

# THE COMMONNESS, AND RARITY, OF SPECIES

F. W. PRESTON

*Preston Laboratories, Butler, Pennsylvania*

## SECTION I. GENERAL

The purpose of this paper is to deduce, from a number of examples and from theoretical considerations, some plausible general law as to how abundance or commonness is distributed among species. Experimentally, this could be done by making a complete *census* of every species, but, with rare exceptions, this procedure is quite impractical. We therefore attempt to deduce the "*universe*" from a *sample*.

Commonness, as understood by ecologists, has several rather different meanings: we are here concerned with (1) the total number of living individuals of a given species, which might be called its global abundance, (2) the total number of individuals living at any instant on a given area, such as on an acre or a square mile, which might be called its local abundance, (3) the ratio which the number of individuals of one species bears to that of another species, i.e., its relative abundance, and (4) the number of individuals *observed*, for example, the number of a moth species counted in a sample from a light trap, its "observed," "apparent," or "sample" abundance. There will not usually be any difficulty in deciding which phase of the subject is under discussion at any time. The Raunkiaer "Index of Frequency" is a measure of ubiquity rather than of commonness as above defined: its relation to our other concepts is discussed later.

As a rule, we are interested in a sample only in so far as it throws light upon the "*universe*" that is being sampled. The sample will be a sufficiently accurate replica of the universe provided (1) it is a perfectly "random" sample, and (2) no species is represented in the sample by less than 20 or 30 individuals.

In most ecological work condition (2)

will never obtain, and much of the present paper will center on this difficulty. Condition (1) will not usually obtain in the broadest sense, and needs a moment's consideration.

A geologist sampling an ore-body, whose boundaries have been delimited accurately by previous exploration, has a known "*universe*" and merely needs information on composition. His universe is permanent. But in ecological work, the "*universe*" changes rapidly. The moths flying tonight are not those that flew a month ago, or will fly a month hence. Those flying this year are a vastly different association from those that flew last year in the same area. The same thing is true of rodent populations, of birds, and of plants. We are dealing with a fleeting and fluctuating assemblage, a "*universe*" continually expanding, contracting, and changing in composition. Thus it is important to recognize at the outset that, for the purposes of our present investigation, the "*universe*" from which the sample is drawn is that universe declared to us by the sample itself, and not our preconceived notion of what the universe ought to be.

Further, it is important to recognize that the randomness we seek is merely randomness *with respect to commonness or rarity*. A light trap is satisfactory in this respect and samples its own universe appropriately. It is definitely *selective* in respect of phototropism, but it is *random* in respect of commonness, i.e., it does not care which of two moths, equally phototropic, it catches, though one may be a great rarity and the other of a very common species. On the other hand, an entomologist, or even an intelligent boy, with a net, is not a satisfactory collector, for he will go after the rarity. For this reason we have to reject Corbet's ('41)

collection of Malayan butterflies, as not being a *random* sample.

*Previous discussions.* In the early years of the present century an extensive literature accumulated in connection with Raunkiaer's Law of Frequency. This was apparently intended as an index of commonness or rarity, more particularly in plant associations but not necessarily confined thereto, and, subject to some assumptions, it is a measure of commonness or rarity. The method of expressing the results, though it appears simple, is in reality very inconvenient, and gives no direct clue to the nature of the universe being sampled.

An analysis of Raunkiaer's findings must be postponed, in the present paper, until after we have established the relationship between sample and universe.

More recently Williams ('43) and his collaborators have taken up the subject from a different point of view.

Williams was struck by the fact that in a "random" collection of butterflies or moths there were usually great numbers of species represented by a single specimen (hereinafter called "singletons"), a much less number of species represented by two specimens ("doubletons") and still fewer species represented by three, four, and so on. He noticed further than if  $N$  be the number of species represented by singletons, the number represented by doubletons was  $N/2$ , the number represented by three was  $N/3$  and so on.

Hence in the collection, if the observation were valid indefinitely, we should have a series

$$N \left( 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \dots \right)$$

as the number of species represented by

1    2    3    4    5    ...

specimens.

This is a harmonic series, and it has the grave theoretical disadvantage that the sum of its terms is infinite, calling for an infinite number of *species* in the collection. As if that were not bad enough, the number of *specimens* is of the order of infinity to the second power.

Williams, aware of these difficulties, apparently sought the help of Fisher ('43) who introduced an arbitrary modifying term into the series, so that it was no longer strictly harmonic, and both species and specimens were reduced to finite numbers. The degree of agreement thereby produced may be judged from the group of papers by Corbet, Williams and Fisher ('43). Even on the collections on which it was tested it does not work too well, and Corbet, in an earlier paper ('41), was aware that as collecting continued, it would represent the facts increasingly poorly.

*Method of graphing.* It has often been a matter of comment by ecologists that one or two species are extraordinarily abundant at a particular time and place: all others seem rare in comparison. Thus in Seamans' moths of Alberta, discussed later, out of a total of 303,251 specimens, 91,502 belonged to a single species; while in Dr. King's moths of Saskatchewan, 38 species accounted for 77,260 specimens, and 237 more accounted for only 9,850 additional individuals.

Under these circumstances it would be logical, merely as a matter of convenience, to try plotting commonness (number of individuals per species) on a logarithmic base. But there is a more cogent basis for so doing. Commonness is a *relative* matter. One species, we say, is twice as common as another, and a natural series of groups representing commonness would run:

Species group	A	B	C	D	E	F	G	H	I	J	etc.
Approximate specimens observed of that species	1	2	4	8	16	32	64	128	256	512	etc.

These groups are a sequence of *octaves* of frequency. It is the most natural grouping possible, but our ultimate justification for adopting it is not its naturalness, but the fact that it works. It is a geometric, or logarithmic series, and this paper produces evidence that commonness of species appears to be a simple Gaussian curve on a geometric base (i.e. "lognormal" curve). The intervals of an octave turn out to be a convenient size of interval, as well as the most readily grasped ones.\*

*Boundaries of the octaves.* An octave is simply an interval of two-to-one. On the piano it may be from C to C' or G to G': similarly in our ecological studies, it is the interval 4-to-8 or 6-to-12. We might choose the numbers 1, 2, 4, 8, etc., as the mid-points of our octaves, but it is more convenient to choose them as the end-points, or boundaries between octaves. Thus denoting the octave 1 to 2 as A, and so on, we have octave D comprising the interval whose boundaries are 8 and 16.

Thus if a given species is represented by 9, 10, 11, 12, 13, 14 or 15 specimens, it clearly falls in octave D. All species falling in octave D may be thought of as having roughly the same degree of commonness, in comparison with those falling for instance in octave J, which are represented by from 513 to 1023 specimens.

If a species is represented by 8 specimens, octave D is credited with half a species, and octave C is credited with the other half. Similarly a species with 16 specimens is credited half to D and half to E. Octave B is comprised of all species having 3 specimens in the sample, plus half the species having 2 and half the species having 4. Octave A is composed of half the species represented by singletons, and half those represented by dou-

bletons. Half the singletons have to be assigned to octaves below A, which we designate by Greek letters and discuss later.

*Procedure.* We obtain a sample (say a catch of moths in a light trap), and make a complete tally of the relevant material (e.g., macrolepidoptera, noctuids, all lepidoptera, or all insects), determining the "representation" (number of individuals) of each species. This assigns the species to its appropriate octave of frequency. We count the number of species in each octave. This is the ordinate of our curve, and the abscissa is a scale of octaves of increasing commonness.

The experimental points thus obtained might, as the reader will doubtless observe, be "graduated" quite well as a Poisson distribution. Indeed, it is possible that that might be the most formal approach, and in the long run the most accurate. However, experience shows that an ordinary Gaussian curve (the "normal curve of errors"  $y = e^{-(ax)^2}$ ) is very satisfactory also, provided the curve is "decapitated" at the veil line as discussed later. The use of this curve has special advantages. Singleton ('44) has worked out a method of fitting a decapitated Gaussian curve in accordance with the principle of Maximum Likelihood: this method was used by Dr. J. L. Glathart (of the Preston Laboratories) to graduate many of the curves we have used. Others are graduated by a graphical method of our own. Both give comparable results, and the methods, for lack of space, will not be discussed in the present paper.

#### *Relation between Williams' harmonic series and the octave plot*

If Williams' unmodified law were true, viz. that if there are  $N$  species represented by one specimen each, there will be  $N/2$  represented by two specimens, and so on, we could calculate how many species would fall in each of our octaves. It comes out as follows:

\* The above procedure is equivalent to taking the ordinary logarithms of the number of specimens per species and grouping the species into the groups whose logarithms extend over an interval of 0.3010 i.e., such groups as 0-0.3010, 0.3010-0.6021, 0.6021-0.9031, etc.

Octave	Number of species in that octave
A (1-2)	0.7500 $N$
B (2-4)	0.7083 $N$
C (4-8)	0.6971 $N$
D (8-16)	0.6941 $N$
E (16-32)	0.6934 $N$
F (32-64)	0.6932 $N$
G (64-128)	0.6932 $N$

All subsequent octaves have approximately 0.6931  $N$ , where  $0.6931 = \log_e 2$ . Except for the first two octaves, the number of species in each octave would be constant to within one per cent. The data are graphed in figure 1.

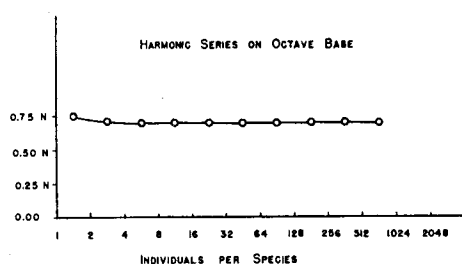


FIG. 1. The harmonic series on an octave base. If the commonness of species followed the (unmodified) Williams' law, then the number of species in each octave would be essentially constant.

Experimental evidence shows that this state of affairs is far from the truth. Theoretical considerations also require its categorical rejection. There are an infinite number of octaves to the right, and if they were all equally filled with species, there would be an infinite number of them, and excessively common species would be as numerous as very moderately common ones, which we know is not the case. Therefore the curve must descend to the right. Williams himself recognizes this, and this is the basis of Fisher's modification of the Harmonic Series Law, which modification produces a descent to the right.

But as will appear later, the real curve also descends to the left. It is a humped curve. This requires a totally different approach.

We are now ready to consider the experimental evidence.

## SECTION II. EXAMPLES

We have one or two examples of bird-counts that meet our essential requirements, and four or five of moths in light traps. Most counts do not meet our requirements, but some throw light on the central problem and seem to extend our findings to other biological universes.

### I. Saunders ('36). *The birds of Quaker Run Valley, western New York State*

Saunders attempted a complete census of the breeding birds on a tract of 16,967 acres. His main reliance was upon a count of singing males, aided by a few auxiliary methods.

Table I-A below is the number of breeding pairs of each species, from commonest to rarest, in the order in which Saunders reported them, but with the names of the species omitted.

Table I-B condenses this information ready for graphing, and the plot is given in Fig. 2.\*

The total number of species here reported is 80, but Saunders states:

"At the present date 141 species of birds are known to occur in Allegany Park and its vicinity during the months of July and August. Of these birds, 27 occur as fall migrants or late summer wanderers. Thirteen breed outside the park itself, but evidently in the near vicinity. Eleven more breed in the park but not, so far as is known, in the Quaker Run Valley. The remaining 90 occur in their breeding season and undoubtedly breed or have bred in the Quaker Run Valley in the past 12 years. Of these 90 species, 11 are irregular in their occurrence and have not been found every year. The remaining 79 are regular breeders and, judging by past experience, can be expected to breed in the area every year."

In the graph (fig. 2) the curve drawn is a Gaussian curve, with its mode (crest) at  $i = 30$  specimens (breeding pairs) per species, that is, near the end of the 16-32 octave: the general equation of such a curve is

$$n = n_0 e^{-(aR)^2},$$

\* The Veil Line shown in this and other graphs will be discussed later.

Birds of Quaker Run Valley

TABLE I-A. Values of *i* (specimens per species)

1670	1656	1196	868	723	723	675	506	477	389	367
324	311	310	288	282	280	270	220	188	181	179
161	160	158	152	138	111	109	91	90	88	79
60	57	56	50	46	46	43	43	35	34	33
32	32	30	28	28	26	24	23	22	17	15
14	12	10	10	10	10	10	8	8	7	6
6	5	5	4	4	4	4	3	3	3	3
2	1	1	?	?	?	?	?	?	?	?
?	?	?								

TABLE I-B. Values of *n* (species per octave)

Octave	<1	1 to 2	2 to 4	4 to 8	8 to 16	16 to 32	32 to 64	64 to 128	128 to 256	256 to 512	512 to 1024	1024 to 2048
Species per octave	>1	1½	6½	8	9	9	12	6	9	11	4	3

where  $n_o$  is the number of species in the modal octave,  $n$  is the number in an octave distant  $R$  octaves from the mode, and  $a$  is a constant to be calculated from the experimental evidence.

In this particular case

$$n = 10e^{-(0.194R)^2}.$$

The curve theoretically extends to infinity, asymptotic to the  $x$ -axis, both to left and right. The area “under” the curve (between curve and  $x$ -axis) from minus infinity to plus infinity, is theoretically the *universe being sampled*, i.e. the total number of species theoretically available for observation.

$$N = \int_{-\infty}^{+\infty} ndR = n_o\sqrt{\pi}/a,$$

where  $\pi = 3.1416$  and  $\sqrt{\pi} = 1.77$ . Putting  $n_o = 10$  and  $a = 0.194$  we get  $N = 91$ , which agrees very closely with Saunders’ own estimate (quoted above) of what it is.

II. Preston and Norris ('47). *Breeding birds of the Frith (Preston Laboratory Grounds)*

From table II of the paper by these authors, we may obtain data of the total pairs of breeding birds, and the total number of “expected” nests (remembering that some species make more nests than one, even if successful with the first, and most species make a second nest if the first is unsuccessful). The count for the combined years 1944 and 1945 gives:

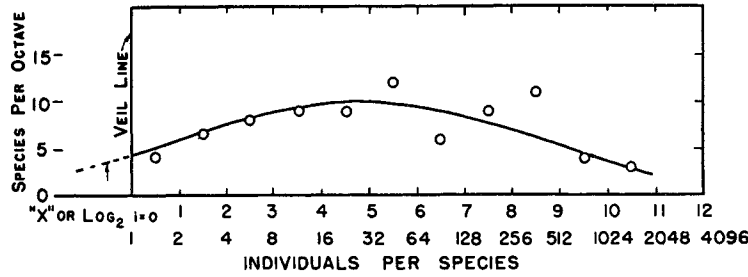


FIG. 2. Saunders’ breeding birds. The octaves are definitely not equally filled, and the curve looks as if it is humped up in the middle.

Octave <i>i</i>	— <1	A 1-2	B 2-4	C 4-8	D 8-16	E 16-32	F 32-64	G 64-128	H 128-256	I 256-512
Species ( <i>n</i> ) Nests	>3½	5	6½	10	7	5	4	4	2	0
Species ( <i>n</i> ) Bird-Pairs	>5	6	7	11½	8½	4	7	1	0	0

TABLE II. Values of *n* (species per octave)

Octave	<1	1 to 2	2 to 4	4 to 8	8 to 16	16 to 32	32 to 64	64 to 128	128 to 256	256 to 512	512 to 1024	1024 to 2048	2048 to 4096	4096 to 8192	8192 to 16384
Species per octave	>19	37	42	49½	45½	42	28½	26½	30	14	9	2	0	2	2

These results are not graphed, as only a few hundred bird-pairs and nests were involved, and statistically this is hardly sufficient. So far as the data go, however, they support a humped distribution with its crest in octave C, both for birds and for nests.

III. Dirks ('37). *Moths in a light trap at Orono, Maine*

Dirks operated a light trap over a period of four years (1931-34). Exclusive of microlepidoptera he caught 56,131 specimens. These he classified as to species and sex, and gives three tabulations: total moths, female moths, gravid female moths. Since most females were gravid, the third tabulation closely resembles the second and will not concern us here.

Condensing his tabulation for total moths produces table II.

This is graphed in figure 3. The first number (for the octave less than 1) is indicated by an upward pointing arrow, meaning that the true number for this range is above the tip of the arrow, a point to which we shall return later. The other points are plotted as circles. The smoothed curve drawn through the points appears to have its crest at about 7 or 8 specimens per species. The left-hand end of the curve is, of course, quite incomplete. It represents the rarer species which were absent from the catch. The general shape of the curve, however, appears clear enough, and we may assume that it is approximately a "normal" probability curve with its mode at, or very near, 8 specimens per species.

The equation of the curve is  $n = 48e^{-(0.207R)^2}$ , and on this basis the total number of species in the universe being

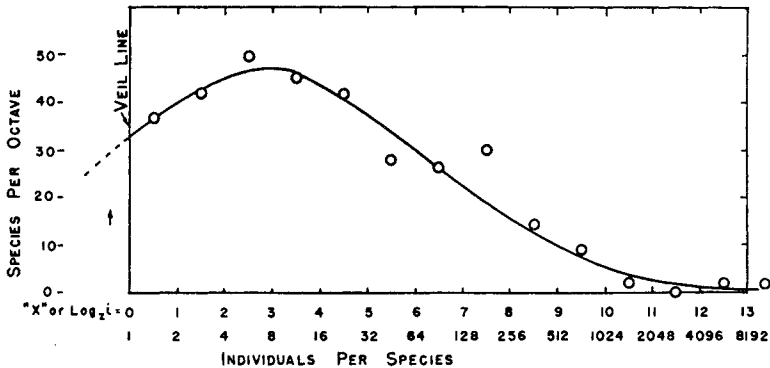


FIG. 3. Dirks' moths of Orono, Maine. The octaves are not equally filled; the curve is humped; we see the right-hand tail, but the left-hand one is hidden behind a veil.

sampled is  
 $48 \sqrt{\pi}/(0.207) = 410$  species  
approximately.

Dirks actually obtained 349 species, so his sample represents about 85% of the species that were theoretically available at Orono, during the four years in question, for trapping by means of such a light trap as he used.

IV. *Dirks ('37). Female moths in a light trap at Orono, Maine*

The condensed statement of results follows:

TABLE III. Values of  $n$  (species per octave)

Octave	<1	1 to 2	2 to 4	4 to 8	8 to 16	16 to 32	32 to 64	64 to 128	128 to 256	256 to 512	512 to 1024	1024 to 2048	2048 to 4096	4096 to 8192
Species per octave	$>27\frac{1}{2}$	42	$39\frac{1}{2}$	$27\frac{1}{2}$	28	23	$19\frac{1}{2}$	9	4	4	0	0	1	1

This is graphed in figure 4, whose mode is at or near  $11\frac{1}{2}$  specimens per species, and equation is

$$n = 42e^{-(0.205R)^2}.$$

On this basis, the available universe comprised 363 species, of which 226 were represented in the sample, in the form of 12,799 individuals. Females in Dirks' sample were only 22.8% of the total, which probably means, not that they were rarer in the outer world, but that they were travelling less than the males. Dirks obtained a sample of 85% of the "total" species (410) in the total universe, but only 62% of the females ( $226/363 = 62\%$ ) which constituted a smaller universe (363 species as against 410).

V. *Williams ('43). Moths in a light trap at Rothamsted, England*

The condensed tabulation is:

TABLE IV. Values of  $n$  (species per octave)

Octave	<1	1 to 2	2 to 4	4 to 8	8 to 16	16 to 32	32 to 64	64 to 128	128 to 256	256 to 512	512 to 1024	1024 to 2048	2048 to 4096
Species per octave	$>17\frac{1}{2}$	23	$27\frac{1}{2}$	36	$27\frac{1}{2}$	33	31	$13\frac{1}{2}$	19	5	6	0	1

The corresponding graph is given in figure 5. Its mode is at about 8 specimens per species and its equation is

$$n = 35e^{-(0.227R)^2}.$$

The total population being sampled is  $N = 35 \sqrt{\pi}/(0.227) = 273$  species. Williams actually trapped, in the four years covered by his report, 240 species, involving 15,609 specimens (macrolepidoptera).

Williams comments that the total Hertfordshire fauna (of this type of lepidoptera) amounts to 461 species, and tries to figure how long it would take to catch

examples of all of them. But if our present theory is valid, the "universe" he was sampling was not the whole Hertfordshire fauna, but a local population comprising about 60% of the Hertfordshire total. This may be compared with the experience of Dirks, who was sampling not the fauna of Maine, but a local population near Orono.

VI. *King (unpublished). Moths in a light trap at Saskatoon, Saskatchewan*

This is a collection of macrolepidoptera, numbering 87,110 individuals, 277 species, caught in Dr. King's "Trap A," over a period of 22 years, from 1923 to 1944 inclusive.

The data are graphed in figure 6, and the curve, calculated by Singleton's method, has its mode at 4 specimens per species and the equation

$$n = 33.0e^{-(0.152R)^2}.$$



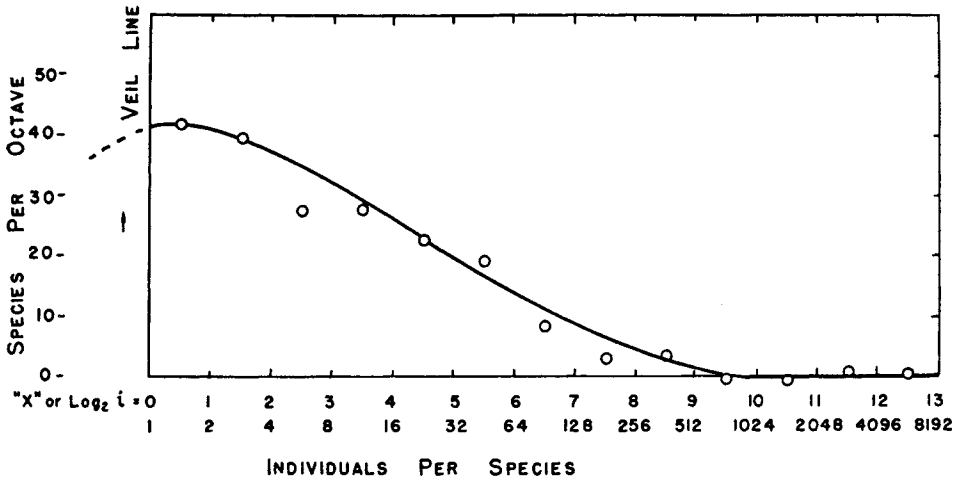


FIG. 4. Dirks' female moths. The hump seems to be very close to the veil. In Sweadner's moths of Wyoming (not illustrated), the veil has moved to the right of the hump, and all we see is the descending side.

The "universe" accordingly comprises 384 species, and since Dr. King caught 277, his trapping is seemingly 72% complete.

It may seem strange that after 22 years the collection should be so incomplete, when 4 years of trapping at other places yield values over 80%. But it must be remembered that what is being sampled during the 22 years is a 22-year universe, not a 4-year universe, and that the moths available for trapping, in Saskatoon at any rate, vary enormously from year to

year. This will be dealt with, briefly, in the later discussion.

Since we have Dr. King's data year by year, it is quite possible to make up sub-samples and plot the results year by year or quadrennium by quadrennium. The curves retain their general shape, though the height at the mode tends to fall to about 24 species per octave instead of the 33 species per octave in the 22-year aggregate. This alone, of course, shows a likelihood that we are dealing with a vastly more extensive universe if we trap

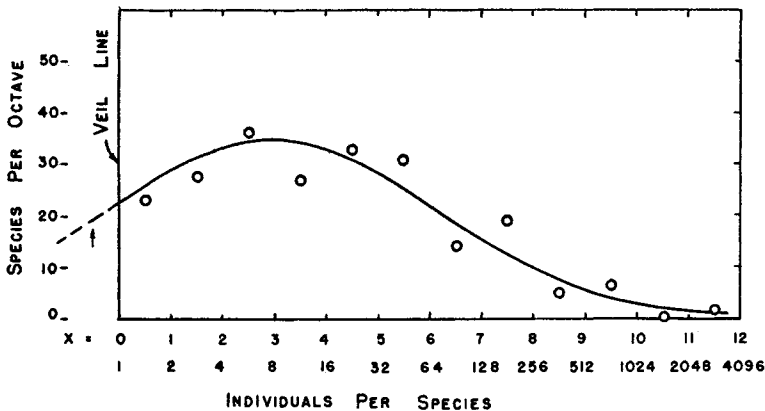


FIG. 5. Williams' moths. This resembles figure 3.



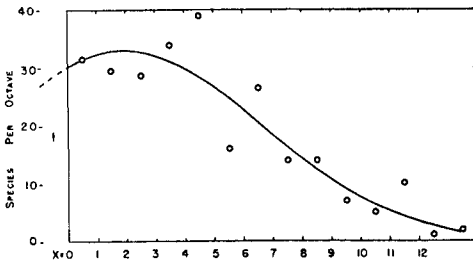


FIG. 6. King's moths of Saskatoon.

for a longer period—at any rate at Saskatoon.

VII. *Seamans (unpublished). Moths in a light trap at Lethbridge, Alberta*

Seamans' records cover the period 1921–43 inclusive, with the exception of the year 1938, when the collections were damaged by mice before they could be catalogued. The number of specimens is 303,251; of species, 291. This is the most extensive collection at present available to me in a completely counted condition.

The graph is given in figure 7: its mode is at about 32 specimens per species, and its equation

$$n = 30.0e^{-(0.160R)^2}.$$

The theoretical universe is accordingly 332 species and the collection is about 88% complete.

With the Lethbridge data, as with those of Saskatoon, it is possible to make up samples for individual years, or for quadrennia. For individual years the modal octave tends to hold about 20 species per octave: for quadrennia it averages about

25: for the 22 years, about 30, but the general shape is very similar. The graphs are accordingly not reproduced here.

These few examples show what the experimental plottings look like, and the degree of fit between the Gaussian curve and the data. In all these examples we have a sample with more than half the universe represented, i.e., the crest of the curve is in sight. Fragmentary collections are available where the crest has not been reached, and probably Raunkiaer's collections (discussed later) were of this type.

With the aid of these actual examples,\* thus very briefly presented, we now pass to a discussion of the general principles, the veil line, the relation between sample and universe, and some deductions about the nature of biological universes in general.

### SECTION III. DISCUSSION

#### *The veil line*

If the universe contains species so rare that the sample, if perfectly proportioned to that universe, should theoretically contain only a fraction of a specimen, it is most likely that the species will be unrepresented in the collection. It cannot be represented by less than one whole specimen, and by definition this would be

\* Hinton ('40) gives data on a collection of adult Mexican water-beetles, and some others. This graduates very well as a decapitated Gaussian curve. The internal evidence suggests it was a random collection, but since Hinton gives no comment on that point, discussion is omitted in the present paper.

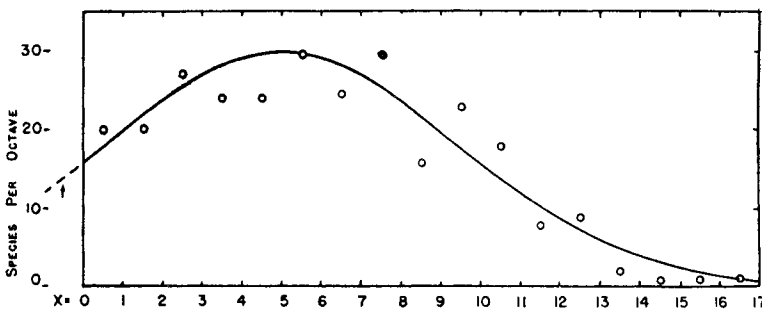


FIG. 7. Seaman's moths of Lethbridge.

over-representation. Some species indeed will achieve this over-representation, just as some species which "ought" to be represented by 1, 2, 3 or even more specimens, may fail of representation. But on the average, those relatively uncommon species whose theoretical representation is appreciably less than one specimen, will be missing from the sample. These species may themselves be of very different degrees of rarity; some "ought" to be represented perhaps by half a specimen, others by  $1/20$  or  $1/200$  of a specimen. In other words, they belong to different octaves, but these octaves are all "missing" in the sample.

There is therefore a veil line at the left end of octave A, at a representation of one specimen. The sample indicates clearly what the universe is like to the right of the veil, but we have to infer what it is like to the left of it.

Now doubling the size of the sample, other things being equal, doubles the number of individuals of each of the commoner species, in the sample. This is equivalent to moving every species one octave to the right, which in turn is equivalent to withdrawing one more octave from under the veil. The process can be repeated until the whole universe, in effect, is exposed. Thus the graph of the universe is identical with the graph of the sample, except that it is not decapitated by a veil line. The graph in fact behaves as if it were drawn complete upon a movable card, and the card slipped with its left-hand end under a veil. Continued sampling, if the universe remains constant, merely withdraws the card progressively from under that veil.

If this last statement is true, it is a very valuable property of our method of representing our results. It will be obvious to all, that it is quite accurately true for those species which, in the sample, are represented by at least a score or two of specimens. These species will be present in the sample in virtually the exact proportion in which they are present in the universe. It will not be obvious that it

will be true of those species whose representation in the sample is small, so that, by the laws of small samples, they have a good chance of being substantially over- or under-represented.

The next page or two, on the Relation between Sample and Universe, show that though any one of the less common species may very well get into the wrong octave, the probability is that every octave has very nearly the correct number of species. This is what we graph, and therefore our graph is not appreciably affected by the interchange of species among the octaves close to the veil line.

In other words, the "shape" of our universe is constant. If however we want to know the exact status of a specified species in that universe, we cannot be sure of it until the sample contains a dozen or two specimens.

#### *Relation of sample to "universe"*

Consider any universe, Gaussian or otherwise, made up of at least several hundred species, and fix attention on any one species, say a melanic, which is not one of the very common ones. Then the numbers of individuals of this species must be less than one in a thousand compared with the non-melanics. Suppose for the moment that it is one in ten thousand.

If we catch, at random, a single specimen from our universe, we have one chance in ten thousand that it is a melanic; and if we make a catch of 10,000 specimens, "perfect representation" would call for one melanic in the collection.

If we now released the specimens unharmed back into the universe, so as not to deplete it, we could make many random collections of 10,000 specimens, and when we tallied up the number of samples (collections) that showed zero, one, two, etc., melanics, we should find they conform to the several terms of the Poisson Series,

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

probability for

$$\begin{array}{ccccccc} 0 & 1 & 2 & 3 & \dots & & \\ & & & & & & \text{melanics.} \end{array}$$

Similarly if perfect representation calls for exactly two melanics in the collection, which would be the case if we collected 20,000 specimens all told, then the probability of our finding zero, one, two, three . . . specimens is given by the terms of the Poisson series

$$e^{-2} \left( 1 + \frac{2}{1!} + \frac{2^2}{2!} + \frac{2^3}{3!} + \text{etc.} \right).$$

More generally if perfect representation calls for " $p$ " \* melanics in the collection, our series is

$$e^{-p} \left( 1 + p + \frac{p^2}{2!} + \frac{p^3}{3!} + \frac{p^4}{4!} + \dots \right).$$

The terms of this series are easily calculated if  $p$  is small or fractional, but it is tedious if  $p$  is large, as the series then diverges at first and converges thereafter very slowly. Fortunately we do not need to use very high values of  $p$ .

What is true for melanics is of course true for any other sort of moth or species which is present to the same extent as the melanic. Thus if a hundred different species are present in identical proportions, and the sampling is carried to the point where we might expect exactly 3 specimens of each, we can tell how many of these species are in fact likely to be represented by 3 specimens, or by any other number.

Next let us suppose that, in the universe, the species in any octave of frequency are all concentrated at the midpoints of the octaves, so that if one species is present to the extent of 6 M where M is some large number—a million say—then there are a number of other species also present to the extent of 6 M, but none to the extent of 4 M, 5 M, 7 M, or 8 M. Similarly the species of the octave

2 M–4 M are all concentrated at 3 M; \* and so on.

Then we can so conduct our collecting that any species occurring in the universe ought to be present, with perfect sampling, to the extent of 3, 6, 12, 24, 48, 96, . . . individuals and not to the extent of any intermediate figure. Similarly the rarer species, if perfectly sampled, would contain fractions of specimens, and would be represented by  $1\frac{1}{2}$ ,  $\frac{3}{4}$ ,  $\frac{3}{8}$ ,  $\frac{3}{16}$ , . . . specimens.

Now in real collections, a species has to be represented by an integral number, or be absent; and the perfect sampling has to be replaced by the most probable sampling given by the Poisson series we have just been discussing. And what is true of fractional examples is equally true of the "perfect" integral numbers. A species that ought to be represented by exactly 3 specimens will, five times out of a hundred, have zero representation, and about one time in a thousand it will have as many as ten specimens to its credit. Otherwise expressed, if there are one thousand species (in a universe containing many thousands of species) which are entitled with *perfect* sampling to a representation of three individuals, then with *random* sampling fifty of these thousand species might very likely be unrepresented, while one species might well be present to the tune of ten specimens.

Perfect sampling can be accomplished only by a conscious agent: random sampling is what happens in the absence of conscious selection.

The whole story may now be reduced to a tabulation (table V).

In this table, the left-hand column gives the midpoints of 12 octaves, six behind the veil line and six out in the open. For instance  $p = 3$  represents the (arithmetical) midpoint of the octave 2–4. Suppose a given species is present in the

\* " $p$ ," as used here, stands for "perfect representation," or "average expectation," more often denoted perhaps by  $c$ . It must not be confused with " $p$  for probability" in the classical binomial  $(p + q)^n$ .

\* There may be some argument that the midpoint of the octave should be  $2\sqrt{2}$  M rather than 3 M, the geometric rather than the arithmetic mean: but the final numerical results will be nearly the same either way.

"universe" to such an extent that a perfect sample would contain exactly three specimens, that is, let the sampling be continued just long enough to justify an expectation of exactly three specimens of this species; then let us return the specimens unharmed to the universe and repeat our sampling many times. Then 12.5% of these many samples would

show the species falling in the range 0-1, 18.7% in the octave 1-2, 42% in its own octave 2-4, and so on.

The second column from the left represents all the times in which we should expect zero representations, and half the times we should expect singletons.

The third column represents half the singletons and half the doubletons.

TABLE V. *Probabilities of representation (expressed as percentages), when "perfect" representation is  $p$  specimens*

(Specimens) actual ( $m$ ) Perfect ( $p$ ) representation	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
3/128	97.6	2.4															
3/64	95.4	4.5	0.1														
3/32	91.1	8.5	0.4														
3/16	83.0	15.5	1.4	0.1													
3/8	68.7	26.1	4.8	0.4													
3/4	47.3	35.5	13.3	3.3	0.6												
1½	22.3	33.4	25.1	12.5	4.7	1.4	0.4	0.1									
3	5.0	15.0	22.4	22.4	16.8	10.0	5.0	2.1	0.8	0.3	0.1						
6	0.2	1.5	4.5	8.9	13.4	16.2	16.2	13.9	10.4	6.9	4.1	2.2	1.1	0.5	0.2	0.1	
12					0.5	1.3	2.5	4.3	6.5	8.7	10.5	11.5	11.5	10.7	9.0	7.2	5.5
24												0.1	0.3	0.5	0.9	1.5	2.2

This table gives the percentage probability that a species will be represented, in the sample, by exactly 0, 1, 2, 3, ... individuals, if its frequency or commonness in the universe (Gaussian or any other) would lead us to expect in the sample an exact representation of " $p$ " specimens. We are interested, however, in knowing something a little less ambitious than this, *viz.*, not the probability of a particular number of specimens, but the probability of the species falling in its proper octave of the sample, or in any other octave. The table can therefore be condensed into a shorter one, covering the first four octaves of the sample. This is done below in table VI.

TABLE VI. *Probabilities of representation in each octave when "perfect" representation would be " $p$ " specimens*

Being a condensation of the previous table (V)  
(Probabilities expressed as percentages)

$p$	Octaves						
	Below 1	1-2	2-4	4-8	8-16	16-32	
3/128	98.8	1.2					— Veil Line
3/64	97.6	2.3	0.1				
3/32	95.3	4.5	0.2				
3/16	90.7	8.5	0.8				
3/8	81.7	15.5	2.8				
3/4	65.0	24.4	10.3	0.3			
1½	39.0	29.3	27.4	4.2			
3	12.5	18.7	42.0	25.9	0.8		
6	0.9	3.0	17.8	58.2	20.3		
12			0.2	11.6	75.1	3.1	
24					4.4		
		107.4	101.6	100.2	100.6		

The fourth represents half the times we should expect doubletons, all the times we should expect three specimens, and half the times we should expect four specimens.

This is in accord with our previous methods of figuring things.

The broken horizontal line is the veil line. A species theoretically represented by  $\frac{3}{4}$  of a specimen should be absent. A species theoretically represented by  $1\frac{1}{2}$  specimens should be represented by at least one.

The sloping interrupted line indicates the percentage probability with which a species should fall in its "own" octave. A species which in a "perfect" sample would have to be represented by  $1\frac{1}{2}$  specimens has less than a 30% chance to be represented in its own octave, while a species theoretically represented by 12 specimens has better than a 75% chance to fall in its own octave.

We thus have (in the table) substantially exact information about the composition of the first four visible octaves of the sample (1-2, 2-4, 4-8, 8-16). Beyond this point the calculations become very tedious and the table has not been computed. But it is easy to see that in the next octave (16-32) it is most probable that well over 90% of all species that should fall in it, *do* fall in it, and that from this point forward (32-64), and onwards, practically every species falls in its own octave, and no octave derives any adventitious support. Therefore it is the first four octaves that are important. Beyond this point the sample is a faithful replica of the "universe," whatever the distribution of commonness and rarity in that universe may be, so long as it is a reasonably smooth continuous sort of function.

We may therefore proceed to calculate what our samples should theoretically be, for two kinds of "universe." The first universe is the one in which all octaves are equally filled, the kind that results from the harmonic series of Professor Fisher and Dr. Williams. The second

universe is the Gaussian one, with the coefficient "*a*" in the exponent set equal to 0.20.

*Harmonic series: octaves in universe equally filled*

This is very easily dealt with. All we have to do is to add up the columns (except the first two) in table VI, as is done at the foot of that table. It will be seen that (allowing for small errors of calculation) all columns add up to just about 100% except the first, where there seems to be about 7% excess. That is, in this column, if we had a right to expect 14 species we should most likely get 15. This error is not very serious, and we might very well say that there are no errors beyond this point. Thus for this sort of universe, the sample is a very faithful replica.

*Gaussian universe*

The calculations here are more tedious, and have to be made for several cases; the veil line in the sample may coincide with the crest of the curve in the universe, or it may be at any point on the ascending or descending slopes. If we consider the universe as fixed, we can bring the veil line to any point we wish by continuing our collecting. Ultimately we collect the whole universe, and then the sample coincides with the universe, and is obviously a complete Gaussian curve if we postulate that the universe also is. We are more interested in the cases where the sample's veil line is within a few octaves of the crest of the "universe."

Let the number of species falling in each octave of the universe be assumed to be

$$n = 100e^{-(0.20R)^2},$$

where *R*, as before, is the number of octaves to left or right of the crest of the curve.

In most of the "universes" we have encountered in practice there is a tendency for "*a*" to be somewhere near 0.20, so it is logical—since we must use some speci-

TABLE VII. *Distribution of species in the Gaussian universe*

<i>R</i> =	0	±1	±2	±3	±4	±5	±6	±7	±8	±9	±10	±11	±12
<i>n</i> =	100	96.1	85.2	69.8	52.7	36.8	23.7	14.1	7.7	3.9	1.8	0.8	0.3

fic value for our calculations—to use this figure. The assumption that the modal octave will have 100 species is of course entirely arbitrary, but if it has some other number, it will merely result in all our figures being multiplied by a constant, both in the universe and in the sample, and will not affect our ultimate result in any way.

Let us now take three cases as follows: (1) with the veil line placed so that the first “exposed” octave in the sample is the modal octave of the universe, (2) with the veil line four octaves to the *left* of the above position, and (3) with the veil line four octaves to the *right* of the first position.

Applying table VII to table VI, we see, in Case (1), that the modal octave ought to be made up of 29.3% of the 100 species that “belong” there, plus (24.4 + 18.7)% of the 96.1 species that theoretically belong in the immediately adjoining octaves, plus smaller contributions from more distant octaves of the universe. After completing the calculations for all the octaves, and for all three cases, we reach the following conclusions:

proximate the veil-line position of Case (1) rather than the others, it appears likely that even in the other octaves the errors will not usually exceed a few per cent. In figures 8A, 8B, 8C and 8D, we show the (species) curve for the postulated universe, and for the three samples of Cases (1), (2) and (3), respectively. In b, c, and d, the solid line is the most probable sample curve, and the broken line the corresponding curve for the universe.

This method of computing the curves is an approximation only, and the postulate that the sample curve is essentially a decapitated version of the universe curve is likewise only an approximation; but it seems that it is sufficiently close for our present purposes.

*Raunkiaer’s law*

Some twenty years ago there was some discussion of Raunkiaer’s law in the pages of *ECOLOGY*, (Kenoyer, ’27; Gleason, ’29; Romell, ’30) and a much more extensive literature elsewhere. This law concerns itself with the question whether a given species occurs, or does not occur.

Sample octave	1-2	2-4	4-8	8-16	Specimens/Species
<i>Case (1)</i>					
Most probable representation	96	96	87.2	—	Number of species in the octave
“Perfect” representation	100	96.1	85.2	—	Number of species in the octave
<i>Case (2)</i>					
Most probable representation	45.5	64.0	81.5	94.7	Number of species in the octave
“Perfect” representation	52.7	69.8	85.2	96.1	Number of species in the octave
<i>Case (3)</i>					
Most probable representation	69.3	44.4	27.3	16.0	Number of species in the octave
“Perfect” representation	52.7	36.8	23.7	14.1	Number of species in the octave

In all cases, though there are errors in the first two or three octaves, representation becomes close to perfect beyond this point: and since most samples tend to ap-

Twenty-five test areas of equal size are examined, which areas may be a square meter in size, or one-tenth of a square meter, or of some other size. The only

question involved is whether species  $S_1$  (say) is present in all quadrats, or only in some, and if so in how many. It does not concern itself (in its original form)

with the question whether it is represented by thousands of specimens or only one.

The species  $S_1, S_2$ , etc., are then clas-

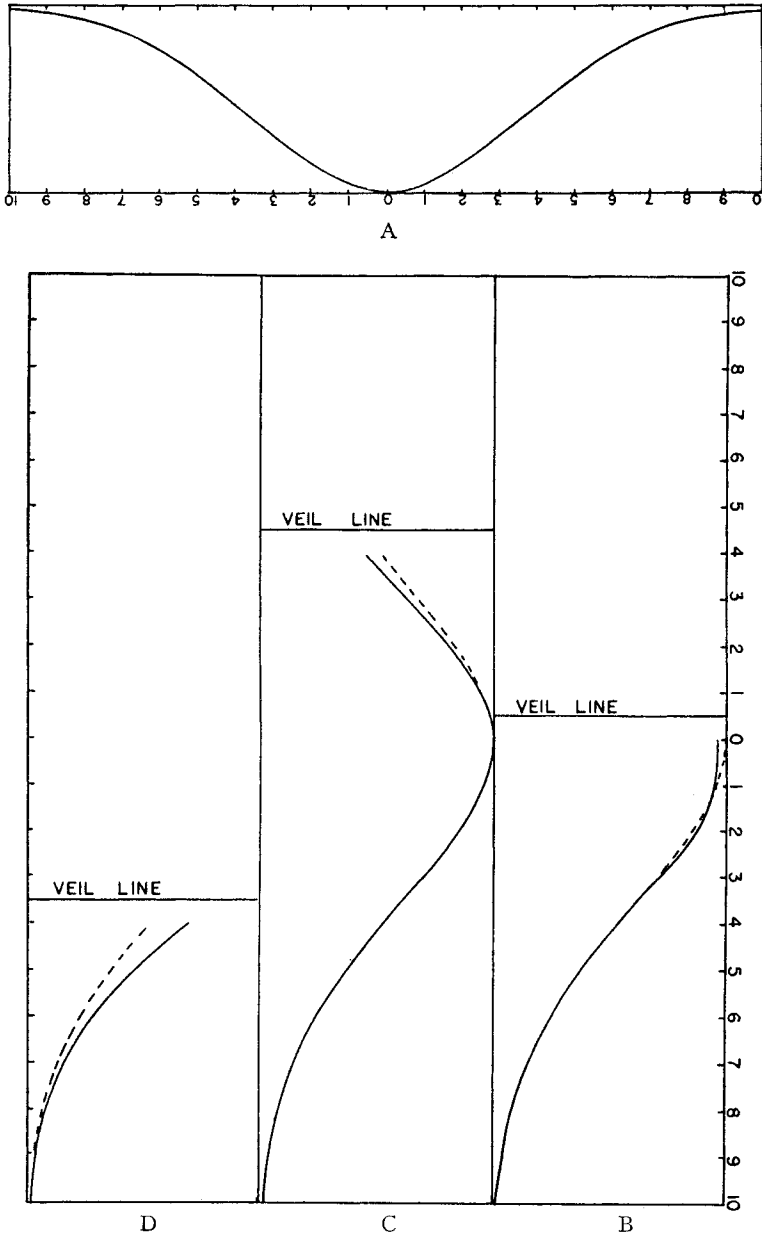


FIG. 8. A: a typical Gaussian curve. B, C and D, approximate relation of sample (solid line) to universe (broken line) with different degrees of decapitation (i.e., different sizes of sample). The samples follow the universe curve fairly closely.



sified into groups: Group A is represented in from 1 to 5 of the quadrats, B in 6 to 10, C in 11 to 15, D in 16 to 20, E in 21 to 25. Raunkiaer found that if the quadrat size were appropriately chosen, then group A was the largest, B less, C less still, but E tended to be larger than D. This sway-backed curve is usually described as "having two peaks."

It will be observed that what is directly measured is not commonness as understood in the present paper, where we are concerned with the numerical representation of a species in the sample or in the universe from which the sample is drawn; what Raunkiaer is concerned with is a measurement of *ubiquity*.

Gleason is quoted by Kenoyer as having called attention in 1920 to the relation existing between the actual number of individuals of the species, and the frequency index (Raunkiaer group) thereof.

But though there is a reasonable presumption that a ubiquitous species is a commoner one than a more local species, there is no *necessary* connection between the two things. We might have two species, buttercups and daisies. One buttercup occurring in each of 25 quadrats will rank the species as of a very high frequency; 25,000 daisies on a single quadrat, but missing from the other 24, will rank it as of a low frequency. Unsocial insects like the praying mantis, which on account of its cannibalism can be found only in a well-scattered arrangement, may be ranked as "frequent" but a colonial species like the bald-faced hornet, or an ant, might be infrequent, though at the nest there might be many thousands of individuals.

However, subject to biological factors of this sort, it still remains true that on the average there will be a definite relation between commonness and ubiquity, and it is easy to pass from the one to the other.

The Raunkiaer Index of Frequency is a measure of the probability that the species will occur (i.e., be represented by one or more specimens) in any one sam-

ple quadrat. Thus if the species occurs in 6 quadrats out of the 25, i.e., in 24% of them, the probability of its occurring in any one quadrat chosen at random is 0.24.

In a Raunkiaer quadrat or its equivalent, there may be one or more species to which the Poisson distribution laws do not apply. The condition for applicability is that there must be room in the quadrat for several times as many individuals as actually occur: this room will be non-existent if the species is already tightly packed in a dense stand, or if the individual bulks so large that it occupies a substantial part of a quadrat: it will also be non-existent for territorially-minded birds, and the like, where biological or ecological factors negate the possibility. The proof of the pudding, however, is in the eating, and experiment shows that normally such species, in a Raunkiaer test, are few or absent, and that therefore we can apply the Poisson distribution series with a sufficiently high degree of accuracy to all species.

In a Raunkiaer analysis, the great majority of species must be *absent* from one or more quadrats out of twenty-five, and in a random set of samples this implies that the *average* number of specimens of any one species in a *single* quadrat will rarely be as high as five.

Thus if  $p$  be the "average expectation" or "perfect representation," the series

$$e^{-p} \left( 1 + p + \frac{p^2}{2!} + \frac{p^3}{3!} + \text{etc.} \right)$$

gives the probability that the species will be represented by

$$0 \quad 1 \quad 2 \quad 3$$

specimens in any one quadrat.

We are concerned now only with the first term, representation by zero specimens. The probability that the species is completely unrepresented in a sample (quadrat) chosen at random is simply  $e^{-p}$ . Therefore the probability that it is *not* unrepresented is  $(1 - e^{-p})$ .

This is true whatever the shape or character of the universe may be.