

SHORT PAPERS

A note on the diffusion approximation for the variance of the number of generations until fixation of a neutral mutant gene

By P. NARAIN†

Institute of Animal Genetics, Edinburgh, U.K.

(Received 1 May 1969)

SUMMARY

A general expression is derived for the variance of time to fixation of a neutral gene in a finite population using a diffusion approximation. The results are compared with exact values derived by matrix methods for a population size of 8.

The average number of generations until fixation of a mutant gene in a finite population has recently been given by Kimura & Ohta (1969) with the help of the diffusion model, whereas Narain (1969) has studied the mean and the variance of the number of generations until fixation using transition matrices.

Kimura & Ohta (1969), however, did not derive an expression for the variance of the length of time until fixation, though they did mention that their method could be adapted to obtain the n th moment of the length of time until fixation in terms of the $(n-1)$ th moment. The object of this note is, therefore, to derive the variance of the time until fixation using the diffusion approximation.

Let us consider a mutant allele A_2 with frequency p (the normal allele A_1 being at frequency $1-p$) in a diploid population of N individuals with variance effective number N_e , which may differ from N if the mating is not random or if the distribution of the number of offspring does not follow a Poisson distribution. N_e is defined as the size of an idealized population that would have the same variance of change in gene frequency as the population under consideration (Kimura & Crow, 1963). Let $u(p, t)$ be the probability that allele A_2 gets fixed by the t th generation starting with frequency p at $t = 0$. Let

$$T_1(p) = \int_0^\infty t \frac{\partial u(p, t)}{\partial t} dt, \quad (1)$$

$$S_1(p) = \int_0^\infty t^2 \frac{\partial u(p, t)}{\partial t} dt. \quad (2)$$

Then

$$M_1(p) = T_1(p)/u(p), \quad (3)$$

$$V_1(p) = S_1(p)/u(p), \quad (4)$$

represent respectively the average and the second moment about the origin of the length of time until the mutant A_2 becomes fixed in the population, excluding the cases in which it is lost from it. Here $u(p)$ is the probability of ultimate fixation such that

$$u(p) = \lim_{t \rightarrow \infty} u(p, t). \quad (5)$$

† On leave from the Institute of Agricultural Research Statistics, (I.C.A.R.), Library Avenue, New Delhi (India).

If $M_{\delta p}$ and $V_{\delta p}$ represent the mean and the variance of the rate of change in the frequency of A_2 per generation, then following Kimura (1962), $u(p, t)$ satisfies the Kolmogorov backward equation

$$\frac{\partial u(p, t)}{\partial t} = \frac{1}{2}V_{\delta p} \frac{\partial^2 u(p, t)}{\partial p^2} + M_{\delta p} \frac{\partial u(p, t)}{\partial p}. \tag{6}$$

Following the technique of Kimura & Ohta (1969), the set of differential equations for $T_1(p)$ and $S_1(p)$ are respectively given by

$$\frac{1}{2}V_{\delta p} \frac{d^2 T_1(p)}{dp^2} + M_{\delta p} \frac{dT_1(p)}{dp} + u(p) = 0, \tag{7}$$

$$\frac{1}{2}V_{\delta p} \frac{d^2 S_1(p)}{dp^2} + M_{\delta p} \frac{dS_1(p)}{dp} + 2T_1(p) = 0. \tag{8}$$

The former differential equation has been derived and solved by Kimura & Ohta (1969) with boundary conditions

$$\lim_{p \rightarrow 0} T_1(p) = \text{finite}, \tag{9}$$

$$T_1(1) = 0. \tag{10}$$

The boundary condition (9) means that in a finite population a single mutant gene which appeared in the population reaches fixation within a finite time, whereas (10) is obvious. If we transform (7) into a differential equation for $M_1(p)$ by differentiating

$$T_1(p) = M_1(p) \cdot u(p)$$

twice and substituting in (7), we get

$$\frac{1}{2}V_{\delta p} \frac{d^2 M_1(p)}{dp^2} + \left[M_{\delta p} + \frac{V_{\delta p} G(p)}{u(p)} \right] \frac{dM_1(p)}{dp} + 1 = 0, \tag{11}$$

$$G(p) = \frac{du(p)}{dp}. \tag{12}$$

Since $\lim_{p \rightarrow 0} u(p)$ is finite and $u(1) = 1$, the boundary conditions for the differential equation in $M_1(p)$ are

$$\lim_{p \rightarrow \infty} M_1(p) = K_1, \tag{13}$$

where K_1 is a finite quantity $M_1(1) = 0. \tag{14}$

In the case of random drift alone, we have

$$M_{\delta p} = 0, \tag{15}$$

$$V_{\delta p} = p(1-p)/2N_e, \tag{16}$$

$$u(p) = p, \tag{17}$$

$$G(p) = 1. \tag{18}$$

The differential equation (11) then reduces to

$$\frac{d^2 M_1(p)}{dp^2} + \frac{2}{p} \frac{dM_1(p)}{dp} + \frac{4N_e}{p(1-p)} = 0. \tag{19}$$

The solution of this differential equation gives the same result as given by Kimura & Ohta (1969) for a selectively neutral gene. In particular

$$K_1 = 4N_e, \tag{20}$$

$$M_1(p) = -4N_e \left(\frac{1-p}{p} \right) \log_e(1-p). \tag{21}$$

For the second moment about the origin of the length of time until fixation of A_2 , we transform (8) into a differential equation for $V_1(p)$ by differentiating $S_1(p) = V_1(p) \cdot u(p)$ twice and substituting in (8). This gives

$$\frac{1}{2} V_{\delta p} \frac{d^2 V_1(p)}{dp^2} + \left[M_{\delta p} + \frac{V_{\delta p} G(p)}{u(p)} \right] \frac{dV_1(p)}{dp} + 2M_1(p) = 0. \tag{22}$$

It may be noted here that (11) and (22) are similar to the set of differential equations respectively for the mean and the second moment (about the origin) of the length of time until homozygosity (Watterson, 1961, 1962) with the difference that $M_{\delta p}$ has been replaced here by $[M_{\delta p} + (V_{\delta p} G(p))/u(p)]$ as for (7) and (11). The boundary conditions to be imposed are, following the same arguments,

$$\lim_{p \rightarrow 0} V_1(p) = K_2, \tag{23}$$

where K_2 is a finite quantity $V_1(1) = 0. \tag{24}$

In the case of random drift, we apply (15) to (18). The differential equation (22), then, reduces to

$$\frac{d^2 V_1(p)}{dp^2} + \frac{2dV_1(p)}{p dp} + \frac{8N_e M_1(p)}{p(1-p)} = 0. \tag{25}$$

The solution of (25), after substituting for $M_1(p)$, is given by

$$V_1(p) = B - \frac{A}{p} - 32N_e^2 \left[\left(1 - \frac{1}{p} - \log_e p \right) \log_e(1-p) - F(p) \right], \tag{26}$$

where A and B are constants of integration and $F(p)$ is given by

$$F(p) = \int \frac{\log_e p}{1-p} dp. \tag{27}$$

Using the boundary conditions, we get

$$B = A + 32N_e^2 F(1), \tag{28}$$

$$K_2 = A - 32N_e^2 \left[1 + \int_0^1 \frac{\log_e p}{1-p} dp \right]. \tag{29}$$

Thus $V_1(p)$ is given by

$$V_1(p) = 32N_e^2 \left[- \int_p^1 \frac{\log_e p}{1-p} dp - \left(1 - \frac{1}{p} - \log_e p \right) \log_e(1-p) \right] + \left(1 - \frac{1}{p} \right) \left[K_2 + 32N_e^2 \left(\int_0^1 \frac{\log_e p}{1-p} dp + 1 \right) \right]. \tag{30}$$

Since $V_1(p)$ is finite as $p \rightarrow 0$, the terms inside the second bracket with the factor $[1 - (1/p)]$ in (30) must vanish. This means

$$K_2 = -32N_e^2 \left[\int_0^1 \frac{\log_e p}{1-p} dp + 1 \right]. \tag{31}$$

We then get

$$V_1(p) = 32N_e^2 \left[- \left(1 - \frac{1}{p} \right) \log_e(1-p) + (\log_e p)(\log_e \overline{1-p}) - \int_p^1 \frac{\log_e p}{1-p} dp \right]. \tag{32}$$

Using the results on dilogarithms given in Abramowitz & Stegun (1965), (32) reduces to

$$V_1(p) = 32N_e^2 \left[\frac{1-p}{p} \log_e(1-p) + \frac{\pi^2}{6} - \sum_{k=1}^{\infty} p^k / k^2 \right]. \tag{33}$$

The variance is then given by

$$\begin{aligned} V &= V_1(p) - [M_1(p)]^2 \\ &= 32N_e^2 \left[\frac{\pi^2}{6} + \left(\frac{1-p}{p} \right) \log_e(1-p) \left\{ 1 - \frac{1-p}{2p} \log_e(1-p) \right\} - \sum_{k=1}^{\infty} p^k / k^2 \right] \end{aligned} \tag{34}$$

and the coefficient of variation is found to be independent of the effective population size.

Also, in the limit, when $p \rightarrow 0$, it follows from Abramowitz & Stegun (1965) that

$$\left. \begin{aligned} \lim_{p \rightarrow 0} V_1(p) &= K_2 \\ &= -32N_e^2 \left[\int_0^1 \frac{\log_e p}{1-p} dp + 1 \right] \\ &= 32N_e^2 [(\pi^2/6) - 1]. \end{aligned} \right\} \tag{35}$$

The variance, in this case, is

$$\left. \begin{aligned} V &= K_2 - K_1^2 \\ &= 16N_e^2 [(\pi^2/3) - 3] \\ &= 4.64N_e^2, \end{aligned} \right\} \tag{36}$$

giving a coefficient of variation of about 54 %.

DISCUSSION

It is apparent from the preceding derivations that an originally rare neutral mutant gene in a population of effective size N_e takes about $4N_e$ generations on an average, with a standard deviation of about $2N_e$ generations, until it spreads in the whole population. According to Kimura & Ohta (1969) neutral mutation and random drift are of fundamental importance in determining the genetic structure of Mendelian populations. The time that is required for establishing the mutant may average four times the effective population size but may also vary considerably.

Using computer results based on transition matrices it was shown in Narain (1969) that the coefficient of variation of 54 % is obtained when p tends to zero and that it is the minimum possible. It increases as the gene frequency increases. The standard deviation, however, decreases as the gene frequency increases. The following table shows the values of the mean and the standard deviation of time until fixation of A_2 as calculated from (21) and (39) and expressed as multiples of the effective population size for various initial frequencies of A_2 .

It is apparent that a larger mean is associated with a larger standard deviation.

The diffusion approximations to the mean and the standard deviations given in Table 1 have been compared with the exact values obtained by the transition matrix approach as developed in Narain (1969). The comparisons are shown in Table 2 for a population of size 8.

Table 1. Mean and standard deviation of time until fixation of mutant A_2 in terms of the effective population size

p	$M_1(p)$	$\sqrt{[V(p)]}$
Tending to zero	4.00	2.1536
0.125	3.74	2.1489
0.250	3.45	2.1302
0.375	3.13	2.0947
0.500	2.77	2.0342
0.625	2.35	1.9344
0.750	1.85	1.7678
0.875	1.19	1.4502
1.000	0.00	0.0000

Table 2. Exact values and diffusion approximation (D.A.) for mean and standard deviation of the number of generations until fixation of A_2 ($N_e = 8$)

p	Mean			Standard deviation		
	Exact	D.A.	Difference	Exact	D.A.	Difference
0.125	28.30	29.92	1.62	16.58	17.19	0.61
0.250	26.06	27.60	1.54	16.46	17.04	0.58
0.375	23.57	25.04	1.47	16.19	16.76	0.57
0.500	20.76	22.16	1.40	15.73	16.27	0.54
0.625	17.50	18.80	1.30	14.96	15.48	0.52
0.750	13.60	14.80	1.20	13.65	14.14	0.49
0.875	8.55	9.52	0.97	11.14	11.60	0.46

It is quite clear from Table 2 that the diffusion approximations overestimate both the mean and the standard deviation. However, while the overestimation for the mean is, on an average, about one generation, it is only about half-generation for the standard deviation. The former observation is consistent with that obtained by Ewens (1963) in regard to the transition matrix results and diffusion approximation for the mean time until homozygosity. He also observed that the mean error in diffusion approximations is approximately unity.

The author is grateful to Drs M. Kimura and T. Ohta for permission to read the manuscript of their paper before publication and to Professor A. Robertson, and Dr W. G. Hill for valuable discussions.

REFERENCES

- ABRAMOWITZ, M. & STEGUN, I. A. Ed. (1965). Handbook of Mathematical Functions. Dover Publications, New York.
- EWENS, W. J. (1963). Numerical results and diffusion approximations in a genetic process. *Biometrika* **50**, 241-249.
- KIMURA, M. (1962). On the probability of fixation of mutant genes in a population. *Genetics* **47**, 713-719.
- KIMURA, M. & CROW, J. F. (1963). The measurement of effective population number. *Evolution* **17**, 279-288.
- KIMURA, M. & OHTA, T. (1969). The average number of generations until fixation of a mutant gene in a finite population. *Genetics* **61**, 763-771.
- NARAIN, P. (1969). Response to selection in finite populations. Ph.D. thesis, Edinburgh University.
- WATTERSON, G. (1961). Markov chains with absorbing states (a genetic example). *Ann. Math. Statist.* **32**, 716-729.
- WATTERSON, G. (1962). Some theoretical aspects of diffusion theory in population genetics. *Ann. Math. Statist.* **33**, 939-957.