The Relation Between the Number of Species and the Number of Individuals in a Random Sample of an Animal Population

R. A. Fisher, A. Steven Corbet, C. B. Williams

THE RELATION BETWEEN THE NUMBER OF SPECIES AND THE NUMBER OF INDIVIDUALS IN A RANDOM SAMPLE OF AN ANIMAL POPULATION

By R. A. FISHER (*Galton Laboratory*), A. STEVEN CORBET (*British Museum, Natural History*) AND C. B. WILLIAMS (*Rothamsted Experimental Station*)

(With 8 Figures in the Text)

CONTENTS

PART 1. RESULTS OBTAINED WITH MALAYAN BUTTERFLIES

By A. STEVEN CORBET (*British Museum, Natural History*)

It is well known that the distribution of a series of biological measurements usually conforms to one of three types:

(*a*) the binomial distribution, where the frequencies are represented by the successive terms of the binomial $(q+p)^n$;

(*b*) the normal distribution, in which the results are distributed symmetrically about the mean or average value, and which is the special case of (*a*) when *p* and *q* are equal;

(*c*) the Poisson series, in which the frequencies are expressed by the series

$$e^{-m}\left(1+m+\frac{m^2}{2!}+\frac{m^3}{3!}+\dots\right),$$

where *m* is the mean and *e* is the exponential base 2·7183.

The usual practice of calculating the arithmetic mean in a set of measurements of a biological nature, such as wing length of a butterfly, assumes a distribution showing no wide departure from normality, although it does not appear that this procedure has been vindicated.

It is the usual experience of collectors of species in a biological group, such as the Rhopalocera, that the species are not equally abundant, even under conditions of considerable uniformity, a majority being comparatively rare while only a few are common. As far as we are aware, no suggestion has been made previously that any mathematical relation exists between the number of individuals and the number of species in a random sample of insects or other animals. Recently, it has been found (Corbet, 1942) that, leaving out of account the commoner species of which no attempt was made to collect all individuals seen, the number of species *S* of butterflies of which *n* individual specimens were collected by a single collector in Malaya was given closely by the expression
$$S=C/n^m,$$

where *C* and *m* are constants.* When *m* is unity, as is the case with the Malayan collection, and has since been found to be a condition which obtains with collections of butterflies from Tioman Island and the Mentawi Islands in which the relation between *S* and *n* follows the above equation, the number of species of which 1, 2, 3, 4, ... specimens were obtained was very close to a series in harmonic progression. Thus, the series can be written
$$C\left(1+\tfrac{1}{2}+\tfrac{1}{3}+\dots\right).$$

Although this relation holds accurately with the rarer species, there is less agreement in the region of the common species; in fact, theoretical considerations preclude an exact relationship here.

Prof. Fisher (see Part 3) has evolved a logarithmic series which expresses accurately the relation between species and individuals in a random sample throughout the whole range of abundance:
$$S=n_1\left(1+\frac{x}{2}+\frac{x^2}{3}+\dots\right),$$

where *S* is the total number of species in the sample, n_1 the number of species represented by single specimens, and *x* is a constant slightly less than unity but approaching this value as the size of the sample is increased.

The total number of individual specimens, *N*, in the sample at all levels of abundance is given by $N=n_1/(1-x)$ and $N(1-x)/x$ is a constant α independent of the size of the sample. As the size of the sample is increased, n_1 approaches α.

The Fisher series has been established for all the entomological collections tested in which there was

---

* This equation may be written
$$\log S=\log C+m\log n,$$
so that the plot of the logarithms of *S* and *n* is a straight line.

reason to believe that the collecting had been un-selective (see Table 1). It is clear that when S and N are known, as is usually the case, the statistics n_1, x and α can be calculated. It is a curious fact that the number of uniques should approach a constant value with increasing size of collection. It is important to ascertain how far this type of distribution of indi-viduals among species holds in other zoological groups, for it would appear that we have here an effective means of testing whether a collection has been made under conditions approaching random sampling or whether some degree of selection has been exercised, a consideration which is often of some importance in faunistic studies.

tinued after 24 specimens had been taken. In such cases, we have the following information:

The total number of species under 25 individuals per species:

$$S_{(1-24)} = n_1 \left(1 + \frac{x}{2} + \frac{x^2}{3} + \ldots + \frac{x^{23}}{24} \right).$$

The total number of individuals at frequencies below 25 per species:

$$N_{(1-24)} = n_1 \left(1 + x + x^2 + \ldots + x^{23} \right).$$

Table 2, which gives the results obtained with the Malayan butterflies, is based on these considerations, and shows the very close relation between the ob-served and the calculated results.

Table 1. *Entomological collections examined showing the relation between the numbers of species and individuals*

	Observations				Calculations			
	Total indi-viduals	No. between 1 and 24	Total species	No. between 1 and 24	x	n_1 calc.	n_1 found	α
Malayan Rhopalocera	—	3306	620	501	0·997	135·05	118	135·47
Rhopalocera from Tioman Island (east coast of Malaya) (Malay collector, 1931)	157	—	41	—	0·887	15·96	19	18·00
Rhopalocera from Mentawi Islands (excluding Hesperiidae) (C. Boden Kloss and N. Smedley, 1924)	1,878	890	135	110	0·983	32·77	37	33·35
Karakorum Rhopalocera (Mme J. Visser-Hooft; *vide* Evans, 1927)	403	195 (1–28)	27	24 (1–28)	0·984	6·42	6	6·52
Mexican Elmidae (Col.) (H. E. Hinton; *vide* Hinton, 1940)	11,798	—	35	—	0·9998	4·72	4	4·72

How far the results obtained with any particular collection can be regarded as representative of the distribution of the same species group in the area in which the collection was made must obviously de-pend on the uniformity or otherwise of the conditions prevailing when the collection was made and to the extent to which these conditions are representative of the habitat. In an equatorial forest-clad island with no mountain heights above 2000 ft., conditions are very uniform as far as such orders as Leipido-ptera are concerned; although even here some allowance must be made for the fact that collections of butterflies made in such regions are usually poor in the crepuscular species. In temperate climates, it is evident that results obtained during one period of the year are usually inapplicable to other seasons or to the year as a whole. It would appear that the results obtained with the moth trap at Harpenden (see Part 2) can be regarded as giving an accurate picture of the distribution frequencies of the photo-tropic moths in the area, and it is probable that the same is true of the collection of Mexican Elmidae.

With many collectors, and for a variety of reasons, the collecting of common species is discontinued once a certain number of specimens of these are obtained. In the case of the Malayan Rhopalocera cited, collecting of all individuals seen was not con-

Table 2. *Calculated and observed distribution frequencies of butterflies collected in Malaya*

The values in the second column are obtained from the Fisher series given on p. 42, taking $x = 0.997$.

n	S (calc.)	S (found)	Deviations
1	135·05	118	17·05
2	67·33	74	−6·77
3	44·75	44	0·75
4	33·46	24	9·46
5	26·69	29	−2·31
6	22·17	22	0·17
7	18·95	20	−1·05
8	16·53	19	−2·47
9	14·65	20	−5·35
10	13·14	15	−1·86
11	11·91	12	−0·09
12	10·89	14	−3·11
13	10·02	6	4·02
14	9·28	12	−2·72
15	8·63	6	2·63
16	8·07	9	−0·93
17	7·57	9	−1·43
18	7·13	6	1·13
19	6·74	10	−3·26
20	6·38	10	−3·62
21	6·06	11	−4·94
22	5·77	5	0·77
23	5·50	3	2·50
24	5·25	3	2·25
			0·82

According to the calculated values, the total species and total individuals at levels between $n = 1$ and $n = 24$, are 501·92 and 3132·24 respectively. The actual values found for the total species and total individuals between $n = 1$ and 24 are 501 and 3306 respectively.

For the above series in Table 2, $\chi^2 = 19·270$, degrees of freedom $= 23$, and P is between 0·90 and 0·80, showing the deviations of the observed results from the calculated values are not significant.

REFERENCES

Corbet, A. S. (1942). 'The distribution of butterflies in the Malay Peninsula.' Proc. R. Ent. Soc. Lond. (A), 16: 101–16.

Evans, W. H. (1927). 'Lepidoptera-Rhopalocera obtained by Mme J. Visser-Hooft of the Hague (Holland) during an exploration of previously unknown country in the Western Karakorum, N.W. India.' Tijds. Ent. 70: 158–62.

Hinton, H. E. (1940). 'A monographic revision of the Mexican water beetles of the family Elmidae.' Nov. Zool. 42: 217–396.

PART 2. RESULTS OBTAINED BY MEANS OF A LIGHT-TRAP AT ROTHAMSTED

By C. B. WILLIAMS

This gives an account of the application of Fisher's series (see Part 3 of this paper) for the frequency of occurrence of species of different levels of abundance in a random sample, to collections of nocturnal Lepidoptera made by means of a light-trap at Rothamsted Experimental Station, Harpenden, Herts, England, during the four years 1933–6. The trap and insects caught in it have already been discussed in a series of papers (Williams, 1939, 1940).

It was necessary to choose, from the material collected, groups in which all or nearly all of the specimens had been identified to species. For this certain families of Lepidoptera were most suited, and the discussion below deals with the captures in the Sphingidae, Noctuidae, Arctiidae, Geometridae and a few other related families. In the Geometridae the genus *Eupithecia* was omitted owing to difficulties of identification. Altogether 15,609 individuals belonging to 240 species were captured. The names and details of numbers for each species will be found in Williams (1939, Tables 6–8).

The frequency of species of different abundance

Table 3 shows the frequency distribution of the species for the four years added together. It will be seen that 35 species were represented by a single individual each; that 85 (including the 35 above) were represented by 5 or fewer individuals; 115 by 10 or fewer; and 205 species by 100 or fewer individuals; leaving therefore 35 species with over 100 individuals per species. The highest total of one species was 2349 individuals of *Agrotis exclamationis*. The results up to 50 individuals per species are represented diagrammatically by the vertical lines in Fig. 1 A, giving a curve closely resembling a hyperbola.

If, however, the log number of species is plotted against the log number of individuals as in Fig. 2 A it will be seen that, while the straight-line distribution expected for a hyperbola holds approximately true for the rarer species, the number of commoner species is distinctly below the hyperbolic expectation

and falls rapidly away from it at higher numbers of individuals per species.

Fisher suggests (see Part 3) that the true series is represented not by

$$n_1, \frac{n_1}{2}, \frac{n_1}{3}, \frac{n_1}{4}, \ldots,$$

which would be the hyperbolic series and which would require an infinite number of species and an infinite number of individuals; but by the series

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

when n_1 is the number of species with 1 individual and x is a number less than 1.

Lepidoptera
in light trap at
Rothamsted

A = total of 4 years
B = 1/8th of year 1933
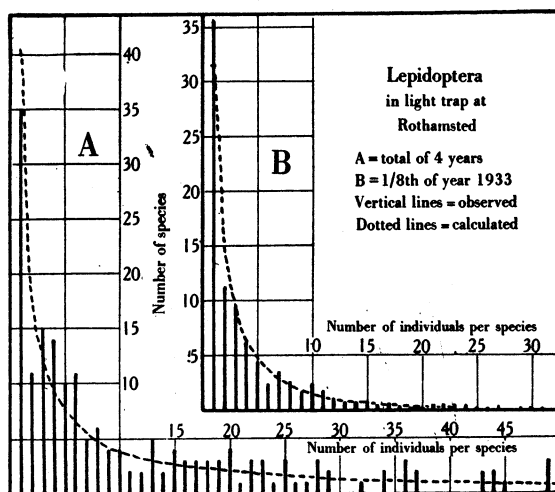Vertical lines = observed
Dotted lines = calculated

Fig. 1

If this is correct he shows that (1) the total number of individuals (N) is finite and $= \dfrac{n_1}{1-x}$, (2) the total number of species (S) is finite and

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

Table 3. *Observed and calculated captures of Macrolepidoptera in a light-trap at Rothamsted*

No. of individuals per species	(1) All 4 years		(2) Average of single year		(3) Average of 1/8 year		(4) Capsidae, all 4 years	
	Obs.	Calc.	Obs.	Calc.	Obs.	Calc.	Obs.	Calc.
1	35	40·14	39·75	37·55	35·63	31·77	18	11·84
2	11	20·03	20·25	18·60	11·38	14·74	8	5·87
3	15	13·32	16·00	12·28	9·63	9·12	3	3·88
4	14	9·96	8·75	9·12	6·24	6·35	4	2·89
5	10	7·95	8·75	7·22	4·50	4·71	2	2·29
	85	91·42	93·50	84·77	67·38	66·68	35	26·77
6	11	6·66	6·25	5·96	2·38	3·64	3	1·89
7	5	5·65	4·25	5·06	3·50	2·90	2	1·61
8	6	4·93	6·00	4·39	2·63	2·55	1	1·40
9	4	4·37	4·00	3·86	1·75	1·93	1	1·23
10	4	3·92	3·50	3·44	2·13	1·62	—	1·10
	30	25·48	24·00	22·70	12·00	12·44	7	7·22
11	2	3·56	2·00	3·10	1·75	1·37	1	0·99
12	2	3·25	2·50	2·81	1·13	1·16	—	0·90
13	5	3·00	2·75	2·57	0·75	0·99	—	0·82
14	2	2·77	2·25	2·36	0·75	0·86	—	0·76
15	4	2·58	3·25	2·19	0·75	0·74	2	0·70
	15	15·16	10·75	13·03	4·18	5·12	3	4·17
16	3	2·42	2·00	2·03	0·50	0·64	—	0·65
17	3	2·27	2·75	1·89	0·63	0·56	—	0·61
18	3	2·14	1·25	1·77	0·38	0·49	—	0·57
19	3	2·02	0·50	1·66	0·25	0·43	—	0·54
20	4	1·92	2·75	1·56	0·25	0·38	—	0·50
	16	10·78	9·25	8·91	2·00	2·53	0	2·87
21	1	1·82	2·75	1·47	0·50	0·34	—	0·48
22	3	1·73	1·75	1·39	0·50	0·30	1	0·45
23	3	1·65	0·75	1·32	0·50	0·27	1	0·43
24	1	1·58	1·00	1·25	0·50	0·24	—	0·41
25	3	1·51	0·50	1·19	0·13	0·21	—	0·38
	11	8·28	6·75	6·62	2·13	1·35	2	2·14
Total to 25	157	151·12	144·25	136·03	87·69	88·12	47	43·17
26	1	1·45	0·25	1·13	0·13	0·19	—	
27	1	1·39	0·25	1·08	0·38	0·17	2	
28	3	1·34	0·75	1·03	—	0·15	—	
29	2	1·29	1·50	0·99	0·13	0·13	—	
30	—	1·24	0·50	0·95	0·25	0·12	—	0·30
	7	6·70	3·25	5·18	0·88	0·76	2	
31	—	1·20	—	0·91	0·13	0·11	—	
32	1	1·16	1·00	0·87	—	0·10	—	
33	—	1·12	1·00	0·83	0·25	0·09	—	
34	2	1·09	0·75	0·80	—	0·08	—	
35	—	1·05	0·25	0·77	—	0·07	—	
	3	5·61	3·00	4·18	0·38	0·51	0	
36	3	1·02	0·75	0·74	0·25	0·06	—	
37	2	0·99	0·25	0·72	0·13	0·06	—	
38	—	0·96	1·00	0·69	—	0·05	—	
39	—	0·93	0·50	0·67	0·13	0·05	1	
40	—	0·91	1·50	0·64	0·25	0·04	—	0·21
	5	4·81	4·00	3·46	0·75	0·26	1	
					Total to 40			
					89·70	89·65		
41	—	0·88	0·75	0·62			—	
42	—	0·86	0·25	0·60			—	
43	2	0·84	0·25	0·58			—	
44	2	0·81	0·25	0·56			—	
45	1	0·80	—	0·54			—	
	5	4·19	1·50	2·91			0	
46	—	0·78	0·25	0·53			—	
47	—	0·76	—	0·51			—	
48	—	0·74	1·50	0·50			—	
49	3	0·72	—	0·48			—	
50	—	0·71	—	0·47			—	0·15
	3	3·71	1·75	2·48			0	
Total to 50	180	176·14	157·75	154·24				

Table 3 (*continued*)
Values above 50

(1)   Total of 4 years: 51 (4), 52, 53, 54 (2), 57, 58 (2), 60 (3), 61, 64, 67, 73, 76 (2), 78, 84, 89, 96, 99, 109, 112, 120, 122, 129, 135, 141, 148, 149, 151, 154, 177, 181, 187, 190, 199, 211, 221, 226, 235, 239, 244, 246, 282, 305, 306, 333, 464, 560, 572, 589, 604, 743, 823, 2349 = 15,609 individuals of 240 species.

(2)   Average of 1 year: 51 (0·75), 52 (0·50), 53 (0·50), 54 (0·25), 58 (0·50), 60 (0·50), 61 (0·25), 64 (0·50), 65 (0·25), 69 (0·25), 73 (0·75), 75 (0·50), 76 (0·25), 77 (0·25), 80 (0·25), 82 (0·25), 83 (0·25), 87 (0·25), 88 (0·50), 90 (0·25), 93 (0·25), 99 (0·50), 100 (0·50), 104 (0·25), 105 (0·50), and the following all (0·25) each: 107, 109, 110, 111, 115, 126, 132, 138, 139, 141, 144, 145, 153, 159, 165, 173, 179, 197, 200, 201, 219, 223, 232, 275, 294, 323, 329, 603, 1799 = 3902 individuals of 176 species.

(3)   Average of ⅛ year: 79 (0·13), 109 (0·13) = 440 individuals of 89·875 species.

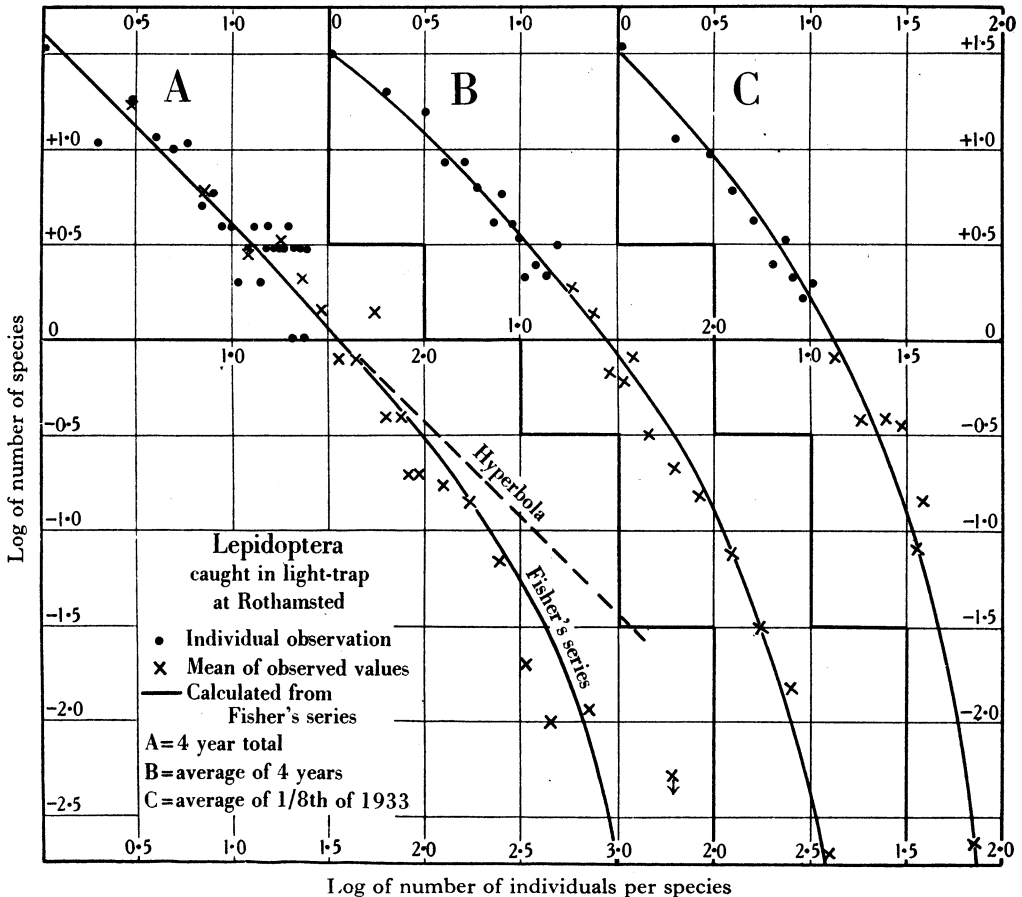(4)   Capsidae = 53, 80, 85, 158, 206, 237, 298 = 1414 individuals of 57 species.

Log of number of species (vertical axis)

Lepidoptera
caught in light-trap
at Rothamsted

•   Individual observation
×   Mean of observed values
—   Calculated from
Fisher's series
A = 4 year total
B = average of 4 years
C = average of 1/8th of 1933

Log of number of individuals per species

Fig. 2

Thus if N and S are known both n_1 and x can be calculated, and hence the whole series is known.

If N is large n_1 tends to become a constant value α, and Fisher shows that for all levels of sampling for the same population $x = \dfrac{N}{N + \alpha}$.

Applying these formulae to our Macrolepidoptera with 15,609 individuals of 240 species we find

$$\alpha = 40\cdot24, \quad n_1 = 40\cdot14, \quad x = 0\cdot997429.$$

The calculated values for n_2, n_3, etc., are shown in Table 3 opposite the observed figures and also by curves in Figs. 1 A and 2 A. In the latter the log of the frequency is plotted against the log of the number of individuals per species.

It will be seen that for both common and rare species the calculated values are very close indeed to the observed. The calculated number of species with one individual is slightly larger (40·14) than the observed (35). By calculation there should be 116·9

species with 10 or fewer individuals and the observed number was 115. The close resemblance at higher frequencies is best seen in Fig. 2 A.

If, instead of adding the four years together, we take the average number of individuals in each year and the average number of species with 1, 2, 3, etc., individuals we find the observed results in the trap to be as in the second column of Table 3. There are 3902 individuals of 176 species. Calculations from these figures show that $x = 0.990356$ and $n_1 = 37.55$. The calculated series is shown up to 50 individuals per species in Table 3 and diagrammatically in Fig. 2 B. The calculated n_1 is slightly smaller than the observed.
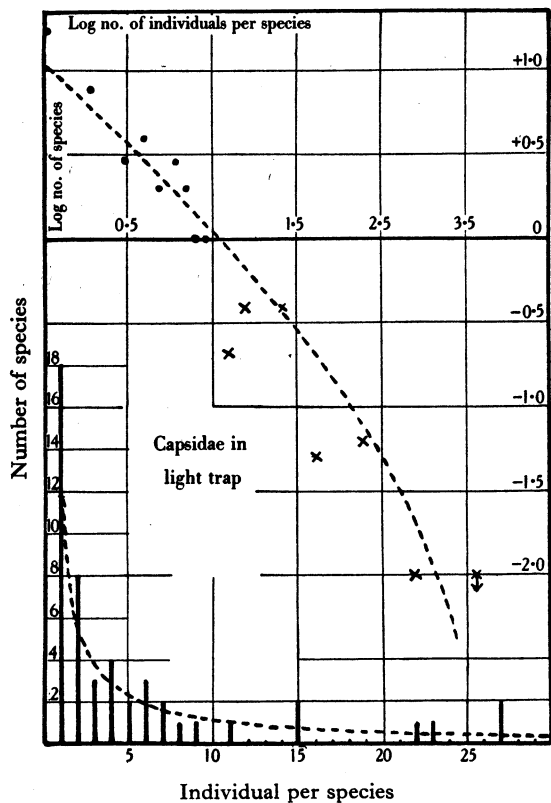

Fig. 3

A still smaller sample was obtained by dividing the catches in the year 1933 into 8 samples, the first including those insects caught in the 1st...9th...17th day, the second on the 2nd...10th...18th day, etc. The average for the 8 samples is shown in the third column of Table 3, together with the calculated figures. The number of individuals was 440, the number of species 89.9, from which $n_1 = 31.77$ and $x = 0.9278$. The calculated n_1 is slightly smaller than the observed. Results are shown diagrammatically in Figs. 1 B and 2 C. The close resemblance in both the two series of the observed and calculated results is very striking.

Thus in the Macrolepidoptera caught in the light-trap at Harpenden samples as large as 15,609 individuals and as small as 440 individuals both agree very closely with the results calculated from Fisher's series.

The only other group of insects captured in the light-trap in which the majority of the specimens were identified was the family Capsidae of the Heteroptera. These were identified by D. C. Thomas (1938). In the four years 1414 individuals of 57 species were captured. The observed and calculated results are shown in the fourth column of Table 3 and in Fig. 3. The calculated n_1 is 11.87 and $x = 0.9916$.

It will be seen that the fit is not so good as in the Lepidoptera. The observed number of species represented by only one individual is 18 and the calculated less than 12. In general there are rather more of the rarer species than the calculated series indicates, and fewer of the commoner species.

Table 4 shows a summary of all the various samples taken from the Lepidoptera and the Capsidae. It will be seen that there is a slight general tendency for the calculated value of n_1 to be below the observed. In the Macrolepidoptera for one or more years this is not so obvious, but it is distinct in the one-eighth year samples. In the Noctuidae only there are four results below and two practically equal to the observed figure; but in the Capsidae there are five below and only one equal. This requires further investigation.
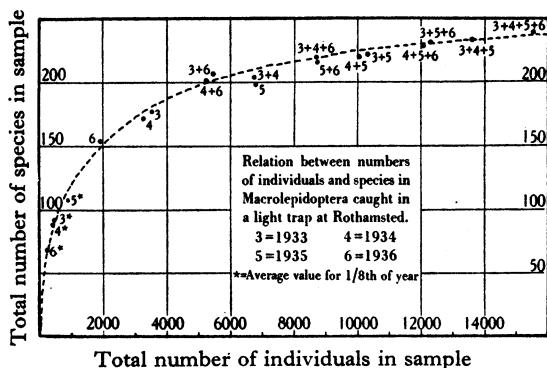
Fig. 4

The relation between the number of species and total number of individuals in the sample

In Tables 4 and 5 are shown a number of samples of different size taken from the light-trap captures with the number of individuals in the sample and the number of species represented.

Fig. 4 shows diagrammatically the number of species in relation to the number of individuals for all the different samples of Lepidoptera. It will be seen that there is a very regular relation between the observed numbers. This form of diagram is, however, not very suitable for showing large samples,

Table 4. *Calculations of x, α and n_1 for various light-trap samples*

	Individuals	Species	x	α	Species with 1 individual — Calc.	Species with 1 individual — Obs.
			Macrolepidoptera			
1933	3,540	178	0·98903	39·15 ± 1·45	38·62	32
1934	3,275	172	0·98834	38·64	38·19	34
1935	6,817	198	0·994425	38·19	37·94	37
1936	1,977	154	0·980595	39·05	38·30	50
4 years: Total	15,609	240	0·997429	40·24 ± 1·06	40·14	35
Average	3,902	176	0·990380	37·90	37·54	39·75
$\frac{1}{8}$ of 1933	561	105	0·93638	38·12	35·69	38
,,	581	104	0·94028	36·90	34·70	39
,,	458	86	0·93609	32·27	29·27	34
,,	357	84	0·91157	34·63	31·57	41
,,	344	80	0·91306	32·75	29·90	35
,,	282	75	0·8941	33·40	29·86	28
,,	503	101	0·92968	38·05	35·37	38
,	427	81	0·93725	29·26	27·42	36
Average of 8 × $\frac{1}{8}$	440	89·5	0·9278	34·24 ± 2·67	31·77	35·63
			Noctuidae only			
1933	1,636	84	0·9887	18·70 ± 1·02	18·49	18
1934	1,894	73	0·9922	14·89	14·77	14
1935	5,413	94	0·997	16·29	16·23	21
1936	1,362	71	0·9886	15·71	15·53	20
4 years: Total	10,304	112	0·9983	17·56 ± 0·63	17·52	20
Average	2,576	80	0·994	15·55	15·46	18·25
			Capsidae			
1933	341	26	0·981	6·60 ± 0·71	6·5	10
1934	479	38	0·98	9·77	9·6	17
1935	446	24	0·988	5·41	5·45	5
1936	148	23	0·95	7·79	7·40	9
4 years: Total	1,414	57	0·9916	11·98 ± 0·76	11·88	18
Average	354	27·75	0·98047	7·04	6·90	10·25

Table 5. *Number of individuals and species in different sized samples from the light-trap*

	Macrolepidoptera Indiv.	Macrolepidoptera Sp.	Noctuidae only Indiv.	Noctuidae only Sp.	Capsidae Indiv.	Capsidae Sp.
1933	3,540	178	1,636	84	341	26
1934	3,275	172	1,894	73	479	38
1935	6,817	198	5,413	94	446	24
1936	1,977	154	1,362	71	148	23
Average of single year	3,902	176	2,576	80·5	353·5	27·75
1933 + 1934	6,815	204	3,530	94	820	46
1933 + 1935	10,357	222	7,049	106	787	37
1933 + 1936	5,517	206	2,998	95	489	32
1934 + 1935	10,092	220	7,307	100	925	45
1934 + 1936	5,252	200	3,256	88	627	48
1935 + 1936	8,794	216	6,775	101	594	31
Average of years in pairs	7,804	211·3	5,152	97·3	707	39·8
1933 + 1934 + 1935	13,632	234	8,943	109	1,266	52
1933 + 1934 + 1936	8,792	219	4,892	299	968	54
1933 + 1935 + 1936	12,334	232	8,411	109	935	42
1934 + 1935 + 1936	12,069	229	8,669	105	1,073	51
Average of years in threes	11,706	228·5	7,728	105·5	1,060·5	49·75
All 4 years	15,609	240	10,305	112	1,414	57

and in Fig. 5 (heavy line) the same data are shown with the number of individuals in the sample expressed as a logarithm.

According to Fisher's theory, any population should have a constant value of α for samples of any size taken from it under identical conditions. Thus it follows that for any population, if the value of α is obtained from one sample, the values of

$$x \left(= \frac{N}{N+\alpha} \right)$$

and hence of S can be calculated for other samples of different sizes from the same population. In

Relation between number of individuals and number of species in a sample
— Observed
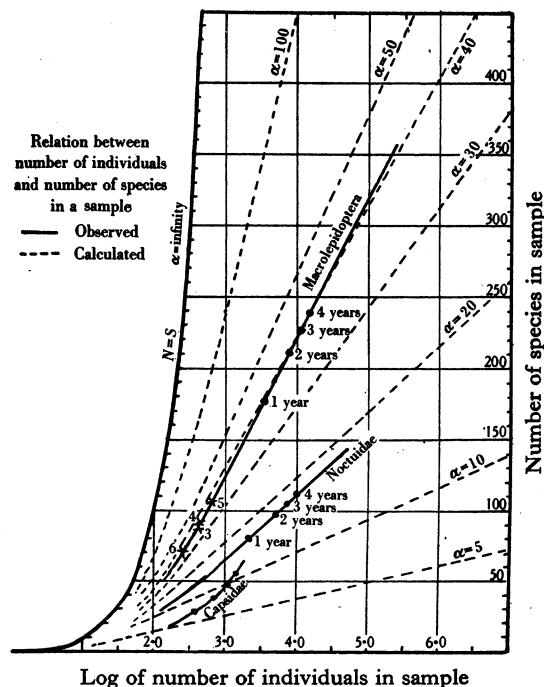---- Calculated

Log of number of individuals in sample

Fig. 5

Fig. 5 there are shown as dotted lines typical calculated relations between the number of individuals and number of species for different values of α. It will immediately be seen that the observed values for the Lepidoptera fit very closely to the calculated relation for $\alpha = 40$. The slight difference between the observed and calculated figures will be discussed below.

It will also be seen that the observed values for the Noctuidae alone fit very closely to a value of $\alpha = 17.5$, but that the observed values for the Capsidae cut rather rapidly across the α lines giving too few species in small samples and too many in large samples (when they are taken, as in the light-trap, in more than one year). The various calculated values of α in the different samples together with their standard deviation as calculated from Fisher's formula are given in Table 4. It is suggested that the para-
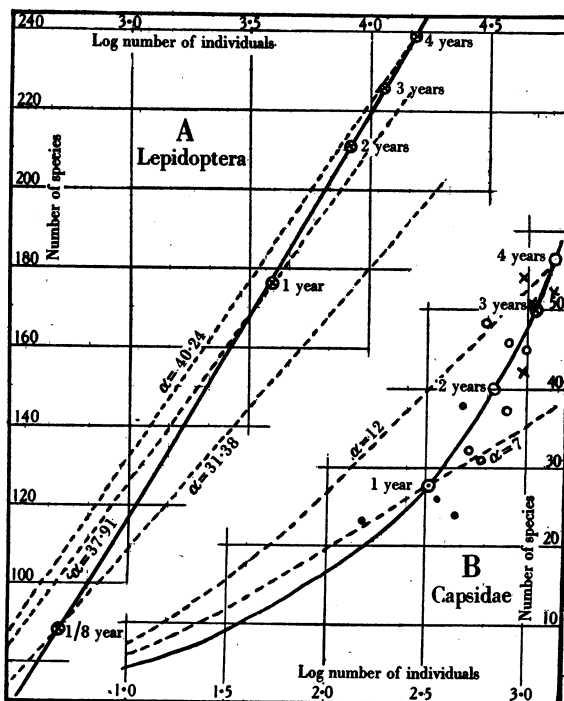
meter α should be known as the 'index of diversity' of the population.

The differences between the observed and calculated values for Lepidoptera and for Capsidae are shown diagrammatically on a larger scale in Fig. 6. In the Lepidoptera (Fig. 6A) it will be seen that more species are obtained in a sample spread over several years than would be expected in a sample of the same size taken in a single year. Thus when a sample of 15,609 (log 4·19) Lepidoptera was captured in four years 240 species were represented. If the same number had been caught in one year it will

Fig. 6

be seen (by finding where the $\alpha = 37.91$ line cuts log 4·19) that only about 226 species would have been expected. Otherwise about 14 extra species have been captured owing to the new biological conditions introduced by sampling over a longer period. The same effect on an even larger scale is obvious between samples captured throughout a whole year and those captured in one-eighth of the year. Thus the departure of the observed from the calculated results is immediately explicable by the alteration of conditions introduced by spreading sampling over a longer period of time.

The Capsidae (Fig. 6B) show the same effect in a more extreme form. With 1414 individuals caught in four years 57 species were obtained, whereas if the same-sized sample had been taken in one year we would only have expected about 36 species. There is thus a considerable increase of species by spreading

the cat h over a longer period. This is indicated also by a fact that I have already pointed out (Williams, 1939, p. 96) that in the Capsidae 50% of the 57 species occurred only in a single year, as compared with a figure of only 19% in the Macrolepidoptera. The Capsidae fauna therefore is much more variable from year to year than is the Lepidoptera.

Since with high number of N the relation between the number of species and log number of individuals becomes practically a straight line (both theoretically and observationally) it is possible to extrapolate the observed curve for the Macrolepidoptera with some interesting and suggestive results. In the first place it should be noted that the number of British species in the families of Macrolepidoptera dealt with is 681, and the number of species already recorded for Hertfordshire (where the trap was situated) was 461.

Thus in four years we have captured approximately 16,000 individuals representing 240 species. After ten years we would have captured approximately 40,000 individuals which would contain about 275 species. After 50 years the 200,000 individuals would

The seasonal changes of α in the Macrolepidoptera

If the Macrolepidoptera captured in the light-trap during the four years are tabulated month by month the numbers of individuals and species in each month (excluding the five winter months when numbers are too small) are shown in Table 6 together with the approximate value of α. The changes in α are shown in Fig. 7.

There is a regular seasonal change which is almost identical in each year; the value of α rises from a low value in April to a maximum in July and back to a low value in October. There is a very much greater difference between the α values for two different months than there is for the same month in two different years. For example, the number of insects caught in July 1935 was almost seven times as great as the number in July 1936, and yet the values of α are almost identical; but rather more insects in August 1936 than in July gives a considerably smaller value, and an almost identical number of insects in

Table 6. *Number of individuals and of species of Macrolepidoptera in each of the summer months of the four light-trap years, with the approximate values of α*

	1933			1934			1935			1936		
	N	S	α	N	S	α	N	S	α	N	S	α
Apr.	70	12	4·0	55	8	2·5	24	9	5·0	16	5	3·0
May	318	38	11·0	131	27	10·0	119	26	10·0	38	16	10·0
June	693	73	20·5	518	53	14·5	1178	67	15·5	434	45	13·0
July	987	90	24·0	1059	97	25·5	3136	119	24·5	449	68	22·5
Aug.	701	59	15·5	591	51	13·5	1920	84	18·0	491	57	16·5
Sept.	389	30	7·5	528	31	7·0	647	36	8·0	448	33	8·0
Oct.	220	14	3·0	152	14	3·5	171	18	5·0	59	14	5·0

include 340 species. After 100 years the 400,000 would include 365 species or only 25 new species in 50 years.

To get the Hertfordshire fauna (461 species) we would have to collect over three million insects, and this would take, at the present rate, nearly 1000 years. When this level was reached only one new species would be added in 20 years' trapping. To get the British fauna (681 species) we would have to trap about 1800 million insects which would take nearly half a million years!

Doubling the number of insects caught (and hence the time of trapping) at any level, except for very small samples, always adds about 30 species to the total.

These figures are based on the curve allowing for captures of 4000 insects per year. If the whole number were caught in one year (assuming that to be possible) the number of species would be less in each case. So that 400,000 individuals would give 350 species (instead of 368), the Hertfordshire fauna would require 8 million insects, and the British fauna 3200 million.
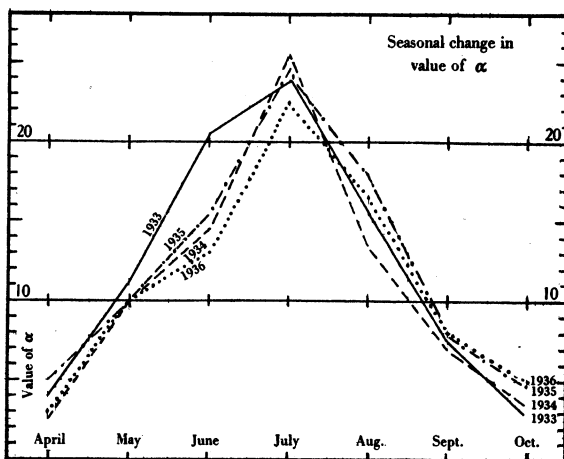

Fig. 7

September 1936 gives a value of α only about one-third as great.

There is no doubt whatever that there is a seasonal

change in α or in the 'richness of species' quite independent of the seasonal change in numbers of insects.

Value of α for a single night in a month

In the month of July 1935, which was a month of high catches of Lepidoptera in the light-trap, 2,586 Noctuidae were captured, belonging to 56 species. This gives α (for Noctuidae only) = 10·09.

The following were the captures on seven selected nights in the month:

Date	N	$\log N$	S
1	429	2·63	18
6	143	2·16	10
11	394	2·60	18
16	11	1·04	7
21	47	1·67	17
26	64	1·80	11
31	53	1·72	10
Mean		1·95	13

From these figures $\alpha = 4\cdot2$. Thus it is considerably smaller for a short period of sampling than for a longer period. This would be expected, as the variety of species available for catching on a single night must be less than that on a series of different nights.

Comparison of results from two light-traps a short distance apart

During the autumn of 1933 two light-traps were kept running simultaneously from 8 August to 31 October. One was in the fields at Rothamsted (the standard trap used in all the previous calculations) and the other was about 400 yards away on the roof on the Entomology building at a height of about 35 ft. from the ground. This overlooked a more varied environment with gardens and mixed vegetation.

The captures of Noctuidae in the two traps were as follows:

1933	Field trap		Roof trap	
	N	S	N	S
8-31 Aug.	419	26	706	36
Sept.	355	18	593	32
Oct.	155	9	557	19
	929	40	1856	58
	$\alpha = 8\cdot51 \pm 0\cdot7$		$11\cdot38 \pm 0\cdot7$	

The difference between the two values of α is 2·8 and the standard error of the difference approximately 1·0, so that difference is probably significant. The roof trap has caught a larger number of individuals but a still larger number of species. This is probably due, as mentioned above, to the considerably more varied vegetation that it overlooked.

The effect on number of species of increasing the size of a sample

Fisher has shown that

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

If the sample is large so that N is large compared with α we can neglect the 1 in comparison with N/α. Hence for large samples

$$S = \alpha \log_e \frac{N}{\alpha},$$

hence

$$S_{2N} - S_N = \alpha \left(\log_e \frac{2N}{\alpha} - \log_e \frac{N}{\alpha}\right) = \alpha \log_e 2.$$

So the number of species added to a large sample by doubling it is $\alpha \log_e 2x = 0\cdot693\alpha$. Similarly, if the sample were increased 10 times the number of new species added $= \alpha \log_e 10 = 2\cdot3\alpha$ approximately. On the other hand, if the size of the sample is multiplied by e ($= 2\cdot7183$) the additional number of species is $\alpha \log_e e$, which is equal to α.

For example, with the Macrolepidoptera in the light-trap $\alpha = $ approximately 40, so doubling a large catch will at any level add approximately 28 new species, multiplying it by 10 will add approximately 92 species.

To add one new species to a large sample $\alpha \log_e z$ must equal 1, when z is the factor by which the sample must be multiplied, i.e.

$$\log_e z = 1/\alpha;$$

for example, if $\alpha = 40$ as in the Macrolepidoptera above

$$\log_e z = 0\cdot025, \quad z = 1\cdot0255;$$

hence the number of insects must be increased by just over 1/39 of the original sample to add one additional species.

Diagram showing the interrelation of N, S, α and the standard error of α

From the formulae provided by Prof. Fisher α and the error of α can be calculated if N and S are known, and similarly S can be calculated if N and α are known. These calculations can be made to any number of decimal places. For many purposes, however, a close approximation to α and its error is all that is necessary, and if this can be obtained rapidly without calculation so much the better.

For this purpose Fig. 8 has been drawn up. In this the vertical axis is the number of species and the horizontal the log of the number of individuals in the sample. A large number of values of α have been inserted as diagonals, with the result that if any two of these three factors is known the third can be easily estimated. Thus with a sample of 1000 individuals ($\log = 3\cdot0$) containing 60 species $\alpha = 14$. Or with $\alpha = 25$ a sample of 10,000 individuals should
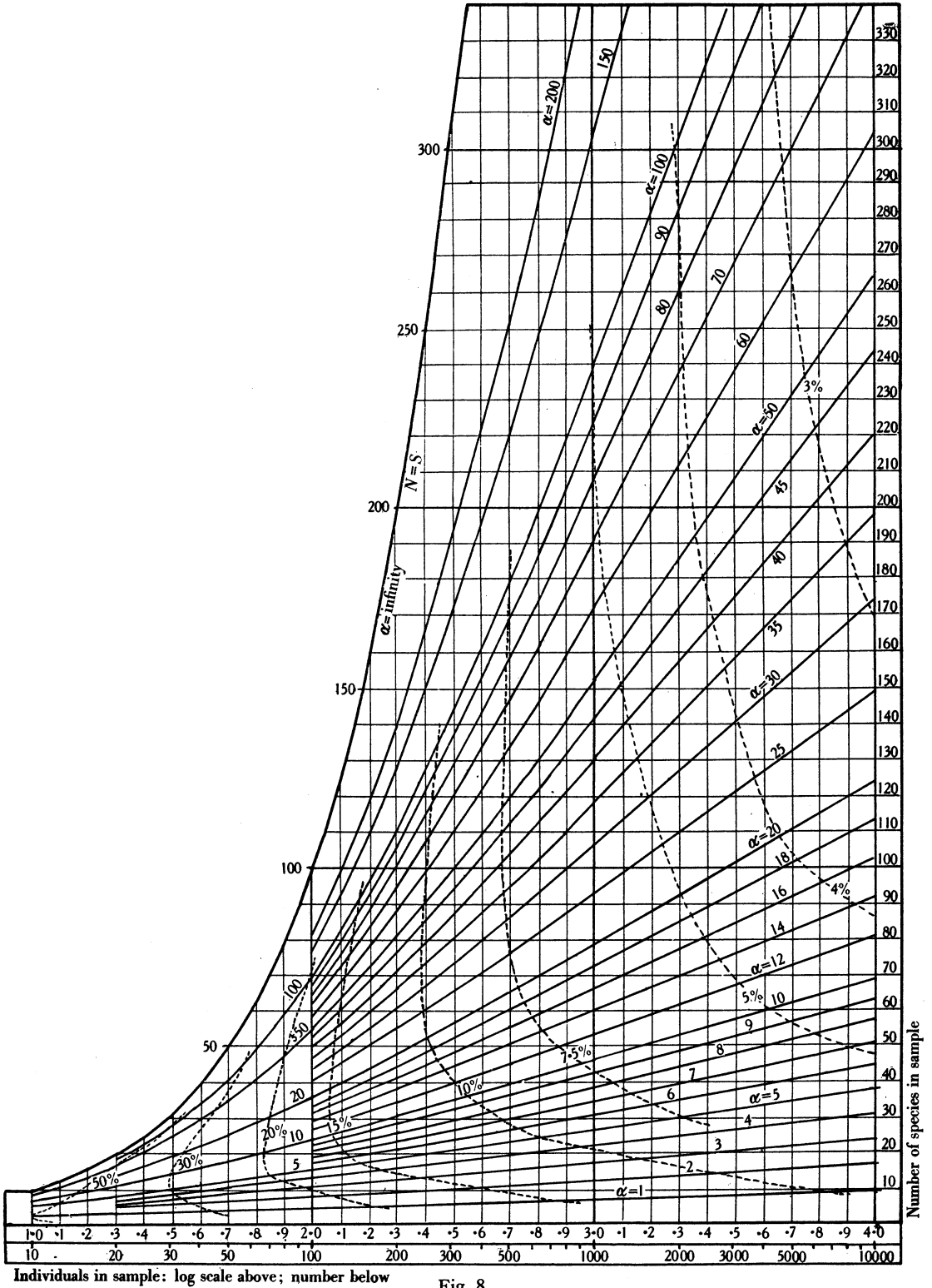
Fig. 8

Individuals in sample: log scale above; number below

Table 7. *Values of S for different combinations of N and α*

α \ N	10	20	50	100	200	500	1000	2000	5000	10,000	100,000
1	2·40	3·00	3·93	4·62	5·30	6·22	6·91	7·60	8·35	9·21	—
2	3·58	4·80	6·52	7·86	9·23	11·05	12·43	13·82	15·65	17·03	—
3	4·40	6·08	8·62	10·61	12·64	15·37	17·44	19·51	22·26	24·34	—
4	5·01	7·17	10·41	13·03	15·73	19·34	22·10	24·87	28·53	31·30	—
5	5·49	8·04	11·99	15·03	18·57	23·08	26·52	29·97	34·54	38·01	49·57
6	5·89	8·80	13·40	17·23	21·21	26·60	30·73	34·87	40·36	44·51	—
7	6·22	9·45	14·68	19·09	23·71	29·98	34·78	39·61	46·01	50·86	—
8	6·49	10·02	15·85	20·82	26·06	33·08	38·79	44·21	51·51	57·06	—
9	6·72	10·53	16·92	22·45	28·30	36·32	42·47	48·67	56·90	63·14	—
10	6·93	10·98	17·92	23·98	30·45	39·32	46·15	53·03	62·17	69·08	92·10
12	7·27	11·76	19·70	26·80	34·46	45·02	53·21	61·46	72·42	80·72	—
14	7·55	12·45	21·28	29·36	38·21	50·44	59·96	69·57	82·33	92·02	—
15	7·67	12·71	21·99	30·42	39·93	53·03	63·22	73·50	87·18	97·56	—
16	7·76	12·98	22·67	31·70	41·65	55·58	66·16	77·38	91·97	103·02	—
18	7·96	13·45	23·92	33·84	44·89	60·48	72·63	84·76	101·36	113·80	—
20	8·11	13·86	25·06	35·84	47·96	65·16	78·64	92·30	110·52	124·34	170·35
25	8·43	14·70	27·48	40·25	54·93	76·13	92·85	109·87	132·58	149·86	—
30	8·61	15·33	29·43	43·98	60·84	86·16	106·05	126·45	153·66	174·36	—
35	8·82	15·82	31·08	47·25	66·64	95·52	118·55	142·21	173·92	198·07	—
40	8·92	16·24	32·44	50·12	71·68	104·12	130·32	157·28	193·44	221·04	—
45	9·04	16·52	33·62	52·65	76·23	112·23	141·53	172·72	212·36	243·36	—
50	9·12	16·85	34·65	54·95	80·45	119·90	152·25	185·7	230·75	265·15	380·07
60	9·27	17·22	36·36	58·86	87·96	133·98	172·32	212·1	266·04	307·32	—
70	9·35	17·64	37·73	62·16	94·50	146·79	191·03	237·1	299·81	347·83	—
80	9·42	17·84	38·80	64·88	100·24	158·48	208·24	260·6	330·80	386·88	—
90	9·48	18·09	39·78	67·23	105·30	169·20	224·46	283·1	363·15	424·71	—
100	9·53	18·20	40·60	69·31	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	9·76	19·06	44·60	81·10	138·6	250·6	258·11	479·6	651·6	786·4	—

contain about 149 species. Or in a population with α = 40, 180 species would be produced by a sample of 1000 individuals.

Further, by following along one of the α curves it is possible to find the number of species in samples of different sizes from the same population. Thus if a sample of 1000 individuals for a population produced 73 species, then a sample of 10,000 from the same population should produce about 112 species, but a sample of 100 individuals would only contain about 34 species (α = 18). In addition, on the diagram the standard error (as a percentage of the value of α) has been indicated by dotted contours. Thus a sample containing 100 individuals and 24 species gives α = 10 ± about 17 %, say 10 ± 1·7; a sample of 1000 individuals and 46 species gives α = 10 ± 0·75 (i.e. 7·5 %); while a sample of 10,000 individuals containing 69 species gives α = 10 ± 0·45 (4·5 %).

Thus the size of sample necessary to give α to any required degree of accuracy can rapidly be obtained from the figure.

Table 7 shows values of S correct to two decimal places for various combinations of α and N, and Table 8 shows the standard error of α to three significant figures for various values of N and α.

Table 8. *Standard error of α for different values of N and α*

α \ N	10	100	1000	10,000	100,000
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

REFERENCES

Corbet, A. S. (1941). 'The distribution of butterflies in the Malay Peninsula.' Proc. R. Ent. Soc. Lond. (A), 16: 101–16.

Fisher, R. A. (1941). 'The negative binomial distribution.' Ann. Eugen., Lond., 11: 182–7.

Thomas, D. C. (1938). 'Report on the Hemiptera-Heteroptera taken in the light trap at Rothamsted Experimental Station, during the four years 1933–36.' Proc. R. Ent. Soc. Lond. (A), 13: 19–24.

Williams, C. B. (1939). 'An analysis of four years captures of insects in a light trap. Part I. General survey; sex proportion; phenology; and time of flight.' Trans. R. Ent. Soc. Lond. 89: 79–131.

Williams, C. B. (1940). 'An analysis of four years captures of insects in a light trap. Part II. The effect of weather conditions on insect activity; and the estimation and forecasting of changes in the insect population.' Trans. R. Ent. Soc. Lond. 90: 227–306.

PART 3. A THEORETICAL DISTRIBUTION FOR THE APPARENT ABUNDANCE OF DIFFERENT SPECIES

By R. A. Fisher

(1) The Poisson Series and the Negative Binomial distribution

In biological sampling it has for some time been recognized that if successive, independent, equal samples be taken from homogeneous material, the number of individuals observed in different samples will vary in a definite manner. The distribution of the number observed depends only on one parameter, and may be conveniently expressed in terms of the number *expected*, m, in what is known as the *Poisson Series*, given by the formula

$$e^{-m}\frac{m^n}{n!}. \qquad (1)$$

Here n is the variate representing the number observed in any sample, m is the parameter, the number expected, which is the average value of n, and need not be a whole number. Obviously, m will be proportional to the size of the sample taken, and to the density of organisms in the material sampled. For example, n might stand for the number of bacterial colonies counted on a plate of culture medium, m for the average number in the volume of dilution added to each plate. The formula then gives the probability of obtaining n as the number observed.

The same frequency distribution would be obtained for the numbers of different organisms observed in one sample, if all were equally frequent in the material sampled.

If the material sampled were heterogeneous, or if unequal samples were taken, we should have a mixture of distributions corresponding to different values of m. The same is true of the numbers of different organisms observed in a single sample, if the different species are not equally abundant.

An important extension of the Poisson series is provided by the supposition that the values of m are distributed in a known and simple manner. Since m must be positive, the simplest supposition as to its distribution is that it has the Eulerian form (well known from the distribution of χ^2) such that the element of frequency or probability with which it falls in any infinitesimal range dm is

$$df = \frac{1}{(k-1)!}p^{-k}m^{k-1}e^{-m/p}dm. \qquad (2)$$

If we multiply this expression by the probability, set out above, of observing just n organisms, and integrate with respect to m over its whole range from 0 to ∞, we have

$$\int_0^\infty \frac{1}{(k-1)!}p^{-k}m^{k-1}e^{-m/p}e^{-m}\frac{m^n}{n!}dm,$$

which, on simplification, is found to have the value

$$\frac{(k+n-1)!}{(k-1)!\,n!}\frac{p^n}{(1+p)^{k+n}}, \qquad (3)$$

which is the probability of observing the number n when sampling from such a heterogeneous population. Since this distribution is related to the negative binomial expansion

$$\left(1-\frac{p}{1+p}\right)^{-k}=\sum_{n=0}^{\infty}\frac{(k+n-1)!}{(k-1)!\,n!}\left(\frac{p}{1+p}\right)^n,$$

it has become known as the *Negative Binomial distribution*. It is a natural extension of the Poisson series, applicable to a somewhat wider class of cases.

The parameter p of the negative binomial distribution is proportional to the size of the sample. The expectation, or mean value of n, is pk. The second parameter k measures in an inverse sense the variability of the different expectations of the component Poisson series. If k is very large these expectations are nearly equal, and the distribution tends to the Poisson form. If heterogeneity is very great k becomes small and approaches its limiting value, zero. This second parameter, k, is thus an intrinsic property of the population sampled.

(2) The limiting form of the negative binomial, excluding zero observations

In many of its applications the number n observed in any sample may have all integral values including zero. In its application, however, to the number of representatives of different species obtained in a collection, only frequencies of numbers greater than zero will be observable, since by itself the collection gives no indication of the number of species which are not found in it. Now, the abundance in nature of different species of the same group generally varies very greatly, so that, as I first found in studying Corbet's series of Malayan butterflies, the negative binomial, which often fits such data well, has a value of k so small as to be almost indeterminate in magnitude, or, in other words, indistinguishable from zero. That it is not really zero for collections of wild species follows from the fact that the total number of species, and therefore the total number not included in the collection, is really finite. The real situation, however, in which a large number of species are so rare that their chance of inclusion is small, is well represented by the limiting form taken by the negative binomial distribution, when k tends to zero.

The limiting value $k=0$ cannot occur in cases where the frequency at zero is observable, for the

distribution would then consist wholly of such cases. If, however, we put $k=0$ in expression (3), write x for $p/(p+1)$, so that x stands for a positive number less than unity, varying with the size of the sample, and replace the constant factor $(k-1)!$ in the denominator, by a new constant factor, α, in the numerator, we have an expression for the expected number of species with n individuals, where n now cannot be zero,

$$\frac{\alpha}{n} x^n. \tag{4}$$

These two relationships enable the series to be fitted to any series of observational data, for if S is the number of species observed, and N the number of individuals, the two equations

$$S = -\alpha \log_e (1-x), \quad N = \alpha x/(1-x),$$

are sufficient to determine the values of α and x. The solution of the equations is, however, troublesome and indirect, so that to facilitate the solution in any particular case I have calculated a table (Table 9) from which, given the common logarithm

Table 9. *Table of $\log_{10} N/\alpha$ in terms of $\log_{10} N/S$, for solving the equation*

$$S = \alpha \log_e \left(1 + \frac{N}{\alpha}\right), \text{ given } S \text{ and } N$$

$\log_{10} N/S$	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

The total number of species expected is consequently

$$\sum_{n=1}^{\infty} \frac{\alpha}{n} x^n = -\alpha \log_e (1-x),$$

so that our distribution is related to the algebraic expansion of the logarithm, as the negative binomial distribution is to the binomial expansion. Next, it is clear that the total number of individuals expected is

$$\sum_{n=1}^{\infty} \alpha x^n = \frac{\alpha x}{1-x}.$$

of N/S, we may obtain that of N/α. Five-figure logarithms are advisable, such as those in *Statistical Tables*. If x be eliminated from the two equations, it appears that

$$N = \alpha (e^{S/\alpha} - 1), \quad S = \alpha \log_e \left(1 + \frac{N}{\alpha}\right),$$

and

$$\frac{N}{S} = (e^{S/\alpha} - 1) \div S/\alpha,$$

from which Table 9 has been constructed.

(3) *Fitting the series*

The use of the table is shown, using Williams's extensive data for the Macrolepidoptera at Harpenden (total catch for four years). Symbols + and − are used to indicate numbers to be added and subtracted respectively.

Symbol	Number	Common logarithm
S	240	−2·38021
N	15609	+4·19338
N/S	—	1·81317

	log (N/S)	log (N/α)
From the table	−1·81	−2·58484
	+1·82	+2·59684
Difference	0·01	0·01200
Proportional parts	0·00317	0·00380
	1·81317	2·58864

Then	Number	Common logarithm
N/α	—	−2·58864
N	—	+4·19338
α	40·248	1·60474

For constructing the distribution we should then calculate

$$x = \frac{N}{N+\alpha} = \frac{15609}{15649 \cdot 248} = 0 \cdot 9974281.$$

The quantity α is independent of the size of sample, and is proportional to the number of species of the group considered, at any chosen level of abundance, relative to the means of capture employed. Values of α from different samples or obtained by different methods of capture may therefore be compared as a measure of richness in species. To this end we shall need to know the sampling errors by which an estimate of α may be affected.

(4) *Variation in parallel samples*

Whatever method of capture may be employed, it is to be expected that a given amount of activity devoted to it, e.g. a given number of hours exposure of a light-trap, or a given volume of sea water passed through a plankton filter, will yield on different occasions different numbers of individuals and of species, and, consequently, varying estimates of α. The amount of variation of these kinds attributable to chance must form the basis of all conclusions as to whether variations beyond chance have occurred in the circumstances in which two or more samples were made.

In strictly parallel samples, i.e. equivalent samling processes applied to homogeneous material, the numbers caught of each individual species will be distributed in a Poisson series, and it easily follows that the same is true of the aggregate number, N, of all species. Since N is a large number of hundreds or thousands, this is equivalent to N being normally distributed with a variance equal to its mean, so that to any observed value N we may attach a standard error (of random sampling) equal to $\pm \sqrt{N}$.

For the variation of S we must obtain the distribution of species according to the number m expected in the sample; modifying expression (2) in the same way as (3) has been modified, this is found to be

$$\alpha e^{-\alpha m/N} dm/m. \tag{5}$$

The probability of missing any species is e^{-m}, so that the contribution to the sampling variance of S due to any one species being sometimes observed and sometimes not, is

$$e^{-m}(1 - e^{-m}).$$

Multiplying this by the frequencies in (5) and integrating over all values of m, we have

$$\alpha \int_0^\infty e^{-m(N+\alpha)/N}\left(1 - \frac{m}{2} + \frac{m^2}{6} - \ldots\right) dm = \alpha \log_e \left(\frac{2N+\alpha}{N+\alpha}\right),$$

which is the sampling variance of S. For large samples this is approximately $(0 \cdot 6931)\,\alpha$.

Variations of S and N in parallel samples are not, however, independent. When present, a species must contribute on the average $m/(1-e^{-m})$ individuals, which exceeds the expectation in all samples by

$$\frac{me^{-m}}{1 - e^{-m}},$$

and as the frequency of occurrence is $1 - e^{-m}$, each species must contribute $m \cdot e^{-m}$ to the covariance of S and N. The covariance is thus found to be

$$\frac{\alpha N}{N+\alpha}.$$

From these three values it is possible by standard methods to find the sampling variance of S in samples having a given number of specimens N, which is

$$V(S), \text{ given } N, = \alpha \log_e \frac{2N+\alpha}{N+\alpha} - \frac{\alpha^2 N}{(N+\alpha)^2}$$

and, the variance of α,

$$V(\alpha) = \frac{\alpha^3 \left\{(N+\alpha)^2 \log_e \frac{2N+\alpha}{N+\alpha} - \alpha N\right\}}{(SN + S\alpha - N\alpha)^2}.$$

We may, therefore, complete the example of the last section by calculating the standard error of α. Using the values obtained, the variance comes to 1·1251, of which the square root is 1·0607.

The estimate obtained for α, 40·248, has, therefore, a standard error of 1·0607, available for comparison with like estimates.

(5) *Test of adequacy of the limiting distribution*

From the manner in which the distribution has been developed it appears that we never have theoretical grounds for supposing that k is actually zero;

but, on the contrary, must generally suppose that in reality it has a finite, though perhaps a very small, value. Our reasons for supposing this small value to be negligible must always be derived from the observations themselves. It is, therefore, essential to be able to test any body of data in respect to the possibility that in reality some value of k differing significantly from zero might fit the data better than the value zero actually assumed.

The most sensitive index or score by which any departure of the series of frequencies observed from those expected can be recognized, is found by the general principles of the Theory of Estimation, as, for example, in the author's *Statistical Methods for Research Workers*, to be

$$S\left\{a_n\left(1+\tfrac{1}{2}+\tfrac{1}{3}+\ldots+\frac{1}{n-1}\right)\right\},$$

when a_n is a number of species observed with n individuals in each. If the values of a_n conformed accurately with expectation, the total score would be equal to

$$\frac{S^2}{2\alpha}.$$

If, on the contrary, the series were better fitted by a negative binomial with a value of k differing from zero, we should expect the difference

$$S\left\{a_n\left(1+\tfrac{1}{2}+\tfrac{1}{3}+\ldots+\frac{1}{n-1}\right)\right\}-\frac{S^2}{2\alpha}$$

to show a positive discrepancy.

Applying this test to Williams's distribution for 240 species of Macrolepidoptera, one finds, after a somewhat tedious calculation,

$$S\left\{a_n\left(1+\tfrac{1}{2}+\tfrac{1}{3}+\ldots+\frac{1}{n-1}\right)\right\} \qquad 724\cdot86$$

$$\frac{S^2}{2\alpha} \qquad\qquad\qquad 715\cdot57$$

Difference $+9\cdot29$

The series, therefore, shows a deviation in the direction to be expected for the negative binomial, though apparently quite a small one. In order to test the significance of such discrepancies, I give in Table 10, for the same range of observable values of the average number of specimens in each species N/S, the values of i/S, where i is the quantity of information, in respect of the value of k, which the data supply.

Table 10. *The amount of information respecting k, supposed small, according to the numbers of individuals (N) and species (S) observed*

$\log_{10} N/S$	i/S	$\log_{10} N/S$	i/S
0·4	0·1971		
0·5	0·2882		
0·6	0·3914	2·1	3·1047
0·7	0·5054	2·2	3·3606
0·8	0·6295	2·3	3·6260
0·9	0·7639	2·4	3·9009
1·0	0·9076	2·5	4·1854
1·1	1·0608	2·6	4·4791
1·2	1·2232	2·7	4·7825
1·3	1·3950	2·8	5·0954
1·4	1·5762	2·9	5·4178
1·5	1·7665	3·0	5·7498
1·6	1·9661	3·1	6·0912
1·7	2·1751	3·2	6·4421
1·8	2·3934	3·3	6·8026
1·9	2·6211	3·4	7·1726
2·0	2·8582	3·5	7·5521

Entering the table with our value 1·81317 for $\log_{10} N/S$ we have $i/S=2\cdot4656$, or $i=591\cdot7$. This quantity may now be used for two purposes. In the first place it is the sampling variance of the discrepancy observed, so that, taking its square root, the standard error is found to be 24·33. This suffices to test the significance of the discrepancy, since $9\cdot29\pm24\cdot33$ is clearly insignificant.

If, on the contrary, a significant discrepancy had been found, an estimate of the value of k required to give a good fit to the data could be made by dividing the discrepancy by i. In fact

$$\frac{9\cdot29}{591\cdot7}=0\cdot016$$

would have been the value of k indicated by the data, if any value other than zero had been required.

REFERENCES

Fisher, R. A. & Yates, F. (1943). 'Statistical tables for biological, agricultural and medical research' (2nd ed.). Edinburgh.

Fisher, R. A. (1941). 'Statistical methods for research workers' (8th ed.). Edinburgh.

SUMMARY

Part 1. It is shown that in a large collection of Lepidoptera captured in Malaya the frequency of the number of species represented by different numbers of individuals fitted somewhat closely to a hyperbola type of curve, so long as only the rarer species were considered. The data for the commoner species was not so strictly 'randomized', but the whole series could be closely fitted by a series of the logarithmic type as described by Fisher in Part 3. Other data for random collections of insects in the field were also shown to fit fairly well to this series.

Part 2. Extensive data on the capture of about 1500 Macrolepidoptera of about 240 species in a light-trap at Harpenden is analysed in relation to Fisher's mathematical theory and is shown to fit extremely closely to the calculations.

The calculations are applied first to the frequency of occurrence of species represented by different numbers of individuals—and secondly to the number of species in samples of different sizes from the same population.

The parameter 'α', which it is suggested should be called the 'index of diversity', is shown to have a regular seasonal change in the case of the Macrolepidoptera in the trap. In addition, samples from two traps which overlooked somewhat different vegetation are shown to have 'α' values which are significantly different.

It is shown that, provided the samples are not small, 'α' is the increase in the number of species obtained by increasing the size of a sample by e (2·718). A diagram is given (Fig. 8) from which any one of the values, total number of species, total number of individuals and index of diversity (α), can be obtained approximately if the other two are known. The standard error of α is also indicated on the same diagram.

Part 3. A theoretical distribution is developed which appears to be suitable for the frequencies with which different species occur in a random collection, in the common case in which many species are so rare that their chance of inclusion is small.

The relationships of the new distribution with the negative binomial and the Poisson series are established.

Numerical processes are exhibited for fitting the series to observations containing given numbers of species and individuals, and for estimating the parameter α representing the richness in species of the material sampled; secondly, for calculating the standard error of α, and thirdly, for testing whether the series exhibits a significant deviation from the limiting form used.

Special tables are presented for facilitating these calculations.