Theoretical and Applied Genetics 44, 247–254 (1974)

Ultimate Probability of Fixation and Time to Fixation or Loss of a Gene under a Variable Fitness Model

R. F. NASSAR and R. D. COOK

Department of Statistics, Kansas State University, Manhattan, and School of Statistics, University of Minnesota, St. Paul (USA)

Summary. We investigated the time to fixation or loss and the probability of fixation of a gene under a discrete diploid model of variable fitness.

Data were presented on the effects of the level of dominance, relative magnitude of the variance in fitness of the three genotypes, and selection intensity on the probability of fixation and time to absorption of a gene. The relationship of our model in the haploid case to that of Kimura and Ohta was discussed.

1. Introduction

The probability of fixation of a gene and its time to fixation or loss in a population of small size have drawn considerable attention from population geneticists because of the significance of those factors in natural and artificial selection. Kimura (1962) was the first to formulate a stochastic model that took into consideration random sampling and random fluctuation in the selection intensity of a gene. For his model Kimura derived a general expression for the ultimate probability of fixation of a mutant gene. The formula was applied for limited cases by Kimura (1962). While our work was in preparation, Ohta (1972) applied Kimura's formula for cases of genic selection where the selection intensity of a mutant fluctuates at random from generation to generation. No work has been done on variable selection in diploids; hence, in this paper, we shall examine the effect of fluctuations in fitness on the probability of fixation and the time to fixation or loss of a gene under a more general case of zygotic selection.

2. Theory and Methods

Consider a single locus with two alleles (A, a) in a population of N diploid individuals that mate entirely at random. Assume that the population size is constant and that, aside from selection, random fluctuations in the selection intensity and random drift are the only pressures present in the population causing gene frequency to change.

Let X_1 denote the frequency of allele A in an infinite (or potentially infinite) population of newly formed zygotes. Let the fitness of the three genotypes AA, Aa and aa be W_1, W_2 and W_3, respectively. Without loss of generality, we take W_1, W_2, and W_3 to be $1 + S_1$, $1 + S_2$ and $1 + S_3$. S_i $(i = 1, 2, 3)$ is a random variable with mean μ_{s_i} and variance v_{s_i}. The frequency of A in the infinite population after selection and before sampling of gametes is

$$X_2 = \frac{X_1^2(1+s_1) + X_1 Y_1 (1+s_2)}{X_1^2(1+s_1) + 2X_1 Y_1 (1+s_2) + Y_2^2 (1+s_3)} \tag{2.1}$$

where $Y_1 = 1 - X_1$. After selection has taken place the remaining zygotes reproduce at random and establish a proportion, x_3, of A alleles in the next generation. In each generation the quantity $2Nx_3$ is an observed value of a random variable Z with binomial probability function

$$P[Z = z] = \binom{2N}{z} x_2^z (1 - x_2)^{2N-z}, \quad z = 0, 1, \ldots, 2N. \tag{2.2}$$

In our model the change in gene frequency from generation to generation is, strictly speaking, Markovian. We will, however, use the diffusion approximation to obtain a solution. A diffusion solution, found to be very good even for a population size as small as 10 (Carr and Nassar, 1970a), has the advantage in this case of simplicity over a Markov chain solution.

Let

$U(x)$ = the ultimate probability that a gene whose initial frequency is x becomes fixed.

$E(x)$ = the mean time until fixation or loss (absorption) of a gene whose initial frequency is x.

$M(x)$ = mean change in gene frequency per generation.

$V(x)$ = variance of change in gene frequency per generation

then $\mu(x)$ and $E(x)$ satisfy the ordinary differential equations

$$\frac{V(x)}{2}\frac{d^2\mu(x)}{dx^2} + M(x)\frac{d\mu(x)}{dx} = 0 \tag{2.3}$$

and

$$\frac{V(x)}{2}\frac{d^2E(x)}{dx^2} + M(x)\frac{dE(x)}{dx} = -1 \tag{2.4}$$

with boundary conditions $\mu(0) = 0$, $\mu(1) = 1$ and $E(0) = E(1) = 0$, respectively (Kimura 1954, Feller 1954). Schematically, the change in gene frequency from any generation to the next may be represented in our model by

$$X_1 \xrightarrow{s} X_2 \xrightarrow{b} X_3$$

Contribution Number 191, Department of Statistics and Statistical Laboratory of the Kansas Agricultural Experiment Station, Kansas State University, Manhattan, Kansas 66506.

where s denotes the change from X_1 to the intermediate frequency X_2 due to selection and b denotes the change from X_2 to the next generation frequency due to binomial sampling.

Let

$\Delta t = X_3 - X_1$, $\Delta_{12} = X_2 - X_1$ and $\Delta_{23} = X_3 - X_2$. The mean change in gene frequency can now be represented by

$$\begin{aligned} M(x_1) &= E(\Delta t) = E(\Delta_{12} + \Delta_{23}) \\ &= E_s E_b (\Delta_{12} + \Delta_{23}|s) \\ &= E_s(\Delta_{12}) = E_s[(X_2) - X_1] \\ &= E_s\left[\frac{X_1(1-X_1)(S_1 X_1 + S_2(1-2X_1) - S_3(1-X_1)}{1 + S_1 X_1^2 + 2 S_2 X_1 (1-X_1) + S_3 (1-X_1)^2}\right] \end{aligned} \tag{2.5}$$

where E_s denotes the expectation with respect to the random variables S_1, S_2, and S_3 and E_b denotes the expectation with respect to the binomial probability function (2.2).

An analogous expression for $V(x)$ can be found by assuming that Δ_{12} and Δ_{23} are independent

$$\begin{aligned} V(x_1) &= Var(\Delta_{12}) + Var(\Delta_{23}) \\ &= Var(\Delta_{12}) + E_s[(X_1 + \Delta_{12})(1 - X_1 - \Delta_{12})/2N] \\ &= Var(\Delta_{12}) + \frac{X_1(1-X_1)}{2N} + \frac{1-2X_1}{2N} E_s(\Delta_{12}) - \\ &\quad - \frac{E_s(\Delta_{12}^2)}{2N} \\ &= Var(\Delta_{12}) + \frac{X_1(1-X_1)}{2N} + \frac{1-2X_1}{2N} E_s(\Delta_{12}) - \\ &\quad - \frac{1}{2N}[(E_s(\Delta_{12}))^2 + Var(\Delta_{12})]. \end{aligned} \tag{2.6}$$

Expressions (2.5) and (2.6) for the mean and variance of the change in gene frequency per generation are difficult to evaluate explicitly because we must obtain the expectation and variance of a ratio involving the random variables S_1, S_2 and S_3. We approximated $E_s(\Delta_{12})$ and $Var(\Delta_{12})$ in (2.5) and (2.6) by a Taylor series expansion to three and two terms, respectively. As thus we obtained

$$\begin{aligned} E_s(\Delta_{12}) &= E_s(H(s_1, s_2, s_3)) \\ &\cong H(\mu_{s_1}, \mu_{s_2}, \mu_{s_3}) \\ &\quad + \frac{1}{2}\sum_{i=1}^{3} \frac{\partial^2 H}{\partial s_i^2} V_{si} + \sum_{\substack{i,j=1 \\ i<j}}^{3} \frac{\partial^2 H}{\partial s_i \partial s_j} Cov_{s_i s_j} \end{aligned} \tag{2.7}$$

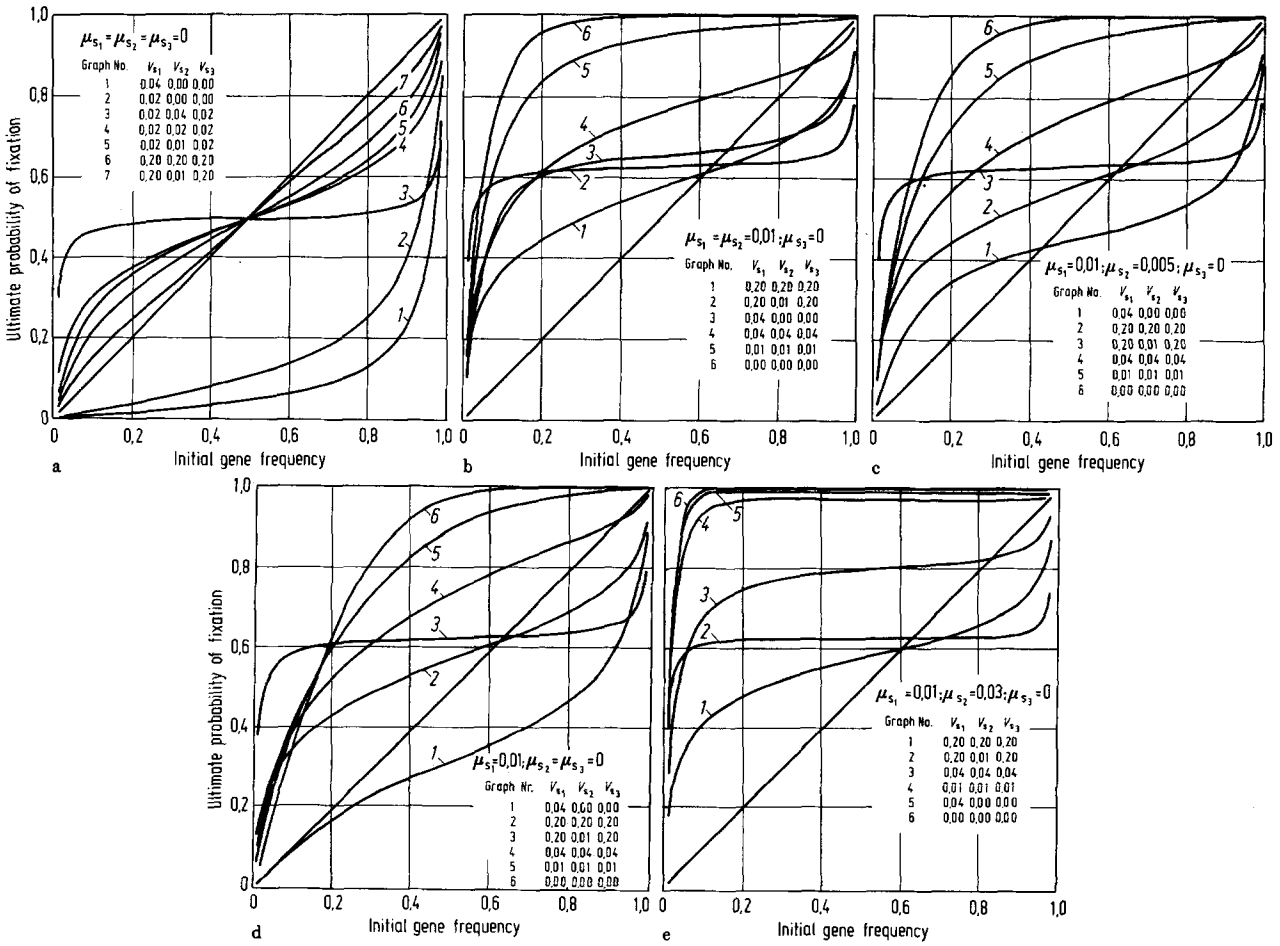

Fig. 1 a, b, c, d, e. Ultimate Probability of Fixation of a Gene as a Function of the Initial Gene Frequency. A Number for each Graph is Used to Designate the Value of the Variance of Fitness for each Genotype

and

$$Var\,(\Delta_{12}) = \sum_{i=1}^{3}\left(\frac{\partial H}{\partial s_i}\right)^2 V_{s_i} + 2\sum_{\substack{i,j=1\\ i<j}}^{3}\frac{\partial H}{\partial s_i}\frac{\partial H}{\partial s_j}\,Cov_{s_i s_j} \quad (2.8)$$

where the first and second partial derivatives are evaluated at $(\mu_{s_1}, \mu_{s_2}, \mu_{s_3})$. With (2.7) and (2.8) as approximate expressions for $M(x)$ and $V(x)$, the solutions to (2.3) and (2.4) were obtained numerically using a one-dimensional difference scheme. The derivatives were approximated using the centered-difference forms and the resulting tri-diagonal system of equations was solved at 100 points (for $N = 500$) in the interval [0, 1] using the Gaussian elimination method (Carnahan, Luther and Wilkes, 1969). For accuracy of the numerical solution we compared it with the exact solution using a finite Markov chain when selection was constant (Carr and Nassar, 1970a, b) and found it to be exact. The difference in values between the two solutions for the expected time to fixation or loss was 0.5 or less. For the probability of fixation, the two solutions were the same for at least two significant places. We will later show that the approximation employed in evaluation $M(x)$ and $V(x)$ had no effect at least on the qualitative nature of the results. We expect also that if the expected values and variance of S_1, S_2 and S_3 are small, as in this study, the effect on the quantitative nature of the results will be negligible.

Results in this paper will be based on the assumption that there is no covariance in fitness among the three genotypes. In other words $Cov_{s_i s_j} = 0$ in expressions (2.7) and (2.8). We will, however, examine later the general effect a covariance in fitness might have on these results.

3. Results

If we consider fig. 1 a, we see that if a gene's selective value fluctuates from generation to generation but is zero on the average ($V_{s_i} \neq 0$, $\mu_{s_i} = 0$), the gene will be selected for or against, depending on its initial frequency. For mutant and low frequency genes the result is that if the fitness of all three genotypes varies ($V_{s_i} \neq 0$), the ultimate probability of fixation is larger than the initial gene frequency, implying that the neutral gene becomes advantageous. For high frequency genes the reverse is true and the gene becomes disadvantageous as a result of fluctuations in the genotypic fitnesses. When $V_{s_1} = V_{s_2}$, a gene at initial gene frequency 1/2 will have an ultimate probability of fixation of 1/2; hence, it is truly a neutral gene. As V_{s_1} is increased relative to V_{s_2} and V_{s_3}, the point of neutrality (point of intersection with the diagonal line) slides back toward the origin. If the homozygote of interest is the only one with variable fitness ($V_{s_1} \neq 0$, $V_{s_2} = V_{s_3} = 0$), then the gene with the favorable fitness becomes disadvantageous at all initial frequencies (graphs 1, 2; fig. 1 a). This implies that if a selectively neutral mutant gene ($\mu_{s_i} = 0$) has a variable fitness while its allele has not, the mutant gene becomes in effect disadvantageous and its fixation probability becomes less than its initial gene frequency. From fig. 1 a it is also clear that relative to the case $V_{s_1} = V_{s_2} = V_{s_3}$, increasing the variance in fitness for the heterozygote over the two homozygotes tended to increase the probability of fixation of low frequency genes and decrease it for high frequency genes. Decreasing the variance of the heterozygotes relative to the two homozygotes tended to decrease the probability of fixation of low frequency genes and increase it for high frequency genes.

With selection ($\mu_{s_i} \neq 0$) the results (fig. 1 b, c, d, e) show that (aside from genes at very low frequency) regardless of the level of dominance the effect of a variance in genotypic fitness was to decrease the ultimate probability of fixation of a gene relative to the case of no variance ($V_{s_i} = 0, 0, 0$). With increased variance (V_{s_i}: (.04, .04, .04) or (.2, .2, .2)), the probability of fixation decreased further. Decreasing the variance of the heterozygote fitness relative to that of the two homozygotes ($V_{s_i} = .2, .01, .2$) tended to equalize the amount of reduction in the fixation probabilites over a wide range of initial gene frequencies.

The magnitude of the effect of a variance in fitness on the fixation probabilities varies with the level of dominance, as seen by comparing graphs of the same V_{s_i} values in b, c, d and e of fig. 1. A value of (.04, 0, 0), as an example, caused the largest reduction in the fixation probability for the recessive and additive cases, and the least reduction for the overdominant case. Similarly fluctuations in fitness (.01, .01, .01; .04, .04, .04) that are not of relatively large magnitudes had little effect on reducing probabilities of fixation of a strongly overdominant gene (fig. 1 e).

At low initial frequencies, as in the case of mutant genes, Table 1 shows that for overdominance there

Table 1. *Ultimate Probability of Fixation of a Gene for Different Selection Models with Varying Genotypic Fitness*

Initial gene frequency:						.01	.02	.03	.04
μ_{s_1}	μ_{s_2}	μ_{s_3}	V_{s_1}	V_{s_2}	V_{s_3}				
.01	.01	0	0	0	0	.169	.309	.424	.518
			.2	.2	.2	.145	.207	.247	.276
			.2	.01	.2	.396	.486	.527	.549
			.04	.04	.04	.153	.242	.303	.350
			.04	0	0	.105	.193	.264	.323
.01	.005	0	0	0	0	.094	.180	.258	.328
			.2	.2	.2	.137	.198	.237	.265
			.2	.01	.2	.385	.476	.518	.542
			.04	.04	.04	.124	.200	.255	.298
			.04	0	0	.037	.071	.102	.129
.01	0	0	0	0	0	.035	.071	.106	.141
			.2	.2	.2	.129	.188	.226	.254
			.2	.01	.2	.373	.466	.509	.534
			.04	.04	.04	.098	.162	.210	.248
			.04	0	0	.009	.019	.028	.037
.01	.03	0	0	0	0	.437	.680	.816	.893
			.2	.2	.2	.177	.247	.289	.319
			.2	.01	.2	.437	.520	.555	.574
			.04	.04	.04	.285	.416	.493	.545
			.04	0	0	.435	.676	.811	.888

was no increase in the fixation probability from (V_{s_i} = 0, 0, 0). For full dominance (μ_{s_i} = .01, .01, 0) there was an increase only for (V_{s_i} = .2, .01, .2). For additive and recessive cases, there was an increase for (V_{s_i} = .2, .2, .2), (V_{s_i} = .2, .01, .2) and (V_{s_i} = .04, .04, .04). Hence for advantageous mutant genes, fluctuations in fitness could increase the ultimate probability of fixation depending on the level of dominance.

Table 2. *Ultimate Probability of Fixation of a Deleterious Gene for an Additive and a Recessive Selection Model with Varying Genotypic Fitness*

Initial gene frequency:						.01	.02	.03	.04	.05
μ_{s_1}	μ_{s_2}	μ_{s_3}	V_{s_1}	V_{s_2}	V_{s_3}					
−.01	0	0	0	0	0	.000	.000	.000	.000	.000
			.02	.02	.02	.009	.016	.022	.027	.031
			.04	.04	.04	.027	.045	.059	.070	.079
			.04	.01	.04	.034	.055	.069	.080	.088
			.04	0	0	.000	.000	.000	.000	.000
			.02	0	0	.000	.000	.000	.000	.000
−.01	−.005	0	0	0	0	.000	.000	.000	.000	.000
			.02	.02	.02	.006	.012	.017	.021	.025
			.04	.04	.04	.022	.037	.049	.059	.068
			.04	.01	.04	.029	.047	.060	.071	.079
			.04	0	0	.000	.000	.000	.000	.000
			.02	0	0	.000	.000	.000	.000	.000

Table 2 shows that for a midly deleterious gene with or without a deleterious heterozygote effect (μ_{s_i}: −.01, 0, 0 and −.01, −.005, 0) a variable genotypic fitness (V_{s_i} = .02, .02, .02) had made the gene less deleterious. Increasing the variances further (V_{s_i} = .04, .04, .04) resulted in making a deleterious gene beneficial, especially if the heterozygote is more stable than either homozygotes in fitness (V_{s_i} = .04, .01, .04). If only the deleterious mutant homozygote is variable in fitness, there is very little effect on changing the fitness of a gene.

In figs. 2a, b, c and d, the ratio [$R(i)$] of the time to fixation or loss for selection and/or variable genotypic fitness to that with no selection and constant genotypic fitness is plotted against the initial gene frequency, i. If $R(i)$ was less than 1, there was acceleration in the time to fixation or loss compared to that of drift alone. A ratio larger than 1 implies retardation in the time to fixation or loss compared with drift. Fig. 1a shows that for a neutral gene (μ_{s_i}: 0, 0, 0) a fluctuation in genotypic fitness accelerated the time to fixation or loss over a large range of initial gene frequencies. For mutant genes, a fluctuation in genotypic fitness can prolong the time to fixation or loss of a gene (graphs 3, 5, 6, fig. 1a), provided the heterozygote is more stable than both homozygotes [(.02, .01, .02), (.2, .01, .2)]. For the case of selection (fig. 2b, c, d), one sees that, compared with a constant fitness (V_{s_i} = 0, 0, 0), two cases of variable fitness increased the time to fixation or loss. In those cases the heterozygote was more stable than either homozygote (V_{s_i} = .2, .01, .2) and the homozygote for the gene in question was the only one with variable fitness. For overdominance (Table 3), any variable fitness seemed to reduce substantially the time to fixation or loss as compared to constant fitness (V_{s_i} = 0, 0, 0).

4. Discussion

Our analysis of the time to fixation or loss and probability of fixation of a gene included all initial gene frequencies at an interval of 0.1 between 0 and 1. Population geneticists and evolutionists are interested in mutant genes that are usually at low initial frequency in the population. In addition to mutant genes, beneficial genes that are not of low frequency,

Table 3. *Ratio of the Time to Fixation or Loss of an Overdominant Gene (μ_{s_1} = .01, μ_{s_2} = .03, μ_{s_3} = 0) with Varying Genotypic Fitness to the Time to Fixation or Loss under Constant Genotypic Fitness and no Selection (Drift)*

(V_{s_1}, V_{s_2}, V_{s_3})	.01, .01, .01	.04, .04, .04	.2, .2, .2	.2, .01, .2	0, 0, 0	.04, 0, 0
Initial gene frequency						
.01	34.25	5.32	.82	20.29	669.58	140.42
.02	28.74	4.26	.62	13.37	576.62	120.86
.03	24.66	3.61	.51	10.24	497.21	104.18
.04	21.62	3.16	.44	8.43	433.55	90.82
.05	19.28	2.83	.40	7.24	383.18	80.26
.10	12.96	1.97	.28	4.55	245.94	51.50
.20	8.58	1.38	.20	2.99	159.69	33.43
.30	7.03	1.15	.17	2.45	130.56	27.31
.40	6.36	1.05	.16	2.23	118.39	24.75
.50	6.15	1.01	.16	2.16	114.90	23.99
.60	6.30	1.02	.16	2.23	118.30	24.66
.70	6.87	1.10	.17	2.45	130.29	27.08
.80	8.18	1.28	.20	2.98	158.63	32.81
.90	11.55	1.75	.27	4.51	235.62	48.92
.95	15.82	2.38	.38	7.11	332.06	73.87
.99	23.56	4.05	.76	19.26	459.41	147.46

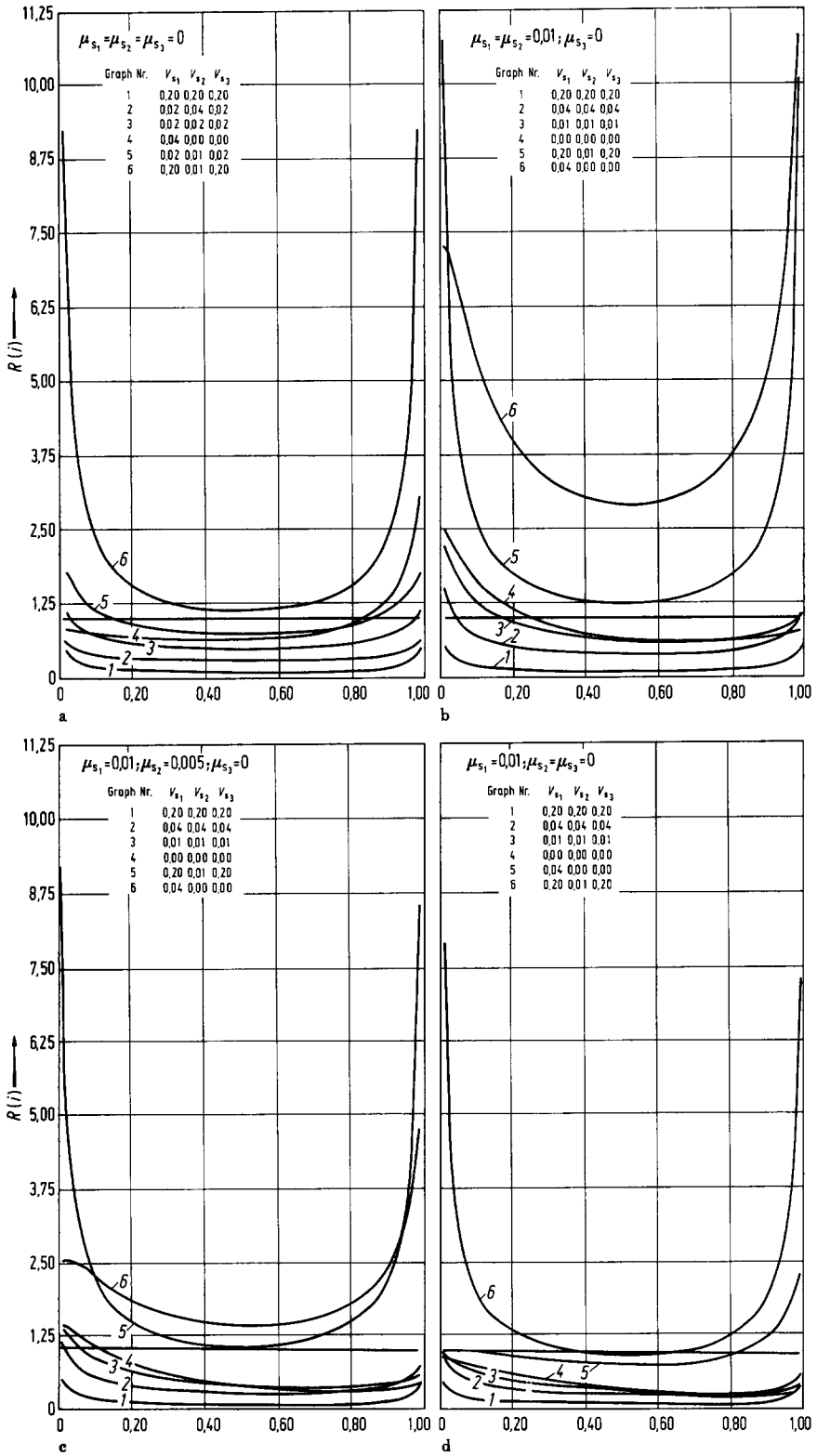

Fig. 2a, b, c, d. Time to Fixation or Loss Relative to that of Drift as a Function of the Initial Gene Frequency. The Ordinate, $R(i)$, is the Ratio of the Time to Fixation or Loss with Selection and Variable Fitness to that with no Selection and with Constant Fitness. A Number for each Graph is Used to Designate the Value of the Variance of Fitness for each Genotype

but rather in the intermediate range between 0 and 1 are of interest to quantitative geneticists and animal and plant breeders. Our results show that for such genes any sort of variance in fitness can be expected to reduce the net fitness of a gene. With regard to time to fixation or loss, which has a direct bearing on the maintenance of genetic variability in a line, our analysis indicates that if the fitness of the two homozygotes varies more than that of the heterozygote, time to fixation or loss for a gene (excluding overdominance) can be expected to be prolonged as compared with the case where no variance in fitness exists.

Also time to fixation or loss can be prolonged if the variance in fitness is only for the homozygote genotype of the gene in question. We believe that this latter type of situation is not likely to occur for intermediate and high frequency genes. Of biological significance, however, is the case where both homozygotes are variable in fitness and the heterozygote less variable than either homozygote. Other conditions of variable fitness for all three genotypes can only substantially accelerate a gene's time to fixation or loss.

It is interesting to note that as dominance increases to the point of strong overdominance, a variance in fitness can only lead to a substantial acceleration (from the case of constant fitness) in the time to fixation or loss.

With regard to neutral genes ($\mu_{s_i}=0$), should a situation arise where the two homozygotes have about equal variance in fitness, genes with initial frequency below 0.5 would become advantageous and those above 0.5 disadvantageous. Under a variable

fitness discrete model there is truly then no neutral genes in fitness except at one point in the initial gene frequency. That point of neutrality is determined by the relative magnitude of the variance in fitness of the two homozygotes (figure 1a).

For low frequency genes, such as mutants, we have shown that a selectively neutral gene ($\mu_{sj} = 0$) can become advantageous if both homozygotes have variable fitness and disadvantageous if only the mutant homozygote has a variable fitness. A slightly deleterious mutant might become near neutral or advantageous in fitness (depending on the magnitude of the variance in fitness to the mean fitness of a genotype) if both homozygotes are variable in fitness; but would remain deleterious if only the mutant homozygote is variable in fitness. A variance in fitness will not always shorten the average time of fixation or loss of a gene. If both homozygotes are variable in fitness and the heterozygote is less variable than either homozygote, the time to fixation or loss is prolonged considerably.

If we assume that the fitness of a mutant homozygote genotype fluctuates but not the fitnesses of the wild type homozygote and heterozygote, then we see that a beneficial mutant can be reduced in fitness to the point of neutrality or to being deleterious if there is weak to no dominance ($V_{s_1} = .04$, 0, 0: Table 1). On the other hand a deleterious gene (Table 2) might become neutral or beneficial. In other words, because of the asymmetry in the selection effect, the limiting fitness, as the variance in fitness increases relative to mean fitness, is not neutrality. Neutrality in fitness is only an outcome in a continuous array of possible fitness values that can arise.

In this light it is proper to ask whether a gene can be selectively neutral. For our model the answer is in the negative. A gene that is neutral in its average fitness will become deleterious or advantageous, depending on the mode of the variance in fitness. Slightly deleterious or advantageous genes on the other hand, might become neutral, but the likelihood is very small because neutrality is not a limiting case, and such genes might go beyond neutrality into the deleterious or advantageous range of fitness, depending on the magnitude of the variance in fitness to mean fitness and on the dominance level as shown by these results.

In view of the current "non-Darwinian" theory on molecular evolution it is relevant that mutant genes that are neutral in mean fitness will become advantageous or deleterious in effect if fitness is a random variable. The theory postulates that most amino acid substitutions have occurred as a result of random fixation caused by drift of selectively neutral mutations. The theoretical arguments in support of this postulate is based on constant fitness. For natural populations one can argue that genes that are neutral on the average will not behave as neutral genes under variable fitness. If one assumes that only the fitness of the mutant is variable, one arrives at the conclusion that most mutants are deleterious. This conclusion agrees with the current knowledge of mutants.

We might ask what effect does our approximations have on the quality of our results? We have approximated $E_s(\Delta_{12})$ and $Var(\Delta_{12})$ by a Taylor series expansion to three and two terms, respectively, thus ignoring terms of the order of $(s_i - \mu_{si})^2$ and higher. For the values of s_i and V_{s_i} used in this study, we think that the approximation cannot be so much in error as to effect the quantitative nature of our results. We have utilized two approaches to show that our results are at least qualitatively correct. In the first approach we simulated the model for the case of $\mu_{si} = 0$ with $V_{si} = 1.0$. We assumed that each s_i has a normal distribution with mean μ_{si} and variance V_{si}. The simulation, based on 300 replications, produced these results:

initial gene frequency:	.01	.10	.5	.9	.99
probability of fixation:	.105	.305	.53	.69	.88

The results agree with those obtained from the diffusion equation.

The second approach is exact and entails using the Jensen's inequality to predict the outcome of the solution of the diffusion equation for the ultimate probability of fixation. Jensen's inequality states that if a function, $H(s_1, s_2, s_3)$ in our case, is concave then

$$E_s[H(s_1, s_2, s_3)] < H(\mu_{s_1}, \mu_{s_2}, \mu_{s_3})$$

and if convex

$$E_s[H(s_1, s_2, s_3)] > H(\mu_{s_1}, \mu_{s_2}, \mu_{s_3}) .$$

In words, the expectation of a concave or convex function is less or greater than the function of the expectation. $H(s_1, s_2, s_3)$ is concave if and only if

$$SAS' < 0 \text{ for every } S = (s_1, s_2, s_3) \text{ vector}$$

and convex if

$$SAS' > 0 .$$

Here

$$A = \begin{bmatrix} \frac{\partial^2 H}{\partial s_1^2} & \frac{\partial^2 H}{\partial s_1 \partial s_2} & \frac{\partial^2 H}{\partial s_1 \partial s_3} \\ \frac{\partial^2 H}{\partial s_1 \partial s_2} & \frac{\partial^2 H}{\partial s_2^2} & \frac{\partial^2 H}{\partial s_2 \partial s_3} \\ \frac{\partial^2 H}{\partial s_1 \partial s_3} & \frac{\partial^2 H}{\partial s_2 \partial s_3} & \frac{\partial^2 H}{\partial s_3^2} \end{bmatrix}$$

is the matrix of second partial derivatives of the function with regard to s_1, s_2 and s_3.

We shall sight only two cases to demonstrate the point

1. $\mu_{s_1} = \mu_{s_2} = \mu_{s_3} = 0$ and $V_{s_1} \neq 0$, $V_{s_2} = V_{s_3} = 0$.
2. $\mu_{s_1} \neq 0$, $\mu_{s_2} = \mu_{s_3} = 0$ and $V_{s_1} \neq 0$, $V_{s_2} = V_{s_3} = 0$.

For the first case of neutral alleles, it can be shown that

$$SAS' = -2 s_1^2 x^4 (1 - x) < 0 ,$$

hence,

$$E_s[H(s_1, s_2, s_3)] < H(0, 0, 0) = 0 .$$

This predicts that the ultimate probability of fixation of a neutral gene is always less than its initial gene frequency, because selection on the average is against the gene. This prediction is born out by graphs 1 and 2 in fig. 1a. For the second case, $SAS' = -2 s_1^2 x^4 (1 - x)$ also. This implies that $E_s[H(s_1, s_2, s_3)] < H(\mu_{s_1}, 0, 0)$. Graph 1 of fig. 1d shows that the probability of fixation for $\mu_{s_1} = .01$, $\mu_{s_2} = \mu_{s_3} = 0$ and for $V_{s_1} = .04$, $V_{s_2} = V_{s_3} = 0$ is in fact less than that of the same selection model with no variance in fitness ($V_{s_i} = 0$).

All along, the results of the diffusion equation agree with the prediction whenever the function can be shown to be concave or convex. For some cases the function is neither concave nor convex (all those cases with $V_{s_i} \neq 0$, $i = 1, 2, 3$). Then, our results reflect this property (graphs 3, 4, 5, 6, 7 in fig. 1a).

Results of this study were for the case where there was zero correlation between the fitnesses of any two genotypes. We have results (unpublished) where the $Cov_{s_i s_j}$ in expressions (2.7) and (2.8) is such that the correlation between s_i and $s_j (r_{ij})$ ranges from -1 to 1. Those results indicate, as expected, that a positive correlation reduces the effect of a variance in fitness on the time to absorption and the probability of fixation of a gene. A negative correlation, on the other hand, increases it. In all cases, the correlation has to be relatively large ($r_{ij} > 1/2$ or $< 1/2$) before its impact can become significant.

Haploid

It is of interest to investigate the properties of our model for the haploid situation. In this case we let the absolute fitness of the two gametes or types (A_1, A_2) in the population be $W_1 = 1 + s_1$ and $W_2 = 1 + s_2$ where s_i $(i = 1, 2)$ is a random variable with mean μ_{s_i} and variance V_{s_i}. Our analysis of this case leads to results that are qualitatively similar to the diploid case. Table 4 shows that the effect of a variable fitness for both types in the population is to cause a gene of neutral fitness ($\mu_{s_i} = 0$) to become advantageous or disadvantageous, depending on its initial gene frequency. If, however, one type is variable in fitness ($V_{s_1} \neq 0$, $V_{s_2} = 0$), it will become disadvantageous for all initial gene frequencies (these results are comparable with those of fig. 1a). For a beneficial gene (Table 4) the results are similar to the diploid results (fig. 1b, c, d and e) in that a variance in fitness reduces the ultimate probability of fixation, and as the variance increases relative to the mean, the ultimate probability of fixation, for genes of high initial gene frequency, becomes less than that for a neutral gene under constant fitness. As Table 4 shows, variable fitness decreases substantially the expected time to fixation or loss.

We can use the Jensen's inequality to check on the nature of our results. In the case of $\mu_{s_1} = 0$, $\mu_{s_2} = 0$, $V_{s_1} \neq 0$, $V_{s_2} = 0$, the rate of change of gene frequency

Table 4. *Ultimate Probability of Fixation under Varying Fitness and the Ratio [R(i)] of the Time of Fixation or Loss with Varying Fitness and Selection to that under Constant Fitness and no Selection in a Haploid Population of Size 500*

μ_{s_1}, μ_{s_2}	0, 0	0, 0	0, 0	.01, 0	.01, 0	.01, 0	0, 0	0, 0	.01, 0	.01, 0	.01, 0
V_{s_1}, V_{s_2}	.1, .1	1, 1	.1, 0	0, 0	.1, .1	.1, 0	.1, .1	.1, 0	0, 0	.1, .1	.1, 0
Initial gene frequency									$R(i)$		
.01	.071	.124	.000	.094	.105	.000	.632	.294	1.444	.701	.343
.05	.194	.252	.001	.391	.271	.004	.375	.198	1.315	.403	.235
.10	.267	.314	.002	.629	.361	.008	.279	.161	1.125	.294	.192
.15	.314	.353	.004	.774	.417	.011	.233	.144	.963	.243	.172
.20	.351	.382	.006	.862	.458	.015	.207	.134	.832	.213	.160
.25	.381	.407	.008	.916	.492	.020	.190	.129	.729	.194	.154
.30	.408	.428	.010	.949	.520	.024	.179	.126	.647	.181	.151
.35	.433	.447	.012	.969	.546	.029	.171	.126	.582	.172	.150
.40	.456	.465	.015	.981	.569	.035	.166	.127	.531	.166	.151
.45	.478	.483	.019	.988	.592	.041	.163	.129	.489	.163	.153
.50	.500	.500	.023	.993	.613	.048	.162	.133	.457	.161	.158
.55	.521	.517	.028	.995	.634	.056	.163	.139	.430	.161	.164
.60	.543	.534	.034	.997	.654	.065	.166	.147	.410	.162	.173
.65	.566	.552	.041	.998	.676	.077	.171	.158	.394	.166	.185
.70	.591	.571	.051	.999	.691	.091	.179	.173	.383	.173	.201
.75	.618	.592	.064	.999	.721	.110	.190	.193	.376	.182	.224
.80	.648	.617	.084	.999	.747	.136	.207	.223	.374	.197	.256
.85	.685	.646	.114	.999	.778	.175	.233	.270	.379	.220	.306
.90	.733	.685	.168	.999	.816	.240	.279	.354	.392	.259	.394
.95	.805	.747	.297	1.000	.870	.379	.375	.544	.426	.343	.585
0.99	.928	.875	.685	1.000	.955	.741	.632	1.072	.505	.561	1.075

per generation is

$$H(s_1) = \frac{s_1 x_1 (1 - x_1)}{1 + x_1 s_1}$$

this function is concave since

$$H''(s_1) = \frac{-2 x_1^2 (1 - x_1)}{(1 + x_1 s_1)^3} < 0$$

hence,

$$E_{s_1}[H(s_1)] < H(\mu_{s_1}) = 0$$

this implies that selection, on the average, is against the A_1 allele and the ultimate probability of fixation ought to be less than its initial gene frequency. Results of Table 4 for $\mu_{s_1} = \mu_{s_2} = 0$, $V_{s_1} \neq 0$, $V_{s_2} = 0$ agree with the prediction.

Our results for the haploid case differ significantly from Ohta's results. For a mutant gene that is neutral on the average ($\mu_{s_1} = 0$), Ohta (1972) concluded that the expected time to fixation is equal to p the initial frequency of the mutant. We have shown that this is not so and that the ultimate probability of fixation will be larger or smaller than p depending on whether both genes have variable fitness or only the mutant is variable in fitness. The discrepancy in results seems to hinge on the level of approximation of $M(x_1)$, the mean rate of change in gene frequency per generation. Ohta approximated $M(x_1)$ by a linear function of fitness. In reality the mean change of gene frequency per generation is a non-linear function of fitness. This is so whether the model is discrete or continuous. To show that, we consider first the haploid model of this section.

Under a discrete generation model the rate of change of gene frequency per generation is

$$M(x_1) = \frac{x_1 (1 - x_1)(w_1 - w_2)}{x_1 w_1 + (1 - x_1) w_2} \tag{4.1}$$

expression (4.1) also holds for a continuous model. In the later case the rate of change in gene frequency is

$$\frac{dx_1}{dt} = x_1 (1 - x_1) \log \frac{w_1}{w_2} \tag{4.2}$$

where w_1 and w_2 are assumed to be constant in an interval of length t; and fluctuate at random from one period of length t to another. If x_1 is the frequency at the beginning of the time interval $(0, t)$, then x_t, the frequency at the end of that interval is

$$x_t = \int dx/dt = \frac{x_1 (w_1/w_2)^t}{x_1 (w_1/w_2)^t + (1 - x_1)}. \tag{4.3}$$

If t is taken to be one generation ($t = 1$), then the rate of change of gene frequency per generation reduces to that of (4.1).

When fitness is Malthusian (as in (4.2)), 4.1 can be approximated by a Taylor series expansion to give

$$M(x_1) = \mu_w x_1 (1 - x_1) + \frac{1}{2} \sigma_w^2 x_1 (1 - x_1)(1 - 2 x_1) \tag{4.4}$$

where

$$\mu_w = E\left[\log_e \frac{w_1}{w_2}\right]$$

and

$$\sigma_w^2 = V\left[\log_e \frac{w_1}{w_2}\right].$$

Ohta took $M(x_1)$ to be the first term in (4.4). This is seen to be inadequate because it does not reflect the non-linear property of the mean change of gene frequency per generation.

Literature

Carr, R. N., Nassar, R. F.: Effects of selection and drift on the dynamics of finite population I. Ultimate probability of fixation of a favorable allele. Biometrics **23**, 41–49 (1970a).

Carr, R. N., Nassar, R. F.: Effects of selection and drift on the dynamics of finite population II. Expected time to fixation or loss of an allele. Biometrics **26**, 221–227 (1970b).

Carnahan, B., Luther, H. A., Wilkes, H. A.: Applied Numerical Methods. New York: John Wiley and Sons 1969.

Feller, W.: Diffusion processes in one dimension. Trans. Amer. Math. Soc. **77**, 1–31 (1954).

Kimura, M.: Process leading to quasi-fixation of genes in natural population due to random fluctuations of selection intensities. Genetics **39**, 280–295 (1954).

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

Ohta, T.: Fixation probability of a mutant influenced by random fluctuation of selection intensity. Genet. Res. **19**, 33–38 (1972).

Received May 21, 1973
Communicated by W. Seyffert

R. F. Nassar
Associate Professor
Department of Statistics
Statistical Laboratory
Calvin Hall 19
Manhattan, Kansas 66506 (USA)

R. D. Cook
School of Statistics
University of Minnesota
St. Paul, Minnesota 55104 (USA)