GENETIC SAMPLING IN A RANDOM MATING POPULATION OF CONSTANT SIZE AND SEX RATIO

By B. D. H. LATTER*

[*Manuscript received March* 23, 1959]

Summary

An expression is derived for the effective breeding size of a random mating population of constant size and sex ratio, which takes into account reproductive inequalities among the breeding individuals. Consideration is given to the problem of estimating the possible genetic drift in the mean of a given quantitative character in an experimental control population.

I. INTRODUCTION

The practice of maintaining a control population of animals, to provide a reference point over a period of generations, is becoming increasingly popular in selection programmes wherever the necessary facilities are available. Whether one is interested to assess short-term fluctuations in the effects of environmental factors on a particular quantitative trait, or to detect and measure long-term trends due to improvements in management, the problem is to hold as constant as possible the genetic constitution of the base population. Gene frequencies in the population will, however, be expected to change from generation to generation, due to the sampling of genes involved in the choice of parents, and to the effects of natural selection. The ways in which the effects of these two agencies can be minimized have been discussed with particular reference to poultry flocks by Gowe, Robertson, and Latter (1959).

If one is able to tag the offspring of each mating within the population and control the contributions of the respective families to the next set of selected parents, the efficiency of such a procedure in reducing the effects of genetic sampling can readily be calculated. It is in fact possible to achieve substantial gains in efficiency in this way, by ensuring that each parent contributes equally to the succeeding set of parents. If, however, the parents are randomly chosen each generation from the survivors to breeding age, changes in gene frequency due to genetic sampling will obviously be a function of differences in reproductive capacity among the breeding individuals. The purpose of this paper is to discuss the genetic sampling process in the latter situation, and to derive a general expression for the effective breeding size of the population. The expression has previously been given without proof by Gowe, Robertson, and Latter (1959).

II. GENERAL APPROACH

The general approach to problems of drift in gene frequency may be illustrated by the derivation (due to Dr. A. Robertson) which follows. Suppose we have N parents from which we sample $n_1, n_2, \ldots, n_N$ gametes respectively, and we consider

* Division of Plant Industry, C.S.I.R.O., Canberra.

only sampling between individuals in its effect on the change in gene frequency. If q_i is the gene frequency at a particular locus in the ith parent (the possible values of q_i being 0, $\frac{1}{2}$, or 1), then the change in gene frequency due to sampling between individuals is

$$\delta q' = \sum_i q_i\left(\frac{n_i}{\Sigma n} - \frac{1}{N}\right),$$

and the expected drift variance is therefore

$$E(\delta q')^2 = \sigma_q^2 E\left[\sum_i\left(\frac{n_i}{\Sigma n} - \frac{1}{N}\right)^2\right]$$

$$= \frac{q(1-q)}{2}\left[\frac{N\sigma_n^2}{(\Sigma n)^2}\right],$$

where q and σ_q^2 are the mean and variance of the q_i, and σ_n^2 the variance of the n_i. In addition, the sampling of gametes within individuals contributes to the drift variance whenever an individual is heterozygous at the locus concerned, giving a total drift variance of

$$E(\delta q)^2 = \frac{q(1-q)}{2}\left[\frac{1}{\Sigma n} + \frac{N\sigma_n^2}{(\Sigma n)^2}\right]. \quad \text{.....(1)}$$

III. Derivation

The population structure with which we shall be concerned is one in which M males and F females are chosen as parents randomly from the surviving animals each generation, and are mated at random to produce the members of the next generation. The overall change in gene frequency from one generation to the next involves genetic sampling along the four paths from male parents to succeeding male parents, from male parents to succeeding female parents, and so on, and we must take account of the fact that the changes in gene frequency along the separate paths are not necessarily uncorrelated.

Consider M male parents contributing x_i males to the next generation of male parents and y_i females to the next generation of female parents ($i = 1, 2, \ldots, M$). Suppose x_i has mean μ_1 and variance σ_1^2, and y_i has mean μ_2 and variance σ_2^2. Denoting by δq_1 the change in gene frequency along the path male $\rightarrow$ male, and by δq_2 the change along the path male $\rightarrow$ female, we have from (1)

$$E(\delta q_1)^2 = \frac{q(1-q)}{2M\mu_1}\left[1 + \frac{\sigma_1^2}{\mu_1}\right],$$

and

$$E(\delta q_2)^2 = \frac{q(1-q)}{2M\mu_2}\left[1 + \frac{\sigma_2^2}{\mu_2}\right]. \quad \text{.....(2)}$$

Since the overall change in gene frequency due to sampling from males is $\delta q_m = \frac{1}{2}(\delta q_1 + \delta q_2)$, the expected value of $(\delta q_m)^2$ involves the covariance of δq_1

and δq_2. It is obvious that only sampling between individuals is involved in this covariance, and we have

$$\delta q_1' = \sum_i q_i\left(\frac{x_i}{M\mu_1} - \frac{1}{M}\right),$$

$$\delta q_2' = \sum_i q_i\left(\frac{y_i}{M\mu_2} - \frac{1}{M}\right),$$

so that the expected value of $(\delta q_1)(\delta q_2)$ is

$$E(\delta q_1)(\delta q_2) = \frac{q(1-q)}{2} E\left[\sum_i \left(\frac{x_i}{M\mu_1} - \frac{1}{M}\right)\left(\frac{y_i}{M\mu_2} - \frac{1}{M}\right)\right]$$

$$= \frac{q(1-q)}{2M} \cdot \frac{\mathrm{cov}(x_i, y_i)}{\mu_1\mu_2}. \qquad \ldots\ldots\ldots\ldots\ldots\ldots (3)$$

If complete pedigrees are available, the required covariance term can be calculated directly, but in order to obtain a general expression for use in the absence of pedigree records, we are lead to postulate the following sequence in the selection of parents in any generation. Suppose each male contributes m_i individuals regardless of sex to a group of progeny which is to be sampled at random (separately within each sex) to provide the succeeding M male and F female parents. Let the mean and variance of m be μ_m, σ_m^2. The m_i offspring of the ith male will then be distributed to the two sexes according to a binomial distribution with parameters m_i and $\frac{1}{2}$ denoted $B\,(m_i, \frac{1}{2})$. If the actual split is n_i males and $m_i - n_i$ females, the probabilities that x_i, y_i individuals are included in the sample following random choice of M males and F females are given by $B\,(n_i, \mu_1/\frac{1}{2}\mu_m)$ and $B\,(m_i - n_i, \mu_2/\frac{1}{2}\mu_m)$.

The required covariance is then

$$\mathrm{cov}(x_i, y_i) = [(\mu_1\mu_2)/\tfrac{1}{4}\mu_m^2]\,\mathrm{cov}(n_i, m_i - n_i)$$

$$= \mu_1\mu_2\left(\frac{\sigma_m^2 - \mu_m}{\mu_m^2}\right).$$

Before combining expressions (2) and (3), we must derive a relationship between σ_m^2 and σ_1^2, σ_2^2. Within the framework of the above assumptions, it turns out that

$$\sigma_1^2 = \mu_1\left[1 + \mu_1\left(\frac{\sigma_m^2 - \mu_m}{\mu_m^2}\right)\right],$$

and

$$\sigma_2^2 = \mu_2\left[1 + \mu_2\left(\frac{\sigma_m^2 - \mu_m}{\mu_m^2}\right)\right],$$

so that we can write

$$E(\delta q_m)^2 = \frac{q(1-q)}{4}\left[\frac{1}{M\mu_1} + \frac{1}{M\mu_2} + \frac{2R_m}{M}\right],$$

where $R_m = (\sigma_m^2 - \mu_m)/\mu_m^2$. In a population of constant size and sex ratio, $\mu_1 = 1$ and $\mu_2 = F/M$, and the expression becomes

$$E(\delta q_m)^2 = \frac{q(1-q)}{4}\left[\frac{1}{M} + \frac{1}{F} + \frac{2R_m}{M}\right]. \quad \ldots\ldots\ldots\ldots (4)$$

A similar expression can be derived for the drift variance due to sampling along the two paths from female → male and female → female, and these gene frequency changes are obviously independent of the two already considered. We have

$$E(\delta q_f)^2 = \frac{q(1-q)}{4}\left[\frac{1}{M} + \frac{1}{F} + \frac{2R_f}{F}\right], \quad \ldots\ldots\ldots\ldots (5)$$

where $R_f = (\sigma_f^2 - \mu_f)/\mu_f^2$, μ_f and σ_f^2 being defined in the same manner as μ_m, σ_m^2. Combining (4) and (5), the average change in gene frequency from one generation to the next has variance

$$E[\tfrac{1}{2}\delta q_m + \tfrac{1}{2}\delta q_f]^2 = q(1-q)\left[\frac{1}{8M}(1+R_m) + \frac{1}{8F}(1+R_f)\right]. \quad \ldots\ldots (6)$$

The derivation of expression (6) has been based on two assumptions in addition to the stipulation that chosen parents be randomly mated each generation; they are:

(1) That the group of progeny to which the parameters μ_m, σ_m^2 and the corresponding μ_f, σ_f^2 refer shall be sampled at random separately within each sex to provide M male and F female breeding individuals each generation; and

(2) That mortality in the population, prior to the point of time to which the abovementioned parameters refer, shall have been independent of sex.

With complete equality among the male parents and among the female parents in reproductive capacity, and random mortality among the offspring, each parent has an equal chance of contributing to the next set of selected parents, and the anticipated distribution of progeny number per parent is one in which the mean and variance are equal (Poisson), i.e. R_m and R_f are zero. Expression (6) then reduces to the familiar formula

$$E(\delta q)^2 = q(1-q)\left[\frac{1}{8M} + \frac{1}{8F}\right].$$

IV. Discussion

An expression has been given by Crow and Morton (1955) for the expected drift variance in gene frequency in a population with unequal numbers in the two sexes which at first sight seemed to provide the solution to the problem we have posed. In the above notation, their equation can be written

$$E(\delta q_m)^2 = \frac{q(1-q)}{2M(\mu_1+\mu_2)}\left[1 + \frac{\mathrm{var}(x_i+y_i)}{\mu_1+\mu_2}\right],$$

which is a direct application of expression (1) above. In the framework we have specified above, this reduces to

$$E(\delta q_m)^2 = \frac{q(1-q)}{4}\left[\frac{4}{M+F} + \frac{2R_m}{M}\right],$$

which is only equivalent to (4) if the numbers of male and female parents in the population are equal. The difference stems from the fact that Crow and Morton have taken the average gene frequency following sampling as a weighted average of the gene frequencies in males and females, whereas the average relevant to our problem is unweighted. Presumably their expression was not intended to apply to a population in which the sex ratio is held constant from generation to generation.

The problem of estimating the possible genetic drift in the mean of a given quantitative character in an experimental control population can best be considered in terms of the "effective" size of the population (N_e), defined such that

$$E(\delta q)^2 = \frac{q(1-q)}{2N_e}.$$

One can also speak of the effective number of male parents (M_e) and the effective number of female parents (F_e), where

$$\frac{1}{N_e} = \frac{1}{4M_e} + \frac{1}{4F_e}.$$

The change in genetic mean performance from one generation to the next due to genetic sampling has variance roughly equal to σ_g^2/N_e where σ_g^2 is the additive genetic variance in the character concerned, so that an estimate of effective population size can be used to provide some measure of the genetic changes in the population mean which are likely over any given period of generations. Such a measure will of course take no account of the effects of natural selection over the period concerned, and is likely to be most useful in the case of populations with no immediate history of artificial selection.

An instance of this approach has been given by Gowe and Johnson (1959). The data which these authors had available were the mean number of eggs per hen laid during the period from which replacements were kept, and the corresponding variance; and the mean hatchability per hen over the same period, with the corresponding variance. Making the assumption that production and hatchability were uncorrelated, it was possible to estimate the mean and variance of number of chicks per hen, and to make use of these statistics to estimate the effective number of females in the control flock concerned. Use of the expression given in this paper was conditional upon the assumptions that mortality from hatching to breeding age was random within each sex, and that embryonic mortality was independent of sex.

V. Acknowledgments

My thanks are due to Dr. A. Robertson, Institute of Animal Genetics, Edinburgh, and to Dr. R. S. Gowe, Central Experimental Farm, Ottawa, for guidance and stimulating discussion. The award of a Walter and Eliza Hall Agricultural Research Fellowship from the University of Sydney is gratefully acknowledged.

VI. References

Crow, J. F., and Morton, N. E. (1955).—Measurement of gene frequency drift in small populations. *Evolution* **9**: 202–14.

Gowe, R. S., and Johnson, A. S. (1959).—Environment and poultry breeding problems. IV. The value of a random-bred control strain in a selection study. *Poult. Sci.* **38**: 443–62.

Gowe, R. S., Robertson, A., and Latter, B. D. H. (1959).—Environment and poultry breeding problems. V. The design of poultry control strains. *Poult. Sci.* **38**: 462–71.