The average number and the variance of generations at particular gene frequency in the course of fixation of a mutant gene in a finite population*

By TAKEO MARUYAMA

National Institute of Genetics, Mishima, Japan

SUMMARY

In the case of an allele which is going to become fixed in a population, the average number of generations for which the population assumes particular gene frequencies is investigated, using the diffusion approximation. Explicit formulas were obtained and they were checked by computer simulations. As a particular case, it is shown that if a new mutant that is selectively neutral is eventually fixed in a population of size N, it spends two generations on average at each of the intermediate frequencies $(1/2N, 2/2N, ..., (2N-1)/2N)$, and the variance at each frequency is four generations.

The number of generations until a mutant gene becomes fixed (excluding the case of eventual loss) by random genetic drift and the selection was given by Kimura & Ohta (1969 *a*, *b*). For the case of a selectively neutral gene, the distribution density of the fixation time was also obtained by Kimura (1970). The above investigations were based on the diffusion model. Narain (1970) obtained the variance of the fixation time for the neutral case also based on the diffusion model, and compared the results with those obtained by the matrix method for the discrete generation model with small population size.

These investigations, however, give no information on the 'sojourn time' at a particular gene frequency in the course of gene fixation. In other words, if a particular allele starting from initial frequency x becomes fixed, then how long does the gene frequency stay (or visit) at particular frequency y? This problem of sojourn time in the cases of both eventual fixation and eventual loss was solved by Kimura (1969). In this note, I shall solve this problem in the eventually fixed cases for two allele situations of panmictic finite populations.

Let us denote the alleles by A_1 and A_2 and let the relative fitnesses of A_1A_1, A_1A_2 and A_2A_2 be $1+s$, $1+sh$ and 1 respectively. We denote by x or y the frequency of A_1 and denote by N the number of diploid individuals in the population. Let $\Phi_1(x, y) \, dy$ be the number of generations for which the population assumes the gene frequency in range $(y - \frac{1}{2}dy, y + \frac{1}{2}dy)$, provided its initial frequency is x and its destination is the fixation of A_1. Let $p(t, x, y)$ be the transition probability that

8-2

the gene frequency of a population moves from x to y in time interval t, and let $u(x)$ be the ultimate fixation probability of A_1 if its initial frequency is x. A population assumes the gene frequency y at time t with probability $p(t, x, y)$ and then its conditional fixation probability is $u(y)$. This means that a population starting from x goes to y at time t and to fixation with probability $p(t, x, y) u(y)$. Therefore we have

$$\Psi(x, y) \equiv u(x)\Phi_1(x, y) = \int_0^\infty p(t, x, y) u(y) \, dt. \tag{1}$$

Since the transition probability satisfies

$$\frac{\partial p(t, x, y)}{\partial t} = \frac{V_{\delta x}}{2} \frac{\partial^2 p(t, x, y)}{\partial x^2} + M_{\delta x} \frac{\partial p(t, x, y)}{\partial t} \tag{2}$$

with $V_{\delta x} = x(1-x)/2N$ and $M_{\delta x} = x(1-x)\{(1-2x) sh - sx\}$, substituting the $\Psi(x, y)$ in (1) in this equation, we have

$$\frac{V_{\delta x}}{2} \frac{d^2\Psi(x, y)}{dx^2} + M_{\delta x} \frac{d\Psi(x, y)}{dx} = -u(y) \delta(x-y) \tag{3}$$

where $\delta(\cdot)$ is Dirac's delta function. The boundary condition that $\Psi(x, y)$ must satisfy is $\Psi(0, y) = \Psi(1, y) = 0$. (For the differential equation (2), see Crow & Kimura, 1970, Ch. 8.) A general treatment of this kind of analysis is given in Maruyama & Kimura (1971).

The differential equation (3) with the above boundary condition can be immediately integrated:

$$\Psi(x, y) = \frac{2u(y) g(0, x) g(y, 1)}{G(y) V_{\delta x} g(0, 1)} \quad \text{for } y > x$$

$$= \frac{2u(y)}{G(y) V_{\delta x}} \left\{ \frac{g(0, x) g(y, 1)}{g(0, 1)} - g(y, x) \right\} \quad \text{for } y < x,$$

where $\quad G(x) = \exp\left\{ -\int_0^x \frac{2M_{\delta \zeta}}{V_{\delta \zeta}} d\zeta \right\}$ and $g(a, b) = \int_a^b G(x) \, dx$,

and $\qquad\qquad\qquad \Phi_1(x, y) = \frac{\Psi(x, y)}{u(x)}$,

where $u(x) = g(0, x)/g(0, 1)$ (Kimura, 1962). If $s = 0$ (neutral case)

$$\Phi_1(x, y) = 4N \quad \text{for } y > x \tag{4}$$

$$= 4N \frac{y(1-x)}{x(1-y)} \quad \text{for } y < x;$$

and if $s \neq 0$ and $h = \frac{1}{2}$,

$$\Phi_1(x, y) = \frac{2S(y) S(1-y)}{sy(1-y) S(1)} \quad \text{for } y > x$$

$$= \frac{2S(y)}{sy(1-y)} \left\{ \frac{S(1-y)}{S(1)} - \frac{S(x-y)}{S(x)} \right\} \quad \text{for } y < x, \tag{5}$$

where $S(y) = 1 - \exp(-2Nsy)$.

The quantity $\Phi_1(x, y)/2N$ gives the average number of generations spent at gene frequency y $(= 1/2N, 2/2N, \ldots, (2N-1)/2N)$ by the mutant allele, provided that it starts with frequency x and is eventually fixed in the population. As a particular case, if a new mutant that is selectively neutral is fixed by random genetic drift, it spends two generations on average at each of the intermediate frequencies (formula (4) with $x = 1/2N$), and it takes altogether $4N$ generations to reach fixation, which agrees with the result of Kimura & Ohta (1969). It is worth noting that in the case of no dominance ($h = \frac{1}{2}$), formula $\Phi_1(x, y)$ is symmetrical in y and $1-y$ and assumes its minimum at $y = \frac{1}{2}$, provided $y > x$. It is also interesting to notice that the formula in this case ($h = \frac{1}{2}$) does not depend on the sign of s and the sojourn time depends only on the value of $|s|$.

We shall next consider the second moment of the sojourn time. Let $\Phi_1^{(2)}(x, y)$ be the second moment of the number of generations for which a population starting from x and destined for fixation of A_1 stays at y. Let $\Psi^{(2)}(x, y) \equiv u(x) \, \Phi_1^{(2)}(x, y)$. Then $\Psi^{(2)}(x, y)$ is the solution of the differential equation

$$\frac{V_{\delta x}}{2} \frac{d^2 \Psi^{(2)}(x, y)}{dx^2} + M_{\delta x} \frac{d\Psi^{(2)}(x, y)}{dx} + 2\delta(x - y)\, \Psi(x, y) = 0 \tag{6}$$

with boundary condition $\Psi^{(2)}(0, y) = \Psi^{(2)}(1, y) = 0$, where $\Psi(x, y)$ is the function defined in (1) (cf. Dynkin, 1965, Chapters 10 and 13). The solution of this equation can be immediately integrated:

$$\Psi^{(2)}(x, y) = \frac{4\Psi(y, y)\, g(y, 1)\, g(0, x)}{G(y)\, V_{\delta y}\, g(0, 1)} \quad \text{for} \quad y > x \tag{7}$$

$$= \frac{4\Psi(y, y)}{G(y)\, V_{\delta y}} \left\{ \frac{g(y, 1)\, g(0, x)}{g(0, 1)} - g(y, x) \right\} \quad \text{for} \quad y < x,$$

and

$$\Phi_1^{(2)}(x, y) = \frac{\Psi^{(2)}(x, y)}{u(x)}.$$

If $h = \frac{1}{2}$,

$$\Phi_1^{(2)}(x, y) = \frac{8S(1-y)^2\, S(y)^2}{s^2 y^2 (1-y)^2\, S(1)^2} \quad \text{for} \quad y > x$$

$$= \frac{8S(y)^2\, S(1-y)}{s^2 y^2\, (1-y)^2\, S(1)^2\, S(x)} \left\{ \frac{S(1-y)\, S(x)}{S(1)} - S(1)\, S(x-y) \right\} \quad \text{for} \quad y < x,$$

and if $s = 0$,

$$\Phi_1^{(2)}(x, y) = 32N^2 \quad \text{for} \quad y > x$$

$$= 32N^2 \frac{y}{x} \quad \text{for} \quad y < x.$$

The quantity $\Phi_1^{(2)}(x, y)/(2N)^2$ gives the second moment of the sojourn time at gene frequency y $(= 1/2N, 2/2N, \ldots, (2N-1)/2N)$, provided that the allele starts with frequency x and is eventually fixed. In the case of a neutral mutant, the average sojourn time in generations at each of the intermediate frequencies is two and its variance is four. If $h = \frac{1}{2}$, formula $\Phi_1^{(2)}(x, y)$ is symmetrical in y and $1-y$, and does not depend on the sign of s, as in $\Phi_1(x, y)$. In this case it is also interesting to note that the variance of the sojourn time at each intermediate frequency is exactly
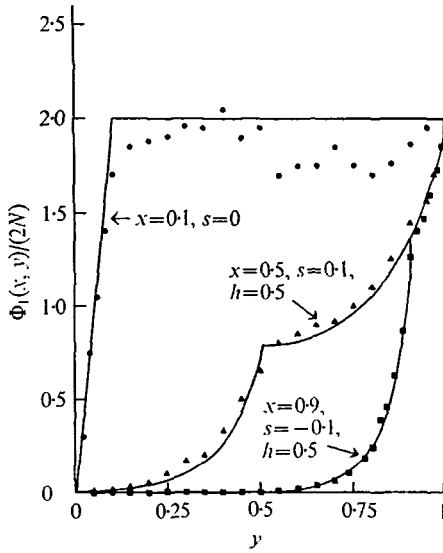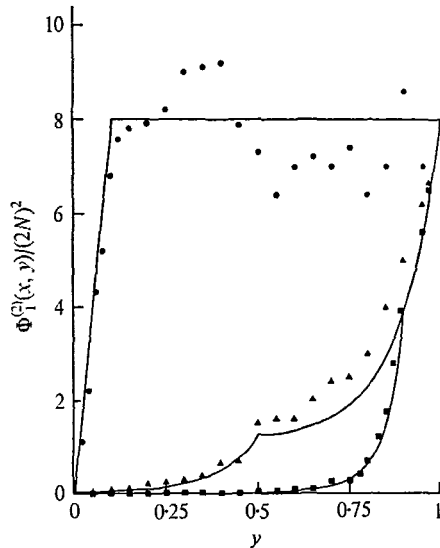
Fig. 1. The relationship between the initial frequency x and the sojourn time at y provided that the allele is eventually fixed in the population. Curves indicate the theoretical expectations and points (circles, triangles or squares) indicate the simulation results. The ordinate is $\Phi_1(x, y)/2N$. In all the three cases, $N = 50$, and each dot indicates the average of about 1000 repetitions.

Fig. 2. The second moment of the sojourn time taken from the same simulations presented in Fig. 1.

the square of the mean provided $y > x$. In other words, the mean and the standard deviation are the same.

Incidentally, if we replace the $\Psi(x, y)$ in (6) by the solution of

$$\frac{V_{\delta x}}{2}\frac{d^2\Psi(x, y)}{dx^2} + M_{\delta x}\frac{d\Psi(x, y)}{dx} + \delta(x - y) = 0,$$

we have the second moment of the sojourn time for populations of both kinds (i.e. eventual fixation or loss of A_1), and the solution is given by (7) with

$$\Psi(y, y) = \frac{2g(0, y)\, g(y, 1)}{G(y)\, V_{\delta y}g(0, 1)}.$$

In particular, if $h = \frac{1}{2}$,

$$\Psi^{(2)}(x, y) = \frac{8S(y)\, S(1-y)^2\, S(x)}{s^2 y^2 (1-y)^2\, S(1)^2} \quad \text{for} \quad y > x$$

$$= \frac{8S(y)\, S(1-y)}{s^2 y^2 (1-y)^2\, S(1)}\left\{\frac{S(1-y)\, S(x)}{S(1)} - S(x-y)\right\} \quad \text{for} \quad y < x;$$

if $s = 0$

$$\Psi^{(2)}(x, y) = \frac{32N^2 x}{y} \quad \text{for} \quad y > x$$

$$= \frac{32N^2(1-x)}{1-y} \quad \text{for} \quad y < x.$$

In order to show the validity of the above formulae, I have carried out several computer simulations based on the discrete generation model of finite size. The agreement of the results with theoretical expectations was satisfactory. Three examples are presented in Figs. 1 and 2. In the figures, the curves indicate the theoretical expectations and the dots indicate the simulation results. Fig. 1 presents the mean $\Phi_1(x, y)/2N$ with initial frequencies $x = 0\cdot1$, $x = 0\cdot5$ and $x = 0\cdot9$, and the corresponding second moments $\Phi_1^{(2)}(x, y)/(2N)^2$ are presented in Fig. 2.

REFERENCES

CROW, J. F. & KIMURA, M. (1970). *An Introduction to Population Genetics Theory.* New York: Harper and Row.

DYNKIN, E. B. (1965). *Markov Processes.* Berlin: Springer-Verlag.

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

KIMURA, M. (1969). The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. *Genetics* **61**, 893–903.

KIMURA, M. (1970). The length of time required for a selectively neutral mutant to reach fixation through random frequency drift in a finite population. *Genetical Research* **15**, 131–133.

KIMURA, M. & T. OHTA (1969a). The average number of generations until fixation of a mutant gene in a finite population. *Genetics* **61**, 763–771.

KIMURA, M. & T. OHTA (1969b). The average number of generations until extinction of an individual mutant gene in a finite population. *Genetics* **63**, 701–709.

MARUYAMA, T. & M. KIMURA (1971). Some methods for treating continuous stochastic processes in population genetics. *Japanese Journal of Genetics* **46**, 407–410.

NARAIN, P. (1970). A note on the diffusion approximation for the variance of the number of generations until fixation of a neutral mutant gene. *Genetical Research*, **15**, 251–255.