SHORT COMMUNICATION

SOME METHODS FOR TREATING CONTINUOUS STOCHASTIC PROCESSES IN POPULATION GENETICS¹⁾

TAKEO MARUYAMA AND MOTOO KIMURA

National Institute of Genetics, Misima 411

Received October 21, 1971

The theory of stochastic processes of the change of gene frequencies in finite populations is occupying an increasingly important place in the mathematical theory of population genetics. It is particularly important for our consideration of the mechanism of evolution at the molecular level (Kimura 1971). The most powerful method so far developed for treating such processes is the method of diffusion equations or "diffusion models" (Kimura 1964; see also Wright 1969, chapters 13 and 14; Crow and Kimura 1970, chapters 8 and 9).

In the present paper, we consider a single locus with a pair of alleles A and A' segregating in a finite Mendelian population. We shall denote the frequency of A by x and that of its allele (A') by $1-x$. We assume no mutations so that $x=0$ and $x=1$ act as "absorbing barriers". In what follows, we regard the stochastic process in the change of gene frequency as a collection of sample paths $\{\omega\}$ and denote by $x(\omega, t)$ the position of a particular path ω at time t . Let $x(\omega, 0) \equiv p$ be the position where the path starts at time 0. We denote by $\tau(\omega)$ the time when the path ω exits from the interval $(0, 1)$. Note that the exit time $\tau(\omega)$ depends on the path ω . Since we assume no mutations, $\tau(\omega)$ is always finite, and $x(\omega, \tau(\omega))=0$ or 1 , depending on the exit.

Now, consider an arbitrary function $f(x)$, and let

$$F^{(n)}(p) \equiv E \left\{ \left[\int_0^{\tau(\omega)} f(x(\omega, t)) dt \right]^n \right\}, \quad (1)$$

where $E\{ \}$ indicates the expectation with respect to the paths. Then $F^{(n)}(p)$ represents the expectation of the n -th moment of the sum of the quantity given by $f(x)$ from $t=0$ until the exit time, given that the initial gene frequency is p . For example, if we want to measure the total heterozygosity, we may put $f(x)=2x(1-x)$. This problem was investigated by Kimura (1969) for $n=1$. In the expression for $F^{(n)}(p)$ given by (1), the sum contains two kinds of paths, one going to $x=0$ and the other to $x=1$. We may however want to distinguish them and calculate such quantity only for those paths going to one destination, say, $x=1$. A typical example is the time until fixation as first

¹⁾ Contribution No. 855 from the National Institute of Genetics, Misima, Shizuoka-ken 411, Japan. Aided in part by a Grant-in-Aid from the Ministry of Education, Japan.

worked out by Kimura and Ohta (1969). Now let

$$F_1^{(n)}(p) \equiv u(p) E \left\{ \left[\int_0^{\tau(\omega)} f(x(\omega, t)) dt \right]^n \middle| x(\omega, \tau(\omega)) = 1 \right\}, \quad (2)$$

where $u(p)$ is the probability of ultimate fixation, that is, the probability of the path reaching $x=1$. The quantity of our particular interest is $F_1^{(n)}(p)/u(p)$ which represents the n -th moment of the sum of $f(x)$ under the condition that the path's ultimate destination is the fixation. The crucial point in the present treatment is that $F^{(n)}(p)$ and $F_1^{(n)}(p)$ satisfy appropriate differential equations to be discussed below. First, consider $F^{(1)}(p)$. It can be shown that it satisfies

$$LF^{(1)}(p) + f(p) = 0, \quad (3)$$

where

$$L = \frac{V_{\delta p}}{2} \frac{d^2}{dp^2} + M_{\delta p} \frac{d}{dp} \quad (4)$$

is a second order differential operator, and $M_{\delta p}$ and $V_{\delta p}$ are respectively the mean and variance of the change of gene frequency p per unit time (generation). In a typical situation, if A has selective advantage s in homozygotes and sh in heterozygotes over its allele A' , $M_{\delta p} = sp(1-p)\{h + (1-2h)p\}$, and if the "variance" effective size of the population is N_e , $V_{\delta p} = p(1-p)/(2N_e)$. The appropriate boundary conditions for equation (3) are $F^{(1)}(0) = F^{(1)}(1) = 0$.

Similarly, it can be shown that $F^{(n)}(p)$ for $n \geq 2$ satisfies the differential equation

$$LF^{(n)}(p) + nf(p)F^{(n-1)}(p) = 0 \quad (5)$$

with boundary conditions $F^{(n)}(0) = F^{(n)}(1) = 0$. A rigorous proof of these theorems can be found in Dynkin (1965, chapter 13). Second, consider $F_1^{(1)}(p)$ of (2). A slight modification of Dynkin's method proves that $F_1^{(1)}(p)$ satisfies

$$LF_1^{(1)}(p) + u(p)f(p) = 0 \quad (6)$$

with $F_1^{(1)}(0) = F_1^{(1)}(1) = 0$, and more generally $F_1^{(n)}(p)$, $n \geq 2$, satisfies

$$LF_1^{(n)}(p) + nf(p)F_1^{(n-1)}(p) = 0, \quad (7)$$

with boundary conditions $F_1^{(n)}(0) = F_1^{(n)}(1) = 0$. Finally, the fixation probability $u(p)$ satisfies

$$Lu(p) = 0 \quad (8)$$

with boundary conditions $u(0) = 0$, $u(1) = 1$ (cf. Kimura 1962).

In general, we are concerned with an ordinary second order differential equation of the form

$$LY(p) + K(p) = 0, \quad (9)$$

where $K(p)$ is a given function of p and $Y(p)$ satisfies the boundary conditions

$$Y(0) = Y(1) = 0. \quad (10)$$

Equation (9) can readily be integrated and we have

$$Y(p) = \{1 - u(p)\} \int_0^p \phi_K(\xi) u(\xi) d\xi + u(p) \int_p^1 \phi_K(\xi) \{1 - u(\xi)\} d\xi, \quad (11)$$

where

$$\begin{aligned} \phi_K(\xi) &= 2K(\xi) \int_0^1 G(\lambda) d\lambda \Big/ \{V_{\delta\xi} G(\xi)\}, \\ u(p) &= \int_0^p G(\xi) d\xi \Big/ \int_0^1 G(\xi) d\xi \end{aligned}$$

and

$$G(\xi) = \exp \left\{ -2 \int_0^\xi (M_{\delta x} / V_{\delta x}) dx \right\}.$$

As an example, let us calculate the sum of heterozygosity for the paths leading to fixation (but excluding the paths leading to loss), which we denote by $H_1^{(1)}(p)$. This corresponds to $K(p) = u(p)f(p) = u(p)2p(1-p)$. To simplify calculations, we shall assume that A is selectively neutral ($s=0$). Then, $M_{\delta\xi} = 0$ and $V_{\delta\xi} = \xi(1-\xi)/(2N_e)$ so that $G(\xi) = 1$, $u(p) = p$, and $\phi_K(\xi) = 8N_e\xi$. Thus, from (11), we obtain $H_1^{(1)}(p) = F_1^{(1)}(p)/p = (4N_e/3)(1-p^2)$. Similarly, the expectation of the second moment of the sum of heterozygosity from $x=p$ until $x=1$ (denoted by $H_1^{(2)}(p)$) can be calculated. To check the validity of the formulae for $H_1^{(1)}(p)$ and $H_1^{(2)}(p)$, we performed Monte Carlo experiments using the computer TOSBAC 3400, and the agreement between theoretical predictions and the experimental results was satisfactory.

The above analyses can be generalized as follows. By choosing two arbitrary values $x_0 < x_1$ in $(0, 1)$ we ask what will be the total sum of a certain quantity before a path reaches either x_0 or x_1 for the first time. The answer to this problem is given by the solution of equation (3) and (5), with boundary conditions $F(x_0) = F(x_1) = 0$. Similarly, if we want to measure the quantity for those paths which reach x_1 before x_0 is reached, then the solution is given by equations (6) and (7), but with boundary conditions $F(x_0) = F(x_1) = 0$. In addition the *first exit* probability through x_1 is given by the solution of (8) with conditions $u(x_0) = 0$, $u(x_1) = 1$.

Finally, if we put $f(x) = \delta(x-y)$, where $\delta(\cdot)$ is the Dirac delta function, then, the unique solution, denoted by $\Phi(p, y)$ of equation (3) with boundary conditions $\Phi(0, y) = \Phi(1, y) = 0$ gives the total sojourn time at y for the paths starting from p . This is equivalent to "the transient function" $t(x)$ of Ewens (1969), and also to the steady flux distribution with $\nu_m = 1$ given by Kimura (1969). Similarly, let $\Phi_1(p, y)$ be the sojourn time at y of the paths starting from p and going to fixation (excluding the paths leading to extinction). Then, this is given by $F_1^{(1)}(p)/u(p)$ where $F_1^{(1)}(p)$ is the solution of (6) with $f(p) = \delta(p-y)$. Thus, using (11), we obtain

$$\left. \begin{aligned} \Phi_1(p, y) &= 2u(y)\{1 - u(y)\} / \{V_{\delta y} u'(y)\}, & (y \geq p), \\ \Phi_1(p, y) &= 2\{1 - u(p)\} u^2(y) / \{u(p) V_{\delta y} u'(y)\}, & (y < p), \end{aligned} \right\} \quad (12)$$

where $u'(y) = du(y)/dy$. In a special but important case in which a single mutant gene is destined to eventual fixation, the average number of generations which the mutant

allele spends in the interval between x and $x+dx$ before fixation is given by $\Phi_1(x)dx$, where

$$\Phi_1(x) = 2u(x)\{1-u(x)\} / \{V_{\delta x}u'(x)\}, \quad (x \geq 1/2N). \quad (13)$$

For a neutral mutant, this reduces to $\Phi_1(x) = 4N_e$, showing that the mutant spends on the average $2N_e/N$ generations at each frequency class until fixation in the population, where N is the actual size of the population.

The present method can also be extended to obtain the moment generating functions. For example, let

$$\phi(\lambda, p) \equiv \sum_{n=0}^{\infty} (\lambda^n/n!) F_1^{(n)}(p), \text{ then } \phi \text{ satisfies}$$

$$L\phi(\lambda, p) + \lambda\phi(\lambda, p)f(p) = 0$$

with condition $F_1^{(0)}(p) = u(p)$. In the special case of a selectively neutral mutant, we can obtain the moment generating function of the probability distribution of the time until fixation by putting $L = \{p(1-p)/4N_e\}d^2/dp^2$, $f(p) = 1$ and $u(p) = p$ in these equations.

LITERATURE CITED

- Crow, J. F., and M. Kimura, 1970 *An Introduction to Population Genetics Theory*. Harper and Row, New York.
- Dynkin, E. B., 1965 *Markov Processes*. Springer-Verlag, Berlin.
- Ewens, W. J., 1969 *Population Genetics*. Methuen, London.
- Kimura, M., 1962 On the probability of fixation of mutant genes in a population. *Genetics* **47**: 713-719.
- Kimura, M., 1964 Diffusion models in population genetics. *J. Appl. Probability* **1**: 177-232.
- 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., 1971 Theoretical foundation of population genetics at the molecular level. *Theoretical Population Biology* **2**: 174-208.
- Kimura, M., and T. Ohta, 1969 The average number of generations until fixation of a mutant gene in a finite population. *Genetics* **61**: 763-771.
- Wright, S., 1969 *Evolution and the Genetics of Populations*. Vol. 2. *The Theory of Gene Frequencies*. University of Chicago Press, Chicago.