Vol. 102, No. 924 The American Naturalist March–April, 1968

ALTERNATE DERIVATIONS OF A SPECIES-ABUNDANCE RELATION

JOEL E. COHEN

Society of Fellows, Harvard University, Cambridge, Massachusetts 02138

Exactly the same numerical predictions of the average relative abundance of species that follow from MacArthur's (1957) "broken stick" model also follow from a "balls and boxes" model (Cohen, 1966) with a different set of assumptions. This paper presents a third model, the "exponential" model, leading to the same numerical predictions. The assumptions of this third model are nearly opposite to those of MacArthur's. The numerical data which have been taken to confirm the broken-stick model in fact confirm all three models equally. Hence, this paper describes some experimental and field obervations which could discriminate among the available models.

THE BROKEN-STICK MODEL

Considering a community of populations of n different species within some taxon, the broken-stick model assumes that some critical (abundance-limiting) factor in the environment is fixed in quantity at, say, s units (per unit time, if the critical factor is a rate). Let $n - 1$ points uniformly distributed between zero and s divide the critical factor into n intervals. The "order statistics" of interval size are the interval lengths rearranged in order of increasing size. The model predicts that the ranked average abundances of the species will be proportional to the expected values of the order statistics of interval size.

The biological interpretation of this model has been to assume that species partition the available, fixed supply of the critical factor: the species divide it into mutually disjoint, exhaustive subsets. If the model is to have any usefulness, the critical factor must be some measurable dimension of species' niches. Energy input to the community is currently favored to be the critical factor. Further explication of the biology of the model appears in MacArthur (1960), Slobodkin (1961), and King (1964). Criticisms that the model seems to predict accurately only in certain narrowly defined circumstances yielding certain kinds of data are evaluated in Cohen (1966) and are quite relevant also to the exponential model to be presented.

Reviewing the numerous empirical studies of species' abundance stimulated by MacArthur's broken-stick model, King (1964) concluded:

There is, in the author's opinion, an adequate body of data to permit the statement that the fit of at least some natural associations to the model is not fortuitous. This is not evidence that the basic stipulation of nonoverlapping and contiguous niches is correct since the former is obviously not true, at least not true in the strictest sense. There may, however, be some critical area, such as food utilization, in which very little overlap is permitted in a stable system. If this is true the limits of such permissible overlap have yet to be defined. Perhaps the studies which will contribute most to the evaluation of the model are studies of natural history and population dynamics of the tested groups. Without these data it is impossible to accept or reject the biological assumptions on which the MacArthur model is based.

THE BALLS AND BOXES MODEL

If n species are considered, the balls and boxes model likens the environment to a target of n boxes. Each box is a "subniche." The set of all boxes occupied by balls of a species at the end of the game is that species' "niche." For example, a waterhole during a certain season, a grove of fruit trees, or insects of a certain size might constitute subniches in a game being played by one group of species. Further explication of the biology of the model is offered in Cohen (1966, chap. iii).

The n species distribute balls into the boxes until each species' "niche" contains a number of subniches different from the number of subniches in the niche of each other species. That is, one species throws balls until it has at least one ball in all n boxes; another throws until it has at least one in any $n - 1$ distinct boxes; and so on down to the least abundant species, which throws just one ball, which must land in some one box. The balls and boxes model predicts that the ranked average abundances of the species will be proportional to the ranked expected values of the numbers of balls thrown by each species.

The biological interpretation of this model, in simplified form, has been that the principle of competitive exclusion is satisfied by letting species' niches overlap as long as there is at least one subniche they do not have in common. It is entirely possible that all the subniches of one species may be among the subniches of a second species, as long as the second has at least one subniche which the first does not. Hence this model requires a considerably weaker form of exclusion than does the broken-stick model, but it leads to the same predictions of ranked average abundances.

THE EXPONENTIAL MODEL

Suppose n quantities (random variables) $X_1, \ldots X_n$ are independently and identically distributed with the cumulative distribution function

$$Pr\{X_i \leq x\} = 1 - e^{-\lambda x}, \qquad \lambda > 0, \qquad x \geq 0, \qquad i = 1, 2, \cdots, n. \tag{1}$$

The X_i are said to be exponentially distributed with scale parameter λ and location parameter zero. Defining

$$R_i = X_i/(X_1 + \cdots + X_n), \qquad i = 1, 2, \cdots, n, \tag{2}$$

and ranking the n values R_i in increasing size gives the n order statistics $R_{(1)}, \ldots, R_{(n)}$. The expected value of the ith order statistic is

$$E(R_{(i)}) = n^{-1} \sum_{j=1}^{i} (n - j + 1)^{-1} \tag{3}$$

exactly as in the broken-stick model and the balls and boxes model. The result (3) is proved in the Appendix.

The exponential model proposes that, at least under the circumstances in which the numerical predictions (3) are confirmed, the abundances of the species considered behave like independently, identically, exponentially distributed random variables. Hence, the average relative abundances are given by (3). The exponential model imposes no restrictions whatsoever on the degree of overlap or similarity between any or all dimensions of any species' "niche." The model refers only to observable abundances and not to niches at all.

The exponential model is convincing only if there is some good reason to suppose that the species' abundances might be exponentially distributed. Feller (1966, chap. i) presents a slightly less than infinite number of ways in which the exponential distribution arises mathematically. Biological interpretations can be attached to many of these. For instance (Feller, 1966, p. 1), suppose the addition of one animal to a species' population has probability $1 - p$ of occurring and probability p of not occurring; suppose that this probability is independent of the existing size of the population. (I do not claim that this is true in general, only that it may be under the circumstances where [3] holds.) If population growth, viewed as a sequence of Bernoulli trials, is supposed to continue until the first animal fails to be added to the population (until the first failure of a Bernoulli trial), population size will be geometrically distributed. For large population sizes and appropriate values of p, the geometric distribution approaches an exponential one.

In general, when species are studied without reference to community structure, abundances seem to be nearly log normally distributed rather than exponentially distributed (Preston, 1948; Whittaker, 1965), so the mode of population growth just proposed certainly does not always hold. Whether the populations of the special communities for which (3) has been confirmed may be considered to grow by the mode just described or by some other mode leading to the exponential distribution is an open question.

IMPLICATIONS

Now that three distinct models predict the same set of ranked average relative abundances for a community of n species, one major point of King's conclusion becomes more important: "The fit of at least some natural associations to the [broken stick] model . . . is not evidence that the basic stipulation of nonoverlapping and contiguous niches is correct" (1964, p. 726). But experimental and field observations could be made in order to discriminate among the three models. Here are some suggestions.

Suppose energy input to the community has been selected as the critical, abundance-determining factor. Equivalently, suppose we wish to test whether energy input is a critical, abundance-determining dimension of species' niches and whether species partition energy according to the broken-stick model. We set up a terrarium or aquarium containing a stable community of *n* (perhaps only two or three) species in some taxon, and we make as sure as possible that the energy inputs to the community are as constant as possible over time. Thus food inputs, heat flows, and light exposure are either kept constant or kept at a fixed level per, say, a 24-hr period. The abundances of the species are observed on each of a set of dates over some long period of time (long relative to the time in which fluctuations in population size occur).

Among the possible measures of abundance are number of animals, biomass, and respiration calories. Naturally any measure which requires destruction of the animals makes it impossible to observe the same community at a later date.

The abundances on each date of observation are then ranked from smallest to largest and divided by the sum of the abundances on that date. This procedure yields the relative abundance on each date. At the end of the experiment, all the smallest relative abundances are averaged together to give the average smallest relative abundance; all the next-to-smallest abundances are averaged together, and so on up to the largest abundances. The resulting *n* average values are to be compared with the predictions of equation (3).

If (3) is confirmed, we then face the problem of discriminating among the broken-stick, balls and boxes, and exponential models; to this problem we turn in a moment.

If the observed distribution is flatter than would be predicted by (3), that is, if the largest average value observed is not as large as the largest average value predicted and if the smallest average value observed is larger than the smallest average value predicted, then we must consider the possibility of a threshold effect—a minimum population size required for each species to survive.

A least squares formula for estimating the threshold from the observed values, assuming the broken-stick model, is given in Cohen (1966, chap. ii) along with a formula for predicting the ranked average abundances according to the broken-stick model when this threshold is taken into account. Kendall and Stuart (1961, p. 97) give the computationally simpler, best linear (minimum variance, maximum likelihood) estimator of the threshold (location parameter) for the exponential distribution. (In the 47 cases for which data are presented in the Appendix to Cohen [1966], the numerical values of the threshold given by the least squares estimator and by the best linear estimator are quite close.) Predicted relative abundances in the exponential case are calculated from (2) above, where now the X_i need not have a location parameter equal to zero. If the predicted relative abundances, adjusted for threshold, correspond well to the observed values, then we may proceed to try to distinguish among the three models.

If, even after this adjustment for threshold, the predictions seem far from the observations, then all three models must be considered not relevant to the experimental situation. Insofar as the experimental situation can be taken as a paradigm of the distribution of abundances in certain parts of nature, the three models must then be considered not relevant to those parts of nature.

Now suppose our predictions, with or without threshold, are confirmed approximately. In order to separate the broken-stick from the other two models, we consider the total abundances on each date of observation. The broken-stick model predicts that the total abundances will be constant (within errors of counting or of measurement) from one date of observation to the next, since the abundance-limiting factor, energy input, has been held constant. The exponential model predicts that the sum of the n species' abundances will not be constant but will have a Γ distribution (or Pearson Type III distribution), that is (Feller, 1966, p. 10), will have density function

$$\frac{\lambda(\lambda x)^{n-1}e^{-\lambda x}}{(n-1)!},$$

where λ is again the scale parameter of the individual exponential functions. The balls and boxes model predicts that the sum of the abundances will be distributed as the sum of independent geometric variables with n different means; the moments of this distribution can be derived from the product of the probability generating functions of the individual geometric variables.

But the main thing to look at is whether the total abundances are constant. If not, set aside the broken-stick model, or at least the belief that energy input is the critical factor in it. If so, the broken-stick model is strikingly confirmed.

To discriminate the balls and boxes model from the exponential model, rather than trying to determine the detailed distribution of the summed abundances, it would probably be more useful to examine the detailed natural history of the species and to decide whether their "niches," however defined operationally, were totally unconstrained in relation to each other (exponential model) or could reasonably be described as consisting of component "subniches," the number of which differed between any two species (balls and boxes models).

Field observations which might discriminate the broken-stick from the exponential model have been suggested by Robert H. MacArthur (personal communication, July 19, 1967). He suggests visiting an island on which, in the taxon being considered, half the mainland species are missing. The broken-stick model assumes that abundances are limited only by the critical factor, which would be presumed to be the same on island and mainland; hence the total number of individuals (per unit area) would be predicted to be the same on island and mainland. The exponential model, as proposed above, assumes that the location and scale parameters are constant over time and across species, for a community with a given set of species; the model applies only to a fixed set of species and hence is consistent with either greater or lesser total abundances when the set of species is changed. On the additional, stronger assumption that the location and scale parameters would be constant even if the species composition of the community were changed, the exponential model would predict, all else being equal, half as many total individuals in the taxon (per unit area) on the island as on the mainland.

Generalizing MacArthur's idea, we could compare any two areas such that, within the taxon studied, the species in one area would be a strict subset of the species in the other area and all other factors which might affect abundance would be equal. The total abundances per unit area could then be compared as before.

One drawback of this approach is that it may be difficult to show convincingly that any two distinct areas with different numbers of species are ecologically similar enough, particularly in their supply of the critical factor, to constitute a "controlled experiment." It is thus not too surprising that a field study which attempted to carry out MacArthur's proposal did not succeed in differentiating decisively between the broken-stick and exponential models.

Crowell (1961, 1962) compared resident land-bird populations on Bermuda and those in similar habitats on the North American mainland. Only a few of the species found on the mainland are found in Bermuda. Crowell does not give the relative abundances of all land-bird species on Bermuda, so it is not possible to test the predictions (3). Hence it is not possible to decide whether his other findings are relevant to a comparative test of the three models considered here.

But, on the assumption that his other results are relevant, Crowell's nicely substantiated finding of "considerable overlap" in species' methods and loci of feeding favors the exponential or balls and boxes models.

Crowell (1961) also found that the 10 common resident land birds of Bermuda "achieve total populations at least as great as those of all species on comparable continental communities. . . . Absence of competition in Bermuda has allowed the few species present to attain far greater densities than they do in North America." If the energy available to the birds were exactly the same in both places, the broken-stick model would predict the same total (all species) populations in both places, not greater populations on Bermuda. So, according to this model, differences in energy supply must be assumed. The exponential model plus the strong assumption of constant parameters would predict that the species found in comparable North American and Bermuda communities would have the same absolute average abundances, respectively, in both, a prediction contradicted by Crowell's findings. But a possible difference in available energy, and the presence in Bermuda of species absent from the mainland, makes the *ceteris paribus* assumption implausible.

The strong assumption of constant parameters for the exponential distribution, regardless of the composition of the community, does not seem especially compelling to me, even in a comparison of two environments identical except for the species present. The assumption is not especially attractive because the presence or absence of one species alters the situation of the others, even if "all else" is equal. I would prefer to see a critical comparison of the three models based on long-term field studies of a number of communities in each of which the species composition was constant over time and for each of which the energy input from year to year, say, was constant. Observations like those suggested above for an experimental community could then be made and analyzed.

Most likely, convincing conclusions can be drawn only from a combination of field and experimental tests.

CONCLUSION

Besides the three models just reviewed, it is probable that still others, with other biological interpretations, would lead to the same predictions of average relative species' abundances (3), and it is certain that a vast variety of models would lead to predictions approximating (3). Hence, confirmation of (3) alone does not confirm the assumptions of one model leading to (3) against the assumptions of another.

The moral is hardly novel: Once a set of assumptions (a model) has been found which accounts for certain observed data, it is necessary to ask what other explanations are available, to determine how these other explanations differ in their observable implications, and to search for data which could discriminate among the explanations. Means of discriminating among the broken-stick, balls and boxes, and exponential models have been suggested.

The arguments and conclusions of this paper apply with at least equal strength to the testing of these models in economics (Cohen, 1966).

ACKNOWLEDGMENT

I am grateful to R. H. MacArthur and L. B. Slobodkin for highly useful criticisms of previous drafts.

APPENDIX

Proof of Equation (3)

Let $S = X_1 + \ldots + X_n$ and $R_{(i)}$ be the ith order statistic of the n values R_i defined in (2) above. Then, since the X_i have independent Γ distributions and since $R_{(i)}$ is a scale-free function of the X_i, the $R_{(i)}$ are independent of S (Kendall and Stuart, 1958, p. 368, citing Pitman, 1937, Cambridge Phil. Soc. Proc. 33:212). Hence $E(R_{(i)})E(S) = E(R_{(i)}S) = E(X_{(i)})$. But

$$E(X_{(i)}) = \frac{1}{\lambda} \sum_{j=1}^{i} (n - j + 1)^{-1}$$

and $E(S) = n/\lambda$ (Sarhan and Greenberg, 1962, p. 343); (3) follows immediately.

An even shorter proof follows from the fact that R_i has the same distribution as the ith interval, from left to right, of a unit line randomly divided as in the broken-stick model (Feller, 1966, p. 75). Since the distributions are identical, the expected values of the order statistics must be also, hence (3).

LITERATURE CITED

Cohen, J. E. 1966. A model of simple competition. Harvard Univ. Press, Cambridge, Mass. 138 p.

Crowell, K. 1961. The effects of reduced competition in birds. Nat. Acad. Sci., Proc. 47: 240–243.

———. 1962. Reduced interspecific competition among the birds of Bermuda. Ecology 43: 75–88.

Feller, W. 1966. An introduction to probability theory and its applications. Vol. 2. Wiley, New York. 626 p.

Kendall, M. G., and A. Stuart. 1958. The advanced theory of statistics. Vol. 1: Distribution theory. 2d ed. Hafner, New York. 433 p.

———. 1961. The advanced theory of statistics. Vol. 2: Statistical inference and statistical relationship. Hafner, New York. 676 p.

King, C. E. 1964. Relative abundance of species and MacArthur's model. Ecology 45: 716–727.

MacArthur, R. H. 1957. On the relative abundance of bird species. Nat. Acad. Sci., Proc. 43: 293–295.

———. 1960. On the relative abundance of species. Amer. Natur. 94: 25–36.

Preston, F. W. 1948. The commonness, and rarity, of species. Ecology 29:254–283.

Sarhan, A. E., and B. G. Greenberg. 1962. Contributions to order statistics. Wiley, New York. 482 p.

Slobodkin, L. B. 1961. Growth and regulation of animal populations. Holt, Rinehart & Winston, New York. 184 p.

Whittaker, R. H. 1965. Dominance and diversity in land plant communities. Science 147(3655): 250–260.