86

THE DISTRIBUTION OF GENE RATIOS FOR RARE
MUTATIONS

Author's Note (CMS 19.204a)

The two subjects of this paper which deserve attention are (i) its bearing upon evolutionary theory, and (ii) the mathematical treatment of a class of functional equations.

In 1922 the author had attempted to examine whether the genetical situation indicated somewhat vaguely by observational data with quantitative characters could be more fully elucidated on the basis of the theoretical concepts of genetics. It was first necessary to form an opinion on the distribution of the gene ratios of factors exposed to different selective conditions, and then to ascertain their respective contributions to the genetic variance, and to selective progress.

The mathematical treatment in 1922 left much to be desired, since on reflection it appeared that more searching questions could be asked, and especially the probable progress of new mutations could be traced statistically, making the distinction between the frequency of the new gene and that of the old. The numerical error to which attention is called in the first section of this paper is of little consequence, since the corresponding distribution is unchanged. It is not on this point that I have differed from Professor Sewall Wright, but in that I do not share his conviction that evolutionary progress is favoured by the subdivision of a species into small, imperfectly isolated populations, save in the case stressed by Darwin in which the environmental conditions of these are sufficiently diverse to induce divergent evolutionary tendencies. Wright, on the other hand, has maintained that random survival in such populations leads to the testing of a greater variety of genotypes, and to the more rapid discovery of successful combinations, while my own studies have not led me to believe in any such effect, as a factor contributing to organic evolution.

To mathematicians the chief interest of the present paper lies in the treatment of the functional equations which arise in the exact examination of the terminal distributions, in the three cases considered, namely, (i) the steady state without mutation or selection, (ii) equilibrium with mutations but without selection, (iii) the equilibrium distribution for mutations having very small selective effects. With these distributions established, the probabilities of mutations of different classes establishing themselves, and their contribution to the frequencies at given gene ratios and to the heritable variance, are all calculable. The evolutionary consequences are developed in *Genetical Theory of Natural Selection* (Oxford, 1930), which is based for these questions on the present paper.

XVII.—The Distribution of Gene Ratios for Rare Mutations. By R. A. Fisher, Sc.D., F.R.S. (Rothamsted Experimental Station, Harpenden, Herts). *Communicated by* Professor GODFREY H. THOMSON.

1. INTRODUCTORY.

IN 1922 the author published a short paper, " On the Dominance Ratio," in the *Proceedings of the Royal Society of Edinburgh* (vol. xlii, pp. 321–341). Among other results, the conclusion was drawn that in the total absence of mutations and of selective survival, the quantity of variation, the variance, of an interbreeding group would decrease by reason of random survival, at a rate such that the "time of relaxation" was $4n$ generations, where n is the number breeding in each generation.

The variance after the lapse of T generations was found to be proportional to $e^{-T/4n}$.

During last year Professor Sewall Wright of Chicago has been good enough to send me in MS. an investigation in which, while confirming many other conclusions of my paper, he arrives at a time of relaxation of only $2n$ generations. Both periods are in most species so enormous that they lead to the same conclusion, namely, that random survival, while of great importance in conditioning the fate of an individual mutant gene, is a totally unimportant factor in the balance of forces by which the actual variability of species is determined. Nevertheless it will, I hope, minimise the confusion which every error is liable to cause if I put on record at once my acceptance of Professor Wright's value, and at the same time eradicate the error of my previous work by giving a more rigorous and comprehensive treatment of the whole subject. I may say that the previous conclusions as to the interpretation of the evidence for Mendelian dominance in the factors contributing to human variability are untouched, but that the rôle of mutations in maintaining the current genetic variability of a species may now be set in a much clearer light.

The error to be corrected lies in the derivation (p. 326) of the differential equation satisfied by the distribution of the frequency ratios of different factors, when none are subject to selective action. If the

two alternative genes in any locus appear in the ratio $p : q$, the variance of p after one generation of random breeding will be

$$\frac{pq}{2n},$$

where n is the number breeding in each generation. To avoid the inconvenience that this variance is a function of p, we may write

$$2p = 1 - \cos\theta, \quad 2q = 1 + \cos\theta$$

when

$$\delta p = \sqrt{pq}\,\delta\theta$$

and the variance of θ is therefore very nearly constant at the value $1/2n$.

Although, n being large, the values of θ after one generation of random breeding will be well represented by a normal distribution with constant variance, yet its mean will differ from zero by an amount of order $1/n$. This was overlooked in the previous treatment; to find the mean of $\delta\theta$ as far as terms in n^{-1}, we may write

$$\delta\theta = \frac{1}{\sqrt{pq}}\delta p - \frac{1-2p}{4pq\sqrt{pq}}(\delta p)^2 \ . \ . \ .,$$

then since the mean value of δp is strictly zero, while that of $(\delta p)^2$ is $pq/2n$, the mean value of $\delta\theta$ is seen to be

$$-\frac{1-2p}{8n\sqrt{pq}} = -\frac{1}{4n}\cot\theta.$$

This, of course, with values of n of many millions, is an exceedingly small quantity, but its effect is not negligible for the discussion required, for if

$$df = y\,d\theta$$

is the distribution of the values of θ for different factors, the flux past every value of θ due to random reproduction in one generation is changed from

$$-\frac{1}{4n}\frac{\partial y}{\partial\theta}$$

to

$$-\frac{y}{4n}\cot\theta - \frac{1}{4n}\frac{\partial y}{\partial\theta},$$

and the differential equation to be satisfied by y becomes

$$\frac{\partial y}{\partial \mathrm{T}} = \frac{1}{4n}\left\{\frac{\partial}{\partial\theta}(y\cot\theta) + \frac{\partial^2 y}{\partial\theta^2}\right\}, \qquad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (1)$$

instead of

$$\frac{\partial y}{\partial \mathrm{T}} = \frac{1}{4n}\frac{\partial^2 y}{\partial\theta^2},$$

the equation previously obtained; in both T is measured in generations.

2. The Solution for Steady Decay.

It so happens that the function of θ which satisfies the true equation in the case when, in the absence of mutations, the variance is steadily decaying owing to chance extinctions at the termini $\theta = 0$, $\theta = \pi$, is the same as the corresponding solution of the original erroneous equation, namely, $y = A \sin \theta$.

Substituting in the true equation we have

$$\frac{\partial A}{\partial T} \sin \theta = \frac{2A \sin \theta}{4n},$$

or

$$A = A_0 e^{-T/2n},$$

in place of

$$A = A_0 e^{-T/4n}$$

originally obtained. This confirms the value of $2n$ generations for the time of relaxation, found by a quite independent method by Professor Wright. The variance will then be halved by random survival in $2n \log 2 = 1\cdot 4n$ generations. The immense length of this period for most species shows how trifling a part random survival must play in the balance of influences which determines the actual variability.

3. Variability Maintained Constant by Mutations in the Absence of Selection.

If in equation (1) we put $\partial y/\partial T$ equal to zero, we may at once integrate the right-hand side in the form

$$\frac{\partial y}{\partial \theta} + y \cot \theta = -4nB,$$

where B is the net number of factors in each generation, the gene ratios of which flow past any specified value of θ, and the differential equation now simply represents the fact that this flux is the same for all values of θ. The equation may now be integrated giving the primitive,

$$y \sin \theta = A + 4nB \cos \theta$$

or

$$y = A \operatorname{cosec} \theta + 4nB \cot \theta. \qquad . \qquad . \qquad . \qquad . \qquad . \quad (2)$$

If we make the convention that mutations are equally frequent in supplying factors with θ near to zero and in supplying factors with θ near to π, the symmetrical solution

$$y = A \operatorname{cosec} \theta$$

R. A. Fisher,

will be appropriate; but, if we suppose all mutations occur at $\theta = 0$, then y should tend to zero at $\theta = \pi$, and the appropriate form is

$$y = 4n\mathrm{B}(\operatorname{cosec} \theta + \cot \theta). \qquad . \qquad . \qquad . \qquad . \qquad . \qquad (3)$$

In either case the integral of y to the limit of its range at $\theta = 0$ fails to converge, so that the relation between the number of factors maintained and the rate of mutation cannot be made out without an investigation of the terminal conditions. Before passing on to consider
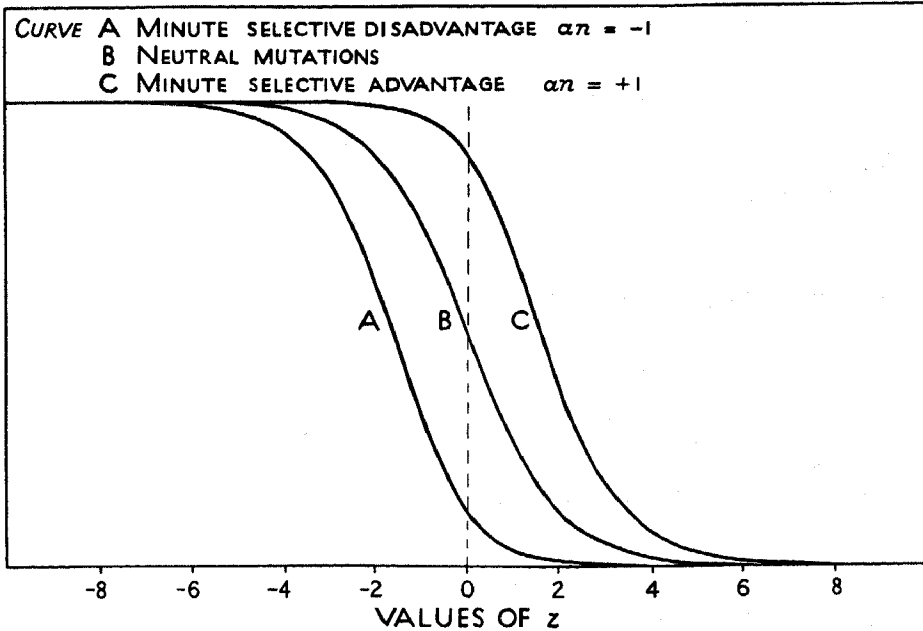

CURVE A MINUTE SELECTIVE DISADVANTAGE $an = -1$
B NEUTRAL MUTATIONS
C MINUTE SELECTIVE ADVANTAGE $an = +1$

VALUES OF z

Fig. 1.—Frequency curves of logarithmic gene ratio, z, for different levels of selective advantage; note that the frequency ordinate is highest for the most extreme admissible negative values of z, and remains nearly constant over a range which is extremely sensitive to small selective intensities.

these, we may consider the distribution now obtained as a distribution not of θ but of the more convenient variate $z = \log (p/q)$, the logarithmic gene ratio. The frequency distribution (3) may be represented on the scale of z by noting that

$$\begin{aligned} df &= 4n\mathrm{B} (\operatorname{cosec} \theta + \cot \theta)d\theta \\ &= 4n\mathrm{B}q\,dz \\ &= 4n\mathrm{B}\frac{dz}{1 + e^z}. \end{aligned}$$

This frequency distribution is illustrated by curve B in fig. 1. The frequency ordinate is nearly constant for values of z less than -4, at

which point the mutant gene occupies nearly 2 per cent. of the available loci; it falls to half its previous value when z is raised to zero, when 50 per cent. of the loci are occupied by mutant genes. For higher gene ratios still, the frequency falls rather rapidly to zero. Since the frequency ordinate is nearly constant for high negative values of z, the total frequency maintained depends on how far the curve may be carried to the left, or how large (negative) values the logarithmic frequency ratio, z, may have. Evidently this will depend on the size of the population, and an exact treatment will evidently require an examination of the terminal conditions.

4. DISTRIBUTIONS EXPRESSED BY FUNCTIONAL EQUATIONS.

A very powerful method of approach was indicated, but not utilised, in the previous paper. If

$$p_0, p_1, p_2, \ldots$$

are the probabilities of an individual gene carried by a member of the species, being represented in the next generation in 0, 1, 2, . . . offspring, we may define a function

$$f(x) = p_0 + p_1 x + p_2 x^2 + \ldots,$$

for values of x between 0 and 1, and it has been shown that to consider the offspring of two individuals instead of one, we have only to substitute

$$\{f(x)\}^2 \quad \text{for} \quad f(x).$$

Consequently, if the number of factors in which the rarer gene occupies 1, 2, 3, . . . loci are given by $\pi_1, \pi_2, \pi_3, \ldots$, and if

$$\phi(x) = \pi_1 x + \pi_2 x^2 + \pi_3 x^3 + \ldots,$$

the effect upon ϕ of random breeding for one generation is to substitute

$$\phi\{f(x)\} \quad \text{for} \quad \phi(x).$$

In practice we shall require to use the form

$$f(x) = e^{x-1},$$

and if we first take the case of extinction of genes without mutation, the distribution of gene frequencies, which maintains its form, while one factor is extinguished in each generation must satisfy the functional equation

$$\phi(e^{x-1}) = \phi(x) + \tfrac{1}{2},$$

for the distribution being symmetrical, half the extinctions may be taken to be reductions from 1, 2, 3, . . . loci to zero, and half to be increases to $2n$.

The corresponding equation for the generating function ϕ, for the case of a distribution in equilibrium with mutations at the rate of one in each generation, is

$$\phi(e^{x-1}) - \phi(x) = 1 - x,$$

for a mutation may be represented as an increase of unity in the number of genes occupying one locus only, and a corresponding decrease of the (indefinite) number occupying no loci.

The solutions of these equations will be shown to correspond with the solutions of the differential equations obtained above, and to admit in addition of an investigation of the terminal condition.

5. The Function u_ν.

In order to solve the functional equations, we define a function u_ν of a single real variable ν, which shall satisfy the condition

$$u_{\nu+1} = e^{u_\nu - 1},$$

starting from the arbitrary value $u_0 = 0$. The values of $u_1, u_2, \ldots$ may now be obtained by direct substitution, and these evidently tend to unity as a limit. To obtain a form for large values of ν, we may put

$$v_\nu = \frac{1}{1 - u_\nu}$$

and obtain

$$v_{\nu+1} = \frac{1}{1 - e^{-1/v_\nu}} = v_\nu + \frac{1}{2} + \frac{1}{12 v_\nu} - \frac{1}{720 v_\nu{}^3} + \frac{B_6}{6!\, v_\nu{}^5} + \ \cdots$$

Where B_6, etc., stand for the Bernoulli numbers

$$B_6 = \frac{1}{42} \qquad B_8 = -\frac{1}{30} \qquad B_{10} = \frac{5}{66} \qquad B_{12} = -\frac{691}{2730}.$$

It appears from the recurrence formula of v that when v is large a first approximation is given by

$$v = \tfrac{1}{2}\nu,$$

and substituting this in the third term of the expression, we obtain the second approximation

$$v = \tfrac{1}{2}\nu + \tfrac{1}{6} \log \nu,$$

the error of which will tend to a finite limit, as v tends to infinity. Equally

$$\tfrac{1}{2}\nu - v + \tfrac{1}{6} \log v$$

must tend to a constant value as v is increased indefinitely.   Let $-C$ stand for this constant, and let

$$w_v = \tfrac{1}{2}v - v + \tfrac{1}{6}\log v + C,$$

then we may obtain an expansion for w in inverse powers of v; for the recurrence formula provides that

$$w_{v+1} - w_v = \frac{1}{2} - \left\{ \frac{1}{2} + \frac{1}{12v_v} - \frac{1}{720v_v{}^3} + \cdots \right\}$$
$$+ \frac{1}{6}\log\left\{ 1 + \frac{1}{2v_v} + \frac{1}{12v_v{}^2} - \frac{1}{720v_v{}^4} + \cdots \right\}$$

and expanding this expression we obtain, dropping the suffix of v,

$$-\frac{v^{-2}}{144} + \frac{v^{-3}}{720} + \frac{v^{-4}}{24 . 720} - \frac{v^{-5}}{42 . 720} - \frac{v^{-6}}{1512 . 720} + \frac{v^{-7}}{1680 . 720} + \frac{1473 v^{-8}}{336 . 720^2} - \frac{v^{-9}}{924 . 72 . 720} \cdots$$

as an expansion of $w_{v+1} - w_v$; the first term shows that the leading term in the expansion of w is $1/72v$ for

$$\frac{1}{v_{v+1}} - \frac{1}{v_v} = -\frac{1}{2v^2} + \frac{1}{6v^3} - \frac{1}{24v^4} + \cdots,$$

and similar expansions may be obtained for $\left(v_{v+1}^{-2} - v_v^{-2} \right)$, and so on.   We thus obtain

$$w = \frac{v^{-1}}{72} + \frac{v^{-2}}{1080} - \frac{v^{-3}}{108 . 144} - \frac{71 v^{-4}}{168 . 72^2} - \frac{8759 v^{-5}}{630 . 720^2}$$
$$+ \frac{31 v^{-6}}{81 . 720^2} + \frac{1637 v^{-7}}{1008 . 720^2} - \frac{20879093 v^{-8}}{9504 . 840 . 720^2}.$$

While the last three coefficients are all less than 10^{-5}, they show no such a decided tendency to decrease as would justify our evaluating the constant C by putting $v = 1$, $\nu = 0$, a substitution which shows C to exceed by unity the limit of the sum of the coefficients.   We may, however, use the larger values of v found by the recurrence formula for somewhat higher integral values of ν.

For example, at $\nu = 5$, the last three terms in w are less than 10^{-9}, so that w will not be much in error in the ninth place of decimals; u is found by direct substitution in the recurrence formula to be $\cdot 73192\ 31844$ and

$$v - \tfrac{1}{6}\log v - \tfrac{1}{2}v + w = 1\cdot 01464,8607$$

gives a value of C nearly right to the last figure.   To improve much upon this, it would be necessary to work to more than 10 places in the calculation of u.   As a check, working to 14 places up to u_{10}, where the last term retained in w is about 2×10^{-12}, the value was found to be $1\cdot 01464\ 86071\ 7$, a value which shows that the apparent precision attained by the series is not illusory.

6. Solutions of the Functional Equations.

If, in the equation

$$\phi(e^{x-1}) - \phi(x) = \tfrac{1}{2},$$

we substitute

$$x = u_\nu$$

we have

$$\phi(u_{\nu+1}) - \phi(u_\nu) = \tfrac{1}{2},$$

which is satisfied if ϕ is the same function of x as $\tfrac{1}{2}\nu$ is of u. But we know that

$$\tfrac{1}{2}\nu = \frac{1}{1-u} + \frac{1}{6}\log(1-u) - C + \frac{1-u}{72} + \frac{(1-u)^2}{1080} + \cdots;$$

hence, apart from a finite fraction of the frequency, ϕ may be expanded in powers of x in the form

$$\left.\begin{aligned}
\phi(x) &= \frac{x}{1-x} + \frac{1}{6}\log(1-x) \\
&= \frac{5}{6}x + \frac{11}{12}x^2 + \frac{17}{18}x^3 + \cdots
\end{aligned}\right\}, \qquad . \qquad . \qquad . \qquad (4)$$

showing that in the distribution of gene ratios appropriate to steady extinction without mutation or selection, the frequency of factors represented in k loci must, when k is large, tend to unity. Since each step increases the gene proportion p by $1/2n$, we have, apart from the extremes of the distribution,

$$\begin{aligned}
df &= 2ndp \\
&= n\sin\theta\, d\theta,
\end{aligned}$$

in agreement with the solution obtained for this case from the differential equation. The total number of factors at all frequencies will be

$$2n - \tfrac{1}{3}(\gamma + \log 2n) - \cdot01464,86071,7,$$

(where γ is Euler's constant $0\cdot5772 15664$), the remainder of which is negligible compared with the first term, twice the number of individuals breeding in each generation, thus verifying the rate of decay to be 1 in $2n$ in each generation.

The exact treatment of the terminal frequencies, which shall account for the distribution of the finite quantity $0\cdot014649$, omitted from expression (4), evidently requires the differential coefficients of $\tfrac{1}{2}\nu$ with respect to u, at the value $u = 0$. Since the series for w in powers of $(1-u)$ is itself doubtfully convergent at this value, its differential coefficients may be still less relied upon to converge; we therefore require reduction formulæ for these coefficients.

From the recurrence formula

$$u_\nu - 1 = \log u_{\nu+1},$$

we have, differentiating with respect to ν,

$$\frac{du_\nu}{d\nu} = \frac{1}{u_{\nu+1}} \frac{du_{\nu+1}}{d\nu}$$

or

$$\frac{d\nu}{du_\nu} = u_{\nu+1} \frac{d\nu}{du_{\nu+1}},$$

from which the value of $d\nu/du$ for a lower value can be obtained with the same relative precision as at the higher.

We may write the relation in the form

$$\nu_0{}^I = u_1 \nu_1{}^I$$

with the understanding that any suffixes differing by unity can be substituted for those indicated. Since also

$$\frac{d}{du_0} = u_1 \frac{d}{du_1}$$

we can at once derive the further relations

$$\nu_0{}^{II} = u_1 \nu_1{}^I + u_1{}^2 \nu_1{}^{II}$$
$$\nu_0{}^{III} = u_1 \nu_1{}^I + 3u_1{}^2 \nu_1{}^{II} + u_1{}^3 \nu_1{}^{III}$$
$$\nu_0{}^{IV} = u_1 \nu_1{}^I + 7u_1{}^2 \nu_1{}^{II} + 6u_1{}^3 \nu_1{}^{III} + u_1{}^4 \nu_1{}^{IV}$$
$$\nu_0{}^{V} = u_1 \nu_1{}^I + 15u_1{}^2 \nu_1{}^{II} + 25u_1{}^3 \nu_1{}^{III} + 10u_1{}^4 \nu_1{}^{IV} + u_1{}^5 \nu_1{}^{V}$$
$$\nu_0{}^{VI} = u_1 \nu_1{}^I + 31u_1{}^2 \nu_1{}^{II} + 90u_1{}^3 \nu_1{}^{III} + 65u_1{}^4 \nu_1{}^{IV} + 15u_1{}^5 \nu_1{}^{V} + u_1{}^6 \nu_1{}^{VI},$$

and so on.

From these it is evident that, knowing the series of differential coefficients of ν with respect to u at any integral value such as $\nu = 5$, the corresponding series may be obtained step by step down to $\nu = 0$. In this way we obtain, for the series of coefficients

$$\frac{1}{2} \cdot \frac{1}{k!} \frac{d^k \nu}{du^k},$$

the values:

k.	True Value.	Approximation.	Error.	Remainder.
1	·818,202,78	·833,333,33	+ ·015,130,55	− ·000,481,94
2	·916,762,37	·916,666,67	− ·000,095,70	− ·000,386,24
3	·944,923,44	·944,444,44	− ·000,479,00	+ ·000,092,76
4	·958,266,12	·958,333,33	+ ·000,067,21	+ ·000,025,55
5	·966,634,08	·966,666,67	+ ·000,032,59	− ·000,007,04
6	·972,225,35	·972,222,22	− ·000,003,13	− ·000,003,91

The table shows in parallel columns (i) the values derived from the reduction formula from those at $\nu = 5$, (ii) the values given by the approximation $1 - 1/6k$, (iii) the differences between these values, (iv) the remainder

of the deviations needed to make up the total $+ \cdot014{,}648{,}61$. The mere fact that this difference decreases at every step, and is finally reduced to a very trifling value, indicates that the errors shown in the first six terms, small as they are, are far greater than those to be anticipated at higher values of k.

The second functional equation, appropriate for variability maintained by a constant supply of mutations, has the form

$$\phi(e^{x-1}) - \phi(x) = 1 - x \, ;$$

substituting here $u_\nu = x$, we have

$$\phi(u_{\nu+1}) - \phi(u_\nu) = 1 - u_\nu \, ;$$

but from the recurrence formula

$$\frac{d}{d\nu} u_{\nu+1} = e^{u_\nu - 1} \frac{d}{d\nu} u_\nu \, ,$$

hence

$$\log \frac{d}{d\nu} u_{\nu+1} - \log \frac{d}{d\nu} u_\nu = u_\nu - 1 \, ;$$

so the functional equation may be written

$$\phi(u_{\nu+1}) + \log \frac{d}{d\nu} u_{\nu+1} = \phi(u_\nu) + \log \frac{d}{d\nu} u_\nu,$$

which is satisfied if

$$\phi(u) + \log \frac{du}{d\nu}$$

is a constant. Since by its definition $\phi(0) = 0$, we thus find that ϕ is the same function of x as

$$\log \nu^I - \log \nu_0{}^I$$

is of u. The approximate form,

$$\nu = \frac{2}{1-u},$$

gives

$$\log \nu^I = \log 2 - 2 \log (1 - u) \, ;$$

so that an approximation is given by

$$\phi(x) = -2 \log (1 - x) = 2x + \tfrac{2}{2} x^2 + \tfrac{2}{3} x^3 + \ldots,$$

which will account for the whole frequency save for

$$\log 2 - \log \nu_0{}^I = \cdot 200{,}645{,}07.$$

The frequency at $p = k/2n$ is now found to be $2/k$ when k is large, or the frequency element to be

$$df = \frac{2dp}{p} = \frac{\sin \theta d\theta}{1 - \cos \theta} = \frac{(1 + \cos \theta) d\theta}{\sin \theta}$$

$$= (\operatorname{cosec} \theta + \cot \theta) d\theta,$$

thus confirming the solution obtained by means of the differential equation. By the present method, however, we can evaluate the total number of factors maintained in the specific variance by one mutation in each generation as

$$2(\gamma + \log 2n) + \cdot 200,645,07,$$

the value of which ranges from $30 \cdot 372$ to $57 \cdot 903$ as n changes from 10^6 to 10^{12}.

The exact terminal frequencies for this case may be obtained from

$$\nu^I = \nu_0{}^I + u\nu_0{}^{II} + \frac{u^2}{2}\nu_0{}^{III} + \ldots;$$

hence

$$\log \frac{\nu^I}{\nu_0{}^I} = \log \left\{ 1 + u\frac{\nu_0{}^{II}}{\nu_0{}^I} + \frac{u^2}{2}\frac{\nu_0{}^{III}}{\nu_0{}^I} + \ldots \right\},$$

which, on expansion in powers of u, yields the frequency coefficients of the following table:

k.	True Value.	Approximation.	Error.	Remainder.
1	2·240,917,26	2·000,000,00	− ·240,917,26	+ ·040,272,19
2	·953,776,16	1·000,000,00	+ ·046,223,84	− ·005,951,65
3	·671,863,62	·666,666,67	− ·005,196,95	− ·000,754,70
4	·501,095,71	·500,000,00	− ·001,095,71	+ ·000,341,01
5	·399,761,71	·400,000,00	+ ·000,238,29	+ ·000,102,72

showing, as in the previous case, that the discrepancy from the approximate formula is confined, for all practical purposes, to the extreme terminal values.

7. THE EFFECTS OF A SMALL SELECTIVE ADVANTAGE OR DISADVANTAGE.

The method of functional equations has now made clear in what way the terminal forms of the solutions of the differential equations should be interpreted; we may therefore now consider the differential equation appropriate to mutations enjoying a small selective advantage, such supplying in all probability the greater portion of the genetic changes taking place in the course of evolution.

If a is the selective advantage of the mutant genes, the flux past any value of θ may be written as

$$\frac{1}{2}ay \sin \theta - \frac{y}{4n}\cot \theta - \frac{1}{4n}\frac{\partial y}{\partial \theta}, \qquad . \qquad . \qquad . \qquad . \quad (5)$$

provided a^2 may be neglected. It should be noted that the equation will only be correct if a^2n is a small quantity, and this limits its application

to very minute selective intensities. For these, however, the equilibrium condition of constant flux yields a differential equation for y of the first order, which may be written

$$y^I - (2an \sin \theta - \cot \theta)y = -4an\mathrm{A}$$

and may be integrated in the form

$$ye^{2an \cos \theta} \sin \theta = 2\mathrm{A}e^{2an \cos \theta} + \mathrm{B}.$$

Since $\cos \theta = -1$ when $\theta = \pi$, the condition that at this terminus, where no mutations are occurring, $y \sin \theta$ should be zero, is that

$$\mathrm{B} = -2\mathrm{A}e^{-2an},$$

giving the solution

$$y = 2\mathrm{A} \operatorname{cosec} \theta(1 - e^{-2an(1+\cos \theta)}).$$

At the terminus $\theta = 0$ this will correspond to the distribution in equilibrium with one mutation per generation if

$$\mathrm{A} = \frac{2}{1 - e^{-4an}},$$

so that the distribution adopted is

$$df = yd\theta = 4 \operatorname{cosec} \theta \frac{1 - e^{-2an(1+\cos \theta)}}{1 - e^{-4an}}d\theta$$

$$= \frac{2}{1 - e^{-4an}}\left\{1 - e^{-4an/(1+e^z)}\right\}dz.$$

Fig. 1, C, shows the distribution on the scale of z for $an = 1$, while the curve A on the same figure shows the curve for factors at a minute selective disadvantage, $an = -1$.

While the curve of continuous distribution represents the frequencies well over that part of the range in which $\pm z$ is considerably less than $\log n$, the termini of the distribution are subject to adjustments similar to those investigated in the absence of selection. Thus at the terminus $\theta = 0$, the frequency element $4 \operatorname{cosec} \theta d\theta$ will be replaced by a series of frequencies for 1, 2, 3 genes given approximately by the series $2, 1, \frac{2}{3}, \ldots$, while at the terminus $\theta = \pi$ we have the frequency element

$$\frac{8an \operatorname{cosec} \theta(1 + \cos \theta)}{1 - e^{-4an}}d\theta,$$

the limit of which is

$$\frac{4an \sin \theta d\theta}{1 - e^{-4an}}.$$

the form appropriate to steady extinction without mutation, the rate of extinction at this terminus being

$$\frac{2a}{1 - e^{-4an}}$$

in each generation; this rate may equally be obtained by substituting in the assumed flux of factors, Aa, the solution

$$A = \frac{2}{1 - e^{-4an}}.$$

The probability of a mutant, enjoying a small selective advantage a, spreading until it establishes itself throughout the entire population is thus found to be $2a/(1-e^{-4an})$; it is easy to see that with an indefinitely large population, or in any case if $4an$ is large, this expression reduces to $2a$. Thus a mutation conferring a selective advantage of 1 per cent.
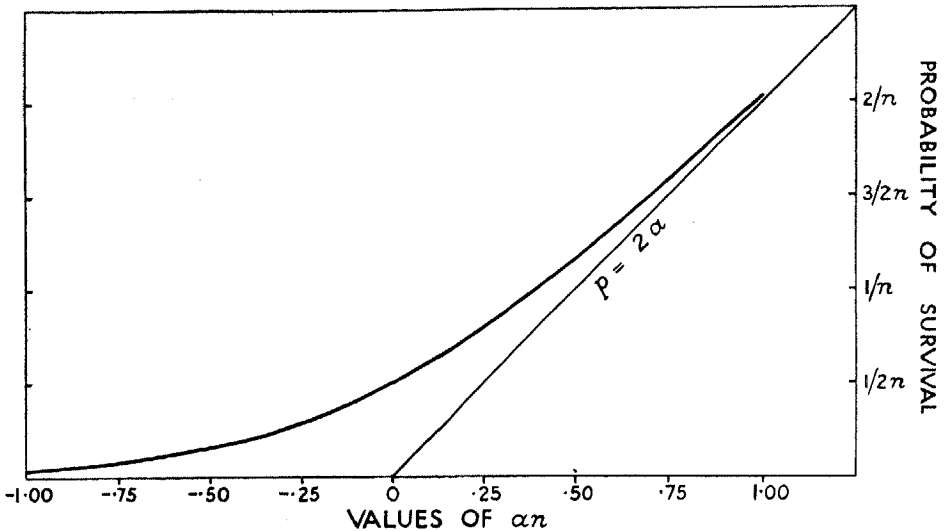

FIG. 2.—Probability of success for mutations having a very minute selective advantage or disadvantage.

will have practically a 2 per cent. chance of establishing itself. The value of this probability affords a means of checking the accuracy of our solution for values of a which, while still small, are large enough to vitiate the condition that a^2n should be small, the condition subject to which the differential equation has been obtained. For, in an indefinitely large population, the exact probability of ultimate survival is given by $1-U$, where U satisfies the equation

$$U = e^{c(U-1)}$$

and $c = e^a$.

Writing P for $1-U$, we have

$$cP = -\log(1-P) = P + \tfrac{1}{2}P^2 + \tfrac{1}{3}P^3 + \ldots$$

which is satisfied by

$$P = 2a - \frac{5}{3}a^2 + \frac{7}{9}a^3 - \frac{131}{540}a^4 \ldots,$$

showing that when a is small, even though a^2n may be large, the value $2a$ is a good approximation to the probability of survival.

When an is not large, the probability $2a/(1-e^{-4an})$ tends to the small but finite value $1/2n$, as a tends to zero, and is finite even for negative values of a; its value changes, however, very rapidly as we pass from small negative to small positive selective advantages. Fig. 2 shows the course of this change. It will be observed that the probability of success increases over fiftyfold (e^4) in passing from $an=-1$ to $an=+1$, that is from distribution A to distribution C of fig. 1.

8. Contributions to the Variance.

In previous work the calculations of the quantitative contribution of different classes of factors to the total variance of the species has been much complicated by the widespread phenomenon of dominance, and by our ignorance of the conditions under which factors may be expected to be dominant or recessive. With the extension of genetical experience it now seems probable that the recessive character is characteristic of deleterious mutations which have long persisted in regular occurrence in the species or group in which they are known; and in the case of stable dimorphism, determined by a simple Mendelian factor, of the less favoured of the two phenotypes (genetic selection being necessarily absent or balanced in such cases). Consequently, it is probable that the new and sometimes favourable mutations on which evolutionary progress must rely are neither dominant nor recessive, but have heterozygotes of an intermediate character. Their contribution to the variance will then be simply proportional to pq or to $\sin^2\theta$, and the total variance supplied by mutations having a selective advantage a, for each one occurring per generation, will be proportional to

$$\int_0^\pi \frac{1-e^{-2an(1+\cos\theta)}}{1-e^{-4an}}\sin\theta\,d\theta$$

or to

$$\frac{2}{1-e^{-4na}}-\frac{1}{2an}.$$

For negative values of an exceeding 2 this is nearly equal to $1/2an$, while for large positive values it approaches a constant value of 2, passing through the value unity when $a=0$. Its course is shown in fig. 3. If in the immediate neighbourhood of neutrality beneficial and harmful mutations are equally frequent, the variance contributed by mutations in a given range of utility will increase sharply as the utility is increased past the

point of neutrality.   For higher values of a there is every reason to suppose that the supply of mutations falls off, so that there will be a maximum in the contributions to the specific variance ascribable to slightly beneficial mutations.   The frequency of harmful mutations probably increases considerably with the extent of the injury up to high values of $-an$; in spite of the decrease in the average contribution of each mutation to the specific
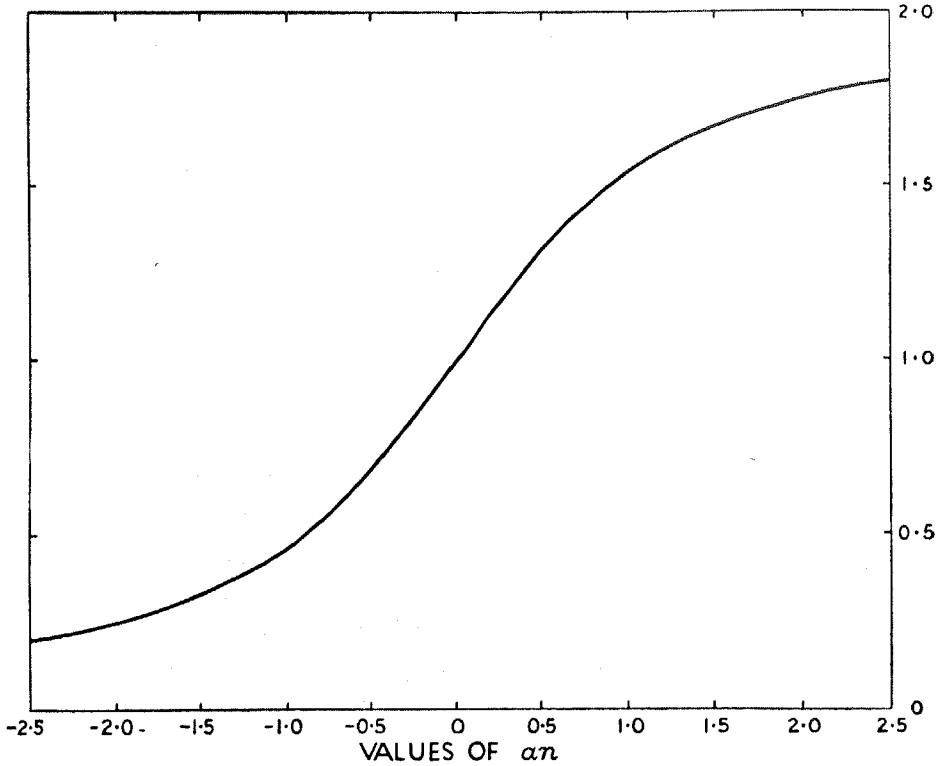

Fig. 3.—Proportionate contribution to the specific variance for factors of varying selective advantage.

variance, there may thus well be a second maximum, representing the contribution of definitely deleterious mutations which are constantly kept rare by counter-selection.   This latter maximum is of no direct importance for evolutionary change, though the effects of Natural Selection in reducing persistent mutants of this class to the recessive condition seem to be of the greatest interest.   The portion of the genetic variance to which evolutionary progress is to be ascribed may be a large or a small portion of the whole observable variance, but seems in any case to be concentrated in groups of factors each determining a very minute selective advantage.

9. Summary.

The discussion of the distribution of the gene ratio of the author's paper of 1922 is amended by the use of a more exact form of the differential equation to be satisfied. It appears that the time needed to halve the variance by random extinction of genes in the total absence of mutations should be 1·4 instead of 2·8 times the number of potential parents in each generation. Either value shows that the loss of variance due to this cause is too trifling to be appreciable in the balance of causes which maintain the actual genetic variability of species.

The same correction alters the distribution appropriate for the maintenance of variability at a fixed level by mutations in the absence of selection. The new solution closely resembles the form previously obtained and now confirmed for the practical case in which selection is present. The method of differential equations, however, fails to deal satisfactorily with these cases, owing to the failure of the integrals to converge at the termini representing cases in which one or other allelomorph is extremely rare.

A method of functional equations is developed for dealing with the termini, and is shown to lead to the same solutions as the amended differential equations in the central portion of the range for which the latter are valid, and further to give the terminal distribution of rare allelomorphs. The method requires the investigation of a continuous function u_ν of argument ν satisfying the recurrence formula

$$u_{\nu+1} = e^{u_\nu - 1}.$$

From the asymptotic form of this function its expansion in the neighbourhood of $u = 0$ is derived, giving the frequencies of the required distributions.

Exceedingly minute values for the selective advantage or disadvantage make a great difference to (i) the chance of success of a mutation and (ii) the contribution of such mutations to the specific variance. The order of magnitude to be considered is the inverse of the population of the species. The neutral zone of selective advantage in the neighbourhood of zero is thus so narrow that changes in the environment, and in the genetic constitution of species, must cause this zone to be crossed and perhaps recrossed relatively rapidly in the course of evolutionary change, so that many possible gene substitutions may have a fluctuating history of advance and regression before the final balance of selective advantage is determined.