Williams, C. B. 1947. The logarithmic series and its application to biological problems. *Journal of Ecology.* 34 (2), pp. 253-272.

The Logarithmic Series and Its Application to Biological Problems
Author(s): C. B. Williams
Source: *Journal of Ecology*, Vol. 34, No. 2 (Aug., 1947), pp. 253-272
Published by: British Ecological Society
Stable URL: https://www.jstor.org/stable/2256719
Accessed: 24-10-2018 12:24 UTC

VOLUME 34, No. 2 AUGUST 1947

THE LOGARITHMIC SERIES AND ITS APPLICATION TO BIOLOGICAL PROBLEMS

BY C. B. WILLIAMS, Sc.D., *Rothamsted Experimental Station, Harpenden*

(*With five Figures in the Text*)

CONTENTS

1. INTRODUCTION

In the course of biological investigations of a numerical character the observer frequently obtains values which can be arranged in a discontinuous series of the type known as a 'frequency series'.

For example, there may be a random collection of a number of insects which have been classified into species. In this case the number of species with one individual, with two, with three individuals, and so on, would form a frequency series.

Alternatively, a collection might be made of a number of rats and on each the number of fleas counted. Then the number of rats with one flea, with two fleas, with three and so on, would again form a frequency series. Many other examples could be given, but most can be put under the general type of units classified into groups, or groups divided into units, which form a series of the numbers of groups with one, two or three, etc., units.

It is with one of the possible mathematical interpretations of such data—the logarithmic series—that we are concerned here. It was first suggested for biological problems by R. A. Fisher in 1943 (**1**), and a certain amount of information has already been published during the war; but owing to paper restrictions only a few reprints were obtained and the supply of them is already exhausted.

Dr R. A. Fisher has generously allowed me to quote freely from his contribution, so that all the relevant information can be collected together here.

Before discussing the mathematical properties of the series it must be pointed out that the data under consideration must be a randomized sample with no selection that would affect the size of the groups, or the number of groups of any one size. For example, a museum collection of butterflies in which every effort had been made to obtain many specimens of the 'rare' species, would not be suitable for consideration.

It is also important to understand that the original randomization of the sample may occur in two different ways, (1) by units and (2) by groups, as shown by the two examples given above.

In a randomized collection of individual insects (as, for example, a number of moths caught in a light trap) which are later classified into species, the catch is randomized on the individuals, and an addition to or an increase in the size of the sample will bring in new individuals to species already represented, i.e. new units in old groups.

If, on the other hand, collections of rats are made, and the number of fleas on each rat counted, then an increase in the number of rats examined will not add any fleas to the rats already counted, i.e. all the new units will be in new groups. In this case the sample is randomized by groups.

It will be shown below that samples taken from a population by these two different methods require different mathematical treatment.

2. The logarithmic series* and its properties

In most elementary text-books of algebra there will be found the proof that

$$\log_e(1+x)=x-\frac{x^2}{2}+\frac{x^3}{3}-\frac{x^4}{4},\ \ldots,\ \text{etc.},$$

and the latter expansion is known as the 'logarithmic series'.

As negative terms have little or no meaning in biology, for our purpose it is better to write the equation

$$x+\frac{x^2}{2}+\frac{x^3}{3}+\frac{x^4}{4}+\ldots=-\log_e(1-x), \tag{1}$$

or as a more general frequency series it can be written

$$n_1,\quad \frac{n_1x}{2},\quad \frac{n_1x^2}{3},\quad \frac{n_1x^3}{4},\ \ldots,$$

where n_1 is the number of groups with 1 unit, and the successive terms those with 2, 3, 4, etc., units.

The series is of course discontinuous and has an infinite number of terms.

The total number of groups

The sum of all terms to infinity, which is the total number of groups, is given by

$$S=\frac{n_1}{x}(-\log_e\overline{1-x}). \tag{2}$$

This is finite if x is less than unity; that is to say the series is then convergent.

The series therefore has two constants or parameters: 'n_1' which is the number of groups with one unit, and 'x' which is a number less than one.

The $r+1$th term is obtained from the rth by multiplying by $rx/r+1$, so the second term of the series is less than half the first, and the third term less than one-third of the first, and so on.

Two series of this form are shown in Table 1 and graphically in Fig. 1, A, and it will be seen from the figure that they form hollow curves of the same general appearance as a hyperbola. The nearer x approaches to unity the closer does the resemblance become; and in the limiting case when $x=1$ the curve is identical with a harmonic series (hyperbola): the series is then divergent and the sum of its terms is infinite.

* For the relation of the logarithmic series to the negative binomial see Fisher *et al.* (1943).

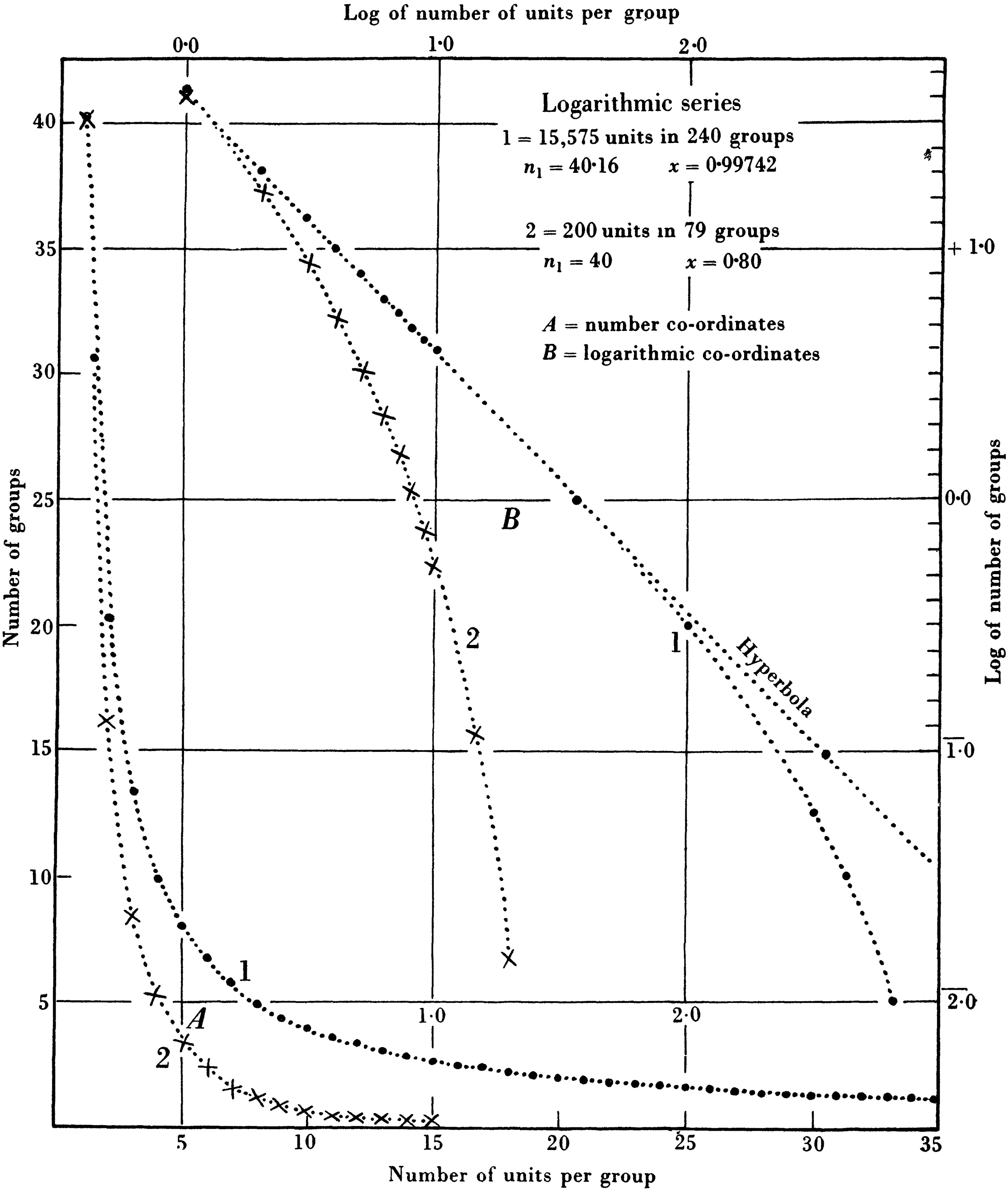

Fig. 1. Two examples of the logarithmic series from Table 1, plotted with number and with logarithmic co-ordinates.

The total number of units

Since the successive terms of the series are the number of groups containing 1, 2, 3, etc., units, it follows that the number of units in each successive term is

$$1n_1, \quad 2\frac{n_1x}{2}, \quad 3\frac{n_1x^2}{3}, \quad 4\frac{n_1x^3}{4}, \text{ etc.,}$$

which equals

$$n_1, \quad n_1x, \quad n_1x^2, \quad n_1x^3, \text{ etc.}$$

Table 1. *Two examples of the logarithmic series*

Term	1 15,575 units in 240 groups	2 200 units in 79 groups	3 Harmonic series* (hyperbola) for comparison
1	40·14	40·0	40·0
2	20·03	16·0	20·0
3	13·32	8·53	13·33
4	9·96	5·12	10·00
5	7·95	3·278	8·00
6	6·66	2·125	6·67
7	5·65	1·498	5·71
8	4·93	1·049	5·00
9	4·37	0·7460	4·44
10	3·92	0·537	4·00
20	1·92	0·02884	2·00
50	0·71	0·00001429	0·80
	$x = 0·99742$	0·80	—
	$\alpha = 40·24$	50·0	—

* Number of groups and of units both infinite.

This is a geometric series, and its sum to infinity (i.e. the total number of units of the sample)* is

$$N = \frac{n_1}{1-x}. \tag{3}$$

This is also finite if x is less than unity.

From (2) and (3) it follows that the average number of units per group equals

$$\frac{N}{S} = \frac{x}{(1-x)\,(-\log_e \overline{1-x})}. \tag{4}$$

Thus for any average number of units per group there is only one possible value of x. When this has been calculated n_1 can be obtained by multiplying N by $(1-x)$.

The relation between x and N/S is shown in Table 2 and graphically in Fig. 2.

Thus for any series of units classified into groups in the form of a logarithmic series, if the total number of units and the total number of groups are both known, it is possible to calculate x and n_1, and so the whole series. In other words, for any definite number of units and of groups only one logarithmic series is possible.

* This is the same as the statistical expression 'first moment', which is the sum of the numbers of groups in each term multiplied by the number of units per group in that term.

The 'second moment', which is sometimes useful, is the sum of the number of groups in each term multiplied by the square of the number of units per group in that term, i.e. it equals

$$1^2n_1 + 2^2\frac{n_1x}{2} + 3^2\frac{n_1x}{3} + 4^2\frac{n_1x}{4}, \text{ etc.,}$$

and this

$$= \frac{N}{1-x} \quad \text{or} \quad \frac{n_1}{(1-x)^2}.$$

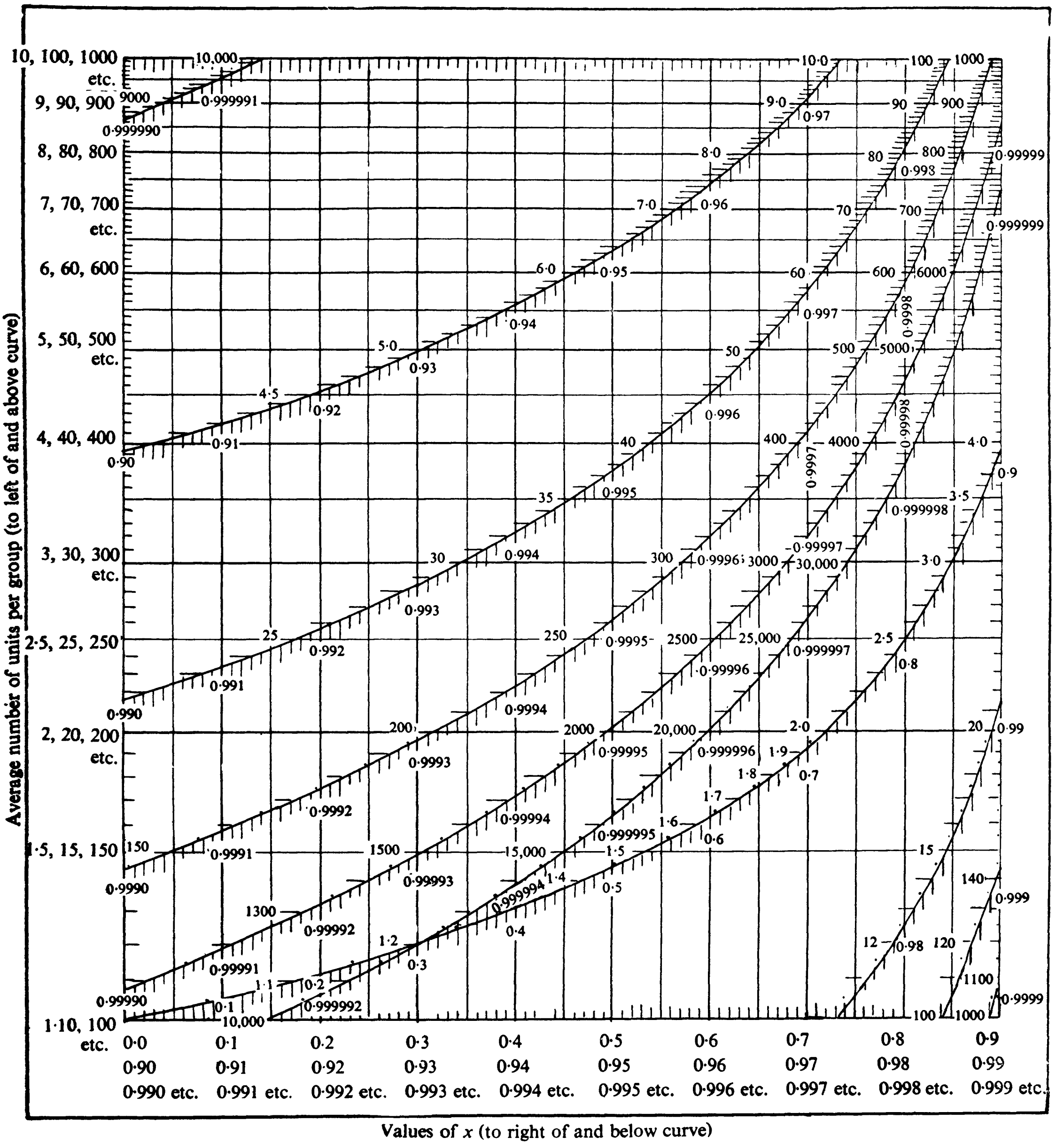

Fig. 2. The relation between the average number of units per group (N/S) and x for all values of the former from 1 to 70,000. For any point on the ordinate there are several values in multiples of 10. The graph to be used for any value indicated by the figures written immediately above the actual curve. Similarly the corresponding value of x to be used on the abscissa is indicated by figures written immediately along and below the actual curve.

Table 2. *Values of x with corresponding values of N/S, and n_1/S*

x	N/S	n_1/S	x	N/S	n_1/S
0·50	1·443	0·7215	0·992	25·68	0·2054
0·60	1·637	0·6548	0·993	28·58	0·2001
0·70	1·938	0·5814	0·994	32·38	0·1940
0·80	2·483	0·4966	0·995	37·48	0·1874
0·85	2·987	0·4480	0·996	45·11	0·1804
0·90	3·909	0·3909	0·997	57·21	0·1716
0·91	4·198	0·3778	0·998	80·33	0·1607
0·92	4·551	0·3641	0·9990	144·6	0·1446
0·93	4·995	0·3496	0·9992	175·1	0·1400
0·94	5·567	0·3340	0·9994	224·5	0·1347
0·95	6·340	0·3170	0·9996	319·4	0·1278
0·96	7·458	0·2983	0·9998	586·9	0·1174
0·97	9·214	0·2764	0·99990	1,086·0	0·1086
0·980	12·53	0·2506	0·99995	2,020	0·1010
0·985	15·63	0·2345	0·999990	8,696	0·0870
0·990	21·47	0·2147	0·999995	16,390	0·0820
0·991	23·38	0·2104	0·9999990	71,430	0·0714

Transformation to logarithmic co-ordinates

If a hyperbolic (harmonic) series of the form n_1, $n_1/2$, $n_1/3$, etc., is transformed graphically to logarithmic co-ordinates, both of numbers of groups and of numbers of units per group, it gives a straight-line relationship (Fig. 1, B). If a logarithmic series is so transformed (same figure) it gives a series of points on a line starting very near the hyperbola transformation but gradually departing from it more and more rapidly in a downward direction, i.e. giving fewer groups containing a large number of units than would be expected from the hyperbolic series. The nearer x is to unity, the longer the transformed series follows closely to the line of the hyperbola. This gives a rapid graphical method of testing if any set of figures is likely or not to represent a logarithmic series.

Different samples from the same population

If several samples of different sizes are taken by the same method from the same population, and if (see Introduction, pp. 253–4) the samples are randomized on units, then as the size of the sample increases the average number of units per group will also increase and hence x will increase.

As the size of the sample increases so also will n_1, the number of groups with one unit, increase; at first (with very small samples) rapidly, but more and more slowly as it approaches a limiting value.

The index of diversity

But for all samples taken from the same population by the same method the ratio of n_1/x is a constant, α. That is,

$$\frac{n_1}{x}=\alpha \quad \text{or} \quad n_1=\alpha x. \tag{5}$$

Since with increasing size of sample x gets nearer to unity but cannot exceed the value, it follows that the number of groups with one unit increases up to the limit of α, but cannot exceed this, however large the sample may be.

The logarithmic series can therefore be written

$$\alpha x, \quad \alpha\frac{x^2}{2}, \quad \alpha\frac{x^3}{3}, \quad \alpha\frac{x^4}{4}, \tag{6}$$

which is sometimes a more convenient form.

In this case the total number of groups (e.g. species) is

$$S=\alpha\,(-\log_e \overline{1-x}), \tag{7}$$

and the total number of units (e.g. individuals) is

$$N=\alpha\,\frac{x}{1-x}. \tag{8}$$

Both n_1 and x therefore vary with the size of the sample, but α is constant for all samples (or summations of samples) from the same population taken by the same method of sampling. It is thus a property of the population sampled. It is high in populations which have a large number of groups relative to the number of individuals and low in populations which have a small number of groups relative to the number of units. We have called it the 'Index of Diversity', as it is a measure of the extent to which the units are associated into groups.

It follows from the above that

$$x=\frac{N}{N+\alpha} \quad \text{or} \quad \alpha=N\,\frac{(1-x)}{x}. \tag{9}$$

Thus if the total number of units and of groups is known in one sample from a given population, the index of diversity can be calculated; and from this it is possible to find the number of groups (and hence the frequency distribution) in any other sample, larger or smaller, from the same population.

The relation between the number of groups (S), the number of units (N), and the index of diversity is given by the formula

$$S=\alpha\,\log_e\left(1+\frac{N}{\alpha}\right). \tag{10}$$

Thus if α is known for any population from any one sample it is only necessary to insert the new N for a second sample of a different number of units, in the above formula, to find the new S or number of groups.

For example, 15,575 moths were captured in a light trap at Harpenden and were found to belong to 240 species. This gives (see below) $\alpha=40{\cdot}25$ approximately. From this we can calculate that had the sample contained only 1000 moths there would have been only about 130 species represented; if, on the other hand, one million moths could have been caught by the same method in the same time, about 405 species could have been expected.

Fig. 3 shows graphically the relation between the values of α and the number of groups and units (e.g. species and individuals) in small samples up to 90 groups and 150 units. Fig. 4 (p. 265) shows larger samples up to 340 groups and 10,000 units. The units are plotted on a logarithmic scale. The relation between $\log N$ and S is approximately linear for each value of α for all large samples.

Fig. 5 (p. 268) shows a similar diagram for still larger samples up to $S=900$ and $N=10{,}000{,}000$.

Table 3 shows some of the basic data from which these tables were prepared.

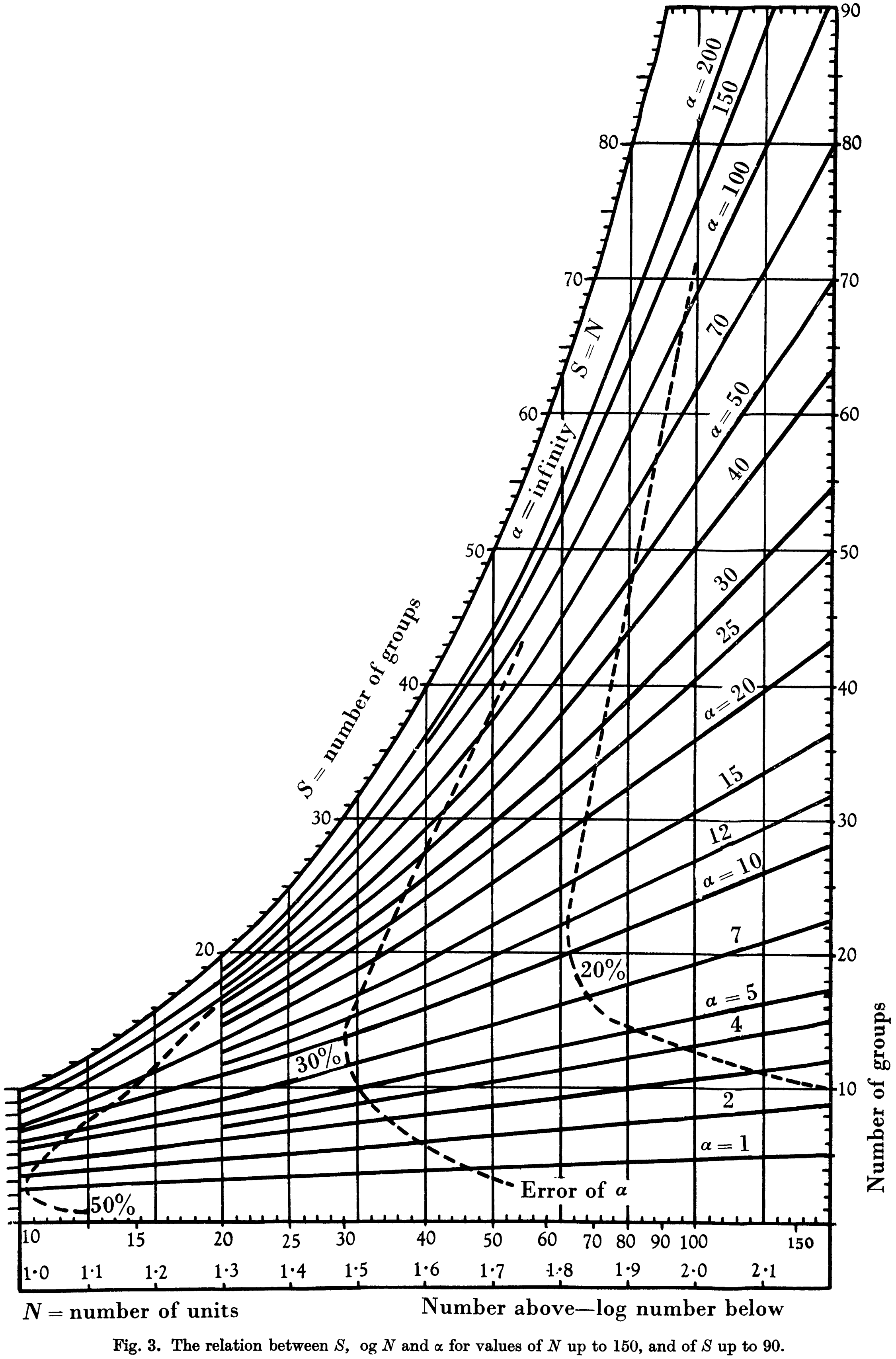

Fig. 3. The relation between S, og N and α for values of N up to 150, and of S up to 90.

Of some biological interest are: (1) the average number of units per group (e.g. individuals per species); (2) the ratio of groups with one unit to the total groups (e.g. the proportion or percentage of monotypic genera); and (3) the proportion of the units which are found in groups of one unit only.

Table 3. *Values of S for different combinations of N and α*

(From *J. Anim. Ecol.* **12**, 53.)

	$N=10$	20	50	100	200	500	1,000	2,000	5,000	10,000	100,000
$\alpha=1$	2·4	3·0	3·9	4·6	5·3	6·2	6·9	7·6	8·4	9·2	—
2	3·6	4·8	6·5	7·9	9·2	11·1	12·4	13·8	15·7	17·0	—
3	4·4	6·1	8·6	10·6	12·6	15·4	17·4	19·5	22·3	24·3	—
4	5·0	7·2	10·4	13·0	15·7	19·3	22·1	24·9	28·5	31·3	—
5	5·5	8·0	12·0	15·0	18·6	23·1	26·5	30·0	34·5	38·0	49·6
6	5·9	8·8	13·4	17·2	21·2	26·6	30·7	34·9	40·4	44·5	—
7	6·2	9·5	14·7	19·1	23·7	30·0	34·8	39·6	46·0	50·9	—
8	6·5	10·0	15·9	20·8	26·1	33·1	38·8	44·2	51·5	57·1	—
9	6·7	10·5	16·9	22·5	28·3	36·3	42·5	48·7	56·9	63·1	—
10	6·9	11·0	17·9	24·0	30·5	39·3	46·2	53·0	62·2	69·1	92·1
12	7·3	11·8	19·7	26·8	34·5	45·0	53·2	61·5	72·4	80·7	—
14	7·6	12·6	21·3	29·4	38·2	50·4	60·0	69·6	82·3	92·0	—
15	7·7	12·7	22·0	30·4	39·9	53·0	63·2	73·5	87·2	97·6	—
16	7·8	13·0	22·7	31·7	41·7	55·6	66·2	77·4	92·0	103·0	—
18	8·0	13·6	23·9	33·8	44·9	60·5	72·6	84·8	101·4	113·8	—
20	8·1	13·9	25·2	35·8	48·0	65·2	78·6	92·3	110·5	124·3	170·4
25	8·4	14·7	27·5	40·3	54·9	76·1	92·9	109·9	132·6	149·9	—
30	8·6	15·3	29·4	44·0	60·8	86·2	106·1	126·5	153·7	174·4	—
35	8·8	15·8	31·1	47·3	66·6	95·5	118·6	142·2	173·0	198·1	—
40	8·9	16·2	32·4	50·1	71·7	104·1	130·3	157·3	193·4	221·0	—
45	9·0	16·5	33·6	52·7	76·2	112·2	141·5	172·7	212·4	243·4	—
50	9·1	16·9	34·7	55·0	80·5	119·9	152·3	185·7	230·8	265·2	380·1
60	9·3	17·2	36·4	58·9	88·0	134·0	172·3	212·1	266·0	307·3	—
70	9·4	17·6	37·7	62·2	94·5	146·8	191·0	237·1	299·8	347·8	—
80	9·4	17·8	38·8	64·9	100·2	158·5	208·2	260·6	330·8	386·9	—
90	9·5	18·1	39·8	67·2	105·3	169·2	224·5	283·1	363·2	424·7	—
100	9·5	18·2	40·6	69·3	109·9	179·2	239·8	304·5	392·2	461·5	690·9
150	—	—	—	—	—	219·9	304·2	399·3	530·3	632·3	—
200	9·8	19·1	44·6	81·1	138·6	250·6	358·1	651·6	479·6	786·4	—

The average number of units per group

This is given by the formula

$$\frac{N}{S}=\frac{e^{S/\alpha}-1}{e^{S/\alpha}} \quad \text{or} \quad =\frac{x}{(1-x)\,(-\log_e \overline{1-x})}.$$

So for all samples with the same value of α (i.e. from the same population) the average number of units per group is dependent on the size of S, or of x; that is, on the size of the sample. It is larger with large samples and smaller with small samples (see Table 2 and Table 4, column 3, for examples).

The proportion of groups with one unit

This is n_1/S (or $100n_1/S$ if expressed as a percentage), and is given by the formula

$$\frac{n_1}{S}=\frac{x}{-\log_e(1-x)}.$$

It is thus dependent on x, which is in turn dependent on the size of the sample for different samples from the same population (i.e. with the same α). It is large in small samples and small in large samples (see Table 2 and Table 4, column 6).

The proportion of units in single groups to the total units

This is n_1/N (or $100n_1/N$ if expressed as a percentage), and is given by the formula

$$\frac{n_1}{N}=1-x.$$

This also varies with x and is large in small samples and very small in large samples from the same population. Examples of the different values of these ratios in different-sized samples from the same population are given in Table 4.

Table 4. *Numbers of species and other properties in different-sized samples from a population with an index of diversity* = 100

N	S	N/S	x	n_1	$100n_1/S$	$100n_1/N$
100,000	691	144·7	0·99902	98	14·2	0·098
10,000	461	21·7	0·9903	97	21·0	0·97
1,000	239	4·18	0·905	95	39·7	9·5
100	69	1·45	0·40	50	72·5	50
10	9·5	1·05	0·05	9·5	99·7	95

The number and percentage of units in series of groups

The number of units in the successive terms has been shown to be a geometric series

$$n_1, \quad n_1x, \quad n_1x^2, \text{ etc.,}$$

and the sum to infinity $N=n_1/(1-x)$. The $r+1$th term is n_1x^r, and the sum of all the units from that term on to infinity is $n_1x^r/(1-x)$. So the total number of units in the first r terms inclusive is

$$\frac{n_1}{1-x}-\frac{n_1x^r}{1-x}=\frac{n_1}{1-x}(1-x^r).$$

The proportion of the total units that are in groups 1 to r is therefore

$$(1-x^r) \quad \text{or} \quad 100\,(1-x^r)\,\%.$$

For example, in a series with $x=0·9$, one-tenth of the units are in groups with one unit each, and 65·2 % are in groups with 10 or less units per group.

If $x=0·99$, one-hundredth of the total number of units are in groups with one unit each, and 9·64 % in groups with 10 or less.

Fifty per cent of the units are found in groups containing up to r units, where

$$r=\frac{\log 0·5}{\log x}.$$

Thus if $x=0·9$, 50 % of the units are in groups with up to 6·6 units, i.e. up to between the sixth and seventh terms of the series.

If $x=0·99$, 50 % of the units are in groups with 68 units or less.

Since for samples from the same population the value of x depends on the size of the sample, it follows that all the above proportions are dependent on the size of the sample.

But the proportions of units in groups of particular sizes is the same for all samples with the same x, and that is in all samples with the same average number of units per group.

It is not possible to get a simple expression for the proportion of *groups* in any series of terms of the logarithmic series.

Sampling from a log series by units

If a sample, randomized by units, is taken from a population arranged in a log series, the sample forms a new log series with the same α, but with a different x. If the new sample contains a proportion p (i.e. $100p\%$) of the number of units in the population sampled, then the new x

$$=\frac{px}{1-(1-p)\,x}.^{*}$$

For example, if one-third of the original population is taken in the sample the new x

$$=\frac{0{\cdot}33x}{1-0{\cdot}67x}.$$

Since a sample of a population which is in a log series gives a log series in the sample, we are justified in assuming that, when we find a log series in a sample, the population sampled is itself in a log series.

Sampling a log series by randomization of groups

If successive samples of a population are based on randomization of groups (see Introduction, pp. 253–4) a different result is obtained. The successive samples have not, as above, the same α and different values of x, but they have the same x and different values of α.

For example, if a random sample of a rat population is taken and 50 rats are found infected with fleas, and on these are found 500 fleas, there will be an average of 10 fleas per rat with fleas. This gives $x=0{\cdot}9732$ (approximately from Fig. 2) and $\alpha=14$ (approximately from Fig. 3). If now a second sample of similar size is taken, one would expect this also to have 50 rats and approximately 500 fleas. The two samples together (i.e. one larger sample) would thus have 100 rats and 1000 fleas, with x as before $0{\cdot}9732$, but $\alpha=28$ approximately.

The error in estimate of α

Fisher has shown ((**1**), p. 56) that the standard error of α is the square root of

$$\alpha^3\,\frac{(N+\alpha)^2\log_e\,(\overline{2N+\alpha}/\overline{N+\alpha})-\alpha N}{(SN+S\alpha-N\alpha)^2}\,.$$

This is complicated to work out, but in Figs. 3 and 4 I have superimposed on the diagram lines of equal percentage error of α. It will be seen that in general the error is high with small numbers of N and S. It is also high if S is large compared with N (a high value of α), or if N is very large compared with S (a low value of α).

Thus the error of estimation of α is about 10% if N is 10,000 and $S=9$; or if N is about 250 and S from 50 to 100. The error is 30% with 30 individuals in 10 groups. With $N=10{,}000$ and $S=170$, the error of α is only about 3%.

Table 5 shows the standard error of α for several different values of N and α.

Table 5. *Error of α for different values of N and α*

(From *J. Anim. Ecol.* **12**, 53.)

	$N=10$	100	1000	10,000	100,000
$\alpha=1$	0·504	0·288	0·141	0·091	—
5	2·785	0·860	0·430	0·282	0·209
10	6·46	1·60	0·719	0·445	0·321
20	15·82	3·19	1·52	0·712	0·495
50	49·87	8·79	2·67	1·359	0·891
100	153·7	20·42	5·04	2·27	1·41

* I am indebted to Mr M. H. Quenouille for this formula.

The index of diversity and Yule's 'characteristic'

I have shown (Williams (**8**)) that the index of diversity is proportional to the reciprocal of the 'Characteristic' defined by Yule in his *Statistical Study of a Literary Vocabulary* (Cambridge, 1944), which he used as a property of the population he was sampling. In his case this was the number of different nouns available for use in the mind of a writer.

Yule's characteristic $= 10{,}000\ (S_2 - S_1)/S_1^2$, where S_1 and S_2 are the first and second moments of the series. The index of diversity for the log series $= S_1^2/(S_2 - S_1)$; but values calculated by this method are much more variable from sample to sample from the same population than values calculated by the methods given below.

3. Methods of fitting a logarithmic series to known data

Two rapid preliminary steps may be taken to see if a logarithmic series is a likely explanation for any set of frequencies.

First by inspection of the numbers it should be checked that the second term should be less than half the first (within limits of error); the third term should be less than one-third of the first, etc.—and there should be a more or less steady fall.

A second rapid test is to plot the successive terms on a logarithmic scale (or on double log paper) and see if they approximate to the transformation shown in Fig. 1, at first near the straight line of the hyperbola and then falling away steadily and more rapidly.

If the total number of units and the total number of groups is known, the logarithmic series can be calculated by several methods varying in accuracy.

Approximate method

Calculate the average number of units per group (N/S), and find x by inspection from Fig. 2, which shows the corresponding values of x for all values of N/S from 1 to 70,000.

Example. 15,575 moths caught in a light trap at Rothamsted were found to belong to 240 species. What is the corresponding logarithmic series?

The average number of individuals per species is 64·9. By inspection (Fig. 2), $x = 0{\cdot}9974$ approximately.

Hence $n_1 = (1 - x)\ N = 0{\cdot}0026 \times 15{,}575 = 40{\cdot}5$.

The series is therefore

$$40{\cdot}5,\quad \frac{40{\cdot}5}{2} \times 0{\cdot}9974,\quad \frac{40{\cdot}5}{3} \times 0{\cdot}9974^2,\ \text{etc.}$$

(For calculating the successive terms see below, p. 267.)

Another approximate method

If N and S are known an approximate value of α can be read off from Figs. 3, 4 or 5.

Example. With the same data as above an inspection of Fig. 4 gives $\alpha =$ approximately 40. Then

$$x = \frac{N}{N + \alpha} = \frac{15{,}575}{15{,}575 + 40} = 0{\cdot}9974.$$

A more accurate method is based on Table 6 (first published by R. A. Fisher in *J. Anim. Ecol.* **12**, 55). The value of log N/S is calculated, and the corresponding value of log N/α is found by interpolation in Table 5. Since N is known, α can then be calculated.

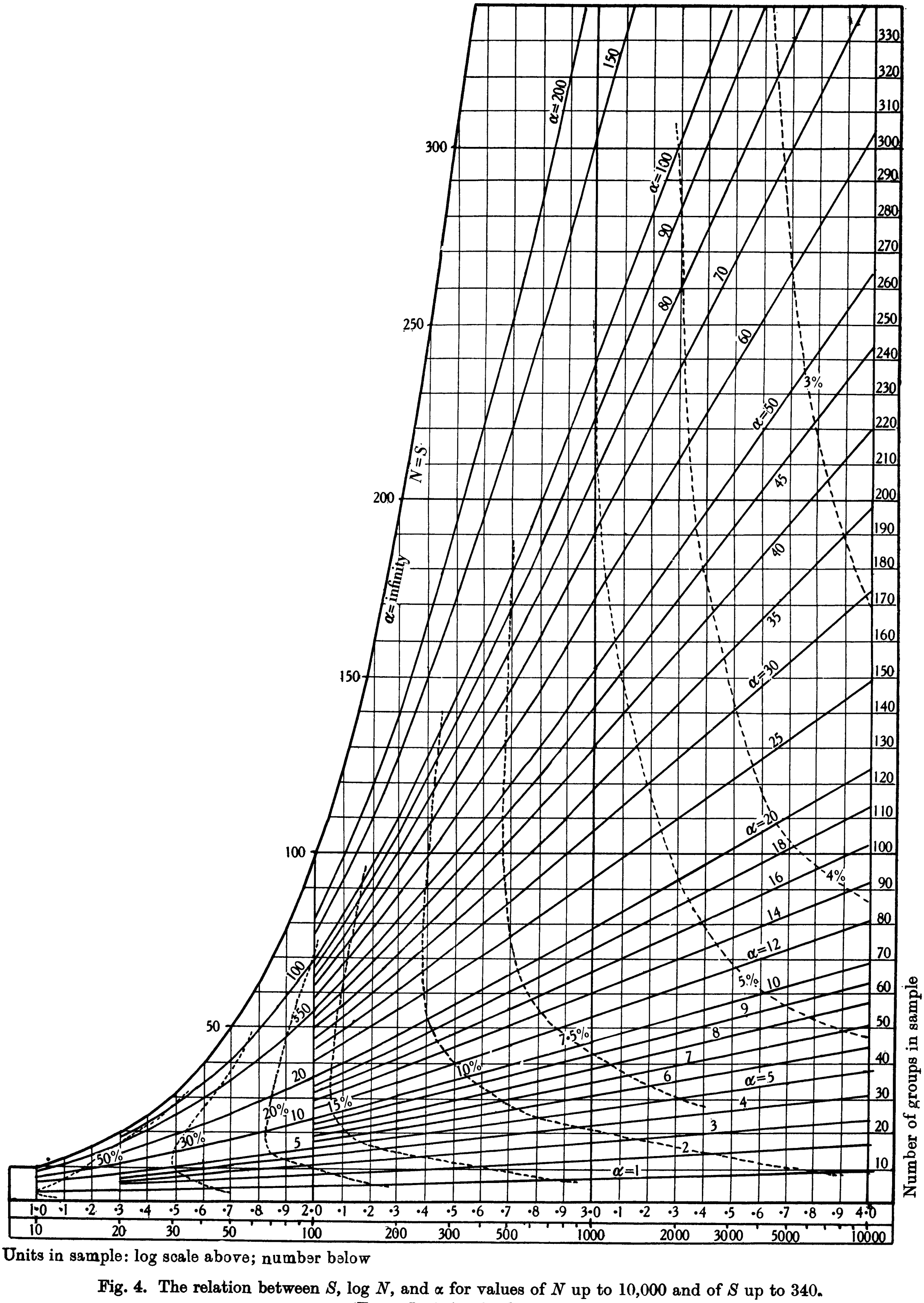

Fig. 4. The relation between S, log N, and α for values of N up to 10,000 and of S up to 340. (From *J. Anim. Ecol.* **12**, 52.)

Example. Data as above, $N = 15{,}575$, $S = 240$:

$$\log N = 4{\cdot}19243, \quad \log S = 2{\cdot}38021.$$

Therefore $$\log N/S = 1{\cdot}81222.$$

From Table 6 it will be seen that:

$$\text{If } \log N/S = 1{\cdot}82, \quad \text{then} \quad \log N/\alpha = 2{\cdot}59684.$$
$$\text{If } \log N/S = 1{\cdot}81, \quad \text{then} \quad \log N/\alpha = 2{\cdot}58484. \qquad (a)$$

Therefore the difference in $\log N/\alpha$ for 0·01 difference in $\log N/S = 0{\cdot}01200$. Since the relation over short distances is approximately linear the difference in $\log N/\alpha$ for 0·00222 in $\log N/S$ is 0·00266.

Adding this to (a) we get:

$$\text{If } \log N/S = 1{\cdot}81222, \quad \text{then} \quad \log N/\alpha = 2{\cdot}58750,$$

but $$\log N = 4{\cdot}19243,$$

therefore $$\log \alpha = 1{\cdot}60492,$$

or $$\alpha = 40{\cdot}2644.$$

Hence $$x = \frac{N}{N+\alpha} = 0{\cdot}9974214 \quad \text{and} \quad n_1 = 40{\cdot}1617.$$

Table 6. *Values of log N/α for different values of log N/S for solving equation $S = \alpha \log_e (1 - N/\alpha)$, given S and N*

(From *J. Anim. Ecol.* **12**, 55.)

$\log_{10} N/S$	0	1	2	3	4	5	6	7	8	9
0·4	0·61121	63084	65023	66939	68832	70701	72551	74382	76195	77990
0·5	0·79766	81526	83271	85002	86717	88417	90105	91779	93442	95092
0·6	0·96730	98356	99973	1·01579	03174	04759	06335	07902	09460	11010
0·7	1·12550	14220	15813	17331	18772	20136	21631	23120	24602	26077
0·8	1·27546	29008	30465	31916	33361	34801	36234	37663	39087	40506
0·9	1·41920	43329	44733	46133	47528	48919	50305	51688	53066	54440
1·0	1·55810	57177	58539	59898	61254	62605	63954	65299	66640	67979
1·1	1·69314	70646	71975	73301	74623	75943	77261	78575	79886	81195
1·2	1·82501	83805	85106	86404	87700	88994	90285	91574	92860	94144
1·3	1·95426	96706	97984	99259	2·00532	01804	03073	04340	05605	06869
1·4	2·08130	09389	10647	11902	13156	14409	15659	16908	18155	19400
1·5	2·20644	21886	23126	24365	25602	26838	28072	29305	30536	31766
1·6	2·32994	34221	35446	36670	37893	39114	40334	41553	42770	43986
1·7	2·45201	46414	47627	48838	50048	51256	52464	53670	54875	56079
1·8	2·57282	58484	59684	60884	62083	63280	64476	65672	66866	68059
1·9	2·69252	70443	71633	72822	74011	75198	76385	77570	78755	79939
2·0	2·81121	82303	83484	84664	85843	87022	88199	89376	90552	91727
2·1	2·92901	94075	95247	96419	97590	98760	99930	3·01099	02267	03434
2·2	3·04600	05766	06931	08095	09259	10422	11584	12745	13906	15066
2·3	3·16225	17384	18542	19699	20856	22012	23168	24323	25477	26630
2·4	3·27783	28936	30087	31238	32389	33539	34688	35837	36985	38133
2·5	3·39280	40426	41572	42717	43862	45006	46150	47293	48436	49578
2·6	3·50719	51860	53001	54141	55280	56419	57558	58696	59833	60970
2·7	3·62106	63242	64378	65513	66648	67782	68915	70048	71181	72313
2·8	3·73445	74577	75707	76838	77968	79097	80227	81355	82484	83611
2·9	3·84739	85866	86992	88119	89244	90370	91495	92619	93743	94867
3·0	3·95991	97114	98236	99358	4·00480	01602	02723	03843	04964	06084
3·1	4·07203	08322	09441	10560	11678	12795	13913	15030	16147	17263
3·2	4·18379	19494	20610	21725	22839	23954	25068	26181	27295	28408
3·3	4·29520	30632	31744	32856	33967	35079	36189	37300	38410	39520
3·4	4·40629	41738	42847	43956	45064	46172	47280	48387	49494	50601
3·5	4·51707	52814	53920	55025	56131	57236	58340	59445	60549	61653

An *accurate method* by trial and error is as follows. In this successive approximations are made to the value of x, and the result tested against the data.

Example. Data as above, $N=15{,}575$, $S=240$. We know that

$$\frac{S}{N}=\frac{1-x}{x}\,(-\log_e \overline{1-x}).$$

If we use logs to the base 10 we get

$$\frac{1-x}{x}\,(-\log \overline{1-x})=\frac{1}{2\cdot30258}\times\frac{N}{S}=\text{(in this case) } 0\cdot0066922.$$

We require to find a value of x that will fit this equation. Starting with a first approximation (from Fig. 2) of $x=0\cdot9974$, we proceed as follows:

If x equals	then $1-x$ equals	then $-\log(1-x)$ equals	$\frac{1-x}{x}$ then $(-\log \overline{1-x})$ equals	Conclusion
		First approximation		
0·9974	0·0026	$-\bar{3}\cdot4150$ $=2\cdot5850$	0·006738	Too large; so make x larger
		Second approximation		
0·99742	0·00258	$-\bar{3}\cdot4116$ $=2\cdot5884$	0·006695	Slightly too large
		Third approximation		
0·997421	0·002579	$-\bar{3}\cdot4114$ $=2\cdot5886$	0·0066932	Very slightly too large

and so on until the required accuracy is reached. The final value taken for x was 0·9974214. Hence

$$n_1=15{,}575\times0\cdot0025786=40\cdot1617,$$

and

$$\alpha=N\,\frac{1-x}{x}=40\cdot2644.$$

To calculate the series, or any one term

This may be done by direct calculation on a calculating machine, but without this the simplest method is to use logs, and in this case the logs of reciprocals, given in most mathematical tables, are helpful.

The log of the rth term $=\log n_1+(r-1)\log x+\log 1/r$. Since $\log x$ is negative its value is subtracted from $\log n_1$ in successive repetitions, and the log of the reciprocal added for each term.

Example. Data as above:

$$n_1=40\cdot2644,\quad x=0\cdot9974214;$$

$$\log n_1=1\cdot604917,\quad \log x=\bar{1}\cdot998787=-0\cdot001213;$$

Term	$\log n_1+(r-1)\log x$		$\log 1/r$		Total log	Number
1	1·604917		—		—	40·2644
2	1·603704	+	$\bar{1}\cdot6990$	=	1·3027	20·08
3	1·602491	+	$\bar{1}\cdot5229$	=	1·1254	13·35
4	1·601278	+	$\bar{1}\cdot3797$	=	0·9800	9·550
5	1·600065	+	$\bar{1}\cdot3010$	=	0·9011	7·963

and so on. For the 50th term

50	1·604917	log 1/50	—	—
	$-(49\times0\cdot001213)$	=		
	$=1\cdot54448$	$\bar{2}\cdot3010$	$\bar{1}\cdot8455$	0·7005

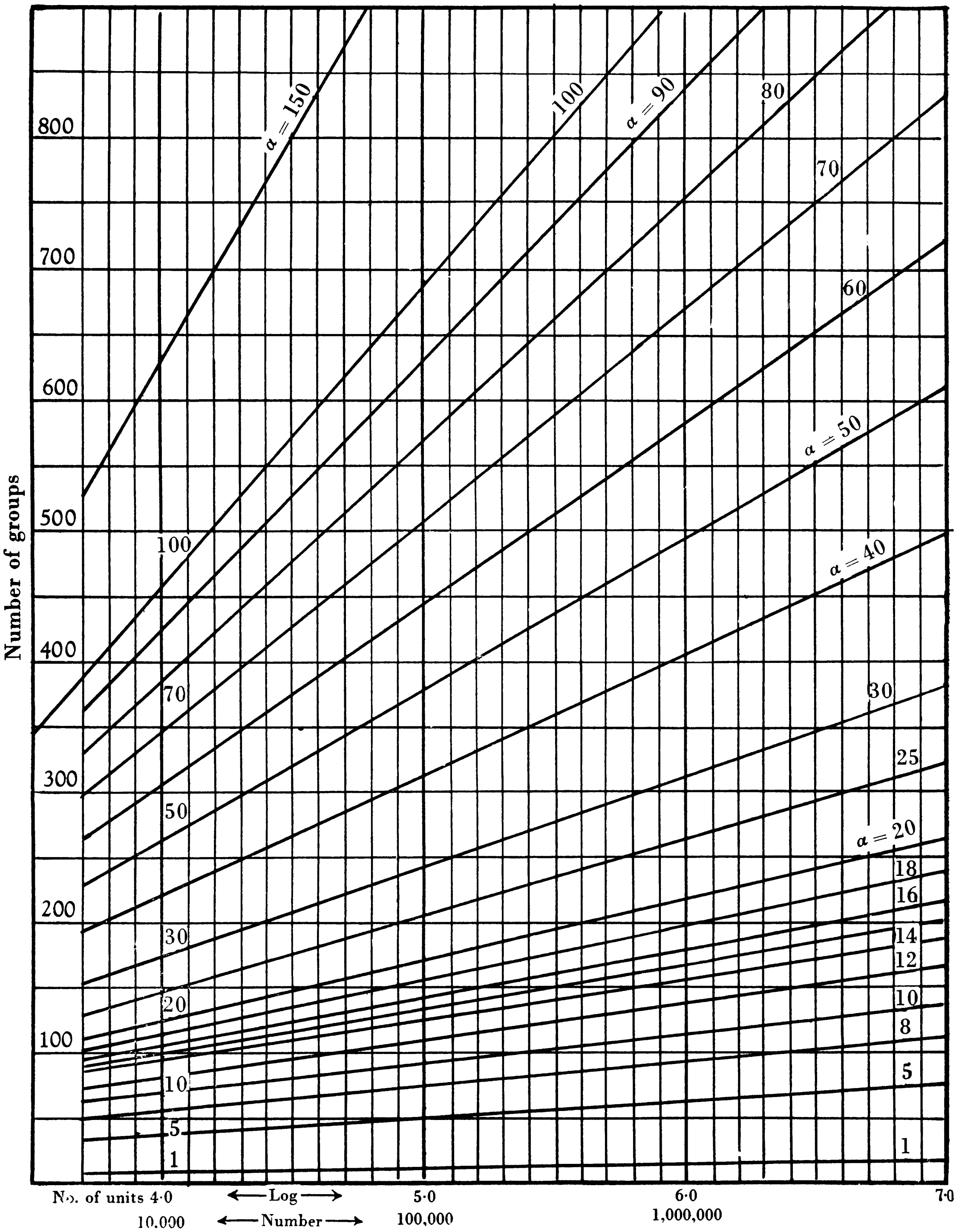

Fig. 5. The relation between S, log N, and α for values of N from 10,000 to 10,000,000, and S up to 900.

Alternate approximate method of finding the index of diversity from two samples of different sizes

An approximate value of the index of diversity may be calculated for a given population, if two or more samples can be obtained of different sizes, neither small.

This method is particularly useful for botanical purposes, as, if the ratio between the different sample sizes is known, the actual number of individuals in each need not be counted. This makes it possible to find α from the number of species of plants on two areas of different size, provided that we can assume, without serious error, that the number of individuals is proportional to the area of the samples. The actual numbers of individuals, however, must be large.

The method is based on the relation

$$S = \log_e \left(1 + \frac{N}{\alpha}\right).$$

When N is very large compared with α we can neglect the 1 in comparison with N/α and say that

$$S \text{ is proportional to } (\log_e N/\alpha).$$

Hence, if two samples from the same population contain N and pN units,

$$S_{pN} - S_N = \left(\log_e \frac{pN}{\alpha} - \log_e \frac{N}{\alpha}\right) = \alpha \log_e p.$$

Thus if a sample size is multiplied by p, the number of groups (e.g. species) is increased by $\alpha \log_e p$.

If the size of a sample is doubled the number of species added is $\alpha \log_e 2 = 0{\cdot}693\alpha$. If the size is multiplied by e ($=2{\cdot}718$) the number of species added equals α. This latter fact could be easily applied to botanical surveys by using two quadrats whose diameters were in the ratio of 100 to 165 ($=\sqrt{2{\cdot}718}$). Then the average increase of species between samples of the two sizes would be a direct measure of the index of diversity. The samples must, however, be large enough to be representative even of the larger plants (see Jones & Williams (**3**)).

Example. Blackman (*Ann. Bot., Lond.*, **49**, 760) states that the number of species of plants found on quadrats of various sizes in a grassland area in England was as follows:

Area in sq.in.	Average no. of species	Increase in no. of species on doubling size of sample
16	11·1	2·5
32	13·6	2·5
64	16·1	2·5
128	18·2	2·1
	Average	2·37

Therefore $\alpha = 2{\cdot}37 \div \log_e 2 = 2{\cdot}37 \div 0{\cdot}693 = 3{\cdot}42$ approximately.

This is probably an underestimate as the samples are small, but as α is also small the error will be relatively less.

Calculation of the number of groups common to two samples

An extension of the above method gives a means of finding the number of groups common to two large samples from the same population on the assumption that it conforms to the logarithmic series (for example, the number of species common to two areas of an ecological association). The samples, or areas must, however, be of different sizes.

Let the two areas be of size A and B, and the number of species in each a and b. If they are from the same population they will have the same α, and if the samples are large we can neglect the 1 in comparison with N/α in the equation $S=\alpha \log_e (1+N/\alpha)$.

Let T be the total number of species in the two samples, then the increase in species by adding B to A

$$=T-a=\alpha \log_e \frac{A+B}{A}.$$

The increase in species by adding A to B

$$=T-b=\alpha \log_e \frac{A+B}{B}.$$

From these two equations T (and α) can be found and the number of species expected to be common to both $=a+b-T$.

Example. The island of Guernsey has 804 species of flowering plants on 24 sq. miles.

The island of Alderney has 519 species of flowering plants on 3 sq. miles.

On the assumption of identity of origin,

$$T-804=\alpha \log_e 1{\cdot}125, \quad T-519=\alpha \log_e 9{\cdot}00.$$

Hence $T=820$ and the number common to the two would be 503. The actual number observed was 480 which shows a high degree of relationship. For fuller discussion, see Williams ((**6**), p. 42), and Williams (**9**).

4. Summary of biological applications of the logarithmic series

The series has so far given reasonably good calculated fits to observed data in the following biological problems. The references are to the bibliography on p. 271.

A. *Individuals classified into species:*

(1) Lepidoptera caught in a light trap at Harpenden, England ((**1**), pp. 44–8).
(2) Capsidae caught in a light trap at Harpenden, England ((**1**), p. 49).
(3) A collection of butterflies made in Malaya by A. S. Corbet ((**1**), p. 42 and (**6**), p. 15).
(4) Butterflies from Mentawi Island, and from Karakorum ((**1**), p. 43).
(5) Butterflies from Tioman Island, Malaya ((**1**), p. 43).
(6) Elmidae (Coleoptera) from Mexico ((**1**), p. 43).
(7) General population of British nesting birds ((**6**), p. 13).
(8) Lepidoptera in light traps in U.S.A. (**7**).
(9) Mosquitoes in light traps in U.S.A. ((**6**), p. 14).
(10) Aphis caught on sticky traps in Derbyshire (Broadbent, data in *Proc. R. Ent. Soc.* ((**21**), pp. 41–6).
(11) Spiders caught in nets (data Freeman, *J. Anim. Ecol.* **15**, 70).

B. *Species and area (especially with plants):*

(1) Grass land in Britain ((**6**), p. 3).
(2) Aspen association in Michigan, U.S.A. ((**6**), p. 5).
(3) Ground vegetation in Tectona forest, Java ((**6**), p. 9).
(4) Species of plants common to related small areas, e.g. Channel Islands ((**6**), p. 42), São Tomé and neighbouring islands (**9**).
(5) General discussion on area and number of species of plants (**4**).

C. *Parasites and host:*

(1) Lice on heads of human beings ((**6**), p. 11).

(2) Fleas on rats (unpublished data from J. L. Harrison).

D. *Species classified into genera:*

(1) Orthoptera of world. Mantidae and Acridiidae ((**6**), p. 17).

(2) Dermaptera of world ((**6**), p. 18).

(3) Coccidae of world ((**6**), p. 18).

(4) British Coleoptera ((**6**), p. 24).

(5) British Lepidoptera ((**6**), p. 24).

(6) British Cicadina ((**6**), p. 28).

(7) British birds ((**6**), p. 29).

(8) Flowering plants of the world ((**6**), p. 38).

(9) British flowering plants ((**6**), p. 32).

(10) British Capsidae (Miridae) (data from China, Ent. Soc. London. Generic names, Brit. Inst. Pt. 8).

(11) Animals and plants in Ecological Communities (**10**).

F. *Miscellaneous applications:*

(1) Number of publications by biologists (**5**).

(2) Number of insects caught in nets at sea (unpublished data from A. C. Hardy).

(3) Species of insects infesting food dumps (**2**).

(4) Number of bacteria in colonies (unpublished data from Jones and Quenouille at Rothamsted).

(5) Larvae of a gall midge in grains of wheat (unpublished data from H. F. Barnes).

REFERENCES

(1) **Fisher, R. A., Corbet, A. S. & Williams, C. B. (1943).** The relation between the number of individuals and the number of species on a random sample of an animal population. *J. Anim. Ecol.* **12**, 42–58.

(2) **Harrison, J. L. (1945).** Stored products and the insects infesting them, as examples of the logarithmic series. *Ann. Eugen., Lond.*, **12**, 280–2.

(3) **Jones, E. W. & Williams, C. B. (1945).** The index of diversity as applied to ecological problems. *Nature, Lond.*, **155**, 390.

(4) **Williams, C. B. (1943).** Area and number of species. *Nature, Lond.*, **152**, 264.

(5) **Williams, C. B. (1944).** The number of Publications written by Biologists. *Ann. Eugen., Lond.*, **12**, 143–6.

(6) **Williams, C. B. (1944).** Some applications of the logarithmic series and the index of diversity to ecological problems. *J. Ecol.* **32**, 1–44.

(7) **Williams, C. B. (1945).** Recent light trap catches of Lepidoptera in U.S.A., analysed in relation to the logarithmic series and the index of diversity. *Ann. Ent. Soc. Amer.* **38**, 357–64.

(8) **Williams, C. B. (1946).** Yule's 'Characteristic' and the 'Index of Diversity'. *Nature, Lond.*, **157**, 482.

(9) **Williams, C. B. (1947).** The logarithmic series and the comparison of island floras. *J. Linn. Soc. Lond.* (in the Press).

(10) **Williams, C. B. (1947).** The generic relations of species in small ecological communities. *J. Anim. Ecol.* **16**, 11–18.

APPENDIX

Summary of formulae connected with the logarithmic series

N = total number of units.

S = total number of groups.

n_1 = number of groups with one unit.

α = the 'Index of Diversity', a constant for all samples from the same population, if randomized on units.

x = a constant for one sample, always less than unity.

N/S = the average number of units per group.

n_1/S = the proportion of groups with one unit.

The logarithmic series is $n_1,\ \frac{n_1 x}{2},\ \frac{n_1 x^2}{3},\ \frac{n_1 x^3}{4}$, etc.,

or $\alpha x,\ \alpha\frac{x^2}{2},\ \alpha\frac{x^3}{3},\ \alpha\frac{x^4}{4}$, etc.,

I. (n_1, α and x): $\alpha=\frac{n_1}{x},\quad n_1=\alpha x,\quad x=\frac{n_1}{\alpha}.$

II. (S, α and x): $S=\alpha\,(-\log_e \overline{1-x}).$

III. (S, n_1 and x): $S=\frac{n_1}{x}(-\log_e \overline{1-x}),\quad \frac{n_1}{S}=\frac{x}{-\log_e \overline{1-x}}.$

IV. (S_1, n_1 and α): $S=\alpha\left(-\log_e \overline{1-\frac{n_1}{\alpha}}\right).$

V. (N, α and x): $x=\frac{N}{N+\alpha},\quad N=\frac{\alpha x}{1-x},\quad \alpha=\frac{N(1-x)}{x}.$

VI. (N, n_1 and x): $n_1=N(1-x),\quad N=\frac{n_1}{(1-x)},\quad x=\frac{N-n_1}{N},\quad \frac{n_1}{N}=1-x.$

VII. (N, n_1 and α): $N=\frac{n_1\alpha}{\alpha-n_1},\quad \alpha=\frac{Nn_1}{N-n_1},\quad n_1=\frac{N\alpha}{N+\alpha}.$

VIII. (N, S and x): $S=\frac{N}{x}(1-x)(-\log_e \overline{1-x}),\quad \frac{N}{S}=\frac{x}{(1-x)(-\log_e \overline{1-x})}.$

IX. (N, S and α): $S=\alpha\log_e\left(1+\frac{N}{\alpha}\right),\quad N=\alpha\,(e^{s/\alpha}-1),\quad \frac{N}{S}=\frac{e^{s/\alpha}-1}{e^{s/\alpha}}.$

X. (N, S and n_1): $S=\frac{Nn_1}{N-n_1}\log_e\frac{N}{n_1}.$

XI. (N, S, n_1 and x): $\frac{n_1}{S}=\frac{N}{S}(1-x).$

XII. The increase in number of groups obtained by multiplying the size of sample by p when the sample is large (i.e. when N is large compared with α) approximates to $\alpha\log_e p$. Doubling the size of the sample therefore adds $\alpha\log_e 2$ groups. Multiplying the size of the sample by e ($=2{\cdot}718$) adds α groups.

XIII. The variance of α is $\alpha^3\frac{(N-\alpha)^2\log_e \overline{(2N+\alpha/N+\alpha)}-\alpha N}{(SN+S\alpha-N\alpha)^2}.$

The standard error of α is the square root of this.

XIV. When $x=0{\cdot}633$, i.e. when the average number of units per group is $1{\cdot}72$, $S=\alpha$.

XV. If S_1 and S_2 are the first and second moments of the series:

$$S_1=N=\frac{n_1}{1-x},\quad S_2=\frac{N}{1-x}=\frac{n_1}{(1-x)^2},\quad \alpha=\frac{S_1{}^2}{S_2-S_1}.$$

XVI. The percentage of units in groups containing $r+1$ or more units is $100x^r$. The percentage in groups containing r or less units is $100\,(1-x^r)$.