed.

2. *Relevant Results from the Theory of Branching Processes*

Let us consider a T-type branching process with the first moment matrix **M** and finite second moments. It will be assumed that **M** is irreducible and aperiodic, so that all the elements of $\mathbf{M}^t$ are positive for some integer t. It can then be shown that **M** has a simple dominant eigenvalue ρ. In all the examples to be discussed in this paper $\rho = 1$. Associated with this dominant eigenvalue are unique left and right eigenvectors, having only positive elements, that satisfy the equations

$$\mathbf{p}' \mathbf{M} = \mathbf{p}' \ , \ \mathbf{M}\mathbf{v} = \mathbf{v}$$

$$\mathbf{p}' \mathbf{1} = \sum_{i=1}^{T} p_i = 1, \ \mathbf{p}'\mathbf{v} = \sum_{i=1}^{T} p_i v_i = 1 \tag{1}$$

(Here and in the remainder of this paper the transpose of a matrix **A** is denoted by **A**′ and a lower-case boldface letter without a prime indicates a column vector).

Now let $\mathbf{e}'_i$ be equal to a row vector that has 1 in the i-th position and zeros elsewhere. Then the matrix of first moments of the process at time t is given by $\mathbf{e}'_i \mathbf{M}^t$. Because 1 is the dominant eigenvalue of **M**, it follows that

$$\mathbf{e}'_i \mathbf{M}^t = \mathbf{e}'_i \mathbf{v}\mathbf{p}' + \mathbf{M}_2^t = v_i \mathbf{p}' + \mathbf{M}_2^t \tag{2}$$

where the sum of the absolute values of the elements of M_2^t is of the same order of magnitude as a^t for some a, $0 < a < 1$. The foregoing results are presented, for example, by Harris [10].

In what follows, the types in the branching process will be individuals with one copy of A in their genotype if there is complete random mating, but if there is partial inbreeding at least one type could have two or more copies of this allele.

3. A Method for Calculating the Effective Population Size

Suppose that the frequency of an allele A in a population with discrete generations is q at time 0. Then if Δq is the random change in the frequency of A between generations 0 and 1,

$$\text{Var}(\Delta q) = \frac{q(1-q)}{2N_e} \approx \frac{q}{2N_e}$$

if q is small. Thus if the number of copies of a gene is K among fertilized eggs of every generation and z is the number of copies of A,

$$\text{Var}(\Delta z) \approx K^2 \frac{z}{K} \frac{1}{2N_e} = \frac{Kz}{2N_e} \tag{3}$$

if $q = z/K$ is small. Note that if N_m and N_f are the numbers of males and females then K is equal to $2(N_m + N_f) = 2N$ if there is an autosomal locus and to $2N_f + N_M$ if the locus is sex-linked.

Now if t_1 is large but $t_1/2N_e$ is small, (3) implies that the variance of the number of copies of A generated in t_1 generations by z copies at time 0 is approximately

$$t_1 \text{Var}(\Delta z) \approx \frac{t_1 Kz}{2N_e} \tag{4}$$

At the level of approximation used to obtain (3) and (4), the branching process, introduced to Section 2, can be used to obtain an approximation to N_e if K is large and there is initially a single ancestral individual with a single copy of A in its genotype. Let the associated element of v be v_1. Then we have from (2) that there are on the average v_1 descendants of this ancestor at time t_1, of which a proportion p_i are expected to be individuals of type i that have allele A.

Now let us first suppose that each type has only one copy of A_1. Let Y_{ij} be the number of offspring of type j of a single representative of type i. Then if t_2 is large and t_2/K is small, the variance of the number of copies of A generated in $t_1 + t_2$ generations by a single ancestor of type 1 is approximately

$$(t_1 + t_2) \ v_1 \sum_{i=1}^{T} p_i \ \text{Var} (\sum_{j=1}^{T} Y_{ij} \ v_j)$$

because different individuals with A are assumed to reproduce independently. This assumption is reasonable while A_1 is rare, and should approximately hold because in the examples to be considered a is much smaller than $1 - 1/K$. Thus if t_1 is replaced by $t_1 + t_2$ in (4), it follows that

$$\frac{K}{2N_e} \approx \sum_{i=1}^{T} p_i \ \text{Var} (\sum_{j=1}^{T} Y_{ij} \ v_j) \tag{5}$$

if K is large and each type in the branching process has only one copy of A.

We now consider a situation in which each representative of type i in the population has n_i copies of A and $v_i = n_i v_1$. Then, in the long run, the expected number of descendants of a single ancestor of type 1 is v_1, whereas the expected number of copies of A among these descendants is

$$v_1 \sum_{i=1}^{T} n_i p_i = \sum_{i=1}^{T} v_i p_i = 1$$

Thus, if $t_1 + 1$ replaces t_1 and $z = 1$, the right side of (4) becomes $(t_1 + 1) K/2N_e$. It is also true, however, that the variance of the number of copies of A generated in $t_1 + 1$ generations by a single ancestor of type 1 is approximately

$$t_1 v_1 \sum_{i=1}^{T} p_i \ \text{Var} (\sum_{j=1}^{T} Y_{ij} n_j) = \frac{t_1}{v_1} \sum_{i=1}^{T} p_i \ \text{Var} (\sum_{j=1}^{T} Y_{ij} v_j)$$

Hence $\qquad \dfrac{v_1 K}{2N_e} \approx \sum_{i=1}^{T} p_i \ \text{Var} (\sum_{j=1}^{T} Y_{ij} \ v_j) \tag{6}$

if K is large and $v_i = n_i v_1$, where n_i is an integer equal to the number of copies of A in type i.

If T is large, it seems at first sight that it may be difficult to verify that $\rho = 1$, even if A is neutral in comparison with other alleles, collectively denoted by $\overline{A}$, and K is assumed not to change with time. If, however, one can find vectors v and p' that satisfy (1), it will not be necessary to calculate all the eigenvalues of **M**. Suppose that $\rho > 1$ and that the right eigenvector with only positive elements that corresponds to ρ were y. Then we would have

$$p'My = p'y = \rho p'y$$

which implies that $p'y = 0$. But this contradicts the assumption that y has only positive elements. Therefore $\rho = 1$ if vectors v and p', which satisfy (1), can be found.

4. Random Mating Dioecious Populations

We assume that there is a large population with, initially, a single copy of A in a fertilized egg, whereas all other copies are $\overline{A}$. Let types 1 and 2 be respectively males and females that have one copy of A in their genotypes. Then, if N_m and N_f are respectively the numbers of males and females in each generation, the first moment matrix of the branching process is

$$\mathbf{M} = \begin{bmatrix} \dfrac{1}{2}b & \dfrac{c}{2}\dfrac{N_f}{N_m} \\[2ex] \dfrac{1}{2}\dfrac{N_m}{N_f} & \dfrac{1}{2} \end{bmatrix} \tag{7}$$

where $b = c = 1$ if the locus under consideration is autosomal and $b = 0$, $c = 2$ if it is sex-linked.

The eigenvectors corresponding to $\rho = 1$ are then

$$p' = \left[\frac{N_m}{N_m + cN_f} , \frac{cN_f}{N_m + cN_f} \right] \tag{8}$$

and

$$v = \left[\frac{N_m + cN_f}{N_m(1 + c)} , \frac{N_m + cN_f}{N_f(1 + c)} \right] \tag{9}$$

It follows from (5), (8) and (9) that

$$\frac{K}{2N_e} \approx \sum_{i=1}^{2} p_i \ \text{Var}\left(\sum_{j=1}^{2} Y_{ij} v_j \right)$$

$$= \frac{N_m + cN_f}{(1+c)^2} \left\{ N_m \left[\frac{1}{N_m^2} \ Var \ (Y_{11}) + \frac{2}{N_m N_f} \ Cov \ (Y_{11}, Y_{12}) \right. \right.$$

$$\left. + \frac{1}{N_f^2} \ Var \ (Y_{12}) \right] + cN_f \left[\frac{1}{N_m^2} \ Var \ (Y_{21}) \right.$$

$$\left. \left. + \frac{2}{N_m N_f} \ Cov \ (Y_{21}, Y_{22}) + \frac{1}{N_f^2} \ Var \ (Y_{22}) \right] \right\} \qquad (10)$$

Now let k_{ij} denote the total number of successful gametes, contributing to offspring of sex j, that emanate from a parent of sex i. Let us suppose first that the locus is autosomal. Then the conditional distribution of Y_{ij}, given k_{ij}, is binomial with k_{ij} trials and a probability of success equal to 1/2.

Hence
$$Var \ (Y_{ij}) = E \ [\ Var \ (Y_{ij} \mid k_{ij}) \ ] + Var \ [\ E(Y_{ij} \mid k_{ij}) \ ]$$

$$= \frac{1}{4} \ [\ E(k_{ij}) + Var \ (k_{ij}) \ ] \qquad (11)$$

where $E \ (k_{ij}) = 2E \ (Y_{ij})$. In addition, Mendelian segregation takes place independently in gametes contributing to male and female offspring, so that

$$Cov \ (Y_{i1}, Y_{i2}) = E \ [Cov \ (Y_{i1}, Y_{i2} \mid k_{i1}, k_{i2})] + Cov \ [\ E \ (Y_{i1} \mid G_{i1}), E \ (Y_{i2} \mid E_{i2})]$$

$$= \frac{1}{4} \ Cov \ (k_{i1}, k_{i2}) \qquad (12)$$

Equations (7), (10), (11) and (12) imply that

$$\frac{1}{N_e} \approx \frac{1}{16 N_m} \left[\ Var \ (k_{11}) + 2 \left\{ \frac{N_m}{N_f} \right\} Cov \ (k_{11}, k_{12}) \right.$$

$$\left. + \left(\frac{N_m}{N_f} \right)^2 Var \ (k_{12}) + 2 \right] + \frac{1}{16 N_f} \left[\ Var \ (k_{22}) \right.$$

$$\left. + 2 \left(\frac{N_f}{N_m} \right) Cov \ (k_{21}, k_{22}) + \left(\frac{N_f}{N_m} \right)^2 Var \ (k_{21}) + 2 \right] \qquad (13)$$

This formula is the discrete generation special case of an expression obtained by Hill ([11], [12]).

When there is a sex-linked locus, equations (11) and (12) still hold for $i = 2$. But males of genotype AY produce no sons with that genotype when they mate with $\overline{AA}$ females, so that Y_{11} is identically equal to 0. Hence $Cov \ (Y_{11}, Y_{12}) = Var \ (Y_{11}) = 0$. In addition, all daughters of an $AY \times \overline{AA}$

mating are of genotype $A\overline{A}$ which implies that $\text{Var}(Y_{12}) = \text{Var}(k_{12})$. Now $c = 2$, so that (7), (10), (11) and (12) lead to

$$\frac{1}{N_e} \approx \frac{1}{9N_f}\left[\text{Var}(k_{22}) + 2\left(\frac{N_f}{N_m}\right)\text{Cov}(k_{21}, k_{22}) + \left(\frac{N_f}{N_m}\right)^2 \text{Var}(k_{21}) + 1 \right]$$

$$+ \frac{1}{9N_m}\left[2\left(\frac{N_m}{N_f}\right)^2 \text{Var}(k_{12}) + 1 \right] \tag{14}$$

This formula is the discrete generation special case of an expression obtained by Pollak ([20], [21]).

5. Monoecious Populations with Partial Selfing

Consider a monoecious population, of size N in every generation, in which there is self-fertilization with probability β_1 and the union of gametes from random separate individuals with probability $1 - \beta_1$. If there is selfing, an individual supplies both a male and a female gamete to a fertilized egg, whereas if there is random mating the average numbers of male and female gametes it contributes to separate offspring are each equal to one. Thus, each individual has, on the average, β_1 offspring from self-fertilization and $2(1 - \beta_1)$ from random mating.

Let types 1 and 2 be defined to be respectively the genotypes $A\overline{A}$ and AA and assume that initially all individuals except one have the genotype $\overline{A}\overline{A}$. The first moment matrix of the relevant branching process is then

$$\mathbf{M} = \begin{bmatrix} \dfrac{\beta_1}{2} + 1 - \beta_1 & \dfrac{\beta_1}{4} \\ 2(1 - \beta_1) & \beta_1 \end{bmatrix} = \begin{bmatrix} 1 - \dfrac{\beta_1}{2} & \dfrac{\beta_1}{4} \\ 2(1 - \beta_1) & \beta_1 \end{bmatrix}$$

Hence

$$\mathbf{p}' = \left(\frac{4(1 - \beta_1)}{4 - 3\beta_1}, \frac{\beta_1}{4 - 3\beta_1} \right)$$

and

$$\mathbf{v} = \left(\frac{4 - 3\beta_1}{2(2 - \beta_1)}, \frac{4 - 3\beta_1}{2 - \beta_1} \right)'$$

from which it follows that

$$\sum_{i=1}^{2} p_i \text{Var}\left(\sum_{j=1}^{2} Y_{ij} v_j \right)$$

$$= \frac{4 - 3\beta_1}{(4 - 2\beta_1)^2}\left[4(1 - \beta_1)\text{Var}(Y_{11} + 2Y_{12}) + \beta_1 \text{Var}(Y_{21} + 2Y_{22}) \right]$$

Note that $Y_{i1} + 2Y_{i2}$ is equal to k_i, the number of copies of allele A that an individual of type i contributes to offspring. If $i = 2$, this is equal to k, the total number of successful gametes produced by an individual. If $i = 1$, the conditional distribution of k_1, given k, is binomial with k trials and a probability 1/2 of success. Since $E(k) = 2$, it follows that

$$\text{Var}(Y_{11} + 2Y_{12}) = \text{Var}\left(\frac{k}{2}\right) + E\left(\frac{k}{4}\right)$$

In this case $K = 2N$ and (6) applies, so that

$$\frac{v_1}{N_e} \approx \frac{v_1}{2(2 - \beta_1)N} [\text{Var}(k) + 2(1 - \beta_1)] \tag{15}$$

Now if there is an infinite population with respective probabilities β_1 and $1 - \beta_1$ of reproduction by selfing and random mating, it was shown by Haldane [8] that the inbreeding coefficient F of a random individual is equal to $\beta_1 / (2-\beta_1)$ in the long run. Thus $1 + F = 2/(2-\beta_1)$, $1 - F = 2(1-\beta_1)/(2 - \beta_1)$, so that (15) can be recast as

$$\frac{1}{N_e} \approx \frac{1}{4N} [(1 + F)\text{Var}(k) + 2(1-F)] \tag{16}$$

which agrees with a result obtained by other methods by Caballero and Hill [1] and Pollak and Sabran [24].

6. *Dioecious Populations Reproducing Partly by Full-Sib Mating*

I assume that there are N/2 permanent couples in every generation, of which a fraction β_2 is expected to be full-sibs and a fraction $1 - \beta_2$ to consist of random male-female pairs. If the population is large and the allele A is initially rare, full-sibs mated to each other do not give rise to independently developing lineages, but, to a good approximation, we can assume that separate couples reproduce independently. Thus mating couples with at least one A in the genotype of at least one mate will be taken to be units in the approximating branching process.

Let us first assume that there is an autosomal locus. Then the parental mating types and the expected numbers of offspring mating types produced by them are listed in Table 1.

An example of how the entries in this table are generated is as follows. First, the expected number of full-sib mating offspring couples produced by a parental couple is one, whereas the expected number of progeny of a couple is two. Thus if there is random mating one couple produces, on the average,

Table 1. Expected numbers of matings of various types among offspring-autosomal locus

Parental mating type	Offspring given sibs do not mate	Mating types given full-sib mating
(1) $A\overline{A} \times \overline{A}\overline{A}$	(1)	$\frac{1}{2}(1) + \frac{1}{4}(3)$
(2) $AA \times \overline{A}\overline{A}$	2 (1)	(3)
(3) $A\overline{A} \times A\overline{A}$	$\frac{1}{2}(2) + (1)$	$\frac{1}{4}(1) + \frac{1}{8}(2) + \frac{1}{4}(3)$ $+ \frac{1}{4}(4) + \frac{1}{16}(5)$
(4) $AA \times A\overline{A}$	$(1) + (2)$	$\frac{1}{4}(3) + \frac{1}{2}(4) + \frac{1}{4}(5)$
(5) $AA \times AA$	2(2)	(5)

two offspring couples. Thus, if there is a parental mating of type (4), it produces the array of offspring $\frac{1}{2} A\overline{A} + \frac{1}{2} AA$, so that if there is random mating we expect one each of matings of types $A\overline{A} \times \overline{A}\overline{A}$ and $AA \times A\overline{A}$. In a shorthand summary notation this can be written as (1) + (2). If, on the other hand, sibs mate, we expect the array of offspring couples to be $\frac{1}{4}(3) + \frac{1}{2}(4) + \frac{1}{4}(5)$.

By using the entries in Table 1, we can calculate the elements of the first moment matrix. Thus we have

$$M = \begin{bmatrix} \dfrac{2-\beta_2}{2} & 0 & \dfrac{\beta_2}{4} & 0 & 0 \\ 2(1-\beta_2) & 0 & \beta_2 & 0 & 0 \\ \dfrac{4-3\beta_2}{4} & \dfrac{4-3\beta_2}{8} & \dfrac{\beta_2}{4} & \dfrac{\beta_2}{4} & \dfrac{\beta_2}{16} \\ 1-\beta_2 & 1-\beta_2 & \dfrac{\beta_2}{4} & \dfrac{\beta_2}{2_2} & \dfrac{\beta_2}{4} \\ 0 & 2(1-\beta_2) & 0 & 0 & \beta_2 \end{bmatrix}$$

It can then be shown that

$$v = v_1 (1, 2, 2, 3, 4)'$$

$$p' = p_3 \left(\frac{16 - 20\beta_2 + 5\beta_2^2}{2\beta_2 (2 - \beta_2)}, \frac{1}{2}, 1, \frac{\beta_2}{2(2-\beta_2)}, \frac{\beta_2(\beta_2 + 2)}{16(1-\beta_2)(2-\beta_2)} \right)$$

and $\qquad 1 = p'v = v_1 p_3 \dfrac{(4-3\beta_2)(4-\beta_2)}{4\beta_2(1-\beta_2)}$ $\qquad\qquad$ (17)

Thus we have

$$\sum_{i=1}^{5} p_i \, \text{Var} \left(\sum_{j=1}^{5} Y_{ij} \, v_j \right)$$

$$= \sum_{i=1}^{5} p_i \, \text{Var} \, [(Y_{i1} + 2(Y_{i2} + Y_{i3}) + 3Y_{i4} + 4Y_{i5}) \, v_1]$$

Now $\qquad Y_{i1} + 2(Y_{i2} + Y_{i3}) + 3Y_{i4} + 4Y_{i5}$

is equal to g_i, the number of copies of A that a couple of type i contributes to offspring. Let g be the total number of successful gametes contributed to offspring by one parental couple. Then g_5 is equal to g and, because all offspring of matings of type 2 are heterozygotes, $g_2 = g/2$.

Now let Bin(n, p) represent a binomial distribution with parameters n and p. Then, because of Mendelian segregation when only one of the mates in a couple is heterozygous, the conditional distributions of g_1, g_3 and g_4 are as follows :

$$g_1 \mid g \sim \text{Bin}\left(\frac{g}{2}, \frac{1}{2}\right), \qquad g_3 \mid g \sim \text{Bin}\left(g, \frac{1}{2}\right), \qquad g_4 \mid g \sim \frac{g}{2} + \text{Bin}\left(\frac{g}{2}, \frac{1}{2}\right)$$

Therefore

$$\text{Var}(g_1) = E\left(\frac{g}{8}\right) + \text{Var}\left(\frac{g}{4}\right) = \frac{1}{2} + \frac{1}{16}\,\text{Var}(g)$$

$$\text{Var}(g_2) = \frac{1}{4}\,\text{Var}(g)$$

$$\text{Var}(g_3) = E\left(\frac{g}{4}\right) + \text{Var}\left(\frac{g}{2}\right) = 1 + \frac{1}{4}\,\text{Var}(g)$$

$$\text{Var}(g_4) = E\left(\frac{g}{8}\right) + \text{Var}\left(\frac{3}{4}g\right) = \frac{1}{2} + \frac{9}{16}\,\text{Var}(g)$$

and $\text{Var}(g_5) = \text{Var}(g)$. Therefore

$$\sum_{i=1}^{5} p_i \, \text{Var}\left(\sum_{j=1}^{5} Y_{ij} v_j\right) = \frac{1}{16}\left(\sum_{i=1}^{5} p_i v_i^2\right) \text{Var}(g)$$

$$+ \frac{v_1^2 p_3}{4}\left[\frac{16 - 20\beta_2 + 5\beta_2^2}{\beta_2(2-\beta_2)} + 4 + \frac{\beta_2}{2-\beta_2}\right]$$

By virtue of (17) this expression can be shown to be equal to

$$\frac{v_1}{4\,(4-3\beta_2)}\Big[\,\text{Var}(g)+8(1-\beta_2)\Big]$$

However, Ghai [7] has shown that in an infinite population with partial full sib mating the inbreeding coefficient is equal to $\beta_2/(4-3\beta_2)$ in the long run. Hence $1+3F=4/(4-3\beta_2)$ and $1-F=4(1-\beta_2)/(4-3\beta_2)$. Consequently, (6) implies that

$$\frac{1}{N_e}\approx\frac{1}{16N}\,[\,(1+3F)\,\text{Var}\,(g)+32(1-F)]\tag{18}$$

But g is equal to twice k, the number of offspring of a couple. Therefore (18) can be rewritten as

$$\frac{1}{N_e}\approx\frac{1}{4N}\,[\,(1+3F)\,\text{Var}\,(k)+2(1+F)]\tag{19}$$

which is consistent with a general result obtained by Caballero and Hill [1].

Now let us consider a sex-linked locus. The parental mating types with allele A and the expected numbers of offspring mating types with that allele are listed in Table 2. It is, of course, assumed that a large population consists almost entirely of $\overline{AA}$ females and $\overline{A}Y$ males.

The notation in this table is the same as in Table 1, and the first moment matrix can now be obtained from its entries. It is equal to

$$M=\begin{bmatrix}\dfrac{2-\beta_2}{4} & \dfrac{2-\beta_2}{4} & 0 & \dfrac{\beta_2}{4} & 0\\[2mm] 1 & 0 & 0 & 0 & 0\\[2mm] 1-\beta_2 & 1-\beta_2 & 0 & \beta_2 & 0\\[2mm] \dfrac{2-\beta_2}{4} & \dfrac{1-\beta_2}{2} & \dfrac{2-\beta_2}{4} & \dfrac{\beta_2}{4} & \dfrac{\beta_2}{4}\\[2mm] 0 & 1-\beta_2 & 1-\beta_2 & 0 & \beta_2\end{bmatrix}$$

The left and right eigenvectors corresponding to $\rho=1$ are, respectively,

$$\mathbf{p}'=p_4\left(\frac{4-3\beta_2}{\beta_2}\,,\,\frac{4-3\beta_2}{2\beta_2}\,,\frac{1}{2}\,,\,1\,,\,\frac{\beta_2}{4(1-\beta_2)}\right)$$

Table 2. Expected numbers of matings of various types among offspring-sex-linked locus

Parental mating type	Offspring mating types	
	Given sibs do not mate	Given full-sib mating
(1) $A\overline{A} \times \overline{A}Y$	$\frac{1}{2}(1) + \frac{1}{2}(2)$	$\frac{1}{4}(1) + \frac{1}{4}(2) + \frac{1}{4}(4)$
(2) $\overline{AA} \times AY$	(1)	(1)
(3) $AA \times \overline{A}Y$	(1) + (2)	(4)
(4) $A\overline{A} \times AY$	$\frac{1}{2}(1) + \frac{1}{2}(2)$ $+ \frac{1}{2}(3)$	$\frac{1}{4}(1) + \frac{1}{4}(3)$ $+ \frac{1}{4}(4) + \frac{1}{4}(5)$
(5) $AA \times AY$	(2) + (3)	(5)

and
$$v = v_1 (1, 1, 2, 2, 3)' \tag{20}$$

From the equation $p' v = 1$ we then obtain

$$v_1 p_4 = \frac{4\beta_2 (1-\beta_2)}{3(2-\beta_2)(4-3\beta_2)} \tag{21}$$

Equation (20) implies that

$$\sum_{i=1}^{5} p_i \, \text{Var} \left(\sum_{j=1}^{5} Y_{ij} v_j \right) = \sum_{i=1}^{5} p_i \, \text{Var} (v_1 g_i)$$

where g_i is the number of gametes of type A that a couple of type i contributes to offspring couples. Now let k_f and k_m respectively denote the number of gametes with X-chromosomes contributed by females to sons and daughters and by males to daughters. Then, given k_f and k_m, the conditional distributions of g_1, g_2, g_3, g_4 and g_5 are as follows :

$$g_1 \mid k_f, k_m \sim \text{Bin}\left(k_f, \frac{1}{2}\right)$$
$$g_2 \mid k_f, k_m \sim k_m$$
$$g_3 \mid k_f, k_m \sim k_f$$
$$g_4 \mid k_f, k_m \sim k_m + \text{Bin}\left(k_f, \frac{1}{2}\right)$$

$$g_5 \mid k_f, k_m \sim k_f + k_m$$

It therefore follows that

$$\mathrm{Var}\,(g_1) = E\left(\frac{k_f}{4}\right) + \mathrm{Var}\left(\frac{k_f}{2}\right) = \frac{1}{2} + \frac{1}{4}\,\mathrm{Var}\,(k_f)$$

$$\mathrm{Var}\,(g_2) = \mathrm{Var}\,(k_m)$$

$$\mathrm{Var}\,(g_3) = \mathrm{Var}\,(k_f)$$

$$\mathrm{Var}\,(g_4) = E\left(\frac{k_f}{4}\right) + \mathrm{Var}\,(k_m + \frac{1}{2}\,k_f)$$

$$= \frac{1}{2} + \mathrm{Var}\,(k_m) + \frac{1}{4}\,\mathrm{Var}\,(k_f) + \mathrm{Cov}\,(k_m, k_f)$$

$$\mathrm{Var}\,(g_5) = \mathrm{Var}\,(k_m) + \mathrm{Var}\,(k_f) + 2\,\mathrm{Cov}\,(k_m, k_f)$$

Thus, by virtue of (6) and (21), we obtain

$$\frac{3Nv_1}{4N_e} \approx \frac{v_1}{3}\left[\frac{4\,(1-\beta_2)}{4-3\beta_2} + \frac{2-\beta_2}{4-3\beta_2}\,\mathrm{Var}\,(k_f) + \mathrm{Var}\,(k_m) + \frac{2\beta_2}{4-3\beta_2}\,\mathrm{Cov}\,(k_m, k_f)\right]$$

$$(22)$$

The approximate equation (22) can be expressed in terms of the inbreeding coefficient in the long run in an infinite population that reproduces partly by full-sib mating. When there is a sex-linked locus, the inbreeding coefficient of a female is the probability that its two separate copies of a gene are identical by descent, and the coancestry of a female and a male is the probability that a randomly chosen copy of a gene in the female is identical by descent to the male's single copy. If a male and a female are mated, the inbreeding coefficient of a random daughter from the mating is the coancestry of her parents. In addition, the coancestry of a male and a female is equal to the coancestry of the female with the mother of the male.

I shall now apply the theory summarized in the foregoing paragraph to full-sib mating. Thus, let x_n and Y_n be respectively a male and a female of generation n, and denote by F_{Y_n} and $\theta_{x_n Y_n}$ the inbreeding coefficient of Y_n and the coancestry of x_n and Y_n. Then, given that there is full-sib mating,

$$F_{Y_{n+2}} = \theta_{x_{n+1} Y_{n+1}} = \theta_{Y_n Y_{n+1}} = \frac{1}{2}\,\theta_{Y_n x_n} + \frac{1}{2}\,\theta_{Y_n Y_n}$$

$$= \frac{1}{2}\,F_{Y_{n+1}} + \frac{1}{4}\,(1 + F_{Y_n})$$

If there is not full-sib mating, $\theta_{x_{n+1} Y_{n+1}} = 0$. Consequently, if F_n is the inbreeding coefficient of a random female of generation n

$$F_{n+2} = \frac{\beta}{4} [1 + F_n + 2F_{n+1}]$$

At equilibrium, which is attained in the long run, we have F in place of F_n and this is equal to $\beta_2 / (4 - 3\beta_2)$, as it is for an autosomal locus. It now follows that $1 - F = 4(1-\beta_2) / (4-3\beta_2)$ and $1 + F = 2(2-\beta_2) / (4-3\beta_2)$. Thus, (22) can be rewritten as

$$\frac{1}{N_e} \approx \frac{2}{9N} [2(1-F) + (1+F) \text{ Var } (k_f) + 2 \text{ Var}(k_m) + 4F \text{ Cov}(k_f, k_m)] \quad (23)$$

Note that if $N_m = N_f = N/2$ and there is random mating, (23) reduces to the special case of (14) that applies when there are equal numbers of males and females, as it should.

7. General Results for Age-Structured Populations

To the best of my knowledge, theory for age-structured populations has until now only been developed for situations in which there is complete random mating. If a population is monoecious the limiting form of (16) when F tends toward 0 is

$$\frac{1}{N_e} \approx \frac{1}{4N} [\text{Var}(k) + 2]$$

which is a result obtained by Haldane [9], and essentially the same as an expression due to Wright [26] if N is large. Hill ([11], [12]) derived the generalization for an age-structured population observed at times 0, 1,. . . and having age classes 0, 1, . . . , T. It is

$$\frac{1}{N_e} \approx \frac{1}{4NL} [\text{Var}(k) + 2] \quad (24)$$

where L is the mean, over individuals in age group 0 in a population having a stable age distribution, of the ages of parents, N is the number of individuals in age group 0, and k is the number of offspring an individual produces in its lifetime. His reasoning differed from the approach taken in this paper, but Pollak [19] was able to obtain (24) in an alternative way by using an approximate expression analogous to (5).

When the population is dioecious it can be divided into age-sex classes (01), (11), ..., $(T_1 1)$, (02), (12), ..., $(T_2 2)$, where the first $T_1 + 1$ of these are associated with males and the remainder with females. Thus, for example, a female is said to be in age-sex class (x2) at time t if it is between x and x+1 units of age at that time. Individuals in these classes that have one A in their genotypes are the types in a branching process.

In what follows a subscript i or j will assume the value 1 if it refers to a male and 2 if it refers to a female. In terms of this notation L_{ij} will then denote the mean age of a parent of sex i when an offspring of sex j is born. The generation interval L is the unweighted average of the L_{ij} 's over all possible values of i and j, so that

$$L = \frac{1}{4} [L_{11} + L_{12} + L_{21} + L_{22}] \tag{25}$$

for an autosomal locus and

$$L = \frac{1}{3} [L_{12} + L_{21} + L_{22}] \tag{26}$$

for an x-linked locus.

If at times 0, 1, ... the numbers of males and females in age group 0 are respectively m and f it turns out that (13) generalizes to

$$\frac{1}{N_e} \approx \frac{1}{L} \left\{ \frac{1}{16m} \left[\text{Var}(k_{11}) + 2\left(\frac{m}{f}\right)\text{Cov}(k_{11}, k_{12}) + \left(\frac{m}{f}\right)^2 \text{Var}(k_{12}) + 2 \right] \right.$$
$$\left. + \frac{1}{16f} \left[\text{Var}(k_{22}) + 2\left(\frac{f}{m}\right)\text{Cov}(k_{21}, k_{22}) + \left(\frac{f}{m}\right)^2 \text{Var}(k_{21}) + 2 \right] \right\} \tag{27}$$

where L is given by (25) and k_{ij} is the mean number of offspring of sex j produced by a parent of sex i throughout its life. This is the formula derived by Hill ([11], [12]). If there is a sex-linked locus the expression corresponding to (14) is

$$\frac{1}{N_e} \approx \frac{1}{L} \left(\frac{1}{9f} \left[\text{Var}(k_{22}) + 2\left(\frac{f}{m}\right)\text{Cov}(k_{21}, k_{22}) + \left(\frac{f}{m}\right)^2 \text{Var}(k_{21}) + 1 \right] \right.$$
$$\left. + \frac{1}{9m} \left[2\left(\frac{m}{f}\right)^2 \text{Var}(k_{12}) + 1 \right] \right\} \tag{28}$$

where L is given by (26). This is the formula obtained by Pollak ([20], [21]), the first time by use of the methods discussed in this paper and the second

by adopting Hill's method. In the first of these papers there was also a derivation of (27) by means of the branching process approximation.

My earlier derivations were direct and lengthy. In a later paper (Pollak [22]) they were shortened by considering a discrete generation branching process corresponding to distributions of offspring males and females produce throughout their lives. However, the argument was completed by invoking an approximate expression for the survival probability of a slightly advantageous allele originally in a single ancestor of sex i and age 0. I would prefer to be able to do this by reasoning entirely in terms of neutral alleles, as in this paper, but have not yet been able to do so.

The results obtained by Pollak ([19], [20], [21]) indicate that the analogue and generalization of (5) that holds when there is, an age-structured random mating population is

$$\frac{K}{2N_eL} \approx \sum_{i=1}^{T} p_i \, Var(\sum_{j=1}^{T} Y_{ij}v_j) \tag{29}$$

8. Discussion

The right side of (5) appears in branching process theory in contexts other than those discussed in this paper. Thus, for example, if z'_t is the vector of numbers of individuals of various types at time t and $\rho = 1$,

$$\lim_{t\to\infty} tP\,[\,z'_t \neq 0' \,|\, z'_0 = e'_j\,] = \frac{v_j}{\sum_i p_i \sum_r \sum_s v_r v_s E\,[\,Y_{ir}Y_{is} - \delta_{rs}Y_{ir}\,]}$$

where $0' = [\,0, 0, \ldots, 0\,]$ and

$$\delta_{rs} = \begin{cases} 1, & r = s \\ 2, & r \neq s \end{cases}$$

provided that all of the second moments of the numbers of offspring at time 1 of parents at time 0 are finite. This was proved by Joffe and Spitzer [14], though Mullikin [18] had earlier derived the limiting form under the assumption that moments up to the third order are finite.

Because

$$E\,[\,Y_{ir}\,Y_{is} - \delta_{rs}\,Y_{ir}\,] = Cov\,(Y_{ir},\, Y_{is}\,) - \delta_{rs}\,m_{ir} + m_{ir}m_{is}$$

where m_{ir} is the element in the ith row and rth column of **M**, it follows from (1) that

$$\sum_i p_i \sum_r \sum_s v_r v_s \, E[\, Y_{ir} Y_{is} - \delta_{rs} Y_{ir}\,]$$

$$= \sum_i p_i \sum_r \sum_s v_r v_s \text{Cov}(Y_{ir}, Y_{is}) - \sum_r p_r v_r^2 + \sum_i p_i v_i^2$$

Hence

$$\lim_{t \to \infty} t\, P[\, z'_n \neq 0' \mid z'_o = e'_j\,] = \frac{v_j}{\sum_i p_i \, \text{Var}(\sum_r Y_{ir} v_r)} \qquad (30)$$

The right side of (5) also plays a role in the calculation of an approximation to the survival probability of a line descended from a single ancestor of type i if ρ is slightly larger than 1.

In this case Hoppe [13] proved that

$$\lim_{t \to \infty} P[\, z'_t \neq 0' \mid z'_o = e'_j\,]$$

$$= \frac{2(\rho - 1)}{\sum_i p_i \, \text{Var}(\sum_r Y_{ir} v_r)} v_j + o(\rho - 1) \qquad (31)$$

if ρ is a function of a small parameter ε and ρ and the extinction probabilities approach 1 as ε approaches 0. It is also assumed that all the probability generating functions and their first derivatives converge uniformly as $\varepsilon \to 0$ for all values of their arguments that are between 0 and 1. Eshel [5] and Pollak [22] had already made use of (31) in special cases that arise in genetics when one wishes to compute the survival probability of a slightly advantageous allele.

Note that, to a good approximation, (5) and (31) imply that

$$\lim_{t \to \infty} P[\, z'_t \neq 0' \mid z'_o = e'_1\,] = \frac{4N_e}{K}(\rho - 1)v_1 \qquad (32)$$

whereas (6) and (31) analogously lead to

$$\lim_{t \to \infty} P[\, z'_t \neq 0' \mid z'_o = e'_1\,] = \frac{4N_e}{K}(\rho - 1) \qquad (33)$$

Expressions (32) and (33) generalize a result obtained by Kimura [15].

In my previous work, I have already used special cases of (5) (see Pollak [22]). My first encounters with theory for nonrandom mating populations led to errors, but these were corrected and correct expressions involving special cases of (6) were given by Pollak and Sabran [24] and Pollak [23]. However, the complete derivations were not given because there were no errors in the original matrix calculations, to which references were made, and these papers were mainly on other properties of nonrandom mating populations.

Two purposes of this paper have been to demonstrate in general why (5) and (6) work for diploid populations with discrete generations, and to present self-contained and complete treatments of various special cases on which I have previously written. In my earlier works the right sides of (5) and (6) appeared as by-products of calculations concerning other properties of finite populations than effective population sizes per se. Because of this, and the fact that units in the branching process were not necessarily the same as individuals, it was not immediately obvious that (6) holds.

I have not previously calculated N_e for the case of partial full-sib mating, with $N/2$ permanent couples and a sex-linked locus. In a paper that has just been published, Wang [25] has shown that in such a population

$$\frac{1}{N_e} = \frac{1}{9N} \; [\; 4(1-\alpha) + 2(1+3\alpha) \; \sigma^2 + 4(1+\alpha) \; \sigma_f^2 - 4\alpha\sigma_m^2 \;]$$

In this expression $\alpha = F$, σ_m^2 and σ_f^2 are respectively the variances of the numbers of male and female progeny of a family, and $\sigma^2 = \sigma_m^2 + \sigma_f^2 + 2\theta$, where θ is the covariance between the numbers of male and female progeny. Noting that k_f and k_m are respectively equal to Wang's $m+f$ and f, it is clear that (23) can be rewritten as

$$\frac{1}{N_e} = \frac{1}{9N} \; [\; 4(1-F) \; + \; 2(1+F) \; \sigma^2 + 4\sigma_f^2 + 4F(\sigma_f^2 + 2\theta + \sigma_m^2 + \sigma_f^2 - \sigma_m^2)]$$

$$= \frac{1}{9N} \; [\; 4(1-F) + 2(1+3F)\sigma^2 + 4(1+F)\sigma_f^2 - 4F\sigma_m^2]$$

Hence the two results are equivalent.

I have also shown how some of my previous work on age-structured populations involves (29), which generalizes (5). I have not yet been able to generalize (6), but if such a generalization were shown to hold, it could prove to be useful.

Finally, a connection has been shown between the ratio of K to $2N_e$ and equation (31) for the probability of the long-term survival of allele A in a

lineage descended from one ancestor with a single copy of A. Thus, at least if one can formulate the problem of calculating survival probabilities in terms of a multitype branching process approximation, (32) or (33) can be used, with N_e / K being calculated either by the methods used in this paper or, say, the method perfected by Caballero and Hill [1].

REFERENCES

[1] Caballero, A. and Hill, W. G., 1992. Effective size of nonrandom mating populations. *Genetics*, **130**, 909-916.

[2] Crow, J. F., 1954. Breeding structure of populations. II. Effective population number. In : *Statistics and Mathematics in Biology* (O. Kempthorne, T. A. Bancroft, J. W. Gowen, and J. L. Lush eds.), 543-556, Iowa State College Press, Ames.

[3] Crow, J. F. and Denniston, C., 1988. Inbreeding and variance effective population numbers. *Evolution*, **42**, 482-495.

[4] Crow, J. F. and Morton, N., 1955. Measurement of gene frequency in small populations. *Evolution*, **9**, 202-214.

[5] Eshel, I., 1984. On the survival probability of a slightly advantageous mutant gene in a multitype population. *J. Math. Biol.*, **12**, 201-209.

[6] Ethier, S. N. and Nagylaki, T., 1980. Diffusion approximations to Markov chains with two time scales and applications to population genetics. *Adv. Appl. Prob.*, **12**, 14 - 49.

[7] Ghai, G. L., 1969. Structure of populations under mixed random and sib mating. *Theor. Appl. Genet.*, **39**, 179-182.

[8] Haldane, J. B. S., 1924. A mathematical theory of natural and artificial selection. Part II. The influence of partial self-fertilisation, inbreeding, assortative mating and selective fertilisation on the composition of Mendelian populations and on natural selection. *Proc. Camb. Phil. Soc. Biol. Sci.*, **1**, 158-163.

[9] Haldane, J. B. S., 1939. The equilibrium between mutation and random extinction. *Ann. Eugen.*, **9**, 400-405.

[10] Harris, T. E., 1963. *The Theory of Branching Processes.* Springer-Verlag, Berlin.

[11] Hill, W. G., 1972. Effective size of populations with overlapping generations. *Theor. Popul. Biol.*, **3**, 278-289.

[12] Hill, W. G., 1979. A note on effective population size with overlapping generations. *Genetics*, **92**, 317-322.

[13] Hoppe, F. M., 1992. The survival probability of a mutant in a multidimensional population. *J. Math. Biol.*, **30**, 567-575.

[14]   Joffe, A. and Spitzer, F., 1967. On multitype branching processes with $\rho \leq 1$. *J. Math. Anal. & Appl.*, **19**, 409-430.

[15]   Kimura, M., 1964. Diffusion models in population genetics. *J. Appl. Probab.*, **1**, 177-232.

[16]   Kimura, M. and Crow, J. F., 1963. The measurement of effective population number. *Evolution*, **17**, 279-288.

[17]   Latter, B. D. H., 1959. Genetic sampling in a random mating population of constant size and sex ratio. *Aust. J. Biol. Sci.*, **12**, 500-505.

[18]   Mullikin, T. W., 1963. Limiting distributions for critical multitype branching processes with discrete time. *Trans. Am. Math. Soc.*, **106**, 469-494.

[19]   Pollak, E., 1979. Effective population numbers and mean times to extinction in monoecious populations with overlapping generations. *Math. Biosci.*, **46**, 87-106.

[20]   Pollak, E., 1980. Effective population numbers and mean times to extinction in dioecious populations with overlapping generations. *Math. Biosci.*, **52**, 1-25.

[21]   Pollak, E., 1990. The effective size of an age-structured population with a sex-linked locus. *Math. Biosci.*, **101**, 121-130.

[22]   Pollak, E., 1992. Survival probabilities for some multitype branching processes in genetics. *J. Math. Biol.*, **30**, 583-596.

[23]   Pollak, E., 1995. Some effects of selection when there is partial full-sib mating. *Genetics*, **139**, 439-444.

[24]   Pollak, E. and Sabran, M., 1992. On the theory of partially inbreeding finite populations. III. Fixation probabilities under partial selfing when heterozygotes are intermediate in viability. *Genetics*, **131**, 979-985.

[25]   Wang, J., 1996. Inbreeding coefficient and effective size for X-linked locus in nonrandom mating populations. *Heredity*, **76**, 569-577.

[26]   Wright, S., 1939. Statistical genetics in relation to evolution. In : *Actualities scientifiques et industrielles*, no. 802, 5-64. Exposes de-Biometrie et de la statistique biologique. XIII. Hermann et Cie. Paris.