

INBREEDING AND VARIANCE EFFECTIVE POPULATION NUMBERS

JAMES F. CROW AND CARTER DENNISTON

Genetics Department, University of Wisconsin, Madison, WI 53706

**Abstract.**—In this paper, a correction and extension of earlier work, we derive expressions for the inbreeding effective number,  $N_{ci}$ , and the variance effective number,  $N_{ev}$ , with various models. Diploidy, random mating, and discrete generations are assumed and formulas for  $N_{ci}$  are given for six situations: isogamous monoecious populations with self-fertilization permitted or excluded; monoecious populations, male and female gametes distinguished, with self-fertilization permitted or excluded; and separate sexes with or without male and female progeny distinguished.  $N_{ev}$  is given for monoecious and separate-sexed populations.

Most higher animals have separate sexes, and male and female progeny are distinguished. Letting the subscript  $t$  represent the generation in which the inbreeding effect is manifest, the inbreeding effective number is

$$\frac{1}{N_{ci}} = \frac{\sigma_{mm,mt} + \mu_{mm}\mu_{mf}}{4N_{m,t-2}\mu_{mm}\mu_{mf}} + \frac{\sigma_{ff,ft} + \mu_{ff}\mu_{ft}}{4N_{f,t-2}\mu_{ff}\mu_{ft}}$$

in which  $\sigma_{m,t}$  is the covariance of the number of male (m) and female (f) progeny of a parent of sex  $s$  ( $s = m$  or  $f$ ),  $\mu_{sm}$  and  $\mu_{sf}$  are the mean numbers of sons and daughters of a parent of sex  $s$ , and  $N_{s,t-2}$  is the number in the grandparental generation.

For the variance effective number

$$\frac{1}{N_{ev}} \approx \frac{1}{4N_{evm}} + \frac{1}{4N_{evf}}$$

in which

$$\frac{1}{N_{evs}} = \frac{1}{2(2N_{s,t-1} - 1)} \left[ \frac{1}{\mu_{sm}} + \frac{1}{\mu_{sf}} + \frac{\sigma_{sm}^2}{\mu_{sm}^2} + \frac{2\sigma_{sm,sf}}{\mu_{sm}\mu_{sf}} + \frac{\sigma_{sf}^2}{\mu_{sf}^2} \right]$$

and  $\sigma_{sm}^2$  and  $\sigma_{sf}^2$  are the variances in the number of sons and daughters of a parent of sex  $s$ . Observations are made at the same age in each generation.

Received December 1, 1986. Accepted December 1, 1987

The effective population number is one of the most widely used devices in population biology. The census number hardly ever reveals accurately the effects of inbreeding and gene-frequency drift. The effective population number is a surrogate population number that correctly reflects these effects. There has been increased interest in effective population size because of its use in the study of rare and endangered species.

The effective number of a natural population is almost always less than the number of adults of reproducing age and may be considerably less, usually for one or more of three reasons: 1) unequal numbers of males and females, 2) temporal variation in population number, and 3) greater than binomial or Poisson variability in the number of progeny per parent. This paper is mainly concerned with the third reason. It is a correction and extension of formulas summarized in Crow and Kimura (1970 pp. 345–

364). We consider only diploid populations with discrete generations.

History

The concept of effective population number was introduced by Wright (1931). In his words, the mathematical theory “has postulated a population of  $N$  diploid individuals, reconstructed each generation from a random sample of  $2N$  gametes. It is important to consider deviations from this ideal situation” (Wright, 1939 p. 20). His effective population number,  $N_e$ , is the size of such an ideal population that has the same rate of increase in homozygosity or gene-frequency drift as the actual population under consideration. Wright dealt with the situation of nonrandom variability in progeny number in the following way (Wright, 1939 p. 21):

Assume that  $N$  individuals furnish varying numbers ( $k$ ) of gametes to the next generation, but a total number of  $2N$ .

The mean number per individual is thus  $\bar{k} = 2$ . There is a variance

$$\sigma_k^2 = \frac{\sum (k - 2)^2}{N}$$

The proportion of cases in which two random gametes come from the same parent is

$$\frac{\sum k(k - 1)}{2N(2N - 1)} = \frac{2 + \sigma_k^2}{4N - 2}$$

This should be compared with the corresponding proportion among pairs of gametes drawn from an indefinitely large number to which all of the  $N$  parents contribute equally. This is obviously  $1/N$ . The effective size of population may thus be written  $(4N - 2)/(2 + \sigma_k^2)$ .

(Wright [1931] and Wright [1939] are reprinted in Provine [1986].)

On consideration of this problem, it became apparent that there are various kinds of  $N_e$ , and that these can sometimes lead to quite different expressions (Crow, 1954).

1) Wright’s procedure is based on the probability of homozygosity because of common ancestry, and hence was called the “inbreeding effective number.” 2) The amount of allele-frequency drift per generation, as measured by its variance, could be used to define a “variance effective number.” 3) One could consider the asymptotic rate of decay of segregating loci, which, when a result of Haldane (1939) was used, led to what was called the effective number for random extinction. Since this is determined by the largest nonunit eigenvalue of the transition matrix, Ewens (1979, 1982) has called it the “eigenvalue effective number.”

This article deals with only the first two kinds of  $N_e$ , as these are the ones most sought in demographic studies of actual populations.  $N_{ci}$ , the inbreeding effective number, is appropriate when the question of interest is the increase of homozygosity due to random drift.  $N_{ev}$ , the variance effective number, is appropriate when the question is the amount of gene-frequency drift or the increase in variance among subgroups.

The concept was developed more fully in later papers (Crow and Morton, 1955; Ki-

mura and Crow, 1963). The formulas for  $N_{ev}$  in these papers are incorrect for a population with separate sexes, because the covariance in the number of male and female progeny of a parent was neglected. This covariance was first taken into account by Latter (1959). Hill (1972, 1979) and Pollak (1980) extended Latter’s results to overlapping generations. Both Latter and Hill dealt with populations of constant size. The problem has also been studied by Moran and Watterson (1959), Chia and Pollak (1974), Pollak (1977), Johnson (1977), Choy and Weir (1978), Emigh and Pollak (1979), Ethier and Nagylaki (1980), and Nagylaki (1981).

Ewens (1979, 1982) has assumed the standard model in which there is an infinite number of populations, each subject to the same process of gamete sampling in every generation. The effective population number then grows naturally out of population-genetics theory. Others (e.g., Pollak, 1977; Nagylaki, 1981) have taken the same approach. Ewens (1982) points out that the eigenvalue effective number can be estimated by tracing the progress of a quantity such as heterozygosity.

This is quite proper for theoretical developments, but we have a different objective. We are seeking practical estimates of effective numbers for species when census data are available, but gene frequency and heterozygosity changes are not known. For this purpose, we take the demographic data as given and ask for the change in homozygosity or gene-frequency variance that would happen in this population at this time. It turns out that the relevant demographic data are census numbers and the means, variances, and covariances of the number of progeny per parent.

Throughout this paper, we assume that the discrete parent and offspring generations are censused at the same stage, conveniently just before reproduction with only those that survive to this stage being counted, although the procedure does not require this. For a discussion of the appropriate stage for enumeration, see Fisher (1939). Crow and Morton (1955) have given formulas to correct the data when offspring are counted at an earlier age than the parents. We shall use the subscript  $t$  to designate the generation

in which the inbreeding or drift effect is manifest.

In addition to discrete generations, the other basic assumptions are random mating, no selection, and no correlation between the fertility of a parent and that of its offspring. Crow and Morton (1955) treated the  $k_i$ 's as random variables. In this paper, we follow the pattern of Kimura and Crow (1963), who regarded them as constants. The population demography was taken from census data, and the question asked was: what is the effective population number of a population with these demographic parameters? In order that this article be self-contained, we have repeated and rederived some formulas from earlier papers.

#### Inbreeding Effective Number

Throughout this paper we ignore problems of estimating the relevant demographic parameters, and treat their values as given—in particular, the number of adults in generation  $t - 1$  and the mean and variance of the number of gametes (progeny) contributed by them to generation  $t$ . We assume that these gametes unite at random to produce generation  $t$ . We then ask for the size of an ideal population that would have the same probability of allelic identity in generation  $t$ .

*Isogamous Monoecious Diploids, Self-Fertilization Permitted.*—We first assume, unrealistically, that there is a random amount of self-fertilization; i.e., self-fertilization occurs at the same rate as if pairs were drawn at random from the gametic pool. Following Wright (1939), we ignore any distinction between egg and sperm, but we remove his restriction that the population be of constant size.

In an ideal population, the probability that two uniting gametes came from the same parent is  $1/N_{t-1}$ , where  $N_{t-1}$  is the number of individuals in the parent generation. Hence, we define  $N_{ei}$  by equating the actual probability to  $1/N_{ei}$ . Letting  $k_i$  be the number of gametes contributed by the  $i$ th parent, the number of pairs of gametes that came from the same parent is  $\sum_i k_i(k_i - 1)/2$ , and the total number of pairs is  $N_{t-1}\mu_k(N_{t-1}\mu_k - 1)/2$ , where  $\mu_k$  is the mean number of gametes produced per individ-

ual. The proportion of cases in which two random gametes uniting to produce generation  $t$  are derived from the same parent in generation  $t - 1$  is the ratio of these, or

$$\frac{\sum_{i=1}^{N_{t-1}} k_i(k_i - 1)}{N_{t-1}\mu_k(N_{t-1}\mu_k - 1)}$$

If the  $i$ th "parent" has no progeny,  $k_i = 0$ .  $\sum k_i = N_{t-1}\mu_k$  and  $\sum k_i^2 = N_{t-1}(\sigma_k^2 + \mu_k^2)$ , where  $\mu_k$  and  $\sigma_k^2$  are the mean and variance of the number of gametes per parent. Again,  $\mu_k$  and  $\sigma_k^2$  are simply functions of the set of fixed  $k_i$  and are not to be thought of as parameters of a probability distribution. Substituting these into the above expression and equating to  $1/N_{ei}$  yields a formula for the inbreeding effective number,  $N_{ei}$ ,

$$\frac{1}{N_{ei}} = \frac{\mu_k - 1 + \frac{\sigma_k^2}{\mu_k}}{N_{t-1}\mu_k - 1} \quad (1)$$

(Crow, 1954). The heterozygosity in generation  $t$ ,  $H_t$ , is given by

$$H_t = \left(1 - \frac{1}{2N_{ei}}\right)H_{t-1}$$

For a population of constant size ( $\mu_k = 2$ ),  $N_{ei}$  reduces to Wright's (1939) formula given above,  $(4N - 2)/(2 + \sigma_k^2)$ . If each of the  $2N_t$  gametes is drawn randomly from a pool to which each parent has contributed equally, the expected distribution of progeny numbers is binomial,  $Bn(1/N_{t-1}; 2N_t)$ . If the observed progeny mean and variance correspond to this distribution,  $\mu_k = 2N_t/N_{t-1}$ ,  $\sigma_k^2 = 2N_t(1/N_{t-1})(1 - 1/N_{t-1})$ ,  $\sigma_k^2/\mu_k = (N_{t-1} - 1)/N_{t-1}$ , and the effective number is the same as the number of parents,  $N_{t-1}$ , as would be expected from Wright's concept of effective population number.

The ratio  $\sigma_k^2/\mu_k$  is an index of the direction and amount by which the effective number deviates from the actual number. When the index is greater than 1 (more accurately,  $>1 - 1/N_{t-1}$ ), the effective number is smaller than the actual, and vice versa.

*Isogamous Monoecious Diploids, Self-Fertilization Prohibited.*—Without self-fer-

tilization, common ancestry for an individual cannot be closer than the grandparental generation,  $t - 2$ . Each individual in generation  $t - 2$  contributes  $k_i$  gametes to generation  $t - 1$ . Therefore the number of pairs of genes that came from the same grandparent is  $\sum_i k_i(k_i - 1)/2$ . There are

$N_{t-2}\mu_k/2$  individuals in generation  $t - 1$ . The number of pairs of individuals is  $\{(N_{t-2}\mu_k/2)[(N_{t-2}\mu_k/2) - 1]\}/2$ , and the number of pairs of genes in different individuals is four times this number, or  $N_{t-2}\mu_k(N_{t-2}\mu_k - 2)/2$ . Thus, the probability of drawing the same grandparental allele from a random sample of different individuals in generation  $t$  is

$$\frac{\sum_{i=1}^{N_{t-2}} k_i(k_i - 1)}{N_{t-2}\mu_k(N_{t-2}\mu_k - 2)}$$

The key assumption in this argument is that the number of progeny produced by an individual is independent of the number produced by its parent. Therefore, the probability of drawing from two individuals in generation  $t - 1$  a pair of gametes that are descended from a single individual in generation  $t - 2$  is not a function of the  $k_i$  in generation  $t - 1$ .

Making the same substitutions as before, we obtain

$$\frac{1}{N_{ei}} = \frac{\mu_k - 1 + \frac{\sigma_k^2}{\mu_k}}{N_{t-2}\mu_k - 2} \quad (2)$$

To compute the increase in the inbreeding coefficient we must consider two generations and apply the standard formula

$$f_t = \left(\frac{1}{2N_{ei}}\right)(1 + f_{t-2}) + \left(1 - \frac{1}{N_{ei}}\right)f_{t-1} \quad (3A)$$

or, letting  $H$  stand for heterozygosity,

$$H_t \approx \left[1 - \frac{1}{(2N_{ei} + 1)}\right]H_{t-1} \quad (3B)$$

*Male and Female Gametes Distinguished, Self-Fertilization Prohibited.*—It is quite

likely that the distributions of successful female and male gametes are different; in many species, especially wind- or insect-pollinated plants, the difference may be great. Let  $k_{ei}$  and  $k_{pi}$  be the numbers of eggs and pollen grains (or sperms) contributed to the next generation by the  $i$ th parent. Then, with random mating the probability that uniting male and female gametes come from the same parent is

$$\frac{\sum_{i=1}^{N_{t-1}} k_{ei}k_{pi}}{\sum_{i=1}^{N_{t-1}} k_{ei} \sum_{i=1}^{N_{t-1}} k_{pi}}$$

Noting that  $\sum_i k_{ei}k_{pi} = N_{t-1}(\sigma_{ep} + \mu_e\mu_p)$ , in

which  $\mu_e$  and  $\mu_p$  are the mean numbers of eggs and pollen grains per parent and  $\sigma_{ep}$  is the observed covariance of the number of eggs and pollen grains from a parent and equating the expression above to  $1/N_{ei}$  gives

$$\frac{1}{N_{ei}} = \frac{\mu_k + \frac{\sigma_{ep}}{\mu_k}}{N_{t-1}\mu_k} \quad (4)$$

in which  $\mu_k = \mu_e = \mu_p$ .

*Male and Female Gametes Distinguished, Self-Fertilization Prohibited.*—In this case, again assuming independence of individual fertility in successive generations, two gametes uniting to produce generation  $t$  have come from two eggs, egg and pollen, pollen and egg, or two pollen grains from generation  $t - 2$  with equal probability. Thus, using the same procedure as was used in deriving (2), the probability that two gametes uniting to produce a zygote in generation  $t$  came from the same grandparent in generation  $t - 2$  is

$$\frac{\sum_{i=1}^{N_{t-2}} (k_{ei} + k_{pi})(k_{ei} + k_{pi} - 1)}{\sum_{i=1}^{N_{t-2}} (k_{ei} + k_{pi}) \left[ \sum_{i=1}^{N_{t-2}} (k_{ei} + k_{pi}) - 2 \right]}$$

But,  $k_i = k_{ei} + k_{pi}$  and  $\sum_i (k_{ei} + k_{pi}) = N_{t-2}\mu_k$ , so using the procedure for deriving (2), we have

$$\frac{1}{N_{et}} = \frac{\mu_k - 1 + \frac{\sigma_k^2}{\mu_k}}{N_{t-2}\mu_k - 2} \quad (2')$$

Notice that this is the same equation as (2) above; if self-fertilization is prohibited, it does not matter whether male and female gametes are distinguished or not.

*Separate Sexes, Progeny Identified as to Sex.*—If there are separate sexes, we can note that the probability of two gametes uniting to form generation  $t$  both coming from male grandparents in generation  $t - 2$  is  $1/4$  and likewise for female grandparents. The probability of both coming from the same male grandparent is  $1/4N_{elm}$ , and the probability of both coming from the same female is  $1/4N_{elf}$ , in which  $N_{elm}$  and  $N_{elf}$  stand for the inbreeding effective numbers of males and females. The probability of both gametes coming from the same individual, regardless of sex, is the reciprocal of the inbreeding effective number, and is

$$\frac{1}{N_{et}} = \frac{1}{4N_{elm}} + \frac{1}{4N_{elf}} \quad (5)$$

To determine  $N_{elm}$  and  $N_{elf}$  we must consider both male and female progeny of each parent. To do this, let

- $N_{s,t-2}$  = the number of adults of sex  $s$  ( $s = m$  or  $f$ ) in generation  $t - 2$ , and
- $k_{smis}, k_{sfi}$  = the number of male (female) progeny from the  $i$ th parent of sex  $s$  in generation  $t - 2$ .

Still assuming that the fertility of a parent in generation  $t - 2$  and that of its progeny in generation  $t - 1$  are independent, the conditional probability,  $P = 1/N_{els}$ , that two gametes uniting to form a zygote in generation  $t$  and coming from an individual of sex  $s$  came from the same individual of sex  $s$  in generation  $t - 2$  is

$$\frac{1}{N_{els}} = \frac{\sum_i k_{smi}k_{sfi}}{\sum_i k_{smi} \sum_i k_{sfi}} = \frac{\sigma_{sm,sf} + \mu_{sm}\mu_{sf}}{N_{s,t-2}\mu_{sm}\mu_{sf}} \quad (6)$$

where  $\mu_{sm}$  and  $\mu_{sf}$  are the mean numbers of male and female progeny from a grandparent of sex  $s$  and  $\sigma_{sm,sf}$  is the observed covariance. Substituting (6) separately for each sex into (5) gives the inbreeding effective number.

*Separate Sexes, Progeny Not Identified by Sex.*—In this case, if we make the reasonable assumption that whether two individuals in generation  $t - 1$  are both male, both female, or one of each is independent of the probability of their having common ancestry in generation  $t - 2$ , we can use (1) separately for each parental sex. Thus

$$\frac{1}{N_{els}} = \frac{\mu_s - 1 + \frac{\sigma_s^2}{\mu_s}}{N_{s,t-2}\mu_s - 1} \quad (7)$$

where  $\mu_s$  and  $\sigma_s^2$  are the mean and variance of the number of gametes produced by individuals of sex  $s$  ( $s = m$  or  $f$ ). As before, substituting (7) separately for each grandparental sex in (5) gives an expression for obtaining  $N_{et}$ . For large and constant  $N_m$  and  $N_f$ , the numbers of males and females, this is the same as given by Moran and Waterson (1959). Nagylaki (1981) has provided analogous formulas for an X-linked locus.

Equations (5) and (7) can be put together to give

$$\frac{1}{N_{et}} = \frac{\mu_k - 1 + \frac{\sigma_k^2}{\mu_k}}{N_{t-2}\mu_k - 2} \quad (2'')$$

in which  $\mu_k = 2m\mu_m = 2f\mu_f$ , where  $m$  and  $f$  stand for the proportions of males and females, and  $\sigma_k^2 = m\sigma_m^2 + f\sigma_f^2 + mf(\mu_m - \mu_f)^2$  (Kimura and Crow, 1963). Again, as perhaps expected, we get Equation (2). The effective number for species with separate sexes is the same as that for monoecious species in which self-fertilization is prohibited, regardless of whether or not male and female gamete numbers have the same distribution.

When the variance in progeny of males and females are those expected from a binomial distribution [ $\sigma_s^2 = (N_s - 1)/N_s$ ], substituting (7) into (5) gives Wright's familiar formula

$$\frac{1}{N_e} = \frac{1}{4N_m} + \frac{1}{4N_f} \quad (8)$$

Of course, if data on the sex of individual offspring is available, this information can be used in Equations (5) and (6) to obtain an estimate involving fewer assumptions.

*Variance Effective Number*

For the variance effective number, we ask for the size of an ideal population that has the same amount of random gene-frequency drift as the observed population. Random allele-frequency drift causes an increase in between-group variance. In an ideal population of size  $N$ , the between-group variance of allele frequency,  $p$ , in generation  $t$  is

$$V_t = \left(1 - \frac{1}{2N}\right)V_{t-1} + \frac{p(1-p)}{2N} \quad (9)$$

We designate by  $\Delta p$  the change in allele frequency in one generation. The variance effective number,  $N_{ev}$ , is then defined by equating  $E\{(\Delta p)^2\}$  to its value in an ideal population,  $p(1-p)/2N_{ev}$  (Crow, 1954). Thus,

$$N_{ev} = \frac{2E\{(\Delta p)^2\}}{p(1-p)} \quad (10)$$

We assume that the number of parents and the number of gametes contributed by each parent are known, but their genotypes are not. We ask for the variance of allele-frequency drift if we sample completely from the parent population with specified numbers of gametes contributed by each parent but with the genotypes allocated among the parents at random. Gene-frequency drift will occur because of both differential reproduction of different genotypes and random segregation in heterozygotes. Again, generations are discrete.

*Isogamous Monoecious Diploids.*—As in the derivation of the inbreeding effective number for this case, we make no distinction between eggs and sperms and assume a random amount of self-fertilization. An alternative derivation was given by Pollak (1977).

The allele of interest is designated by  $A_1$ , and its frequency in the parent generation is  $p = P_{11} + P_{1x}$ , where  $P_{11}$  is the proportion of  $A_1A_1$  and  $2P_{1x}$  is the proportion of  $A_1A_x$ . We focus our attention on  $A_1$ ;  $A_x$  stands for all the other alleles at this locus. Let  $n_{11} = N_{t-1}P_{11}$  be the number of  $A_1A_1$  individuals

in the parent generation,  $t - 1$ , and  $n_{1x} = 2N_{t-1}P_{1x}$  be the number of heterozygotes in which one allele is  $A_1$ .  $N_{t-1}$  is the total number of individuals in the parent generation, most conveniently (but not necessarily) enumerated at the beginning of the reproductive period. Let  $k_i$  be the number of gametes contributed by the  $i$ th parent.

To facilitate the derivation, we define three random variables,  $l$ ,  $a$ , and  $b$ ;  $l_i$  is binomially distributed;  $a$  and  $b$  are characteristic functions for  $A_1$  homozygotes and heterozygotes, respectively; i.e.,

$$l_i \sim \text{Bn}\left(\frac{1}{2}, k_i\right)$$

$$a_i = \begin{cases} 1, & \text{if } i \text{ is } A_1A_1 \\ 0, & \text{otherwise} \end{cases}$$

$$b_i = \begin{cases} 1, & \text{if } i \text{ is } A_1A_x \\ 0, & \text{otherwise} \end{cases}$$

Then,  $\sum_i a_i = n_{11}$ ,  $\sum_i b_i = n_{1x}$ ,  $E l_i = k_i/2$  and  $E(l_i - k_i/2)^2 = k_i/4$ .

Assume  $n_{11}A_1A_1$  and  $n_{1x}A_1A_x$  individuals in the parent generation and random assignment of genotype to each individual, and (for momentary convenience) let  $N = N_{t-1}$ . In what follows, all expectations involving  $a$  and  $b$  refer to a sample space in which the  $N$  genotypes (of which  $n_{11}$  are  $A_1A_1$ ,  $n_{1x}$  are  $A_1A_x$ , and the rest are  $A_xA_x$ ) are assigned to the  $Nk$ 's in all possible ways, each equally likely. Then

$$E(a_i) = E(a_i^2) = \frac{n_{11}}{N}$$

$$E(b_i) = E(b_i^2) = \frac{n_{1x}}{N}$$

$$E(a_i b_i) = 0$$

and for  $i \neq j$ ,

$$E(a_i a_j) = \frac{n_{11}(n_{11} - 1)}{N(N - 1)}$$

$$E(b_i b_j) = \frac{n_{1x}(n_{1x} - 1)}{N(N - 1)}$$

$$E(a_i b_j) = \frac{n_{11}n_{1x}}{N(N - 1)}$$

In addition, we define the following functions of the  $k_i$ 's:

$$N\mu_k = \sum_i k_i + \sum_i b_i \left( l_i - \frac{k_i}{2} \right) \quad (11)$$

$$N\sigma_k^2 = \sum_i (k_i - \mu_k)^2$$

$$N(N-1)\sigma_{kk'} = \sum_{i \neq j} (k_i - \mu_k)(k_j - \mu_k).$$

Here, and in what follows, all summations are over all individuals in the parent generation, i.e.,  $i, j = 1, 2, \dots, N$ . The number of  $A_i$  gametes contributed to the next generation is  $\sum_i a_i k_i + \sum_i b_i l_i$ .

Recall that the  $k$ 's are fixed while the  $a$ 's,  $b$ 's, and  $l$ 's are random variables. The change in frequency of allele  $A_i$  then is

$$\Delta p = \frac{\sum_i a_i k_i + \sum_i b_i l_i}{N\mu_k} - p$$

$$= \frac{\sum_i a_i k_i + \sum_i b_i l_i}{N\mu_k} - \frac{2 \sum_i a_i + \sum_i b_i}{2N}$$

So, we have

$$N\mu_k \Delta p = \sum_i a_i k_i + \sum_i b_i l_i - \left[ \sum_i a_i \mu_k + \left( \frac{1}{2} \right) \sum_i b_i \mu_k \right]$$

$$= \sum_i a_i (k_i - \mu_k) + \sum_i b_i \left( l_i - \frac{\mu_k}{2} \right)$$

$$= \sum_i a_i (k_i - \mu_k) + \sum_i b_i \left( l_i - \frac{k_i}{2} + \frac{k_i}{2} - \frac{\mu_k}{2} \right)$$

$$= \sum_i a_i (k_i - \mu_k) + \sum_i b_i \left( l_i - \frac{k_i}{2} \right) + \left( \frac{1}{2} \right) \sum_i b_i (k_i - \mu_k)$$

Squaring and taking expected values with respect to  $l, a$ , and  $b$ , we get

$$(N\mu_k)^2 E(\Delta p)^2 = E \left\{ \sum_i a_i (k_i - \mu_k) + \left( \frac{1}{2} \right) \sum_i b_i (k_i - \mu_k) + \sum_i b_i \left( l_i - \frac{k_i}{2} \right) \right\}^2$$

$$= E \left\{ \sum_i a_i (k_i - \mu_k) \right\}^2 + \left( \frac{1}{4} \right) E \left\{ \sum_i b_i (k_i - \mu_k) \right\}^2 + E \left\{ \sum_i b_i \left( l_i - \frac{k_i}{2} \right) \right\}^2 + E \left\{ \sum_i a_i (k_i - \mu_k) \sum_i b_i (k_i - \mu_k) \right\} + 2E \left\{ \sum_i a_i (k_i - \mu_k) \sum_i b_i \left( l_i - \frac{k_i}{2} \right) \right\} + E \left\{ \sum_i b_i (k_i - \mu_k) \sum_i b_i \left( l_i - \frac{k_i}{2} \right) \right\}.$$

The last two terms are zero because  $b_i$  and  $l_i$  are independent and  $E \left\{ \left( l_i - \frac{k_i}{2} \right) \right\} = 0$ . Dropping these and continuing, we have

$$(N\mu_k)^2 E(\Delta p)^2 = E \left\{ \sum_i a_i^2 (k_i - \mu_k)^2 + \sum_{i \neq j} a_i a_j (k_i - \mu_k)(k_j - \mu_k) \right\} + \left( \frac{1}{4} \right) E \left\{ \sum_i b_i^2 (k_i - \mu_k)^2 + \sum_{i \neq j} b_i b_j (k_i - \mu_k)(k_j - \mu_k) \right\} + E \left\{ \sum_i b_i^2 \left( l_i - \frac{k_i}{2} \right)^2 \right\}$$

$$+ \sum_{i \neq j} b_i b_j \left( l_i - \frac{k_i}{2} \right) \left( l_j - \frac{k_j}{2} \right) \Big\} + E \left\{ \sum_i a_i b_i (k_i - \mu_k)^2 + \sum_{i \neq j} a_i b_j (k_i - \mu_k)(k_j - \mu_k) \right\}$$

$$= N\sigma_k^2 E \{ a_i^2 \} + N(N-1)\sigma_{kk'} E \{ a_i a_j \} + \left( \frac{1}{4} \right) N\sigma_k^2 E \{ b_i^2 \} + \left( \frac{1}{4} \right) N(N-1)\sigma_{kk'} E \{ b_i b_j \} + \sum_i E \left\{ \left( l_i - \frac{k_i}{2} \right)^2 \right\} E \{ b_i^2 \} + \sum_{i \neq j} E \left\{ \left( l_i - \frac{k_i}{2} \right) \left( l_j - \frac{k_j}{2} \right) \right\} E \{ b_i b_j \} + N\sigma_k^2 E \{ a_i b_i \} + N(N-1)\sigma_{kk'} E \{ a_i b_j \}$$

$$= n_{11}\sigma_k^2 + n_{11}(n_{11}-1)\sigma_{kk'} + \frac{n_{11}\sigma_k^2}{4} + \frac{n_{11}(n_{11}-1)\sigma_{kk'}}{4} + \frac{n_{11}\mu_k}{4} + 0 + 0 + n_{11}n_{11}\sigma_{kk'}. \quad (12)$$

The sixth term is zero because meioses in different individuals are independent, and the seventh term is zero because  $E \{ a_i b_j \} = 0$ . In (12),  $\sigma_k^2$  is the observed variance in the number of progeny per parent and  $\sigma_{kk'}$  is the covariance between progeny numbers of different parents. This covariance is negative, since the total number of progeny,  $N\mu_k$ , is fixed. Its value may be obtained by noting that  $\sum_i (k_i - \mu_k) = 0$ , and therefore

$$\sum_i (k_i - \mu_k)^2 + \sum_{i \neq j} (k_i - \mu_k)(k_j - \mu_k) = N\sigma_k^2 + N(N-1)\sigma_{kk'} = 0$$

from which

$$\sigma_{kk'} = \frac{-\sigma_k^2}{N-1} = \frac{-\sigma_k^2}{N_{i-1}-1}. \quad (13)$$

Substituting (13) into (12) and writing  $N_{i-1}$  for  $N$  gives

$$(N_{i-1}\mu_k)^2 \sigma^2(\Delta p) = \frac{\sigma_k^2 \left[ N_{i-1} \left( n_{11} + \frac{n_{1x}}{4} \right) - \left( n_{11} + \frac{n_{1x}}{2} \right)^2 \right]}{N_{i-1} - 1} + \frac{n_{1x}\mu_k}{4}. \quad (14)$$

It is now convenient to employ  $\alpha$  (equivalent to Wright's  $F$ ) as a measure of departure from Hardy-Weinberg ratios in the parent generation. Then,

$$n_{11} = N_{i-1}[p^2 + p(1-p)\alpha] \quad (15)$$

$$n_{1x} = N_{i-1}[2p(1-p)(1-\alpha)].$$

We also let  $s_k^2 = \sigma_k^2 [N_{i-1}/(N_{i-1}-1)]$ ; that is,  $s_k^2$  is  $\sigma_k^2$  with the Gaussian correction. Substituting (15) into (14) and rearranging gives

$$\sigma^2(\Delta p) = p(1-p) \left[ \frac{(1-\alpha)\mu_k + (1+\alpha)s_k^2}{2N_{i-1}\mu_k^2} \right]. \quad (16)$$

From the definition of variance effective number,

$$\sigma^2(\Delta p) = \frac{p(1-p)}{2N_{ev}}. \quad (17)$$

Equating (16) and (17) and noting that  $N_{i-1}\mu_k = 2N_i$ , we obtain the result we seek:

$$\frac{1}{N_{ev}} = \frac{1-\alpha + \frac{(1+\alpha)s_k^2}{\mu_k}}{2N_i} \quad (18)$$

in agreement with Kimura and Crow (1963).

When the mean and variance of the number of progeny per parent are equal to binomial expectations,  $s_k^2 = \mu_k$  and  $N_{ev} = N_i$ ; the population behaves like an ideal population. The variance effective number is more naturally related to the number of progeny than to the number of parents, as expected.

When  $\alpha = 0$  and  $s_k^2 = 0$ ,  $N_{ev} = 2N_i$ , showing that, in a population of stable size with binomial progeny distribution and with the parents in Hardy-Weinberg proportions, half the variance in gene-frequency drift is due to variable progeny numbers, and half is due to segregation from heterozygotes.

Usually we do not know the parental genotypes, and random mating is assumed. If the alleles in the parent generation are allocated randomly among the  $N_{t-1}$  individuals, then the expected value of  $\alpha$  is  $-1/(2N_{t-1} - 1)$  (see Kimura and Crow, 1963). Recalling that  $N_{t-1}\mu_k = 2N_t$ , (18) becomes

$$\frac{1}{N_{ev}} = \frac{2\left(1 + \frac{\sigma_k^2}{\mu_k}\right)}{4N_t - \mu_k} \quad (19)$$

in agreement with Kimura and Crow (1963 eq. 22) and Pollak (1977 eq. 4.1). Since the variance effective number is concerned only with allele-frequency changes, differences such as whether there is self-fertilization and whether male and female gametes are recognized do not matter.

When the variance of the progeny numbers corresponds to the binomial expectations ( $\sigma_k^2/\mu_k = 1 - 1/N_{t-1}$ ),  $N_{ev} = N_t$ . We saw the same result for the inbreeding effective number, except that the effective number is equal to the number of the parent generation,  $N_{t-1}$ . There is one other circumstance in which the formulas for  $N_{ei}$  and for  $N_{ev}$  [Eqs. (1) and (19)] are the same. This is when the population size is constant ( $\mu_k = 2$ ). In both cases, the effective number is  $(4N - 2)/(2 + \sigma_k^2)$ , as first given by Wright (1939).

**Separate Sexes.**—For our model of separate sexes, we assume that the total numbers of progeny of each sex are separately fixed. In assessing allele-frequency change between generations, we use an unweighted average of the frequencies in the two sexes. This is natural, since the total contributions of the two sexes must be equal.

Assume that there are  $N_{s,t-1} = N_s$  adults of sex  $s$  ( $s = m$  or  $f$ ) in the parent generation. Let  $k_{smi}$  and  $k_{sfi}$  be the numbers of sons and daughters of the  $i$ th parent of sex  $s$ ;  $k$  can, of course, be zero. Then, from (16) the variance of allele-frequency change from parents of sex  $s$  to their sons is

$$\sigma^2(\Delta p_{sm}) = p_s(1 - p_s) \frac{(1 - \alpha_s)\mu_{sm} + (1 + \alpha_s)s_{sm}^2}{2N_s\mu_{sm}^2} \quad (20)$$

in which  $\alpha_s$  is the departure from Hardy-Weinberg proportions in parents of sex  $s$  and  $\mu_{sm}$  is the mean number of sons of a

parent of sex  $s$ . (The first subscript always indicates the sex of the parent and the second that of the progeny.)

There will generally be a covariance between the number of sons and daughters of a parent. To assess this, we let  $n_{11}$  and  $n_{1s}$  stand for the numbers of homozygotes and heterozygotes for allele  $A_1$  in individuals of sex  $s$  and use definitions analogous to those of the previous section but applicable to parents of one sex. We define

$$N_s\sigma_{smi,sfi} = \sum_i (k_{smi} - \mu_{sm})(k_{sfi} - \mu_{sf})$$

and

$$N_s(N_s - 1)\sigma_{smi,sfj} = \sum_{ij} (k_{smi} - \mu_{sm})(k_{sfj} - \mu_{sf})$$

in which  $\mu_{sm}$  and  $\mu_{sf}$  are the mean numbers of sons and daughters of a parent of sex  $s$ , and  $N_s$  is the number of parents of sex  $s$ . In the summations,  $i$  and  $j$  go from 1 to  $N_s$ .

Then, using the same procedure as was used for deriving (12) and defining  $l$ ,  $a$ , and  $b$  in the same way as before, we write

$$\begin{aligned} N_s^2\mu_{sm}\mu_{sf}E(\Delta p_{sm}\Delta p_{sf}) &= E\left\{\sum a_i(k_{smi} - \mu_{sm})\right. \\ &\quad + \left(\frac{1}{2}\right)\sum b_i(k_{smi} - \mu_{sm}) \\ &\quad + \left.\sum b_i\left(l - \frac{k_{smi}}{2}\right)\right\} \\ &\quad \cdot \left\{\sum a_i(k_{sfi} - \mu_{sf})\right. \\ &\quad + \left(\frac{1}{2}\right)\sum b_i(k_{sfi} - \mu_{sf}) \\ &\quad + \left.\sum b_i\left(l - \frac{k_{sfi}}{2}\right)\right\}. \end{aligned}$$

This can be expanded as we did before. For example, the product of the first terms in each expression is

$$E\left\{\sum a_i(k_{smi} - \mu_{sm})\sum a_i(k_{sfi} - \mu_{sf})\right\} = E\left\{\sum a_i^2(k_{smi} - \mu_{sm})(k_{sfi} - \mu_{sf}) + \sum_{i \neq j} a_i a_j (k_{smi} - \mu_{sm})(k_{sfi} - \mu_{sf})\right\}$$

$$\begin{aligned} &= N_s\sigma_{smi,sfi}E\{a_i^2\} \\ &\quad + N_s(N_s - 1)\sigma_{smi,sfj}E\{a_i a_j\} \\ &= n_{11}\sigma_{smi,sfi} + n_{1s}(n_{11} - 1)\sigma_{smi,sfj}. \end{aligned}$$

Making similar computations for the other terms and simplifying leads to

$$\begin{aligned} N_s^2\mu_{sm}\mu_{sf}E(\Delta p_{sm}\Delta p_{sf}) &= n_{11}\sigma_{smi,sfi} + n_{1s}(n_{11} - 1)\sigma_{smi,sfj} \\ &\quad + \frac{n_{11}n_{1s}\sigma_{smi,sfj}}{2} + \frac{n_{1s}n_{1s}\sigma_{sfi,smj}}{2} \\ &\quad + \frac{n_{1s}\sigma_{smi,sfi}}{4} + \frac{n_{1s}(n_{1s} - 1)\sigma_{smi,sfj}}{4} \end{aligned} \quad (21)$$

in which  $\sigma_{smi,sfi}$  is the covariance between the number of sons and daughters of a parent of sex  $s$ , and  $\sigma_{smi,sfj}$  is the covariance between the number of sons and daughters of different parents, both of sex  $s$ .

We would expect  $\sigma_{smi,sfi}$  to be positive and possibly large if there is large variability in fertility. It could be negative if the variance of the number of progeny per parent is less than the binomial expectation, but in nature this must surely be rare. On the other hand,  $\sigma_{smi,sfj}$  will generally be very small. It can be obtained by the same device used to derive (13) and is

$$\sigma_{smi,sfj} = \frac{-\sigma_{smi,sfi}}{N_s - 1} \quad (22)$$

Substituting (22) and (15) into (21) yields

$$\begin{aligned} \sigma(\Delta p_{sm}, \Delta p_{sf}) &= \frac{p_s(1 - p_s)(1 + \alpha_s)\sigma_{sm,sf}}{2(N_s - 1)\mu_{sm}\mu_{sf}} \\ &= \frac{p_s(1 - p_s)(1 + \alpha_s)s_{sm,sf}}{2N_s\mu_{sm}\mu_{sf}} \end{aligned} \quad (23)$$

in which  $\sigma_{sm,sf}$  is the covariance of the number of sons and daughters of a parent of sex  $s$  and  $s_{sm,sf} = \sigma_{sm,sf}N_s/(N_s - 1)$ .

We can now write the variance of the allele-frequency change from generation  $t - 1$  to generation  $t$ . Let  $\Delta p_s$  be the change in allele frequency between a parent of sex  $s$  and its progeny. Then, weighting the two sexes equally,

$$\Delta p_s = \frac{\Delta p_{sm} + \Delta p_{sf}}{2} \quad (24)$$

The variance of  $\Delta p_s$  is given by

$$\begin{aligned} \sigma^2(\Delta p_s) &= \frac{[\sigma^2(\Delta p_{sm}) + \sigma^2(\Delta p_{sf}) + 2\sigma(\Delta p_{sm}, \Delta p_{sf})]}{4}. \end{aligned} \quad (25)$$

Substituting into (25) from (20) and (23), and equating  $\sigma^2(\Delta p_s)$  to  $p_s(1 - p_s)/2N_{evs}$ , we obtain

$$\begin{aligned} \frac{1}{N_{evs}} &= \frac{1}{4N_s} \left[ (1 - \sigma_s) \left( \frac{1}{\mu_{sm}} + \frac{1}{\mu_{sf}} \right) + (1 + \alpha_s) \right. \\ &\quad \left. \cdot \left( \frac{s_{sm}^2}{\mu_{sm}^2} + \frac{2s_{sm,sf}}{\mu_{sm}\mu_{sf}} + \frac{s_{sf}^2}{\mu_{sf}^2} \right) \right] \end{aligned} \quad (26)$$

where  $N_s$  is the number of individuals of sex  $s$  in generation  $t - 1$ .

Weighting the two sexes equally, the gene-frequency change from  $t - 1$  to  $t$  is  $\Delta p = \sum (\Delta p_s)/2 = (\Delta p_m + \Delta p_f)/2$ . Since we assume that the total numbers of progeny of each sex are separately fixed, these changes are independent. The allele-frequency variance from the contributions of both sexes is

$$\sigma^2(\Delta p) = \frac{[\sigma^2(\Delta p_m) + \sigma^2(\Delta p_f)]}{4} \quad (27)$$

Therefore,

$$\begin{aligned} \frac{4p(1 - p)}{2N_{ev}} &= \frac{1}{4} \left[ \frac{p_m(1 - p_m)}{2N_{evm}} + \frac{p_f(1 - p_f)}{2N_{evf}} \right] \end{aligned}$$

and

$$\frac{1}{N_{ev}} \approx \frac{1}{4N_{evm}} + \frac{1}{4N_{evf}} \quad (28)$$

in which  $N_{evm}$  and  $N_{evf}$  are given by (26). The approximation is usually very good, because we are dealing with autosomal loci, which means that there is no cumulative male-female difference in gene frequency and that the difference  $|x(1 - x) - y(1 - y)|$  is less than  $|x - y|$ . Equation (28) combines the two separate  $N_{ei}$ 's into an approximation for the whole population. One might, however, be more interested in describing the population by using the effective numbers for each sex separately.

Equation (28) may be substituted into (9)

to give an approximation to the change in allele-frequency variance or into a diffusion approximation. This gives an estimate for this generation only; it might be necessary to repeat the calculation each generation. When the variance of each gamete class (mm, mf, fm, and ff) is equal to its binomial expectation ( $s_{sm}^2/\mu_{sm} = s_{sf}^2/\mu_{sf} = 1$  and  $s_{sm,sf} = 0$ ), (28) reduces to Wright's formula (8), as expected.

Usually we assume that the parents are derived by random mating. In this case, when the numbers of male and female parents are separately fixed,  $\alpha_s = -1/(2N_s - 1)$ . Using this in (26) yields

$$\frac{1}{N_{evs}} = \frac{1}{2(2N_s - 1)} \left[ \frac{1}{\mu_{sm}} + \frac{1}{\mu_{sf}} + \frac{\sigma_{sm}^2}{\mu_{sm}^2} + \frac{2\sigma_{sm,sf}}{\mu_{sm}\mu_{sf}} + \frac{\sigma_{sf}^2}{\mu_{sf}^2} \right] \quad (29)$$

If the sex ratio is constant from generation to generation, we let  $R$  and  $R'$  stand for the male:female and female:male sex ratios, respectively. Then,  $\mu_{mm} = \mu_{ff} = \mu$ , and  $\mu_{fm} = R\mu$ ,  $\mu_{mf} = R'\mu$ , where  $\mu$  is the ratio of the total number in generation  $t$  to that of generation  $t - 1$ . From (28) and (29), we have

$$\frac{1}{N_{ev}} \approx \frac{1}{8(2N_m - 1)\mu^2} \cdot [\mu(1 + R) + \sigma_{mm}^2 + 2R\sigma_{mm,mf} + R^2\sigma_{mf}^2] + \frac{1}{8(2N_f - 1)\mu^2} \cdot [\mu(1 + R') + R'^2\sigma_{ff}^2 + 2R'\sigma_{fm,ff} + \sigma_{ff}^2] \quad (30)$$

in which  $\mu$  is  $N_t/N_{t-1}$  and  $N_m$  and  $N_f$  are the numbers of males and females in generation  $t - 1$ . Note that  $\mu_k = 2\mu$ . In a population of constant size,  $\mu = 1$ . If we ignore the distinction between  $2N_s - 1$  and  $2N_s$  and note that  $R/N_m = 1/N_f$  and  $R'/N_f = 1/N_m$ , we recover the discrete-generation special case of the results of Hill (1979 eq. 9).

#### Male and Female Progeny not Identified

In this case, it is reasonable to assume, as an approximation, that the total number of progeny of each parent is fixed but that the sex ratio is a random variable, as was first done by Latter (1959). Letting  $m_i$  and  $f_i$  ( $m_i + f_i = 1$ ) be the proportions of male and female progeny, we have the usual formulas (see, for example, Crow and Morton [1955 p. 207])

$$\begin{aligned} \sigma_{mm}^2 &= m_i f_i \mu_m + m_i^2 \sigma_m^2 \\ \sigma_{mf}^2 &= m_i f_i \mu_m + f_i^2 \sigma_m^2 \\ \sigma_{mm,mf}^2 &= -m_i f_i \mu_m + m_i f_i \sigma_m^2. \end{aligned} \quad (31)$$

Substituting (31) into (29) and replacing  $2N_s - 1$  by  $2N_s$ , we obtain

$$\frac{1}{N_{ev}} \approx \frac{1}{8N_{mf}} \left[ 2f_i + m_i \left( \frac{\sigma_m^2}{\mu_m} + \frac{\sigma_f^2}{\mu_f} \right) \right] + \frac{1}{8N_{ff}} \left[ 2m_i + f_i \left( \frac{\sigma_m^2}{\mu_m} + \frac{\sigma_f^2}{\mu_f} \right) \right]. \quad (32)$$

But  $m_i/N_{mf} = f_i/N_{ff}$ , and (32) becomes

$$\frac{1}{N_{ev}} \approx \frac{1}{4N_{mf}} \left( f_i + m_i \frac{\sigma_m^2}{\mu_m} \right) + \frac{1}{4N_{ff}} \left( m_i + f_i \frac{\sigma_f^2}{\mu_f} \right). \quad (33)$$

When the population size is constant ( $m = 1/\mu_m$  and  $f = 1/\mu_f$ ), (33) agrees with equation 6 in Latter (1959) and approximately with Pollak (1977). Equations (32) and (33) should replace equation 31 on p. 359 in Crow and Kimura (1970).

#### Overlapping Generations

We have considered only populations with discrete generations. There have been a number of papers devoted to the effective number in populations with overlapping generations (Nei and Imaizumi, 1966; Nei, 1970; Crow and Kimura, 1971; Felsenstein, 1971; Johnson, 1977; Pollak, 1977, 1980; Choy and Weir, 1978; Emigh and Pollak, 1979). The simplest results are those of Hill (1972, 1979). He showed that, for a population of constant size and sex ratio and with a stable age distribution, the effective number is the same as that for a discrete-gen-

eration population having the same variance in lifetime progeny numbers and the same number of individuals entering the population each generation.

For example, in (30) one has only to let  $\mu = 1$  (for a population of constant size) and replace  $N_m$  and  $N_f$  by  $ML$  and  $FL$ , where  $M$  and  $F$  are the numbers of males and females reaching breeding age in each time unit and  $L$  is the number of time units per generation (i.e., mean age of reproduction). Similar arguments apply to the inbreeding effective number (Hill, 1979).

To derive Hill's formulas one must assume that the population is of constant size and has a stable age distribution. The latter assumption is the most critical, and it seems reasonable that Hill's results would also be approximately correct for populations that are growing or contracting, provided that they have a stable age distribution. Thus, in (33),  $N_{mf}$  and  $N_{ff}$  could be replaced by  $LM$  and  $LF$ , where  $L$  is the mean age of reproduction and  $M$  and  $F$  are the geometric means of the number of males and females entering reproductive age in each time unit during the current generation. Of course any change of growth rate must be gradual enough for the age distribution to be in approximate equilibrium. This would rule out a great many populations whose rapid fluctuations keep the age structure permanently out of equilibrium.

#### DISCUSSION

The variance effective number provides an assessment of the amount of allele-frequency drift currently occurring in the population. If one is interested in conserving genetic variance, as with rare species or zoo animals, it is the most appropriate effective number. It is also appropriate when diffusion equations are used for the study of stochastic processes.

The inbreeding effective number is more useful if gene identity is the property of interest. For example, this is the relevant effective number to use in consideration of current models of evolution of multigene families (Ohta, 1980; Nagylaki, 1984). In domestic and zoo animals, common ancestry is often important, so in this case  $N_{ei}$  is the relevant effective number. However, if

actual pedigree information is available, direct calculation of the inbreeding coefficient is always preferable.

Ewens (1979 p. 111) says that the variance effective number is defined assuming only two alleles and that therefore the inbreeding effective number is possibly superior. We would emphasize that the two numbers measure different properties, and the choice of which to measure depends on the use to which the measurement is put. The variance effective number is appropriate for assessing the frequency change of an allele; it does not matter how many other alleles there are.

Ewens (1979) notes that with separate sexes a variance effective number as defined by Equation (10) strictly does not exist, and the allele-frequency variance of generation  $t$  cannot be given in terms of the allele frequency in generation  $t$  alone. This is indeed correct, as Equation (28) exemplifies, but the equation still serves our purpose of finding an approximation useful for studying natural populations. Ewens also departs from the Wright tradition and defines effective numbers in terms of alleles rather than diploid individuals. This is appropriate when genotypes can be observed but not when only demographic data are available.

In deriving formulas for the inbreeding effective number when self-fertilization is precluded or there are separate sexes, we assumed independence of the fertility of a parent and its progeny. This will clearly not be true in some cases, in particular if fertility is heritable (see Nei [1966] for one approach in this situation); but the heritability of such a fitness trait is likely to be very small. More likely are environmental correlations, which could be positive if parent and progeny both occupy an especially favorable niche. In contrast, highly fertile parents may produce overcrowding of their progeny, leading to a negative correlation. It would be desirable to have two-generation data on means and variances of fitness that would permit the measurement of such correlations. Taking them into account would yield an improved theory. In the absence of such information, the assumption of independence seems to us to be the only workable one. It is possible to work out the effective population number in other more complex cases, but there

comes a point of diminishing returns where it is more feasible to work out the allele frequency or homozygosity change directly.

Our definition of variance effective number uses an unweighted average of the allele frequencies in the two sexes. As stated earlier, this seems the most natural thing to do to take account of the biological constraint that each sex must contribute equally to the next generation. It is possible that a definition based on weighted averages could be developed, but we have tried and have not found a consistent, biologically meaningful formula. One often measures the variance and mean of the number of progeny per male and female parent in a separate experiment. Then the equations given here can be applied to estimating the relationship between the effective number and the census number. The ratio of the variance of the progeny number to its mean is usually the most important quantity.

In practice, the parameters that we have used have to be estimated from population studies. We have not considered the statistical problems associated with these estimates. This is a topic meriting a separate treatment (Ewens, 1982).

#### ACKNOWLEDGMENTS

Our indebtedness to Sewall Wright goes without saying. The clarity of presentation has been substantially improved by suggestions from anonymous reviewers. This is paper number 2912 from the Laboratory of Genetics, University of Wisconsin.

#### LITERATURE CITED

- CHIA, A. B., AND E. POLLAK. 1974. The inbreeding effective number and the effective number of alleles in a population that varies in sizes. *Theoret. Pop. Biol.* 6:149-172.
- CHOY, S. C., AND B. S. WEIR. 1978. Exact inbreeding coefficients in populations with overlapping generations. *Genetics* 89:591-614.
- CROW, J. F. 1954. Breeding structure of populations. II. Effective population number, pp. 543-556. *In* O. Kempthorne, T. A. Bancroft, J. W. Gowen, and J. L. Lush (eds.), *Statistics and Mathematics in Biology*. Iowa State College Press, Ames.
- CROW, J. F., AND M. KIMURA. 1970. An Introduction to Population Genetics Theory. Burgess, Minneapolis, MN.
- . 1971. The effective number of a population with overlapping generations: A correction and further discussion. *Amer. J. Hum. Genet.* 24:1-10.

- CROW, J. F., AND N. E. MORTON. 1955. Measurement of gene frequency drift in small populations. *Evolution* 9:202-214.
- EMIGH, T. H., AND E. POLLAK. 1979. Fixation probabilities and effective population numbers in diploid populations with overlapping generations. *Theoret. Pop. Biol.* 15:86-107.
- ETHIER, S. N., AND T. NAGYLAKI. 1980. Diffusion approximations of Markov chains with two time scales and applications to population genetics. *Adv. Appl. Prob.* 12:14-49.
- EWENS, W. J. 1979. *Mathematical Population Genetics*. Springer-Verlag, Berlin, W. Ger.
- . 1982. On the concept of effective population size. *Theoret. Pop. Biol.* 21:373-378.
- FELSENSTEIN, J. 1971. Inbreeding and variance effective numbers in populations with overlapping generations. *Genetics* 68:581-597.
- FISHER, R. A. 1939. Stage of development as a factor influencing the variance in the number of offspring, frequency of mutants, and related quantities. *Ann. Eugen.* 9:406-408.
- HALDANE, J. B. S. 1939. The equilibrium between mutation and random extinction. *Ann. Eugen.* 9:400-405.
- HILL, W. G. 1972. Effective size of populations with overlapping generations. *Theoret. Pop. Biol.* 3:278-289.
- . 1979. A note on effective population size with overlapping generations. *Genetics* 92:317-322.
- JOHNSON, D. L. 1977. Inbreeding in populations with overlapping generations. *Genetics* 87:581-591.
- KIMURA, M., AND J. F. CROW. 1963. The measurement of effective population number. *Evolution* 17:279-288.
- 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.
- MORAN, P. A. P., AND G. A. WATTERSON. 1959. The genetic effects of family structure in natural populations. *Aust. J. Biol. Sci.* 12:1-15.
- NAGYLAKI, T. 1981. The inbreeding population number and the expected homozygosity for an X-linked locus. *Genetics* 97:731-737.
- . 1984. Evolution of multigene families under inter-chromosomal gene conversion. *Proc. Nat. Acad. Sci. USA* 81:3796-3800.
- NEI, M. 1966. Effective population size when fertility is inherited. *Genet. Res.* 8:257-260.
- . 1970. Effective size of human populations. *Amer. J. Hum. Genet.* 22:694-695.
- NEI, M., AND Y. IMAIZUMI. 1966. Genetic structure of human populations. II: Differentiation of blood group gene frequencies among isolated populations. *Heredity* 21:183-190, 344.
- OHTA, T. 1980. *Evolution and Variation of Multigene Families*. Springer-Verlag, Berlin, W. Ger.
- POLLAK, E. 1977. Effective population numbers and their interrelationships, pp. 115-144. *In* Proceedings of the Washington State University Conference on Biomathematics and Statistics. Department of Pure and Applied Mathematics, Washington State Univ., Spokane, WA.
- . 1980. Effective population numbers and mean times to extinction in dioecious populations with overlapping generations. *Math. Biosci.* 52:1-25.

- PROVINE, W. B. (ed.). 1986. *Evolution, Selected Papers*. Univ. Chicago Press, Chicago, IL.
- WRIGHT, S. 1931. *Evolution in Mendelian populations*. *Genetics* 16:97-159.
- . 1939. *Statistical Genetics in Relation to Evo-*

lution. *Exposés de Biométrie et de Statistique Biologique*. Herman & Cie. Paris, France.

Corresponding Editor: M. C. Uyenoyama