



Minimal Community Structure: An Explanation of Species Abundance Patterns

George Sugihara

The American Naturalist, Vol. 116, No. 6. (Dec., 1980), pp. 770-787.

Stable URL:

<http://links.jstor.org/sici?sici=0003-0147%28198012%29116%3A6%3C770%3AMCSAEO%3E2.0.CO%3B2-9>

The American Naturalist is currently published by The University of Chicago Press.

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

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/ucpress.html>.

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

The JSTOR Archive is a trusted digital repository providing for long-term preservation and access to leading academic journals and scholarly literature from around the world. The Archive is supported by libraries, scholarly societies, publishers, and foundations. It is an initiative of JSTOR, a not-for-profit organization with a mission to help the scholarly community take advantage of advances in technology. For more information regarding JSTOR, please contact support@jstor.org.

MINIMAL COMMUNITY STRUCTURE: AN EXPLANATION
OF SPECIES ABUNDANCE PATTERNS

GEORGE SUGIHARA

Department of Biology, Princeton University, Princeton, New Jersey 08544

Submitted October 12, 1978; Accepted October 19, 1979

Here I attempt to provide qualitative insight into the structure of natural communities through an investigation of regularities in species abundance relations. In particular, I aim to show that the ubiquitous canonical lognormal distribution of Preston (1962) and the biogeographical species-area constant, $z \approx 1/4$ (MacArthur and Wilson 1967), are not mathematical artifacts as has been previously suggested (Connor and McCoy 1979; May 1975), but can reflect a broad and very simple underlying form of community organization.

Few propositions in ecology have as much empirical support as Preston's (1962) canonical lognormal hypothesis of species abundance (fig. 1). This statistical abundance distribution has been observed for a diverse array of organisms including diatoms (Patrick 1968; Patrick et al. 1954), soil arthropods (Hairston and Byers 1954), lepidoptera, birds, and mammal faunas (Preston 1962), and for areas ranging in size from diatometer slides to entire continents. It is the general rule for collections with large numbers of taxonomically related species, and has special importance for the theory of island biogeography (MacArthur and Wilson 1967) since it can generate the familiar species-area constant, $z \approx 1/4$, for the relation, $S = cA^z$ where S and A are species count and area, respectively, and c and z are constants. (May [1975] offers the clearest derivation of this relationship.) Repeated investigations show that when z is treated as a fitted constant, it often agrees with the theoretically derived value, $z \approx 1/4$, which rests directly on the canonical hypothesis (see Preston 1962; MacArthur and Wilson 1967; May 1975; Connor and McCoy 1979 for a catalogue of examples).

By convention, a lognormal species curve is canonical if the parameter γ has a constant value of 1, where γ is the ratio of the position of the individual's curve mode, R_N , to the upper truncation point of species curve, R_{\max} (fig. 1). This is essentially an empirical relationship deriving from the observation that R_N and R_{\max} have a tendency to coincide. Notice that when $\gamma = 1$, the independent variables of the general lognormal distribution become coupled to yield a specific one-parameter family of lognormal curves. For example, the specific coupling between species count, S , and the variance of the lognormal, σ^2 , can be written approximately as

$$S = \sigma \sqrt{\pi/2} \exp \left[\frac{(\sigma \ln 2)^2}{2} \right] \quad (1)$$

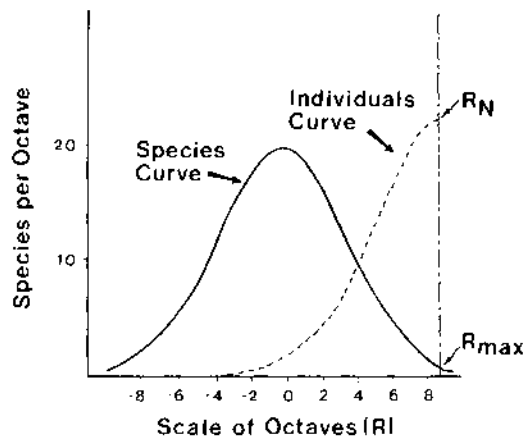


FIG. 1.—The canonical lognormal distribution for an ensemble of 178 species (after Preston 1962). By convention, the x -axis is scaled as logarithmic (base 2) abundance classes or "octaves" of individuals/species, adjusted to have a mean of zero. The species curve denotes the number of species in each octave and the individual's curve shows the number of individuals in each abundance class. This particular distribution is canonical because the mode of the individual's curve, R_N , coincides with the upper truncation point of the species curve, R_{\max} (i.e., $\gamma = R_N/R_{\max} = 1$). Setting $\gamma = 1$ couples species count and variance in a specific way (see fig. 2).

(after eq. [1A]). This equation contains the same information as the phenomenological rule $\gamma = 1$ and is perhaps more meaningful since it shows that under the canonical hypothesis, a large variance in the distribution accompanies a large species count. In the general lognormal distribution, on the other hand, these two parameters are completely independent. The canonical hypothesis, therefore, describes a specific positive coupling between the number of species contained under the distribution and the size of its variance.

In view of the importance and predictive success of this relationship, especially in the species-area context, it is puzzling that little serious effort has been made toward a theoretical justification. Connor and McCoy (1979) have tried to explain the island biogeographic constant, $z = 1/4$, as an artifact of linear regression. They argue that z , which in practice is reckoned from the slope of a linear regression, should behave in a null model as the product of two uniform (0, 1) random variables, i.e., $z = \rho(\sigma_y/\sigma_x)$ where the correlation coefficient (ρ) and the ratio of the standard deviations of the independent and dependent variables (σ_y/σ_x) are uniform on the interval (0, 1). This treatment addresses the species-area relation out of the context of its canonical lognormal underpinnings. Although the expectation of their null model approaches $z \approx 1/4$, the predicted distribution of z values does not agree with the distribution observed. Rather, using the empirical results that they present, one is forced to reject the null model at the 99% level (Sugihara, in press), leading to the conclusion that $z \approx 1/4$ is not an artifact of the regression system.

An attempt at a more complete explanation, aiming directly at Preston's can-

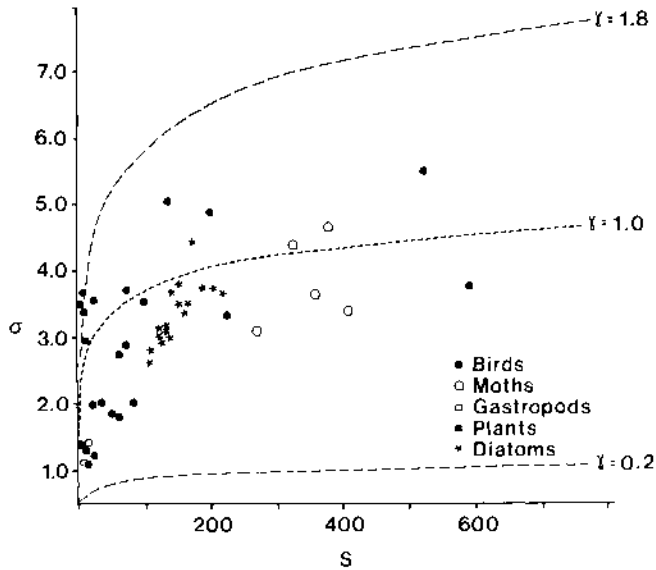


FIG. 2.—A graphical illustration of the validity of the canonical hypothesis, $\gamma = 1$, for an arbitrary set of data on birds, moths (Preston 1948, 1962) gastropods (Kohn 1959), land plants (Oosting 1948), and diatoms (Patrick et al. 1954; Patrick 1968; Sugihara, unpublished data). The relationship between the standard deviation of logarithmic abundances and species count adheres to the rule $\gamma = 1$, rather than to some other close value of γ . In this figure, the curves for $\gamma = 1.8$ and 0.2 were computed from the approximation given in eq. (1A), whereas the canonical relationship ($\gamma = 1$) was determined from values given in Preston (1962).

onical lognormal hypothesis, suggests that $\gamma = 1$ is a robust consequence of general lognormal distributions (May 1975). Although this seems reasonable in principle, it can be shown to be insufficient. In Appendix A, for the unrestricted lognormal curve, it is shown that γ is not confined to some close neighborhood of 1 but can vary freely between 0 and $+\infty$. For ensembles containing a reasonably finite number of individuals and species, γ will range from 0.1 to 6.8. Hence, although the insensitivity of this parameter must contribute to its success, the rule $\gamma \approx 1$ does not cover lognormal curves in general, and strictly speaking the canonical hypothesis is not an artifact of lognormal distributions. Figure 2 illustrates this result empirically. The points shown do not scatter independently as in the general lognormal distribution, but follow the canonical trend, adhering specifically to the relationship given by $\gamma \approx 1$ (after Preston 1962) rather than to some other close value of γ (eq. [1A]). This figure shows that Preston's hypothesis reflects real regularities in the shape of the distribution and cannot be explained simply as an artifact arising from the insensitivity of γ . This is important because it means that a full explanation of the lognormal abundance curve must also account for its canonical form, ruling out many of the traditional explanations for lognormality based on random multiplicative effects (see e.g., MacArthur 1960). These traditional explanations not only fail to account for the canonical variance, but more importantly, when the multiplicative impulses are assumed to be independent one

encounters the serious shortcoming that the variance of the distribution expands monotonically, and without bound, with each successive moment (Aitchison and Brown 1966). Hence, the lognormal species abundance distribution, its particular canonical form, and the species-area constant remain without a substantial theoretical basis.

In what follows, a model is developed which aims to explain these regularities as a consequence of a general form of community structure involving hierarchically related species niches. This hypothesis, which is motivated by simple evolutionary and ecological considerations for generating species diversity, proposes that a minimal form of community organization involving hierarchically related niches can explain the canonical lognormal abundance pattern. The model will be tested first against data for small assemblages containing two and three species and then extrapolated to account for the patterns observed in large ensembles. Evidence is given which suggests that niche apportionment is multidimensional and that the canonical lognormal distribution is not simply an artifact of classification.

FOUNDATION

A clue to these regularities may be culled from the fact that they seem to apply only to taxonomic collections having some measure of ecological homogeneity (Preston 1962, 1980). This demonstrates that a certain degree of evolutionary or ecological similarity is required in order for them to operate. Therefore, explanations based purely on statistical artifact, although attractive for their generality, seem less likely to apply; instead, the evidence suggests that one should seek to understand these phenomena in terms of very general biological models for generating species diversity.

A reasonable approach toward understanding patterns of species abundance is to interpret them in terms of an underlying niche structure. Indeed, because a niche translates ultimately into numbers of organisms (or biomass), observed abundance patterns can offer a useful standard measure of niches, allowing legitimate comparisons to be made between different types of species niches.

It is commonly believed that the relative abundance of a species is a reflection of the amount of limiting resources it controls (Motomura 1932; MacArthur 1957; Whittaker 1965, 1969, 1972, 1977; May 1975, 1976; Pielou 1975). Although the classical niche apportioning theories require a uniform set of limiting resources, it is also plausible to consider apportionment in a heterogeneous resource pool, involving the subdivision of several different sets of niche axes. This allows the apportionment analogy to be extended to large species ensembles which do not possess a uniform set of governing factors.

Suppose a communal niche space (the total niche requirements of a community in Whittaker's [1977] sense) is likened to a unit mass which has been sequentially split up by the component species so that each fragment denotes relative species abundance. The successive subdivisions may correspond to apportioning on different sets of niche axes, which could be driven by either ecological or evolutionary forces. This is similar in spirit to the MacArthur (1957, 1960) broken-

stick model with the important exception that breakage occurs sequentially rather than simultaneously (Bulmer 1974; Pielou 1975). Kolmogoroff (1941) has shown that such sequential fractures can lead asymptotically to lognormal size-frequency distributions. The breakages themselves do not have to be random; however, the magnitude and frequency of breakage must be independent of particle size. Thus, for example, gravel fragments resulting from repeatedly crushing rocks often tend to be lognormal. Instead of smashing rocks one can imagine dividing up relative species abundance in a way that reflects a sequentially divided niche space. Therefore, large numbers of taxonomically related species should tend to have a lognormal distribution of abundance.

The biological motivation for this mechanism depends principally on two propositions. First, the underlying structure of niches should be reflected in the relative abundance pattern (Motomura 1932; MacArthur 1957; Whittaker 1965, 1969, 1972, 1977; May 1975, 1976; Pielou 1975), and second, in general, the minimal niche structure for communities should be hierarchical. This latter proposition derives from the essential differences between species niches, which allows one to sort them into natural groups according to increasing niche similarity. Such a pattern is illustrated, for example, in a niche overlap dendrogram (fig. 3) where communities are sequentially subdivided into smaller and more tightly related functional groups of species. The sequential aspect of the breakage metaphor, therefore, corresponds to this underlying niche hierarchy, with each bifurcation representing a break point (fig. 3). Although evolution is not necessary, it is a sufficient condition for generating this pattern. Whittaker (1977) has proposed that community diversity may evolve by the sequential partitioning and dispersal of species populations in a communal niche space. To some extent, speciation itself can be characterized as the successive carving up and elaboration of a taxon's niche. Such processes should lead in the end to a community structure consisting of subdivided taxonomic guilds. This inevitable tree structure or niche hierarchy is in fact the minimal kind of community structure since, apart from evolution, it could arise naturally from a gradient in niche similarities.

In terms of the breakage metaphor, this situation can be described by the simple case involving successive single fractures. Therefore, the initial unit mass is broken randomly to produce two fragments and one of these is chosen randomly and broken to yield the third, and so on. This process, where particles are sequentially chosen and broken at random, is intended to reflect a hierarchical niche pattern. According to this model, total abundance may expand or shrink proportionally for each species or may remain constant through time (cf. Van Valen's [1974] red queen hypothesis or Levinton 1979). Because the sequential aspect here corresponds to the underlying niche structure, it may be incorrect to interpret a colonization event as the subdivision of an existing hierarchy, since an entirely new set of relationships can be formed.

Although this model is similar in spirit to the MacArthur broken-stick model, it differs from it in several important respects. First, unlike MacArthur's model this mechanism involves sequential rather than simultaneous random breakages. The unit mass is split up by repeated breakage events rather than divided instantaneously with one hammer blow. It should be emphasized that simultaneous breakage

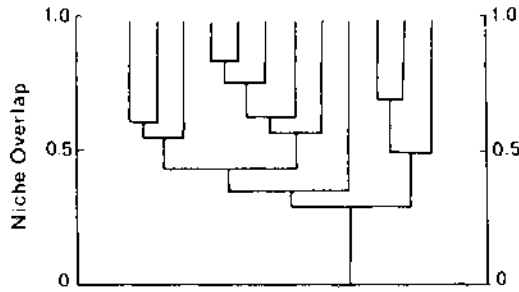


FIG. 3.—A hypothetical niche overlap dendrogram demonstrating hierarchical (minimal) community structure. In the sequential breakage model, each bifurcation in the underlying niche structure corresponds to a subdivision of the communal biomass (abundance).

does not lead to a lognormal distribution but generates a pattern which is more equitable than the one produced by sequential breakage. Second, as will be explained further in the next section, the present model incorporates complex random breakages involving the translation of several niche axes into abundance. This differs from the broken-stick model which involves the uniform breakage of a single resource axis.

Thus, this model not only leads to lognormal distributions but it is also intuitively consistent with evolutionary and ecological conditions for generating species diversity. Encouragingly, it also agrees, at least qualitatively, with Preston's canonical relationship as variance will tend to grow with the convolution of additional species to the community.

MULTIDIMENSIONAL BREAKAGE

Consider first the two-species case involving a single random breakage. At this stage, there is no difference between sequential and simultaneous breakage; therefore, expected proportional abundances can be computed from MacArthur's formula

$$E(p_i) = \frac{1}{S} \sum_{k=i}^S \frac{1}{S - k + 1} \quad (2)$$

where $E(p_i)$ is the expected fraction of the i th most abundant species and S is the number of species in the ensemble. For assemblages with two species, the dominant member will assume values anywhere from 0.5 to 1.0 with uniform probability; and the expectation given by equation (2) is simply the midpoint or 0.75. As Pielou (1975) pointed out, uniform breakage means that all fractional abundances are equally likely; therefore it is essentially meaningless to seek individual examples in nature which agree with equation (2). This casts a shadow over more than a decade of such attempts (Deevey 1969; King 1964; Goulden 1969; Kohn 1959; Longuet-Higgins 1971; MacArthur 1957, 1960; Tramer 1969; Tsukada 1972). Rather, to test the hypothesis it is necessary to consider a distribution of values taken from many assemblages. Swingle's (1950) studies of bass-bluegill combi-

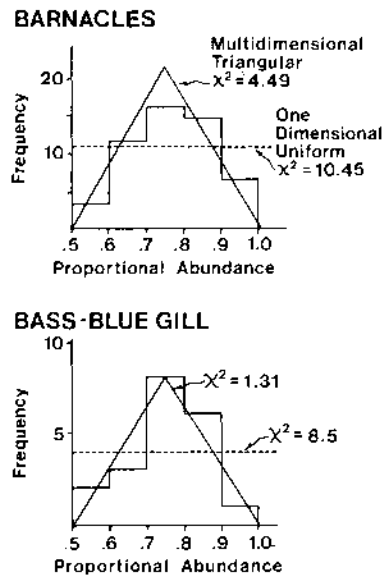


FIG. 4.—A test of one-dimensional breakage (MacArthur 1957, 1960) versus multidimensional breakage for two-species assemblages of barnacles and fish. This figure shows that the relative abundance of the most abundant species of fish (in biomass) and barnacles (in numbers of individuals) is not distributed uniformly but appears to cluster about the expected figure of 0.75. In both cases, the hypothesis of uniform breakage on one niche dimension is rejected at the 95% level.

nations are ideally suited for this as they offer many replicate data from experimental ponds which have come to equilibrium over periods of 2 to 30 yr. There was no sampling error in these studies because the ponds were either completely drained or poisoned, and total biomass could be used to calculate frequencies (fig. 4). Dayton's (1971) barnacle data, although less well suited, are also used because they contain a good number of replicates, and his total exclusion cages ensure that interaction is limited to only two species. These data, however, only represent barnacle associations in the first 3 to 5 mo after settlement, thus the temporal definition of community may be somewhat artificial. Nonetheless, there were no consistent shifts in proportional frequency (calculated from numbers of individuals) at most localities during this period. These were the only studies of two-species associations that I encountered in an arbitrary survey of the literature which had data appropriate for this analysis.

Data for the relative abundance of the dominant species of fish and barnacle are plotted in figure 4. The distributions for the less abundant members have been omitted as they are simply mirror images of the ones shown. The important result is that, rather than a uniform distribution, which is predicted if a single limiting factor is divided up randomly, there is a clear tendency in both cases for values to cluster about the expected figure of 0.75. The hypothesis of uniform breakage on a single limiting factor is rejected for both fish and barnacles at the 95% level.

Whereas individual breakages generate a flat distribution of proportions, com-

posite breakages averaged over several dimensions can explain clustering about the mean. For example, although single throws of a fair die turn up all outcomes uniformly, pairs of throws averaged will accumulate around the expected figure of 3.5. The clustering will tighten as each point includes more throws, rising in the limit to a normal distribution with a shrinking variance. In a similar manner, the relative abundance of a species pair can be reckoned as the mean of random breakages on several dimensions, representing different resources or discrete environmental regimes which the organisms divide up. Accordingly, the fractional abundance, p_s , for k -dimensional apportioning between two species is a simple expectation,

$$p_s = \sum_{r=1}^k \alpha_r p_{r,s} \quad (3)$$

(where α is a weighting factor such that $\sum \alpha = 1$ and $p_{r,s}$ is the relative fraction that species s gets on dimension r), and the distribution of p_s will modulate around the mean value.

If, for simplicity, two equally important factors determine fractional abundance, then there are three possibilities to consider. *Case 1.* If the dominant species tends to get the larger share of both factors, then this is roughly equivalent to averaging two uniform fractures in the interval $[0.5, 1)$. The distribution will be triangular around a mean value of 0.75 (fig. 5A). *Case 2.* On the other hand, the ranked shares may be negatively correlated so that the species which gets a larger fraction on one dimension gets the smaller part of the other. In this case, a narrow triangular distribution centered about 0.5 will result (fig. 5B). *Case 3.* Intermediate between these two extremes, the rank on either dimension may be completely independent yielding a wide triangular distribution about 0.5 (fig. 5C). Data for both fish and barnacles are most consistent with case 1, and the agreement is especially good ($P < .05$) for the bass-bluegill combinations. It is not surprising that the more abundant species should get the larger share of several resources. Unfortunately, one cannot say that only two factors are involved since a similar degree of central tendency is realized with many more factors which are differentially, rather than equally, important.

Variation about the expected value of 0.75 may be further reduced if samples are large and heterogeneous so that each homogeneous local area is roughly an independent broken-stick variable. Large samples which encompass more area are simply an average of these local values, and the relative proportions should therefore cluster more tightly around the mean. This is not likely to be important, however, in the fish and barnacle data, because the experimental ponds and the small exclusion cages are essentially homogeneous.

It should be mentioned that one cannot explain clustering around the mean value in terms of sampling along a gradient. For example, in a hypothetical transect, if species A varies in abundance linearly from 0 to K and B goes from K to 0, random sampling along this interval will yield a flat distribution of proportions rather than a modal one. Similarly, a stochastic reformulation of generalized competition models having random coefficients will not generate the observed clustering of abundances (G. Sugihara and L. Nunney, in prep.). However,

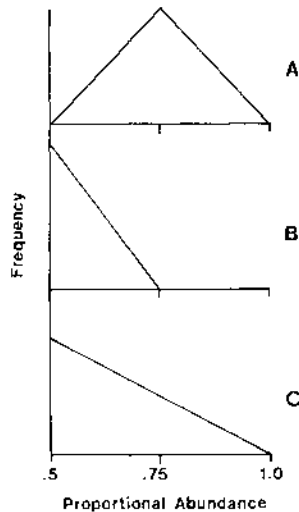


FIG. 5.—Three distinct abundance distributions are possible if two equally important factors determine fractional abundance. *A*, *Case 1*. If the dominant species gets the larger share of both factors this is equivalent to averaging two uniform variables in the interval $[0.5, 1)$. *B*, *Case 2*. If the niche dimensions are negatively correlated so that the species which gets the larger share of one dimension, also gets the smaller share of the other, then this is equivalent to averaging two uniform variables, with one in the interval $(0, 0.5]$ and the other in the interval $(0.5, 1)$. *C*, *Case 3*. When the two dimensions are completely independent, this can be represented as the average of two uniform variables in $(0, 1)$.

recognizing the possibility of an alternative recondit mechanism, the development that follows will depend only on the empirical fact of the pattern observed in the two-species case and only indirectly on its theoretical underpinnings.

INSTANTANEOUS VERSUS SEQUENTIAL BREAKAGE

Results observed for the two-species case will now be extended to predict ratios for assemblages with three species. Taking case 1 as the empirically validated rule for breakage, figure 6 presents the two alternative ways of generating ranked expectations depending on whether the larger or smaller fraction is broken the second time. The method used to calculate the expected sizes of the fractions generated for each breakage pathway is given in Appendix B. The final ranked expectation is simply the average of the ranked expectations for the two possible pathways.

In table 1, I compare the predictions of the sequential hypothesis and MacArthur's instantaneous breakage model for an arbitrary collection of three-species associations including molluscs (Fuller 1972), trees (Keever 1973; Jackson and Faller 1973), fish (Swingle 1950), and barnacles (Dayton 1971). Notice that these data do not represent single assemblages, but are expectations averaged over numerous collections. Although the sequential model fits more closely ($P \approx .05$) than the MacArthur model ($P \approx .40$), the difference is not compelling. A more

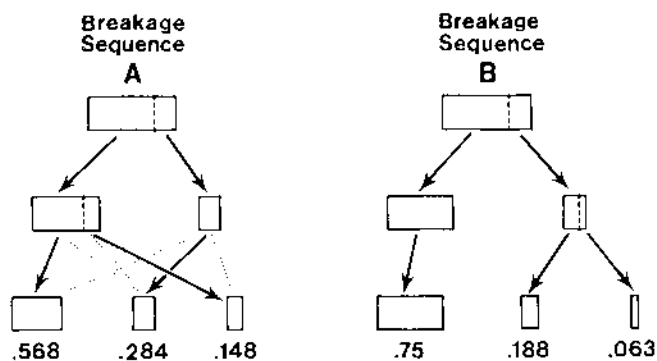


FIG. 6.—Two possible breakage sequences for generating proportional abundances for a three-species assemblage. The values at the bottom of each sequence are the expected proportional sizes (see Appendix B for the method of calculation). The solid arrows in sequence A represent the most probable sequence, and the dashed arrows indicate other possible arrangements depending on the relative sizes that pieces break into with a second breakage.

TABLE 1
COMPARISON OF THE BROKEN-STICK AND SEQUENTIAL BREAKAGE MODELS

SOURCE	RANKED PROPORTIONAL ABUNDANCE			χ^2
	p_1	p_2	p_3	
MacArthur broken-stick611	.277	.111	1.02
Sequential hypothesis659	.235	.105	.10
Observed data675	.227	.095	
(Molluscs, trees, barnacles, fish) $n = 57$				

NOTE.— $n = 57$ is the total number of three-species assemblages used in the calculations. p_i is the proportional abundance of the i th most abundant species. Values for the sequential hypothesis were obtained by averaging the expected values for the two possible breakage pathways (fig. 6).

powerful test was not attempted because the exact forms of the distributions are unknown.

Better evidence for sequential breakage comes from the observation of frequencies within a single community. If the breakage path is ecologically meaningful, or reflects the evolutionary partitioning of niches, then a given type of community should follow one of the two pathways shown in figure 6. Of the results surveyed, only the studies on fish (Swingle 1950), barnacles (Dayton 1971), and trees (Jackson and Faller 1973) offered a sufficient number of data to test this. These values, shown in table 2, appear to agree nicely with the sequential hypothesis: The conifers of Wizard Island follow breakage sequence A, while fish and barnacles follow B. Not only are the expected proportions in close accord with the model, but more importantly the observed variances for these p_i 's are almost precisely those which would be generated by sequential breakage. For breakage

TABLE 2
DATA FROM TABLE 1 SORTED INTO SEQUENCES A AND B FROM FIGURE 6

SOURCE		RANKED PROPORTIONAL ABUNDANCE		
		p_1	p_2	p_3
Breakage Sequence				
A	Model prediction (approximate)	.568 $\sigma^2 = .011$.284 $\sigma^2 = .006$.148 $\sigma^2 = .004$
	Trees (Wizard Island) $n = 15$.566 $\sigma^2 = .010$.300 $\sigma^2 = .008$.144 $\sigma^2 = .006$
Breakage Sequence				
B	Model prediction (exact)	.750	.188 $\sigma^2 = .004$.063 $\sigma^2 = .001$
	Barnacles $n = 18$.758 $\sigma^2 = .006$.165 $\sigma^2 = .004$.077 $\sigma^2 = .002$
	Fish $n = 12$.757 $\sigma^2 = .006$.174 $\sigma^2 = .004$.069 $\sigma^2 = .002$

NOTE.—Model variances predicted for sequence A are approximate since they assume that the proportions for each successive breakage were chosen from a triangular distribution (cf. fig. 4). This assumption was not necessary in sequence B where it was possible to use the observed variance of p_1 (i.e., the observed variance for a single breakage = .006) to predict the variances of p_2 and p_3 (both resulting from a second breakage).

sequences B (fig. 6) the variances of p_2 and p_3 are conditional values predicted for the fractional abundances resulting from a second breakage. In this sequence, the larger piece, p_1 , results from a first breakage and the two smaller ones are produced by a second breakage. From the observed variance in the larger fraction (first breakage) one can calculate the conditional variances for the two smaller pieces (see Appendix C for method of calculation). This conditional argument, however, does not apply to sequence A where one cannot know the exact variance of the fractions resulting from a first breakage. In this sequence, involving the breakage of the larger segment, none of the p_i 's will result purely from a single fracture. For example, the largest segments, which are averaged to determine p_1 , may have resulted from either a first or second breakage (fig. 6). Therefore it is necessary to approximate the variances for the ranked proportions in this sequence by assuming that the breakage fractions were chosen from a triangular distribution as observed in the two-species case (Appendix B).

GENERATION OF THE CANONICAL LOGNORMAL AND THE SPECIES-AREA CONSTANT

Because the number of possible breakage sequences is $(S - 1)!$, it is unreasonable to discuss specific pathways for larger assemblages. Instead, the relationship

TABLE 3
COMPARISON OF THE SEQUENTIAL BREAKAGE MODEL WITH
THE CANONICAL LOGNORMAL HYPOTHESIS

S	CANONICAL HYPOTHESIS σ	SEQUENTIAL BREAKAGE HYPOTHESIS	
		$\hat{E}(\sigma)$	$\widehat{\text{VAR}}(\sigma)$
25	2.97	2.80	.4509
50	3.37	3.23	.4714
100	3.72	3.61	.4318
200	4.05	3.97	.4106
400	4.36	4.32	.3856
800	4.66	4.63	.3522
1,000	4.75	4.73	.3368

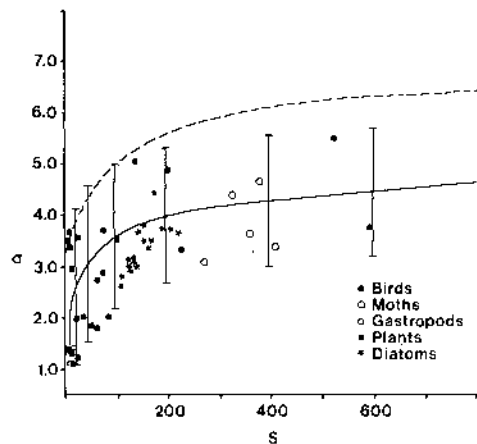


FIG. 7.—Graphical evidence for the sequential breakage hypothesis for an arbitrary assortment of taxonomic collections (see fig. 1). The solid line and error bars represent the mean and two standard deviations of the mean predicted by the sequential breakage model. The dashed line represents the trend generated by biologically trivial or arbitrary sequential breakages, i.e., where the successive breakage fractions are chosen uniformly from the interval (0,1). This latter trend is identified with the null hypothesis that species are artifacts of classification in the nominalist sense.

between the standard deviation of logarithmic abundances, σ , and species count, is examined with the aim of determining whether the coupling of these parameters agrees with the empirically based canonical relationship.

The results of a numerical simulation of the sequential breakage model are compared in table 3 to the canonical hypothesis. The correspondence is surprisingly good. An even stronger statement is made in figure 7 which demonstrates an almost perfect match between the variance in the model and the scatter of points found in nature. These numerical results are generated by assuming that the breakage fractions are precisely 0.75, 0.25, which is roughly what should be expected if the collections are large and heterogeneous, and if pairwise rank is

preserved, i.e., if species A is more abundant than B in all independent localities where they occur together. Under these circumstances, because of the modulating effect of sample size, the values should peak into a spike around the expected figure. Virtually identical results can be obtained, however, by assuming that some breakages are variable as in the two-species case, and that pairwise rank is not always maintained. This can be portrayed approximately as an intergrading mixture of point fractures (0.75, 0.25, and 0.5, 0.5) and breakage distributions shown in figures 5A and 5C to yield a distribution with a peak roughly midway between 0.5 and 0.75. The relationship between S and σ generated by this regime is virtually indistinguishable from that presented in table 3. Both of these possibilities are plausible extrapolations of the two- and three-species cases previously observed. On the other hand, if one arbitrarily assumes that the successive breakages are uniform in the interval (0,1), then the canonical relationship is not obtained (fig. 7). This rules out an alternative and biologically trivial interpretation of breakage proposed by Aitchison and Brown (1967) in which arbitrary hierarchical systems of classification are conceived of as a sequential breakage operation. Clearly then, the canonical lognormal distribution cannot be explained as an artifact of classification in the nominalist sense, and species do not represent arbitrary subdivisions of biomass.

These results demonstrate that a reasonable extension of the two- and three-species cases can generate the ubiquitous canonical lognormal distribution, whereas arbitrary sequential breakages will not account for the relationship. Furthermore, because the model generates the canonical lognormal distribution, it can also explain the Preston-MacArthur-Wilson species-area constant $z = 1/4$. It is surprising and encouraging that a simple induction of the patterns observed in small ensembles can account so well for such large-order regularities, and reassuring that the generality of this hypothesis corresponds to the ubiquity of the patterns which it aims to explain.

CONCLUDING REMARKS

The evidence suggests that the canonical lognormal abundance pattern and the species-area constant may be most simply explained as consequences of a hierarchically structured communal niche. This model is admittedly phenomenological; however in view of the regularities addressed, this is perhaps inevitable. It is intended strategically to capture the first-order effects of the most minimal kind of community structure, and exactly how literally it may be interpreted remains to be seen.

In constructing my argument the results of the two- and three-species cases were induced to make predictions about the multispecies case. If the hypothesis of minimal community structure is correct then it should be possible to recover roughly the pattern observed for two species from a large ensemble by constructing a dendrogram of niche overlaps and considering each bifurcation as the two-species case. If the area sampled is large enough, then the distribution of proportions thus obtained should be roughly triangular with a peak midway

between 0.5 and 0.75. Therefore, an independent test of the hypothesis may be accomplished with a dendrogram containing the abundance of each species.

For the Wizard Island trees (table 2), although the breakage path was conserved, the rank for each species varied in different localities. Because these associations were from a range of altitudes and soil conditions, the lack of a fixed ordering is not surprising, and demonstrates that niche subdivision is not unique for a particular species ensemble and depends on the specific environmental context. The species rank-list for barnacles supports this idea: For tide levels under 1.3 m, the rank-list is *Balanus glandula* > *B. cariosus* > *Cthamalus dalli*, whereas at stations higher in the intertidal, *C. dalli* tends to predominate over *B. cariosus*. On the other hand, Swingle's fish associations, all from similar pond environments, tend to maintain their ordering. These examples align most consistently with ecological forces determining breakage, but this should not reduce the plausibility of evolutionary niche subdivision. It seems likely that the principle fractures which divide a large community into major functional groups will reflect evolutionary lines of niche apportionment, and that finer partitioning, within smaller more closely related species clusters, will be guided more strongly by ecological forces.

It is interesting that the breakage pathway for a given set of species was conserved (table 2). Jackson and Faller's trees cleaved the most abundant species (sequence A) while Swingle's fish and Dayton's barnacles cleaved the least abundant one (sequence B). One possible explanation, in terms of competition, is that among trees the incentive for sharing a large resource pool may have outweighed the disadvantage of competing with an abundant species, whereas the opposite may have held for fish and barnacles. The full ecological significance of this observed invariance in breakage pathway may offer interesting biological insights and deserves further consideration.

SUMMARY

Recent proposals that the canonical lognormal distribution and the resulting species-area constant, $z = 1/4$, are artifacts of the general lognormal curve and regression techniques, are shown to be inadequate. An alternative hypothesis is suggested which accounts for these regularities in terms of a hierarchical community structure represented by a sequentially divided niche space. This hierarchical pattern, which can be considered to be a minimal form of community structure, derives from evolutionary and ecological considerations for generating species diversity, and is shown to account for the observed abundance structures of small ensembles as well as large natural communities. Evidence is presented which implies that niche apportionment between species may involve the random division of more than one resource, and an interesting invariance in the pattern of apportionment is observed for assemblages with three species. The possibility that the canonical lognormal distribution is a conceptual artifact resulting from arbitrary systems of classification is considered and shown to be false. Aside from its intuitive appeal, the model presented should be of interest because it offers

explanations of two ubiquitous patterns in nature: the canonical lognormal and the resulting species-area constant.

ACKNOWLEDGMENTS

I thank D. A. Livingstone, R. M. May, H. S. Horn, M. Huston, G. F. Estabrook, J. H. Vandermeer, and R. H. Whittaker for their valuable advice and discussions in preparing this manuscript. This work was supported in part by a Princeton University Prize Fellowship.

APPENDIX A

This section contains a counterargument to the proposition that $\gamma = 1$ is a general property of lognormal distributions.

Following May (1975):

S = total number of species in the lognormal assemblage.

J = total number of individuals/number of individuals of the rarest species.

$a = 1/\sqrt{2\sigma^2}$ where σ^2 is the variance of the lognormal species curve.

$\gamma = R_n/R_{\max}$ such that if the species curve mode is at zero, R_n = the position of the individual's curve mode, and R_{\max} = the position of the upper truncation point of the species curve.

Upper and lower limits for γ are obtained conveniently from the relation

$$\gamma \approx \frac{\ln 2}{2a\sqrt{(\ln KS)}} \quad (1A)$$

where $K = 2a/\sqrt{\pi}$. This is similar to May's eq. (3.8) with the exception that the area contained under the upper and lower limbs of the distribution is assumed to sum to 1 (after Preston 1962) rather than 2. Holding a fixed, it follows that $\gamma \rightarrow 0$ as $S \rightarrow \infty$. Because $a > 0$ and $\ln KS > 0$, γ is strictly positive. Hence, $\gamma_{\min} = 0$ is a greatest lower bound. An upper limit for γ is obtained from eq. (1A) simply by setting $S = c/2a$ where $c > \sqrt{\pi}$, and allowing $a \rightarrow 0$, whence, $\gamma \rightarrow \infty$.

Collecting these remarks yields the following:

$$0 < \gamma < \infty. \quad (2A)$$

Hence, $\gamma = 1$ is not a necessary property of the general lognormal.

It is now reasonable to ask whether restricting S and J to realistic limits leads to lognormal distributions with $\gamma = 1$. For $50 < S < 10^4$ and $10^4 < J < 10^{12}$ the corresponding range for γ is roughly

$$0.1 < \gamma < 6.8 \quad (3A)$$

where (3A) follows from equation (1A), and the relations: $\ln J = \Delta^2(1 + \gamma)^2$ for $\gamma < 1$, and, $\ln J = 4\gamma\Delta^2$ for $\gamma > 1$, where $\Delta^2 = \ln KS$ (May 1975). Therefore, the property $\gamma = 1$ is not inevitable for lognormal species frequency curves, even within reasonable limits for S and J .

APPENDIX B

In sequence B, which involves the division of the smaller fragment (fig. 6) the ranked expectations are easily computed as follows:

$$\begin{aligned} E(p_1) &= .75 \\ E(p_2) &= (.25)(.75) = .1875 \\ E(p_3) &= (.25)(.25) = .0625 \end{aligned}$$

This simple computation is possible because there is no ambiguity in the rank ordering of fractions in relation to their sequence of generation. That is, p_1 is always generated by a single breakage and p_2 and p_3 always involve a second breakage. The situation is not as clear cut in sequence A, however, where the larger piece is divided a second time. To see this, let x_1 and x_2 be the larger fractions used in the first and second breakages, respectively, and let $(1 - x_1)$ and $(1 - x_2)$ represent the smaller fractions. Because it is the larger piece from the first breakage that is divided, three cases must be considered:

- 1) $(1 - x_1) > x_1x_2 > x_1(1 - x_2)$.
- 2) $x_1x_2 > (1 - x_2) > x_1(1 - x_2)$.
- 3) $x_1x_2 > x_2(1 - x_2) > (1 - x_1)$.

The terms in each of these inequalities correspond to p_1 , p_2 , and p_3 . One can see the ambiguity in p_1 , for example, since it can involve pieces resulting from a first $(1 - x_1)$ and/or second (x_1x_2) breakage. Therefore, in calculating the ranked expectations and variances for sequence A the above three cases must be considered. The expectations may be computed as follows:

$$\begin{aligned} E(p_1) &= \int_{\frac{1}{2}}^1 \int_{\frac{1}{2}}^{\frac{1}{1+x_2}} (1 - x_1) dF_{x_1} dF_{x_2} \\ &\quad + \int_{\frac{1}{2}}^1 \int_{\frac{1}{1+x_2}}^1 x_1 dF_{x_1} dF_{x_2}. \\ E(p_2) &= \int_{\frac{1}{2}}^1 \int_{\frac{1}{2}}^{\frac{1}{1+x_2}} x_1x_2 dF_{x_1} dF_{x_2} \\ &\quad + \int_{\frac{1}{2}}^1 \int_{\frac{1}{1+x_2}}^{\frac{1}{2-x_2}} (1 - x_1) dF_{x_1} dF_{x_2} \\ &\quad + \int_{\frac{1}{2}}^1 \int_{\frac{1}{2-x_2}}^1 x_1(1 - x_2) dF_{x_1} dF_{x_2}. \\ E(p_3) &= \int_{\frac{1}{2}}^1 \int_{\frac{1}{2}}^{\frac{1}{1+x_2}} x_1(1 - x_2) dF_{x_1} dF_{x_2} \\ &\quad + \int_{\frac{1}{2}}^1 \int_{\frac{1}{1+x_2}}^{\frac{1}{2-x_2}} x_1(1 - x_2) dF_{x_1} dF_{x_2} \\ &\quad + \int_{\frac{1}{2}}^1 \int_{\frac{1}{2-x_2}}^1 (1 - x_1) dF_{x_1} dF_{x_2}, \end{aligned}$$

where F_{x_1} and F_{x_2} are the cumulative densities of the triangular distribution on the interval $[\frac{1}{2}, 1]$. The calculation of variances follows trivially from the relations above.

APPENDIX C

This section illustrates a method for computing the conditional variances for fractions resulting from a second breakage in sequence B (see fig. 6). As a specific example, we will compute the variance of p_2 from breakage path B (see fig. 5). This is the larger fraction resulting from the second breakage.

Let $y = (1 - x_1)$ denote the size of the smallest fraction from the first breakage, and $p_2 =$ the size of the largest piece from the second fracture. The distribution of $p_2, f(p_2)$, is conditional on the distribution of y , and one can write

$$f(p_2) = \int f(p_2|y)f(y) dy. \quad (1B)$$

Thence, the variance of $p_2, V(p_2)$, is simply the sum of the expectation of the conditional variance plus the variance of the conditional expectation,

$$V(p_2) = E[V(p_2|y)] + V[E(p_2|y)]. \quad (2B)$$

If $V(y)$ is observed to have a value of a , then $V(p_2|y) = ay^2$ and $E(p_2|y) = (3/4)y$, whereupon

$$V(p_2) = (9/16)a + aE(y^2). \quad (3B)$$

Because, in general, $V(t) = E(t^2) - E(t)^2$, therefore $E(y^2) = a(a + 1/16)$ and equation (3B) becomes $V(p_2) = a(a + 5/8)$. For example, in the fish data $a = .006$ and $V(p_2) = .004$.

LITERATURE CITED

- Aitchison, J., and J. A. C. Brown. 1966. The lognormal distribution. Cambridge University Press, Cambridge.
- Bulmer, M. G. 1974. On fitting the Poisson lognormal distribution to species-abundance data. *Biometrics* 30:101-110.
- Connor, E. F., and E. D. McCoy. 1979. The statistics and biology of the species-area relationship. *Am. Nat.* 113:791-833.
- Dayton, P. K. 1971. Competition, distribution, and community organization: the provision and subsequent utilization of space in a rocky intertidal community. *Ecol. Monogr.* 41:351-389.
- Deevey, E. S., Jr. 1969. Specific diversity in fossil assemblages. Pages 224-241 in *Diversity and stability in ecological systems*. Brookhaven Symp. Biol. No. 22.
- Goulden, C. E. 1969. Temporal changes in diversity. Pages 96-102 in *Diversity and stability in ecological systems*. Brookhaven Symp. Biol. No. 22.
- Hairston, N. G., and G. W. Byers. 1954. The soil arthropods of a field in southern Michigan: a study in community ecology. *Contrib. Lab. Vertebr. Biol. Univ. Mich.* 64:1-37.
- Jackson, M. T., and A. Faller. 1973. Structural analysis and dynamics of the plant communities of Wizard Island, Crater Lake National Park. *Ecol. Monogr.* 43:441-461.
- Keever, C. 1973. Distribution of major forest species in southeastern Pennsylvania. *Ecol. Monogr.* 54:303-327.
- King, C. E. 1964. Relative abundance of species and MacArthur's model. *Ecology* 45:716-727.
- Kohn, A. J. 1959. The ecology of *Conus* in Hawaii. *Ecol. Monogr.* 29:47-90.
- Kolmogoroff, A. N. 1941. Über das logarithmisch normale Verteilungsgesetz der Dimensionen der Teilchen bei Zerstückelung. *C. R. (Doklady) Acad. Sci. URSS.* 31:99-101.
- Levington, J. S. 1979. A theory of diversity equilibrium and morphological evolution. *Science* 204:335-336.
- Longuet-Higgins, M. S. 1971. On the Shannon-Weaver index of diversity, in relation to the distribution of species in bird censuses. *Theor. Popul. Biol.* 2:271-289.
- MacArthur, R. H. 1957. On the relative abundance of bird species. *Proc. Natl. Acad. Sci. USA* 43:293-295.
- . 1960. On the relative abundance of species. *Am. Nat.* 94:25-36.
- MacArthur, R. H., and E. O. Wilson. 1967. *The theory of island biogeography*. Princeton University Press, Princeton, N.J.

- May, R. M. 1975. Patterns of species abundance and diversity. Pages 81–120 in M. L. Cody and J. M. Diamond, eds. *Ecology and evolution of communities*. Harvard University Press, Cambridge, Mass.
- . 1976. Patterns in multi-species communities. Pages 142–162 in R. M. May, ed. *Theoretical ecology, principles and applications*. Saunders, Philadelphia.
- Motomura, I. 1932. A statistical treatment of associations (In Japanese). *Jpn. J. Zool.* 44:379–383.
- Oosting, Henry J. 1948. *The study of plant communities*. Freeman, San Francisco.
- Patrick, R. 1968. The structure of diatom communities in similar ecological conditions. *Am. Nat.* 102:173–183.
- Patrick, R., M. Hohn, and J. Wallace. 1954. A new method of determining the pattern of the diatom flora. *Not. Nat.* 259.
- Pielou, E. C. 1975. *Ecological diversity*. Wiley, New York.
- Preston, F. W. 1948. The commonness, and rarity, of species. *Ecology* 29:254–283.
- . 1962. The canonical distribution of commonness and rarity. *Ecology* 43:185–215, 410–432.
- . 1980. Noncanonical distributions of commonness and rarity. *Ecology* 61:88–97.
- Sugihara, G. 1981. $S = CA^z$, $z = \frac{1}{2}$; a reply to Connor and McCoy. *Am. Nat.* (in press).
- Swingle, H. S. 1950. Relationships and dynamics of balanced and unbalanced fish populations. Agricultural Experiment Station of the Alabama Polytechnic Institute. No. 274:1–74.
- Tramer, E. J. 1969. Bird species diversity; components of Shannon's formula. *Ecology* 50:927–929.
- Tsukada, M. 1972. The history of Lake Nojiri, Japan. *Trans. Conn. Acad. Arts Sci.* 44:337–365.
- Van Valen, L. 1973. A new evolutionary law. *Evol. Theory* 1:1–30.
- Whittaker, R. H. 1965. Dominance and diversity in land plant communities. *Science* 147:250–260.
- . 1969. Evolution of diversity in plant communities. Pages 178–196 in *Brookhaven Symp. Biol.* No. 22.
- . 1972. Evolution and measurement of species diversity. *Taxon* 21:213–251.
- . 1977. Evolution of species diversity in land communities. *Evol. Biol.* 10:1–66.