Random selective advantages of genes and their probabilities of fixation*

By LOUIS JENSEN†

Statistical Laboratory, Iowa State University, Ames, Iowa, U.S.A.

SUMMARY

The question of what is meant by random fluctuations in selection intensities in a finite population is re-examined. The model presented describes the change in the frequency of a gene in a haploid population of size M. It is assumed that in any generation the adaptive values of A and a are equally likely to be $1+s:1$ or $1:1+s$. If s is the selective advantage and x the frequency of gene A, then the first two moments of the change in frequency are found to be $m(\Delta x) = x(1-x)(1-2x)\,\theta/2M$ and

$$v(\Delta x) = x(1-x)/M + x^2(1-x)^2\,\theta/M,$$

where $E(s^2) = \theta/M$. The ultimate probability of fixation is computed, showing that variability in selection increases the chance of fixation of a rare gene. A more general form for $m(\Delta x)$ also is obtained. This form is compared with the equation currently used in describing random fluctuations in selection intensities.

It has been observed that populations, both large and small, are subjected to varying selective intensities from generation to generation. These variations are primarily due to changing environmental conditions. Fisher & Ford (1947) conjectured that varying selection intensities are capable of producing greater fluctuations in gene ratios than could be attributed to random sampling alone. In this paper, a model is presented that describes gametic selection, due to changes in environment, which varies from generation to generation in a random manner, and the ultimate probability of fixation is computed.

Let us consider a large haploid population of constant size M, with genes A and a at a single locus. If the initial frequency of gene A is p, then it is known that $u(p)$, the ultimate probability that A becomes fixed in the population, satisfies

$$\tfrac{1}{2}v(\Delta p)\frac{d^2}{dp^2}u(p) + m(\Delta p)\frac{d}{dp}u(p) = 0 \quad (0 \leqslant p \leqslant 1), \tag{1}$$

with $u(0) = 0$ and $u(1) = 1$. The functions $m(\Delta p)$ and $v(\Delta p)$ are, respectively, the first and second moments of the change in gene frequency in one generation (Kimura (1962)).

15-2

Let s be the selective advantage of gene A over a. In describing the change in the frequency of gene A due to random sampling in the presence of random fluctuations of selection intensities, s is a random variable with mean μ/M and 'variance' σ^2/M. The first two moments of the change in gene frequency are currently thought to be

$$m(\Delta x) = x(1-x)\,\mu/M,$$
$$v(\Delta x) = x(1-x)/M + x^2(1-x)^2\,\sigma^2/M. \tag{2}$$

Recently, Ohta (1972) investigated the probability of ultimate fixation with $m(\Delta x)$ and $v(\Delta x)$ given by (2). She computed the fixation probability for various values of the two parameters μ and σ^2. In particular, she was interested in the case when the fluctuations in selection are 'nearly neutral'; that is, when μ, in some sense, is small. Her conclusion was that variability in s reduces the chance of fixation of a rare gene. Unfortunately, the motivation for the form of the mean function given in (2) seems unjustified, and it is not clear what actually is meant by random selection being 'nearly neutral'.

A model that probabilistically incorporates random fluctuations in the selection intensities, which are, in a sense, 'neutral on the average', was proposed by Jensen & Pollak (1969). The model is as follows: Let x_t denote the frequency of gene A in the adult population at time t. It is assumed that the total change in gene frequency in one generation, $\Delta x_t = x_{t+1} - x_t$, can be represented as

$$\Delta x_t = \Delta_r x_t + \Delta_s x_t, \tag{3}$$

where Δ_r and Δ_s denote the change due to random sampling and due to random selection, respectively. The two types of change act independently of each other.

It is well known that

$$E(\Delta_r X_t) = 0 \quad \text{and} \quad E((\Delta_r X_t)^2) = (1/M)\,X_t(1-X_t) + O(M^{-2}). \tag{4}$$

The higher moments of $\Delta_r X_t$ are of the form $O(M^{-2})$ and can be neglected for large M.

Random selection is introduced thusly: The two genotypes A and a have adaptive values $1+s:1$ or $1:1+s$, respectively, each chosen in any generation with probability $\frac{1}{2}$. Thus, the advantage of a particular genotype is only a function of the environment, and the environment has two states occurring with equal probabilities. This type of selection is termed 'random selection being zero on the average', since either genotype is equally likely to be advantageous. If s is fixed,

$$E(\Delta_s X_t | s) = (s^2/2)\,X_t(1-X_t)\,(1-2X_t) + O(s^3),$$
$$E((\Delta_s X_t)^2 | s) = s^2 X_t^2(1-X_t)^2 + O(s^3).$$

Let s be a random variable with $E(s^2) = \theta/M$ ($\theta > 0$) and higher moments of the form $O(M^{-2})$ so that they can be neglected if M is large. By using conditional expectation, it follows that

$$E(\Delta_s X_t) = E_s E(\Delta_s X_t | s)$$
$$= \frac{\theta}{2M}\,X_t(1-X_t)\,(1-2X_t) + O(M^{-2}),$$
$$E((\Delta_s X_t)^2) = \frac{\theta}{M}\,X_t^2(1-X_t)^2 + O(M^{-2}). \tag{5}$$

The first and second moments of the total change in gene frequency are obtained by using equations (3), (4) and (5). That is,

$$E(\Delta X_t) = E_s E(\Delta X_t | s)$$
$$= \frac{\theta}{2M} X_t(1-X_t)(1-2X_t) + O(M^{-2}),$$

$$E((\Delta X_t)^2) = \frac{1}{M} X_t(1-X_t) + \frac{\theta}{M} X_t^2(1-X_t)^2 + O(M^{-2}). \tag{6}$$

Neglecting any terms of the form $O(M^{-2})$, $m(\Delta x)$ and $v(\Delta x)$ are approximately

$$m(\Delta x) = \frac{\theta}{2M} x(1-x)(1-2x),$$

$$v(\Delta x) = \frac{1}{M} x(1-x) + \frac{\theta}{M} x^2(1-x)^2. \tag{7}$$

The general form of the mean function, being a function of θ, has also been obtained by Gillespie (1973). He presents a model for random fluctuations in selection that differs from the one presented here. The results, however, are almost identical.

In this special case, the solution to equation (1) (i.e. the ultimate probability that gene A is fixed in the population) is found to be

$$u(p) = \frac{\ln\left(\dfrac{p-\lambda_2}{\lambda_1-p}\right)}{2\ln\left(-\dfrac{\lambda_1}{\lambda_2}\right)} + \tfrac{1}{2} \quad (\theta > 0), \tag{8}$$

where
$$\lambda_1 = \tfrac{1}{2}\{1 + \sqrt{[1+(4/\theta)]}\},$$
$$\lambda_2 = \tfrac{1}{2}\{1 - \sqrt{[1+(4/\theta)]}\}.$$

If $\theta = 0$, the probability of fixation is $u(p) = p$ (Kimura, 1962). Equation (8) is symmetric in p; that is, $u(p) + u(1-p) = 1$ and $u(\tfrac{1}{2}) = \tfrac{1}{2}$ irrespective of the value of θ.

An interesting consequence of the effect of this type of random selection is that as $\theta \to \infty$

$$u(p) \uparrow \tfrac{1}{2} \quad \text{for} \quad 0 < p < \tfrac{1}{2},$$
$$u(p) \downarrow \tfrac{1}{2} \quad \text{for} \quad \tfrac{1}{2} < p < 1.$$

This indicates that the ultimate probability of fixation is forced to be near $\tfrac{1}{2}$ independent of the initial frequency, provided that $0 < p < 1$ and θ is large. The consequence is that variability in selection increases the chance of fixation of a rare gene. This is contrary to Ohta's findings.

Figure 1 illustrates the probability of fixation for several initial gene frequencies and for various values of θ. Only values of p between 0 and $\tfrac{1}{2}$ are presented. The probability of fixation for other values of p may be obtained by symmetry.

If, originally, the two adaptive values $1+s:1$ or $1:1+s$ were chosen in any generation with probabilities q and $1-q$, respectively, then it is straightforward to show
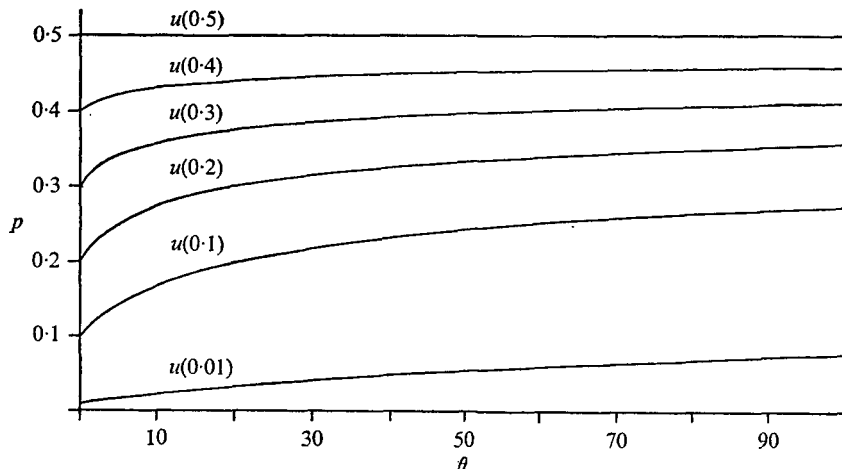
Fig. 1. The probability of fixation, $u(p)$, for several initial gene frequencies, p,
as a function of θ.

that $E((\Delta X_t)^2)$ is identical to that given in equation (6). However, the mean of total change in gene frequency becomes

$$E(\Delta X_t) = (2q-1)\frac{\mu}{M}X_t(1-X_t)+\frac{\theta}{M}X_t(1-X_t)(1-q-X_t)+O(M^{-2}), \qquad (9)$$

where $E(s) = \mu/M$ and θ/M is as before. For no choice of the three parameters, q, μ and θ, is this equation of the form $E(\Delta X_t) = KX_t(1-X_t)+O(M^{-2})$ for some constant K. In the special case when $q = 1$, equation (9) reduces to

$$E(\Delta X_t) = \frac{\mu}{M}X_t(1-X_t)-\frac{\theta}{M}X_t^2(1-X_t)+O(M^{-2}). \qquad (10)$$

Thus, if the model for random selection is gametic selection with random fluctuations in the selection intensities, then the mean of the change in gene frequency is given by equation (10). It seems there is little justification for using

$$m(\Delta x) = x(1-x)\,\mu/M$$

as the first moment of the change in gene frequency due to random fluctuations in selection. If equation (9) or (10) is used as the first moment function, the resulting differential equation for the probability of fixation is not amenable to solution. The author, however, feels that the only interesting situation is when $q = \frac{1}{2}$, since, in this case, the fluctuation are in a sense 'neutral'.

REFERENCES

FISHER, R. A. & FORD, E. B. (1947). The spread of a gene in natural conditions in a colony of the moth *Panaxia dominula* L. *Heredity* **1**, 143–174.

GILLESPIE, J. (1973). Haploid alleles in uncorrelated environments. *Genetical Research* **21**, 115–120.

JENSEN, L. & POLLAK, E. (1969). Random selective advantages of a gene in a finite population. *Journal of Applied Probability* **6**, 19–37.

KIMURA, M. (1962). On the probability of fixation of mutant genes in a population. *Genetics* **47**, 713–719.

OHTA, T. (1972). Fixation probability of a mutant influenced by random fluctuation of selection intensity. *Genetical Research* **19**, 33–38.