ECOLOGY

VOL. 43 SPRING 1962 No. 2

THE CANONICAL DISTRIBUTION OF COMMONNESS AND RARITY: PART I

F. W. PRESTON
Preston Laboratories, Butler, Pennsylvania

POSTULATES AND THEORY

Introduction

In an earlier paper (Preston 1948) we found that, in a sufficiently large aggregation of individuals of many species, the individuals often tended to be distributed among the species according to a lognormal law. We plotted as abscissa equal increments in the logarithms of the number of individuals representing a species, and as ordinate the number of species falling into each of these increments. We found it convenient to use as such increments the "octave," that is the interval in which representation doubled, so that our abscissae became simply a scale of "octaves," but this choice of unit is arbitrary. Whatever logarithmic unit is used, the graph tended to take the form of a normal or Gaussian curve, so that the distribution was "lognormal." We called this the "Species Curve."

In the present paper we take up a point merely mentioned in 1948 that not only is the distribution lognormal, but the constants or parameters seem to be restricted in a peculiar way. They are not fixed, but they are interlocked. The nature of this restriction and interlocking is the main theme of the present paper.

In the earlier paper we graduated the experimental results with curves of the form

$$y = y_0 e^{-(aR)^2} \quad (1)$$

where y is the number of species falling into the R^{th} "octave" to the right or left of the mode, y_0 is the number in the modal octave, and a was treated as an arbitrary constant, to be found from the experimental evidence. This constant is related to the logarithmic standard deviation σ by the formula

$$a^2 = 1/2\sigma^2 \quad (2)$$

and we noted that it had a pronounced tendency to come out at a figure not far from 0.2, so that σ would have a value of about 3.5 octaves, and, since there are 3.3 octaves to an "order of magnitude," σ would be a little more than an order of magnitude.

If we now make a 2nd graph in which, using the same abscissae, we plot as ordinate not the number of species (y) that fall in each interval but the number of individuals which those y species comprise, we get another lognormal curve with the same standard deviation as the first graph, but with its mode or peak displaced to the right. This we call the "Individuals Curve."

Though we can use a Gaussian curve to "graduate" the observed points of the species curve, the curve extends infinitely far to left and right, while the number of observed points is necessarily finite. This is a common situation in statistical work and usually causes no complications, but in our problem there results an additional piece of information. The Individuals Curve necessarily lacks at least part of the descending limb. It terminates over the last observed point of the Species Curve, and this is long before the Individuals Curve begins to become asymptotic to the horizontal axis. In the earlier paper we noted that, as a matter of observation, it seems to terminate at its crest, so that, in effect, only half of the curve is present.

What I failed to observe in 1948 was that when the Individuals Curve terminates at its crest or very close to it, the value of "a" in equation (1) and of the standard deviation "σ," is fixed within narrow limits, and this value is in fact the one actually observed. This does not mean that "a" is a true constant, but only that it is not independent of y_0 or of the total number of species, N. It may be said that "a" is a function of y_0, so that given one, the other is settled. Thus we are reduced from 2 seemingly disposable parameters or constants to one. More generally, given any one piece of information about our collection, for instance given either the total number of species or the total number of individuals, everything else is fixed.

The word "Canonical"

I have ventured to call such an equation "canonical." It appears likely that this term was introduced into mathematical physics by J. Willard Gibbs: I quote from the preface to Volume 2 of his Collected Works (1931): "We return to the consideration of statistical equilibrium . . . we consider especially ensembles of systems in which the logarithm of probability of phase is a linear function of the energy. This distribution, on account of its unique importance in the theory of statistical equilibrium, I have ventured to call *canonical*" (italics his).

By a sort of rough analogy, I have designated as "canonical," for ecological purposes, that particular lognormal distribution of the abundances of the various species (or genera, families, etc.) whose "Individuals Curve" terminates at its crest. This way of describing it is probably imperfect, and, as shown later, it apparently corresponds to a situation in space or time where the individuals, or pairs, are distributed at random, not clumped on the one hand nor over-regularized on the other. Thus a better definition may be possible, and in that case preferable; but, however defined, it is a distribution that seems to have special importance in the general theory of ecological ensembles, and so I have ventured to call it "canonical."

In the present paper we trace the consequences of assuming the distribution canonical. We examine how nearly the experimental results fit the purely theoretical curves. These experimental results, though some of them involve "collections" having many millions of individuals, take us only as far as a few hundred species. We attempt to estimate what would happen if we had thousands or scores of thousands of species, where we have no actual counts of individuals but may sometimes be able to estimate them roughly, and we draw such other tentative conclusions as occur to us.

As the scale of abscissae we may use octaves, which is equivalent to taking "logarithms to the base 2," as we did in 1948, or we may use "orders of magnitude" which is more convenient when dealing with very large numbers. This is equivalent to taking logarithms to base 10 as James Fisher (1952) did. For theoretical purposes a scale of natural logarithms (i.e to base 2.718) would be most convenient. Williams (1953) found it convenient to work with logarithms to base 3.

Consequences of assuming that the individuals curve terminates at its crest

This matter may be stated briefly, anticipating the more detailed statement of the next heading, as follows: As shown in Preston (1948) the distance between the crests of the Species and Individuals Curves is $\ln 2/(2a^2)$ or $(\ln 2)\sigma^2$, where σ is the logarithmic standard deviation in octaves.

Though we describe our distribution as lognormal, it actually is finite, and species and individuals are not found infinitely distant from the mode either to the right or the left. In industrial "quality control" work it is customary to say that not more than one specimen out of a thousand should be expected beyond the 3-sigma limit, but in none of the biological examples we have yet encountered do we have as many as a thousand species to work with. We should therefore expect the finite distribution to end short of the 3-sigma limit. In fact, with the number of species in our examples to date we should expect it to terminate at about 2.5 to 2.8 sigma. If we take the latter value, the assumption that the distributions terminate where the Individuals Curve reaches its crest is equivalent to setting

$$(\ln 2)\sigma^2 = 2.8\sigma, \text{ or } \sigma = 4.0 \text{ octaves approximately.}$$

But this is just about what we find in our observations; it corresponds to an "a" value of 0.175.

Thus by an appeal to observation, but not by pure theory, we can reach the conclusion that very often the finite distribution does in fact end just about where the Individuals Curve reaches its crest. This seems to make it advisable to restate the matter more formally, in order to cover the complete range of possible values of the number of species involved.

The Individuals Curve

In the Species Curve each octave contains a certain number of species and each of these is represented by roughly the same number of individuals. Multiplying the one figure by the other gives the total number of individuals that have, in effect, been assigned to that octave. By making this computation for each octave we can construct the "Individuals Curve." This can be done for the observed points, but here we are concerned with its theoretical form.

For the Species Curve we have, as in equation (1) above

$$y = y_o e^{-a^2R^2}.$$

Let the number of individuals per species, the "representation" of the species, be n_o at the modal octave. Then at the R^{th} octave from the mode it is $n_o 2^R$ individuals per species, and the octave holds $yn_o 2^R$ individuals, or:

$$Y = (n_o y_o)[2^R e^{-a^2 R^2}] = (n_o y_o)[e^{R \ln 2} e^{-a^2 R^2}]$$
$$= n_o y_o e^{\left(\frac{\ln 2}{2a}\right)^2} e^{-a2\left(R - \frac{\ln 2}{2a^2}\right)^2} \qquad (3)$$

This is a lognormal curve with the mode displaced by an amount $\frac{\text{In} 2}{2a^2}$ octaves: it has the same dispersion constant as the Species Curve, and it has the modal height $Y_o = n_o y_o e^{\left(\frac{\ln 2}{2a}\right)^2}$.

(See Figure 1.)

Equations like (3) become somewhat simpler in appearance if we use "natural orders of magnitude" in place of "octaves," i.e. if we use intervals in which the frequency or abundance of a species increases in the ratio 2.718 instead of 2.0, and if we use σ, the standard deviation in orders of magnitude, instead of the coefficient a.

This method of working is convenient if we have available adequate tables of natural logarithms, for then the observed frequencies are easily classified into their natural orders of magnitude. I think, however, that it will be more convenient if we continue, as we have begun, by using "octaves," which do not depend on the availability of such tables, or on the alternative method of converting ordinary logarithms to natural ones.

The effects of a finite number of species

Referring to the Species Curve in Figure 1, we note that in theory this curve extends infinitely far both to left and to right, but the long "tails" are exceedingly close to the R axis as asymptote. The area under the curve, or the integral of the curve, represents the number of species we have accumulated, as we go from minus infinity to any given point. This area is at first so small that not until we are within about 9 octaves of the mode (for this particular case, where we have a total of 178 species) have we accumulated enough area to correspond to a single species. This is the beginning of the real, finite, distribution. As we continue to the right we accumulate species rapidly; then we pass the mode and accumulate them increasingly slowly. Finally we reach a point some 9 octaves to the right of the mode where the remaining area is scarcely enough to hold one more species. In practice,

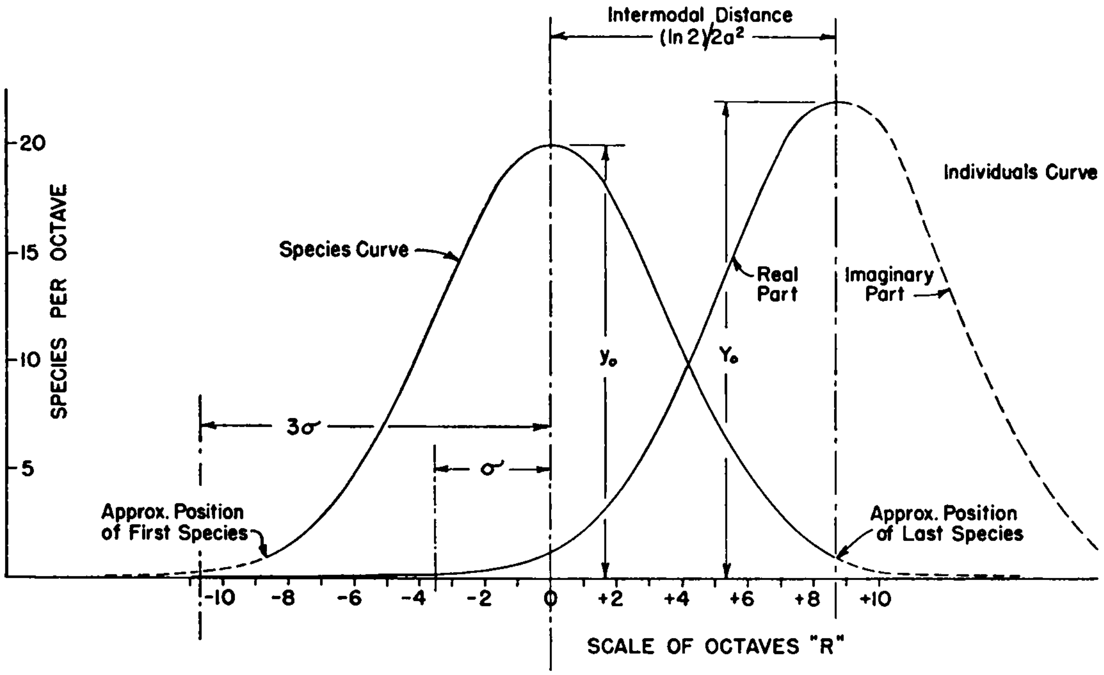

FIG. 1. The canonical distribution for an ensemble of 178 species. For this number of species, the coefficient a = 0.200 and the standard deviation σ is 3.53 octaves for both "Species" and "Individuals" curves. The modal height of the species curve is $y_0 = 20$ species. The ordinate scale for the individuals curve is arbitrary: see text for explanation. The intermodal distance is 8.68 octaves. The real part of each curve is drawn solid: the first (rarest) and last (commonest) species ought to lie at about this distance (8.68 octaves or 2.45 σ) from the mode of the species curve, or perhaps a little, but only a little, more.

the finite distribution ends at, or near, this point.

The numerical value of the integral between any specified limits can be found from published tables, but the integral itself cannot be expressed readily in analytical form, suitable for finding a general solution to our problems. Nor, if it could, would it define the end of the distribution with real precision; at best it can only give an idea of the most probable position of the end of the distribution. We have taken as the most probable position that point where the remaining area under the tail of the curve corresponds to half a species. This mean that there is a 1:1 chance that we have not quite reached the end or that we have just passed it.

In Table I we have given, for various values of the total number, N, of species involved, that value of x, or R_{max}/σ,[1] which makes

$$\frac{1}{\sqrt{2\pi}}\int_{-x}^{x} e^{-q^2/2}\,dq = \frac{N-1}{N} \qquad (4)$$

That is to say, we have left one species out of N to be divided between the two tails, to left of $-x$ and to right of $+x$, or half a species per tail. The values are taken from Lowan (1942).

TABLE I. The parameters of the canonical ensemble

N	x	σ	a	y_o	R_{max}	r	r^2	I/m
100...	2.576	3.72	0.190	10.7	9.58	765	5.8×10^5	2.69×10^6
200...	2.807	4.05	0.175	19.7	11.4	2,720	7.40×10^6	3.75×10^7
400...	3.024	4.36	0.162	36.6	13.2	9,410	8.85×10^7	4.82×10^8
800...	3.227	4.66	0.152	68.5	15.0	32,770	1.07×10^8	6.23×10^9
1,000...	3.291	4.75	0.149	84.0	15.6	49,670	2.47×10^9	1.47×10^{10}
2,000...	3.481	5.02	0.141	158.9	17.5	185,400	3.44×10^{10}	2.16×10^{11}
4,000...	3.662	5.28	0.134	302	19.3	645,500	4.17×10^{11}	2.75×10^{12}
8,000...	3.836	5.54	0.128	576	21.2	2,409,000	5.80×10^{12}	4.02×10^{13}
10,000...	3.891	5.61	0.126	711	21.8	3,651,000	1.33×10^{13}	0.93×10^{14}
100,000...	4.418	6.38	0.111	6,130	28.2	2.88×10^8	8.29×10^{16}	6.61×10^{17}
1,000,000...	4.892	7.06	0.100	56,500	34.5	2.43×10^{10}	5.90×10^{20}	5.21×10^{21}

(Note. ln 2 = 0.69315 and the reciprocal of this is 1.443) σ is to be ascertained by multiplying x by 1.443.
Then a is to be ascertained by dividing 0.707 by σ.
Then y_o is to be ascertained from the formula $y_o = 0.3989\ N/\sigma$.
R_{max} is to be ascertained by multiplying x by σ.
$r = 2\ R_{max}$
$I/m = 1.25\ r^2\ \sigma$

It can be shown that this point is very nearly the same as would be obtained by setting $y = 0.4$ (species per octave) in equation (1), which would give

$$y = 0.4 = y_o\, e^{-(aR)^2} \text{ or } (aR)^2 = \ln(2.5\, y_o) \qquad (5)$$

We now have to trace the consequences of assuming that the crest of the Individuals Curve coincides with the finite end of the Species Curve.

Estimating the canonical constants

We have seen that the distance between the modes of the species and individuals curves is $\ln 2)/2a^2 = \sigma^2 \ln 2$, and from Table I we have the values of R_{max}/σ or "x." Equating the two we have

$$R_{max} = x\sigma = \sigma^2 \ln 2 \text{ or } \sigma = x/\ln 2 = 1.44\ x \qquad (6)$$

whence

$$R_{max} = 1.44\ x^2. \qquad (7)$$

Since we already have the values of x for various values of N, the total number of species, we are now in a position to add to Table I the values of σ and of R_{max}, and this has been done. Further, since $a^2 = 1/2\sigma^2$ or $a = 0.49/x$, we can also add the value of a. Again, the number of species (y_o) in the modal octave of the Species Curve can also be computed, for

$$N = y_o\, \sigma\sqrt{2\pi}$$

so that

$$y_o = 0.399\ N/\sigma = 0.277\ N/x. \qquad (8)$$

This value, also, is therefore added to Table I and this completes all the unknowns. Given the total number of species, the first column of Table I, we can ascertain all the constants of equation (1). The basic equation of the distribution has therefore become canonical, in the sense that nothing is left to chance; once N is specified, everything is determined.

The relation between total individuals and total species

The canonical equation implies that there is a definite relationship between these 2 quantities, and if "m," as defined below, is known, it is easily calculated. When we permit ourselves the liberty of making this theoretical estimate, we can expect only rough agreement with it in practice in any particular instance. The actual termination of the Species Curve may be some distance from where our estimates place its "most probable" position and, as shown later, the crest of the Individuals Curve is not precisely at its termination if "contagion" is present. The computation however is worth making; it may lead to some understanding of the problem.

We have denoted by n_o the number of individuals in the modal octave representing a single species. Let us denote by R_{max} the "range," in octaves over which the finite distribution of species extends to left or right of the mode of the Species Curve. In a complete "universe" or lognormal curve, the range should be the same to left or right, though samples usually show a truncation at the left end.

[1] Note that this defines x as the half-range in terms of the logarithmic standard deviation, σ, as the unit.

The commonest species, located at $+R_{max}$, ought to have $n_0 2^{R_{max}}$ individuals. Similarly, the rarest, located at $-R_{max}$, ought to have $n_0/2^{R_{max}}$. If we denote the number $2^{R_{max}}$ by r, the commonest species has n_0r individuals, and the rarest n_0/r. The ratio of the numbers of individuals of the commonest to those of the rarest is then r^2. Since we know R_{max}, approximately, we can add to Table I the values of r and of r^2.

Now the rarest species must include one individual or one pair, and if we suppose the species to be viable, we probably must assume that it is somewhat more than this. We can call its number of individuals (or pairs), "m," for minimum, where m is probably a rather small number, without committing ourselves immediately to an estimate of m. Then the modal species will have mr individuals per species, and the commonest will have about mr^2. In equation (3) we have denoted mr^2 by the symbol Y_0.

The total number of Individuals in the whole ensemble is

$$I = \tfrac{1}{2}\sqrt{\pi}\, Y_0/a = \tfrac{1}{2}\sqrt{2\pi}\, Y_0\, \sigma$$

whence

$$I/m = \tfrac{1}{2}\sqrt{2\pi}\, r^2\, \sigma = 1.25\, r^2\, \sigma \qquad (9)$$

This value of I/m, as a function of the total number of species N, is given in the last column of Table I.

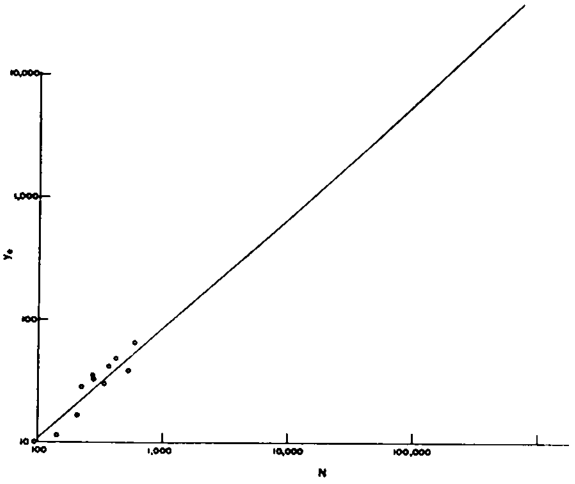

FIG. 2. The relation between the number of species (y_0) in the modal octave and the total number of species (N) in the ensemble. The solid line is the computed relationship from equation 18, and is pure theory. The observed points are taken from Table II.

In Figure 2 we have plotted y_0 as a function of N, using "log-log" plotting. The curve is substantially a straight line. Observed points from earlier work, and from computations made on the estimates of Fisher and of Merikallio, are plotted at the lower end of the curve, which seems to lie reasonably well among them.

Note that the line we have drawn is a purely theoretical one. It is not drawn to fit the observed points which were added after the curve had been drawn. Unless the theory bore some relation to the facts we might very well find all the points to one side of the curve and not indicating much the same slope as that of the theoretical line. The fact that in position and in slope the theoretical line, at its lower end, lies close to where we should place an empirical line drawn among the points, is encouraging.

In Figure 3 we plot the logarithmic standard deviation σ and the coefficient "a," against N, using semi-logarithmic plotting. In Figure 4 we plot the value of R_{max}, or half-ranges against N. It appears that R_{max} is nearly a rectilinear function of log N. In Figure 5 we plot r, r^2, and I/m as functions of N, log-log ploting. All 3 lines are nearly straight.

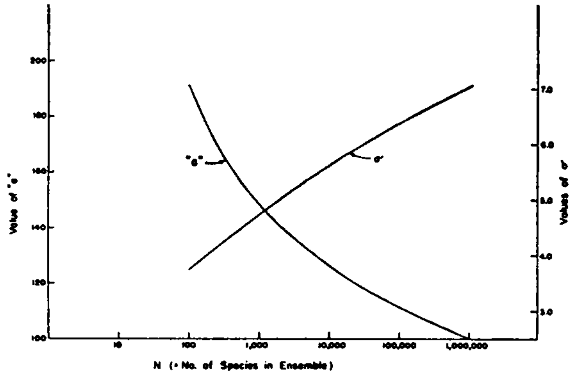

FIG. 3. The relation between the coefficient "a" in equation 7, or the standard deviation σ, and the total number of species (N) in the ensemble. (Computed relationship)

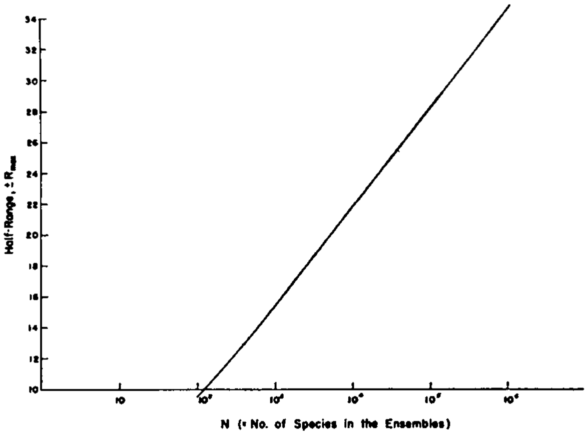

FIG. 4. Half-range (± R_{max}) as a function of N, the total number of species. The "Range" is the number of octaves which the finite distribution may be expected to cover.

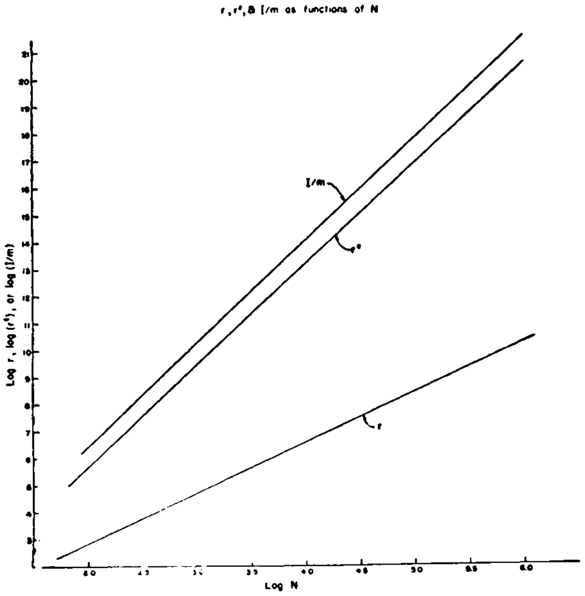

FIG. 5. The relation of r, r^2, and I/m to N. Here r is the value of $2R_{max}$, I/m is the total number of individuals divided by that minimum number of individuals (m) that may be assumed necessary to keep a species in existence, and N is the total number of species in the ensemble.

The equations of these straight lines, as determined by a least-squares fitting to the data of Table I are

$$\log r = 1.872 \log N - 0.875 \quad (10)$$

$$\log r^2 = 3.745 \log N - 1.751 \quad (11)$$

$$\log I/m = 3.821 \log N - 1.21 \quad (12).$$

These equations are, as we have seen, all intimately connected, and if one has to be modified, all must.

Note that these equations are only approximations, because the lines are not quite straight. Equations (10) and (11) are probably close enough for any purposes that I can foresee; equation (12), particularly in its inverted form where it relates log N with log (I/m), may usefully be modified slightly for different ranges of the value of N (See below, under "The Species-Area Curve.").

The relation between I and N in a complete ensemble

We have, in equation (12) a relation between I and N if we can make a reasonable estimate of the value of m. We have tentatively identified m as the number of individuals or pairs in the rarest species and, since species are all the time being exterminated, we may expect that m will be a number not far from unity. In practice we often find it so as shown below. But a full discussion of this would involve many biological considerations. For instance we find in practice that m is frequently less, even appreciably less, than unity, and the temporary interpretation we have given then has no meaning. Another, related, meaning can be given to m but the simplest interpretation for the present is a purely mathematical one. Any 2 of the 3 quantities N, y_o, and σ (or "a"), define the size and shape of the Species-Curve, but they do not define its position along the axis of x, i.e. of R. The quantity "m" specifies this position. It thus determines not merely the number of individuals theoretically representing the rarest species, but also the number of individuals representing any of the other species, including the commonest. In consequence, the relation is really between (I/m) and N, not between I and N directly, but if we know, or can estimate, I and N, we can get an estimate of m.

The Species-Area equation

There is one other formula that may be useful to us. In some cases we may regard individuals (or pairs of birds), as being distributed uniformly, statistically speaking, over wide areas. Let the density of individuals (or pairs) be ρ per acre. Then the formula

$$I = \rho A \quad (13)$$

gives the number of individuals or pairs to be expected on an area of A acres.

We can substitute this in equation (12) and obtain

$$\log N = 0.262 \log (\rho A/m) + 0.316 \quad (14)$$

$$N = 2.07 (\rho/m)^{.0262} A^{.0262} \quad (15)$$

This is the Species-Area Equation under "ideal" conditions such that the area we consider is populated with a complete, not a truncated, lognormal ensemble, and that the density of the population (ρ) does not change substantially over the range of areas we are considering. The first stipulation is important because in contiguous areas, for which Species-Area Curves are often propounded (and this includes my own companion paper on "Time and Space and the Variation of Species"), the smaller areas act much like "samples" of larger ones.

If we are dealing with isolates that take the form of complete canonical ensembles, and if ρ and m are substantially unchanged from isolate to isolate, then equation (15) may be written.

$$N \propto A^{0.262} \quad (16)$$

and since, 0.262 is not far from 1/4, this may be written

$$N \propto \sqrt[4]{A} \text{ approximately} \qquad (17)$$

which I have referred to as "the fourth-root law." This is useful for quick calculations.

As a matter of fact, as mentioned earlier, the exponent 0.262 is an average index over the whole range shown in Fig. 5 or Table I. It was obtained by comptuing a theoretical regression line from that table. Over the range for which we have some observational data, i.e. when the number of species is between 100 and 1000, the index as computed from theory is more nearly 0.270, and our equation becomes

$$N = 1.83\ (\rho/m)^{0.270}\ A^{0.270} \qquad (18)$$

At the present time, therefore, equation (18) is more useful in practice than (17), and this is graphed in Figure 6 for various values of ρ/m, and tabulated in Table II.

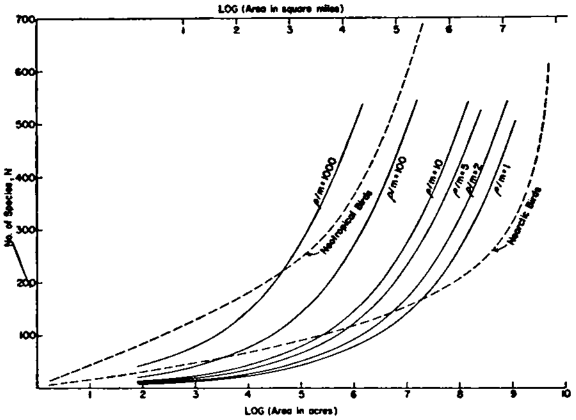

FIG. 6. The Species Area Curve for Isolates (solid line) and for Samples (broken line). Here ρ is the areal density (e.g. number of birds per acre) and m is the minimum number of individuals assumed necessary to keep a species in existence.

TABLE II. Number of species to be expected on area (A) of different sizes when given the density (ρ) of individuals per unit area and the minimum number (m) of individuals needed to keep a species in existence

A in acres ρ/m	10	10^2	10^3	10^4	10^5	10^6	10^7	10^8	10^9	10^{10}
1....	3.4	6.3	11.8	22	41	76	142	265	493	917
2....	4.1	7.6	14.2	27	49	92	171	320	595	1107
5....	5.5	10.3	19.2	36	67	124	230	430	796	1490
10....	6.3	11.8	22	41	76	142	265	492	917	1710
100....	11.8	22	41	76	142	265	492	917	1710	3180
1,000....	22	41	76	142	265	492	917	1710	3180	5920
10,000....	41	76	142	265	492	917	1710	3180	5920	11×10^3
100,000....	76	142	265	492	917	1710	3180	5920	11×10^3	20.5×10^3
1,000,000....	142	265	492	917	1710	3180	5920	11×10^3	20.5×10^3	38.3×10^3

In all probability, for the range from 50 to 200 species, the index is slightly higher still, perhaps 0.280. A number of our counts are within this range, and we have to consider them also when we deal with families rather than species. However the differences between 0.262 and 0.270 or even 0.280 are small compared with the uncertainties of our experimental or observational information. When we are dealing with areas that are merely samples of larger areas the index will tend, except for very small samples, to fall well below 0.262.

Fictive areas

Not all areas of equal size are equal in carrying capacity. A thousand square miles of the Greenland ice-cap are not equivalent to a thousand square miles of Neotropical forest, for instance. The treatment given above assumes that the areas we consider are reasonably comparable in fertility or, more generally, in "carrying capacity." The problems presented when this is not the case are discussed below where the concept of equivalent or "Fictive" areas is introduced.

Theoretical Species-Area Curves (for complete lognormal ensembles)

The equation is $N = 1.83\ (\rho/m)^{0.270}\ A^{0.270}$ or $\log N = 0.262 + 0.270 \log (\rho/m) + 0.270 \log A$. Table II gives the values of N over the range $10 < A < 10^{10}$, or $1 < \log A < 10$, and for $\rho/m = 1, 2, 5, 10, 100, 1{,}000, 10{,}000, 100{,}000$, and 1,000,000. In Fig. 6 we graph the entries of Table II. The solid curves are all alike, merely displaced to left or right. For comparison there are added the observed curves for Nearctic and Neotropical birds, taken from the companion paper (Preston 1960). These 2 curves, in broken lines, represent the behavior of "samples," not of complete canonical ensembles, and their form is discussed later in the present paper.

COMPARISON OF THEORY AND OBSERVATION, AND THE SPECIES-AREA CURVE FOR ISOLATES

In this section and the next we shall be dealing with the properties of isolated "universes" or isolated "populations," hereinafter called simply "isolates," as contrasted with the properties of "samples." In general the isolate tends to give a distribution-curve that is symmetrical and looks like a fair approximation to a complete lognormal. On the other hand, if we have a "sample," which is a small fraction of an isolate and is a random sample thereof, the distribution tends to look like a truncated or decapitated lognormal, an unsymmetrical distribution, of which we showed examples in Preston (1948).

We have shown above the mathematical consequences of 2 assumptions; viz, that abundance is typically distributed lognormally among species, and that this distribtuion is Canonical in the sense that not all lognormals meet the requirement, but only those in which a definite relation exists between the number of species, (N), the number of species in the modal octave (y_0), and the logarithmic standard deviation (σ). We now consider what degree of confirmation or refutation is available from observation.

The Canonical Parameters

The relative constancy of σ and "a"

In any Gaussian distribution or, in our case, any lognormal distribution, we have from equation (8),

$$N = \sqrt{2\pi}\,(y_0\,\sigma) \qquad (19)$$

That is, the total number of species in the "universe" is proportional to the product of the number of species in the modal octave and the logarithmic standard deviation. As the number of species increases, y_0 or σ, or both, must increase. When the distribution is canonical both increase, but σ increases only slowly while y_0 increases rapidly. In fact if we double N, y_0 increases by about 85%, but σ by only 8% in the range of most interest to us.

Similarly, since "a," the "modulus of precision," is related to σ by the formula $\sigma^2 = 1/(2a^2)$, "a" also is relatively constant. This is what we found in Preston (1948).

The numerical values of "a" and σ

Not only are these values relatively constant over the accessible range of values of N, but Table III shows that, in this range, where N averages perhaps two or three hundred species, "a" is about 0.175 and σ is about 4 octaves, or something like one and a fifth orders of magnitude.

Now, referring to Table I, we see that for 314 species we should theoretically have an average value of "a" of about 0.169. This agreement is close, perhaps fortuitously so (see below on contagious distributions), but it warrants a few comments.

The only attempts to get a picture of the complete ensemble by direct observation are those of Fisher (1952) and Merikallio (1958). There are considerable difficulties with the experimental work and some uncertainties, some of which the authors have indicated. The other results come from estimates of what the "universe" is like as a result of studying a sample. Indeed neither Fisher nor Merikallio was studying a perfect "isolate," though they approximated it. The sample theoretically has the same modal height and the same dispersion as the universe, and it has also a 3rd variable, the position of the "Veil-line," or what is the same thing in the end, the abscissa of the mode. This 3rd disposable variable makes our estimates of the other 2 more uncertain than they would otherwise be, and therefore agreement in our estimate of "a" or σ within about 6% seems in part fortuitous. For a discussion of the fitting of truncated Gaussian distributions see Hald (1952). Furthermore, when we are dealing with truncated ensembles, we do not directly observe the value of N, the total number of species, but have to estimate it from the sample. This throws a further strain upon the interpretation of our observations.

TABLE III. (Observed Relationships). N is the number of species estimated, on the basis of the sample, to be present in the total "universe" or "population": y_0 is the number of species in the modal octave, and "a" is the "modulus of precision" of the lognormal distribution

Instance	N (estimated)	y_0	a	Reference
Saunders (birds).........	91	10	.194	Preston 1948
Dirks (moths)...........	410	48	.207	Preston 1948
Dirks (female moths).....	363	42	.205	Preston 1948
Williams (moths)........	273	35	.227	Preston 1948
King (moths)...........	277	33	.152	Preston 1948
Seamans (moths)........	332	30	.160	Preston 1948
Maryland birds..........	233	28	.213	Preston 1957
Nation-wide bird count...	530	38	.129	Preston 1958
Nearctic estimate........	600	65	.19	Preston 1948
Land Birds of England and Wales..........	142	11.2	.14	Fisher 1952
Breeding Birds of Finland.	204	16.5	.146	Merikallio 1958

The relation between y_0 and N

To the extent that we find the correct relationship between σ and N we must necessarily find a correspondingly correct relationship between y_0 and N but, since σ is so nearly constant over the observable range of N while y_0 varies rather rapidly, y_0 may throw some further light on the matter. Table III gives the observed values of y_0 for various values of N, most of which are estimated from incomplete or truncated distributions: Figure 7 shows the data in graphical form. It should be once more emphasized that the line is purely theoretical. The small circles represent observed points from Table III. The line is not "fitted" to the points and then extrapolated; the line is from theory, the points from observation. But it will be observed that the line passes neatly among the observed points which lie in a narrow band straddling the line so that if we did "fit" a line to the observations it would not be very different from the theoretical one. The observed points not only lie reasonably near the line, but they parallel its course. This gives us an additional point of agreement between theory and observation, beyond what we get from considering the relationship of σ and N.

The relation between N and I in a complete ensemble

Here we exclude discussion of the relation between N and I in "samples." The best way to get complete ensembles is probably to deal with "isolates" such as the fauna and flora of islands which are in internal equilibrium, but not necessarily in equilibrium with, or even appreciably affected by, the populations of other land masses. However, so far as I know, we have no counts of individuals for such islands, nor even good estimates of the numbers of individuals. Fortunately when our areas become reasonably large, of the order perhaps of 50,000 to 100,000 square miles, the distribution begins to approximate closely to that of a complete ensemble (in the case of birds) even though the area is in intimate contact with neighboring land areas. We have at least 2 fairly good estimates of the number of species and of individuals on such areas: James Fisher's (1952) estimate of the *land* birds of England and Wales, and Merikallio's (1958) estimate of the breeding birds of Finland.

The land birds of England and Wales

Fisher lists some 142 land birds as regular breeding species, and the distribution looks reasonably complete and symmetrical (see his Figure I). From our equation (12), we compute that I/m = 10,000,000 very closely. Fisher estimates that I is 63,000,000 *individuals,* hence m should be approximately 6 individuals or 3 pairs. This is, I think, a reasonable estimate of the number of pairs of the rarest regularly-breeding species.

The value of r^2 may be computed from formula (11) and comes out very close to 2,000,000. Then $mr^2 = 6.3 \times 2 \times 10^6 = 12.6$ million individuals. Fisher gives a value of 10 million approximately, so again the agreement is good. The value of r may be computed from formula (10) or taken as the square root of 2 million. Then we get mr = 8,900 individuals per species in the modal octave. Our computation from his data (which are not given in great detail) is 8,500. The agreement is almost too good.

The breeding birds of Finland

Merikallio (1958) lists about 204 species. On this basis we should expect I/m to be about 41 million. Merikallio gives a value of I of 31 million *pairs,* hence our computation suggests in this case that m is about one pair (0.76 pairs). We can also compute that r = 2,810, whence mr = 2.1×10^3 for the modal species and $mr^2 = 6.0 \times 10^6$ for the commonest. The figure Merikallio gives for his commonest species is 5.7×10^6 pairs, where once more the agreement is accidentally too good, and our computation from his figure gives the modal representation as 4.26×10^3, which is fair.

The relation between N and A: the Species-Area Curve for complete Canonical ensembles

In general there can be no complete count of individuals on areas large enough to approximate complete ensembles. Both Fisher and Merikallio use methods of estimating, not counting, individuals except for the rarest species. Their methods could of course be extended to other areas and other taxonomic groups and it might then be seen how nearly the theory fits the observations.

We can however assume that in certain parts of the world we have isolates, or near-isolates, of some taxonomic groups on islands of different sizes between which there is very limited floral or faunal interchange, and that on each of these islands there may be a fair approximation to a canonical distribution. These islands ought to be numerous enough for us to strike a reasonably good statistical correlation; they should be similar climatically and not too dissimilar in vegetation cover and soil. There are 2 obvious groups of such islands, the East Indies and the West, and there are taxonomic groups such as land mammals, land reptiles, and amphibia that do not readily cross stretches of sea. Birds qualify partially, and are usually better known. Unfortunately, the number of species on the smaller islands is usually far below what we like for statistical work. Elsewhere (Preston 1957) I have suggested that we ought to have something like 200 species, and this is out of the question for the groups mentioned. We must accordingly do the best we can with smaller numbers.

The mammals of the East Indies

Darlington (1957, p. 480) discusses the mammals of Sumatra, Borneo and Java. These islands are comparable in latitude and climate (a matter that cannot be neglected) and their areas are given as 167,000; 290,000; and 49,000 square miles respectively, while their "units" of mammals are 55, 47, and 33. "Units" in this case are a mixture of species and genera. Striking an average of the 2 larger islands gives us an imaginary island 228,000 mi^2 and 51 species. Comparing this with the smaller island of Java by means of the formula

$$z = \log (N_1/N_2)/\log (A_1/A_2) \text{ (from eq. (18)) } \quad (20)$$

where the suffix 1 applies to the one island and the suffix 2 to the other, we find that

$$z = 0.28$$

which is very close to what is called for by the theory of Equation (20).

Amphibia and reptiles of the West Indies

Darlington (1957, p. 483) in discussing the Anura and the lizards and snakes of Cuba, Hispaniola, Jamaica and Puerto Rico, with some support from the small islands of Montserrat and Saba in the Lesser Antilles, concludes that "division of area by ten divides the fauna by two." (The "herps" on 40,000 mi^2 total about 80, and on 4,000 mi^2 about 40.) Darlington's statement implies that

$$z = \log 2/\log 10 = \log 2 = 0.301$$

which is quite close to the theoretical value of 0.28. The smaller islands support Darlington's views, but since the number of species is very small, neither agreement nor disagreement would carry much weight.

It may be noted that if we take frogs and toads alone, since the legitimacy of lumping these with lizards and snakes may be questioned, we have an average of 25½ species on the 2 larger islands, with an average of 35,000 mi^2, and an average of 14½ species on the two smaller islands, with an average of 4000 mi^2. The computation then gives $z = 0.26$, which is also the theoretical value. The reptiles (snakes and lizards combined) give a value of 0.318; the lizards alone give a value of 0.294. The snakes are too few to be used safely for a caculation.

Birds of the West Indies

The populations of birds, particularly seafowl, on the various islands of the West Indies are probably somewhat less isolated from one another than the frogs and lizards. On the other hand there is quite likely not much gene-flow among a large proportion of the land-birds. The populations of each island may therefore approximate to canonical lognormals and the exponent z may be expected to approach 0.26-0.28, the theoretical value, or to be a little lower.

Bond (1936 and 1956, with supplements) has given detailed accounts of these birds, from which I attempted to compile a list of the breeding species of all the major islands and a number of the smaller ones. The areas I took from the Encyclopedia Britannica, edition of 1949. In the case of a few species, while Bond leaves no doubt that they breed within the West Indies, there may be some uncertainty whether they breed regularly on some of the islands. This is probably unimportant in the case of the major islands, where the birds are probably better known, but I may be in error in connection with some of the smaller ones. In order to permit others to correct me, I give in Table IV a list of the islands I have used, their areas and my estimates of the numbers of breeding species of birds, and in Figure 7 I have graphed the results. Note that I have excluded introduced species. Note also that the Isle of Pines has an abnormally rich fauna for its size. This probably comes about from its not being an "isolate," but rather a "sample" of Cuba, probably a truncated distribution.

The line is not a calculated regression line, but

TABLE IV. Birds of the West Indies

Island	Area in Square Miles	Breeding Species of Birds
Cuba............	43,000	124
(Isle of Pines).....	(11,000)	(89)
Hispaniola........	47,000	106
Jamaica..........	4,470	99
Puerto Rico......	3,435	79
Bahamas.........	5,450	c74
Virgin Islands.....	465	35
Guadalupe........	600	37
Dominica.........	304	36
St. Lucia.........	233	35
St. Vincent.......	150	35
Grenada..........	120	29

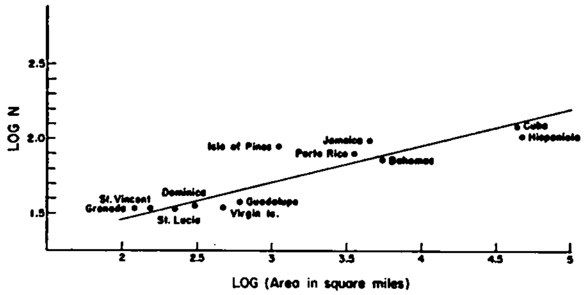

FIG. 7. Birds of the West Indies: Species-Area Curve. The abscissa gives the areas of the various islands; the ordinate is the total number of bird species breeding on each island.

is drawn "by eye" and seems reasonable; it is scarcely worth more refinement in view of the errors I may have made in my estimates or the general question of how legitimate it may be to include seafowl in a study of "isolates." Its slope corresponds to z = 0.24, which seems to agree well with expectations.

Birds of the East Indies

The "East Indies" here includes 2 or 3 islands of the Sunda Shelf, and several other islands not strictly on the shelf or even part of the same zoogeographical province, and one island, Ceylon, far removed to the west. However, the islands are all essentially tropical and climatically not too dissimilar. I have depended heavily on information supplied by Dr. Kenneth Parkes who has used, besides his own knowledge, the findings of Delacour and Mayr. The 1949 Edition of the Encyclopedia Britannica gives somewhat different counts of species.

A list of the islands with their areas and breeding species as given in Table V, and the results are graphed in Figure 8. The slope of the curve is z = 0.288, close to the theoretical value which in this case should be about 0.27. The curve is drawn "by eye," but seems likely to be as close as the observational evidence warrants. Good counts on some smaller islands might help but, in view of differences of opinion among taxonomists as to what constitutes a valid species, and in view also of the fact that with small islands statistical fluctuations or "errors" are likely to be important, I am not sure that much would be gained.

TABLE V. Breeding birds of some East Indian islands

Island	Area in Square Miles	Breeding Species of Birds
New Guinea......	312,000	540
Borneo...........	290,000	420
Phillipines (excluding Palawan)	144,000	368
Celebes..........	70,000	220
Java.............	48,000	337
Ceylon...........	25,000	232
Palawan..........	4,500	111
Flores............	8,870	143
Timor............	300 × 60 = 18,000	137
Sumba...........	4,600	108 (or 103)

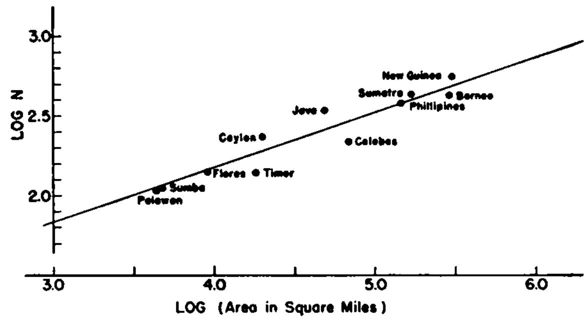

FIG. 8. Birds of the East Indies: Species-Area Curve. The abscissa gives the areas of the various islands; the ordinate is the total number of bird species breeding on each island.

It may be noted that Java is rich for its size, while Celebes is poor. Ceylon is relatively "rich," but it may be acting to some extent as a sample of India and thus be somewhat enriched. New Guinea is famous for its wealth of species but some fraction of these may be an enrichment from the Australian mainland. Mayr (1944) comments that Timor is poor in birds, being both peripheral and arid.

The birds of Madagascar and the Comoro Islands

Madagascar is one of the large islands of the world, with an area of 229,000 mi^2. The Comoro Islands lie off its northeast coast, in the Mozambique Channel, in Latitude 12°S. They are 4 small islands, averaging about 200 mi^2 apiece, and have received their bird fauna predominantly from Madagascar (Benson 1960). Madagascar has apparently been an isolated island since early Mesozoic (Triassic) times while the Comoros date back to the mid-Tertiary (Miocene) times. The 1949 edition of the Encylopedia Britannica gives about 260 species of birds for Madagascar. Rand (1936) gives his estimate as 237 breeding species. I have comprised on 250. Benson (1960, p. 18) gives the breeding bird tally as follows: on Grand Comoro (366 square miles), 35 species; on Moheli (83 square miles), 34 species; on Anjouan (178 square miles), 35 species; on Mayotte (170 square miles), 27 species.

Madagascar is roughly a thousand times as large as the average Comoro Island and there are no islands of intermediate size in the immediate neighborhood. We can, however, make the assumption that over the ages, many species of birds from Madagascar, Africa, and occasionally elsewhere have made landfalls on the Comoros, and that the islands have as large a number of species as they can support. On that basis we can plot them on a graph along with the avifauna of Madagascar, and this has been done in Figure 9. The line has been drawn by eye, and its equation is

$$N = 7.95\ A^{0.28}$$

The exponent, 0.28, is very slightly above the theoretical value, and the coefficient, 7.95, only a little below the theoretical value of 10, appropriate

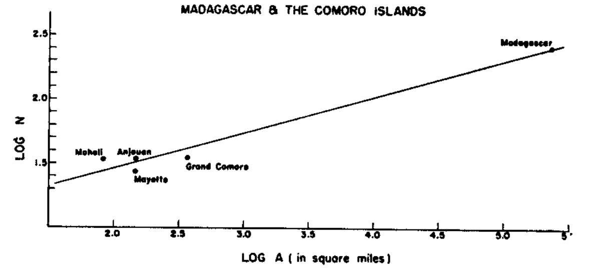

FIG. 9. Birds of Madagascar and the Comoro Islands: Species-Area Curve. The abscissa gives the areas of the various islands; the ordinate is the total number of bird species breeding on each island.

when A is in square miles and ρ/m is assumed to be unity.

The land vertebrates of islands in Lake Michigan

Hatt *et al.* (1948, p. 149) list 12 islands in Lake Michigan ranging in area from 2.5 acres to 37,400 acres, and in species-count of land vertebrates from 4 to 120. The 2 smallest islands with 8 and 4 species respectively are of dubious value for our purposes, but the remaining 10 islands, with a minimum of 19 species, seem satisfactory. Land Vertebrates are defined as amphibians, reptiles, birds and mammals. Table VI gives the information.

Figure 10 presents the data in graphical form for the 10 larger islands and a log-log plot yields a fair approximation to a straight line. The curve as drawn has the equation $N = 10\ A^{0.24}$ (where A is in acres). This implies a high density of individual "vertebrate animals" compared with what we encounter among birds. The slope, or exponent 0.24, is not far below the theoretical 0.27. If we include the 2 lowest points, a very questionable procedure, and give them equal weight with the other points, the slope of the curve increases to about 0.30. Giving them even a little weight brings the exponent close to the theoretical value. I am inclined to suspect that most islands have achieved a rough approximation to internal equilibrium, i.e. to a canonical distribution, but that there is a modest exchange of individuals among the islands or with the mainland, perhaps by water in summer or ice in winter. This tends to depress the index slightly below the value for strict isolates, and makes the islands to some extent "samples" of the mainland.

TABLE VI. Land vertebrates of islands in Lake Michigan

Area in Acres	Species	Area	Species	Area	Species	Area	Species
37,400	120	3,400	68	130	27	16	19
13,000	88	895	38	115	36	3	4
5,000	77	270	37	75	42	2.5	8

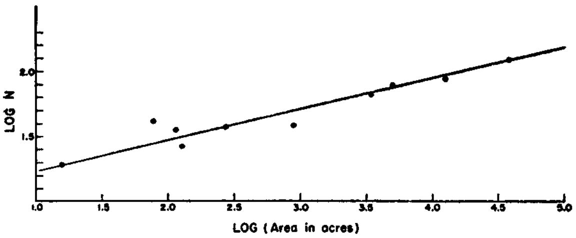

FIG. 10. Land Vertebrates of Islands in Lake Michigan: Species-Area Curve. The abscissa gives the areas of the various islands; the ordinate is the total number of bird species breeding on each island.

The land plants of the Galapagos Islands

So far we have been considering faunas, and it may be well to consider a flora. Kroeber (1916) has made a detailed study of 18 of the Galapagos Islands. He does not give the areas of any of the islands, and I have obtained these from other sources, which are not always in very exact agreement, or by estimates from the U.S. Hydrographic Survey maps. One small island, which he calls Brattle, I have not been able to identify with assurance, and this has therefore been omitted, reducing our count to 17 islands. Since we shall have occasion to analyse Kroeber's results further, I have here included an outline map of the islands as Fig. 11.

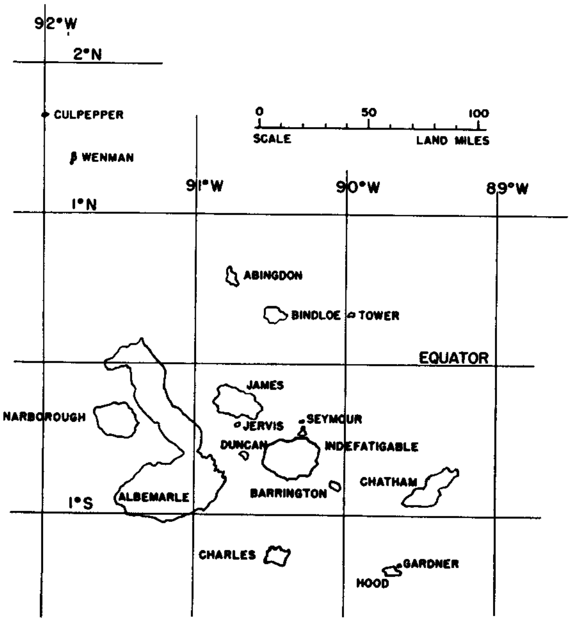

FIG. 11. Outline map of the Galapagos Islands.

We shall work principally with Kroeber's Table VI, from which our Table VII has been prepared,

TABLE VII. The land plants of the Galapagos Islands

No.	Name	Area (Sq. mi.)	Species
1.	Albemarle	2249	325
2.	Charles	64	319
3.	Chatham	195	306
4.	James	203	224
5.	Indefatigable	389	193
6.	Abingdon	20	119
7.	Duncan	7.1	103
8.	Narborough	245	80
9.	Hood	18	79
10.	Seymour	1	52
11.	Barrington	7.5	48
12.	Gardner	0.2	48
13.	Bindloe	45	47
14.	Jervis	1.87	42
15.	Tower	4.4	22
(16).	(Brattle)	—	(16)
17.	Wenman	1.8	14
18.	Culpepper	0.9	7

with an identifying number added for convenience, and with my estimate of the area of each island also added. There is a good deal of scatter, because area is not the only factor that decides the richness of faunas and floras, and accordingly we have computed 2 regression lines, one for the 12 larger islands and one for all 17. The two are scarcely distinguishable but this is a coincidence; too much attention should not be paid, as a rule, to very small populations. The regression line drawn in Figure 12 is the one for the 12 larger islands, but the "points" for the others are indicated. The slope of this line is z = 0.325, a little higher than the theoretical value of 0.27 or thereabouts.

Summary of the z-values for isolates

We now have 7 independent estimates of the index z of Species-Area curves for "isolates," all of them relating to islands. Some of these are oceanic islands, some are islands in a fresh-water lake, some are in tropical, some in temperate latitudes. We have floras and faunas, and among the faunas we have mammals, birds, reptiles, amphibians, and "land vertebrates."

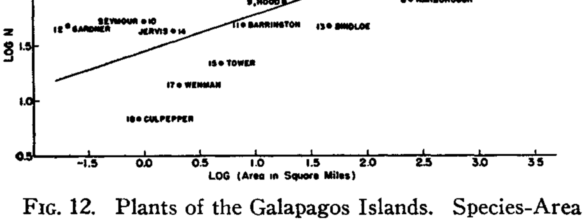

FIG. 12. Plants of the Galapagos Islands. Species-Area Curve. The abscissa gives the areas of the various islands; the ordinate is the total number of bird species breeding on each island.

Table VIII summarizes our results. It is clear that the average of the 7 z-values comes very close to the theoretical figure. The individual departures from this average figure are in general modest, and the agreement may be regarded as satisfactory. Departures may be expected because of several factors. For instance, large islands are more likely to have high mountains, and therefore special habitats, than small ones, and then area alone would not adequately describe the opportunities for faunas. Other factors are discussed below.

TABLE VIII. The Species-Area curve for isolates (The coefficient "z" of the Species Area Curve-Summary

Fauna or Flora	Locality	z-value	Fig. No.
Mammals.	East Indies	0.280	none
Amphibian and Reptiles.	West Indies	0.301	none
Birds.	West Indies	0.240	7
Birds.	East Indies	0.333	8
Birds.	Madagascar and Comoros	0.280	9
Land Vertebrates. . .	Islands in Lake Michigan	0.239	10
Land Plants.	Galapagos Islands	0.325	12
	Average (of 7). . .	0.285	
Land Plants.	Many areas	0.222	13

The flowering plants of the world

Williams (1943b) has plotted on a log-log basis all the data available to him for flowering plants of adequately defined areas, and these areas range from a few square centimeters to the land area of the globe. He has divided the information into 6 categories, the flora of arctic, temperate, subtropical, tropical, and desert regions and of oceanic islands, and he comments that over the linear portion of the graph, which extends from about 0.1 km^2 to 10^2 km^2, "to double the number of species, the area must be increased by thirty-two times." This may be compared with Darlington's statement that the area must be increased about ten-fold. Williams' statement amounts to saying that our exponent z = 0.20 approximately. Williams is working with the upper boundary of his scatter plot, the boundary between the area where there are observed points and the area where there are none. It happens that this boundary is quite well defined and Williams marks it with a broken line. The lower boundary on the other hand is nebulous. Williams treats the upper boundary as a statement of optimum conditions, i.e. as indicating the maximum number of species that can be accommodated on various sizes of "quadrat."

The slope is actually a little steeper than Williams' estimate; I make it 0.222. In Figure 13 I have reproduced all of Williams' points for areas between 0.1 and 10^8 km². Below this range many or most of the points must refer to "samples," i.e. to truncated distributions; it is virtually certain that a large proportion of the remainder do likewise. This gives us 135 points with which to work. I have drawn 2 lines on the graph, one marking the boundary as estimated by eye, and the other the regression line of "y on x" (log N on log A) computed by orthodox methods. The 2 lines turn out to have identical slopes, z = 0.222. The equation of the regression line is in fact

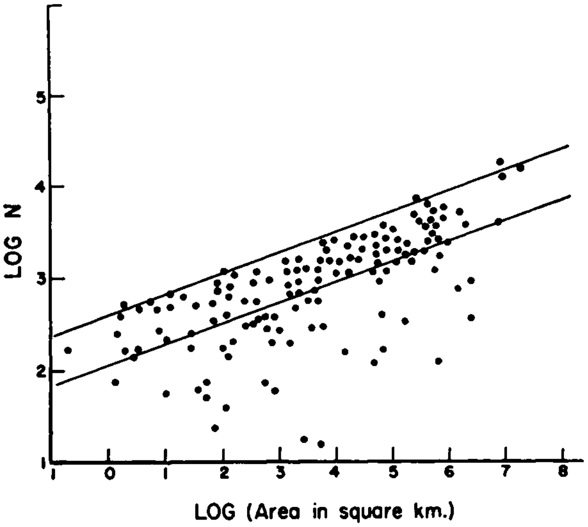

Fig. 13. Re-plotting of Williams' plants of many islands and other areas. Species-Area relationship. The upper curve is Williams' upper limit of maximum richness. The lower is the regression line for all plotted points. N is the number of species reported from the various areas.

$$N = 118\ A^{0.222} \qquad (21)$$

$$\log N = 2.07 + 0.222 \log A. \qquad (22)$$

By separating the categories into which Williams divides his data, it is possible to compute or estimate values of z for the tropical, temperate, and desert floras. I have merely estimated the slopes of the upper boundary lines. For tropical and desert regions, the slope seems to be about 0.25, for temperate regions, 0.22. The regression lines may be a little flatter. In any case they are all roughly similar, and all a little below the theoretical value, which over this range, which extends to more than 10,000 species, is about 0.265. We may speculate that though, on unit area we have approximately complete canonical distributions, yet some samples are truncated, thus depressing the index below its theoretical value, from 0.26 down to 0.22. It may be worth noticing that the depression is by no means so great as we found in Preston (1960) for the bird faunas of the Neotropical region (0.16+) or Nearctic (0.12+) where the truncation is presumably more pronounced.

We may also note that while the statements of Darlington and Williams, that it takes 10 times, or 32 times, the area to double the number of species, sound considerably at variance, yet when expressed as an index the difference is only between 0.22 and 0.30. It will not escape notice that the average of these two is 0.26, almost exactly the theoretical value, but I think this is accidental.

Our justification for ignoring those areas in Williams' graph below 0.1 km² lies less in the fact that Williams himself did so than in the fact that we here have many counts involving less than 50 species, and scarcely any with more than a hundred, and this brings us to about the limit for satisfactory statistical work. (This restriction makes it difficult to work with the extensive tabulations of Vestal 1949. However see below.)

There is another calculation we can make, of a somewhat more risky kind, since the index 0.222 differs appreciably from the theoretical one of 0.262. We can compute the theoretical number of individuals per square meter and see how it agrees with what Williams has to say on the subject.

The regression line not far from the middle of the range we have studied gives 1000 species at about log A = 4.2 or A = 1.6 × 10^4 km². From Table I we should have I/m = 1.47 × 10^{10} individuals when N = 1000 species. Then if we assume tentatively that m = 1 or thereabouts, this involves approximately one flowering plant to the square meter, averaged over tropical, temperate, and arctic areas. The boundary line would give as an "optimum" or maximum concentration something like 30 cm² to a flowering plant.

Williams says, on p. 260, "Available information indicates that there is about one plant per square centimeter in temperate grassland, about one per square meter in woodland, and about one per hundred square meters upwards in semi-desert areas." Thus our estimates fall well within the range of concentrations given by Williams, and since we have evidence that m can vary at least over the range from 10 to 1/10 or rather more, the agreement can only be regarded as satisfactory.

Summarizing this section, we might say that there are a number of cases where the theory of isolated canonical ensembles seems to be confirmed by the exprimental, or observational, evidence; or at least the observations seem to accord with the theory.

The Properties of "Samples"

Introduction

In the previous two sections we examined first the theory of complete canonical ensembles and then the available observations on ensembles that seemed likely to be substantially complete. Under ideal conditions we think of these as "isolates." Information of this sort is rare, and in general we have to content ourselves with "samples" drawn from a presumably complete canonical universe. The relationship between sample and universe was discussed briefly in Preston (1948). What we get, if the sample is a random one, is a truncated lognormal distribution. To a close approximation the sample has the same modal height (y_0) as the universe, and it indicates correctly the logarithmic standard deviation (σ). From these figures the total number of species (N) in the unknown universe may be computed, though we may have observed or collected, in the sample, only 70 or 80% of them.

A large part of the field work of an ecologist consists of taking samples, so the properties of samples have great practical importance. The purpose of the sampling is in nearly all cases to deduce something about the "population," "universe," or "stand" or "community" that is being sampled, so the relation between sample and universe is a matter of prime concern. Some of the problems that arise are of broad scientific interest. In large samples, which typically should contain at least a couple of hundred species and at least forty or fifty thousand individuals, the problems are largely analytical in nature; in small or very small samples, biological peculiarities, especially "contagion," positive or negative (See Cole 1946), become very important; with larger samples the "contagion" tends to smooth out and disappear.

A "sample" may be obtained by catching moths in a light trap. Obviously we do not catch all the moths that are on the wing, and it is clear that we are dealing with a sample, though it is far from clear what "universe" we are sampling. It is important not to assume that we know this from general considerations; in particular, in this instance, we must not assume that it is the population of a definite geographical area. Some of the noctuid and hawk moths we catch may be very strong fliers and may have crossed the English Channel or some larger body of water from foreign parts; some of the geometrids may be very weak fliers and originate entirely close at hand. Similar considerations apply to birds and even to plants, and even more obviously to wind-blown pollen. The universe being sampled is simply what the sample says it is; it must be ascertained from the internal evidence, not from assumptions. This applies to "quadrats" of plants, and hence to the Species-Area curves thereof, just as it applies to moths in a light trap.

Fundamental differences between sample and universe

There are 2 very obvious differences between a sample and a universe: first, the ratio of species to individuals is vastly higher in the sample than in the universe, and second, there are vastly more species represented by a single specimen ("singletons"), or by a few specimens, in the sample if it is a "random" one.

Let us suppose that we have a "fixed" universe which will not expand as we collect our random sample, and let us examine our collection as we collect it.

The first moth into a light trap may be one of the commonest species, but it is more likely to be one of some other, since the commonest species is not as plentiful as all the other species put together; hence it may very well be that at first we collect half a dozen moths all of different species. Thus we accumulate species, at first, as rapidly as we accumulate individuals. As we continue collecting this situation ceases to obtain, and after a time the addition of new individuals piles up indefinitely, while the addition of a new species is a rare event. We have reached the point of diminishing returns.

In Preston (1948) we showed that doubling the catch of individuals was approximately equivalent to withdrawing a graph of the lognormal distribution from under a Veil-line to a distance of one more octave. This is a very close approximation for most of the cases of practical importance, or at any rate those cases we were considering in 1948. But actually withdrawing the graph by one octave rather more than doubles the count of individuals. It doubles the count for all those octaves that were previously withdrawn and it adds in addition one individual for each species in the new octave now unveiled. This modification of our 1948 statement is of significance only at the beginning of collecting.

The index "z" in the Arrhenius Equation, when collecting at random from a fixed universe

The Arrhenius equation, in ecological work, is the statement that the number of species (N) is connected with the area (A) by the equation

$$N = KA^z$$

where K and z are constants. This equation is not given by Arrhenius but is implied by his work. (See Preston 1960). This is of the same form as our equation (16). Note that we are not referring to that other Arrhenius equation that concerns the rate of chemical reactions and their variation with temperature.

Suppose in our collecting we have begun at the octave R_{max} and have reached octave R, going from right to left across our Species-Curve.

The number of species we have collected is

$$Q = y_R + y_{R+1} + y_{R+2} + \dots + y_{R_{max}} = \sum_{y_R}^{y_{R_{max}}} (y) \qquad (23)$$

The number of individuals collected is

$$I = y_R + 2y_{R+1} + 2^2y_{R+2} + 2^3y_{R+3} + \dots + 2^{(R_{max}-R)}\, y_{R_{max}} \qquad (24)$$

Let us now collect till the next octave is fully exposed. The additional number of species collected is

$$\Delta Q = y_{R-1} \qquad (25)$$

The additional number of individuals collected is

$$\Delta I = I + y_{R-1} \qquad (26)$$

For a given number of species (N) in the fixed universe, we can, provided it is canonical, obtain the value of y at any octave, and also the value of I/m.

Assuming that m does not change as we collect, and there is no reason why in a fixed universe it should, we can compute the index z in the Arrhenius equation as

$$z = (\Delta \log N / \Delta \log I) \text{ (See Preston 1960)} \qquad (27)$$

This has been done for the case where $N = 200$ species in the universe, and where in consequence $\sigma = 4.07$ octaves if the universe is canonical. The tabulation is not reproduced here, but in Figure 14 we give a graph of the Index z as ordinate against the percentage of species collected. It will be seen that at first we collect species as fast as we collect individuals ($z = 1$), but by the time we have collected 15 or 16 species (7 or 8% of the total of 200 species available), z has fallen to 0.6, and from this point on it continues to fall steadily till it reaches $z = 0.0$ as the last species is collected.

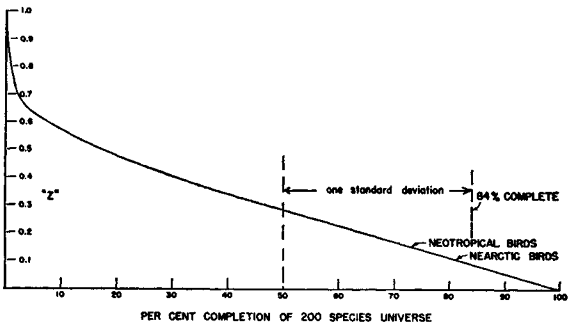

FIG. 14. The index "z" of the Arrhenius equation for a truncated distribution.

The inconstancy of z

It follows that "z" in the Arrhenius equation is not by any means a constant under the conditions we have stipulated, and since the argument applies to increasing sizes of quadrat, i.e. to a species-area curve, as well as to a sample of moths in a light trap, it follows that a log-log plot of a species-area curve will not give a straight line, at least not under the stipulated conditions.

The reason we sometimes find a fair approximation to a straight line in the log-log plot is that, after we reach a certain degree of completeness, the universe begins to expand as fast as the sample. This seems to happen typically when we have collected some 75 or 85% of the initially-available species. The remaining species are very rare, and it requires a prodigious effort and the capturing of innumerable individuals to get these rarities by strictly random collecting. We are likely to obtain a lot of "strays," "casuals," and "accidentals" before we succeed in netting all the rare species that are legitimately present.

In Figure 14, Williams index of 0.22 would be reached at 64% completion, but his data probably relate to a mixture of plant-isolates (with a z index of 0.27) and of "samples" with an index well below 0.22. The index we found (Preston 1960) for the neotropical avifauna, $z = 0.16$, is reached at 73% completion, and the figure of $z = 0.12$ (nearctic avifauna) is reached at 83% completion. This last point is at just about one standard deviation beyond the mode. Collecting by random methods beyond this point is a rather unprofitable procedure; deliberate searching now pays dividends.

The Species-Area Curve for samples

Even when the sample-area seems to be well defined and our methods systematic rather than random, as in determining the breeding birds of a county or some such area, if we succeed in locating every pair and in counting them, we shall, in general, have a sample rather than a universe. Usually it will approximate to being a sample of a much larger area, perhaps of an area as large as a state.

From the same data as before we can construct a Species-Area curve for such samples. Let us hide the canonical curve completely behind the veil line, and then withdraw it octave by octave. At each step of the withdrawal we add a number of individuals, slightly more than doubling, in fact, the number we have already counted, as given by equation (26). A number of individuals may be taken as proportional to an area. At the same time we add a number of species given by equation (25). We plot the accumulated area logarithmically as abscissa, and the accumulated species either arithmetically (Gleason) or logarithmically (Arrhenius) as ordinate. In Figure 15 we have used the latter method of plotting.

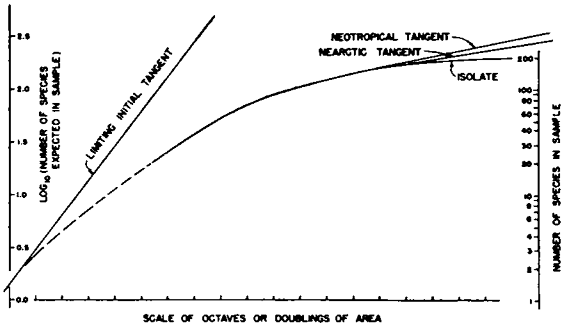

FIG. 15. Theoretical Species-Area curve for samples.

In this figure, as in the previous one, we assume that the universe initially sampled consists of 200 species. The left-hand end of the curve is shown as a broken line, till we have captured or observed 20 species; below this level we must in practice expect all sorts of erratic results, and indeed for some little distance beyond it. The initial part of the curve is fairly straight and quite steep, but not so steep as the limiting tangent, except at the very origin. This tangent is indicated; it represents the start where every new specimen is likely to be of a new species. See Preston (1960) for a discussion of such tangents.

At the right-hand end of the curve, we have 3 lines. The lowest is the theoretical behavior of an isolate or fixed universe. There are only 200 species all told, so the curve flattens to horizontal at this level. In practice, however, before we reach this point the effort to collect the few remaining rare species by random methods results, as we have seen, in causing the universe itself to expand. The middle line, therefore, is drawn as a straight line, tangent to the curve at about 80% completion (i.e. when we have collected 160 species) and having a shape corresponding to z = 0.12. This resembles the situation we encountered in our study of the Species-Area Curve of Nearctic breeding birds (Preston 1960). In that case it seemed as though initially our universe consisted of somewhere around 75 or 100 species and by the time we had collected 80% of them the universe was expanding as fast as the sample.

The upper line is the corresponding tangent for the neotropical fauna, with a slope of 0.16. As we go farther to the right (we are not now limited to the right hand end of the graph at 200 species) both of these tangents will ultimately curl upwards in practice, as discussed in Preston (1960).

The general appearance of the curves, with the double-logarithmic or Arrhenius type of plotting, is that of a pair of more-or-less straight lines meeting at an obtuse angle, but rounded off one into the other, the steeper of the lines being toward the left. The graph resembles fairly well the 2 curves found in practice in the previous paper.

The size of the universe in terms of the properties of the sample

Hidden beyond the Veil-line of our truncated distribution, which is our sample, are a number of octaves which are needed to complete our universe. So far as species are concerned, this may amount to only a modest addition, but for individuals each octave represents a doubling of the population. The complete universe is therefore usually many times as large as the sample.

In Table I the quantity x is the half-range in terms of the standard deviation as unit, i.e. $x = R_{max}/\sigma$. To a first approximation, for the cases we have encountered and are likely to encounter in practice, with N between 200 and 800, x is close to 3.0.

When we have collected as far as one standard deviation beyond the mode, we have accumulated 84% of the species in that universe, but we still lack 2 σ or thereabouts of having the complete curve. Since σ is usually around 4 or 5 octaves, we still lack some 8 or 9 octaves. The size of the universe is therefore 2^8 or 2^9 times the size of the sample, or 250 to 500 times as large.

This statement is better put in this form: we should have to collect 250 to 500 times as many specimens before we had a reasonable chance of collecting one specimen (or pair) of the rarest species. Better yet, the statement is true in its original form if $m = 1$, and it frequently seems to be near this.

This would permit us, for instance, to estimate the overwintering population of the birds of the nearctic, using the Audubon Christmas Bird Counts (Preston 1958), if we consider only those species that were found by random methods. Suppose for the sake of illustration, that in a typical single year about 10^7 individuals were seen and that these represented 84% of the complete distribution and about 500 species. Then σ should be about 4.5 octaves and x about 3.1. The end of the distribution would be about 9 octaves away, and the total number of overwintering individuals would be about 500 times the size of the sample, or 5×10^9 birds. (Again we are assuming that $m = 1$.) This is roughly equivalent to saying that, since the nearctic is roughly 4 or 5×10^9 acres, we have one bird per acre wintering in the nearctic. They will of course be far from uniformly distributed, most of them being along the seacoasts or the more southern parts of the country, where some acres will be black with redwings and starlings. Whether the estimate is fairly close or somewhat high may depend on whether different parties of observers saw the same large flocks at different places, and on various other problematical matters. The computation merely illustrates the possibility that we may be able, by more refined methods and especially, perhaps, with a surer knowledge of the value of "m," to estimate the total wintering population.

We can also compute what fraction of a population must be observed or collected in order to reach the mode, i.e. to collect 50% of the species and get at least a rough idea of the height of the mode. With $m = 1$, this fraction is simply 1/r, which may be ascertained from Table I. So, with a population of 200 species, we reach the mode when we have captured or observed 1/2720 of the individuals. With 600 species we need to observe or capture only 1/17,000 of the total.

Graphical construction

Suppose we have collected beyond the mode and the truncated distribution appears as shown in Figure 16. From the point A where the curve intersects the Veil-line, draw a horizontal line to intersect the curve again at B. From B draw the vertical line BB_1 intersecting the R axis in B_1. Let C_1 be the end of the finite distribution actually observed. B_1C_1 is the number (R) of octaves theoretically missing to the left of A, the curve being assumed quite symmetrical. Then 2^R is the ratio of size of universe to size of sample.

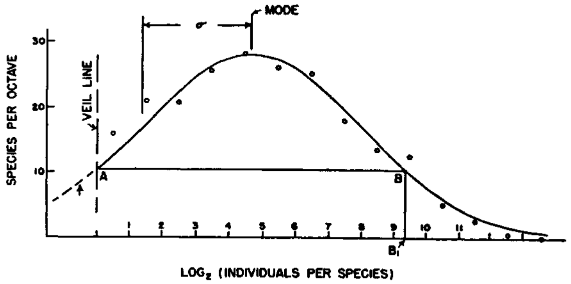

FIG. 16. Maryland State-wide Bird Count: graphical estimate of degree of completeness.

This particular plot is taken from Preston (1957) and represents the State-wide bird counts of Maryland.

Small samples

For our purposes most samples are to be regarded as small if the number of species involved is less than about a hundred. The number of individuals may be extraordinarily diverse. If our ensembles were canonical, the number of individuals for any given number of species would be ascertainable from Table I. But small biological "universes," are greatly distorted by "contagion." This is stated succinctly by Cain and Castro (1959): Most plants "are more or less clumped or contagiously distributed" and by Hopkins (1955): "Results support the view that most plant individuals are aggregated." A recent paper by Hairston (1959) may almost be regarded as a warning that contagion or clumping is so general and widespread that it is scarcely worth while constructing theories about noncontagious distributions. The warning is sound, and was given earlier by Cole (1946), who emphasizes the fact that biological material, whether animal or vegetable, is rarely distributed at random, but is nearly always more or less "contagiously distributed," and he examines the possibility of handling such distributions adequately by mathematical methods. But whereas Hairston's emphasis is on the "clumped" distribution, which he believes is characteristic more particularly of the rarer species, Cole emphasizes that "contagion" can be "positive" (clumped) or "negative" (over-regularized).

A great deal depends on the nature of the biological material with which one is working, and even upon the season of the year. Thus in the breeding season most passerine birds are, owing to their "territorial" propensities, over-regularized spatially, and therefore on a small area they are over-regularized in abundance. (See below, under Thomas, Hicks, Williams, Walkinshaw, etc.). But in the winter they "flock" and become "clumped," and it is not solely the rarer forms that do this.

Plants are perhaps the most obviously clumped material, and sometimes approximate to pure cultures, as with Hopkins' (1955) *Zostera* community. However, this can be matched among colonially nesting birds, like Beebe's (1924) "pure culture" of boobies in the Galapagos Islands, or the Emperor Penguin (*Aptenodytes forsteri*) in Antarctica. Therefore it seems to me that we are justified in examining the properties of randomly distributed populations, and treating them as a "norm," and we may treat contagion, whether positive or negative, as introducing a disturbance, modification, or perturbation into our calculations.

Notwithstanding Hairston's (1959) comment that, so long as we are dealing with a single community, "departure from randomness increases with sample size," I think we are justified in assuming that in this paper we shall rarely be dealing with a single community, but with what he calls heterogenous material, in which departure from randomness decreases with increasing sample size. Thus in order to study contagion in our present context we must examine the properties of rather small samples.

Since in the present paper we are concerned with abundance-distributions and want to know whether they conform to the lognormal type, and in particular to the canonical lognormal, a small sample for our purposes is primarily one that has rather few species, say a hundred or less. For with less than about one hundred we cannot plot the distribution graphically with much success. It is true we might use analytical methods rather than graphical ones, as being more powerful tools, but statistical fluctuations are not thereby prevented from confusing the issue, and so in Preston (1957) I suggested that an adequate sample called for something like a minimum of 200 species and something like a minimum of 40,000 individuals.

Very little in the way of plant material meets the requirement of 200 species, or even 100, and much of it, in fact most of it, appears to have less than 50, often much less. The individuals may be very numerous, especially in our northern latitudes where the floras are poor compared with the tropics (Cain and Castro, 1959) but, because the majority of plant distributions are "positively" contagious, we have an abnormally high number of individuals without accumulating many species. Thus, in order to deal at all with the published literature on plant distributions, i.e., their relative abundances in a "community" or their Species-Area curves, we have to compile a tabulation and make a graph of what to expect of samples containing less than 100 species.

Criteria for small samples

With small samples in this sense, it is easier and perhaps more accurate to reduce them to graphs, and the easiest computation to make concerns the (logarithmic) standard deviation σ. Therefore in Table IX below we use the methods of our first section to get estimates of the value of σ of I/mN for canonical ensembles ("universes") with less than 100 species.

TABLE IX. Properties of small canonical ensembles

Number of species N	Logarithmic standard deviation σ (octaves)	Ratio of individuals to species I/N or more properly I/mN
100..........	3.72	26,600
80..........	3.62	12,100
60..........	3.48	7,500
40..........	3.26	3,300
20..........	2.84	390
10..........	2.37	65
6..........	2.00	20
3..........	1.40	3.7
2..........	0.97	1.5

These values should be regarded as approximate only. They are graphed in Figure 17. On that same graph are shown a number of experimental points exhibiting contagion, positive and negative, which will be discussed later. It should be understood that these figures relate to complete, non-truncated, canonical ensembles; but where small samples seem not to be too severely truncated we can compute σ as if the distribution were normal in order to get a rough picture of the effects of contagion and the "distortion" produced thereby. We shall see later that the observed departure of σ from its expected or canonical value gives an estimate of the degree of contagion, as indeed is otherwise obvious, and that contagious distributions do not end at the crest of the "Individual Curve."

Skewness as a criterion

It is very likely that contagious distributions are somewhat skewed, the mode lying to the right of the mean, i.e. at a higher abundance of individuals per species than the mean for negative contagion. This criterion, being subject to com-

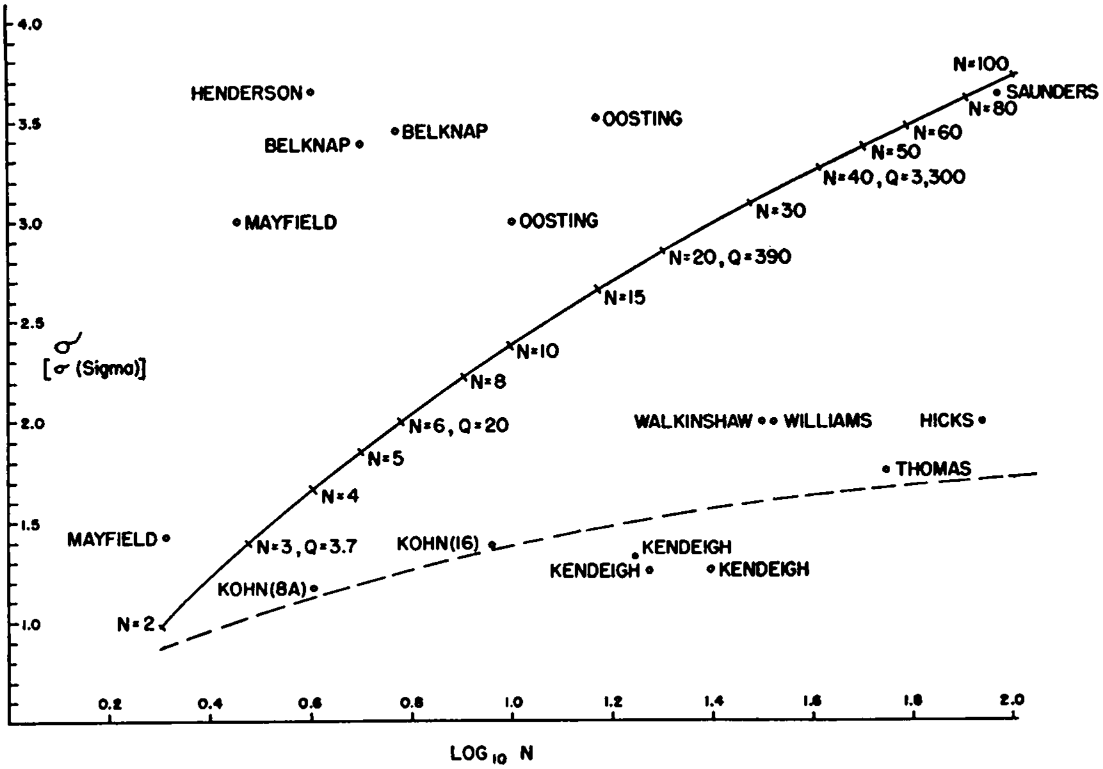

Fig. 17. The logarithmic standard deviation for small samples and universes, as a function of the number of species: N less than 100. The full line is the Canonical expectation, the broken line is the MacArthur "broken-stick" expectation. The Canonical line seems to divide positively contagious aggregates (above) from over-regularized aggregates (below).

putation, might sometimes be useful, but it is subject to greater statistical uncertainty than the standard deviation, which for the present seems sufficient for our purpose.

Examples of negative contagion

There are available a number of useful counts of breeding birds, which, being disposed to "defend a territory," tend to be distributed more uniformly over the countryside than we should otherwise expect. There is more resistance to the incoming of an additional pair of a species that is already common in the area than there is to the coming of a species not yet well represented. Thus on a given area, the species are more nearly of equal abundance than they would otherwise be. This amounts to saying that the standard deviation in practice falls below its canonical value, and the points should lie below the solid line of Fig. 17.

Thomas. The breeding birds of Neotoma, south-central Ohio

This report, given in Preston (1960), is valuable because we have the counts for 10 separate years on the same 65 acres. If we take each year and list the numbers of singletons, doubletons, and so on, discarding the names of the species, and then strike an average for the 10 years, gather the results into "octaves" and graph the outcome, we obtain Figure 18. Such a curve is somewhat typical of one-year counts and other small samples (Preston 1948), but in this case it does not make use of all the available informa-

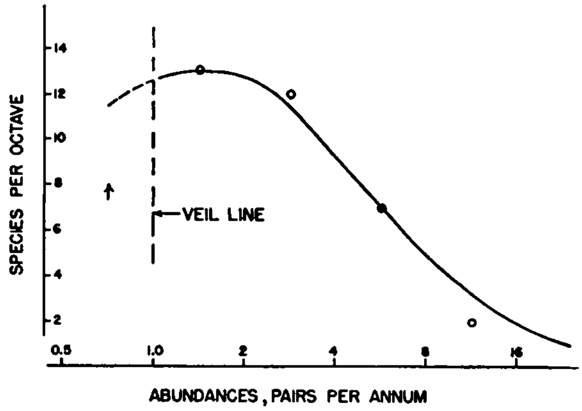

Fig. 18. A typical year's count of the breeding birds of "Neotoma," by E. S. Thomas. (Species Curve)

tion. In particular, it does not discriminate between singletons (say) of species that occur as singletons every year and of species that occur only once in 10 years.

If we therefore retain the names of the birds and make this distinction, we can distinguish between species that have a 1 :1 chance of appearing in any one year and those that have only one chance in 4 or one in 8. This produces Figure 19, where, so to speak, we get a peep behind the veil. Although the curve as drawn lies reasonably well among the observed points, which generally alternate above and below it, there could be some argument that a better curve might be one skewed to the right, descending abruptly and terminating at abundance 16. We shall not debate the point, but merely note that the number of species involved (N) is 56 excluding the cowbird, that σ is 1.73 octaves as against the canonical expectation of 3.37, and I/N is less than 3. The canonical expectation in a complete universe would give I/mN = 5000 or thereabouts. We may also note that the individuals curve Figure 20 continues past its own crest into its descending limb for an octave or so, a necessary consequence of the low σ value of the Species-Abundance curve.

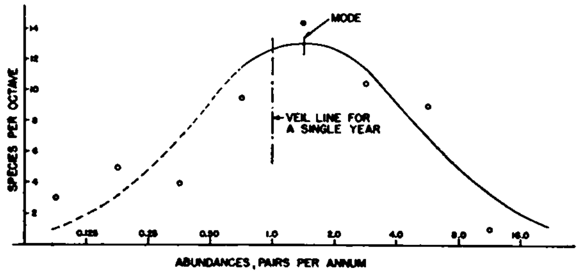

FIG. 19. Birds of "Neotoma," a typical year reconstructed from a knowledge of what species were present in each of 10 years. (Species Curve)

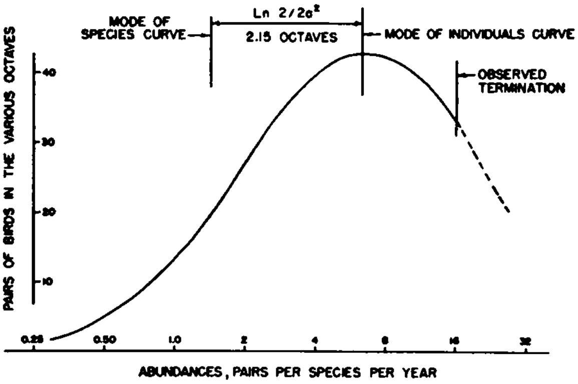

FIG 20. The "individuals" curve for an average year at "Neotoma." The observed termination is well beyond the crest, as it should be for a negatively-contagious or over-regularized distribution.

Hicks (1935). Breeding birds near Westerville, Ohio

This also is a 10 year count, in the course of which 86 species were observed, though the average for a single year was 63 or less. In Figure 21 we give the graph, on the same basis as Fig. 19, except that I have used the accumulated totals for the 10 years and have not struck an average by dividing this result by 10. The value of σ is almost exactly 2.0 octaves, the average value of I/N is about 3.5 for a single year, and there is a suggestion of skewness, as expected.

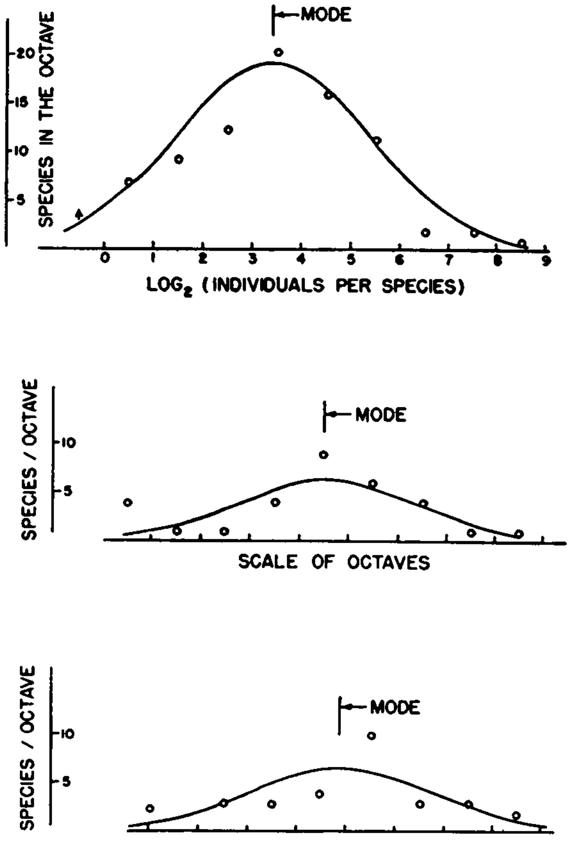

FIG. 21. Top, Hicks. Breeding Birds near Westerville, Central Ohio. Ten year count. Center, Walkinshaw. Breeding Birds near Battle Creek, Michigan. Ten year count. Bottom, Williams. Breeding Birds near Cleveland, Ohio. Fifteen year count.

Walkinshaw (1947). Breeding birds on 83 acres of brushy fields

This is another 10 year count. Some 31 species were accumulated. The distribution is graphed in Fig. 21. The standard deviation is almost exactly 2.0.

Williams (1947). Breeding birds of a beech-maple forest near Cleveland, Ohio

This is a 15 year count and 33 species, including the Cowbird, were accumulated on 65 acres. The value of σ is approximately 2.0, and I/N is about 4.2 pairs per species per annum. (Bottom Fig. 21.)

The counts of the individuals curves for Hicks, Walkinshaw and Williams all seem to lie about 1½ to 2 octaves to the left of the termination of the curve, which again is what is expected.

Kendeigh (1946). Breeding bird counts in New York State

These counts extended, in form usable by us, over 3 years in the early 1940's. The results from Kendeigh's tables I, II, and III, may be condensed for our purposes into the form of Table X below, with Thomas' results for "Neotoma" added for comparison.

TABLE X. Kendeigh's breeding bird counts in New York state

Kendeigh's Table Number..	I	II	III	"Neotoma"
Character of Woodland.....	Beech Maple Hemlock	Hemlock Beech	Beech Maple Hemlock	Mixed Deciduous
Acreage involved..........	8	21	62	69
Number of species known breeding...............	24	17	18	56 in 10 years
Av. pairs per Species per Annum (I/N).........	1.22	2.48	4.0	—
Standard Deviation (σ) observed σ............	1.26	1.32	1.22	1.76
Standard Deviation for a canonical ensemble with same number of species σ^1.....................	3.0	2.75	2.80	3.45

So far we have dealt solely with birds, and have tentatively ascribed the departure from the canonical results to "negative contagion," or, in biological language, to the territorial demands of the birds. Let us now turn to an entirely different fauna, marine gastropods.

Kohn (1959). Gastropods (Conus) in Hawaii

Kohn examines the abundance-distribution of cone-shell species from several collecting localities, and plots them in terms of the MacArthur hypothesis discussed below. I have chosen to examine here two instances, Kohn's Figure 8A, because it is the first he gives, and his Fig. 16, which is the one that agrees most closely with the MacArthur distribution.

For his Fig. 8A there are 4 species, and 136 individuals, so that I/N (Kohn's m/n) is 34 individuals per species. This is more than we expect in a canonical distribution. Yet the standard deviation σ appears to be only 1.11 octaves, far below the 1.67 octaves expected for a canonical distribution. For his Fig. 16 there are 9 species and 182 individuals, giving I/N = 20.2. This is much less than the canonical expectation, unless our "m" (not his) is not unity but 2 or 3, as it may perhaps be. The standard deviation is about 1.38 octaves, far below the canonical 2.3 octaves.

These results suggest, but do not prove, that species of *Conus* may be over-regularized in distribution, so that the various species in a quadrat are of more uniform commonness than a random distribution would give. In correspondence Dr. Kohn says that *Conus* is not known to "defend a territory," but not much is known about it in this respect. We do know however that the genus is carnivorous, but the different species have different food-preferences, some eating marine worms, some fishes, and some other animals. Thus there is a possibility that food supply of itself might induce a non-random distribution, either directly, or indirectly by causing quasi-territorial behavior. All these examples of presumed negative contagion (over-regularized distribution) are plotted on Fig. 17, and it will be noted that they all fall below the line, often a very long way below it.

Examples of positive contagion

Oosting (1942). Plants of the Carolina piedmont

From Oosting's Table #1, dealing with fields abandoned for one year, I took Field #4, which has 15 species reported. The abundances are given in terms of "densities," and the standard deviation works out at about $\sigma = 3.52$ octaves, as against the 2.65 expected for a canonical ensemble.

Oosting (1942). Shrubs and vines in a 15 year old stand of the Carolina piedmont

There are 10 species, and the standard deviation is $\sigma = 3.0$ octaves, as against the canonical 2.37.

I think these examples from Oosting are sufficient. He provides many other tables, and it seems, from a rather casual inspection, that most of the other tabulations might corroborate these 2.

The Herons of West Sister Island, in Lake Erie off Toledo

This information comes from Mr. Harold Mayfield. American Egrets first nested, so far as is known, in 1946. In that year a count showed that the heronry consisted of about 1500 Black Crowned Night Herons, 200 Great Blue Herons, and 10 American Egrets. This gives a σ-value of 2.95 octaves, far above the canonical value (1.40) for 3 species. The next year the number of Egrets doubled, and if everything else remained constant, the σ-value would increase a little. In 1945 there were no Egrets, presumably, and so only two species nested. The σ-value may presumably have been 1.44, which is still well above the canonical expectation of 0.97 for 2 species.

Belknap (1951). Breeding birds on an Island in Lake Ontario.

Belknap censused a small island of one acre from 1948 to 1951. He gives his count for 1951 and comments briefly on earlier years. This island, like West Sister Island above, presumably behaves as an isolate, and this feature may be valuable. There were, in 1951, six species and a comparatively large number of individuals or pairs (965 occupied nests), giving a value of $I/N = 160.8$, far above the canonical expectation of 20 for $N = 6$. In Table XI we give the tabulation, with the figures predicted by the MacArthur hypothesis for comparison; the latter will be discussed later in this section.

TABLE XI. Belknap's breeding birds on an island in Lake Ontario in 1951

Species	Observed Nests	MacArthur Hypothesis Prediction
Ring-billed Gull	793	394
Common Tern	117	233
Herring Gull	34	153
Double-crested Cormorant	19	99
Black Crowned Night Heron	1	59
Black Duck	1	27
	965	965

The standard deviation is $\sigma = 3.47$ octaves, against the 2.0 of canonical expectation and a still lower figure for the McArthur distribution.

Belknap. The same island in 1950

Belknap does not give this count, but states that there were no Night Herons in 1950, that there were less than half as many terns, that there were twice as many Ring-billed Gulls, and that the Herring Gulls were "relatively stable." This leaves us with 5 species and a rough estimate of the 1950 population, given in Table XII.

TABLE XII. Belknap's Island in 1950

Species	Estimated Nests	MacArthur Prediction
Ring-billed Gull	1600	780
Common Tern	55	439
Herring Gull	34	267
Double-crested Cormorant	19	154
Black Duck	1	69
	1709	1709

The logarithmic Standard Deviation for the first column is 3.4 octaves, compared with 1.85 for the canonical ensemble of 5 species. Belknap's description of the 1948 and 1949 situations would lead to a similar conclusion.

Henderson. A communal roost of passerines near Oberlin, Ohio

Henderson describes this as a roost of "blackbirds," though it includes 500 Robins (*T. migratorius*), and the Redwinged Blackbirds were not positively identified. The pattern appears to be approximately as shown in Table XIII.

TABLE XIII. Henderson's blackbird roost

Redwings	5 (not positively identified)
Cowbirds	50
Robins	500
Grackles	5,000 to 10,000
Starlings	50,000

If we omit the Redwings which "were suspected but never surely identified," the logarithmic standard deviation comes out at about 1.1 orders of magnitude or 3.65 octaves. If we include the Redwings, the standard deviation will be greater, but it is almost off the map (Fig. 17) anyway. In this connection we may note that, while by ordinary standards the roost is a communal one, there is often in such cases a partial segregation of the species, just as there may be a negro quarter in a town.

These half dozen values of "clumped" or "colonial" distributions are plotted as points on Fig. 17. They all lie far above the line, and this is presumably true of "positively contagious" distributions generally.

Seemingly contagion-free examples

It seems only fair to note here that in looking for examples of contagious distributions, I found

3 that did not act quite as expected, but instead fell close to the line. One was Saunders' tally of the birds of Quaker Run Valley (Fig. 17). The explanation may well be the one advanced by MacArthur, that there are really several communities involved and mixing them tends to produce such a result. Hairston has come close to saying the same thing.

Another example was a count of plants by Buell & Cantlon (1951). The same reason is possibly valid, though the catagion to be removed is of the opposite sign.

Finally I took the counts made by Mr. H. H. Mills and myself of the day-flying herons of the heronry of Stone Harbor, New Jersey, in 1959 (unpublished). Six species were involved and the count was of the homing herons and egrets at sundown, not of their nests. The difficulty here is the impracticability of distinguishing in the gloaming between Snowy Egrets and immature Little Blue Herons, for it was late in the season and the immatures were largely on the wing. One estimate of the partition gave a point nearly on the line. Another estimate would place it distinctly above the line.

In any case the general picture seems clear. Gregarious ensembles tend to fall above the line, territory-guarding ones tend to fall below it. Thus the canonical distribution seems to justify itself as a "norm" corresponding to randomly-distributed species devoid of both positive and negative contagion, which can be regarded as perturbations of the canonical.

Possible restatement of the canonical hypothesis

We are now in a position to make a surmise that the canonical hypothesis is a statement, perhaps only an approximate one, of the behavior of lognormal ensembles when the individuals in the ensemble act completely independently and neither attract nor repel others of their own species. If I can interpret Hairston's views in the light of my own, it is the situation that obtains when the "community" is completely devoid of "organization." When "organization" appears, it results in "contagion" and not solely the "clumping" or positive contagion that is Hairston's primary concern, but also the "negative contagion" of Cole and the "regularity" of Hopkins (1955 and 1957).

This suggests that our graphical description of a canonical ensemble as one whose Individuals Curve crests at its termination may very well be replaced by an analytical description of a lognormal ensemble whose individuals (or pairs) completely "ignore" others of their own species, that is, are unaffected by their proximity or distance. Such a reformulation might be very useful, as well as more satisfying esthetically than the one we have used, but I do not know whether it would prove so simple to handle mathematically.

The MacArthur distribution

A recent suggestion by MacArthur (1957), originally based on the analogy of the probable lengths of the fragments of a randomly broken stick, proposes that a population of m individuals may be apportioned among n species according to the law

$$m_r = (m/n) \sum_{i=1}^{r} [1/(n - i + 1)] \qquad (28)$$

where m_r is the number of individuals assigned to the r^{th} rarest species and m/n is our I/N.

The actual operation of computing the abundance of the various species may be carried out as follows: suppose, for example, that n = 57 species. The rarest species will have 1/57 (m/n): the next rarest (1/57 + 1/56)(m/n): the next (1/57 +1/56 + 1/55)(m/n) and so on. Now for comparison with the lognormal distribution, having obtained the MacArthur figures for each species we take the logarithms thereof and gather them into octaves for plotting, or compute the logarithmic standard deviation by orthodox methods.

In Fig. 22 we graph the distribution for 4 instances, where N = 5, 10, 100, and 1000 species respectively. The scale of octaves at the bottom applies to all 4 graphs as abscissa, but the scale of ordinates varies, and is given in each case. With this method of plotting, the curve is single-humped, tangent to the x-axis on the left but cutting it abruptly on the right, skewed somewhat to the right and therefore only a rough approximation to a Gaussian or "normal" curve. We may, however, calculate the standard deviation "as if" the curves were normal, and we find that for 10 species $\sigma = 1.49$ octaves, and for 100 species $\sigma = 1.72$ approximately. We may also easily compute σ for 2 species as 0.79 octave and for 3 species as 1.00 octave. This permits us to draw, as a broken line in Fig. 17, the approximate position of the standard deviation as a function of the number of species in the "community," as predicted by the MacArthur hypothesis. It lies far below the value predicted by the canonical lognormal hypothesis. It passes very close to Kohn's 2 points, and lies among the points for antigregarious or territorially-minded breeding birds.

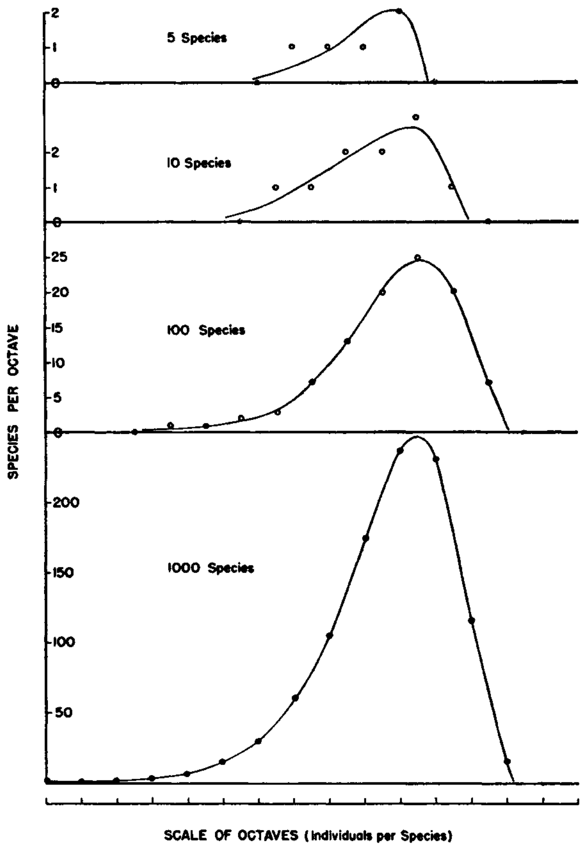

FIG. 22. The MacArthur distribution for 5, 10, 100, and 1000 species.

It lies at an immense distance below the points for gregarious or colonially-nesting birds and for the "contagious" plant communities. Since the effects of contagion may be expected to smooth out, and die out, with very large areas or very large ensembles of species, in the overall picture, while the MacArthur distribution, even for 1000 species (see Fig. 22) gives a σ-value not much higher than for 100 species, the distribution would not apply to such aggregrations of species, as MacArthur himself has said.

I therefore suggest somewhat tentatively that the MacArthur distribution is not the norm, and that the frequent agreement of Kohn's diagrams therewith is due to the cone-shells being somewhat over-regularized in their spacial distribution. We may note here Kohn's own comment that in a number of his diagrams "common species are too common and rare ones too rare" to agree with the MacArthur prediction. This statement amounts to saying that the logarithmic standard deviation of abundances is too high, which means that his points (not those I have plotted in Fig. 17) are often above the MacArthur line, perhaps approaching the canonical line.

It was, I think, MacArthur's original view that a genuine "community" might correspond approximately with his predictions, while fortuitous aggregations would more likely come close to the lognormal. This presents us with the problem of defining a genuine community. Superficially one might imagine that a heronry or a colony of gulls and terns and cormorants like Belknap's is a "community" since in some sense the birds "attract" one another in much the same way as human communities are brought together. But we see from Fig. 17 and from the data in Tables XI and XII, that it is precisely such communities that depart most from the MacArthur prediction. We should have to define a community as a group of persons, animals, or plants that insist on holding their fellows at arm's length, or that are repelled by one another.

Margaret Perner's breeding birds at Cleveland, Ohio

Since the MacArthur formula seems more appropriate to territorial species than others, I thought we might examine one instance in more detail, and for that purpose use Perner's (1955) data on 25 species of nesting birds. This time we plot simply the nests per species as ordinate against ordinal rank (of increasing commonness) as abscissa (Figure 23), and the MacArthur prediction is plotted on the same graph. The departures of the observations from the prediction are not random, but systematic, so that this aggregation of species, though negatively contagious, does not agree well with prediction.

Other pecularities of the Species-Area curve for samples

Samples, especially small ones, have a number of peculiarities, partly because they are samples and therefore likely to be incomplete or truncated distributions, and partly because they are small, especially in numbers of species, and therefore likely to be affected by contagion, positive or negative. Whereas the Species-Area curves for isolates can be understood comparatively easily, those for samples may present greater difficulties. Even in the matter of graphing the plots there are problems, and in interpreting the results there are worse uncertainties.

Methods and problems of plotting the Species-Area curves

This subject is almost a monopoly of plant ecologists, who set out "quadrats" of various sizes and count the species in the quadrats. They

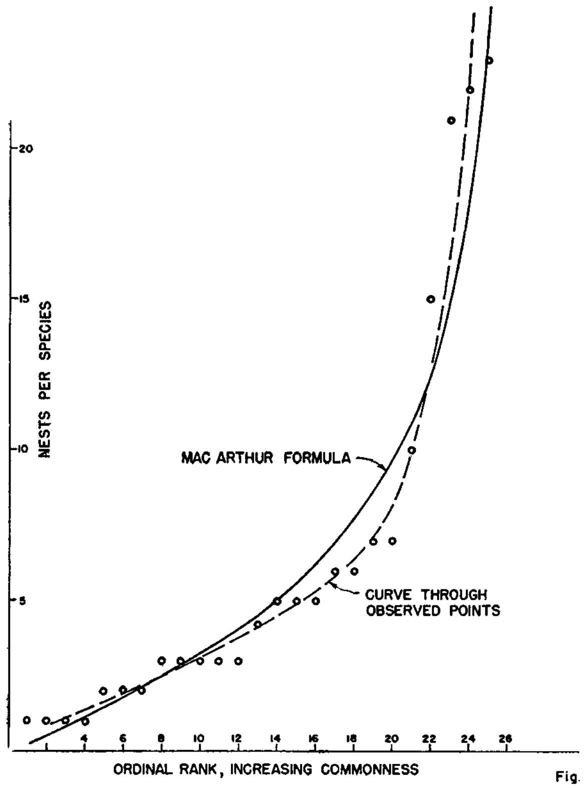

FIG. 23. Perner. Breeding Birds of a Cleveland, Ohio, park.

usually use quadrats that increase in size in a more or less geometric progression, for instance in the area ratios of 1, 2, 4, 8, 16, ... This is an advisable procedure, because the species count increases only slowly with increasing area. The geometric progression implies that psychologically or subconsciously, the operator is working with the logarithm of the area and not with the area itself, yet many investigators have plotted the results on an "arithmetical" basis with number of species as ordinate against area as abscissa (e.g. Cain 1938, Hopkins 1957). This results in most of the points being crowded into a narrow, nearly vertical, line near the origin and being sparsely distributed farther to the right.

The kink in Cain's curve

Numerous plottings by the above writers and others are available to any inspector, so I have chosen to use one that Cain (1959, p. 110) did not plot, though he gave the data from which it may be plotted (Figure 24). Braun-Blanquet and others (see Hopkins 1957, pp. 441-443) as well as Cain at one time, believed there was a definite "break" in this curve, the initial part being essentially vertical and the later part being essentially horizontal, as if the curve "saturated" at a certain level that gave the total species in the "community" or "stand."

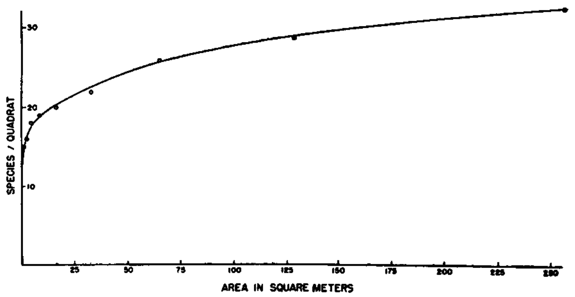

FIG. 24. Cain & Castro. Species-Area curve for savanna vegetation in Para, Brazil.

An examination of this and the many published curves suggests strongly that the curve approximates rather closely to a "generalized parabola," whose equation is

$$y^n = ax \qquad (29)$$

where n is an exponent greater than unity, and not necessarily an integer.

The ordinary "conic-section" or "quadratic" parabola has $n = 2$ in the above equation. The generalized parabola has other values, the most interesting ones being higher than 2, often much higher. The quartic parabola for instance has $n = 4$, and the curve we have illustrated strongly suggests a quartic. Since y, or N, the number of species, is necessarily positive, and so is x, or A, the area of the various quadrats, the index n may take all positive values and the curve will remain real. Whatever the value of n, when limited to positive values the curve will pass through the origin and, provided n is greater than unity, it will there be tangential to the y-axis. The higher the value of n the closer it will hug the axis and the sharper will be the bend when it breaks away from it. Thus in a sense the break is real. The curve never becomes parallel to the x-axis, but continues indefinitely to climb, though always at a decreasing rate with increasing size of area.

In order to help visualize the "higher" parabolas, I have graphed several of them in Figure 25 and by setting $a = 1$ in equation (29), I have caused all curves to pass through the points 0,0 and 1,1. This leaves n as the only variable parameter, and it will be seen that as n increases the curve is steeper near the origin and flatter at the larger coordinates.

It is a property of the ordinary quadratic parabola that its curvature is greatest at the vertex, or origin in this case, but this is not true of the higher parabolas. Nonetheless all of them have a single

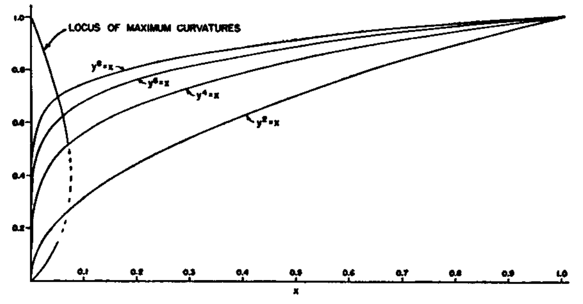

FIG. 25. A family of generalized parabolas, and the locus of points of maximum curvature.

point of sharpest curvature. In this sense there is a theoretical "break" in the curve, though it is not a discontinuity of any sort.

The curvature of any curve at any point is given by

$$\frac{1}{R} = \frac{d^2y}{dx^2}\Big/\left\{1 + \left(\frac{dy}{dx}\right)^2\right\}^{3/2} \qquad (30)$$

and if we differentiate this once more and set

$$\frac{d(1/R)}{dx}_2 = 0 \text{ or } \frac{dR}{dx}_2 = 0 \qquad (31)$$

and solve for x, we can define the point of maximum curvature in terms of the abscissa, or by solving for y we can define it in terms of the ordinate, which is usually (in our problem) more satisfactory.

In this case, we find that the point is given by

$$y^{2n-2} = \frac{a^2}{n^2}\left(\frac{n-2}{2n-1}\right) \qquad (32)$$

On Fig. 25 I have plotted the locus of the maximum curvature. It is probably the break-point that Cain was seeking. In practice, however, it is not easy to determine such a point by graphical methods. It is much easier, even with a curve perfectly free from experimental or statistical errors, to find the point by equation (32), and for that purpose we have first to find the value of n. Clearly, if the curve really is a parabola, this is most easily done by taking logarithms of both ordinate and abscissa, whereon the equation becomes

$$n \log y = \log a + \log x \qquad (33)$$

which is a linear relation between log (Species) and log (Area): i.e. the curve ought to be a straight line. This log-log plotting I have elsewhere (Preston 1960) called an Arrhenius plotting. Using it on the data of G.A. Black and S. A. Cain as reported in Cain and Castro (1959), for savanna vegetation in Pará, Brazil, we get Figure 26, which, except for the first 2 points, is a pretty fair straight line. The slope is given by $k = 1/n = 0.14$, or $n = 7$ approximately, and the curve, over 2½ orders of magnitude, is a pretty good generalized parabola.

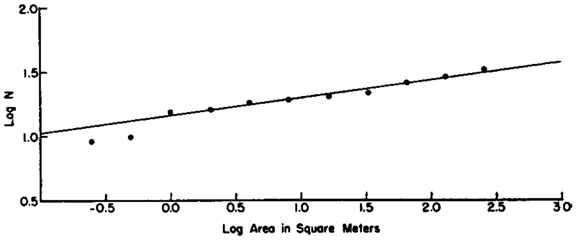

FIG. 26. The Cain & Castro Species-Area curve on a log-log (Arrhenius) basis.

The Gleason plotting

Plant ecologists seem to have made less use of the Arrhenius "log-log" plotting than of the Gleason "semi-log" plotting. It is necessary, in order to space the points fairly uniformly across the graphs, to take the logarithm of the area, but it is not essential to take logarithms of the number of species. Furthermore, because species increase so slowly with increasing area, the logarithm of the number of individuals is a linear function of the number of species over rather wide intervals. Thus if the Arrhenius plot gives a straight line, so will the Gleason plot, over an interval of an order of magnitude or more in the size of quadrat or time of observation.

In Figure 27 we illustrate this point by means of Thomas' data on the 10-years of breeding bird counts on the 65 acres of Neotoma in south-central Ohio. The same data are plotted by the Arrhenius and by the Gleason methods. In each case the abscissa is a logarithmic scale of years-of-observation. The ordinate for the lower curve is

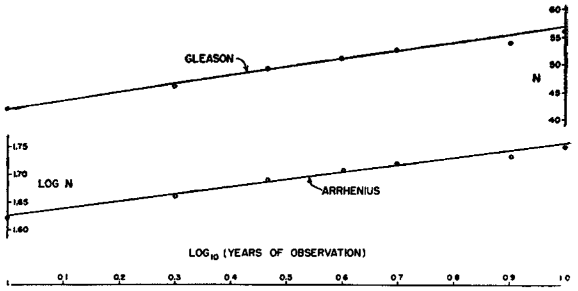

FIG. 27. Gleason and Arrhenius curves compared over a range of one order of magnitude (3.3 doublings or octaves).

[2] I am indebted to Dr. R. E. Mould of Preston Laboratories, Inc., now American Glass Research, Inc., for this result.

logarithmic and is indicated at the left; for the upper curve it is arithmetical and is indicated at the right. Both curves are satisfactorily straight. On a log-log basis the index k = 1/n = 0.132.

If the range of abscissae is large enough, both curves cannot continue to be straight, unless k is so small that both curves are substantially horizontal. For if the log-log curve is straight, the Gleason curve will be concave upwards, and if the Gleason curve is straight, the Arrhenius curve will be concave downwards.

Brian Hopkins' curves (1955)

In order to examine whether the Gleason or the Arrhenius plot is more satisfactory in practice, on a purely empirical basis, it is necessary to have data over a range of areas of far more than one or 2 orders of magnitude. Hopkins (1955, 1957) gives a tabulation of a dozen communities or stands of plants over a nominal range of 8 orders of magnitude or a little more. In his 1955 paper he uses the Gleason semi-log plotting, and the curves tend to be concave upwards and of steadily increasing slope. This suggests that our Arrhenius log-log plot might be better, or at least more instructive, but in neither paper does Hopkins mention it.

With any form of plotting or computing, we run into trouble with the smallest areas, of 0.01 cm^2 and 0.07 cm^2 respectively. Here only 1 or 2 species, or even less than one species, are involved. These points I have had to ignore. The situation is better, but still statistically not too happy, with the other areas below one m^2, for we usually have less than 10 species, sometimes only 4 or 5. These, however, may conceivably make a "community" or at any rate a "stand."

The use of the log-log plot, though logical, is itself a source of some suspicion, to the extent that a liberal use of logarithms tends to reduce any monotonic function to a straight line, and this is especially the case when the dependent variable (the number of species) increases slowly with the independent variable (area) as it does in vegetation stands. The curve is going to approximate not only a straight line, but a horizontal line. Fortunately, Hopkins covered a very wide range of areas, and in most cases took 50 samples at each size of area in order to strike an average for his "point." Thus comparatively small departures from the graduating line will usually be meaningful, especially if the departures are systematic, to one side of the line for several points in succession.

We may summarise the outcome thus:

Hopkins' stand #2. Grassland at Wrynose Pass, Lake District, England: 55 species (bottom, Fig-

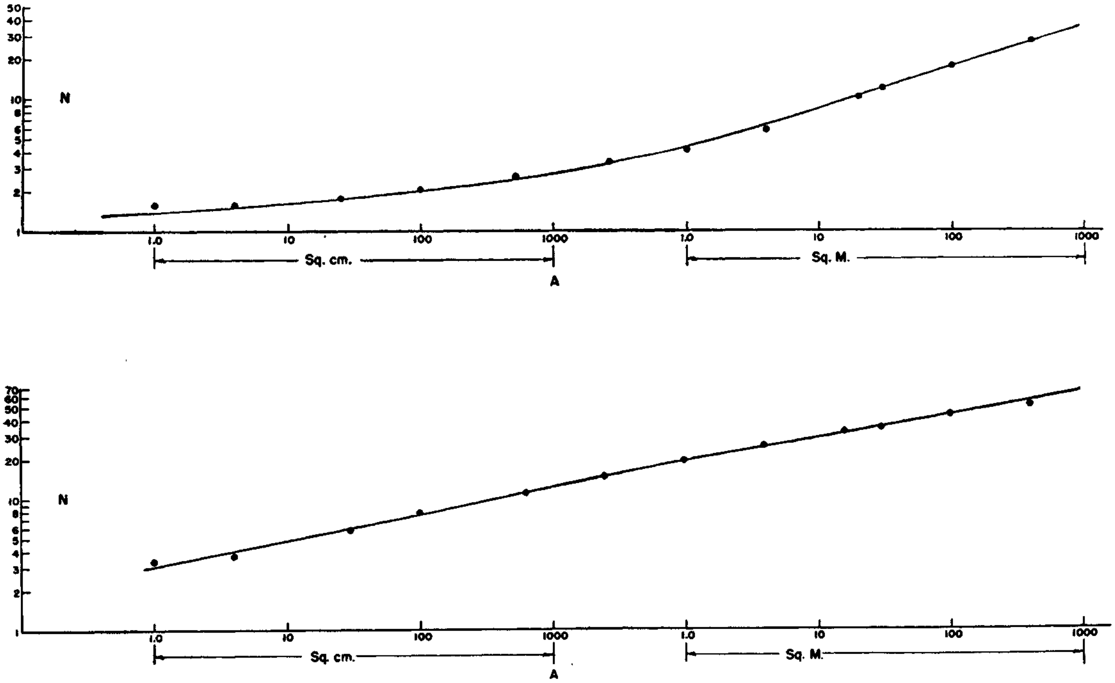

FIG. 28. Lower, Hopkins, Stand #2. Grassland in the Lake District, England. Species-Area Curve. In the following Figures, the ordinate (N) is the number of species on the various observed areas, plotted logarithmically. These are all log-log plottings (i.e. "Arrhenius plottings"). Upper, Hopkins, Stand #3, Beech Wood, Chiltern Hills.

ure 28). This is a good straight line over its whole length of 6½ orders of magnitude. The slope, k, is 0.193.

Hopkins' stand #3. Beechwood in the Chiltern Hills: 28 species. The curve seems to be concave upwards, steepening towards the right, but the last 5 or 6 points lie quite well on a straight line with a slope, k = 0.327. This is an unexpectedly high value, more appropriate to a set of isolates than to a sample.

Hopkins' stand #4. Blanket Bog in Mayo: 47 species. From 1 cm^2 to 4 m^2 the curve is close to a straight line, with a slope of k = 0.245. Then it changes abruptly to another much flatter slope, with k = 0.097 (Figure 29).

Hopkins' stand #5. Bog at Rannoch, Perthshire: 48 species. This somewhat resembles #4, the initial slope is k = 0.203, changing abruptly at 1/4 m^2 to a flatter slope, k = 0.135. In both #4 and #5 there is evidence of systematic departure on this flatter slope, as if it is slightly concave upwards (Figure 30).

Hopkins' stand #6. Pine woods at Rannoch, Perthshire. This set of points is successfully graduated with 2 straight lines, but here the upper slope is the steeper: k = 0.159 changing to k = 0.229. This shows that the curve can be "concave" upwards or downwards.

Hopkins' stand # 11. Blanket bog in the Pennine uplands, England: 46 species. Like the other bogs, this is initially a good straight line (from 1 cm^2 to 25 m^2) changing abruptly then to a flatter slope, defined only by 3 points. The initial slope, valid over 5½ orders of magnitude gives k = 0.218.

Thus our k values come out as follows:

	N_{max}	k
Grassland (#2)	55	0.193
Woodland (#3, Beech)	28	0.327, less at low areas
Woodland (#6, Pine)	42	0.229, less at low areas, where k = 0.159
Bog (#4)	47	0.245, less at greater areas, where k = 0.097
Bog (#5)	48	0.203, less at greater areas, where k = 0.135
Bog (#11)	46	0.218, less at greater areas

The average slope of the more trustworthy-looking sections of the curves is k = 0.24, a little above Williams (1943b) and not far below the theoretical value for a series of isolates, but well below it if we take account of the flatter slopes frequently present.

If we can interpret these curves at all, it seems as though they tend on the average to come close to being straight lines. Some are quite straight, the 2 woodland areas are concave upwards and the 3 bogs are concave downwards. The average, in fact, of all 12 of Hopkins' stands is very close to

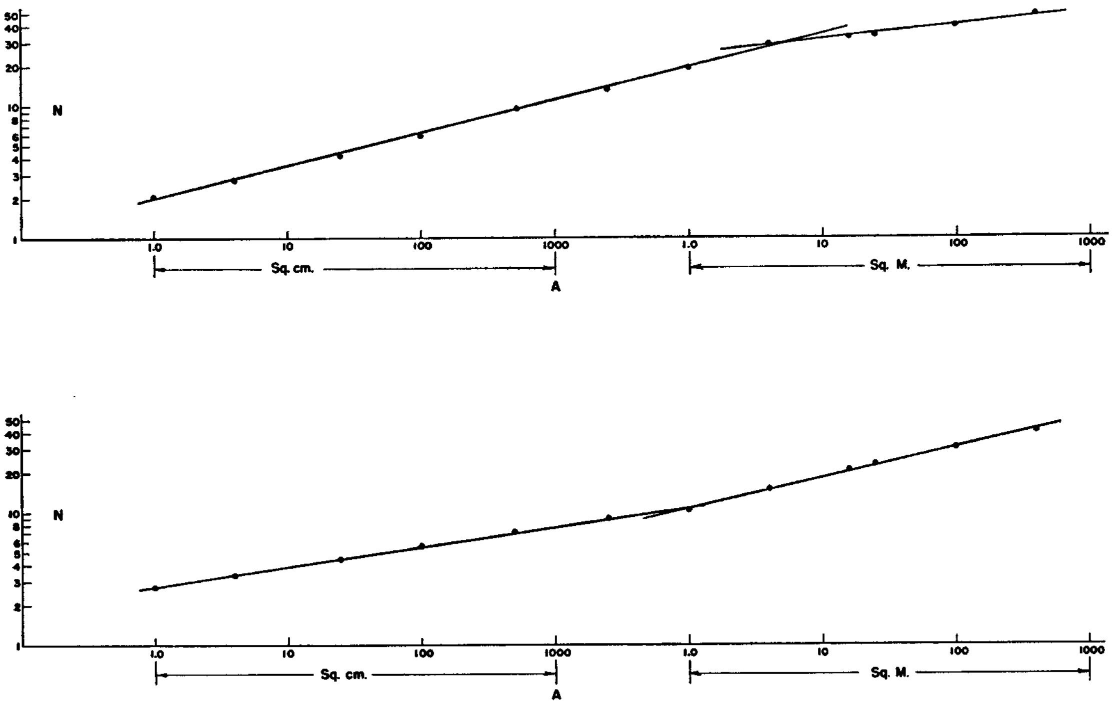

Fig. 29. Lower, Hopkins, Stand #6, Pine Woods. Perthshire. Species-Area Curve. Upper, Hopkins, Stand #4, Blanket Bog, County Mayo.

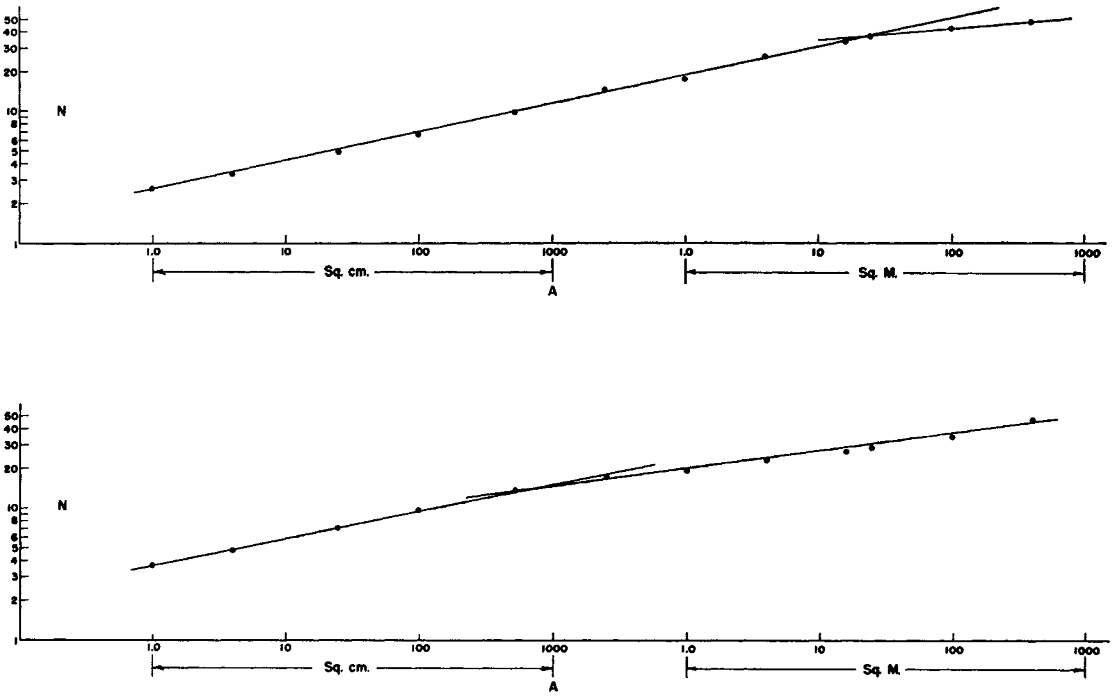

FIG. 30. Lower, Hopkins, Stand #5, Bog in Perthshire. Species-Area Curve. Upper, Hopkins, Stand #11, Blanket Bog, Pennine Uplands.

a straight line over nearly 6 orders of magnitude, with an exponent of about $k = 0.16$.

It would seem possible that bogs in Britain may have a limited flora, and that by the time we reach a few square meters in area we are beginning to exhaust the flora. On the other hand, woodlands there seem to get off to a slow start, and begin to show what they can do about the time we decide, at a hundred square meters or so, to call off our survey. The grassland seems to strike a very happy balance between the two.

It should be noticed that all samples are "small" in number of species and, therefore, positive or negative contagion can markedly affect the slope which can assume a new value, higher or lower (depending on the sign of the contagion), when the contagion begins to smooth out. However that is not the only thing that can bring about a change in slope since, as we have seen earlier, a constant slope for a long distance should imply, in a lognormal ensemble, that the universe is expanding as fast as the sample. To expect it to keep exact pace with the sample for any great distance seems unreasonable. Thus the curve may steepen and flatten from time to time, and one of Hopkins' stands, his #8, actually appears to do this. The rest seem to exhibit a single change, usually somewhat abrupt, but in one case gradual.

The interpretation I tentatively put upon these results, and upon the fact that most of the examples have an index or exponent of k ($= 1/n$) between about 0.15 and 0.24, is that from each area we have a sample of 60% to 80% of this species in the "universe" we are instantaneously sampling.

Preston. Birds of the nearctic and neotropical regions

Lest the botanists think they have a monopoly on problems dealing with species-area curves, we may note that the birds of 2 major zoogeographic regions produce similar results. Over a very wide range of areas, we found (Preston 1960) that the index k for the nearctic was around 0.12 and for the neotropical was about 0.16, and again we interpreted this to mean that after we had "collected" some 70% or 80% of the universe we were sampling, the universe started to expand pari-passu with our further observations.

Vestal's sigmoid

In 1949 Vestal in the U.S.A. and Archibald in Britain (quoted by Hopkins 1955) reached the conclusion that in the case of vegetation stands there was a tendency for the species-log area (Gleason) curve to be sigmoid; it began at a low slope, steepened considerably, and then became less steep. Hopkins (1955) is very dubious about the reality of this. I see no reason why the effect should not sometimes exist; indeed if the dominant factor in controlling the slope is the phenomenon of the "expanding universe," then failure of the universe and sample to keep exact step must produce steepenings and flattenings of the curve which can sometimes take the rather simple form of sigmoids, or S-shaped curves.

The evidence as presented by Vestal and by Archibald (or Hopkins 1955, Fig. 13) is on a Gleason-plot basis, but when the range of area is modest, the same phenomenon would appear in a log-log (Arrhenius) plot. Archibald's curve however covers a wide range. The possibility that such curves may exist can hardly be disputed on theoretical grounds; how often they occur in practice is a matter for observation.

We may note, however, that when the curve gets off to a slow start, an almost horizontal line at or near one species per quadrat, suggests that the "universe" is nearly a pure stand. We must expect that sooner or later, as we expand our quadrats, there will be a marked steepening of slope showing that the stand is not a pure one. This happened with Hopkins' example #3, our Figure 28, a woodland, and when the curve steepened it acquired so high a slope that it is doubtful if it could keep it up for long. If the quadrats could have been extended a few more orders of magnitude (which may not have been physically possible in recent centuries) it is almost certain that the slope would have flattened again to some important extent, and then we should have had a Vestal sigmoid.

All the species-area curves here replotted from Hopkins' data are plotted to the same scale, and if we superpose them we get Figure 31. It is difficult to resist the speculation that if we had data beyond the experimental limit of 400 m^2, say up to 10^6 m^2, we should find all the curves following closely a single line with a slope or index of 0.169, and indicating that at 1 km^2 the flora of the British Isles tends to amount to 250 species or thereabouts.

If the same index continued valid up to 2000 km^2 (= 800 square miles), the area of such a county as Leicestershire, the number of species should be about 880. An earlier count gave 890, but Horwood and Gainsborough (1933) report a total of about 1400, including, however, the "aliens," which are numerous.

The curves are actually closer together at the upper limit of observation (400 square meters) than they are at 1 cm^2, and much closer than at 1 m^2. It is not till we get above 100 m^2 that all the curves represent 20 species or more, and this suggests that the divergences below this area are largely the properties of samples that are statistically too small. This seems to be an almost universal property, or shortcoming, of botanical quadrats.

Summary

We see then that "samples" have some properties in common with isolates, and in some respects they differ radically from them. One of the most important of these differences is the slope of the Arrhenius plot. For isolates it is theoretically about 0.26 to 0.28 depending on the number of species involved, and we found that in practice it is often somewhere near this figure. For samples it is much less, sometimes no more than half this figure. This leads to an important zoogeographical conclusion to be discussed in the next section.

—To Be Continued—

Literature Cited and Acknowledgments will be found at the end of Part II of this paper.

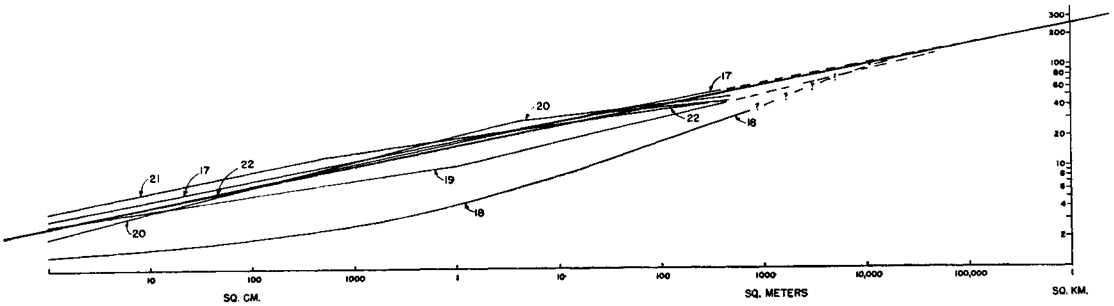

FIG. 31. Hopkins' 6 curves superposed. The heavier line that projects beyond both ends may be a sort of average to which all 6 ultimately trend.