A Mathematical Theory of Evolution, Based on the Conclusions of Dr. J. C. Willis, F.R.S.
Author(s): G. Udny Yule
Source: *Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character*, Vol. 213 (1925), pp. 21-87
Published by: The Royal Society
Stable URL: http://www.jstor.org/stable/92117
Accessed: 05/09/2010 19:04

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/page/info/about/policies/terms.jsp. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/action/showPublisher?publisherCode=rsl.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

The Royal Society is collaborating with JSTOR to digitize, preserve and extend access to *Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character*.

http://www.jstor.org

II.—*A Mathematical Theory of Evolution, based on the Conclusions of Dr. J. C. Willis, F.R.S.*

By G. UDNY YULE, *C.B.E., M.A., F.R.S., University Lecturer in Statistics, and Fellow of St. John's College, Cambridge.*

(Received June 11, 1923.—Read February 7, 1924.)

CONTENTS.

INTRODUCTION.

The following work is founded on that conception of evolution, the most recent and precise formulation of which is due to Dr. J. C. WILLIS,* and represents an attempt to develop the quantitative consequences of the conception.

By his statistical studies of distribution Dr. WILLIS was led to two conclusions :—

(1) Species occupying large areas are, *on the whole*, older than those occupying small areas, provided that allied forms are compared.

It is to be noted that by the "area occupied" by a species is meant, not the total acreage actually covered by individuals of the species in question, but the area of the contour drawn round its outermost stations—the "areal range" of the species as it might be termed.

(2) The larger genera, *i.e.*, the genera with most species, are, *on the whole*, those occupying the larger areas.

The first is the conclusion which Dr. WILLIS briefly summarises as the principle of "Age and Area," the second as the principle of "Size and Space." From these statistical inductions the deduction follows, time being the ruling factor in both cases :—

(3) The larger genera are, *on the whole*, the older.

---

* 'Age and Area,' Cambridge University Press, 1922, and numerous papers therein cited published during the past fifteen years.

[Published May 14, 1924.]

The size of the genus, that is to say, is not an absolute measure of its age but is an *index* to it, very much as the total number of children born to a marriage is an index to the duration of that marriage.

On the Darwinian view that species are continually dying out—that a species rises, flourishes and dies, superseded by the more advantageous form—a species occupying a very small area may be young, but it is equally likely or more likely to be old (a dying species). On Darwin's own view that the whole body of individuals in a species becomes altered together,* the young, species must be found occupying a large area at once, and the species occupying a small area could only be a "dying" species. On the Darwinian view therefore either there need be no relation between Age and Area, or there would be a negative relation, species occupying small areas being on the whole the oldest. Similarly, on the Darwinian view a genus of a few, or of only one, species may be either young or old—a dying genus—and there need be no necessary relation between Age and Size. That species occupying very small areas, and the species of monotypic genera are mainly "relic" forms, is, I gather, the predominant Darwinian view. Dr. WILLIS's conclusions are inconsistent with that view.

We are accordingly led directly to the mutational view of evolution that has been held by more than one writer both before Darwin and after: the view that specific differences arise, not cumulatively by the natural selection of slight favourable variations, but at once *per saltum* as "mutations." On this view a new form must necessarily occupy a small area, and the required correlation between age and area follows at once. The new form may possibly differ so largely from the parent stock as to be classed not merely in a new species, but even in a new genus or a new family. If the new form is badly unsuited to its immediate surroundings, it will be killed out at once; natural selection will strangle the species at birth. But if the species hold out long enough to spread over any considerable range it is highly unlikely to be extinguished by natural selection, for it has a chance to find the niches that suit it best. The operation of natural selection, on Dr. WILLIS's view, is not denied: what is denied is *the origin of species by natural selection,* and what is affirmed is that natural selection operates mainly on the very young species before it has time to spread. The species that is killed out at birth we may regard as a non-viable mutation, and non-viable mutations may be to all intents and purposes disregarded: they are not likely to be seen in life or to be found in the geological record. When we speak in future therefore of a "specific mutation" or of a "generic mutation" it may be understood that a viable mutation is implied. If attention is confined to viable mutations, we have on the present view little or no further concern with the extinction of species by natural selection, or indeed with any extinction of species, so long as conditions are constant. It seems doubtful, at the least, whether we have any reason to predicate death as normal for a species in the same sense that it may be normal for one of the higher plants or the higher animals.

---

* Letter to G. Bentham, November 25, 1869: "More Letters of Charles Darwin," vol. I, 379.

How then, it is naturally asked, do you account for the number of extinct species in the geological record ? The answer is by their having been killed out ; in the main, as it seems to me, by agencies which act less continuously than spasmodically and may fairly be described as cataclysmic, *e.g.*, the sinking of land under the sea, the onset of a glacial epoch, or other great change of climate, such as desiccation. I am not for a moment denying uniformitarianism or demanding catastrophism, but the term " cataclysm " seems to me quite applicable. In the case of land sinking below sea-level it applies in the literal, derivative, sense. But " cataclysm " has now come to carry the meaning of any overwhelming disaster, not a deluge alone, with some implication, I think, of the disaster being sudden and extensive. Such " disasters " as I have in mind are certainly extensive in their range. They are not " sudden " on a scale of time on which a year is regarded as a long period, but on the geological time-scale they may well be regarded as sudden. So far as I can judge from the evidence, the age of the flowering plants is of the order of 100 million years (*cf.* Section VI). Suppose this length of time to be shown on a scale 100 inches long. On such a scale one-hundredth of an inch represents 10,000 years, and if changes of climate are shown by a curve to such a scale, a change say from temperate to glacial or from glacial to temperate accomplished in 10 or 20,000 years will appear to be absolutely sudden. This is the order of time estimated to have elapsed since the culmination of the last glacial epoch.

Now a " cataclysm " in the sense explained would kill out the whole or a great part of the organic life existing in the region over which it swept. It would necessarily act *differentially*, for some only of all the species in the world would lie within its range, but it would not act *selectively* if the cataclysm was overwhelming and the extermination complete : the species exterminated would be killed out not because of any inherent defects but simply because they had the ill-luck to stand in the path of the cataclysm. Only in so far as extermination was not complete there would be selective action, the species surviving being on the whole the fitter to survive in the new circumstances. But even so, the selection would be only interspecific, and selective of already existing species.

If we then put aside for the moment " cataclysmic " destruction of species, we may consider " free " or undisturbed evolution as proceeding without any species being killed out or " dying " out, the number of species continually increasing without a break as mutations occur.

With this brief introduction, the assumptions made for the mathematical work may be stated as follows :—

(1) Within any species, in any interval of time, an " accident " may happen that brings about a (viable) " specific mutation," *i.e.*, the throwing of a new form which is regarded as a new species, but a species within the same genus as the parent.

The chance of this occurrence in any assigned interval of time (an hour, or a year or a century) is taken as the same for all species within the group considered and as constant for all time.

(2) Within any genus, in any interval of time, an "accident" may happen that brings about the throwing of a (viable) "generic mutation," *i.e.*, a new form so different from the parent that it will be placed in a new genus.

The chance of this occurrence in any assigned interval of time is similarly taken as the same for all genera within the group considered and as constant for all time.

These statements call for some comment.

(*a*) The assumptions that the chances of specific (or generic) mutation are identical for all forms within the group considered and constant for all time are unlikely to be in accordance with the facts, but have to be made to simplify the work.

(*b*) The chance of a generic mutation or a specific mutation occurring is not taken as dependent on the number of individuals in the genus or species. It cannot be said with certainty whether it is so dependent or no till we know the nature of a mutation, though there very probably is such dependence, and there must be dependence if the "accident" that brings about a mutation is of the nature of a breakdown in the mechanism which forms the germ-cells. But in any case, and apart from the great additional complexity it would introduce, I was compelled to ignore the number of individuals as I could feel no confidence in any assumption as to the function which number of individuals, that is the "size of the species," would be of the age of the species. At first, as the species spreads, the number of individuals must tend to increase. But over the very long periods which have to be considered there must be a countervailing tendency to ultimate decrease in the number of individuals, owing to the increase in the number of species. The area available being limited, the tendency, as it seems to me, must be towards greater and greater numbers of species and fewer individuals in each.* Possibly the fact that ignoration of the number of individuals leads to results in close accordance with the data may only indicate this tendency first to increase and then to decrease in the size of the species.

(*c*) The possible effect of size of genus (number of species in the genus) on the chance of a generic mutation is also ignored. This assumption may or may not be correct, but was deliberate. The generic characters are regarded as representing a main position of stability, and the chance of occurrence of a transfer from one main position of stability to another is regarded as independent of the number of minor positions of stability (species) which may have been taken up within the main position (genus).

(*d*) If A, B, C, D are the existing genera and one of them throws a generic mutation, it is assumed that this will represent a new genus E. The possibility is ignored that A, for example, may throw a generic mutation which will be classed under the already existing genus B. Again this is an assumption based on complete ignorance: it seems the simplest assumption that one can make, and that is all that can be said.

We are completely ignorant as to the nature of specific mutations, and if my conclusions

---

* During the discussion on Darwinism at the Hull meeting of the British Association, 1922, the point was made that indefinite increase in the number of species was impossible owing to the limited space available. I see no practical limit to the number of species, but only to the total number of individuals.

as to their extreme rarity are correct I do not think we are at all likely in any near future to obtain direct information as to their nature. The whole scheme has therefore, I think justifiably, been simplified to the greatest possible extent to enable us to get a first notion, in quantitative terms, of the possible consequences of the conceptions, and to compare them as far as possible with the facts.

I will now endeavour to summarise the conclusions reached in general terms which I hope may be comprehensible to the non-mathematical biologist. The conclusions result, of course, from the assumptions, and simply exhibit their consequences. In so far as the deductions do not agree with the known facts, the assumptions are probably incorrect or incomplete. In so far as we find agreement, or the more nearly we find agreement, the assumptions are *probably* correct. It is only by a full development of the consequences that an effective test of the assumptions can be made.

In Section I the question is considered, what is the average law of increase of the number of species within the genus. Within any single genus, starting at zero time from one primordial species, the increase in the number of species will, of course, proceed with complete irregularity, the occurrence of specific mutations being assumed a matter of chance. But if we suppose some large number N of primordial species, each belonging to a different genus but similar in so far that we can take the chance of a specific mutation occurring as the same for all (*cf.* above, p. 23), to start together at zero time, we can then ask what is the *average* law of increase. The answer obtained is that, on an average, the number of species (that is the size of the genus) increases in geometric progression: if at zero time there is 1 species and at time 1 there are 2 species, at time 2 there will be 4 species, at time 3 there will be 8 species, at time 4 there will be 16 species and so on, the number of species, on an average, doubling in successive equal intervals of time. This law is important: for it suggests the "doubling-period for species within the genus" as a natural unit of time in investigations of the present kind. Specific mutations may occur much more frequently within one group than another, and the doubling period will then be much shorter in the first group than in the second; but if in each case we take the doubling-period as the unit of time matters are reduced to the same relative scale and will proceed *pari passu* in the two groups. In most of the following work, consequently, the doubling-period for species within the genus is taken as the unit of time and it must be remembered that in terms of years it varies from group to group.

But the above law is only an average. The further question therefore arises, what is the "frequency distribution," as the statistician terms it, of the sizes of these N genera, which all started as monotypic genera from primordial species at zero time, after any given time has elapsed? It is not enough to say that after one doubling-period the average size of a genus will be 2 species: we want to known how many of the genera have remained monotypic, how many of them have become ditypic, how many have become genera of 3, 4, 5 species, and so on. The answer obtained to this question is that the proportional frequencies of genera of 1, 2, 3 species may be expected to form a descending geometric series (*i.e.*, a series with a common ratio less than unity). After one doubling interval the frequencies should be proportional to $\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{8}$, $\frac{1}{16}$, and so on, the common ratio being $\frac{1}{2}$: after two doubling-periods the series should be $\frac{1}{4}$, $\frac{3}{16}$, $\frac{9}{64}$, and so on, the common ratio being $\frac{3}{4}$; if a is the first term of the series the common ratio must be $(1-a)$ to make the sum of the series unity. Clearly we are not in a position to compare this deduction with the facts, for we cannot pick out *any* group of genera, let alone a group of similar genera, which are all precisely of the same age.

We have, therefore, to proceed to the much more general problem. In the above, generic mutations are ignored, attention being fixed on the primordial genera alone. But during the long lapse of time generic mutations as well as specific mutations will be thrown, and any observed distribution will include both the derived and the primordial genera: what will be the form of the distribution of size of genus in the resulting aggregate? This question is considered in two successive sections of the paper. In Section II the limiting form is found when the time elapsed from the beginning of evolution is taken as infinite, and then in Section III the practical problem is solved for finite time. The mathematical results here are complicated: I would refer the non-mathematical reader in the first place to Table II, p. 43, where a series of calculated distributions is given. Evolution is assumed to start as before at zero time, and the form of the frequency distribution for size of genus is shown at times (expressed in terms of doubling-periods for species within the genus) 1, 2, 3, 4, 5, and 6·28 (the time found for a certain case below), the table concluding with the limiting form of distribution when an infinite time has elapsed.

To render the distributions comparable, the number of genera at each stage has been taken as 1,000; as the figures are rounded off to the nearest unit they are necessarily subject to abrupt variation in the tail of the distribution. Besides the time, it is necessary to fix one other quantity in order to determine the distribution, namely the ratio ρ of the chance of a specific mutation occurring in a given small interval of time to the chance of a generic mutation occurring in the same small interval of time: this has been taken at its value in an actual case (the Chrysomelid beetles, Tables IV, V, pp. 54, 56, and Appendix, Table A, p. 85), namely 1·925. Those who are familiar with the forms of distribution found in practice, or will refer to the typical tables given in the Appendix, will see that the distributions of Table II are at least of the right form so far as the eye can judge: monotypic genera are the most frequent, ditypic genera less frequent, tritypic genera less frequent still, and so on, the numbers gradually tailing away as the size of genus is increased. The length of this "tail" grows very rapidly with time, and the fully characteristic form (at least for such tables as those given in the Appendix) is not reached until two, three or more doubling-periods have elapsed. Thus after one doubling-period the percentage of monotypes is as high as 57, and only some 3 per 1,000 of all genera have 9 species or more. After two doubling-periods the percentage of monotypes has fallen to 42 and some 5 per 1,000 of all genera have 17 species or more. After three doubling-periods the percentage of monotypes has fallen further to 37, and there is a considerable proportion of the larger genera such as we seem more usually to find.

The mathematical form of the distribution may be better illustrated graphically. If we form a chart in which the number of genera of a given size is plotted vertically and the size of the genus horizontally, not to ordinary scales but to logarithmic or ratio scales, that is scales on which numbers that bear equal *ratios* to each other (like 1, 2, 4, 8, 16) stand at equal distances apart, the resulting points in any actual case run rather irregularly but fairly closely round a straight line, usually up to genera of 30 species or so, sometimes even up to genera of 100 species or more, after which the points fall rather abruptly away from the line. Three specimen charts so drawn are given in figs. 1, 2 and 3 (pp. 45, 46), to which the reader should refer, as well as to the adjacent explanatory text; two others will be found on pp. 241–2 of 'Age and Area.' If, now, similar charts are drawn for the calculated distributions of Table II they will be found to run as in figs. 4 to 9, pp. 47–49). It will be seen that at first, when time is very short—no more than one or two doubling periods—the graph drops away rapidly; but as time is increased the graph soon takes on the form noted for actual data, at first nearly straight and then falling away rather abruptly. For the longest time considered, 6·28 doubling-periods, the point at which the abrupt fall begins lies outside the chart on the right. So far as the graphic test goes, accordingly, the theory gives very well indeed precisely the form of distribution required.

In Section IV the test of agreement between theory and fact on this point is completed by fitting a calculated distribution to the actual distribution in four cases. To fit a given distribution, we require to determine from it the two constants that determine its *form*, the time τ (in doubling-periods) elapsed since the commencement, and the ratio ρ of the chance of a specific mutation to the chance of a generic mutation: we also require the initial number N_0 of primordial genera, which determines the total number of genera existing. The first two constants are found from the proportion of monotypes and the mean size of genus in the data: N_0 is then given by the number of species (or genera) in the data.* The reader will find the results in Tables V to VIII (pp. 56–58) and the values of τ, ρ, and N_0 in Table IV (p. 54). The numbers of monotypic genera observed and calculated must agree within a decimal point or so owing to the method of fitting: but I think the reader who studies Tables V to VIII will admit that the agreement between observation and calculation is throughout extraordinarily close. It is in fact better than one has any right to expect. I admit very considerable difficulties of interpretation and would refer to the discussion on pp. 58–62. Here I would only direct attention to the rather large number of primordial genera found in each case (Table IV, line 7, p. 54): to the comparatively limited range of values of τ (4·26 to 6·28, *ibid.* line 5): and to the comparatively limited range also of the values of ρ (1·188 to 1·925, *ibid.* line 6). Subject to the admitted difficulties of interpretation, the results of this test, on the one point on which direct comparison can be made with the facts, could hardly be better.

---

* If there are N genera in the data, and *one* primordial genus would only yield n genera after the given lapse of time, we must have $N_0 = N/n$.

The next point considered (Section V) is the frequency distribution of ages for genera of a given size: given, that is to say, N existing genera all of the same size, we require to know how many of them are likely to be of each successive age. The non-mathematical reader will do best to turn first to fig. 10, p. 66, showing the limiting forms of these distributions for genera of 1 to 10 species after an infinite time has elapsed from the beginning of evolution: ρ is taken as 1·5, and the unit of the scale of time is the doubling-period for species within the genus as usual. For monotypic genera the curve is a simple logarithmic curve, falling way steadily from age zero: the most probable age of a monotypic genus is zero, though the average age is, of course, something greater. For the ditypic genera the form of the distribution is quite different. The curve rises abruptly from zero to a maximum at a certain "modal" (most frequent) age, in this case 0·68 units as shown in Table IX, p. 67, and then falls away to a long tail. For the tritypic genera or genera of three species the form changes slightly again, the curve not rising from zero abruptly but starting tangentially to the base. The maximum frequency is reached at a mode of 1·14 units and then falls away again. For larger genera the form of the distribution remains the same, but the mode and mean are thrown further and further towards the right, increasing with the size of the genus. It will be noticed, however, that the increase in the mean, which is at first fairly rapid, gets slower and slower: the distribution for genera of 4 species is evidently pushed more to the right as compared with the distribution for genera of 3 species than is the distribution for genera of 10 species as compared with the distribution for genera of 9 species. The means are tabulated in column 2 of Table IX for genera of 1 to 10 species, and thence by intervals of 10 to genera of 100 species, and these bring out the point very clearly.

The mean age of monotypic genera, in the given case, is 0·87 of a unit, of genera of 10 species 3·22 units, giving an increase of 2·35 units of age for an increase of 9 units in the size of the genus. But a further increase of 10 units in the size of the genus only gives a mean age of 4·14 units, or an increment of 0·92 unit of age, and the next increment of 10 units in size of genus only gives an increment of 0·56 unit of age, and so on. For the larger genera of, say, 10 species and over, the mean age varies nearly as the logarithm of the size: *i.e.*, if we increase the number of species in the genus from s to s^a we must multiply the mean age by a (or a value very near it). Thus the mean age of genera of 100 species (6·40) is not 10 times but barely twice the mean age of genera of 10 species (3·22), since 100 is 10^2. The reader should note this approximate result and also how very largely the successive distributions of fig. 10 overlap each other. Apart from a further complication introduced when time is limited and not infinite, they suggest that, if we want to compare say areas occupied by genera of different sizes, (1) we should group the genera, not in groups of 1–10, 11–20, 21–30, etc., species, but in groups round 2, 4, 8, 16, 32, 64, 128, etc., species or some other series in geometric progression; (2) we must have as many genera as possible in each group in order to average out their varying ages; and even so (3) we must expect considerable fluctuations of sampling as the ages of genera even of one and the same size differ largely *inter se*. The chart, fig. 10, is drawn, as stated, for $\rho = 1\cdot5$. A larger value of ρ would give rather more scattered distributions with higher means and modes, but the general forms would remain the same. Conversely, a lower value of ρ would give more contracted distributions with lower means and modes.

When only a limited time has elapsed since the commencement of evolution, matters are not quite so simple. After an infinite lapse of time the number of derived genera—genera, which have arisen by generic mutations—becomes indefinitely great as compared with the number of the primordial genera, which are, so to speak, lost in the mass. The age distributions are, therefore, simply the limiting forms of the age distributions of derived genera. When time is limited, and as we have seen the four illustrations taken show times varying only from 4·26 to 6·28 doubling-periods, the genera really form two distinct groups: (*a*) the derived genera, for which the frequency distributions of ages are of the forms shown by fig. 10, but truncated at the assigned time, say τ, since the age of a derived genus obviously cannot exceed τ; (*b*) the primordial genera, all of age τ. The larger a genus, the greater is its probable age, and the greater therefore is the probability that it may be a primordial genus.

Table X and fig. 11 illustrate the case, using the constants found for the Chrysomelid beetles (Table V, p. 56). Turning first to Table X (p. 70), columns 4 to 7 give the essential figures: first, the mean age of the derived genera (column 4), then in columns 5 and 6 the proportions of derived and primordial genera respectively (or if we prefer to read it so, the chance that a genus is derived or primordial, terms which apply even if there is only one primordial genus), and finally in column 7 the mean age of the derived and primordial genera together, found from the figures of the three preceding columns. Thus for the monotypic genera, the most probable distribution is a proportion 0·9961 of derived genera of mean age 0·94, and a proportion 0·0039 of primordial genera all of age 6·28, giving a mean age for the entire aggregate

$$(0\cdot9961 \times 0\cdot94) + (0\cdot0039 \times 6\cdot28) = 0\cdot96.$$

The proportion of primordial genera, or the chance that a genus may be primordial, increases rapidly with the size of the genus. For the monotypic genera it is only 0·0039, for genera of 10 species 0·0842, for genera of 60 species 0·5045, for genera of 100 species 0·6549. It is roughly an even chance accordingly that a genus of 60 species is primordial; and the odds are nearly 2 to 1 (65 to 35) that a genus of 100 species is primordial.

Fig. 11, p. 71, shows the frequency distributions of age for genera of 1, 2, 3, 10, 60 and 100 species. The distributions for the derived genera are of the general forms shown in fig. 10, but rather more scattered, since ρ is larger, and truncated at the limiting age 6·28: the area is also reduced from unity to the value given by the figure in column 5 of Table X. The proportion of primordial genera is shown by a square to the right of each distribution, the area on the same scale being given by the figure of column 6, Table X, so that the area of the curve and the square taken together is equal to unity

Tables XI, XII and XIII, pp. 72, 74, give similar information for the three other examples and are illustrated similarly by figs. 12, 13 and 14, pp. 73, 75. The reader should refer to the adjacent text.

In the concluding section, VI, an attempt is made to estimate the order of magnitude of the doubling-period in the case of the flowering plants, and the present rate of occurrence of specific mutations. It will be remembered that in Section I it was found that the law of increase for species *within the genus* was geometric, the numbers tending to double in successive equal intervals of time, so that this "doubling-period" could be taken as the usual unit of our scale. In the entire *aggregate* of genera and species evolving from one or more primordial genera this law does not, on the given assumptions, hold good, but the divergence is only marked during the earlier stages of the evolution. The total number of species in the aggregate more than doubles during each successive doubling-period for species within the genus, but the ratio rapidly falls and tends to approach 2 more and more closely. If the reader will refer to the figures in the last column of Table XIV, p. 78, giving to the nearest unit the total number of species in such an aggregate at successive units of time when $\rho = 1 \cdot 5$, he will see that after some 5 or 6 doubling-periods the law approximates fairly closely to the geometric law of increase with a common ratio of 2. It was judged sufficient therefore to assume, for the rough estimation required in this section, that within the aggregate as within the genus the law might be taken as of this simple form.

On this assumption an estimate of the doubling-period is not difficult if we can fix: (1) the approximate age of the flowering plants, (2) the number of existing species. The latter Dr. WILLIS has placed at roundly 160,000; the former I have taken as roundly 100 million years. On this basis, if evolution has proceeded freely and without any destruction of species, the doubling-period works out at roundly some 6 million years, and the present rate of production of (viable) specific mutations at 1 in some 50 or 60 years, amongst all species of flowering plants on the whole surface of the globe (Table XV, section 1, p. 81).

Destruction of species will clearly tend to lower our estimate of the free doubling-period and raise our estimate of the present rate of production of specific mutations, for if species are destroyed from time to time by cataclysms or otherwise the rate of production during the undisturbed intervals must be increased in order to give the required final number of 160,000. The important question, therefore, arises how far variations in the time-incidence of destruction, a subject on which we have little knowledge, may also affect our estimate. On this point the interesting conclusion is reached that variations in the time-incidence have no effect, *so long as the chance of a species surviving from zero time to the time of observation is kept constant*, on the estimate of either the doubling-period or the present rate of production of specific mutations, but only on the estimated number of species killed-out. If we fix the number of extinct species but alter the time-incidence of destruction, assuming first, for example, that the killing-out has been effected almost continuously throughout the lapse of time and then that it has been effected by a series of highly destructive cataclysms occurring at intervals, we will arrive at different alternative estimates for the doubling-period and the rate of production of specific mutations at the present time. If, on the other hand, we fix the chance of a species surviving from zero time to the time of observation, an alteration in the time-incidence of destruction will vary the estimated number of extinct species alone.

In sections 2 to 5, Table XV, a series of comparisons is given on the first method. In each section of the table two cases are taken, in the first of which the destruction of species is cataclysmic, in the second continuous and of such severity as to give the same total of species killed-out; the severity of the destruction is raised from each section of the table to the next. I need hardly say that no importance is to be attached to the precise number of cataclysms taken and the (very large) proportion of the then-existing species killed-out by each: the scheme is simply so arranged as to illustrate the effect of very different time-incidences of destruction.

Of the four columns on the right of Table XV, the first gives the number of species killed-out. The second shows a chance that I have termed for brevity in the heading the chance of a species surviving for 10 million years: it is, in fact, the tenth-root of the chance of a species surviving for the whole period of 100 million years, which we have assumed is the time elapsed since the genesis of the flowering plants. If p is the chance of surviving 10 million years, p^2 is the chance of surviving 20 million years under the same conditions, p^3 the chance of surviving 30 million years, and so on: since p^{10} in several of the examples is vanishingly small, it is more convenient to give p. The third column gives the doubling-period, and the fourth the present rate of production of specific mutations. By increasing the destruction of species (section 5) to such an extent that over 700,000 must have been killed-out, which would require if there had been 100 equally destructive cataclysms that each should have killed-out no less than one-third of the then-existing species, the estimated length of the doubling-period has been lowered to a little over one million years, and the estimated rate of production of specific mutations raised to 1 in some 10 or 12 years: it makes no practical difference whether the killing-out has been continuous or cataclysmic. But there is little to guide us as to the whereabouts of the truth on the scale of results from section 2 to section 5 of the table, except one broad consideration. We know that the chance of a species surviving over the whole 100 million years must be practically zero—a vanishingly small quantity. We also know, from such cases as *Ginkgo*, that a species *may* survive for enormously long periods of geological time. It therefore seems probable that sections 2 and 3 underestimate the amount of extinction, for they give too high a value for the chances of survival; while section 5, it seems probable, overestimates the amount of extinction, giving too low a value.

The tentative conclusion is confirmed by sections 6 and 7 of the table, based on a table of Mrs. CLEMENT REID'S (reproduced on p. 83 below) showing the approximate percentages of species which have become extinct in certain deposits of various ages in the Pliocene: they suggest a figure somewhere between 1·7 and 3·2 million years for the doubling-period, and between 1 in 15 years and 1 in 29 years for the present rate of production of (viable) specific mutations. The reader should refer to the more detailed discussion at the end of Section VI of the paper.

Considering the roughness of the basis the final conclusion is extraordinarily definite, and leaves no doubt as to the order of magnitude of the required figures. To quote from p. 84, "If the age of the flowering plants is 100 million years, or thereabouts, the doubling-period for species is probably of the order of some 2 or 3 million years: it is, say, almost certainly over 1 million and less than 6 millions. The present rate of production of (viable) specific mutations, amongst all flowering plants on the whole surface of the globe, is almost certainly less than 1 in 10 years and more than one in 60 years: it probaby lies between 1 in 15 and 1 in 30 years." The results are, of course, of the nature of averages, since the flowering plants are a very heterogeneous group.

It is clear that specific mutations must be events of the very greatest rarity; and no argument, as it seems to me, can be based on the fact that we have no knowledge or experience of such phenomena. My work on the point entirely confirms the conclusion of Dr. WILLIS.*

If any conclusions stated in this Introduction or in the body of the paper seem to be too confident, or at all dogmatic, I hope the reader will attribute the appearance to inadvertence of wording, or a simple desire to avoid the wearisome reiteration of qualifying phrases. I have no desire at all to be dogmatic: but Dr. WILLIS's conclusions do appear to me to explain a great range of facts, to be natural and reasonable in themselves, and to present for the first time a conception of evolution at once so simple and so definite that it can be expressed—even if only roughly owing to the gaps in our knowledge—in quantitative terms, and quantitative deductions drawn. Admitting all the difficulties of interpreting some results (Section IV)—and as I have said, they are considerable—these may yield to further work, to the co-operation of biologists with more competent mathematicians, or more likely still to the mathematically trained biologist. I may be optimistic, but it seems to me that the future holds the possibility of great developments.

To Dr. WILLIS himself my acknowledgments are difficult adequately to express. The paper is founded on his work, but would hardly have been carried through without the encouragement and stimulus of personal intercourse. I have, in particular, to thank him for placing at my disposal the data given in Tables A to E of the Appendix and utilised as illustrations.

---

* 'Age and Area,' p. 212.

1. The Increase in the Number of Species within the Genus with Time, and the Frequency Distribution of Sizes of Genera all of the same Age.

Let the chance of a species "throwing" a specific mutation, *i.e.*, a new species of the same genus, in some small assigned interval of time be p, and suppose the interval so small that p^2 may be ignored compared with p. Then, putting aside generic mutations altogether for the present, if we start with N prime species of different genera, at the end of the interval we will have $\mathrm{N}q$ genera ($q = 1-p$) which remain monotypic and $\mathrm{N}p$ genera of two species.

The new species as well as the old can now throw specific mutations and the matter becomes more complex. Of the $\mathrm{N}q$ monotypics a proportion q will again fail to throw in the second interval and we will have, at the end of the interval, $\mathrm{N}q^2$ monotypics.

Of the same $\mathrm{N}q$ a proportion p will throw, contributing $\mathrm{N}pq$ to the genera of 2 species. Of the $\mathrm{N}p$ genera already possessing 2 species at the end of the first interval, a proportion q^2 will fail to throw from either of the species they already possess, giving $\mathrm{N}pq^2$ genera of 2 species at the end of the second interval, or a total of $\mathrm{N}pq\,(1+q)$ ditypic genera (*cf.* the scheme below).

Of the same $\mathrm{N}p$ genera that already possessed 2 species at the end of the first interval, the proportion in which one *or* other will throw a new species in the second interval is $2pq$, the second term in the expansion of $(q+p)^2$, giving $\mathrm{N}.\ 2p^2q$ genera of 3 species.

The proportion in which *both* of the two will throw a new species is p^2 giving $\mathrm{N}p^3$ genera of 4 species.

Proportional frequencies of numbers of genera with 1, 2, 3 species after 0, 1, 2, 3 intervals of time, the genera being initially monotypic.

Number of species in the genus.	Intervals of Time.			
	0	1	2	3
1	1	q	q^2	q^3
2	—	p	$pq\,(1+q)$	$pq^2\,(1+q+q^2)$
3	—	—	$2p^2q$	$2p^2q^2\,(1+q+q^2)$
4	—	—	p^3	$p^3q\,(1+q+6q^2+q^3)$
5	—	—	—	$2p^4q^2\,(3+2q)$
6	—	—	—	$2p^5q\,(1+3q)$
7	—	—	—	$4p^6q$
8	—	—	—	p^7

The work may be continued on the same simple lines. Thus at the end of the third interval, omitting the common factor N, we will have genera—

$$\begin{array}{rl}
\text{With 1 species :} & q^2 \times q \quad . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; q^3 \\
\text{With 2 species :} & \\
\text{From the 1's :} & \left.\begin{array}{l} q^2 \times p \\ pq\,(1+q) \times q^2 \end{array}\right\} \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; pq^2\,(1+q+q^2) \\
\text{From the 2's :} & \\
\text{With 3 species :} & \\
\text{From the 2's :} & \left.\begin{array}{l} pq\,(1+q) \times 2pq \\ 2p^2q \times q^3 \end{array}\right\} \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; 2p^2q^2\,(1+q+q^2) \\
\text{From the 3's :} & \\
\text{With 4 species :} & \\
\text{From the 2's :} & \\
\text{From the 3's :} & \left.\begin{array}{l} pq\,(1+q) \times p^2 \\ 2p^2q \times 3pq^2 \\ p^3 \times q^4 \end{array}\right\} \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; p^3q\,(1+q+6q^2+q^3) \\
\text{From the 4's :} & \\
\text{With 5 species :} & \\
\text{From the 3's :} & \left.\begin{array}{l} 2p^2q \times 3p^2q \\ p^3 \times 4pq^3 \end{array}\right\} \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; 2p^4q^2\,(3+2q) \\
\text{From the 4's :} & \\
\text{With 6 species :} & \\
\text{From the 3's :} & \left.\begin{array}{l} 2p^2q \times p^3 \\ p^3 \times 6p^2q^2 \end{array}\right\} \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; 2p^5q\,(1+3q) \\
\text{From the 4's :} & \\
\text{With 7 species :} & \\
\text{From the 4's :} & p^3 \times 4p^3q \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; 4p^6q \\
\text{With 8 species :} & \\
\text{From the 4's :} & p^3 \times p^4 \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; p^7
\end{array}$$

The formulæ, summarised in the scheme above, may be verified by checking that they total to unity in each column.

We must now proceed to the limit, taking the time-interval Δt as indefinitely small but the number of such intervals n as large, so that the time $n \,.\, \Delta t = t$ is finite. We may write

$$p = s \,.\, \Delta t \qquad\qquad pn = st$$

and we have the usual approximation

$$q^n = (1-p)^n = (1-st/n)^n \sim e^{-st}.$$

Omitting the common factor N, the first term is q^n or in the limit e^{-st}: that is, if f_1 is the proportional frequency of monotypic genera at time t,

$$f_1 = e^{-st} \quad . \; . \qquad (1)$$

As regards the second term, after n intervals we have

$$pq^{n-1}\,(1+q+q^2+ \;.\;.\;. + q^{n-1}) = q^{n-1}\,(1-q^n)$$

or in the limit, f_2 being the proportional frequency of genera with 2 species at time t,

$$f_2 = e^{-st}\,(1-e^{-st}). \qquad \ldots\ldots\ldots\ldots \tag{2}$$

But this expression may be derived by a process which is more convenient for the higher terms. If ${}_nf_2$, ${}_{n+1}f_2$ are the proportional frequencies of ditypic genera after n and $n+1$ intervals respectively, and so on,

$${}_{n+1}f_2 = p_nf_1 + q^2{}_nf_2.$$

That is

$${}_{n+1}f_2 - {}_nf_2 = p_nf_1 + (q^2-1)_nf_2,$$

or in the limit, putting $p = s\,.\,dt$, ${}_nf_1 = e^{-st}$, $1 - q^2 = 2p = 2s\,.\,dt$

$$\frac{df_2}{dt} + 2s\,.f_2 = s\,.\,e^{-st},$$

$$\frac{d}{dt}\,(e^{2st}f_2) = s\,.\,e^{st}$$

$$f_2 = e^{-st} + Ce^{-2st}.$$

But when t is zero f_2 must be zero, and therefore C must be -1, so that finally

$$f_2 = e^{-st}\,(1-e^{-st})$$

as before.

Considering f_3 in the same way, we have

$${}_{n+1}f_3 = 2pq_nf_2 + q^3{}_nf_3.$$

Or proceeding to the limit, when $1-q^3 = 3p$ and $pq = p$

$$\frac{df_3}{dt} + 3sf_3 = 2s\,(e^{-st} - e^{-2st}),$$

$$\frac{d}{dt}\,(e^{3st}f_3) = 2s\,(e^{2st} - e^{st}),$$

$$f_3 = e^{-st} - 2e^{-2st} + C\,e_3{}^{-3st}.$$

Since f_3 must be zero when t is zero, C must be $+1$, and therefore finally

$$f_3 = e^{-st}\,(1-e^{-st})^2 \qquad \ldots\ldots\ldots\ldots \tag{3}$$

Evaluating f_4 in the same way we find

$$f_4 = e^{-st}\,(1-e^{-st})^3 \qquad \ldots\ldots\ldots\ldots \tag{4}$$

The general form of the law is now obvious. We have

$$\left.\begin{aligned} f_1 &= e^{-st} \\ f_2 &= e^{-st}\,(1-e^{-st}) \\ f_3 &= e^{-st}\,(1-e^{-st})^2 \\ &\ldots\ldots\ldots \\ f_n &= e^{-st}\,(1-e^{-st})^{n-1} \end{aligned}\right\} \qquad \ldots\ldots\ldots\ldots \tag{5}$$

That is to say, if N prime genera start together at zero time when they are all monotypic, after time t we will find the numbers that have 1, 2, 3, . . . species given by a geometric series of which the common ratio is $1-e^{-st}$, s being a constant proportional to the chance of a specific mutation occurring in a given time.

The sum of the series (5) to infinity is unity as it must be. As regards the mean, consider the general geometric series

$$S = 1 + r + r^2 + r^3 + \ldots$$

where S is the sum to infinity. We have

$$dS/dr = (1-r)^{-2} = 1 + 2r + 3r^2 + \ldots$$

which is the first moment about zero, so that the mean is $1/(1-r)$. Therefore the mean of (5), say ${}_tM_s$ is given by

$${}_tM_s = e^{st} \quad \ldots\ldots\ldots\ldots \quad (6)$$

That is to say, while the number of species within any single genus increases, of course, quite irregularly as chance mutations occur, the mean number within a group of genera all starting at the same time increases as an exponential function of the time.

As regards the increase in the number of *genera*, the whole process will proceed on precisely the same lines. If N genera, belonging to distinct families, start simultaneously at zero time, at time t the numbers of families with 1, 2, 3, . . genera will be given by N multiplied by the successive terms of the series

$$\left.\begin{aligned} f_1 &= e^{-gt} \\ f_2 &= e^{-gt}(1-e^{-gt}) \\ f_3 &= e^{-gt}(1-e^{-gt})^2 \\ &\ldots\ldots\ldots \\ f_n &= e^{-gt}(1-e^{-gt})^{n-1} \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots \quad (7)$$

where g is a constant proportional to the chance of a generic mutation (or mutation from genus to genus) occurring in a given time. The mean number of genera in a family at time t in these circumstances will be given by

$${}_tM_g = e^{gt} \quad \ldots\ldots\ldots\ldots \quad (8)$$

the mean increasing as an exponential function of the time.

The series (5) gives the frequency distribution of sizes of genera for genera all of the same age t. We have next to proceed to the question, what is the form of the frequency distribution for genera of all ages? If we start at zero time with a group of monotypic genera, both generic and specific mutations will be thrown as time goes on. At time t therefore we will actually observe, not merely the frequency distribution of the sizes of the primordial genera as given by (5), but the frequency distribution of the primordial and the derived genera together, and it is this frequency distribution which is now required. We will first consider the limiting form of the distribution when the time t is infinite, and the primordial genera become practically negligible in numbers as compared with the derived genera.

II. THE LIMITING FORM OF THE FREQUENCY DISTRIBUTION FOR THE SIZES OF DERIVED GENERA WHEN TIME IS INFINITE.

We first require to know how many out of the totality of genera existing at any given time, say T, are of any assigned age x. From (8) the total number of genera at time t is $\mathrm{N}e^{gt}$. The number coming into existence during the interval $\pm\frac{1}{2}dt$ round time t is therefore $\mathrm{N}ge^{gt}\,dt$, and the number of age x at time T is

$$\mathrm{N}g\,e^{g(\mathrm{T}-x)}\,dx.$$

The *proportion* aged x at time T is therefore $ge^{-gx}\,dx$. Note that these are the *derived* genera only, ignoring the prime genera with which we started; but, as stated at the end of Section I, since we are going to take time as infinite the number of derived genera will be infinitely great as compared with the number of primordial genera, and the latter may legitimately be ignored during the present stage of the work.

We have now got to take the series (5), writing x for t throughout, term by term, multiply each term by $ge^{-gx}\,dx$ and integrate from zero to infinity. The first term will give

$$g\int_0^\infty e^{-(g+s)x}\,dx = g\,(g+s)^{-1} = (1+\rho)^{-1} \quad \ldots\ldots\ldots \quad (9)$$

where

$$\rho = s/g. \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (10)$$

As we can only suppose that specific mutations are more frequent than generic mutations ρ must be greater than unity.

The second term will give

$$\begin{aligned} & g\int_0^\infty \left\{e^{-(g+s)x} - e^{-(g+2s)x}\right\} dx \\ &= g\,[(g+s)^{-1} - (g+2s)^{-1}] \\ &= \rho\,(1+\rho)^{-1}\,(1+2\rho)^{-1} \quad \ldots\ldots\ldots\ldots \quad (11) \end{aligned}$$

Proceeding similarly for the further terms we find for the limiting form of the frequency distribution of sizes of genera of all ages after infinite time the series

$$\left.\begin{aligned} f_1 &= \frac{1}{1+\rho} \\ f_2 &= \frac{1}{1+\rho}\,\frac{\rho}{1+2\rho} \\ f_3 &= \frac{1}{1+\rho}\,\frac{\rho}{1+2\rho}\,\frac{2\rho}{1+3\rho} \\ &\ldots\ldots\ldots\ldots \\ f_n &= \frac{1}{1+\rho}\,\frac{\rho}{1+2\rho}\,\frac{2\rho}{1+3\rho}\,\cdots\,\frac{(n-1)\rho}{1+n\rho} \end{aligned}\right\} \quad \ldots\ldots\ldots \quad (12)$$

The series has been written in the form most convenient for calculation: the fractions $1/(1+\rho)$, $\rho/(1+2\rho)$, etc., are first run out on the calculating machine and $f_1, f_2, f_3 \ldots$ are then obtained by successive multiplication.

The series can be summed to n terms. We have

$$f_n = \frac{(n-1)\rho}{1+n\rho} f_{n-1},$$

or

$$\begin{aligned}
(1+\rho)f_n &= (n-1)\,\rho\,(f_{n-1}-f_n) \\
(1+\rho)f_{n-1} &= (n-2)\,\rho\,(f_{n-2}-f_{n-1}) \\
\vdots\quad & \qquad\quad \vdots \\
(1+\rho)f_2 &= \qquad\quad \rho\,(f_1-f_2) \\
(1+\rho)f_1 &= 1.
\end{aligned}$$

Hence if S_n is the sum of the first n terms

$$\begin{aligned}
(1+\rho)\,S_n &= 1+\rho\,(f_1+f_2+\ldots+f_{n-1})-(n-1)\,\rho f_n \\
&= 1+\rho\,S_n-\rho n f_n,
\end{aligned}$$

or

$$S_n = 1-\rho n f_n. \quad \ldots\ldots\ldots\ldots\ldots \quad (13)$$

This result (which with much of the following I owe to the kindness of Mr. F. P. White, Fellow of St. John's College) is exceedingly useful for checking calculations: it also follows from it at once that S_∞ is unity, as it should be.

The series may be written in the form

$$\begin{aligned}
&\frac{1}{1+\rho}\left\{1+\frac{\rho}{1+2\rho}+\frac{1\,.\,2\rho^2}{(1+2\rho)\,(1+3\rho)}+\ldots\right\} \\
&=\frac{\gamma-1}{\gamma-2}\left\{1+\frac{1}{\gamma}+\frac{1\,.\,2}{\gamma\,(\gamma+1)}+\ldots\right\}
\end{aligned}$$

where $\gamma = 2+\rho^{-1}$. Apart from the common factor, the series is therefore a hypergeometrical series in which

$$\begin{aligned}
\alpha &= \beta = x = 1 \\
\gamma &= 2+\rho^{-1}.
\end{aligned}$$

Consider the series

$$1+2\frac{1}{\gamma}+3\frac{1\,.\,2}{\gamma\,(\gamma+1)}+\ldots+n\frac{1\,.\,2\,.\,3\ldots(n-1)}{\gamma\,(\gamma+1)\ldots(\gamma+n-2)} \quad \ldots\ldots \quad (14)$$

giving the first moment of (12) about zero. In this series

$$\frac{u_{n-1}}{u_n} = 1+\frac{\gamma-2}{n}.$$

As $\gamma > 2$ the ratio exceeds unity, only approaching it as n increases. But

$$\mathop{\underline{\big|\quad}}\left\{n\left(\frac{u_{n-1}}{u_n}-1\right)\right\} = \gamma-2$$

which is less than unity, so the series (14) is divergent. The frequency distribution given by the terms of (12) is therefore one of those paradoxical distributions in which, though the median, etc., are finite, the mean is infinite. This is, of course, as it should be, for on our assumptions the mean size of a genus after infinite time must itself be infinite.

Since the limiting distribution (12) is the foundation of all subsequent work, even when time is limited and not infinite, it may be as well to examine its form in some detail. It evidently gives, as required by all the data for size of genus that have yet been tabulated, a distribution with a maximum frequency for monotypic genera, the frequency tailing off at first rapidly and then more slowly as the successive multipliers $(n-1)\rho/(1+n\rho)$ approach unity. The greater ρ, the smaller is the frequency of monotypes and the more slowly do the frequencies fall away. Table I shows the values of the successive terms (giving only every tenth term after f_{20}) for $\rho = 1 \cdot 925$, the value suggested by the Chrysomelid beetles (*cf.* below, p. 54).

TABLE I.—Values of $10^7 f_n$ in the ρ-series (12) and of $-\Delta \log f_n / \Delta \log n$, for $\rho = 1 \cdot 925$.

r.	$10^7 f_n$.	$\frac{-\Delta \log f_n}{\Delta \log n}$.	r.	$10^7 f_n$.	$\frac{-\Delta \log f_n}{\Delta \log n}$.
1	341 8803	1·3331	15	7 3329	1·4951
2	135 6948	1·3939	16	6 6584	1·4965
3	77 1107	1·4244	17	6 0810	1·4978
4	51 1856	1·4430	18	5 5820	1·4989
5	37 0945	1·4554	19	5 1475	1·4999
6	28 4489	1·4644	20	4 7663	1·5008
7	22 7002	1·4712	30	2 5906	1·5068
8	18 6515	1·4764	40	1 6786	1·5099
9	15 6744	1·4807	50	1 1982	1·5118
10	13 4103	1·4841	60	9095	1·5131
11	11 6414	1·4870	70	7202	1·5139
12	10 2285	1·4895	80	5883	1·5146
13	9 0789	1·4916	90	4922	1·5152
14	8 1288	1·4935	100	4196	1·5156

Consider the approximate form of the tail of the distribution. We have

$$f_n = \frac{(n-1)!\,\rho^{n-1}}{(1+\rho)(1+2\rho)\ldots(1+n\rho)}$$

$$= \frac{(n-1)!}{\rho\,(n+\rho^{-1})(n-1+\rho^{-1})\ldots(2+\rho^{-1})(1+\rho^{-1})}$$

$$= \frac{\Gamma(1+\rho^{-1})}{\rho}\;\frac{\Gamma(n)}{\Gamma(n+1+\rho^{-1})}.$$

If n be large, we may write by Stirling's theorem

$$\log\Gamma(n) \sim (n-\tfrac{1}{2})\log n - n + \tfrac{1}{2}\log 2\pi$$

$$\log\Gamma(n+a) \sim (n+a-\tfrac{1}{2})\log n\,(1+a/n) - n\,(1+a/n) + \tfrac{1}{2}\log 2\pi$$

$$\sim (n+a-\tfrac{1}{2})\log n - n + \tfrac{1}{2}\log 2\pi$$

where a is assumed small compared with n: whence in the limit

$$f \sim \frac{\Gamma(1+\rho^{-1})}{\rho}\,n^{-(1+\rho^{-1})} \quad \ldots\ldots\ldots\ldots \quad (15)$$

This formula holds fairly even for moderate values of n, the approximation rapidly becoming exceedingly close, as shown by the following comparison with the true values of f_n as given by (12) in Table I:

	Value of f_n as given by	
n	(12)	(15)
10	0·013 410	0·013 932
20	0·004 766	0·004 860
50	0·001 198	0·001 208
100	0·000 4196	0·000 4212

It follows that, if we plot $\log f_n$ to $\log n$, the resulting points lie nearly on a straight line: how nearly is shown by the dotted line in figs. 4–9, in which $\log f_n$ is plotted to $\log n$ for the data of Table I. The dotted line is actually concave towards the base, but the curvature is so small that it is hardly appreciable to the eye, and for no statistical data which followed the law would such a divergence from linearity appear significant. The values of $-\Delta \log f_n / \Delta \log n$ given in Table I bring out better than the charts the gradual increase in the slope of the curve, and the approach of the slope to the limiting value $1 + \rho^{-1}$ or 1·5195.

When work on the frequency distributions of sizes of genera was first begun, considerations of a very rough kind suggested that the limiting form of the distribution for infinite time should approach this logarithmic-linear law. The generation of species from species, or genera from genera, seemed closely parallel to the generation of offspring in a given stock in which mortality might be ignored: hence it seemed reasonable to assume that the proportion y of genera of age t at any given epoch (apart from the primordial genera which might be ignored if time were infinite) would follow the law

$$y \propto e^{-gt}\, dt.$$

Similarly it seemed reasonable to assume that the number of species x in a genus of age t would be given by

$$x \propto e^{st}.$$

Now suppose that chance can be ignored, that the number of species in a genus can be taken as a continuous variable, and that the above can be taken as absolute functional relations. The size of a genus is then absolutely determined by its age, and we can find the number of genera of each size by eliminating t from the first relation. We have

$$t = \log x^{1/s}, \qquad dt = sx^{-1}dx$$

whence

$$y \propto x^{-(g+s)/s}\, dx$$
$$\propto x^{-(1+\rho^{-1})}\, dx.$$

The method of approach was obviously exceedingly crude, but it suggested logarithmic plotting of the data. Further trial showed that the law did in fact hold very approximately, both for animals and plants, up to genera of 30 species or so, after which the points began to drop away more or less rapidly from the line, *i.e.*, the observed frequencies of the larger genera were smaller than those suggested by the law : figs. 2 and 3 (below, pp. 45, 46) bring out the point very well. The approximation was, however, sufficiently close to encourage further work. As will be seen from the following, the fundamental mistake was the notion that time might be regarded for practical purposes as infinite : it must in fact be regarded as quite short.

III. THE FREQUENCY DISTRIBUTION FOR SIZES OF GENERA (DERIVED AND PRIMORDIAL) WHEN TIME IS FINITE.

If time is finite we can no longer, as at the beginning of Section II, ignore the number of prime genera as compared with the number of derived genera. At time T, out of unit total, we will have $ge^{-gx}\,dx$ derived genera of age x, together with e^{-gT} of age T—the term which vanishes if T is infinite.

To obtain the required frequency distribution there will, therefore, be two corrections to make to the limiting distribution (12). (a) We must *add* to each term of (12) the corresponding term of (5) (writing T for t) multiplied by e^{-gT}. (b) We must subtract from each term of (12) the value given by making the upper limit of the integral T instead of infinity. The additive corrections to $f_1, f_2, f_3, \ldots$ are therefore

$$\begin{aligned}
&e^{-(g+s)T}\\
&e^{-(g+s)T}(1-e^{-sT})\\
&e^{-(g+s)T}(1-e^{-sT})^2\\
&\quad\vdots \qquad \vdots
\end{aligned}$$

and the subtractive corrections

$$\begin{aligned}
&g\,(g+s)^{-1}\,e^{-(g+s)T}\\
&g\,(g+s)^{-1}\,e^{-(g+s)T}-g\,(g+2s)^{-1}\,e^{-(g+2s)T}\\
&g\,(g+s)^{-1}\,e^{-(g+s)T}-2g\,(g+2s)^{-1}\,e^{-(g+2s)T}+g\,(g+3s)^{-1}\,e^{-(g+3s)T}\\
&\quad\vdots \qquad\qquad\qquad \vdots \qquad\qquad\qquad \vdots
\end{aligned}$$

Bringing the two together, the entire corrections are :—

$$\left.\begin{aligned}
c_1 &= s\,(g+s)^{-1}\,e^{-(g+s)T}\\
c_2 &= s\,(g+s)^{-1}\,e^{-(g+s)T}-2s\,(g+2s)^{-1}\,e^{-(g+2s)T}\\
c_3 &= s\,(g+s)^{-1}\,e^{-(g+s)T}-4s\,(g+2s)^{-1}\,e^{-(g+2s)T}+3s\,(g+3s)^{-1}\,e^{-(g+3s)T}\\
&\quad\vdots \qquad\qquad\qquad \vdots \qquad\qquad\qquad \vdots
\end{aligned}\right\}\quad\ldots\quad(16)$$

It will be noticed that these may be written in the form

$$\left.\begin{aligned} c_1 &= c'_1 \\ c_2 &= c'_1 - c'_2 \\ c_3 &= c'_1 - 2c'_2 + c'_3 \\ c_4 &= c'_1 - 3c'_2 + 3c'_3 - c'_4 \end{aligned}\right\}, \quad \ldots\ldots\ldots\ldots \quad (17)$$

and so on, where one new term c' is introduced at each stage, and the numerical coefficients are the coefficients of $(1 - 1)^{n-1}$.

For numerical purposes the terms c'_1, c'_2, c'_3, etc., must be put in a different form. As before, we write $\rho = s/g$. As the unit of time we will take the average time λ in which the number of species within the genus doubles, writing

$$e^{g\mathrm{T}} = e^{\alpha\mathrm{T}/\lambda} = e^{\alpha\tau},$$

where

$$\alpha = \log_e 2 = 0 \cdot 6931472. \quad \ldots\ldots\ldots\ldots \quad (18)$$

Then

$$g = \alpha/\rho$$

$$g + ns = \alpha\,(1 + n\rho)/\rho,$$

and c'_1, c'_2, c'_3, . . . may be written, τ being the time in the new units

$$\left.\begin{aligned} c'_1 &= \rho\,(1 + \rho)^{-1}\, e^{-\alpha\rho^{-1}(1+\rho)\tau} \\ c'_2 &= 2\rho\,(1 + 2\rho)^{-1}\, e^{-\alpha\rho^{-1}(1+2\rho)\tau} \\ c'_3 &= 3\rho\,(1 + 3\rho)^{-1}\, e^{-\alpha\rho^{-1}(1+3\rho)\tau} \\ &\vdots \qquad \vdots \qquad \vdots \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots \quad (19)$$

Given ρ and τ, the values of c'_1, c'_2, c'_3, . . . are calculated from (19), thence the values of c_1, c_2, c_3, . . . from (17): the values of the limiting frequencies ${}_\infty f_1$, ${}_\infty f_2$, ${}_\infty f_3$, ... are found from (12), and finally the frequencies at time τ are given by

$$\left.\begin{aligned} {}_\tau f_1 &= {}_\infty f_1 + c_1 \\ {}_\tau f_2 &= {}_\infty f_2 + c_2 \\ {}_\tau f_3 &= {}_\infty f_3 + c_3 \\ \vdots \quad &\quad \vdots \qquad \vdots \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots \quad (20)$$

Since sufficiently extensive tables of the binomial coefficients were not, so far as I could find, available, tables were calculated giving all the coefficients of $(1 + 1)^n$ up to $n = 31$, and the coefficients up to the twelfth for values of n from 32 to 101. The correcting series on the right of (17) converge with fair rapidity in the illustrative examples that have been tried (*cf.* Section IV), and twelve terms usually give more than sufficient precision. The corrections c_1, c_2, c_3, . . . are at first positive, but decrease steadily in value and sooner or later become negative: such a change of sign must, of course, occur since $\Sigma_1^{\infty}(c_r)$ must be zero. After reaching a negative maximum the corrections again diminish until c_r approximates to $-{}_{\infty}f_r$.

To illustrate the change in the form of the distribution with time, Table II has been calculated, showing the successive distributions when $\tau = 1, 2, 3, 4, 5,$ and $6 \cdot 28$ units (doubling-periods for species within the genus) and ρ is $1 \cdot 925$ ($\rho = 1 \cdot 925$, $\tau = 6 \cdot 28$, being the values found for the *Chrysomelidæ,* below, p. 54); the limiting distribution for $\tau = \infty$ is also given. The table shows in a very interesting way the rapid growth in the tail of the distribution towards large genera, and the quick approximation of the first part of the distribution towards the limiting form for infinite time. These features speak for themselves. It may be added that the correcting terms, which are necessarily positive at the beginning of the series, when $\tau = 1$ become negative with c_4, c_7 reaching the negative maximum. When $\tau = 2$, c_9 is the first correction to become negative and the negative maximum is reached with c_{15}. When $\tau = 3$, c_{18} is the first to become negative and the negative maximum lies (not far) beyond c_{31}. When $\tau = 4$, c_{37} is the first to become negative, and when $\tau = 5$ the first correction to become negative lies somewhere between c_{60} and c_{80}. When $\tau = 6 \cdot 28$, as for the *Chrysomelidæ,* the correcting terms are still positive at the limit to which calculation was carried, viz., f_{102}.

TABLE II.—Showing, at successive epochs, the calculated numbers of genera with 1, 2, 3, . . . species out of 1,000 genera in all at each stage when $\rho = 1 \cdot 925$, the value found for the Chrysomelid beetles (Table V, p. 56, and Appendix, Table A). The table shows the distribution after 1, 2, 3, 4 and 5 doubling-periods for species within the genus; then for 6·28 doubling-periods, the time found for the *Chrysomelidæ*; and finally the limiting form of the distribution after an infinite lapse of time.

Number of species in the genus.	Time τ in doubling-periods.							
	1	2	3	4	5	6·28	∞	
1	571	422	370	352	345	343	342	1
2	227	192	159	145	139	137	136	2
3	104	115	97	85	80	78	77	3
4	50	77	68	59	54	52	51	4
5	24	53	51	44	40	38	37	5
6	12	38	40	35	31	29	28	6
7	6	27	33	29	26	24	23	7
8	3	20	27	24	21	19	19	8
9	1	15	22	21	18	16	16	9
10	—	11	19	18	16	14	13	10
11	—	8	16	16	14	12	12	11
12	—	6	13	14	13	11	10	12
13	—	4	12	13	11	10	9	13
14	—	3	10	11	10	9	8	14

(*Table continued overleaf.*)

TABLE II—(continued).

Number of species in the genus.	Time τ in doubling-periods.							
	1	2	3	4	5	6·28	∞	
15	—	2	8	10	9	8	7	15
16	—	2	7	9	9	7	7	16
17	—	1	6	9	8	7	6	17
18	—	1	5	8	7	6	6	18
19	—	—	5	7	7	6	5	19
20	—	—	4	7	6	5	5	20
21	—	—	4	6	6	5	4	21
22	—	—	3	6	6	5	4	22
23	—	—	3	5	5	4	4	23
24	—	—	2	5	5	4	4	24
25	—	—	2	4	5	4	3	25
26	—	—	2	4	4	4	3	26
27	—	—	2	4	4	4	3	27
28	—	—	1	3	4	3	3	28
29	—	—	1	3	4	3	3	29
30	—	—	1	3	4	3	3	30
31	—	—	1	3	3	3	2	31
32	—	—	—	3	3	3	2	32
33	—	—	—	2	3	3	2	33
34	—	—	—	2	3	3	2	34
35	—	—	—	2	3	3	2	35
36	—	—	—	2	3	2	2	36
37	—	—	—	2	3	2	2	37
38	—	—	—	2	2	2	2	38
39	—	—	—	2	2	2	2	39
40	—	—	—	2	2	2	2	40
41	—	—	—	1	2	2	2	41
42	—	—	—	1	2	2	2	42
43	—	—	—	1	2	2	2	43
44	—	—	—	1	2	2	1	44
45	—	—	—	1	2	2	1	45
46	—	—	—	1	2	2	1	46
47	—	—	—	1	2	2	1	47
48	—	—	—	1	2	2	1	48
49	—	—	—	—	2	2	1	49
50	—	—	—	—	1	2	1	50
Over the last frequency given above...	2	3	6	11	43	85	116	—
Total ...	1,000	1,000	1,000	1,000	1,000	1,000	1,000	—

The gradual approximation of the double logarithmic graph of the frequency distribution towards its limiting, nearly linear, form is very interesting. Before the present theory had been developed the form of the graphs observed seemed rather puzzling. The initial part of the graph, say up to genera of 30 species or more, appeared extraordinarily nearly linear; but after a certain point there was always a more or less rapid falling away of the data from the line. Figs. 1, 2 and 3 are given to illustrate the point,

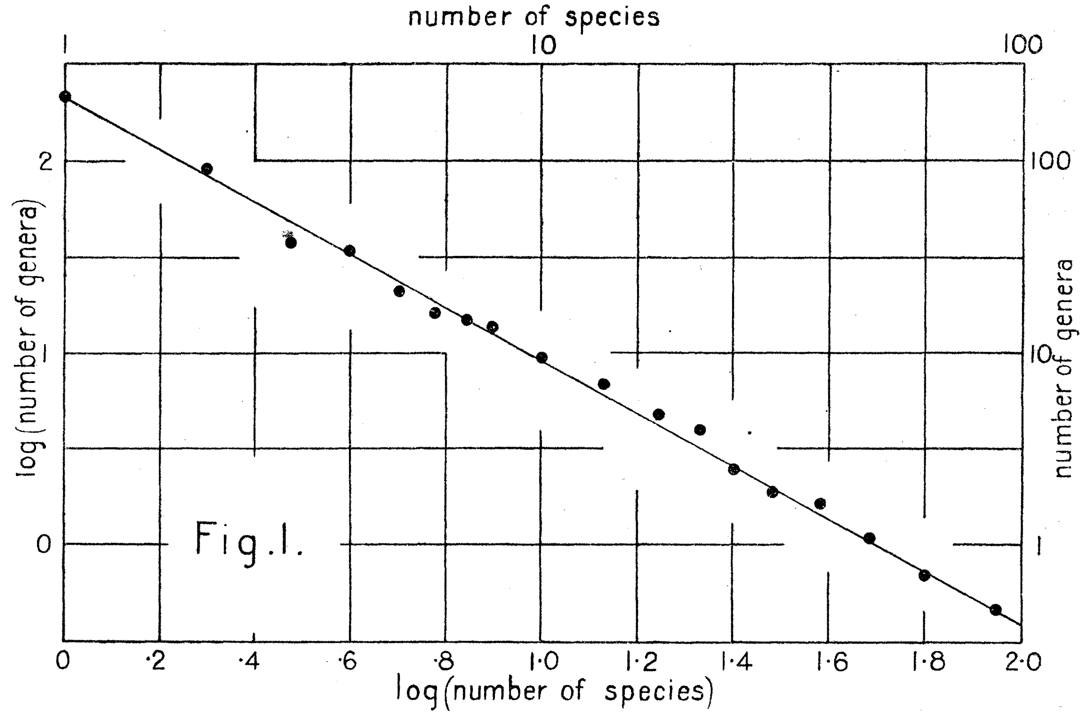

Fig. 1.—Double logarithmic chart for the frequency distribution of sizes of genera in the *Chrysomelidæ*: logarithm of the number of genera plotted on the vertical to logarithm of the number of species on the horizontal. Data in Appendix, Table A.

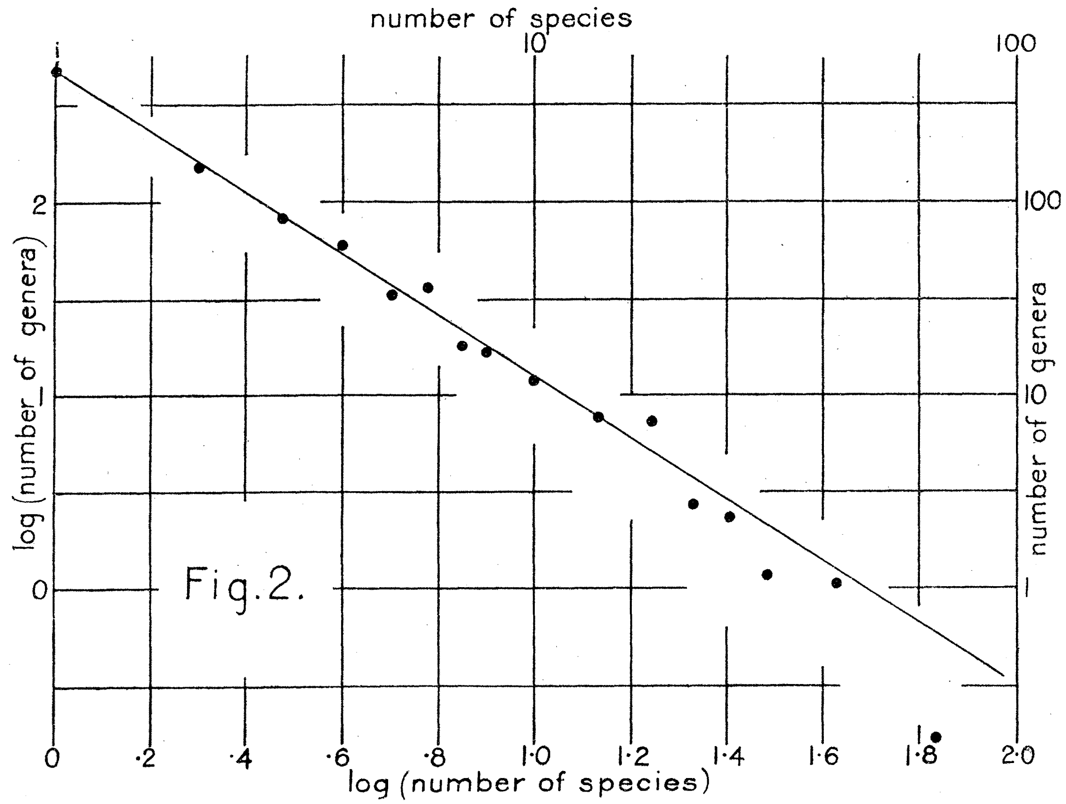

Fig. 2.—Double logarithmic chart for the frequency distribution of sizes of genera in the *Cerambycinæ*. Data in Appendix, Table B.

Fig. 1 is for the Chrysomelid beetles (data in Table A of the Appendix). Here the data are suggestive of nothing but a straight line right up to the limit of the chart, *i.e.*, up to genera of 100 species; calculation shows, however, that beyond this point there is a heavy deficiency of the larger genera as compared with the numbers that would be given by the logarithmic-linear law. Fig. 2 for the *Cerambycinæ* (beetles) (data in Table B of the Appendix) is suggestive of linearity only up to genera of some 20 or 30 species: the points given by the frequencies of the larger genera lie well below the line. In fig. 3, for the *Leguminosæ*, the falling away occurs at much the same point and is very abrupt.* The forms of such graphs seemed so odd, it appeared so unlikely that the natural form of the graph could be nearly linear over the first part of the range and then rapidly curved, that they gave rise in one's mind to all kinds of speculations—*e.g.*, the possibility of the natural distributions being truncated by the last glacial epoch! But when the logarithmic graphs are drawn for the distributions of Table II it will be seen that such forms are precisely those to be expected. The graphs, drawn from a larger number of significant figures than are given in Table II, are shown in figs. 4 to 9, the graph for each value of τ being given by the full line and the limiting graph for $\tau = \infty$ by the broken line.

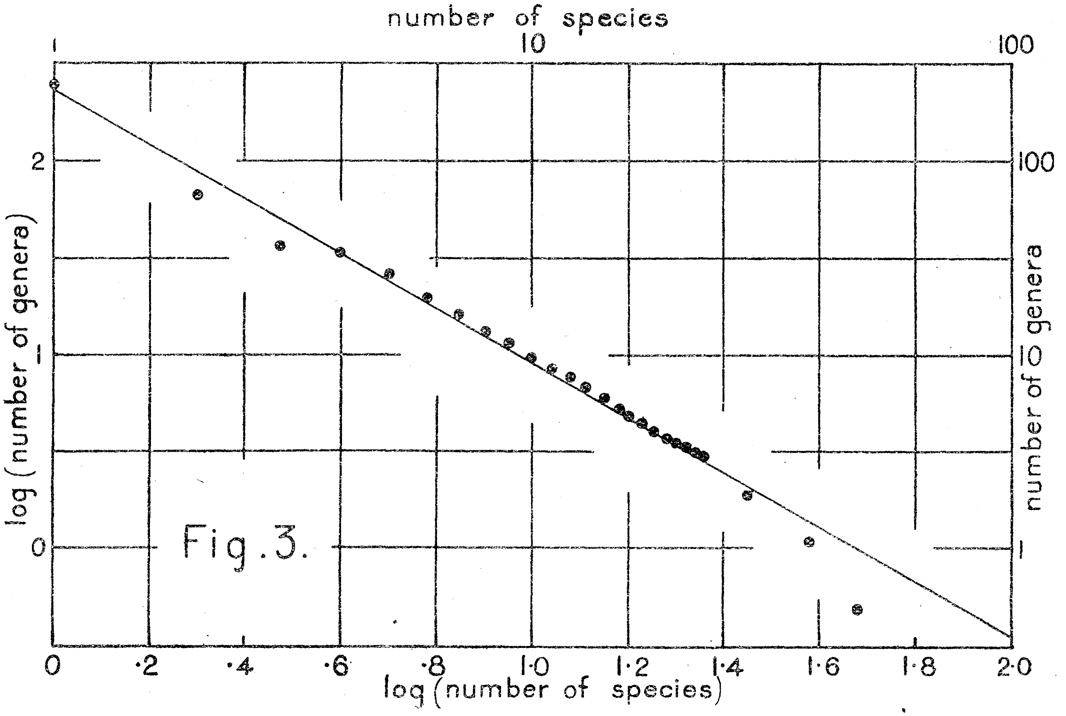

Fig. 3.—Double logarithmic chart for the frequency distribution of sizes of genera in the *Leguminosæ*. Data in Appendix, Table E.

---

* It is of course always necessary to group or graduate the data for such charts, at least in the tail of the distribution. The following were the actual groupings, etc., used in the three charts shown. *Chrysomelidæ*: ungrouped to 8; thence in groups 9–11, 12–15, 16–19, 20–23, 24–27, 28–33, 34–43, 44–53, 54–73, 74–103. *Cerambycinæ*: the same way up to 53, but terminating with 54–83. *Leguminosæ*: 1, 2, 3 ungrouped; *graduated* from 4 to 23 by the two groups 4–13, 14–23; thence grouped 24–33, 34–43, 44–53.

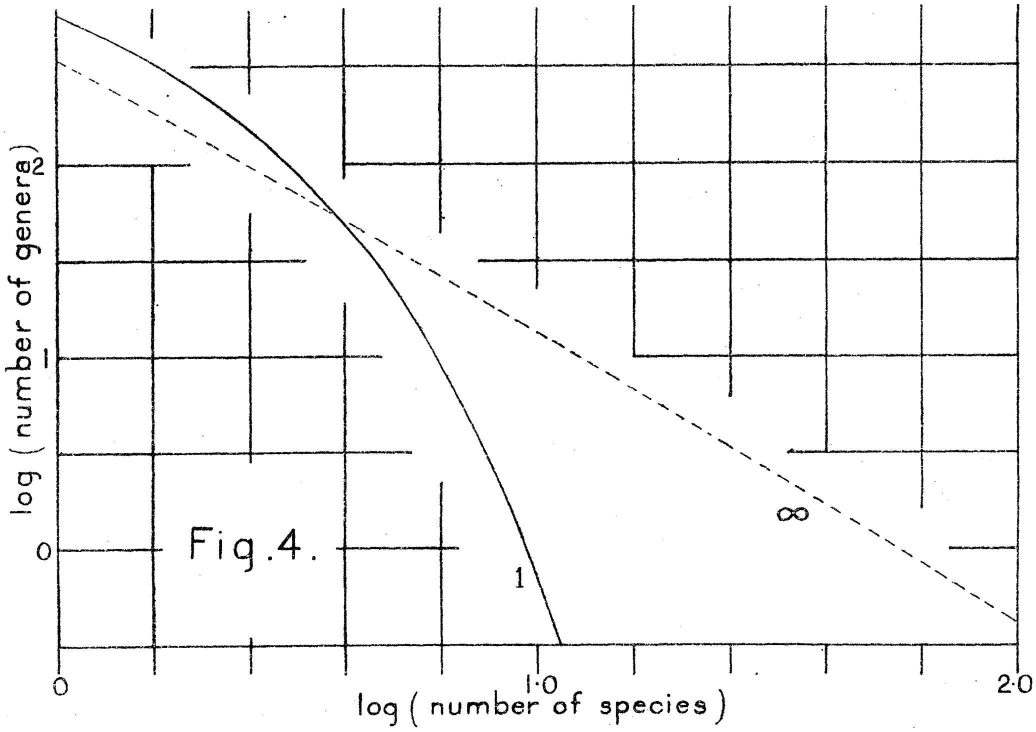

Fig. 4.

Fig. 4.—Double logarithmic graph of the frequency distribution of size of genus for $\tau = 1$, $\rho = 1 \cdot 925$, full line ; the broken line showing the limiting form at $\tau = \infty$ for comparison. Figs. 5 to 9 show the gradual approximation of the form of the graph towards the limiting, nearly linear, form as the time is increased (*cf.* Table II).

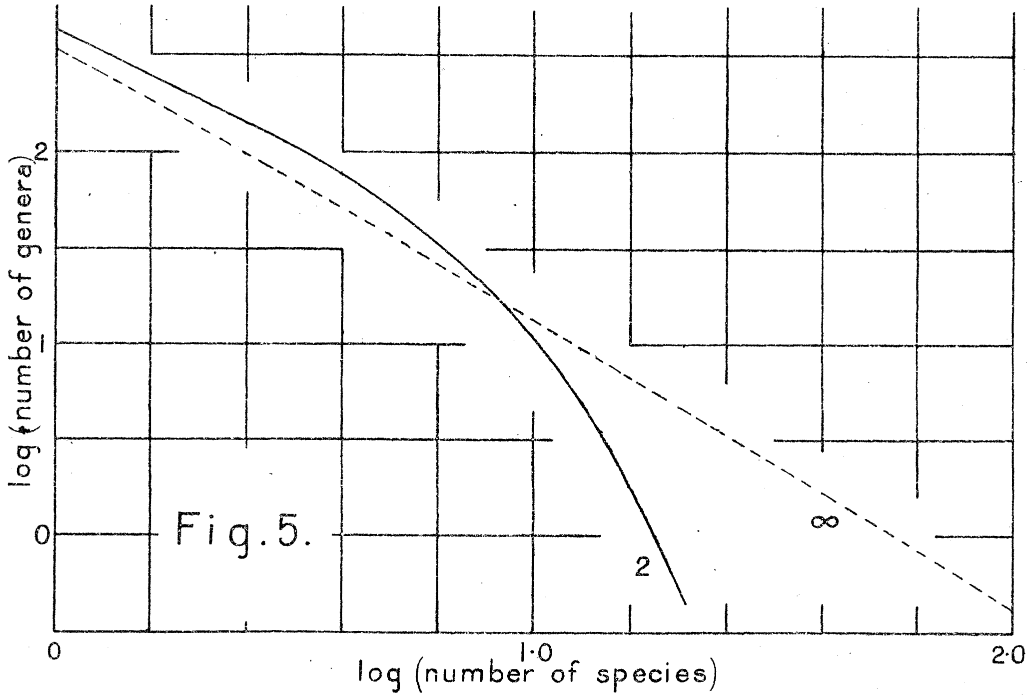

Fig. 5.

Fig. 5.—Double logarithmic graph (full line) of the frequency distribution of size of genus for $\tau = 2$, $\rho = 1 \cdot 925$.

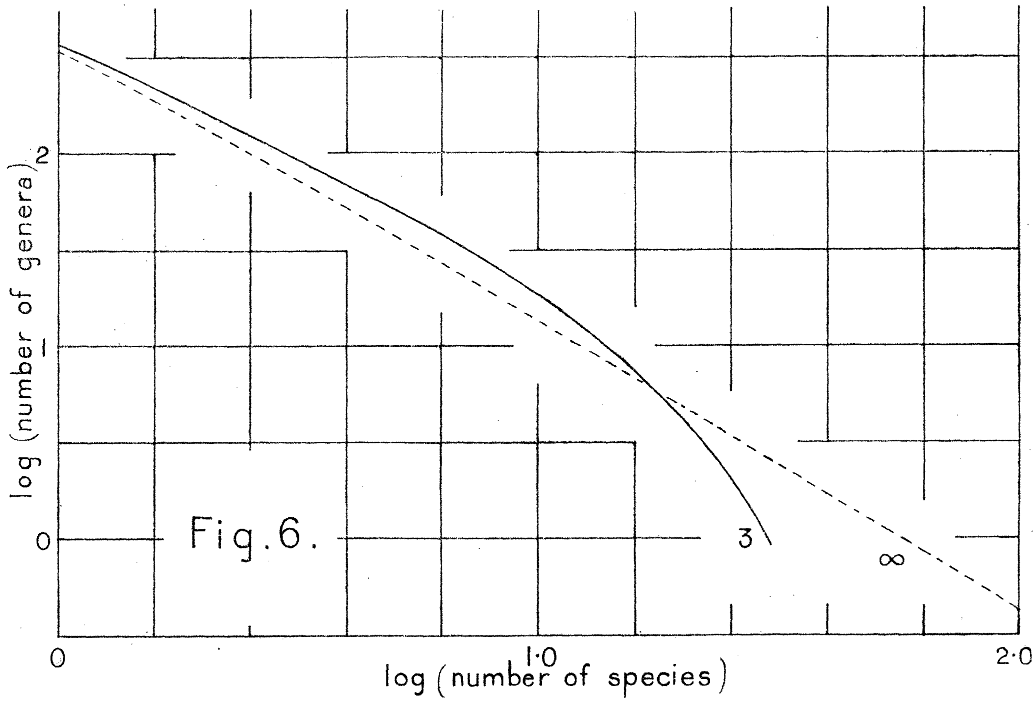

Fig. 6.—Double logarithmic graph (full line) of the frequency distribution of size of genus for $\tau = 3$, $\rho = 1\cdot925$.

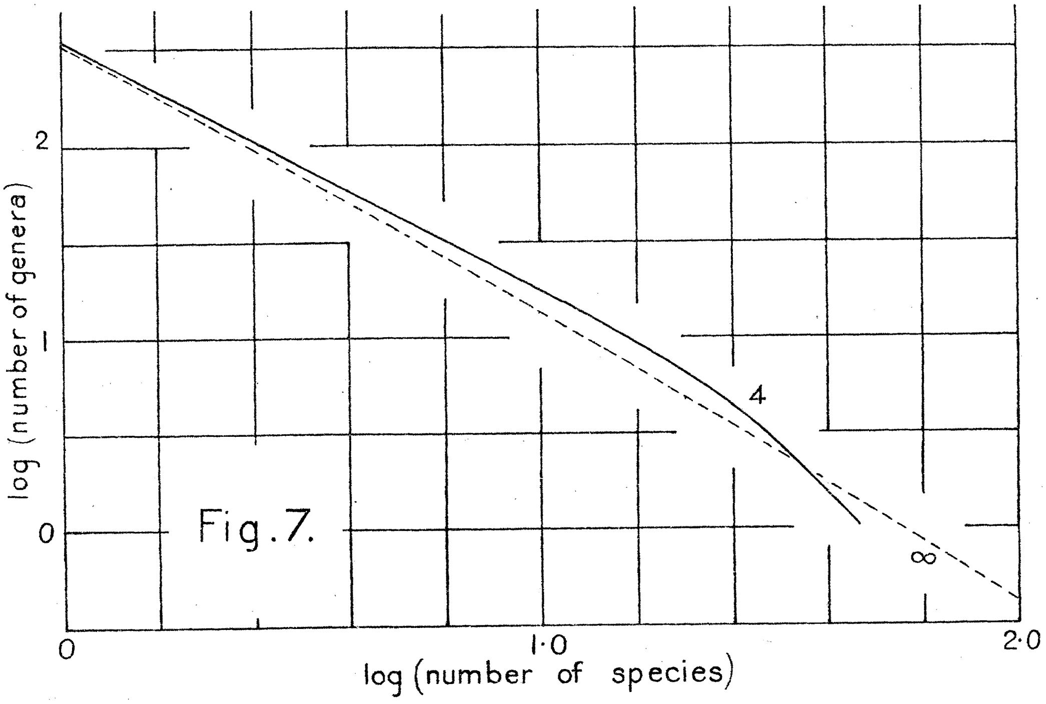

Fig. 7.—Double logarithmic graph (full line) of the frequency distribution of size of genus for $\tau = 4$, $\rho = 1\cdot925$.

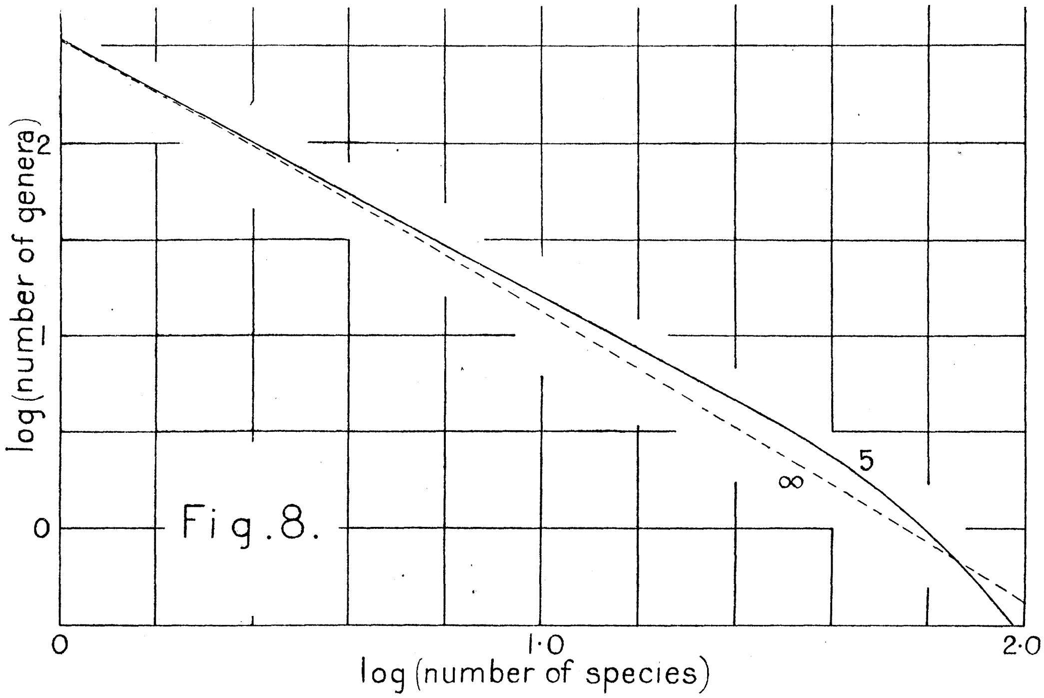

Fig. 8.

Fig. 8.—Double logarithmic graph (full line) of the frequency distribution of size of genus for $\tau = 5$, $\rho = 1 \cdot 925$.

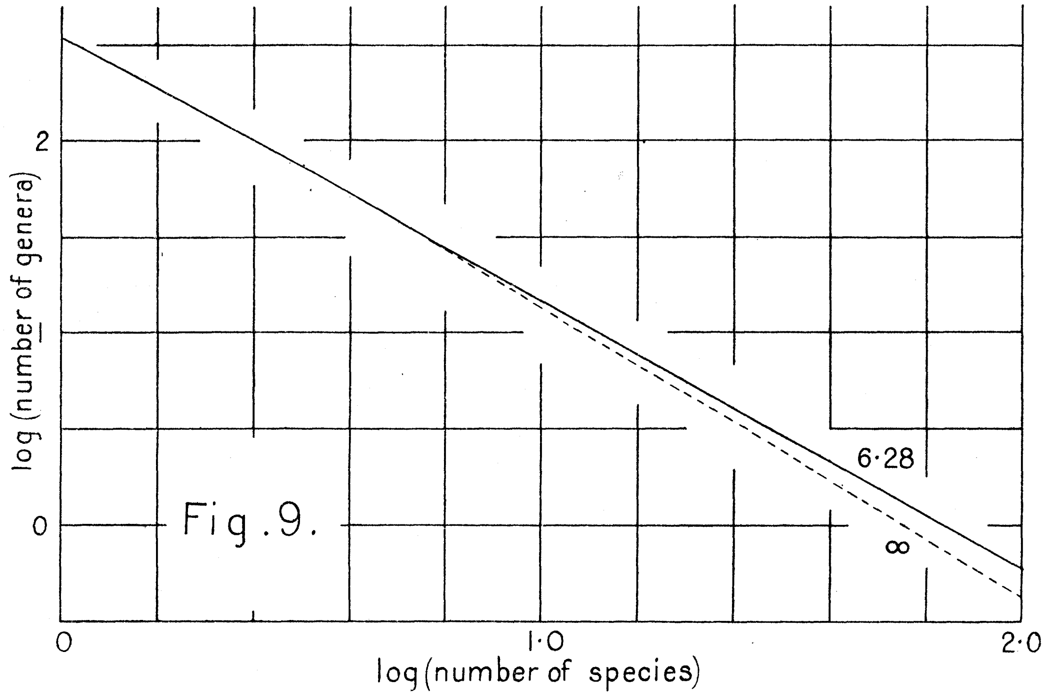

Fig. 9.

Fig. 9.—Double logarithmic graph (full line) of the frequency distribution of size of genus for $\tau = 6 \cdot 28$, $\rho = 1 \cdot 925$.

It will be seen that at first, for very short values of the time, the graph is more or less suggestive of an arc of a parabola, but the curve lies rather close to its tangent for some distance from the start. As time increases the last feature becomes more and more conspicuous, the graph being almost straight for some portion of its length and then falling away very rapidly: fig. 7 ($\tau = 4$) is very suggestive of the sort of graph given by the *Cerambycinæ* (fig. 2) or the *Leguminosæ* (fig. 3). When τ increases to 6·28, as for the *Chrysomelidæ* (figs. 1 and 9), the "straight" portion of the graph covers the whole chart and the point at which falling away becomes conspicuous lies outside the picture on the right. Within the limits shown (genera of 1 to 100 species) no actual data following the law of fig. 9 could well suggest anything but a linear law; the divergence from linearity with 1000 or 2000 genera available in all would never appear significant. It should be noted, however, that the line fitted to the data within the given range would have a smaller slope than the limiting "line," and hence if ρ were estimated therefrom on the assumption that time could be regarded as infinite too high a value would be assigned to that constant. The values of ρ estimated in this way and given at the meeting of the Linnean Society on February 2nd, 1922, when papers were read by Dr. WILLIS and myself, were in error from this cause.

Since we have not obtained the expression (20) for the frequency distribution in any simple form, the deduction of the mean size of genus from the distribution would be complex, but it may be directly deduced from quite simple considerations.

The total number of genera at time T is N_0e^{gt} where N_0 is the number of primordial genera.

The total number of species at time T is the number in genera of all ages, the number in a genus aged x being e^{sx}. Hence, ${}_sN_T$ being the required number (*cf.* the first paragraph of Section II),

$$
\begin{aligned}
{}_sN_T &= N_0e^{sT} + N_0g\int_0^T e^{g\,(T-x)}\,e^{sx}\,dx \\
&= N_0\{s\,(s-g)^{-1}\,e^{sT} - g\,(s-g)^{-1}e^{gT}\} \quad \ldots\ldots\ldots\ldots \quad (21)
\end{aligned}
$$

which checks by putting $T = 0$, when ${}_sN_T = N_0$.

The mean number of species per genus, dividing by the number of genera N_0e^{gT} is

$$
{}_sM_T = (s-g)^{-1}\{s\,e^{(s-g)\,T} - g\} \quad \ldots\ldots\ldots\ldots\ldots \quad (22)
$$

In terms of the notation when the doubling-period for species within the genus is taken as unit the equations may be written:

$$
{}_gN_\tau = N_0e^{\alpha\rho^{-1}\tau} \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (23)
$$

$$
{}_sN_\tau = N_0(\rho-1)^{-1}\{\rho e^{\alpha\tau} - e^{\alpha\rho^{-1}\tau}\} \quad \ldots\ldots\ldots\ldots \quad (24)
$$

$$
{}_sM_\tau = (\rho-1)^{-1}\{\rho e^{\alpha(1-\rho^{-1})\tau} - 1\} \quad \ldots\ldots\ldots\ldots \quad (25)
$$

These expressions increase continuously with the time and are infinite for $\tau = \infty$: but it is to be noted that the *total* number of species ${}_sN_\tau$ increases not at a steady (percentage) rate but with decreasing rapidity. The second term, depending on the number of genera, becomes of less and less importance as the proportion of genera to species decreases; while the total number of species more than doubles in the first unit of time, the percentage increase falls in every following interval and rapidly asymptotes towards 100.

It will of course be noted that we have not considered any group of higher order than the genus. If the assumption that generic mutations give fresh starting points for specific mutations is correct, the effect shown by equation (24) or (25) must follow, and it would seem, therefore, that mutations of a higher order still, such as would give rise to new families, must emphasise the result. But I have not developed the matter further.

IV. The Fitting to Data of the Expression obtained in Section III.

Given an actual frequency distribution for size of genus, the problem is to determine from it the values of ρ and τ. The method of moments would not, with distributions of the present form, be a good method to use even if it led to a simple solution, for the very long "tail" of the distribution implies high probable errors in the moments.* But I have not seen my way to any simple and *direct* solution. The method finally adopted was—(1) to fix on the proportion of monotypes and the mean number of species to the genus as the characteristics to be used for determining ρ and τ; (2) to draw up a table giving the proportion of monotypes (f_1) and the mean number of species to the genus (M) for all values of ρ and τ that seemed likely to occur, and (3) in any given case to determine ρ and τ from this table by inverse interpolation.

Table III is the fundamental table, giving f_1 and M for all values of ρ from 1·0 to 3·0 by tenths of a unit, and for all values of τ from 1 to 10 by units. As examples of fitting I have taken four tables kindly given me by Dr. Willis and shown in full, by his permission, in the Appendix to this paper. They refer to the *Chrysomelidæ* (beetles), the *Cerambycinæ* (beetles), the Snakes and the Lizards respectively. Table IV shows the numbers of genera and species in each group, the values of M and f_1 given by the data, the approximate values of ρ and τ determined from Table III by interpolation as shown below, and N_0 the number of primordial genera. Lines 8 and 9 are simply a check on the work and the precision of interpolation in Table III: such interpolation is not accurate, and the calculation of M and f_1 from the values of ρ and τ determined by interpolation is a desirable check. Lines 10 to 12 give the results of applying the χ^2 test of "goodness of fit" for the respective groupings of the data shown in Tables V to VIII: the P-tables (*Tables for Statisticians and Biometricians*) are entered with n' taken as 2 less than the number of groups, since two constants have been fitted to the data.†

As an example of the detailed work the *Cerambycinæ* may be taken. M is 5·584 and

---

* *Cf.* R. A. Fisher, "On the mathematical foundation of theoretical statistics," 'Phil. Trans.,' A, vol. 222, pp. 309–368.

† R. A. Fisher, "On the interpretation of χ^2 from contingency tables and the calculation of P," 'Jour. Roy. Stat. Soc.,' vol. 85, p. 87 (1922).

f_1 is 0·45801. Looking over Table III to find a pair of values near these, we find that they must be taken between $\rho = 1\cdot1$ and $\rho = 1\cdot2$, for a value of τ very near 5. Using simple interpolation only, we find that the observed value of f_1 is given by ($\tau = 6$, $\rho = 1\cdot185165$) and by ($\tau = 5$, $\rho = 1\cdot188224$), or approximately by any values of ρ and τ subject to the equation

$$\rho + 0\cdot003059, \tau = 1\cdot203519 \quad . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \quad (a)$$

Similarly $M = 5\cdot584$ is given by ($\rho = 1\cdot1$, $\tau = 5\cdot520408$) and by ($\rho = 1\cdot2$, $\tau = 4\cdot908312$), or approximately by any values of ρ and τ subject to the equation

$$\rho + 0\cdot163373, \tau = 2\cdot001886 \quad . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \quad (b)$$

Solving (a) and (b) we find $\tau = 4\cdot980$, $\rho = 1\cdot188$.

TABLE III.—Values of the proportion of monotypes (f_1) and of the mean size of genus (M) for values of the time τ from 1 to 10 (the unit being the doubling period for species within the genus) and of the ratio ρ (of the chance of a specific mutation to the chance of a generic mutation) from 1·0 to 3·0 (decimal point omitted before f_1).

Ratio ρ of chance of specific mutation to chance of generic mutation.		Time.									
		1	2	3	4	5	6	7	8	9	10
1·0	f_1	62,500	53,125	50,781	50,195	50,049	50,012	50,003	50,001	50,000	50,000
	M	1·693	2·386	3·079	3·773	4·466	5·159	5·852	6·545	7·238	7·931
1·1	f_1	61,566	51,333	48,608	47,882	47,689	47,638	47,624	47,620	47,619	47,619
	M	1·715	2·477	3·289	4·153	5·074	6·054	7·098	8·211	9·395	10·656
1·2	f_1	60,761	49,750	46,660	45,793	45,549	45,481	45,462	45,457	45,455	45,455
	M	1·735	2·560	3·485	4·524	5·691	7·000	8·470	10·119	11·971	14·049
1·3	f_1	50,060	48,343	44,905	43,897	43,601	43,514	43,489	43,481	43,479	43,479
	M	1·752	2·634	3·669	4·883	6·309	7·981	9·944	12·247	14·949	18·121
1·4	f_1	59,444	47,084	43,318	42,170	41,820	41,713	41,681	41,671	41,668	41,667
	M	1·767	2·701	3·840	5·229	6·921	8·985	11·500	14·566	18·304	22·860
1·5	f_1	58,899	45,953	41,875	40,591	40,186	40,059	40,018	40,006	40,002	40,001
	M	1·780	2·762	4·000	5·560	7·524	10·000	13·119	17·049	22·000	28·238
1·6	f_1	58,413	44,930	40,559	39,141	38,682	38,533	38,485	38,469	38,464	38,462
	M	1·792	2·818	4·149	5·876	8·115	11·018	14·784	19·667	25·999	34·212
1·7	f_1	57,977	44,001	39·353	37,807	37,293	37,122	37,065	37,046	37,040	37,038
	M	1·802	2·869	4·289	6·178	8·690	12·032	16·479	22·393	30·262	40·730
1·8	f_1	57,584	43,154	38,245	36,575	36,007	35,814	35,748	35,726	35,718	35,716
	M	1·812	2·916	4·420	6·465	9·249	13·037	18·191	25·205	34·750	47·738
1·9	f_1	57,228	42,379	37,224	35,434	34,813	34,597	34,523	34,497	34,488	34,484
	M	1·820	2·960	4·542	6·739	9·790	14·027	19·910	28·080	39·426	55·180

TABLE III—(continued).

			Time.									
			1	2	3	4	5	6	7	8	9	10
Ratio ρ of chance of specific mutation to chance of generic mutation.	2·0	f_1	56,904	41,667	36,280	34,375	33,702	33,464	33,379	33,350	33,339	33,335
		M	1·828	3·000	4·657	7·000	10·314	15·000	21·627	31·000	44·255	63·000
	2·1	f_1	56,607	41,010	35,404	33,389	32,664	32,404	32,311	32,277	32,265	32,261
		M	1·836	3·037	4·765	7·248	10·819	15·953	23·335	33·948	49·206	71·144
	2·2	f_1	56,335	40,403	34,590	32,468	31,695	31,412	31,309	31,272	31,258	31,253
		M	1·842	3·072	4·866	7·485	11·307	16·885	25·027	36·909	54·251	79·561
	2·3	f_1	56,084	39,840	33,831	31,608	30,786	30,482	30,369	30,327	30,312	30,306
		M	1·849	3·104	4·962	7·710	11·777	17·795	26·698	39·872	59·364	88·204
	2·4	f_1	55,852	39,316	33,122	30,801	29,932	29,607	29,485	29,439	29,422	29,416
		M	1·854	3·134	5·052	7·925	12·230	18·681	28·345	42·826	64·522	97·030
	2·5	f_1	55,638	38,828	32,458	30,044	29,129	28,783	28,652	28,602	28,583	28,576
		M	1·860	3·162	5·137	8·130	12·667	19·543	29·965	45·763	69·707	106·000
	2·6	f_1	55,438	38,372	31,835	29,332	28,373	28,006	27,865	27,811	27,791	27,783
		M	1·864	3·189	5·218	8·326	13·087	20·381	31·556	48·675	74·901	115·078
	2·7	f_1	55,252	37,944	31,250	28,660	27,659	27,271	27,122	27,064	27,041	27,032
		M	1·869	3·214	5·294	8·512	13·492	21·196	33·115	51·557	80·089	124·233
	2·8	f_1	55,079	37,544	30,699	28,027	26,984	26,576	26,418	26,356	26,331	26,322
		M	1·873	3·237	5·366	8·691	13·882	21·987	34·643	54·404	85·259	133·436
	2·9	f_1	54,916	37,167	30,179	27,428	26,344	25,918	25,750	25,684	25,658	25,648
		M	1·877	3·259	5·435	8·861	14·257	22·755	36·137	57·211	90·399	142·663
	3·0	f_1	54,764	36,812	29,687	26,860	25,738	25,293	25,116	25,046	25,018	25,007
		M	1·881	3·280	5·500	9·024	14·619	23·500	37·598	59·976	95·500	151·890

These values, it will be seen from Table IV, reproduce f_1 almost with precision (0·457979 against 0·45801) but give by equation (25) a mean 5·593 instead of 5·584, a small and hardly material difference. The value not being quite precise, however, we have two alternative values for the initial number of genera. Putting $N_0 = 1$, the numbers of genera and of species at time τ are found by equations (23) and (24) to be 18·2765 and 102·2133 respectively: hence the initial number of genera is either

$$1024/18{\cdot}2765 = 56{\cdot}03$$

or

$$5718/102{\cdot}2133 = 55{\cdot}94$$

or roundly 56·0. The interpolation is less precise in the case of the *Chrysomelidæ*, where the two alternatives are 65·35 and 66·05, the number corresponding to the estimate based on the genera being given in Table IV.

TABLE IV.—Giving particulars respecting four frequency distributions for size of genus. The distributions are given in a condensed form in the following Tables V to VIII and in full in the Appendix.

——	Chrysomelidæ.	Cerambycinæ.	Snakes.	Lizards.
1. Number of genera	627	1,024	293	259
2. Number of species	9,997	5,718	1,475	1,580
3. Mean species per genus, M	15·94	5·584	5·034	6·100
4. Proportion of monotypes, f_1	0·3429	0·4580	0·4471	0·4054
5. τ	6·28	4·980	4·260	4·281
6. ρ	1·925	1·188	1·253	1·496
7. N_0	65·4	56·0	27·8	36·0
8. Mean from ρ, τ	15·77	5·593	5·038	6·130
9. Proportion of monotypes from ρ, τ ...	0·3428	0·4580	0·4466	0·4049
10. χ^2	11·21	13·76	13·77	3·50
11. n'	16	14	11	11
12. P	0·74	0·39	0·18	0·96

From the given value of ρ, the values of the successive terms of the ρ-series (12) are calculated as there described, and checked from equation (13) by summation at intervals. The first few values in the present case are :—

1	0·457 038
2	0·160 830
3	0·083 727
4	0·051 878
5	0·035 522

The series for the *Cerambycinæ* was calculated up to f_{65}.

Next, the values of c_1', c_2', c_3', . . . are calculated from equations (19). As already mentioned, the tables of binomial coefficients were only calculated up to the twelfth in the final part of my table, so it was no use going beyond c_{12}' : as many correcting terms as this may not always be wanted, but it saves time to calculate all the twelve at once in case they may be needed.

c_1'	$10^{-3} \times 0\cdot941\ \ 3398$
c_2'	$10^{-4} \times 0\cdot386\ \ 6267$
c_3'	$10^{-5} \times 0\cdot135\ \ 9285$
c_4'	$10^{-7} \times 0\cdot455\ \ 6658$
c_5'	$10^{-8} \times 0\cdot149\ \ 5844$
c_6'	$10^{-10} \times 0\cdot485\ \ 6397$
c_7'	$10^{-11} \times 0\cdot156\ \ 2737$
c_8'	$10^{-13} \times 0\cdot503\ \ 0638$
c_9'	$10^{-14} \times 0\cdot161\ \ 1061$
c_{10}'	$10^{-16} \times 0\cdot514\ \ 8885$
c_{11}'	$10^{-17} \times 0\cdot164\ \ 3085$
c_{12}'	$10^{-19} \times 0\cdot523\ \ 7340$

Retaining seven decimal places in the work,* so as to be fairly confident of correctness in the sixth place of the result, terms were required up to c_{11}' in the corrections to f_{57} onwards. I have found it convenient to arrange the work of equations (17) in the following form: the decimal places have been reduced to six for the illustration, and the decimal point and zeros thereafter are omitted for brevity as they were in working.

f	1 $+$	2 $-$	3 $+$	4 $-$	Total correction.	ρ-series.	Total.	Multiplied by 1024.
1	941	—	—	—	+941	457,038	457,979	469·0
2	941	39	—	—	+902	160,830	161,733	165·6
3	941	77	1	—	+865	83,727	84,592	86·6
4	941	116	4	—	+829	51,878	52,707	54·0
5	941	155	8	—	+794	35,522	36,317	37·2
6	941	193	14	—	+762	25,960	26,721	27·4
7	941	232	20	1	+728	19,863	20,592	21·1

Each component correcting term c' is taken in turn and multiplied by the corresponding successive binomial coefficients, the products being entered in the vertical columns headed 1, 2, 3, 4, etc. For c_1' the binomial coefficients are unity throughout; for c_2' they run 1, 2, 3, . . . beginning with f_2; for c_3' they are 1, 3, 6, 10, 15, . . . beginning with f_3; for c_4' 1, 4, 10, 20, 35, . . . beginning with f_4 and so on. The signs of the products are given at the heads of the columns. The summation of these terms then gives the total correction, in the next column, the sign of which is always positive at the commencement, and added therefore to the corresponding term of the ρ-series gives the required value of ${}_{r}f_n$. Finally, multiplying this by the observed number of genera we have the calculated number of genera of each size. In the case of grouped distributions like Tables V to VIII the grouped frequency has been calculated by adding the values of ${}_{r}f_n$ so as to obtain an answer correct in the decimal place. The distribution for the *Chrysomelidæ* was the most laborious to calculate as the series is very extensive: terms were calculated direct up to f_{102}, but an estimate was also made by extrapolation of the portion of the remaining frequency lying between f_{101} and f_{150} inclusive. As two different methods of extrapolation gave nearly the same answer it is hoped the result is fairly close to the truth.

With this digression on the method of fitting we may now revert to Table IV and consider first the particulars respecting goodness of fit in lines 10–12. It will be seen that the values of P are highest for the Lizards and the *Chrysomelidæ*, the formulæ giving in both these cases a most excellent fit to the data. For the *Cerambycinæ* and the Snakes the fits are not quite so good, but still well within the limits of fluctuations of

---

* In most of the work at this stage six decimal places only were retained, not seven. In calculating the ρ-series it is desirable to retain seven significant figures so as to avoid accumulating errors.

sampling; even in the case of the Snakes one would expect to get a worse fit, merely owing to the chances of sampling, once in some five or six trials. Reference to Tables V to VIII fully confirms the impression given by the values of P. For the *Chrysomelidæ* the fit is worst for genera of 3 species, of which there are only 38 against an expectation of 49, and genera of 101–150 species, of which there are only 7 against an expectation of over 12: of all genera with more than 100 species there are only 16 against an expectation of 23·6. The signs of the divergences from expectation are fairly well scattered except over the range 7 to 30 where (for the given grouping) all the observed frequencies are slightly in excess of expectation.

TABLE V.—*Chrysomelidæ*: observed and calculated numbers of genera of each size.

Number of species in genus.	Number of genera.	
	Observed.	Calculated.
1	215	214·9
2	90	85·6
3	38	48·9
4	35	32·6
5	21	23·8
6	16	18·3
7	15	14·7
8	14	12·2
9 to 11	28	27·0
12 to 14	20	18·6
15 to 20	30	24·9
21 to 30	32	25·0
31 to 40	13	15·9
41 to 50	14	11·4
51 to 75	17	18·5
76 to 100	13	11·1
101 to 150 ...	7	12·3*
151 upwards ...	9	11·3*
Total ...	627	627·0

* The frequency of genera of 101 species and upwards was subdivided by extrapolation.

If we re-group the frequencies of Table V according to the runs of sign of the differences from expectation, using groups 1–2, 3, 4, 5–6, 7–30, 31–40, 41–50, 51–75, 76–100, 101 upwards, there are 10 groups; n' is 8, χ^2 is 9·59, and P 0·21, so that the distribution stands the severe test very well.

For the *Cerambycinæ* the most marked divergence is in the group of genera with 15–20 species, of which there are 40 against an expectation of only 27·4: reference to the detailed data in the Appendix will show that there is here a marked "hump" in the data which could not be covered by any smooth curve. Re-grouping Table VI by the runs of sign of the differences between observation and expectation only reduces P from 0·39 to 0·32.

TABLE VI.—*Cerambycinæ*: observed and calculated numbers of genera of each size.

Number of species in genus.	Number of genera.	
	Observed.	Calculated.
1	469	469·0
2	152	165·6
3	82	86·6
4	61	54·0
5	33	37·2
6	36	27·4
7	18	21·1
8	17	16·8
9 to 11	36	35·1
12 to 14	23	22·4
15 to 20	40	27·4
21 to 30	21	24·1
31 to 40	15	13·0
41 to 50	8	8·0
51 to 65	4	7·0
66 upwards ...	9	9·5
Total ...	1,024	1,024·2

TABLE VII.—Snakes: observed and calculated numbers of genera of each size.

Number of species in genus.	Number of genera.	
	Observed.	Calculated.
1	131	130·9
2	35	47·2
3	28	25·2
4	17	16·0
5	16	11·2
6	9	8·3
7	8	6·5
8	8	5·2
9 to 11	13	11·1
12 to 14	3	7·2
15 to 20	7	8·8
21 to 34	14	9·2
35 upwards ...	4	6·2
Total ...	293	293·0

For the Snakes the fit is clearly less satisfactory, and the data (Appendix, Table C) are irregular. It is true that we have passed from a family and a sub-family amongst the beetles to the members of an entire order, but the comparative poorness of the fit can

TABLE VIII.—Lizards : observed and calculated numbers of genera of each size.

Number of species in genus.	Number of genera.	
	Observed.	Calculated.
1	105	104·9
2	44	39·9
3	23	22·2
4	14	14·5
5	12	10·5
6	7	8·0
7	6	6·4
8	4	5·2
9 to 11	13	11·4
12 to 14	4	7·7
15 to 20	10	9·8
21 to 34	11	10·8
35 upwards ...	6	7·7
Total ...	259	259·0

hardly be due to this cause, for the Lizards give the best fit of the four tables, the calculated figures being extraordinarily close to those observed. It may be noticed that in each table the numbers of the largest genera, those in the terminal group, are in defect as compared with expectation, but the deficiency is very small in one case at least (Table VI), and the experience too limited to suggest a general rule. In the case of the *Chrysomelidæ*, moreover, the lack of precision in interpolation has partly contributed to the result, for it will be seen that the mean of the fitted distribution (Table IV, lines 3 and 8) is slightly too low. Desire of the systematist to break up a genus which he regarded as unwieldy might well tend to cause a deficiency of very large genera, but such a deficiency can hardly be held to be proved by the present tables.

So far as the tests go I think it must be admitted that the formula given is capable of representing the facts with considerable precision, more closely indeed than we have any right to expect. One might well have expected the personal factor in classification, the practically cataclysmic destruction of species at numerous epochs in geological time, and all the varied changes that have diversified the organic history of our planet, to have left so many irregularities in the distribution that any formula could at most have given a very rough analogy with the general run of the data. But apparently the formula arrived at can do far more than this. Why this is so seems to me to be a point which requires some discussion. I return to it briefly below.

Having shown, however, that the formula is capable of closely describing the facts, we can revert to Table IV, and the actual values of ρ and τ. It will be seen in the first place that the values found for τ range only from 4·26 to 6·28, the unit of time being the doubling-period for species within the genus *in each several case.* Time must, therefore, be regarded as not merely finite, but short: it is indeed obvious, once the point is considered, that when the mean size of genus is only some 5 to 15 species time cannot well be anything but short. It may be as well to emphasise that our unit of time being a relative unit, its equivalent in years or in geological time will vary from group to group and can only be determined if the results can effectively and without fallacy be collated with the geological record.

But line 7 of Table IV suggests that great caution will have to be used in interpreting doubling-periods in terms of geological time. In every case the number of genera at zero time required to fit the data is very substantial: roundly, some 10 per cent. of the existing number for the *Chrysomelidæ* and the Snakes, 5 per cent. of the existing number for the *Cerambycinæ*, and nearly 15 per cent. for the Lizards. What does this mean? Such a result would presumably be shown in any case where the origin of the group was polyphyletic, but I think it may have another significance and one which will render the interpretation of the figures a matter of difficulty.

Consider the effect on a distribution of the advent of a "cataclysm," *e.g.*, of a glacial epoch, killing off a large number of species. Every genus will be reduced in size; a genus of 20 species may be reduced to a genus of 5 species, a genus of 10 species to a genus of 2, and so on. Some—possibly many—genera will be killed off outright. In the limit, if the cataclysm be very severe, *most* genera will be killed out entirely and no genera will be left with more than a single species. On the passing of such a cataclysm of the utmost possible severity and the restoration of conditions favourable to life, evolution will start again *de novo*, but from an initial number of N_0 monotypic genera instead of a single one. When, long ages after, the biological statistician examines the frequency distribution for sizes of genus in the group evolved from the survivors of the cataclysm, he will therefore find, as we have found, a number of primordial genera N_0 in excess of unity: and further he will have to remember that the time τ will be measured, not from the origin of the group, but from the passing of the cataclysm.

In fact, of course, we must expect matters to be far more complex even than this. The action of a "cataclysm" of less than limiting severity might be represented, in highly simplified terms, somewhat as follows. For any species the chance of survival is p, of destruction q $(p+q=1)$. The chance of 0, 1, 2, . . . n species surviving out of a genus of n species will then be given by the successive terms of $(q+p)^n$. Given the pre-cataclysmic distribution and the value of p, it will then be possible, though a lengthy piece of work, to calculate the post-cataclysmic distribution. It is desirable to carry out such a test on assumed data to see whether, if the pre-cataclysmic distribution be of the form above derived, the post-cataclysmic distribution will or will not also be closely fitted by the same formula, and, if it is closely fitted, what is the effect on the constants. It is evident that the distribution will remain of the same general form to the eye, with a maximum frequency for the monotypics, and I am inclined to suspect that it may be fairly closely of the same mathematical form, but have not yet carried out the test. It would be further desirable to find the form of the post-cataclysmic distribution after a subsequent period of evolution. These are problems still unsolved.

I have, however, carried out two brief empirical tests to find the effect of compounding the distribution. A cataclysm might—or probably would in general—cover part only of the area occupied by a given organic group. The post-cataclysmic distribution would then be compound as regards time, one portion starting *de novo*, if the cataclysm were of limiting severity, one portion remaining untouched.

Suppose that at the time of observation the constants for the two portions are :—

$$\left.\begin{array}{l}\rho = 2\\ \tau = 2\end{array}\right\}\qquad\qquad \left.\begin{array}{l}\rho = 2\\ \tau = 8\end{array}\right\}$$

and that there is the same number of genera in each of the two groups. We will then have

		f_1	M
$\rho = 2$	$\tau = 2$	0·41667	3·000
$\rho = 2$	$\tau = 8$	0·33350	31·000
Compound		0·37508	17·000

the values of f_1 and M for the compound distribution being the means of the values for the components. Using the method of interpolation described above, we would find for the compound distribution the constants

$$\rho = 1\cdot668 \qquad\qquad \tau = 7\cdot203$$

τ tends accordingly to be nearer the higher value of the time and ρ is reduced below its true value. The first 10 terms of the respective series work out as follows, taking the total as 10,000 :—

$\rho =$	2	2	Compound.	1·668
$\tau =$	2	8		7·203
1	4,167	3,335	3,751	3,750
2	1,917	1,335	1,626	1,444
3	1,162	764	963	803
4	775	510	642	525
5	540	371	455	375
6	385	286	335	285
7	278	229	253	225
8	203	189	196	184
9	149	159	154	153
10	110	137	123	131

The fitted distribution is, it will be seen, a poor fit to the actual distribution, the fitted frequencies being badly in defect over the range from f_2 to f_8 inclusive. Over this range observation would give a total frequency 4,470, while the fitted distribution would give only 3,841—a deficiency of 14 per cent.

But the question also arises, what is the effect of a distribution being compound as regards ρ? In all the preceding work we have assumed ρ to be the same for all genera in the group, an assumption which is on the face of it very unlikely to be in accordance with the facts. To carry out a test on this point I took the two distributions

$$\left.\begin{array}{l}\rho = 1\cdot3\\ \tau = 8\end{array}\right\} \qquad \left.\begin{array}{l}\rho = 2\cdot0\\ \tau = 8\end{array}\right\}$$

and assumed them to be compounded in equal proportions, giving :—

		f_1	M
$\rho = 1\cdot3$	$\tau = 8$	0·43481	12·247
$\rho = 2\cdot0$	$\tau = 8$	0·33350	31·000
Compound		0·384155	21·6235

Fitting the compound as before, I find approximately

$$\rho = 1\cdot603 \qquad \tau = 8\cdot293$$

Here ρ takes a value between those of the components, while τ is thrown up above its true value. The first ten terms of the respective series run as shown in the table below.

$\rho =$	1·3	2	Compound.	1·603
$\tau =$	8	8		8·293
1	4,348	3,335	3,842	3,842
2	1,570	1,335	1,453	1,465
3	833	764	798	809
4	524	510	517	525
5	364	371	367	373
6	269	286	277	282
7	208	229	218	222
8	166	189	177	180
9	136	159	148	150
10	114	137	125	127

The misfit over the first part of the range is now reversed, the fitted frequencies being in excess, but the excess is relatively small.

For the four Tables V to VIII, I find :—

	Sum of f_2 to f_8 inclusive. Observed.	Calculated.
Chrysomelidæ	229	236·1
Cerambycinæ	399	408·7
Snakes	121	119·6
Lizards	110	106·7

The differences are all small, the fitted frequencies being in excess for the first two tables and in defect for the last two. There is no clear evidence of either type of compounding, and the goodness of fit remains rather puzzling.

The four tables given as illustrations in this section were the first, and at the time of writing the only, tables which I had completely fitted: they are in no way tables selected to show exceptionally good fit. A good deal of the preliminary work with purely graphic charts was done with data for the flowering plants, but these are not satisfactory for a precise test of theory. The flowering plants have not been completely catalogued and the numbers of species in the larger genera are estimates rounded off to the nearest 5 or 10 (or even for the very large genera the nearest 50 or 100). In these circumstances all that can be expected, at the best, is a fair fit of the formulæ to data which have been submitted to a preliminary graduation. Dr. WILLIS had compiled an estimated distribution for the aggregate of all the flowering plants, 12,571 genera and 160,171 species, giving 38·605 per cent. of monotypes and mean species per genus 12·741. These data give $\rho = 1\cdot594$, $\tau = 6\cdot484$, $N_0 = 750$ approximately (750 to 756). But these constants do not give a good fit, expectation exceeding observation for genera of 2 and 3 species and being continuously in defect of the graduated data from genera of 4 to genera of 34 species, which was as far as I carried the work. The group is so heterogeneous that I do not think a good fit was to be expected: but the type of misfit is not very clearly analogous with either type of misfit shown in the two illustrations above for τ compound and ρ compound respectively. We might reasonably, of course, expect compoundedness as regards both elements.

Brief and inadequate as it is, this discussion is at least sufficient to indicate, I hope, that the interpretation of the values of ρ and τ arrived at in any given case or series of cases is not a simple matter nor likely to be quite straightforward.

To all the other difficulties of interpretation is to be added the fact that errors of sampling in τ and ρ are very high. This became evident at an early stage of the work, since mere inspection of Table III showed that alterations of f_1 and M well within the limits of fluctuations of sampling would produce relatively large changes in τ and ρ. The conclusion is fully confirmed by the following investigation.

We determine τ and ρ virtually by solving the equations

$$f_1 = \phi(\rho, \tau) \qquad\qquad M = \psi(\rho, \tau).$$

Hence

$$\left.\begin{aligned} df_1 &= \frac{\partial\phi}{\partial\rho}\, d\rho + \frac{\partial\phi}{\partial\tau}\, d\tau \\ dM &= \frac{\partial\psi}{\partial\rho}\, d\rho + \frac{\partial\psi}{\partial\tau}\, d\tau \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots \quad (26)$$

Writing for brevity

$$\partial\phi/\partial\rho = a_1 \qquad\qquad \partial\psi/\partial\rho = a_2$$

$$\partial\phi/\partial\tau = b_1 \qquad\qquad \partial\psi/\partial\tau = b_2$$

$$k = a_2 b_1 - a_1 b_2$$

and solving (26) for $d\rho$ and $d\tau$ we have

$$d\tau = k^{-1}(a_2\,df_1 - a_1\,dM)$$
$$d\rho = k^{-1}(b_1\,dM - b_2\,df_1).$$

Whence, squaring and summing,

$$\left.\begin{aligned}\sigma_\tau^2 &= k^{-2}(a_1^2\sigma_m^2 + a_2^2\sigma_f^2 - 2a_1a_2\,r_{mf}\,\sigma_m\sigma_f)\\ \sigma_\rho^2 &= k^{-2}(b_1^2\sigma_m^2 + b_2^2\sigma_f^2 - 2b_1\,b_2\,r_{mf}\,\sigma_m\sigma_f)\end{aligned}\right\}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (27)$$

Where σ_m, σ_f are the standard errors of M and f_1 and r_{mf} is the correlation between errors of sampling in these two constants, I find

$$\left.\begin{aligned}a_1 &= -\rho^{-1}(1+\rho)^{-1}(1+\alpha\tau\rho^{-1}) + f_1\rho^{-1}\{(1+\rho)^{-1} + \alpha\tau\rho^{-1}\}\\ a_2 &= +\rho^{-1}(\rho-1)^{-1}(1+\alpha\tau\rho^{-1}) - M\rho^{-1}\{(\rho-1)^{-1} - \alpha\tau\rho^{-1}\}\\ b_1 &= -\{f_1 - (1+\rho)^{-1}\}\alpha(1+\rho)\rho^{-1}\\ b_2 &= +\{M + (\rho-1)^{-1}\}\alpha(\rho-1)\rho^{-1}\end{aligned}\right\} \quad . \quad . \quad . \qquad (28)$$

and the only undetermined quantity in (27) is r_{mf}. If $f_1, f_2, f_3, \ldots f_n$ are the proportional frequencies, so that

$$f_1 + f_2 + f_3 + \ldots + f_n = 1$$

we have

$$f_1 + 2f_2 + 3f_3 + \ldots + nf_n = M.$$

If, now, there is an error δ in f_1 the compensating errors in the remaining frequencies will most probably be distributed over those frequencies in proportion to their respective magnitudes, so that the new distribution will be

$$f_1 + \delta,\ f_2 - f_2\delta/(1-f_1),\ f_3 - f_3\delta/(1-f_1),\ \ldots,\ f_n - f_n\delta/(1-f_1)$$

the mean of which reduces to

$$M - \delta(M-1)(1-f_1)^{-1}.$$

Hence the regression of errors in M on errors in f_1 is $-(M-1)(1-f_1)^{-1}$ or

$$r_{mf}\,\sigma_m/\sigma_f = r_{mf}\,\sigma f_1^{-\frac{1}{2}}(1-f_1)^{-\frac{1}{2}} = -(M-1)(1-f_1)^{-1}$$

where σ is the standard deviation of size of genus: that is

$$\left.\begin{aligned}r_{mf} &= -(M-1)\sigma^{-1}f_1^{\frac{1}{2}}(1-f_1)^{-\frac{1}{2}}\\ r_{mf}\,\sigma_m\sigma_f &= -N^{-1}(M-1)f_1\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (29)$$

I have failed to get any effective simplification by inserting the values of a_1, a_2, b_1, b_2, and r_{mf} in (27), and it seems best to conduct the calculation by working out the values of the partial differential coefficients from (28), checking from Table III to see that they are approximately correct, and inserting the numerical values in (27).

The following statement shows the results of the calculation for the four illustrations used above. It will be seen that in each case the standard error of τ is of the order of half a unit, and the standard error of ρ of the order of 0·1 to 0·2. For the *time* especially the determination is wholly lacking in precision: the high standard error renders a better method of fitting, which would reduce the standard error, very desirable.

Chrysomelidæ	$f_1 = 0\cdot342903$	$M = 15\cdot9442$	$\sigma = 46\cdot2341$
$N = 627$	$\sigma_f = 0\cdot01896$	$\sigma_m = 1\cdot8464$	$r_{mf} = -0\cdot2335$
$\rho = 1\cdot925$	$a_1 = -0\cdot11550$	$a_2 = +11\cdot6068$	
$\tau = 6\cdot28$	$b_1 = -0\cdot0010771$	$b_2 = +5\cdot6706$	
	$\sigma_\tau = 0\cdot418$	$\sigma_\rho = 0\cdot167$	
Cerambycinæ	$f_1 = 0\cdot458008$	$M = 5\cdot5840$	$\sigma = 11\cdot5871$
$N = 1024$	$\sigma_f = 0\cdot01557$	$\sigma_m = 0\cdot3621$	$r_{mf} = -0\cdot3637$
$\rho = 1\cdot188$	$a_1 = -0\cdot20614$	$a_2 = +6\cdot1426$	
$\tau = 4\cdot98$	$b_1 = -0\cdot0012378$	$b_2 = +1\cdot1960$	
	$\sigma_\tau = 0\cdot408$	$\sigma_\rho = 0\cdot077$	
Snakes	$f_1 = 0\cdot447099$	$M = 5\cdot0341$	$\sigma = 9\cdot2945$
$N = 293$	$\sigma_f = 0\cdot02905$	$\sigma_m = 0\cdot5430$	$r_{mf} = -0\cdot3903$
$\rho = 1\cdot253$	$a_1 = -0\cdot18975$	$a_2 = +4\cdot1762$	
$\tau = 4\cdot260$	$b_1 = -0\cdot0052924$	$b_2 = +1\cdot25775$	
	$\sigma_\tau = 0\cdot576$	$\sigma_\rho = 0\cdot164$	
Lizards	$f_1 = 0\cdot40541$	$M = 6\cdot1004$	$\sigma = 13\cdot9294$
$N = 259$	$\sigma_f = 0\cdot03051$	$\sigma_m = 0\cdot8655$	$r_{mf} = -0\cdot3023$
$\rho = 1\cdot496$	$a_1 = -0\cdot15291$	$a_2 = +3\cdot8879$	
$\tau = 4\cdot281$	$b_1 = -0\cdot0055153$	$b_2 = +1\cdot8652$	
	$\sigma_\tau = 0\cdot563$	$\sigma_\rho = 0\cdot211$	

V. THE FREQUENCY DISTRIBUTION OF AGES, AND THE MEAN AGE, ETC., FOR GENERA OF A GIVEN SIZE.

On the assumptions we have made the size of a genus is not an absolute *measure* of its age, for chance enters largely into the matter, but it is an *index* to age. On an average, within a homogeneous group, genera of 2 species are older than monotypes, genera of 3 species older than genera of 2 species, and so on. It is therefore of importance to determine the frequency distribution of ages for genera of a given size, *i.e.*, to determine what proportion of genera of n species, with assigned values of ρ and τ, are likely to possess any assigned age.

We will first take the limiting case when time is infinite, since the main effect of limiting time is simply to truncate the frequency distribution of the ages of the derived genera at age τ. The required expressions for genera of 1 and 2 species have, in fact, already been given under the integrals of equations (9) and (11) at the beginning of Section II, p. 37. Thus for monotypic genera the distribution of ages is given by

$$y = y_0 \, e^{-(g+s)x},$$

where y_0 must be assigned the value that will make the total unity, so that

$$y = (g+s) \, e^{-(g+s)x}.$$

Similarly for genera of 2 species

$$y = (g + s)\,(g + 2s)\,s^{-1}\,e^{-(g+s)x}\,(1 - e^{-sx}),$$

and so on. Generally, for genera of n species, the proportion y of genera of each age x, time being infinite, is given by

$$\left.\begin{aligned} y &= g f_n^{-1}\,e^{-(g+s)x}\,(1 - e^{-sx})^{n-1} \\ &= \alpha\rho^{-1} f_n^{-1}\,e^{-\alpha(1+\rho^{-1})\xi}\,(1 - e^{-\alpha\xi})^{n-1} \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots \quad (30)$$

where f_n is the nth term of the ρ-series (12), e^{α} as before is 2, and ξ is the age measured in terms of the doubling-period for species within the genus.

The mean age of genera of n species, in terms of the same relative unit of time, is (time infinite)

$$_{\infty}\bar{\xi}_n = \frac{\rho}{\alpha}\left\{\frac{1}{1+\rho} + \frac{1}{1+2\rho} + \ldots + \frac{1}{1+n\rho}\right\}. \quad \ldots\ldots \quad (31)$$

As regards the most frequent or modal age, when $n = 1$, (30) reduces to a simple exponential, and consequently the most probable age of monotypes is zero. When n is 2 or more, the modal age is (common logarithms)

$$_{\infty}\xi^{0}_{n} = (\log 2)^{-1}\{\log(1 + n\rho) - \log(1 + \rho)\} \quad \ldots\ldots \quad (32)$$

the modal frequency being

$$\frac{\alpha}{\rho f_n}\left(\frac{1+\rho}{1+n\rho}\right)^{1+\rho^{-1}}\left(\frac{(n-1)\,\rho}{1+n\rho}\right)^{n-1}.$$

For the standard deviation of ages of genera of n species, I find

$$_{\infty}\sigma^{2}_{\xi n} = \frac{\rho^2}{\alpha^2}\left\{\frac{1}{(1+\rho)^2} + \frac{1}{(1+2\rho)^2} + \ldots + \frac{1}{(1+n\rho)^2}\right\} \quad \ldots\ldots \quad (33)$$

From (31) and (33) it will be seen that for a given size of genus the mean age and the standard deviation of ages (measured in terms of our relative unit of time) are both the larger, the greater ρ; and the modal age naturally increases with the mean age.

As an illustration of the general character of the limiting frequency distributions of age and of the variation in the limit values of means, modes and standard deviations with size of genus, fig. 10 has been drawn and the data of Table IX calculated with $\rho = 1\cdot5$ (nearly the mean of the four values 1·925, 1·188, 1·253, 1·496 found for Tables V to VIII).

Fig. 10 shows the age distributions for genera of 1, 2, 3, . . . 10 species. For monotypic genera the curve is, as already pointed out, a simple exponential, the most probable age being zero. The mean age is 0·87 of a doubling-period (Table IX), and the standard deviation is 0·87 also; the relative frequencies of ages 0, 1, 2 and 3 units are 1·155, 0·364, 0·115, and 0·036.

Genera of 2 species show quite a different form of distribution. It rises abruptly, with finite slope, from a frequency zero at age 0, rises to a maximum at age 0·68 (Table IX), and then tails away slowly. The mean age is 1·41, considerably in excess of the mode.

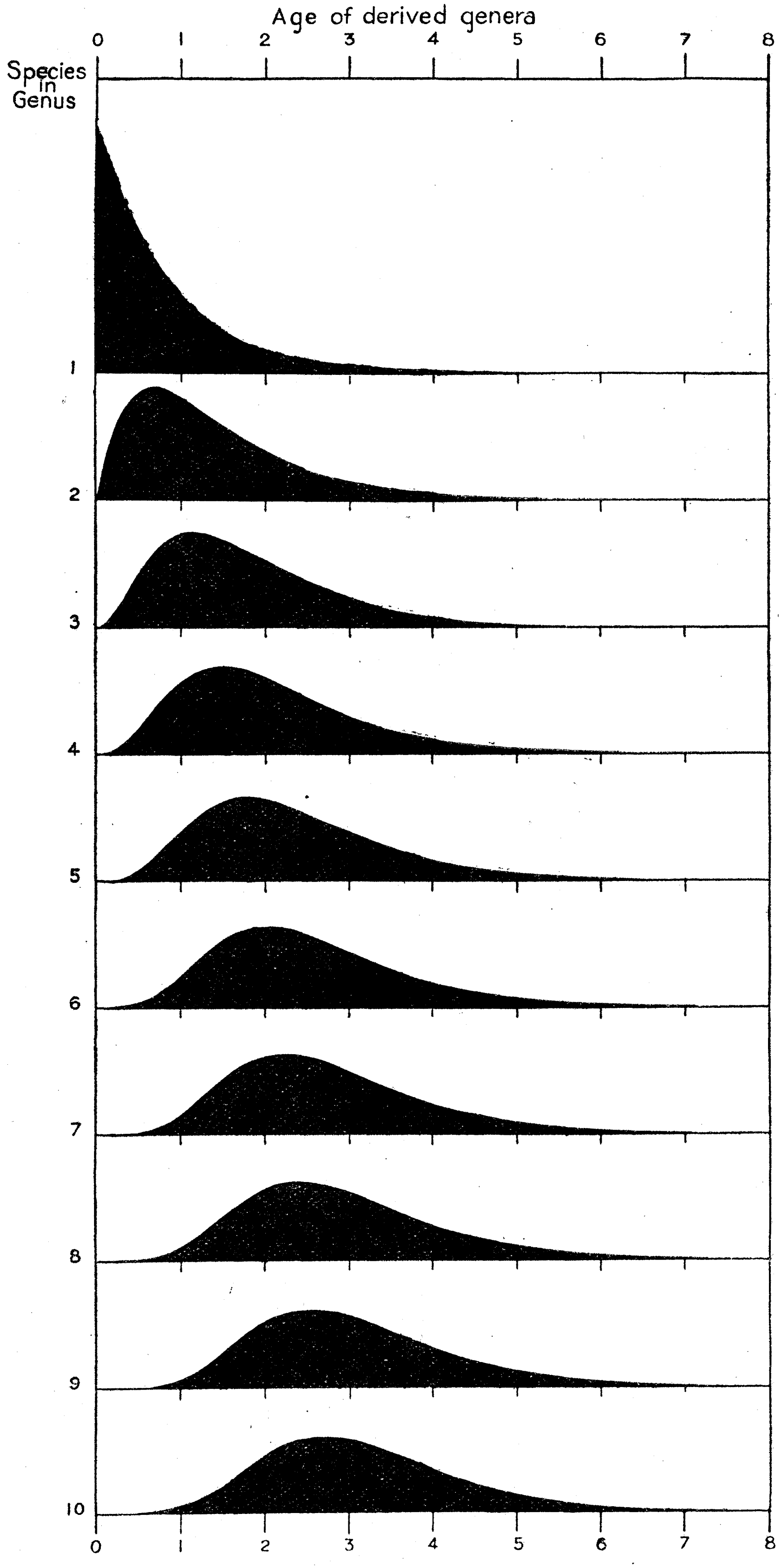

Fig. 10.—Limiting forms of the frequency distributions of age for genera of 1, 2, 3, . . . 10 species when time is infinite ; $\rho = 1 \cdot 5$ (*cf.* Table IX).

TABLE IX.—Limit values (time infinite) of mean and modal ages and standard deviations for genera of 1 to 100 species: $\rho = 1 \cdot 5$. Unit of time, the doubling-period for species within the genus.

1 Number of species in the genus.	2 Mean age.	3 Most probable (modal) age.	4 Standard deviation of ages.
1	0·87	0	0·87
2	1·41	0·68	1·02
3	1·80	1·14	1·09
4	2·11	1·49	1·14
5	2·36	1·77	1·16
6	2·58	2·00	1·18
7	2·77	2·20	1·20
8	2·93	2·38	1·21
9	3·08	2·54	1·22
10	3·22	2·68	1·23
20	4·14	3·63	1·26
30	4·70	4·20	1·28
40	5·10	4·61	1·28
50	5·42	4·93	1·29
60	5·67	5·19	1·29
70	5·89	5·41	1·29
80	6·08	5·60	1·29
90	6·25	5·77	1·29
100	6·40	5·92	1·29

For genera of 3 species the distribution is tangential to the base at the start, the frequency is a maximum for the modal age 1·14, and the mean is 1·80. From genera of 3 species onwards the general form of the distribution remains the same, but mean and mode increase with the size of the genus, at first fairly rapidly and them more slowly. The nature of the variation is well shown by Table IX, which is carried up to genera of 100 species.

When the number of species in the genus n is moderately large, say over 20, we have to a high degree of approximation—

$$\rho\alpha^{-1} \sum_{n_1}^{n_2} (1 + n\rho)^{-1} \sim \rho\alpha^{-1} \int_{n_1 - \frac{1}{2}}^{n_1 + \frac{1}{2}} (1 + \rho z)^{-1}\, dz$$
$$\sim (\log 2)^{-1} \{\log [1 + (n_2 + \tfrac{1}{2}) \rho] - \log [1 + (n_1 - \tfrac{1}{2}) \rho]\} \quad . \; . \quad (34)$$

an expression which was used for calculating the means for genera of 40 species and over in Table IX. Using the true mean by (31) at 10 species, and thence calculating the mean age for genera of 20 species by (34), the error is only 3 in the fourth decimal place. It follows then as n becomes large the mean age varies nearly as log n; the mean age of genera of 100 species (6·40), for example, is very nearly double the age of genera of 10 species (3·22), but slightly less.

The rate of increase of the standard deviation also falls off rapidly as n increases. To carry the figures to another place of decimals beyond those given in the table, it is 1·286 for genera of 50 species and no more than 1·294 for genera of 100 species. Here again the similar approximation, obtained by substituting integration for summation, is useful:—

$$\rho^2\alpha^{-2}\sum_{n_1}^{n_2}(1+n\rho)^{-2} \sim \rho\alpha^{-2}\left\{\left[1+(n_1-\tfrac{1}{2})\rho\right]^{-1}-\left[1+(n_2+\tfrac{1}{2})\rho\right]^{-1}\right\} \quad . \; . \; . \quad (35)$$

It can be safely used, with an error in the sixth place of decimals only, for genera of 30 species onwards.

All the above are limiting results, true only when the time elapsed from the genesis of the primordial species is infinite, and proceedings have continued ever since without break or disturbance. All that can be said by way of comment at the present stage is that they do not look unreasonable. It may be particularly emphasised that the size of the genus is *by no means* a precise measure of its age.

When time is not infinite but limited, the genera of any given size fall into two distinct groups: (1) the primordial genera, all of age τ; (2) the derived genera, the frequency distribution of which is of the form (30) but truncated at age $\xi=\tau$.

As we saw at the beginning of Section III, the proportion of primordial genera of n species to all genera *of every size* is

$$e^{-(g+s)\mathrm{T}}\left(1-e^{-s\mathrm{T}}\right)^{n-1}$$

or in terms of the relative notation

$$e^{-\alpha(1+\rho^{-1})\tau}\left(1-e^{-\alpha\tau}\right)^{n-1}.$$

If, then, q be the proportion of primordial genera of n species to all genera *of n species*

$$q={}_{\tau}f_n^{-1}\,e^{-\alpha(1+\rho^{-1})\tau}\left(1-e^{-\alpha\tau}\right)^{n-1} \quad . \; . \; . \; . \; . \; . \; . \; . \; . \quad (36)$$

where ${}_{\tau}f_n$ is the proportional frequency of genera of n species at time τ (equation 20). The proportion of derived genera is then given by

$$p=1-q \quad . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \quad (37)$$

and if ${}_{\tau}\mathrm{M}_{dn}$ is the mean age of the derived genera of n species, ${}_{\tau}\mathrm{M}_n$ the mean age of all genera of n species

$${}_{\tau}\mathrm{M}_n=p\,{}_{\tau}\mathrm{M}_{dn}+q\tau \quad . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \quad (38)$$

Equation (30) leads to very complex expressions for ${}_{\tau}\mathrm{M}_{dn}$ by integration, and as in any case it is desirable to have the frequency distributions, it seemed to me simpler to calculate the average directly (*i.e.*, arithmetically) from the distribution, which is very rapidly done on the machine and sufficiently precise except for the monotypic genera (*see* below).

The procedure adopted to calculate the mean ages of genera of each size for the four illustrations of the last section was accordingly as follows.

(1) The proportion of primordial genera q was found by (36) for genera of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, . . . species, and thence p by (37).

(2) The age distributions

$$e^{-\alpha(1+\rho^{-1})\xi}(1-e^{-\alpha\xi})^{n-1}$$

were calculated for the same genera, using values of ξ 0·125, 0·375, 0·625, 0·875, 1·25, 1·75, 2·25, 2·75, and so on, and the value for the centre of the final group in the tail. *E.g.*, for the *Chrysomelidæ,* where $\tau = 6 \cdot 28$, the final value of ξ is 6·14, the centre of the interval 6·00 to 6·28.

(3) The mean age of the derived genera $_{\tau}\mathrm{M}_{dn}$ was then calculated from these distributions on the machine.

(4) The mean age of *all* genera of n species was obtained by (38).

(5) In order to plot diagrams of the frequency distributions, the ordinates calculated under (2) were multiplied by the factor necessary to give a total area p.

For the monotypic genera the mean age of the derived genera found by the above process is not very accurate and integration gives a simple result. The area from 0 to τ of the curve

$$y = e^{-\alpha(1+\rho^{-1})\xi}$$

is

$$\alpha^{-1}(1+\rho^{-1})^{-1}[1-e^{-\alpha(1+\rho^{-1})\tau}]$$

which gives the factor necessary for the reduction under (5), and the mean is

$$\alpha^{-1}(1+\rho^{-1})^{-1} - \tau e^{-\alpha(1+\rho^{-1})\tau}[1-e^{-\alpha(1+\rho^{-1})\tau}]^{-1}.$$

Table X summarises the results for the *Chrysomelidæ,* the averages, etc., being carried up to genera of 100 species—a limit of size exceeded by 16 only out of the 627 genera. Turning first to columns 5 and 6, it will be seen that as the size of genus is increased the proportion of primordial genera also increases, at first with increasing and then with decreasing rapidity. The larger a genus, the older is it likely to be, and as the primordial genera are the oldest of all they will be *most likely* to be caught in the net by picking out the larger genera. Of the monotypic genera only 0·4 per cent. (roundly) are primordial, *i.e.*, the odds are about 250 to 1 against a monotypic genus being primordial. Of the genera of 10 species only 8·4 per cent. are primordial, or the odds are about 11 to 1 against a genus of 10 species being primordial. For a genus of 60 species it is about an even chance whether the genus is derived or primordial, and for a genus of 100 species the odds are nearly 2 to 1 that it is primordial.

The mean ages of the derived genera are given in column 4, and are necessarily less than the limiting values of the means, for infinite time, given for the sake of comparison in column 2. The modal ages of the derived genera are given in column 3: it will be seen that even for genera of 100 species the mode falls below the limit 6·28, so that all the frequency distributions of age, up to genera of 100 species, show a true modal age for the derived genera.

From columns 4, 5 and 6 the mean age of derived and primordial genera together was then calculated by (38), the values obtained being given in column 7. It will be seen that, for the small genera, the effect of limiting the time is slightly to increase the average age as compared with the limiting value in column 2: the difference increases up to genera of 20 species or thereabouts and then falls rather abruptly. For a genus of somewhere near 45 species the two averages are equal, and for larger genera the mean age at time τ steadily falls more and more below its limiting value, asymptoting, of course, to τ, that is 6·28 in the present case.

TABLE X.—*Chrysomelidæ.* Table showing mean ages, etc., for genera of each size: unit of time, the doubling-period for species within the genus in the Chrysomelidæ. $\tau = 6\cdot28$, $\rho = 1\cdot925$.

1	2	3	4	5	6	7
Number of species in genus.	Limit mean age.	Modal or most frequent age.	Mean age of derived genera.	Proportion of		Mean age of derived and primordial.
				Derived genera.	Primordial genera.	
1	0·95	0	0·94	0·9961	0·0039	0·96
2	1·52	0·73	1·50	0·9903	0·0097	1·55
3	1·93	1·21	1·90	0·9832	0·0168	1·98
4	2·25	1·57	2·21	0·9752	0·0248	2·31
5	2·51	1·86	2·45	0·9664	0·0336	2·58
6	2·73	2·10	2·66	0·9570	0·0430	2·82
7	2·93	2·31	2·84	0·9472	0·0528	3·02
8	3·10	2·49	2·99	0·9370	0·0630	3·20
9	3·25	2·65	3·13	0·9265	0·0735	3·36
10	3·38	2·79	3·26	0·9158	0·0842	3·51
20	4·32	3·76	4·05	0·8064	0·1936	4·48
30	4·88	4·33	4·47	0·7071	0·2929	5·00
40	5·28	4·74	4·75	0·6231	0·3769	5·33
50	5·59	5·06	4·95	0·5534	0·4466	5·54
60	5·85	5·32	5·10	0·4955	0·5045	5·69
70	6·07	5·54	5·22	0·4474	0·5526	5·80
80	6·26	5·73	5·31	0·4070	0·5930	5·89
90	6·43*	5·90	5·39	0·3735	0·6265	5·95
100	6·58*	6·05	5·46	0·3451	0·6549	6·00

* Beyond the limiting value of the age, 6·28.

Fig. 11 shows as illustrations the age distributions of genera of 1, 2, 3, 10, 60 and 100 species. The distributions of the derived genera are of the same general form as the distributions of fig. 10, but slightly altered owing to the different value of ρ, and truncated at $\tau = 6\cdot28$; the area of the truncated curve is also made equal to p and not unity. The proportion of the primordial genera is shown by a square just to the right of the limit, the area of this square being q, so that the square and the curve make up unit area when taken together. For genera of 1, 2 and 3 species the truncation is scarcely perceptible and the square is very small. For genera of 10 species the truncation becomes appreciable and the square is larger. For genera of 60 species the area of the square is little more than that of the curve, and finally for genera of 100 species most of

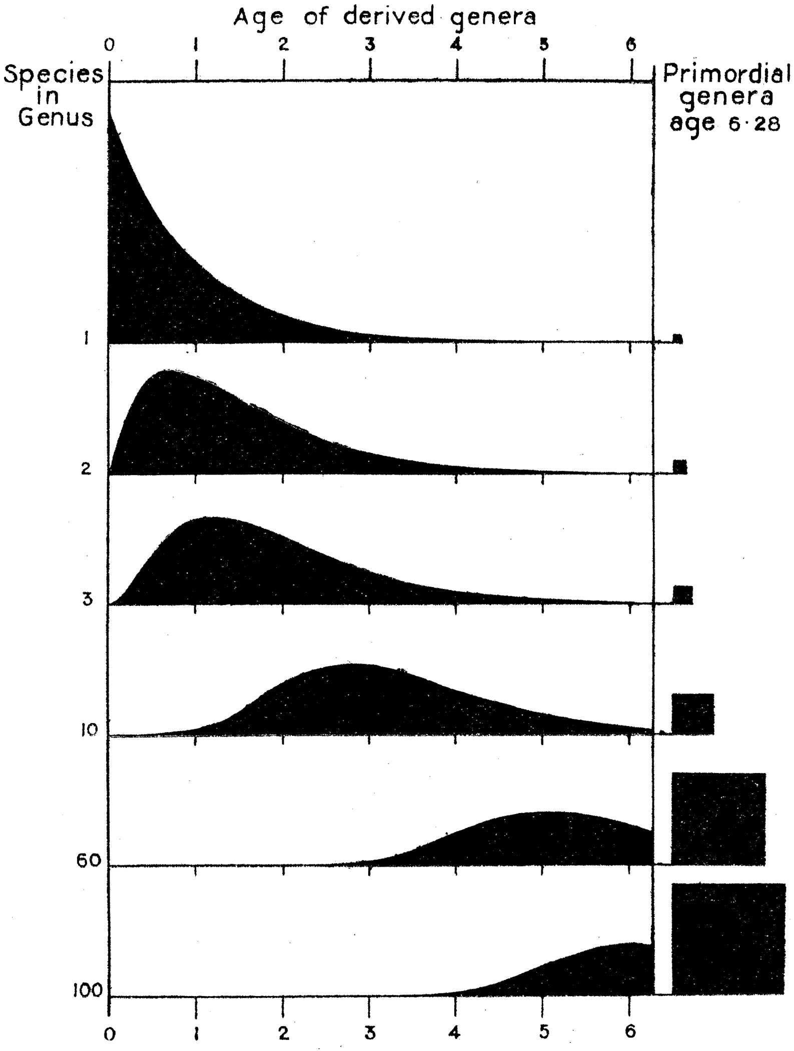

Fig. 11.—*Chrysomelidæ*. Frequency distributions of age for genera of 1, 2, 3, 10, 60 and 100 species. The curves give the age distributions of the derived genera, the squares on the right showing the proportional frequency of primordial genera (*cf.* Table X).

the derived distribution is cut off, the modal age lies just to the left of the limit and the square has nearly twice the area of the small portion of the curve that is left. The odds are nearly 2 to 1 that a genus of 100 species or more is one of the primordial genera (or genera reckoned as such, *cf.* the discussion at the end of Section IV). As a matter of fact the number of genera with 100 species or more in the *Chrysomelidæ* is only 16 (Table V) and we have to class some 65 genera as primordial (Table IV).

Tables XI, XII and XIII show data arranged in the same form for the *Cerambycinæ*, the Snakes and the Lizards, figs. 12, 13 and 14 showing the age distributions for a few selected sizes of genus in the same way. The tables and charts will repay some study but do not seem to call for special comment. It may once more be emphasised that the unit of time is different in each case if measured in years: it is the doubling-period for species within the genus *in the given group.*

TABLE XI.—*Cerambycinæ.* Table showing mean ages, etc., for genera of each size; unit of time, the doubling-period for species within the genus in the *Cerambycinæ.* $\tau = 4{\cdot}98$, $\rho = 1{\cdot}188$.

1	2	3	4	5	6	7
Number of species of genus.	Limit mean age.	Modal or most frequent age.	Mean age of derived genera.	Proportion of		Mean age of derived and primordial.
				Derived genera.	Primordial genera.	
1	0·78	0	0·77	0·9962	0·0038	0·79
2	1·29	0·63	1·27	0·9896	0·0104	1·30
3	1·67	1·06	1·63	0·9808	0·0192	1·70
4	1·96	1·39	1·91	0·9701	0·0299	2·00
5	2·21	1·67	2·14	0·9580	0·0420	2·26
6	2·42	1·89	2·32	0·9448	0·0552	2·47
7	2·61	2·09	2·49	0·9306	0·0694	2·66
8	2·77	2·26	2·63	0·9157	0·0843	2·83
9	2·92	2·42	2·75	0·9004	0·0996	2·97
10	3·05	2·56	2·86	0·8846	0·1154	3·11
20	3·96	3·50	3·55	0·7290	0·2710	3·94
30	4·52	4·07	3·85	0·6012	0·3988	4·30
40	4·91	4·47	4·04	0·5046	0·4954	4·51
50	5·23*	4·79	4·17	0·4310	0·5690	4·63
60	5·48*	5·05*	4·25	0·3764	0·6236	4·71

* Beyond the limiting value of the age, 4·98.

Charts such as those in figs. 11 to 14 suggest the question whether an examination of the larger genera in some group might not be of interest: the age distribution of such genera ought to be markedly compound, the genera being in part primordial and in part derived. Is it beyond the bounds of possibility that one familiar with the group might be able to effect at least a tentative separation of the two?

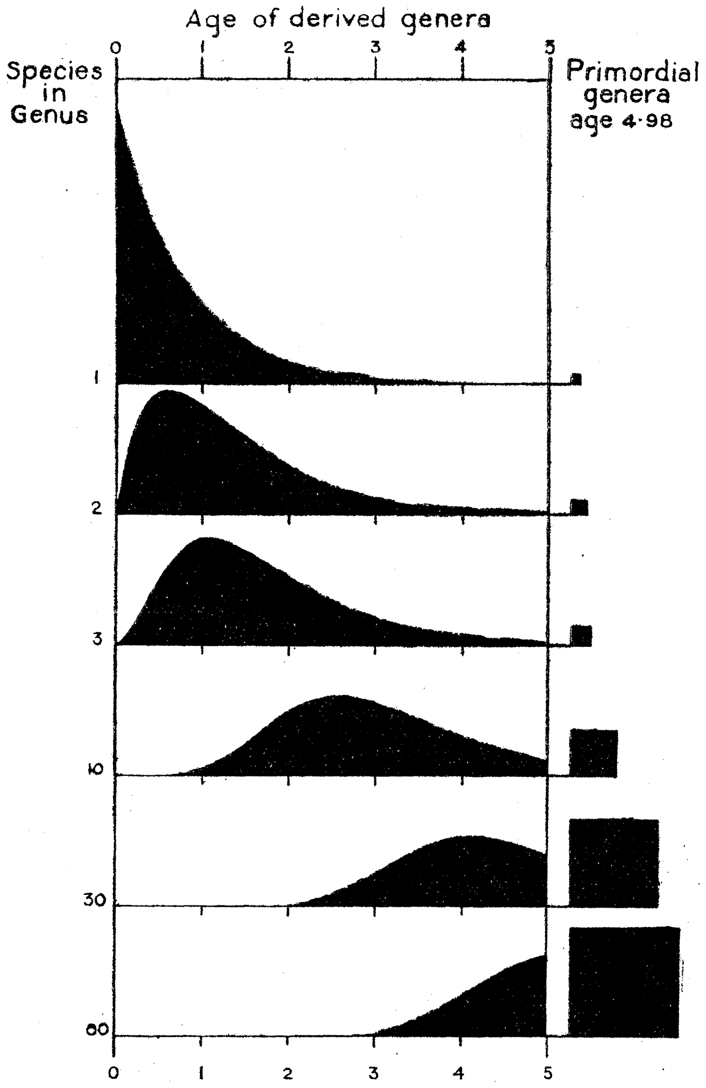

Fig. 12.—*Cerambycinæ*. Frequency distributions of age for genera of 1, 2, 3, 10, 30 and 60 species (*cf.* Table XI).

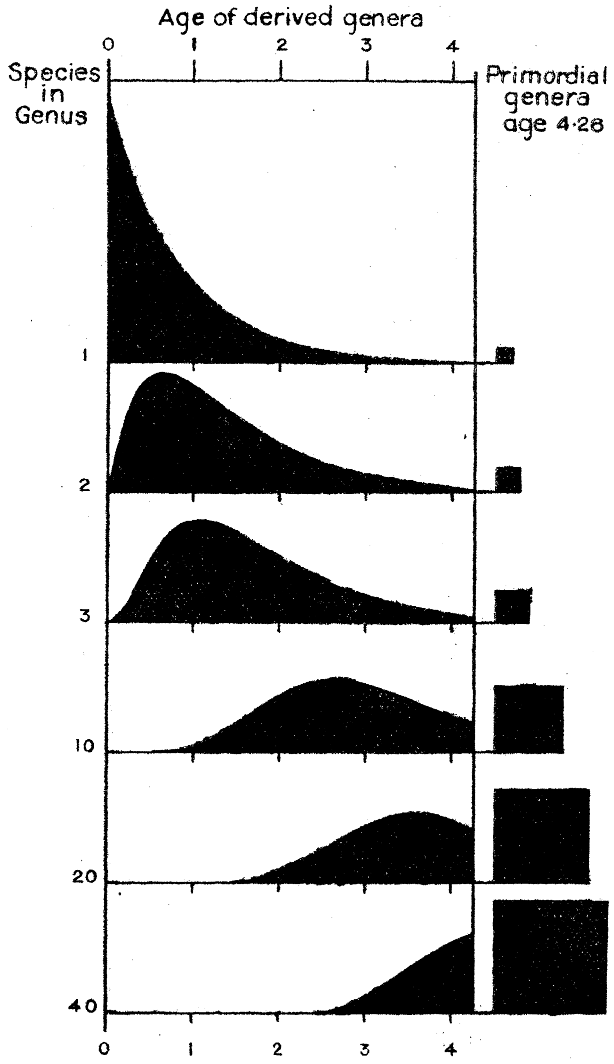

Fig. 13.—*Snakes*. Frequency distributions of age for genera of 1, 2, 3, 10, 20 and 40 species (*cf.* Table XII).

TABLE XII.—Snakes. Table showing mean ages, etc., for genera of each size: unit of time, the doubling-period for species within the genus in the Snakes. $\tau = 4{\cdot}26$, $\rho = 1{\cdot}253$.

1	2	3	4	5	6	7
Number of species in genus.	Limit mean age.	Modal or most frequent age.	Mean age of derived genera.	Proportion of		Mean age of derived and primordial.
				Derived genera.	Primordial genera.	
1	0·80	0	0·78	0·9889	0·0111	0·82
2	1·32	0·64	1·26	0·9709	0·0291	1·35
3	1·70	1·08	1·62	0·9483	0·0517	1·75
4	2·00	1·42	1·88	0·9227	0·0773	2·06
5	2·25	1·69	2·09	0·8953	0·1047	2·31
6	2·46	1·92	2·26	0·8670	0·1330	2·52
7	2·64	2·12	2·40	0·8384	0·1616	2·70
8	2·81	2·29	2·53	0·8100	0·1900	2·86
9	2·96	2·45	2·64	0·7823	0·2177	2·99
10	3·09	2·59	2·73	0·7552	0·2448	3·11
20	4·00	3·53	3·29	0·5403	0·4597	3·73
30	4·56*	4·10	3·54	0·4086	0·5914	3·97
40	4·96*	4·50*	3·69	0·3253	0·6747	4·07

* Beyond the limiting value of the age, 4·26.

TABLE XIII.—Lizards. Table showing mean ages, etc., for genera of each size: unit of time, the doubling-period for species within the genus in the Lizards. $\tau = 4{\cdot}281$, $\rho = 1{\cdot}496$.

1	2	3	4	5	6	7
Number of species in genus.	Limit mean age.	Modal or most frequent age.	Mean age of derived genera.	Proportion of		Mean age of derived and primordial.
				Derived genera.	Primordial genera.	
1	0·86	0	0·83	0·9825	0·0175	0·89
2	1·41	0·68	1·33	0·9564	0·0436	1·46
3	1·80	1·14	1·69	0·9256	0·0744	1·88
4	2·11	1·48	1·95	0·8923	0·1077	2·20
5	2·36	1·76	2·16	0·8580	0·1420	2·46
6	2·58	2·00	2·33	0·8238	0·1762	2·67
7	2·77	2·20	2·47	0·7904	0·2096	2·85
8	2·93	2·38	2·59	0·7581	0·2419	3·00
9	3·08	2·53	2·70	0·7270	0·2730	3·13
10	3·22	2·68	2·79	0·6976	0·3024	3·24
20	4·14	3·63	3·33	0·4796	0·5204	3·82
30	4·53*	4·20	3·57	0·3566	0·6434	4·03
40	5·10*	4·61*	3·71	0·2814	0·7186	4·12

* Beyond the limiting value of the age, 4·28.

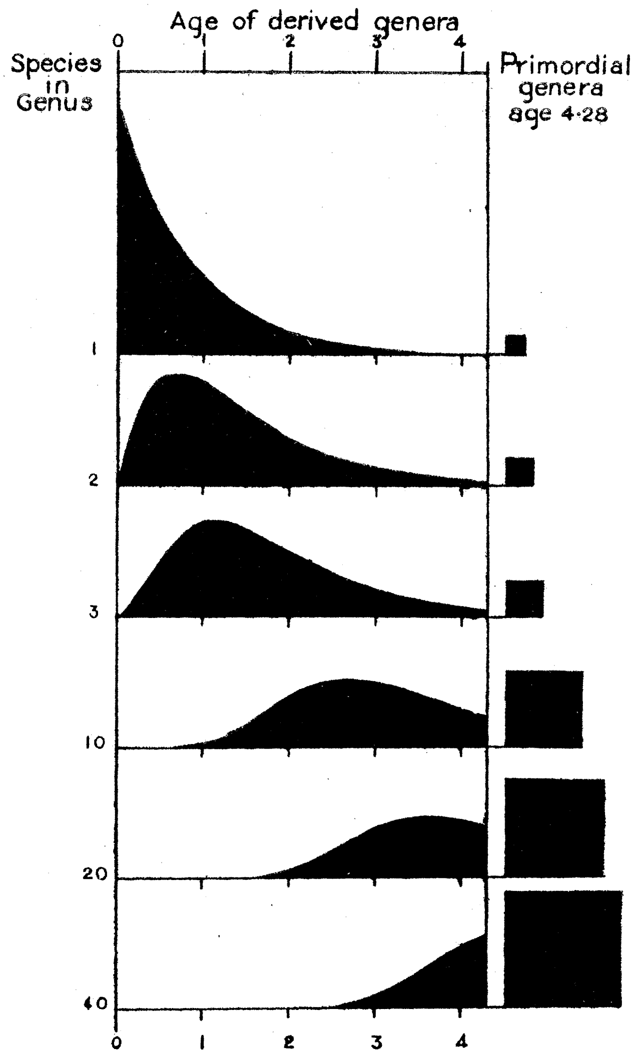

Fig. 14.—*Lizards*. Frequency distributions of age for genera of 1, 2, 3, 10, 20 and 40 species (*cf.* Table XIII).

VI. An Attempt to Estimate the Order of Magnitude of the Doubling-Period for Species in the case of the Flowering Plants, and the Present Rate of Occurrence of Specific Mutations.

Objection has been raised to the assumption of " specific mutations " on the ground that no such phenomena have been observed. " Though undeniable as possibilities "— Dr. Bateson wrote in a review of ' Age and Area ' ('*Nature*,' January 13th, 1923)— " we have to consider what warrant for such guesses " (as *Coleus elongatus* being the immediate parent of *C. barbatus*, and so forth) " can be drawn from the observed facts of variation. The answer is quite clear that up to the present scarcely anything comparable has been observed." To such an argument the reply seems to me to be that we are not likely to observe the actual occurrence of a viable specific mutation—it is far too rare an event—unless or until we discover how to stimulate such mutations artificially. This is in fact the answer already given by Dr. WILLIS ('Age and Area,' p. 212): "Lord RAYLEIGH has estimated the period since the Eocene alone, which covers but a portion of that occupied in the evolution of the higher plants, at 30,000,000 years. But if we suppose one mutation in 50 years to survive, we should get the whole of the existing 160,000 species of flowering plants in 8,000,000 years, which is only 26 per cent. of that time."

Let us look at the matter more closely. In the above illustration Dr. WILLIS has taken the mutations as occurring uniformly throughout the period of evolution, so that the total number of species would increase in arithmetic progression. But we concluded that the total number of species increases, not in arithmetic progression, but according to a law which gradually approximates to geometric progression (equations 21 and 24, p. 50).

Let us suppose that for the present very rough calculations it will suffice to take the number of species y as given by

$$y = \mathrm{A}\, e^{at} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (39)$$

where t is the time in years. Let Y be the known number of species at the present time T. Then

$$a = (\mathrm{T} \log e)^{-1} (\log \mathrm{Y} - \log \mathrm{A}) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (40)$$

Further, if λ be the doubling period $e^{a\lambda} = 2$ or

$$\lambda = (\log 2)\,(a \log e)^{-1} = \mathrm{T} \log 2\, (\log \mathrm{Y} - \log \mathrm{A})^{-1} \quad . \quad . \quad . \quad . \quad . \quad (41)$$

For the present rate of occurrence of specific mutations (viable specific mutations) we have

$$\left(\frac{dy}{dt}\right)_{\mathrm{T}} = a\mathrm{Y} = (\mathrm{T} \log e)^{-1}\, \mathrm{Y}\, (\log \mathrm{Y} - \log \mathrm{A}). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (42)$$

Ignoring for the moment the killing out of species, let us see what values (41) and (42) give us for λ and for dy/dt at the present time: the value so obtained for λ will be an upper limit, and for dy/dt a lower limit. For the time that has elapsed since the origin of the flowering plants I propose to use the round figure of 100,000,000 years. Lord RAYLEIGH's figure of 30 million years since the Eocene, used by Dr. WILLIS, is based on the helium ratio and must be regarded as a lower limit for the time that has elapsed since that epoch. Where the helium ratio gives 146 million years for the age of the Carboniferous, the lead ratio gives 340 millions.* In the absence of a definite figure for the Lower Cretaceous, 100 million years may probably be taken as sufficiently near the truth for a calculation in which we are really only concerned with the order of magnitude of the result: the very roundness of the figure will remind us that no precision is implied. The constant A (*cf.* equation 24) is $\rho/(\rho-1)$: as ρ is apparently about $1 \cdot 5$ (p. 62) we may place its value at 3. Y, as mentioned above, is taken by Dr. WILLIS as roundly 160,000.

---

* ARTHUR HOLMES, 'The Age of the Earth' (Harpers, 1913).

With these data, we have for the doubling-period

$$\lambda = \frac{10^8 \times 0 \cdot 301}{5 \cdot 204 - 0 \cdot 477} = 6 \cdot 4 \times 10^6$$

and for the present rate of occurrence of specific mutations

$$\left(\frac{dy}{dt}\right)_{\mathrm{T}} = \frac{16 \times 10^4 \, (5 \cdot 204 - 0 \cdot 477)}{10^8 \times 0 \cdot 4343} = 0 \cdot 0174 = 1/57 \cdot 5.$$

The doubling-period (either for species within the genus or for all species at the present time) is then, on these assumptions, no less than 6·4 million years, and the present occurrence of viable specific mutations at the rate of 1 in some 57 or 58 years—amongst all species of flowering plants on the whole surface of the globe. Any alteration in ρ, it may be noted, makes a relatively small change in these results. If $\rho = 2$ instead of 1·5, the doubling-period is lowered to 6·1 million years and the rate of occurrence of viable specific mutations raised to 1 in some 55 years. Raising ρ to infinity only lowers λ to 5·8 millions and raises dy/dt to 1 in some 52 years. If, on the other hand, ρ is *lowered* to 1·2, the doubling-period is raised to 6·8 million years and the rate of occurrence of mutations lowered to 1 in some 61 years. The order of magnitude of the result is not affected at all.

A doubling-period of 6·4 million years would give between 15 and 16 doubling-periods in the 100 million years taken as having elapsed since the genesis of the flowering plants. A rough estimate of the length of the doubling-period might have been given off-hand by anyone who kept in front of him a table of the powers of 2 (as in column 2 of Table XIV)—a useful thing to do when considering questions of this kind. On this table 160,000 lies between the 17th and 18th powers of 2 : so that if there had been simple geometric increase in the number of species (corresponding in strictness to $\rho = \infty$) the doubling-period would have been between 100/17 and 100/18 or 5·9 and 5·5 million years and nearer to the former than the latter (5·8 millions, as stated above). The only effect of the more complex law of equation (24), with ρ taken as 1·5 is, as shown by column 3 of Table XIV, to reduce the number of doubling-periods necessary to give 160,000 species from 17 or 18 to 15 or 16 and proportionately increase the length of the doubling-period.

If the flowering plants had a polyphyletic origin, starting from 2, 4, or 8 viable mutations which occurred (geologically speaking) at about the same time, the only consequence would be further to reduce the number of doubling-periods necessary to give the required total of 160,000 species from 15 or 16 to 12, 13 or 14, and further to *increase* the estimated length of the doubling-period to possibly as much as 8 or 9 million years.

In all this argument, however, the killing-out of species is ignored. When species are being killed-out, either more or less continuously during the whole lapse of time or more or less cataclysmically at intervals, it at once becomes doubtful how far equation (24) applies. I propose, therefore, to take it that, for present purposes, it suffices to

TABLE XIV.

Time τ in doubling-periods.	Powers of 2.	${}_sN_\tau$ by equation (24) : $\rho = 1\cdot5$.
1	2	3
0	1	1
1	2	3
2	4	7
3	8	16
4	16	35
5	32	76
6	64	160
7	128	333
8	256	687
9	512	1,408
10	1,024	2,869
11	2,048	5,821
12	4,096	11,776
13	8,192	23,763
14	16,384	47,862
15	32,768	96,256
16	65,536	193,357
17	131,072	388,055
18	262,144	778,240
19	524,288	1,559,860
20	1,048,576	3,125,085

assume the law of free increase as simply logarithmic from the beginning, and write A in (39) as unity. As we have just seen, the only effect of such an assumption in the above case is to *reduce* the estimated doubling-period from 6·4 to 5·8 million years.

I will first suppose the killing-out to be practically continuous, though as stated in the Introduction I do not think this at all closely represents the facts. Let the numbers at successive small intervals of time be

$$1, \quad pr, \quad p^2r^2, \quad p^3r^3, \; . \; . \; . \; p^nr^n$$

where p is the proportion of survivors and r gives the free rate of increase. If

$$p = 1 - \delta\theta$$

$$r = 1 + a\theta$$

in the limit when the time-interval θ is made very small we may write

$$y = e^{(a-\delta)t}. \qquad (43)$$

The total number of deaths from the beginning to the time of observation is

$$\begin{aligned} \mathrm{D} &= qr + qpr^2 + qp^2r^3 + \; . \; . \; . + qp^{n-1}r^n \\ &= qr\,(p^nr^n - 1)\,(pr - 1)^{-1} \\ &= qr\,(pr - 1)^{-1}\,(\mathrm{Y} - 1) \end{aligned}$$

That is approximately, Y being large compared with unity and a and δ very small,

$$\mathrm{D} = \delta\,(a - \delta)^{-1}\,\mathrm{Y}.$$

Let

$$a/\delta = k. \qquad (44)$$

Then finally

$$\mathrm{D} = (k - 1)^{-1}\,\mathrm{Y}. \qquad (45)$$

If, for example, $k = 2$, that is, if the number of species killed-out in a small interval of time is half the number of new species coming into existence in that time, the total number of species that has been killed-out is equal to the number at present existing. We have now

$$a = k \log \mathrm{Y}\,[\,(k - 1)\,\mathrm{T} \log e]^{-1} \qquad (46)$$

and the free doubling-period, the period in which the number of species would double apart from the killing-out, is given by

$$\lambda = \log 2\,(a \log e)^{-1} = (k - 1)\,\mathrm{T} \log 2\,(k \log \mathrm{Y})^{-1}. \qquad (47)$$

Finally

$$\left(\frac{dy}{dt}\right)_{\mathrm{T}} = a\,(k - 1)\,k^{-1}\,\mathrm{Y} = \mathrm{Y} \log \mathrm{Y}\,(\mathrm{T} \log e)^{-1}.$$

But this consists of two parts :—

Rate of occurrence of new species :—

$$\left.\begin{aligned} {}_1\!\left(\frac{dy}{dt}\right)_{\mathrm{T}} &= a\mathrm{Y} = k\,(k - 1)^{-1}\,\mathrm{Y} \log \mathrm{Y}\,(\mathrm{T} \log e)^{-1} \qquad (a) \\ {}_2\!\left(\frac{dy}{dt}\right)_{\mathrm{T}} &= -\frac{a}{k}\,\mathrm{Y} = -(k - 1)^{-1}\,\mathrm{Y} \log \mathrm{Y}\,(\mathrm{T} \log e)^{-1} \ (b) \end{aligned}\right\} \qquad (48)$$

Rate of killing-out of species :— (the second line above, (b))

Were we in a position to make even a rough estimate of the number of species of flowering plants that has been killed-out, (45) would give us an approximate value for k. But unfortunately we are not at present in a position to give even a lower limit for this figure, which would give an upper limit for k. If, merely as a numerical illustration, we take k as 2, λ takes half the previous limiting value ($\rho = \infty$) or 2·9 million years, new species come into existence at about the rate of 1 in 26 years, and species are killed out at about the rate of 1 in 52 years.

So much for the scheme of continuous killing-out. Let us now turn to a scheme of cataclysmic killing: supposing that a series of practically instantaneous cataclysms occurs, each sweeping off a certain proportion of the then existing species. However crude, it seems to me that such a scheme is a slightly closer representation of the facts ; at all events, a comparison between the consequences of such a scheme and the scheme of continuous killing will show whether extreme differences in the time-incidence of destruction on species have or have not any important effect on the estimated doubling-period.

Suppose then that each cataclysm kills off the same proportion q of the existing species, a proportion p surviving: that the cataclysms recur at regular intervals ϕ: that there are n such intervals in the time T: and that observation is made just after the nth cataclysm, *e.g.*, the last glacial epoch. During the intervals between cataclysms it will be assumed as before that increase in the number of species may be taken as approximately logarithmic. Then the changes in the number of species will take place as follows:—

Time in intervals.	Number of species.
0	1
1	$e^{a\phi} \longrightarrow pe^{a\phi}$
2	$pe^{2a\phi} \longrightarrow p^2e^{2a\phi}$
3	$p^2e^{3a\phi} \longrightarrow p^3e^{3a\phi}$

During the first interval the number of species will increase to $e^{a\phi}$, of which only $pe^{a\phi}$ survive after the first cataclysm. During the second interval these will increase to $pe^{2a\phi}$, only $p^2e^{2a\phi}$ surviving after the second cataclysm, and so on. Hence

$$\mathrm{Y} = p^n e^{na\phi} = p^n e^{a\mathrm{T}} \qquad (49)$$

Whence

$$a = (\log \mathrm{Y} - n \log p)\,(\mathrm{T} \log e)^{-1} \qquad (50)$$

where $\log p$ is, of course, essentially negative. The free doubling-period is

$$\lambda = \mathrm{T} \log 2\,(\log \mathrm{Y} - n \log p)^{-1} \qquad (51)$$

and

$$\left(\frac{dy}{dt}\right)_{\mathrm{T}} = \mathrm{Y}\,(\log \mathrm{Y} - n \log p)\,(\mathrm{T} \log e)^{-1}. \qquad (52)$$

The number of species killed-out is

$$\begin{aligned} \mathrm{D} &= qe^{a\phi}\,(p^n e^{na\phi} - 1)\,(pe^{a\phi} - 1)^{-1} \\ &= qp^{-1}\,\mathrm{Y}^{1/n}\,(\mathrm{Y} - 1)\,(\mathrm{Y}^{1/n} - 1) \end{aligned}$$

or as Y is large, very nearly

$$\mathrm{D} = qp^{-1}\,\mathrm{Y}^{(n+1)/n}\,(\mathrm{Y}^{1/n} - 1) \qquad (53)$$

Compare equations (47) and (51). In (47) T log 2 is divided by $\log \mathrm{Y}^{k/(k-1)}$, that is, by $\log (\mathrm{Y}/e^{-\delta \mathrm{T}})$. In (51) the divisor of T log 2 is $\log (\mathrm{Y}/p^n)$. But $e^{-\delta\mathrm{T}}$ in the first case and p^n in the second is the chance, say p_{T}, of a species surviving from the origin of the flowering plants to the time of observation. Hence, so long as p_{T} is the same it is of no consequence whether the killing-out is continuous or discontinuous; λ is unaltered by the changed incidence of destruction.

A comparison of equations (48*a*) and (52) shows that a similar statement holds good for the present rate of production of specific mutations. The rate is the same so long as p_{T} is the same, whether the killing-out is continuous or cataclysmic. But the number of species killed-out, it must be noted, is not the same in the two cases: (45) and (53) do not lead to the same result for the same value of p_T, (45) being the limiting value of (53) when n is made indefinitely great.

Table XV gives a conspectus of the results for various assumed values of the killing-out. In the first section of the table are given, for the sake of comparison with the remainder of the table, the values arrived at for the doubling-period and for the present rate of production of species if there is no killing-out at all. In the first line of section 2, discontinuous killing is assumed to the extent of 10 equidistant cataclysms each of

TABLE XV.—Estimates of the doubling-period for species in the flowering plants, and of the present rate of production of specific mutations on various assumptions.

Case assumed: Time elapsed since the origin of the flowering plants, 100 million years: present existing species, 160,000.	Total number D of species killed-out.	Change of a species surviving 10 million years.	Doubling-period in million years.	Present rate of production of specific mutations, 1 in years stated below.
1. (*a*) No killing-out $\rho = 1\cdot2$	Nil	Unity	6·8	61
(*b*) " " $\rho = 1\cdot5$	Nil	Unity	6·4	57·5
(*c*) " " $\rho = 2\cdot0$	Nil	Unity	6·1	55
(*d*) " " $\rho = \infty$	Nil	Unity	5·8	52
In all the following increase is taken as simply logarithmic ($\rho = \infty$)				
2. (*a*) Discontinuous killing: 10 cataclysms each killing-out $\frac{1}{3}$ of existing species	114,600	0·6667	4·3	39
(*b*) Continuous killing with same total killed-out: $k = 2\cdot396$	114,600	0·4238	3·4	30
3. (*a*) Discontinuous killing: 20 cataclysms each killing-out $\frac{1}{3}$ of existing species	177,500	0·4444	3·4	33
(*b*) Continuous killing with same total killed-out: $k = 1\cdot901$	177,500	0·2645	2·7	25
4. (*a*) Discontinuous killing: 50 cataclysms each killing-out $\frac{1}{3}$ of existing species	375,400	0·1317	2·1	19
(*b*) Continuous killing with same total killed-out: $k = 1\cdot426$	375,400	0·0600	1·7	16
5. (*a*) Discontinuous killing: 100 cataclysms each killing-out $\frac{1}{3}$ of existing species	708,400	0·0173	1·3	12
(*b*) Continuous killing with same total killed-out: $k = 1\cdot226$	708,400	0·0050	1·1	10
6. (*a*) Discontinuous killing: 100 cataclysms each killing-out $\frac{1}{4}$ of existing species	472,300	0·0562	1·7	15
(*b*) Continuous killing with same value of p_T: $k = 1\cdot416$	384,600	0·0562	1·7	15
7. (*a*) Discontinuous killing: 100 cataclysms each killing-out 0·092 of existing species ...	143,500	0·3805	3·2	29
(*b*) Continuous killing with same value of p_T: $k = 2\cdot241$	129,000	0·3805	3·2	29

which sweeps out of existence one-third of the then existent species. This would give the total number of species extinct as 114,600 : the free doubling-period would be reduced to 4·3 million years, and the present rate of production of specific mutations would be raised to 1 in some 39 years. The chance p_T of a species surviving for the whole 100 million years would in several of the cases taken be vanishingly small, hence in the second column of the table I have given for the sake of readier comparison $p_T^{0 \cdot 1}$ or the chance of a species surviving 10 million years. Since 10 million years, in the present case, only cover a single cataclysm, this is 2/3 or 0·6667. In the second line of section 2 are given the corresponding figures when the killing-out is continuous, and of such severity as to make, not p_T the same, but the total number of species killed-out the same. Equation (45) then gives k, and (47) and (48a) give the doubling-period and the present rate of production of specific mutations. It will be seen that on these assumptions λ is further reduced to 3·4 million years, the present rate of production of specific mutations is raised to 1 in some 30 years, and the chance of a species surviving for 10 million years is reduced to 0·4238. It must be remembered that under 2a we assume observation just after the last cataclysm : our 160,000 species would then be the survivors of 240,000 which existed immediately prior to the cataclysm, no less than 80,000 of the 114,600 species extinct having been killed-out in the final cataclysm. When the killing-out is continuous, the deaths are spread over the whole curve of increase, and p_T must naturally be smaller to give the same total of species extinct.

In sections 3, 4, and 5 of the table similar comparisons are made for successively increased severity of the destruction of species. In section 5, with 100 cataclysms each killing-out one-third of the then existing species, the total number of species killed-out would be over 700,000, the free doubling-period would be lowered to 1·3 million years, and the present rate of production of specific mutations would be raised to 1 in some 12 years.

But unfortunately, as already stated, there seems no basis at present for estimating even roughly the total number of extinct species, so as to estimate whereabouts on the table the truth in fact lies. That line of approach, at present at all events, is not possible. But another and more hopeful line is afforded by the second column—the chance of a species surviving. We know that the chance of a species surviving from the origin of the flowering plants to the present time must be infinitesimally small, for no species has so survived. On the other hand, we know that a species may survive through very long periods of geological time. Hence we might reasonably conjecture that p_T is too large in sections 2 and 3 of the table, and possibly too small in section 5. The conjecture seems to be confirmed by utilising some valuable data given by Mrs. CLEMENT REID in Chapter XIV of 'Age and Area' (p. 144) and reproduced below.

Taking, *e.g.*, the lowest figure showing 10 per cent. of *survivals* from the base of the Pliocene to the present time—it is founded on a single local deposit only and obviously too much weight must not be attached to it—can we use it to suggest a value for p_T?

Percentages of Extinct Species belonging to the Chinese-North American Association of Plants in the West European Pliocene at Successive Periods.

Deposit.	Age of Deposit.	Percentage of extinct species (approximate).
Cromerian	Top of Pliocene ...	0
Teglian	Upper Pliocene ...	35
Castle Eden	Middle Pliocene ...	44
Reuverian	Lower Pliocene ...	70
Pont-de-Gail	Base of Pliocene ...	90

If we can fix the approximate time to which it relates, we can readily do so. Lord RAYLEIGH'S figure for the age of the Pliocene is 2·5 million years as determined by the helium ratio, 6·3 millions for the Miocene, figures which would correspond to about 6 and 15 millions respectively on the lead-ratio scale. Taking these as approximately central figures for the Pliocene and the Miocene respectively, the age of the base of the Pliocene must be somewhere between 6 and 15 million years. Suppose we call it 8 millions—probably rather a low figure. Then we have:—

$$p_T{}^{0\cdot 08} = 0\cdot 1$$

whence

$$\log p_T = -12\cdot 5 = \overline{13}\cdot 5$$

while

$$0\cdot 1 \log p_T = \overline{2}\cdot 75 \qquad p_{10} = 0\cdot 0562$$

$$0\cdot 01 \log p_T = \overline{1}\cdot 875 \qquad p_1 = 0\cdot 75$$

If then we assume killing-out to have been effected by 100 cataclysms, the chance of a species surviving the cataclysm must be taken as 3/4, instead of 2/3 as in section 5, and this gives the values of the doubling-period and the present rate of production of species shown in section 6 of the table. The only figure altered if we keep p_T the same but assume killing-out to have been practically continuous is the number of extinct species, which is lowered from 472,000 odd to 384,000 odd.

But the data in Mrs. REID'S table, though they run consistently from the top to the bottom of her table, suggest that the more recent figures for the percentages of extinct species are too low (*cf.* the zero with which the table begins) or possibly the percentages for the older deposits too high. If 0·1 is the chance of survival from the base of the Pliocene, the chance of survival from an epoch only half as distant in time should under uniform conditions be in the neighbourhood of $\sqrt{0\cdot 1}$ or 0·32. But the percentage of survivals from the Middle Pliocene is as high as 56, and this is much more than half as distant in time. If we call the age of the Middle Pliocene 6 million years, this percentage of survivals gives the results shown in section 7 of the table—roundly nearly double the figures shown by section 6. I have purposely taken the age rather low in the first case and possibly rather high in the second, so as to give limiting results. On the general reasoning used above the value of p_T for this second case (section 7), approximately 7×10^{-5}, seems rather too high, as it gives a probability of about 1 in 14,000 for a species surviving right through from the origin of the flowering plants to the present time. On the other hand, the value of p_T for the first case (section 6 of Table XV), roundly 3×10^{-13}, is perhaps rather too low, since we know from such instances as *Ginkgo* that species may survive through very long periods of geological time.

In any case the figures are quite definite as to order of magnitude. If the age of the flowering plants is 100 million years, or thereabouts, the doubling-period for species is probably of the order of some 2 or 3 million years: it is, say, almost certainly over 1 million and less than 6 millions. The present rate of production of viable specific mutations, amongst all flowering plants on the whole surface of the globe, is almost certainly less than 1 in 10 years and more than 1 in 60 years; it probably lies between 1 in 15 and 1 in 30 years. The assumption of a polyphyletic origin for the flowering plants would not very greatly affect these figures. Specific mutations must, therefore, be such exceedingly rare events that no valid argument, as it seems to me, can be based on the fact that we have no experience of such occurrences. My calculations fully confirm Dr. WILLIS's conclusion in this respect.

It should perhaps be added that of course the figures obtained, for a group so heterogeneous as the aggregate of the flowering plants, are of the nature of averages and may well differ considerably for different families and genera. But it does not seem probable that the order of magnitude is wholly different.

APPENDIX.

TABLE A.—*Chrysomelidæ*: Numbers of genera with 1, 2, 3, . . . species. (Compiled by Dr. J. C. WILLIS from Cat. Coleopt., Gemminger and Harold, t. XI, 1874, and t. XII, 1876.)

Species.	Genera.	Species.	Genera.	Species.	Genera.
1	215	32	1	74	1
2	90	33	1	76	1
3	38	34	1	77	1
4	35	35	1	79	1
5	21	36	3	83	1
6	16	37	1	84	3
7	15	38	1	87	2
8	14	39	2	89	1
9	5	40	2	92	2
10	15	41	1	93	1
11	8	43	4	110	1
12	9	44	1	114	1
13	5	45	1	115	1
14	6	46	1	128	1
15	8	49	2	132	1
16	6	50	4	133	1
17	6	52	1	146	1
18	3	53	1	163	1
19	4	56	1	196	1
20	3	58	1	217	1
21	4	59	1	227	1
22	4	62	1	264	1
23	5	63	3	327	1
24	4	65	1	399	1
25	2	66	1	417	1
26	3	67	1	681	1
27	1	69	1		
28	3	71	1		
29	3	72	1		
30	3	73	1	Total	627

TABLE B.—*Cerambycinæ*: Numbers of genera with 1, 2, 3, . . . species. (Compiled by Dr. J. C. WILLIS, from Coleopt. Cat., Junk and Schenkling, Part 39, 1912.)

Species.	Genera.	Species.	Genera.	Species.	Genera.
1	469	21	2	46	1
2	152	22	5	47	1
3	82	23	1	49	1
4	61	24	3	50	1
5	33	25	3	52	1
6	36	26	3	53	1
7	18	27	1	57	1
8	17	28	1	59	1
9	14	30	2	66	1
10	11	31	3	67	1
11	11	32	1	69	1
12	4	34	3	89	1
13	10	35	2	95	1
14	9	36	1	104	1
15	8	37	1	107	1
16	7	39	2	120	1
17	11	40	2	125	1
18	6	42	1		
19	5	43	2		
20	3	44	1	Total	1,024

TABLE C.—Snakes: Numbers of genera with 1, 2, 3, . . . species. (Compiled by Dr. WILLIS from Boulenger, Cat. of the Snakes in the Brit. Mus., 1893.)

Species.	Genera.	Species.	Genera.	Species.	Genera.
1	131	10	4	27	1
2	35	13	3	31	1
3	28	15	2	33	1
4	17	17	2	40	1
5	16	18	3	45	1
6	9	21	4	74	1
7	8	22	5	97	1
8	8	23	1		
9	9	26	1	Total	293

TABLE D.—Lizards : Numbers of genera with 1, 2, 3, . . . species. (Compiled by Dr. WILLIS from Boulenger, Cat. of the Lizards in the Brit. Mus., 1885.)

Species.	Genera.	Species.	Genera.	Species.	Genera.
1	105	13	2	31	1
2	44	14	1	32	1
3	23	15	1	37	1
4	14	16	3	41	1
5	12	17	1	44	1
6	7	18	2	66	1
7	6	19	3	106	1
8	4	21	2	159	1
9	5	22	1		
10	5	24	1		
11	3	25	3		
12	1	27	2	Total	259

TABLE E.—*Leguminosæ* : Numbers of genera with (approximately) 1, 2, 3, . . . species, with graduated or averaged figures. (Compiled by Dr. WILLIS from Dictionary of the Flowering Plants.)

Species.	Genera.	Graduated or averaged figures.	Species.	Genera.	Graduated or averaged figures.
1	245	245	34	—	1·1
2	66	66	35	5	
3	36	36	40	6	
4	24	34·5	43	—	
5	28	25·4	44	—	0·5
6	30	19·8	45	1	
7	7	16	50	4	
8	13	13·3	53	—	
9	—	11·3	55	1	—
10	27	9·8	60	3	—
11	3	8·6	65	2	—
12	20	7·6	70	6	—
13	1	6·8	75	1	—
14	2	6	80	2	—
15	18	5·4	90	2	—
16	4	4·9	100	4	—
17	—	4·6	110	1	—
18	2	4·1	120	3	—
19	—	3·8	150	3	—
20	15	3·6	160	2	—
21	—	3·4	170	1	—
22	1	3·2	175	1	—
23	—	3	220	1	—
24	3	1·9	290	1	—
25	8		300	1	—
27	1		350	1	—
30	6		400	2	—
33	1		500	1	—
			1,600	1	—
			Total	617	