PROCESS LEADING TO QUASI-FIXATION OF GENES IN NATURAL POPULATIONS DUE TO RANDOM FLUCTUATION OF SELECTION INTENSITIES * [1]

MOTOO KIMURA

Department of Animal Husbandry, Iowa State College, Ames, Iowa

Received September 7, 1953

AMONG factors that may produce random fluctuation of gene frequencies in natural populations, random sampling of gametes and random fluctuation of selection intensities may be especially important in relation to evolution. On the process of " random drift " that will be realized in finite populations due to the random sampling of gametes in reproduction, not only precise mathematical studies have been carried out (WRIGHT 1931, 1945; FISHER 1930; HALDANE 1939; FELLER 1950) but also several model experiments have been undertaken (cf. HOUSE 1953). Since 1931, WRIGHT has repeatedly emphasized the evolutionary significance of random drift in a natural population which is subdivided into many partially isolated sub-groups. His theory is now accepted by many evolutionists such as HALDANE (1949), MULLER (1949), DOBZHANSKY (1951) and others.

On the other hand, no special attention seems to have been paid to the random fluctuation of gene frequencies due to the random fluctuation in the selection intensities until FISHER and FORD (1947) emphasized its prevalence in natural populations and challenged the theory of Wright by denying any significance of random drift due to small population number in evolution. This led to a polemic (cf. WRIGHT 1948; FISHER and FORD 1950; WRIGHT 1951). Experimental studies on natural populations have been carried out by the school of FISHER and FORD (e.g., SHEPPARD 1951; DOWDESWELL and FORD 1952) and by LAMOTTE (1952).

In spite of these, no mathematical investigations seem to have been worked up on the process of change due to the random fluctuation of selection intensities, except a short article reported by the present author (KIMURA 1952a), though WRIGHT (1948) gave a distribution of gene frequencies in steady state for a special case.

In his report, the present author proved, using a method of transformation and approximation, that the process can be regarded as a deformed Gaussian process. In the present paper, a pair of alleles lacking dominance will be assumed. The process of change of their frequencies when their selection coefficients fluctuate fortuitously from generation to generation around a mean value

* Part of the cost of the mathematical formulae has been paid by the GALTON AND MENDEL MEMORIAL FUND.

[1] Contribution No. 57 of the National Institute of Genetics, Mishima-shi, Japan.

0 is simplest for mathematical treatment. Investigation of this process is a main subject of this paper. The process of change which will be found in terminal portions of the frequency distribution curve is especially important in this connection, so that a precise analysis of it will be undertaken. Through this analysis the reader will be led to new concepts of " quasi-fixation " and " quasi-loss " of an allele. Comparison of this process with that of random drift due to small population number is another important subject in the present report. Though there are many theoretical studies on the process of random drift, they are usually restricted to the state that will be realized after a sufficient number of generations. In that state the distribution curve assumes a fixed form and the probabilities of all heterallelic classes decrease at a constant rate of 1/2N, with fixation and loss of the gene occurring at the same rate.

Hence more extensive studies may be needed to make such a comparison. In the present paper an asymptotic solution for the process of random drift due to small population number will be presented for the first time.

PROCESS OF CHANGE OF FREQUENCIES OF ALLELES WHICH ARE NEUTRAL ON THE AVERAGE AND LACKING DOMINANCE

Consider a very large randomly mating population and assume a pair of alleles A and A'. If x is the relative frequency of the gene A in the population and s is the selection coefficient of A, the rate of change of the gene frequency due to selection is approximately

$$\delta x = sx(1-x)$$

per generation, when s is small and there is no dominance. If there is random fluctuation in the selection intensity, s and therefore δx are random variables, and a certain irregularity is expected in the process of change in gene frequency from generation to generation. When the rate of change is small, this process may be safely treated as a continuous Markov process.

If $\phi(x,t)dx$ is the probability that the gene frequency lies between x and $x+dx$ in the t-th generation, it can be proved that $\phi(x,t)$ satisfies the partial differential equation,

$$\frac{\partial \phi(x,t)}{\partial t} = \frac{\partial^2}{\partial x^2}\left[\frac{V_{\delta x}}{2}\phi(x,t)\right] - \frac{\partial}{\partial x}\left[M_{\delta x}\,\phi(x,t)\right], \tag{1}$$

where $M_{\delta x}$ and $V_{\delta x}$ represent respectively the mean and the variance of δx. This equation which is known by mathematicians as " Kolmogorov's forward differential equation " is usually called " Fokker-Planck equation " by physicists, though this type of equation was already used by LORD RAYLEIGH (cited from FUSHIMI 1941). However, we are indebted principally to SEWALL WRIGHT (1945) for the application of this equation to the problem of population genetics.

A meaning of this equation can easily be understood by noting that the left hand side of this equation represents the rate of change of the relative proba-

bility of any class per generation and this can be decomposed into two parts as represented by two terms in the right; namely the part due to the random fluctuation (first term) and the one due to the directed change (second term).

If the gene A is selectively neutral on the average such that the mean value of its selection coefficient over very long periods is zero,

$$M_{\delta x} = 0 \quad \text{and} \quad V_{\delta x} = V_s x^2 (1-x)^2$$

where V_s is the variance of s. In this case equation (1) is written in the form:

$$\frac{\partial \phi}{\partial t} = \frac{V_s}{2} \frac{\partial^2}{\partial x^2} \{x^2 (1-x)^2 \phi\}. \tag{2}$$

This is a partial differential equation with singularities at the boundaries, so that no arbitrary conditions can be imposed there. But as will be seen in the following operations a continuous stochastic process satisfying the equation (2) is uniquely determined if an initial condition $\phi(x,0)$ is given.

As was demonstrated in the previous report (KIMURA 1952a), if the gene frequency x is transformed into a variate ξ by the relation:

$$\xi = \log\left(\frac{x}{1-x}\right),$$

ξ changes continuously from $-\infty$ to $+\infty$ as x changes from 0 to 1 and the distribution of ξ becomes approximately normal; that is, the process of change of ξ is approximately represented by a Gaussian process.

To solve the equation (2), the same transformation turns out to be very useful: Putting

$$u = \frac{1}{2} e^{\frac{V_s}{8} t} x^{\frac{3}{2}} (1-x)^{\frac{3}{2}} \phi$$

and

$$\xi = \log\left(\frac{x}{1-x}\right),$$

we obtain the heat conduction equation,

$$\frac{\partial u}{\partial t} = \frac{V_s}{2} \frac{\partial^2 u}{\partial \xi^2}. \tag{3}$$

It is already established that this equation has an unique solution which is continuous over $-\infty$ to $+\infty$ when $t \geqq 0$ and which reduces to $u(\xi,0)$ when $t = 0$.

$$u(\xi,t) = \frac{1}{\sqrt{2\pi V_s t}} \int_{-\infty}^{\infty} e^{-\frac{(\xi-\eta)^2}{2V_s t}} u(\eta,0)\, d\eta.$$

Therefore, if the initial distribution of gene frequencies $\phi(x,0)$ is given, the unique solution which satisfies (2) and is continuous between 0 and 1 is

$$\phi(x,t) = \frac{1}{\sqrt{2\pi V_s t}} \frac{e^{-\frac{V_s}{8}t}}{[x(1-x)]^{3/2}} \int_0^1 e^{-\frac{\left[\log \frac{x(1-y)}{(1-x)y}\right]^2}{2V_s t}} \sqrt{y(1-y)}\, \phi(y,0)dy. \quad (4)$$

If the initial condition is not a continuous distribution $\phi(x,0)$, but is a given gene frequency x_0, the relative probability that the gene frequency in the t-th generation will be between x and $x+dx$ is given by the formula:

$$\phi(x,t) = \frac{1}{\sqrt{2\pi V_s t}} \exp\left\{-\frac{V_s}{8}t - \frac{\left[\log \frac{x(1-x_0)}{(1-x)x_0}\right]^2}{2V_s t}\right\} \frac{[x_0(1-x_0)]^{1/2}}{[x(1-x)]^{3/2}}. \quad (5)$$

The process of change of the distribution curve with generations is illustrated in figure 1 assuming that the initial gene frequency in the population is 50%. In this figure the variance of selection coefficient is 0.0483. This is a value which WRIGHT (1948) obtained for the *medionigra* gene in an isolated colony of *Panaxia dominura* (FISHER and FORD 1947), assuming that observed variance of change in gene frequency per year were due wholly to fluctuations in selection. As will be seen in the figure the distribution curve is unimodal before the 28th generation after which it becomes bimodal. In the 100th generation gene frequencies that give maximum probability (corresponding to peaks) are approximately 0.0007 and 0.9993, where the height of the curve (ϕ_{max}) is about 11.37. This is 28.7 times higher than the height at the valley (about 0.397) in the middle part of the distribution. So the distribution curve looks like an U-shaped curve. The more precise form of the terminal part of the distribution curve where the gene frequency is very small is illustrated in figure 2.

With passage of time, the distribution curve becomes nearly U-shaped. The process of change is rather rapid and in the 1000th generation, the peaks of the distribution curve become so high, the gene frequencies corresponding to them become so close to the two termini of the distribution and the valley becomes so deep that it is practically impossible to illustrate the distribution curve in figure 1. More generally, if the initial gene frequency is 50%, the distribution curve is unimodal if the number of generations is less than $4/(3V_s)$ but becomes bimodal if it exceeds this value.

The mean of the distribution is always

$$\int_0^1 x\phi(x,t)dx = x_0, \quad (6)$$

But the variance,

$$V_t = \int_0^1 (x-x_0)^2 \phi(x,t)dx,$$

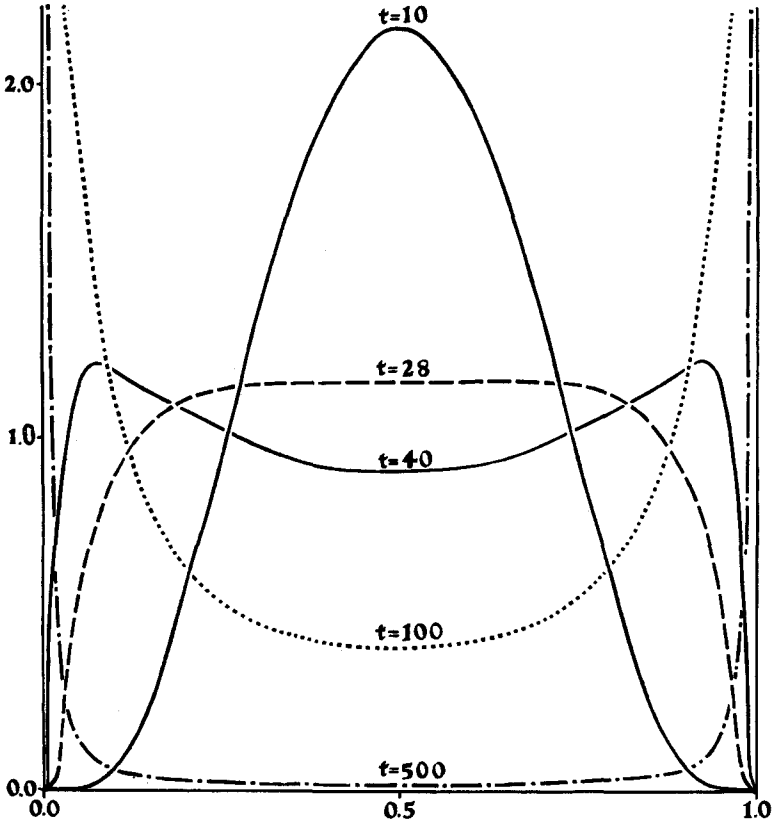

FIGURE 1.—A graph illustrating the process of the change in the distribution of gene frequencies with random fluctuation in the selection intensities. In this illusration it is assumed that the gene is selectively neutral when averaged over a very long period, that there is no dominance, that the initial gene frquncy of the population is 0.5 and that the variance of the selection coefficient is 0.0483. (Abscissa: gene frequency. Ordinate: relative probability.)

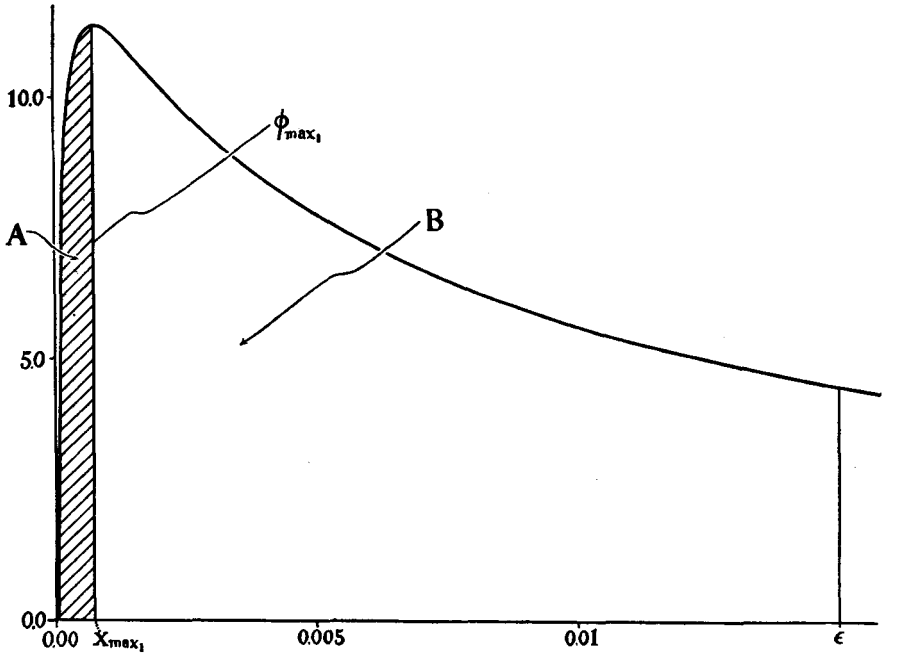

FIGURE 2.—A terminal portion of the distribution curve in the 100th generation where the gene frequency is very low. In this illustration it is assumed that the variance of the selection coefficient is 0.0483 and the initial gene frequency in the population is 0.5. The gene frequency that gives a maximum value in the distribution curve (x_{max_1}) is approximately 0.0007 and corresponding height (ϕ_{max_1}) of the curve is about 11.37. A denotes the probability that the frequency of the gene is smaller than x_{max_1}. This is about 0.007 in this case. ϵ is an arbitrarily chosen gene frequency which is larger than this value. B stands for the probability that the gene frequency in a population is larger than x_{max_1} but smaller than ϵ. If we put ϵ as 0.015, B is approximately 0.098. B becomes ½ at the limit of $t \to \infty$. (Abscissa: gene frequency. Ordinate: relative probability.)

increases in successive generations and for large t it is represented asymptotically by the formula

$$V_t = x_0(1-x_0) - \sqrt{\frac{\pi x_0(1-x_0)}{2V_s t}}\, e^{-\frac{V_s t}{8}} + O\left(\frac{e^{-\frac{V_s t}{8}}}{t\sqrt{t}}\right) \tag{7}$$

Therefore its final rate of approach to the limiting value is very close to $V_s/8$ per generation.

CHANGE IN THE TERMINAL PARTS OF THE DISTRIBUTION AND THE PROCESS OF QUASI-FIXATION

As shown above, classes with the highest probability shift toward the terminals indefinitely with time so that the distribution curve appears to be U-shaped. But it is not a true U-shaped curve, since its value at either terminal is always 0. So it will be important to investigate how the distribution curve will continue to change after a sufficient number of generations, with special reference to its terminal parts.

First let us fix our attention to the terminal portion of the distribution where the frequency of the gene A is very low (see figure 2).

The gene frequency x_{max_1} that gives maximum relative probability ϕ_{max_1} is asymptotically

$$x_{max_1} \sim \frac{x_0}{1-x_0}\, e^{-\frac{3}{2}V_s t} \longrightarrow 0 \quad (t \longrightarrow \infty). \tag{8}$$

While the corresponding relative probability is

$$\phi_{max_1} \sim \frac{1}{\sqrt{2\pi V_s t}} \frac{(1-x_0)^2}{x_0}\, e^{V_s t} \longrightarrow \infty \quad (t \longrightarrow \infty). \tag{9}$$

The gene frequency that gives the maximum value in the distribution curve will approach indefinitely to one terminal point (0), elevating indefinitely the corresponding height of the distribution curve.

Let A be the probability that the gene frequency in the population is lower than x_{max_1}. To calculate this, we will start from a more general relation: The probability that the gene frequency in the population falls between two assigned values a and b is

$$\Pr\{a < x < b\} = \int_a^b \phi(x,t)dx$$

$$= \frac{1-x_0}{\sqrt{2\pi}} \int_{\alpha(a)}^{\alpha(b)} e^{-\frac{1}{2}\lambda^2}\, d\lambda + \frac{x_0}{\sqrt{2\pi}} \int_{\beta(a)}^{\beta(b)} e^{-\frac{1}{2}\lambda^2}\, d\lambda, \tag{10}$$

where

$$\alpha(\lambda) = \frac{\log\left(\frac{\lambda}{1-\lambda}\right) - \log\left(\frac{x_0}{1-x_0}\right)}{\sqrt{V_s t}} + \tfrac{1}{2}\sqrt{V_s t}$$

and

$$\beta(\lambda) = \frac{\log\left(\frac{\lambda}{1-\lambda}\right) - \log\left(\frac{x_0}{1-x_0}\right)}{\sqrt{V_s t}} - \tfrac{1}{2}\sqrt{V_s t}.$$

Using this relation it can be easily shown that

$$\begin{aligned} A &= \Pr\{0 < x < x_{max_1}\} \\ &\sim \frac{1-x_0}{\sqrt{2\pi}} \int_{\sqrt{V_s t}}^{\infty} e^{-\frac{1}{2}\lambda^2} d\lambda + \frac{x_0}{\sqrt{2\pi}} \int_{2\sqrt{V_s t}}^{\infty} e^{-\frac{1}{2}\lambda^2} d\lambda \\ &= \frac{1-x_0}{\sqrt{2\pi V_s t}} e^{-\frac{1}{2}V_s t} + O\left(\frac{e^{-2V_s t}}{\sqrt{V_s t}}\right) \longrightarrow 0. \end{aligned} \tag{11}$$

That is, this probability vanishes at the limit; $t \to \infty$.

On the other hand, Let B stand for the probability that the gene frequency in the population is larger than x_{max_1} but smaller than ϵ, where ϵ is an arbitrarily chosen gene frequency larger than x_{max_1}. Using the relation (10),

$$\begin{aligned} B = \int_{x_{max_1}}^{\epsilon} \phi(x,t)dx &\sim \frac{1-x_0}{\sqrt{2\pi}} \int_{-\sqrt{V_s t}}^{\frac{\log\epsilon - \log\left(\frac{x_0}{1-x_0}\right)}{\sqrt{V_s t}} + \frac{1}{2}\sqrt{V_s t}} e^{-\frac{1}{2}\lambda^2} d\lambda \\ &+ \frac{x_0}{\sqrt{2\pi}} \int_{-2\sqrt{V_s t}}^{\frac{\log\epsilon - \log\left(\frac{x_0}{1-x_0}\right)}{\sqrt{V_s t}} - \frac{1}{2}\sqrt{V_s t}} e^{-\frac{1}{2}\lambda^2} d\lambda. \end{aligned}$$

Therefore, for any ϵ, however small, B can be brought arbitrarily close to $1 - x_0$ by taking t sufficiently large such that

$$V_s t \gg -\log \epsilon.$$

This may be made more clear by the following relation:

$$\begin{aligned} B &= \Pr\{x_{max_1} < x < \epsilon\} \\ &\sim \frac{1-x_0}{\sqrt{2\pi}} \int_{-\sqrt{V_s t}}^{\frac{1}{2}\sqrt{V_s t}} e^{-\frac{1}{2}\lambda^2} d\lambda + \frac{x_0}{\sqrt{2\pi}} \int_{-2\sqrt{V_s t}}^{-\frac{1}{2}\sqrt{V_s t}} e^{-\frac{1}{2}\lambda^2} d\lambda \end{aligned}$$

$$= (1 - x_0) - \frac{2(1 - 2x_0)}{\sqrt{2\pi V_s t}} e^{-\frac{V_s t}{8}} + O\left(\frac{e^{-\frac{V_s t}{2}}}{\sqrt{V_s t}}\right) \longrightarrow 1 - x_0 \ (t \longrightarrow \infty). \quad (12)$$

This shows that after a sufficient number of generations B approaches to $1 - x_0$ with the rate of $V_s/8$ per generation. Figure 2 illustrates the terminal portion of the distribution curve when $V_s = 0.0483$ and $x_0 = 0.5$.

Similar relations hold for the other terminal portion of the distribution where the frequency of the gene is very close to 1: It x_{max_2} stands for the gene frequency giving the maximum value in the distribution curve and ϕ_{max_2} stands for the corresponding relative probability,

$$x_{max_2} \sim 1 - \frac{1 - x_0}{x_0} e^{-\frac{3}{2} V_s t} \longrightarrow 1 \quad (13)$$

and

$$\phi_{max_2} \sim \frac{1}{\sqrt{2\pi V_s t}} \frac{x_0^2}{1 - x_0} e^{V_s t} \longrightarrow \infty \quad (14)$$

as $t \to \infty$. The probability A′ that the frequency of the gene exceeds x_{max_2} vanishes as t approaches infinity;

$$A' = Pr\{x_{max_2} < x < 1\} \sim O\left(\frac{e^{-\frac{1}{2} V_s t}}{\sqrt{V_s t}}\right) \longrightarrow 0. \quad (15)$$

On the other hand, even if $\epsilon' (> 0)$ is taken however small, the probability B′ that the gene frequency of the population will fall between $1 - \epsilon'$ and x_{max_2} approaches to x_0 with the rate of $V_s/8$ per generation at the limit of $t \to \infty$;

$$B' = Pr\{1 - \epsilon' < x < x_{max_2}\} \sim x_0 + O\left(\frac{e^{-\frac{V_s}{8} t}}{\sqrt{V_s t}}\right). \quad (16)$$

The gene frequency giving the minimum of this pseudo-U-shaped distribution curve (x_{min}) approaches ½ at $t \to \infty$ even if the initial gene frequency is not 50%:

$$x_{min} \sim \frac{1}{2} - \frac{\log \frac{x_0}{1 - x_0}}{\frac{3}{2} V_s t - 4}. \quad (17)$$

The corresponding relative probability ϕ_{min} vanishes at the limit:

$$\phi_{min} \sim \sqrt{\frac{x_0(1-x_0)}{2\pi V_s t}} \left\{ \frac{1}{4} - \left(\frac{\log \frac{x_0}{1-x_0}}{\frac{3}{2}V_s t - 4} \right)^2 \right\}^{-3/2} \exp\left\{ -\frac{V_s t}{8} - \frac{9}{8}\left(\frac{\log \frac{x_0}{1-x_0}}{\frac{3}{2}V_s t - 4} \right)^2 V_s t \right\} \longrightarrow 0. \quad (18)$$

That is, the valley in the distribution curve deepens until the bottom reaches the abscissa.

As will be seen from the relation;

$$\lim_{\epsilon \to 0+} \int_{\epsilon}^{1-\epsilon} \phi(x,t) = 1, \quad (19)$$

the random fluctuation of selection intensities by itself cannot lead to the complete fixation or loss, in the strict sense, of the gene contrary to the case of random drift due to small population number. But as has been shown through (8) – (18), there is a strong tendency that the gene frequency will move toward either terminus with increasing time. In other words, after a sufficient number of generations almost all populations will be in such a situation that the gene is either almost fixed in the population or almost lost from it. To distinguish this from the fixation or loss in the case of small effective population number, the terms " quasi-fixation " and " quasi-loss " are proposed. As will be seen from (12) and (16), their rate can be taken as $V_s/8$.

In the long run, this process of quasi-fixation or -loss will be checked by the opposing mutation pressure.

In this state of statistical equilibrium, if the mutation rates of the gene A to and from its allele A' are u and v respectively, the frequency distribution of A in the population is given by the formula:

$$\phi(x) = C \frac{e^{-\frac{2}{V_s}\left(\frac{v}{x} + \frac{u}{1-x}\right)}}{x^2(1-x)^2} \left(\frac{x}{1-x} \right)^{-2\left(\frac{u-v}{V_s}\right)}, \quad (20)$$

where C is a constant chosen such that $\int_0^1 \phi(x)dx = 1$. WRIGHT (1948) derived essentially the same formula assuming migration (p. 292).

COMPARISON WITH THE PROCESS OF RANDOM DRIFT DUE TO SMALL POPULATION NUMBER

In a finite population, owing to random sampling of gametes in reproduction, there occurs random fluctuation of the gene frequency from generation to generation. This process, as is well known, will finally lead to the complete fixation or loss of the gene if such factors as mutation, migration and selection are absent.

If N is the effective number of reproducing individuals in the population and p is the initial gene frequency the nth moment of distribution about zero in the t-th generation is given by the following formula if the order of t is not smaller than N (cf. KIMURA 1952b):

$$\mu_n'^{(t)} = p - 3pq\frac{n-1}{n+1}(1-\lambda_1)^t - 5pq(p-q)\frac{(n-2)(n-1)}{(n+1)(n+2)}(1-\lambda_2)^t$$
$$- 7pq(-5pq+1)\frac{(n-3)(n-2)(n-1)}{(n+1)(n+2)(n+3)}(1-\lambda_3)^t$$
$$- 9pq(14pq^2-7pq+p-q)\frac{(n-4)(n-3)(n-2)(n-1)}{(n+1)(n+2)(n+3)(n+4)}(1-\lambda_4)^t$$
$$+ O\{(1-\lambda_5)^t\}, \quad (21)$$

where $q = 1-p$. From this we can derive the probability that the gene will have become fixed in the population by the t-th generation:

$$f_t(1) = p - 3pq(1-\lambda_1)^t - 5pq(p-q)(1-\lambda_2)^t - 7pq(-5pq+1)(1-\lambda_3)^t$$
$$- 9pq(14pq^2-7pq+p-q)(1-\lambda_4)^t + O\{(1-\lambda_5)^t\} \quad (22)$$

The corresponding probability of complete loss is:

$$f_t(0) = q - 3pq(1-\lambda_1)^t + 5pq(p-q)(1-\lambda_2)^t - 7pq(-5pq+1)(1-\lambda_3)^t$$
$$+ 9pq(14pq^2-7pq+p-q)(1-\lambda_4)^t + O\{(1-\lambda_5)^t\} \quad (t\to\infty). \quad (23)$$

In these formulae

$$\lambda_1 = \frac{1}{2N}, \lambda_2 = \frac{3}{2N}, \lambda_3 = \frac{6}{2N}, \lambda_4 = \frac{10}{2N}, \lambda_5 = \frac{15}{2N}, \ldots.$$

In general λ's are given by the formula;

$$\lambda_i = \frac{i(i+1)}{4N} \quad (i = 1, 2, \ldots).$$

The frequency of the gene in this case may take any one of a series of discontinuous values:

$$0, \frac{1}{2N}, \frac{2}{2N}, \ldots, 1-\frac{1}{2N}, 1.$$

Usually, however, the number of reproducing individuals (N) in a population is so large that practically the gene frequency (x) can be treated as a continuous variable with good approximation. Variance of the rate of change in gene frequency due to the random sampling of gametes is

$$V_{\delta x} = \frac{x(1-x)}{2N}.$$

Therefore if $\phi(x,t)$ is the relative probability that the frequency of the gene in the population will take any value between x and $x+dx(0<x<1)$ in the t-th generation, $\phi(x,t)$ satisfies the following partial differential equation:

$$\frac{\partial \phi}{\partial t} = \frac{1}{4N}\frac{\partial^2}{\partial x^2}\{x(1-x)\phi\}, \tag{24}$$

which is easily derivable from equation (1). To solve this, if we put

$$\phi \propto X_i(x)e^{-\lambda_i t} \qquad (i = 1, 2, 3, \ldots),$$

we obtain the ordinary differential equation;

$$x(1-x)\frac{d^2X_i}{dx^2} + (2-4x)\frac{dX_i}{dx} - (2-4N\lambda_i)X_i = 0,$$

where λ_i corresponds to the eigen value of equation (24). Noting that $\lambda_i = i(i+1)/4N$, this becomes

$$x(1-x)\frac{d^2X_i}{dx^2} + (2-4x)\frac{dX_i}{dx} - (1-i)(i+2)X_i = 0. \tag{25}$$

This type of equation is known as Gauss's differential equation and (25) is satisfied by the following hypergeometric series:

$$\begin{aligned} X_i &= F(1-i, i+2, 2, x) \\ &= 1 + \frac{(1-i)(i+2)}{1\cdot 2}x + \frac{(1-i)(2-i)\cdot(i+2)(i+3)}{1\cdot 2\cdot 2\cdot 3}x^2 \\ &+ \frac{(1-i)(2-i)(3-i)\cdot(i+2)(i+3)(i+4)}{1\cdot 2\cdot 3\cdot 2\cdot 3\cdot 4}x^3 \\ &+ \frac{(1-i)(2-i)(3-i)(4-i)\cdot(i+2)(i+3)(i+4)(i+5)}{1\cdot 2\cdot 3\cdot 4\cdot 2\cdot 3\cdot 4\cdot 5}x^4 \\ &+ \ldots \quad (i = 1, 2, 3, \ldots) \end{aligned}$$

Therefore the asymptotic solution of (24) for large t is;

$$\begin{aligned} \phi(x,t) = {} & C_1e^{-\lambda_1 t} + C_2(1-2x)e^{-\lambda_2 t} + C_3(1-5x+5x^2)e^{-\lambda_3 t} \\ & + C_4(1-9x+21x^2-14x^3)e^{-\lambda_4 t} + O(e^{-\lambda_5 t}). \end{aligned} \tag{26}$$

To determine the constants C_1, C_2, C_3, . . . we can use the relation that the n-th moment of distribution obtained by this formula,

$$\int_0^1 x^n \phi(x,t)\,dx,$$

must be equal to

$$\mu'_n{}^{(t)} - 1^n \cdot f_t(1),$$

since the homallelic classes are excluded from the distribution curve to be given by (24). Thus we obtain the following values:

$$C_1 = 6\,pq, C_2 = -30\,pq(p-q),\ C_3 = 84\,pq(-5\,pq+1),$$
$$C_4 = -180\,pq(14\,pq^2 - 7\,pq + p - q).$$

How the distribution curve represented by (26) changes with generations is illustrated in figure 3, assuming that the initial gene frequency p is 0.1. As may be seen from this figure, the curve becomes gradually flat until finally every heterallelic class has equal probability and falls with the rate of 1/2N per generation. In this final stage, the fixation of the gene proceeds at the same rate, the correct value of which was first obtained by WRIGHT to be 1/2N (see WRIGHT 1931) by using a different method of calculation. This rate is usually known as the rate of fixation due to random sampling of gametes.

The probability Ω_t that the alleles A and A′ coexist in the population in the t-th generation can be obtained from (26):

$$\Omega_t = \int_0^1 \phi(x,t)dx$$
$$= 6\,pq\, e^{-\frac{1}{2N}t} + 14\,pq\,(-5\,pq+1)\, e^{-\frac{6}{2N}t} + O\left(e^{-\frac{15}{2N}t}\right) (t \to \infty). \quad (27)$$

Contrary to what was shown in (19) either complete fixation or loss of alleles is expected in this case and Ω_t vanishes at the limit of $t \to \infty$.

Variance of the distribution in the t-th generation is from (22), (23) and (26),

$$V_t = pq - pq\, e^{-\frac{1}{2N}t}, \quad (28)$$

namely the variance approaches its limiting value pq at the rate 1/2N per generation.

As has been demonstrated above, the process of change due to random fluctuation of selection intensity is quite different from that due to the random sampling of gametes. Therefore comparison of their effects must be made from various angles as WRIGHT (1948) did in analyzing the data of *medio-nigra* gene in Panaxia.

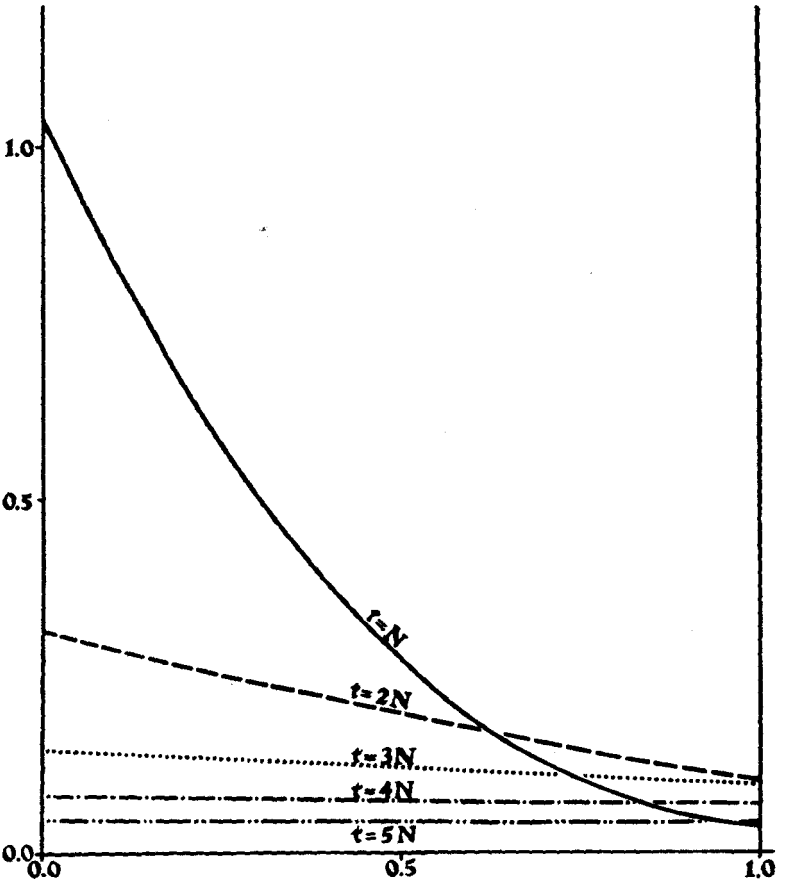

FIGURE 3.—Graphs showing the process of change of the frequency distribution curve due to the random drift in small populations. In this illustration it is assumed that the initial gene frequency in the population is 0.1. It will be seen that the distribution curve becomes more and more flat as the number of generations increases. (Abscissa: gene frequency in the population. Ordinate: relative probability. t: time in generation. N: effective population size.)

Thus, if we consider the process of change which will be realized after a sufficient number of generations, the rate of quasi-fixation, $V_s/8$, may be compared with the rate of fixation due to random sampling, $1/2N$, for the same purpose. Suppose V_s is known to be 0.0483 as in figure 1, the equivalent N is calculated to be about 83 by using a equivalence relation;

$$\frac{1}{2N} \doteq \frac{V_s}{8},$$

though the applicability of this formula is rather restricted.

So far we have treated the two factors separately. But in nature not only these two factors but also systematic factors may work concurrently. The present writer (1951) reported briefly on the distribution of gene frequencies for such case. The more precise account will appear elsewhere, but the main conclusion derived from the analysis of the distribution curve is not difficult to present here: The effect produced by the random fluctuation in natural selection is relatively unimportant for small populations. But in large populations it has a remarkable effect in such ways that in the case of no dominance the distribution curve is modified markedly in the parts where the frequency of either allele is low and in the case of complete dominance in a part where the frequency of the recessive gene lies inside a certain range of higher frequencies. Also the product NV_s is an important quantity. To estimate not only the distribution of population size (N) in nature but also the variability (V_s) of selection intensity for important loci may be an important task left for the future experiments.

ACKNOWLEDGMENT

The author is greatly indebted to MR. Y. KAWAGUCHI who made all the laborious computations from which the figures in this paper were constructed. Thanks are due to DR. SEWALL WRIGHT who read the manuscript and gave valuable criticisms and suggestions. The author wishes to express deep appreciation to DR. JAMES F. CROW without whose advice and help this paper could not have been published.

SUMMARY

When there are random fluctuations in selection intensity, the process of change in gene frequency in a population is represented by a stochastic process. In this paper an analysis of this process is presented for a gene lacking dominance and selectively neutral on the average. Especially interesting is the process of change that can be observed in the terminal portions of the distribution curve. Contrary to the case of random drift in small populations, if the population is very large, complete fixation or loss of an allele, in the strict sense, will not be realized. But there exists a strong tendency toward the state of almost fixation or almost loss. That is, if we allow a sufficient number of generations a situation will almost surely be realized in which the allele is either almost fixed in the population or almost lost from it. To distinguish this

from the fixation or loss in the case of random drift in small populations, terms " quasi-fixation " and " quasi-loss " were proposed. Their rate per generation can be taken as $V_s/8$, where V_s is the variance of the selection coefficient. Comparison of this process with that of random drift in small populations is another important subject in the present paper. In spite of many studies on the process of the drift very little is known about the process of the change before the fixation and loss of an allele proceeds at the constant rate of 1/2N. In this paper an asymptotic solution for this process is presented for the first time.

REFERENCES

DOBZHANSKY, TH., 1951 Genetics and the origin of species. Columbia Univ. Press, New York.

DOWDESWELL, W. H., and E. B. FORD, 1952 The distribution of spot-number as an index of geographical variation in the butterfly *Maniola jurtia* L. (Lepidoptera: Satyridae). Heredity **6**: 99–109.

FELLER, W., 1950 Diffusion processes in genetics. Proceedings of Second Berkeley Symposium on Mathematical Statistics and Probability, Univ. of California Press, pp. 227–246.

FISHER, R. A., 1930 The genetical theory of natural selection. Oxford. Clarendon Press.

FISHER, R. A., and E. B. FORD, 1947 The spread of a gene in natural conditions in a colony of the moth *Panaxia dominula* L. Heredity **1**: 143–174.

1950 The " Sewall Wright Effect." Heredity **4**: 117–119.

FUSHIMI, K., 1941 Theories of probability and statistics (Japanese). Tokyo.

HALDANE, J. B. S., 1939 The equilibrium between mutation and random extinction. Ann. Eug. **9**: 400–405.

1949 Human evolution: Past and future. Genetics, paleontology and evolution, Princeton Univ. Press, pp. 405–418.

HOUSE, V. H., 1953 The use of the binomial expansion for a classroom demonstration of drift in small populations. Evolution **7**: 84–88.

KIMURA, M., 1951 Effect of random fluctuation of selective value on the distribution of gene frequencies in natural populations. Ann. Rep. Nat. Inst. Gen. **1**: 45–47.

1952a Process of irregular change of gene frequencies due to the random fluctuation of selection intensities. *Ibid.* **2**: 56–57.

1952b On the process of decay of variability due to random extinction of alleles. *Ibid.* **2**: 60–61.

LAMOTTE, M., 1952 Le rôle des fluctuations fortuites dans la diversité des populations naturelles de *Capaea nemoralis* (L.). Heredity **6**: 333–343.

MULLER, H. J., 1949 Redintegration of the symposium on genetics, paleontology, and evolution. Genetics, paleontology and evolution, Princeton Univ. Press, pp. 421–445.

SHEPPARD, P. M., 1951 Fluctuations in the selective value of certain phenotypes in the polymorphic land snail *Cepæa nemoralis* (L.). Heredity **5**: 125–134.

WRIGHT, S., 1931 Evolution in mendelian populations. Genetics **16**: 97–159.

1945 The differential equation of the distribution of gene frequencies. Proc. Nat Acad. Sci. **31**: 382–389.

1948 On the roles of directed and random changes in gene frequency in the genetics of populations. Evolution **2**: 279–294.

1951 Fisher and Ford on " The Sewall Wright Effect." Amer. Scientist **39**: 452–479.