Theoretical and Applied Genetics 39, 145–149 (1969)

Distribution of Gene Frequencies under the Case of Random Genetic Drift with and without Selection

RAJA F. NASSAR

Department of Statistics and Computer Science, Kansas State University, Manhattan, Kansas

Summary. Studies by computer simulation to determine applicability of the theory of the distribution of gene frequencies at the steady state of decay to small populations and the effect of linkage on the distribution revealed:

1. For random genetic drift with no selection theory and observations agree well for independent assortment and linkage.

2. For random genetic drift with selection theory and observations agree well for independent assortment and full dominance. Linkage with relatively large $N\,s$ decreased the mean gene frequency of unfixed classes and the steady rate of fixation of the favorable allele.

3. With independent assortment and additive gene effects agreement between theory and observations for the rates of fixation and loss at the terminal classes was good only for effective population number of at least 42. Small effective population numbers increased variance and decreased the steady rate of fixation and loss at the terminal classes. Linkage had no effect for small $N\,s$, but for relatively large $N\,s$, it decreased the mean of gene frequencies and increased the steady rate of fixation and loss at the terminal classes from that of independent assortment.

Two types of factors that produce random fluctuation in gene frequencies exist. The first is random sampling of gametes in reproduction; the second, random fluctuation in systematic pressures, where random fluctuation in selection intensity may be important. The random processes, alone or coupled with directed processes, operate to determine a probability distribution of gene frequencies (this paper is concerned with only the random drift process and selection as a directed process). The first attempt to determine distribution of gene frequencies was made by FISHER (1922, 1930) using the heat diffusion equation. He considered the random drift process, assuming no selection. WRIGHT (1931) studied the same problem using the method of integral equations. He has since published several papers on the probability distribution, using the Fokker-Plank equation. KIMURA (1955a, 1955b, 1957), using a continuous Markov process in time and space (gene frequency), arrived at explicit solutions for the distribution of gene frequencies under the case of random genetic drift alone and with selection. KIMURA (1957) states that the continuous treatment is applicable if the population number N is large so $1/N$ is negligible compared with one. It is of interest, therefore, to investigate the applicability of the continuous treatment to small populations. Furthermore, since there is no general theory for linkage in the multi-locus case, what happens with linkage by comparison with the independent assortment for which theory is available is of importance.

The purpose of this study is twofold. To investigate by computer simulation a) the applicability of KIMURA's diffusion treatment to small populations and b) the effect of linkage on the distribution of gene frequencies.

Material and Methods

Distributions of gene frequencies of a single gene with two alleles under selection and random genetic drift were given by KIMURA (1955b, 1957) for the cases of no dominance and full dominance. Of particular interest here is the distribution of gene frequencies of unfixed classes in the state of steady decay. For no dominance the density function is

$$f(Z) = e^{c(1-z)}\,[f_0^0\,T_0^1(z) + f_2^0\,T_2^1(z) + f_4^0\,T_4^1(z) + \cdots] \quad (1)$$

where

$c = N\,s$

$z = 1 - 2\,x$; x = gene frequency of the favorable allele

f_n's are constants corresponding to the smallest eigen value λ_0.

$T_n(z)$'s are the Gegenbauer polynomials, expressed in terms of the hypergeometric function.

The steady rate of decay of the frequency distribution of unfixed classes is given by the smallest eigen value λ_0, which can be expressed as

$$4\,N\,\lambda_0 = c^2 - B_{1,0} \quad (2)$$

where

N is the effective population number and $B_{1,0}$'s are separation constants obtained from tables listed in STRATTON et al. (1941). For a slightly different way of presenting (2) based on the new table of spheroidal wave functions by STRATTON et al. (1956) see KIMURA (1964) p. 29.

The rate of fixation at the terminal class ($q = 1$) equals $\lambda_0\,\mu(q)$, $\mu(q)$ being the probability of fixation of the favorable allele as given by KIMURA.

For full dominance the density function of the distribution at the state of steady decay is

$$f(x) = e^{2\,c\,x(1-x/2)}\,w_0 \quad (3)$$

where

$$w_0 = d_0\,T_0^1(z) + d_1\,T_1^1(z) + d_2\,T_2^1(z) + d_3\,T_3^1(z) + \cdots.$$

The d_n's are given by KIMURA (1957). They can be evaluated in terms of d_0 and c for the smallest eigen value λ_0, which corresponds to the state of steady decay of the distribution.

The rate of fixation and loss at the terminal classes is given respectively by

$$f(1)/4N \quad \text{and} \quad f(0)/4N. \tag{4}$$

When c = 0 (no selection), the steady rate of fixation at the terminal classes is $1/2N$.

The densities in (1) and (3), expanded to the $T_4(Z)$ term, were adjusted over the range 0 to 1 (fixed classes excluded), by ROMBERG's numerical integration procedure (BAUER et al. 1936,) so the area under each curve was unity and the first two moments for the unfixed classes were evaluated. The rates of fixation and loss at the terminal classes were evaluated from (2) and (3). For the case of random genetic drift with no selection, the mean and variance of the distribution of gene frequencies at any time t were derived following the matrix approach used by A. ROBERTSON (1952). This gave a mean $\mu_1 t$ equal to q, the initial gene frequency in the population; and a variance

$$\sigma_t^2 = q\,(1-q)\left[1-\left(1-\frac{1}{2N}\right)^t\right].$$

The mean, variance and rates of fixation and loss at the terminal classes as evaluated for each distribution were compared with those obtained from simulation under independent assortment and linkage.

For a population under selection, the selection coefficient was calculated from HALDANE (1931), using the formula

$$s = \frac{2\mu k}{\sigma_p}$$

where $2\mu = 1 =$ the difference between the genotypic values of the two homozygotes at a locus.
$k =$ selection intensity in phenotypic standard deviation,
$\sigma_p =$ phenotypic standard deviation in the population in the first generation.
s, strictly speaking, is not the same for dominance and additive gene effects, since σ_p is different. However, here the difference was negligible because the genetic variance was a small fraction of the phenotypic variance. It therefore, was ignored. By the same argument any increase in s due to reduced genetic variance as a result of selection was not considered significant. The variance effective population number was calculated from KIMURA and CROW (1963)

$$N_e = \frac{(N_{t-1}-1)\bar{k}}{1+V_k/\bar{k}}$$

where
$\bar{k} =$ the mean progeny number per parent
and
$V_k =$ the variance in progeny number.

The effective population number, as defined, is for genes under no selection. It is realized that N e would probably decrease with directional selection. Since heritability is low, that was considered inconsequential.

The simulation program was written for the 1604 CDC Computer. There were 47 loci and two alleles per locus. An allele at a locus was assigned at random with an expected frequency of 0.5. Two cases were studied a) No linkage and b) All 47 loci linked on one chromosome of 50 map units with equal distances between loci. The genetic models were additive and complete dominance. For each model an individual phenotypic value in each generation was the sum of its 47 equivalent gene contributions plus or minus a random environmental effect. Populations of different sizes were generated and pair mating was at random with one sire mated to one dam. Selection was directional or random and the size of the population was kept constant. To illustrate, a 4—1—4 population (table 1) stands for four pair mating with 4 offsprings (20, 2 ♀) per mating. For the population size to remain constant, four males and four females are selected as parents of the next generation. A population under consideration was regarded as being initially a random sample from a hypothetically large population in linkage equilibrium. From the foregoing it is evident that the simulated cases agree well with that of KIMURA except for 1. linkage and 2. the small population numbers. Thus any deviations in the numerical results from that of KIMURA can be interpreted as being due to these two factors. The small population number, aside from giving rise to a discrete change in gene frequency, can cause random fluctuations in s from generation to generation as well as between loci within a generation. Also, the linkage disequilibrium arising from a small sample of parental genotypes will in the presence of dominance cause pseudo-overdominance. Selection is also likely to generate a slight negative linkage disequilibrium (NASSAR and COMSTOCK, 1966 and unpublished results). All these factors are inevitable consequences of any selection program in small populations; and taken together can cause deviations from KIMURA's assumption of a continuous change in gene frequency with a fixed s. Hence, any effect due to a small population number must be regarded as a combination of effects due to these inseparable factors.

The computer runs were carried to fixation for small populations and beyond 4 N generations for relatively large populations. The data gathered on each population included the mean and variance of gene frequency and the rate of fixation and loss at the terminal classes. The rate of fixation at the terminal class (1) at time t was computed as the difference between the numbers of loci fixed at that class in time t and t—1 over the number of loci that were segregating at time t—1. The rate of loss at the terminal class (0) was calculated in the same manner. The state of steady decay in generations was determined from regressing the rate of fixation or loss of

Table 1. *Simulated populations with their respective selection coefficients (s), variance effective population numbers (Ne) and generation interval at the state of steady decay. a = 0 is additive and a = 1 is complete dominance*

Population	Ne	s	Ns	a	Generations
4—1—4	10	.1034	1.034	0	15—35
				1	15—35
4—1—20	8.32	.243	2.02	0	10—20
				1	16—32
16—1—4	42.0	.0238	1.0	0	84—104
				1	84—104
4—1—4	10	0	0		20—30
4—1—40	8.15	0	0		16—26
8—1—4	20.66	0	0		40—50
16—1—4	42	0	0		84—94

a terminal class (the rate for each generation was an average of 20 replications) on generation number and finding the generations for which the linear regression coefficient was not significantly different from zero. Table 1 gives the generation interval at which the rates of fixation and loss were taken as constant by virtue of the zero regression coefficient. For populations with no selection this was beyond 2N generations in agreement with theoretical results by KIMURA (1955a) for an initial gene frequency of a half. In all cases about one half of the number of loci was fixed at the beginning of the interval of steady decay. The interval could have been extended to ultimate fixation, but to do so would have increased the standard error of the estimate. Consequently the interval was terminated when about 5 loci were still segregating.

For each population a statistic (except for variance of gene frequency under no selection) was an average of all generations in the state of steady decay and of 20 replications. For the variance of gene frequency under no selection, the average was over 20 replications at a single generation. In all cases the standard error of an estimate was based on 20 replications.

Results

For the case of random genetic drift and no selection theory and observations agreed well for the independent assortment, as shown in table 2. It is also seen that linkage caused no deviations from theoretical expectations. However, there was, in general, an increase in the variance of the rates of fixation at the terminal classes and of gene frequency as can be seen by comparing the standard errors for estimates under independent assortment and linkage for any one population (table 2).

The data in table 3 for random genetic drift and selection shows good agreement between theory and observations for independent assortment and full dominance except for the variance of gene frequency of 4−1−4, 4−1−20 and the mean of 4-1-20. All of those were less than their theoretical values. In the additive case, the 16−1−4 population agreed well with the theoretical rates of fixation at the terminal classes. All three populations showed an observed mean gene frequency less than the theoretical mean and a variance greater than the theoretical variance. For the 4−1−4 and 4−1−20 populations the rates of fixation at the terminal classes were considerably less than the theoretical rates.

For linkage and additive gene effect there was no significant change for $N\,s \doteq 1.0$ from that of independent assortment. For $N\,s \doteq 2.0$, however, mean gene frequency was significantly reduced and rate of fixation and loss at the terminal classes and of the variance of gene frequency were significantly increased (table 3). Considering full dominance and linkage,

Table 2. *Observed and expected values for the mean and variance of gene frequency (q) and rates of fixation at the terminal classes (1,0) for random genetic drift and no selection*

Population	$\bar{q}$	μ_q	Generation	S_q^2	σ_q^2	Observed rate of fixation		Expected rate of fixation	
						Class (1)	Class (0)	Class (1)	Class (0)
Independent Assortment									
4−1−4	.503±.014	.5	20	.157±.0034	.160	.026±.004	.026±.002	.025	.025
			25	.181±.0037	.180				
			30	.193±.0030	.196				
4−1−40	.540±.014	.5	16	.151±.0039	.159	.029±.0035	.029±.0026	.0306	.0306
			21	.176±.0032	.184				
			26	.195±.003	.202				
8−1−4	.491±.013	.5	40	.154±.0044	.156	.0096±.0012	.0111±.0001	.0121	.0121
			50	.171±.0045	.177				
			60	.191±.0041	.193				
16−1−4	.495±.013	.5	84	.158±.0030	.159	.0048±.00047	.00539±.00073	.00595	.00595
			104	.180±.0030	.178				
			124	.195±.0034	.193				
Linkage									
4−1−4	.485±.015		20	.159±.0072		.0228±.004	.0244±.0053		
			25	.173±.007					
			30	.188±.0069					
4−1−40	.525±.015		16	.154±.0063		.0248±.0044	.0213±.0039		
			21	.172±.0059					
			26	.191±.0063					
8−1−4	.505±.012		40	.151±.0048		.0092±.00156	.0098±.00215		
			50	.164±.0056					
			60	.185±.0039					
16−1−4	.492±.0097		84	.153±.0035		.0047±.00067	.0069±.0015		
			104	.175±.0046					
			124	.194±.0047					

Table 3. *Observed and expected values for the mean and variance of gene frequency (q) of unfixed classes and the rates of fixation at the terminal classes (1,0) for additive (a = 0) and dominance (a = 1) under random genetic drift with selection*

Population	a	$\bar{q}$	μ_q	S_q^2	σ_q^2	Observed rate of fixation		Expected rate of fixation	
						Class (1)	Class (0)	Class (1)	Class (0)
Independent Assortment									
4−1−4	0	.534±.0029	.665	.0711±.0087	.0693	.0404±.0031	.0123±.0017	.0522	0.194
	1	.578±.0116	.588	.0579±.0015	.0725	.0324±.0025	.0112±.0009	.0336	.0119
4−1−20	0	.661±.016	.790	.0597±.0029	.0424	.0627±.0042	.01012±.0015	.1369	.0185
	1	.629±.007	.668	.051±.0022	.0567	.0481±.0033	.00457±.00074	.0467	.00459
16−1−4	0	.588±.017	.664	.0797±.0040	.0695	.0113±.001	.0038±.0006	.0124	.0046
	1	.571±.0131	.587	.0732±.0031	.0726	.00719±.0008	.00217±.00054	.008	.00286
Linkage									
4−1−4	0	.552±.013		.0701±.004		.0348±.004	.0137±.0027		
	1	.541±.008		.0613±.0029		.0243±.0032	.0124±.0018		
4−1−20	0	.607±.0134		.0704±.004		.0772±.0033	.0249±.0030		
	1	.596±.0075		.0502±.0025		.0328±.0033	.0046±.0008		
16−1−4	0	.575±.0092		.0878±.0041		.0105±.0017	.00355±.00063		
	1	.569±.011		.0753±.0034		.0095±.0015	.0023±.0004		

there was for $N\,s \doteq 2$, a significant decrease in the mean gene frequency and steady rate of fixation of the favorable allele from that of independent assortment.

Discussion

In the case of independent assortment and dominance, KIMURA's continuous treatment is applicable to very small populations except for the variance of gene frequency. In the additive case the continuous treatment was applicable for the rates of fixation at the terminal classes for population size 42, still relatively small. The effect of small population size is to incresae variance and reduce mean gene frequency of unfixed classes and the steady rate of fixation and loss at the terminal classes as compared to theory. The larger the $N\,s$ value, the larger the discrepancy. Linkage seems not to disturb results from that of independent assortment for small N s ($N\,s \leq 1.0$). However, for relatively large N s the most important effect is an increase in rates of fixation and loss at the terminal classes with a greater proportionate increase in rate of loss. That means that time to fixation and probability of fixation of the favorable allele at the limit are decreased from that of independent assortment. For dominance and linkage compared with independent assortment the rate of fixation of the favorable allele is reduced, which implies an increase in time to fixation, and a reduction in probability of fixation at the limit. These findings are in accord with results by NASSAR and COMSTOCK (1966) in that in the multilocus case linkage with selections (N s $>$ 1) reduced the probability of fixation of the favorable allele in the limit for additive and dominance gene effects and increased the time to fixation for dominance while decreasing it for additive gene effects. HILL and ROBERTSON (1966) and LATTER (1966) have also found in the two locus case that linkage reduced the probability of fixation in the limit for additive genes.

Linkage caused no deviations from theoretical results for the case of drift and no selection. The effect of linkage seems to be in its interaction with selection. A relatively large N s is needed for the effect to be significant.

Zusammenfassung

Simulationsexperimente am Computer zur Untersuchung der Anwendbarkeit der Theorie der Verteilung von Genfrequenzen mit einer stetigen Verfallsrate auf eine kleine Population unter Berücksichtigung des Einflusses der Koppelung auf die Verteilung ergaben:

1) Bei zufälliger genetischer Drift ohne Selektion stimmen Theorie und Beobachtung sowohl für die Annahme freier Spaltung als auch für die der Koppelung gut überein.

2) Bei zufälliger genetischer Drift mit Selektion stimmen Theorie und Beobachtung für die Annahme freier Spaltung und vollständiger Dominanz gut überein. Bei relativ großen Werten von N s reduziert Koppelung die mittlere Genfrequenz unfixierter Klassen und die stetige Fixierungsrate begünstigter Allele.

3) Bei freier Spaltung und additiver Genwirkung war die Übereinstimmung zwischen Theorie und Beobachtung hinsichtlich der Fixierungsrate und des Verlustes in terminalen Klassen nur für eine effektive Populationsgröße von mindestens 42 Individien gut. Kleinere effektive Populationsgrößen steigern die Varianz und reduzieren sowohl die stetige Fixierungsrate als auch den Verlust in den terminalen Klassen. Bei kleinen Werten von N s hat Koppelung keine Wirkung, jedoch reduziert sie die mittlere Genfrequenz und erhöht die stetige Fixierungsrate und den Verlust in den terminalen Klassen im Vergleich zu freier Spaltung dann, wenn N s relativ groß ist.

Literature Cited

1. BAUER, F. L., H. RUTISHAUSER and E. STIEFEL: New aspects in numerical quadrature. Proceedings of Symposium in Applied Mathematics Vol. XV, 199–218 (1936). — 2. FISHER, R. A.: On the dominance ratio. Proc. Roy. Soc. Edin. **42**, 321–341 (1922). — 3. FISHER, R. A.: The distribution of gene ratios for rare mutations. Proc. Roy. Soc. Edin. **50**, 205–220 (1930). — 4. HALDANE, J. B. S.: A mathematical theory of natural and artificial selection. VII. Selection intensity as a function of mortality rate. Proc. Camb. Phil. Soc. **27**, 131–136 (1931). — 5. HILL, W. G., and A. ROBERTSON: The effect of linkage on limits to artificial selection. Genet. Res. **8**, 269–294 (1966). — KIMURA, M.: Solution of a process of random genetic drift with a continuous model. Proc. Nat. Acad. Sci. **41**, 144–150 (1955a). — 7. KIMURA, M.: Stochastic Processes and distribution of gene frequencies under natural selection. Cold Spring Harbor Symp. **20**, 33–53 (1955b). — 8. KIMURA, M.: Some problems of stochastic processes in genetics. Ann. Math. Stat. **28**, 882–901 (1957). — 9. KIMURA, M.: Diffusion models in population genetics. Methuen's Review Series, Vol. 2. London: Methuen and Co. Ltd. 1964. — 10. KIMURA, M., and J. F. CROW: The measurement of effective population number. Evolution **17**, 279–288 (1963). — 11. LATTER, B. D. H.: The interaction between effective population size and linkage intensity under artificial selection. Genet. Res. **7**, 313–323 (1966). — 12. NASSAR, R. F., and R. E. COMSTOCK: Simulation studies of the effects of linkage and selection on the probability of gene fixation in small populations. Biometrics **22**, 954–955 [abstract] (1966). — 13. ROBERTSON, A.: The effect of inbreeding on the variation due to recessive genes. Genetics **37**, 188–207 (1952). — 14. STRATTON, J. A., P. M. MORSE, L. J. CHU, and R. A. HUNTER: Elliptic cylinder and spheroidial wave functions. New York: John Wiley and Sons 1941. — 15. STRATTON, J. A., P. M. MORSE, L. J. CHU, J. D. C. LITTLE, and F. J. CORBATÓ: Spheroidal Wave Functions. New York: Technology Press of M.I.T. and John Wiley, 1956. — 16. WRIGHT, S.: Evolution in mendelian populations. Genetics **16**, 97–159 (1931).

Received November 6, 1968

Communicated by R. C. LEWONTIN, Chicago

Dr. RAJA F. NASSAR
Department of Statistics and Computer Science
Kansas State University
Manhatten, Kansas 66502 (USA)