

*THE DIFFERENTIAL EQUATION OF THE DISTRIBUTION OF
GENE FREQUENCIES*

BY SEWALL WRIGHT

DEPARTMENT OF ZOÖLOGY, THE UNIVERSITY OF CHICAGO

Communicated October 15, 1945

The first attempt to determine the mathematical form of the distribution of gene frequencies in populations was based on the setting up of differential equations for certain special cases (Fisher, 1922,¹ 1930²). A correction and extension of these results came from expression of the conditions in an integral equation (Wright, 1929,³ 1931⁴). A general solution has since been obtained for fully stationary distributions by a third method (Wright, 1937,⁵ 1938⁶). The case of uniform flux has been treated less generally (Wright, 1938,⁶ 1942⁷). Dr. A. Kolmogorov⁸ has recently been kind enough to send me a reprint of an important paper on this subject which was published in 1935 but which had not previously come to my attention. While the application is restricted to a particular stationary distribution, the method of approach points to a more systematic formulation than before.

The situation discussed by Kolmogorov is that of a large population, consisting of many subgroups of size n , each of which receives a certain number (k) of immigrants from the general population but otherwise breeds within itself. The average rate of change of the gene frequency of subgroups, in which p is the frequency of a given gene, is represented by $A = \Sigma(\Delta p) = (k/n)(\bar{p} - p)$ where \bar{p} is the mean value of p in the whole population. The variance of p , due to accidents of sampling in one generation, is represented by $B = \Sigma(\Delta p)^2 = pq/2n$, $q = 1 - p$. It is stated, without demonstration, that the distribution $u(p)$ of gene frequencies among subgroups after a stationary state has been reached, answers to the differential equation

$$\frac{1}{2} \frac{\partial^2}{\partial p^2} (Bu) - \frac{\partial}{\partial p} (Au) = 0. \quad (1)$$

The pertinent solution is given as

$$u(p) = p^{4k\bar{p}-1} q^{4k\bar{q}-1} / B(4k\bar{p}, 4k\bar{q}). \quad (2)$$

The effect of selection in this situation is discussed briefly without, however, modifying $u(p)$ by introduction of the selection term, $\alpha p^2 q$, into A .

It is noted that the same formula (2) had previously been derived by the present author⁴ by a different method. Equation (1) has, however, broader implications if valid for the general case $A = \Sigma(\Delta p)$, $B = \Sigma(\Delta p)^2$ and not merely for the particular case $A = (k/n)(\bar{p} - p)$, $B = pq/2n$.

The Immediate Factors of Evolutionary Change.—The immediate factors that tend to cause systematic changes (Δq) in gene frequency (q) may be listed exhaustively^{4, 5} as (a) mutation pressure, $\Delta q = v(1 - q) - uq$, where v and u are the rates of mutation to and from the gene in question, (b) immigration pressure, $\Delta q = m(q_i - q)$ (Kolmogorov's *A*) where m is the proportion of replacement by immigrants and q_i is the gene frequency in these, and (c) selection pressure, which may take widely diverse forms but in the important case of constant relative selective values (W) for each multiple factor genotype in a random breeding population takes the form $\Delta q = q(1 - q) \frac{\partial \bar{W}}{\partial q} / r\bar{W}$, where r is 1 in haploids, 2 in diploids, the

usual case, 1.5 for sex linked genes (if equal numbers of males and females), 4 in tetraploids, etc.^{7, 9} In addition to these systematic pressures are (d) the random variations, δq , due to accidents of sampling, the variance of which is $\sigma_{\delta q}^2 = q(1 - q)/rN$ in a population of effective size N (Kolmogorov's *B*). The diploid case ($r = 2$) will be assumed in what follows.

The Stationary Distribution of Gene Frequencies.—Systematic pressure toward the gene frequency, at which $\Delta q = 0$ and the cumulative effects of accidents of sampling determine a probability curve $\varphi(q)$ describing the frequencies which would be exhibited in the long run by the value of q for a particular gene in a population subject to constant conditions. This distribution may also be interpreted as that exhibited at one time by the values of q in a group of populations that are all subject to the same conditions (as in the case of Kolmogorov's $u(p)$). The deviations from the binomial square formula for genotypic frequencies in the total population, depend on the variance of $\varphi(q)$ under this interpretation.^{4, 8, 10} In other cases $\varphi(q)$ may be used as the distribution at any time within either a class of non-allelic genes or an extensive series of multiple alleles,¹¹ all subject to the same conditions.

That equation (1) is, in fact, completely general for the stationary form of distribution may be shown by a slight modification of a method^{5, 6} that has been used for derivation of $\varphi(q)$.

The conditions for stability of the distribution (including the terminal classes $q = 0, q = 1$) may be represented by two equations expressing the persistence of the mean and variance, respectively

$$\int_0^1 (q + \delta q + \Delta q) \varphi(q) dq = \int_0^1 q \varphi(q) dq. \tag{3}$$

$$\int_0^1 (q - \bar{q} + \delta q + \Delta q)^2 \varphi(q) dq = \int_0^1 (q - \bar{q})^2 \varphi(q) dq. \tag{4}$$

Noting that the mean value of δq is zero, and that δq is not correlated with q or Δq , these equations reduce to the following if the term in $(\Delta q)^2$ in (4) may be ignored. It may be noted in this connection that this term is negligible if Δq is of the same order as $\sigma_{\delta q}^2$ or less, while if of higher order,

systematic pressure dominates the results so completely that the distribution formula itself becomes unimportant.

$$\int_0^1 \Delta q \varphi(q) dq = 0. \quad (3a)$$

$$2 \int_0^1 (q - \bar{q}) \Delta q \varphi(q) dq + \int_0^1 \sigma_{\Delta q}^2 \varphi(q) dq = 0. \quad (4a)$$

Putting $\Delta q \varphi(q) dq = d\chi(q)$ these conditions become

$$\chi(1) - \chi(0) = 0. \quad (3b)$$

$$2 \int_0^1 \chi(q) dq - 2[\chi(1) + \bar{q}(\chi(1) - \chi(0))] - \int_0^1 \sigma_{\Delta q}^2 \varphi(q) dq = 0. \quad (4b)$$

Substituting (3b) in (4b) the latter becomes

$$\int_0^1 [2\chi(q) - 2\chi(1) - \sigma_{\Delta q}^2 \varphi(q)] dq = 0. \quad (4c)$$

A solution is obtained by removing the integral sign since the resulting equation not only satisfies (4c) but also (3b) (noting that $\sigma_{\Delta q}^2 = 0$ if $q = 0$ or if $q = 1$, there being no sampling variance without alternatives in the sample).

$$\chi(q) - \chi(1) = 1/2 \sigma_{\Delta q}^2 \varphi(q). \quad (5)$$

This can be solved for $\varphi(q)$ by differentiating the logarithm of the left-hand number and making the appropriate substitutions.⁶

$$\varphi(q) = (C/\sigma_{\Delta q}^2) e^{2 \int (\Delta q / \sigma_{\Delta q}^2) dq} \quad (6)$$

where C is a constant such that $\int_0^1 \varphi(q) dq = 1$.

Since q increases by steps of $1/2N$ in a population of size N , the frequency of a given value of q is $f(q) = \varphi(q)/2N$. From a study⁴ of simple cases ($N = 2$ or 3) in which the frequencies in the stationary state can be determined algebraically and from a more elaborate investigation by R. A. Fisher² of the subterminal region in certain cases, it appears that the frequencies are given with considerable accuracy by the formula except for the terminal classes, $q = 0$, $q = 1$. Consideration of the exchanges which occur between the terminal and neighboring classes leads⁴ to the following approximate estimate for the terminal class, $q = 0$. That for $q = 1$ is analogous.

$$f(0) = f(1/2N)/4N[mq_t + v]. \quad (7)$$

The differential equation for the completely stationary case is given by differentiation of (5). It comes under equation (1).

$$\frac{1}{2} \frac{d}{dq} (\sigma_{\Delta q}^2 \varphi(q)) - \Delta q \varphi(q) = 0. \quad (8)$$

Since $\Delta q \varphi(q)$ is the proportion of the distribution which tends to be carried past a specified value of q by the systematic pressure Δq , the other

term must represent the net proportion which tends to be carried in the opposite direction by accidents of sampling in each generation.

The Case of Steady Flux.—There may be a practically stationary state of the proportions in all intermediate values of q in spite of steadily increasing frequency of one terminal class at the expense of the other, provided that the proportion lost by the donor terminal class is negligible. This cannot be the case if either mutation rate or immigration rate is appreciable, but may hold in the presence of strong selection pressure since selection pressure is nil in populations in which $q = 0$ or $q = 1$.

The differential equation for the case of steady flux must differ from (8) by a constant term (D), the net proportion of the total (excluding the recipient class) that is carried past each value of q in each generation.

$$\frac{1}{2} \frac{d}{dq} (\sigma_{sq}^2 \varphi(q)) - \Delta q \varphi(q) + D = 0. \tag{9}$$

This is the general form given by one integration of (1) which is therefore the general differential equation for a steady state of the intermediate classes. It may be reduced to a linear equation of the first order by making the substitution, $y = \sigma_{sq}^2 \varphi(q)$.

$$\frac{dy}{dq} - 2 \left(\frac{\Delta q}{\sigma_{sq}^2} \right) y + 2D = 0. \tag{10}$$

The solution for $\varphi(q)$ is as follows:

$$\varphi(q) = [e^{2\int (\Delta q / \sigma_{sq}^2) dq} / \sigma_{sq}^2] [C - 2D \int e^{-2\int (\Delta q / \sigma_{sq}^2) dq} dq]. \tag{11}$$

The simplest special case is that in which Δq may be treated as zero (although there could be no flux if it were absolutely zero).

$$f(q) = \frac{C}{q(1 - q)} - \frac{2D}{1 - q}. \tag{12}$$

The case under (12) that is most important genetically is that of irreversible mutation at a rate so low that the donor class ($q = 0$, or $q = 1$) is not appreciably depleted. According to direction of mutation,

$$f(q) = 2v/q, \text{ or } f(q) = 2v/(1 - q). \tag{13}$$

The ratio of the subterminal classes ($1/2N$ in this case) gives the probability that a single neutral mutation may reach fixation instead of elimination.

Returning to (12) the case in which $D = 0$ yields the corresponding simplest solution for a completely stationary state

$$\varphi(q) = 1/[2(0.577 + \log 2N)q(1 - q)] \text{ (terminal classes excluded)}. \tag{14}$$

The case in which there are constant relative selection coefficients for all genotypes ($\Delta q = q(1 - q) \frac{\partial \bar{W}}{\partial q} / r\bar{W}$) gives an apparently simple but in general rather refractory form (assuming a given set of frequencies of other genes).

$$\varphi(q) = [\bar{W}^{2N} / \sigma_{sq}^2] [C - 2D \int \bar{W}^{-2N} dq]. \quad (15)$$

It will be convenient for later reference to cite the less general case $\Delta q = q(1 - q)(s + tq)$, $\sigma_{sq}^2 = q(1 - q)/2N$ which allows for any degree of dominance, provided s and t are both small.

$$f(q) = [e^{4Ns_q + 2Ntq^2} / q(1 - q)] [C - 2D \int e^{-(4Ns_q + 2Ntq^2)} dq]. \quad (16)$$

Non-stationary States.—The general case, in which the proportion at each value of q is a function of time as well as of q itself, is given by the following, of which equation (1) is the case in which the left-hand member is zero. Time (T) is measured in generations.

$$\frac{\partial \varphi(q, T)}{\partial T} = \frac{1}{2} \frac{\partial^2}{\partial q^2} [\sigma_{sq}^2 \varphi(q, T)] - \frac{\partial}{\partial q} [\Delta q \varphi(q, T)]. \quad (17)$$

This can be reduced to an ordinary differential equation in the case in which the distribution has reached stability of form, with all classes (except the terminal ones) falling off at the same rate. Let $K = - \frac{1}{\varphi(q, T)} \frac{\partial \varphi(q, T)}{\partial T}$ be the rate of decay per generation.

$$\frac{1}{2} \frac{d^2}{dq^2} (\sigma_{sq}^2 \varphi(q)) - \frac{d}{dq} (\Delta q \varphi(q)) + K \varphi(q) = 0. \quad (18)$$

It may easily be verified that for the case in which fixation is occurring under the uncomplicated effect of inbreeding ($\Delta q = 0$, $K = 1/2N$) the only solution that does not involve negative frequencies is

$$\varphi(q) = 1, \quad \text{or } f(q, T) = C_0 e^{-T/2N}. \quad (19)$$

In the case of irreversible mutation at an appreciable rate, $\Delta q = v(1 - q)$ the rate of decay is easily shown to be $K = v$. Equation (18) is satisfied by the following value, originally derived by a different method.

$$f(q) = 2vq^{4Nv-1}. \quad (20)$$

An analogous solution applies to the effect of swamping by immigration from a population in which the gene in question is fixed ($\Delta q = m(1 - q)$, $K = m$)

$$f(q) = 2mq^{4Nm-1}. \quad (21)$$

Comparison with Results by Other Methods.—The first attempt at determining the distribution of gene frequencies was made by R. A. Fisher¹ who arrived at differential equations for certain special cases, in terms, however, of a different variable than gene frequency, $\vartheta = \cos^{-1}(1 - 2q)$, used in order to make the sampling variance constant. A discrepancy between the rate of decay ($K = 1/4N$), derived by him for the case in which $\Delta q = 0$, and the value, $1/2N$, given by a general method^{12, 13} for determining the rate of fixation of genes under any system of mating, led the present author^{3, 4} to a different approach. The condition for a stationary state of the intermediate classes except for possible decay at rate K , was represented by the following equation in which q and x are recipient and donor classes, respectively, in the exchanges which occur from one generation to the next.

$$(1 - K) \frac{\varphi(q)}{2N} = \frac{(2N)!}{(2Nq)![2N(1 - q)]!} \int_0^1 (x + \Delta x)^{2Nq} (1 - x - \Delta x)^{2N(1-q)} \varphi(x) dx. \quad (22)$$

It could easily be seen that if $\Delta x = 0$, the equation is satisfied by $\varphi(q) = \varphi(x) = 1$, $K = 1/(2N + 1)$, the latter at least a close approximation to the rate of decay expected in this case. For the simplest stationary state, $K = 0$, $\Delta q \doteq 0$, the expression $\varphi(q) = Aq^{-1} + B(1 - q)^{-1}$ is indicated (cf. 12). Approximate solutions could also readily be obtained for the linear pressures of mutation and migration. Selection presented more difficulty.

On inspection of these results in manuscript, Fisher² was able to correct and extend his equations to obtain the following:

I. Case of uniform decay ($\Delta q = 0$)

$$\frac{\partial y}{\partial T} = \frac{1}{4n} \left[\frac{\partial^2 y}{\partial \vartheta^2} + \frac{\partial}{\partial \vartheta} (y \cot \vartheta) \right]. \quad (23)$$

$$y = A_0 e^{-T/2n} \sin \vartheta \quad (\text{cf. (19)}). \quad (24)$$

II. Stationary state, no selection ($\Delta q \doteq 0$)

$$\frac{dy}{d\vartheta} + y \cot \vartheta = -4nB. \quad (25)$$

$$y = A \operatorname{cosec} \vartheta + 4nB \cot \vartheta \quad (\text{general, cf. (12)}). \quad (26)$$

$$y = A \operatorname{cosec} \vartheta \quad (\text{symmetrical case, cf. (14)}). \quad (27)$$

$$y = 4nB(\operatorname{cosec} \vartheta + \cot \vartheta) \quad (\text{unidirectional mutation, cf. (13)}). \quad (28)$$

III. Stationary state, selection, no dominance, $\Delta q = aq(1 - q)$

$$\frac{dy}{d\vartheta} - (2an \sin \vartheta - \cot \vartheta)y = -4anA. \quad (29)$$

$$y = \operatorname{cosec} \vartheta (2A + Be^{-2an \cos \vartheta}) \quad (\text{general, cf. (16), } t = 0). \quad (30)$$

$$y = 4 \operatorname{cosec} \vartheta \frac{(1 - e^{-2an(1 + \cos \vartheta)})}{(1 - e^{-4an})} \quad (\text{unidirectional mutation}). \quad (31)$$

In cases I and II, the results agreed with those obtained from the integral equation (22), as may be seen by making the substitutions $\cos \vartheta = 1 - 2q$, $y d\vartheta = \varphi(q) dq$, $d\vartheta/dq = 1/\sqrt{q(1 - q)}$. In case III it was the author's turn to make a correction in the selection term (published first³ as $e^{2Ns q}$), by taking cognizance of a series of small terms erroneously thought to be negligible but which actually doubled the exponent. With this correction, there was agreement.⁴

The most general result⁵ obtained for the completely stationary state by solution of (22) took into account all of the factors of change in the form $\Delta q = v(1 - q) - uq - m(q - q_i) + q(1 - q)(s + tq)$, $\sigma_{sq}^2 = q(1 - q)/2N$.

$$\varphi(q) = Ce^{4Ns q + 2Nt q^2} q^{4N(mq_i + v) - 1} (1 - q)^{4N[m(1 - q_i) + u] - 1}. \quad (32)$$

This agrees with that obtained by substituting these values of Δq and of σ_{sq}^2 in (6).

The most general result^{5,6} obtained by this method for the case of steady flux was for $\Delta q = q(1 - q)(s + tq)$.

$$f(q) = [e^{4Ns q + 2Nt q^2}/q(1 - q)] [C - 2Dq e^{-(2Ns q + Nt q^2)} \psi(2Ns q, 2Nt q^2)]. \quad (33)$$

where

$$\psi(a, 0) = 1 + \frac{a^2}{3!} + \frac{a^4}{5!} + \frac{a^6}{7!} \dots = (e^a - e^{-a})/2a$$

$$\psi(0, b) = 1 + \frac{b}{3!} + \frac{7b^2}{5!} + \frac{27b^3}{7!} \dots E_m b^m$$

$$E_m = (E_{m-1} + E_{m-2})/2m(m + 1).$$

No recurrence formula was recognized for the joint terms, $\psi(2Ns q, 2Nt q^2)$ but the coefficients were calculated⁷ up to those pertaining to q^9 .

The probability of fixation of a single mutation ($C = 2v$, $D = ve^{2Ns + Nt}$ / $\psi(2Ns, 2Nt)$ for irreversible mutations from class $q = 0$, or $C = 0$, $D = -ve^{-(2Ns + Nt)}$ / $\psi(2Ns, 2Nt)$ for irreversible mutations from class $q = 1$), could be calculated from the ratios of the subterminal classes, (Prob. = $\sqrt{s/2N}$ for a recessive mutation with selective advantage s , Prob. = $2s$ for a dominant mutation with selective advantage s , or for a semidominant

with selective advantage s in the heterozygote). The last agrees with Fisher's conclusion.² Equation (33), with $t = 0$, is indeed equivalent to (31).

Comparison of (33) with (16) shows that if the former is correct, the following must hold:

$$\psi(2Nsq, 2Ntq^2) = (e^{2Nsq + Ntq^2}/q) \int e^{-(4Nsq + 2Ntq^2)} dq. \quad (34)$$

This was tested by expanding the two exponentials in (34), integrating each term of the second one and combining. The coefficients were in all cases identical with those published⁷ for $\psi(2Nsq, 2Ntq^2)$.

Equation (22) also gave the solution (20) for the case of uniform decay under an appreciable mutation rate.⁴

The integral equation (22) and the differential equation (18) are clearly equivalent to a close approximation. They are not exact mathematical equivalents, however, as may be seen from the fact that K must be put $1/(2N + 1)$ in (22) if $\Delta x = 0$ to give the solution $\varphi(q) = 1$, while it takes its true value $1/2N$ in (18) to give the same result. In the other cases (except (12)) second order terms have been omitted in the series, obtained as solutions of the integral equation, which do not appear in the solutions of the differential equation. Neither equation, of course, represents the natural conditions exactly since integration is substituted for summation and differentials for minimal steps ($1/2N$) in gene frequency.

¹ Fisher, R. A., *Proc. Roy. Soc. Edinburgh*, **42**, 321-341 (1922).

² Fisher, R. A., *Ibid.*, **50**, 205-220 (1930).

³ Wright, S., *Amer. Naturalist*, **63**, 556-561 (1929).

⁴ Wright, S., *Genetics*, **16**, 97-159 (1931).

⁵ Wright, S., these PROCEEDINGS, **23**, 307-320 (1937).

⁶ Wright, S., *Ibid.*, **24**, 253-259 (1938).

⁷ Wright, S., *Bull. Amer. Math. Soc.*, **48**, 223-246 (1942).

⁸ Kolmogorov, A., *C. R. de l'Acad. des Sciences de l' U.R.S.S.*, **3** (7), 129-132 (1935).

⁹ Wright, S., these PROCEEDINGS, **24**, 372-377 (1938).

¹⁰ Wahlund, S., *Hereditas*, **11**, 65-106 (1928).

¹¹ Wright, S., *Genetics*, **24**, 538-552 (1939).

¹² Wright, S., *Ibid.*, **6**, 111-178 (1921).

¹³ Wright, S., *Amer. Naturalist*, **61**, 330-338 (1922).