

Original article

The stochastic community and the logistic-J distribution

A.K. Dewdney *

*Department of Biology, University of Western Ontario, London, Ont., Canada N6A 5B7,
Department of Computer Science, University of Waterloo, Waterloo, Ont., Canada N2L 3G1*

Received 18 June 2002; accepted 21 December 2002

Abstract

A new formal model called the multispecies logistical (MSL) system produces species/abundance distributions that are compared statistically with those found in natural communities. The system, which is able to handle thousands of individuals from hundreds of species, iteratively selects random pairs of individuals and transfers a unit of biomass (or energy) between the respective species. Several elaborations of the model, including those with trophic compartments, appear to produce the same distribution. The theoretical distribution underlying the MSL system is a hyperbolic section, here called the logistic-J distribution. In the study reported here, the logistic-J distribution has been fitted to the species-abundance histograms of 125 randomly selected taxocoenoses. Since the overall chi square score of the logistic-J achieved near-optimality in this study, it cannot be distinguished statistically from the J-curves observed by field biologists. For comparison, the log-series distribution was given the same test and scored significantly higher (more poorly) than the mean logistic-J score. If there is a single, major distribution underlying natural communities, it is not the log-series distribution. Nor, owing to a mathematical error in its formulation, can it be the lognormal distribution. In the MSL system each species follows a “stochastic orbit” about the mean abundance producing, in consequence, the logistic-J distribution. Such orbits are produced by any system in which the probabilities of reproduction and death are approximately equal. Accordingly, the “stochastic communities hypothesis” is proposed here as the overall mechanism governing abundances in all natural communities. It is not a single mechanism, per se, but the net effect of all environmental influences.

© 2003 Éditions scientifiques et médicales Elsevier SAS. All rights reserved.

Keywords: Species-abundance distribution; Dynamical system; Individual-based model; Taxocoenoses; Stochastic populations; Stochastic communities

1. Introduction

At the present time several theoretical distributions have been proposed to describe the abundances of species in natural communities. They include the log-series of Fisher et al. (1943), the lognormal of Preston (1948), the broken-stick model of MacArthur (1957), as well as the negative binomial of Pielou (1975), and a few others described in Magurran (1988). Of these distributions, only the first two could be described as serious proposals. The remainder may have a slight theoretical interest but appear to be used less often in the field. Recently, I have proposed a new species abundance pattern, the logistic-J distribution (Dewdney, 1997, 1998, 2000).

On the plethora of proposed theoretical distributions (up to 1988), Magurran (1988) writes:

“While it is possible that someone may come up with a model which will revolutionize our understanding of species abundance relationships, at present it seems best to agree with Gray (1988) who concludes that ‘the search for yet more models is unlikely to give any insights into factors structuring biological assemblages’ ”.

The quote is discouraging for a very good reason. Not all of these distributions can be simultaneously valid. Indeed, Magurran seems to have allowed for the possibility that none of the distributions extant at the time were valid.

Fig. 1 shows the two most widely accepted distributions currently in use. In the log-series distribution the heights of the columns represent the expected numbers of species having abundances 1, 2, 3, and so on. The lognormal distribution has the same interpretation, but in a continuous sense. The expected number of species of abundance k would be found as the integral over one unit of the abundance axis at position k .

* Corresponding author.

E-mail address: akd@uwo.ca (A.K. Dewdney).

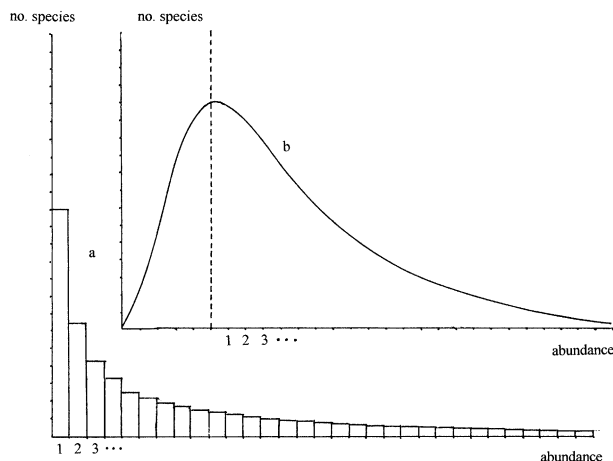


Fig. 1. The discrete log-series distribution (a) and the continuous lognormal distribution (b).

If someone handed us a theoretical distribution for natural communities, how would we test it? Would we not attempt a goodness-of-fit test against abundance distributions observed in the field? So far, with only one exception (that of Hughes, 1984, 1986), known to the author, no proposed distributions has been tested on more than a relative handful of such field distributions and in no case was a goodness-of-fit or other statistical test employed. Fisher et al. (1943) compared the log-series distribution with just two taxocoenoses, the comparison being essentially visual. Preston (1948) presented only a mild improvement, comparing his lognormal distribution visually with seven data sets and conducting no statistical tests. Even Hughes, with an appropriately large set of empirical distributions, did no direct statistical testing.

Standard goodness-of-fit tests such as the Kolmogoroff–Smirnov or the chi square (Hays and Winkler, 1971) are used primarily to reject fits. Nevertheless, attempts to reject hypotheses form a foundation of all scientific methodology. When it comes to “acceptance”, or using such tests to ascertain the presence of an underlying distribution, it is possible to use a goodness-of-fit test in a more positive mode. Such testing requires not a mere handful of field distributions, but a great many.

The present study focuses on the “community”, meaning a set of species living in a restricted volume of space over a restricted period of time. In this context, one may speak of a “species” and its “population”, interchangeably. It is also useful to invoke the “supercommunity”, the set of *all* species (populations) occupying that space and time. The dynamic that drives population levels in a given species may be divided into two kinds of interaction: with the physical environment and with the supercommunity.

The results of an extensive test of the logistic-J and the log-series distributions against 125 field surveys drawn from 123 randomly selected papers are reported here. An important finding that has emerged from the metastudy is that species abundance curves observed in the field have a J-shape

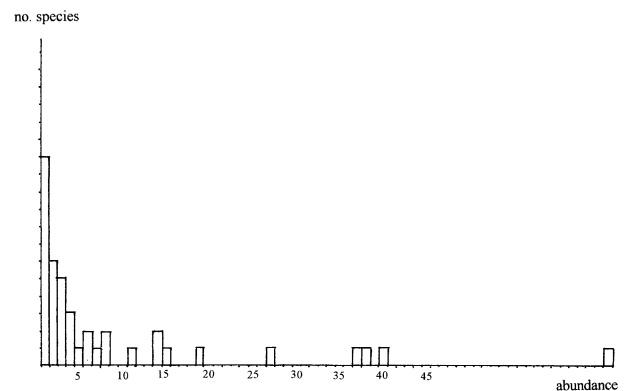


Fig. 2. A typical species abundance frequency distribution.

with probability better than 0.99. The example shown (Fig. 2) of a study by Garcia-Raso (1990) is typical. Almost every empirical species abundance histogram examined in the course of the metastudy (nearly 200 randomly selected papers) showed the greatest number of species at the lowest abundances, thereafter tapering to near-zero at the high abundance tail of the distribution. The same thing is true, of course, of the 123 papers admitted to the study under the acceptance criteria (see Section 2). There were perhaps three exceptions, all involving fewer than 30 species (the minimum richness allowed by the metastudy).

Field biologists sometimes call this general shape the “J-curve” (a generic term that does not seem to appear in the literature), resembling a backward letter J. Williams (1964) calls it the “hollow curve”. The distribution is somewhat ragged, as field data invariably are. Whatever the underlying distribution may be, the goodness of fit will vary for such distributions, according to the dictates of normal statistical variation.

Another important finding, this one theoretical, is that species abundance curves that result from sampling echo the shape of the distribution of the particular community being sampled (Dewdney, 1998). There has been considerable confusion on this point since 1948 when E.F. Preston launched his lognormal distribution with its accompanying concept of a “veil line”, now known to be based on a misunderstanding of the sampling process itself (Dewdney, 1998). There is no veil line, per se. There is what might be called a “veil curve” that represents the numbers of species from each abundance class in the community that are apt to be missing from a sample of that community. Sample distributions reflect those of the community from which they are drawn, differing only in a change of parameter value. In view of this theorem, the fact that nothing like an untruncated lognormal distribution ever shows up in species-abundance histograms pretty well obviates the need for testing. An additional remark might be appropriate in this context: when a species-abundance distribution is transformed logarithmically into “octaves” (Preston, 1948), most of the information inhering in the numbers

is destroyed and, moreover, the J-shape is transformed into a (roughly) bell-shaped curve. This is precisely what happens when the logistic-J distribution is subjected to this treatment, but the resulting curve is *not* the normal distribution.

Another problem with both the log-series and lognormal distributions is that neither has an accompanying biologically realistic model from which it was derived. Both appear to be essentially guesses, bolstered by some after-the-fact observations (Pielou, 1975; Sugihara, 1980) that support them only weakly.

The present paper examines the logistic-J distribution (Dewdney, 1997, 1998, 2000), its derivation from a dynamical system, its close statistical fit to field data, and a surprisingly simple mechanism underlying the dynamical system. All the three components form the pillars of a unified theory that cannot, in the face of evidence presented, be rejected as a candidate for the putative “model that will revolutionize our understanding” (Gray, 1988). This paper extends and completes the metastudy reported by Dewdney (2000) in which (the first) 50 randomly selected taxocoenoses were compared with the logistic-J distribution. It also includes a more detailed analysis of underlying processes and their probable presence in nature.

2. Materials and methods

2.1. The multispecies logistical system

The logistic-J distribution arises from a dynamical, individual-based model of species interactions called the multispecies logistical (MSL) system. Like the logistic difference system of May (1976), the MSL system is dynamical but accommodates not just two species, but an arbitrarily high number. Unlike May’s system, the MSL system is not continuous, but individual-based (Judson, 1994), each individual being represented by a numerical token that can be interpreted variously as an individual or as a packet of biomass/energy, depending on the level of interpretation. The individuals are distributed among n “species”.

In the MSL system, time proceeds in discrete steps, as in standard critical event simulations (Emshoff and Sisson, 1970). At each step, two individuals are selected at random. If the first individual belongs to species j and the second belongs to species k , the first individual is transferred, in effect, from species k to species j . Consequently, after the step, species j has one more individual than it had before, while species k has one less. The transfer of individuals from one species to another is interpreted as trophism, whether predation by animals or light competition by plants. Energy contained in (or destined for) one individual ends up in another which, in consequence, reproduces. Since the system is wide open and because the transfer involves whole individuals, the version of the MSL described below lacks biological realism. Yet it is merely the “core” system and more

sophisticated versions exist which do not suffer this shortcoming.

Since the MSL system neither creates nor destroys individuals, it could be called “conservative” in the physical sense. The word “logistic”, which appears both in the name of the system and in the name of the distribution it produces, refers to the essentially logistic consequence of this conservation. A species that becomes especially abundant in the system finds it increasingly difficult to become more abundant, owing to diminishing resources.

I cannot recommend too strongly that those who wish to understand the claims made in this article should run some version of the program. It takes little programming skill to write a program based on the algorithmic outline in the author’s website. Those without programming experience may copy the program listed there (see the end of this article). I implemented the MSL system as a Turbo Pascal (6.0) program running on a 486 computer. Besides the foregoing essential features, the program has a display loop, an extinction switch, and a basic statistical reporting package. The system takes no more than $100N$ steps to reach equilibrium from an initial spike distribution, N being the total number of individual organisms. This program and three of its variants have been run literally hundreds of times under a very wide variety of conditions and with many different experimental aims.

The program displays a species abundance histogram every 100 iterations of the basic pair selection procedure. The extinction switch operates as follows. With the switch on, a species that reaches abundance 1 will, if selected by the system as “donor”, be extirpated from the system after the predation. With the extinction switch off, any exchange involving such a species as donor is canceled. Typically, the switch is turned off when operating the MSL at population levels that are typical of samples. Otherwise it is turned on.

The MSL system was analyzed mathematically to determine the underlying distribution at equilibrium (Dewdney, 2001). The resulting probability density function (pdf) may be written:

$$f(x) = c(1/x - \delta); \quad \varepsilon \leq x \leq \Delta, \\ = 0; \quad x \geq \Delta$$

The constants δ ($=1/\Delta$) and ε are parameters of the distribution and c is a constant of integration that gives the area under the curve of the pdf a value of unity. The constant c is a function of ε and δ and is not a statistical parameter. The parameters ε and Δ are called the inner and outer limits of the distribution, respectively. The outer limit represents the non-zero extent of the distribution function, acting as an average maximum abundance for the MSL system, when in operation. It could be called a “soft limit” in this sense. The inner limit does not have such a simple interpretation but its inverse is proportional to the predicted number of species in the lowest abundance category. I distinguish this formulation of the logistic-J distribution from what follows by calling it the “special” logistic-J distribution.

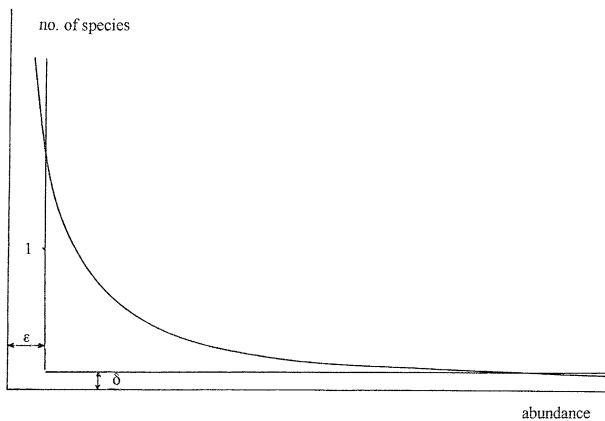


Fig. 3. The pdf of the general logistic-J distribution.

A nearly equivalent formulation is the “general” logistic-J distribution (pdf), which involves a slight change in the variable x :

$$f(x) = c(1/x' - \delta); \quad 0 \leq x' \leq \Delta, \\ = 0; \quad x \geq \Delta,$$

where $x' = x + \varepsilon$ and c is defined as before. Both the special and general forms of the logistic-J pdf are sections of the hyperbola (Fig. 3).

The parameters ε and δ determine what portion of the hyperbola is to be used for the distribution. The values implicit in the figure would be typical of samples of a community. The community itself would be represented by a relatively larger value of ε (reducing the size of the initial peak) and a smaller one for δ (extending the maximum abundance).

The special distribution is easier to work with and fits field data as well as the general distribution. The latter, however, is preferred as the best descriptor of abundances in communities where, typically, the numbers are larger by an order of magnitude or more. The logistic-J distribution is fitted to the field data by using two empirical parameter values (average abundance and the number of species in the lowest abundance category) as the basis of estimates for ε and δ . A pair of transfer equations produces the estimates. With these values in hand, one may write the distribution function for the special logistic-J distribution, for example, as

$$F(x) = Rc(1/x - \delta), \quad \varepsilon \leq x \leq \Delta$$

having the value zero beyond Δ . The quantity R represents the richness (number of species) in the field sample.

A significant version of the MSL system illustrates that the unlimited trophism of the core model, while biologically unrealistic, is something of a red herring. In a second version of the system, species are partitioned into trophic compartments: primary producers (e.g. plants), herbivores, carnivores, and recyclers (e.g. fungi and bacteria). This version allows only interactions appropriate to trophic status. Thus an herbivore may “eat” only primary producers, a carnivore may only eat only herbivores or other carnivores. Recyclers may eat anything. To keep the number of individuals con-

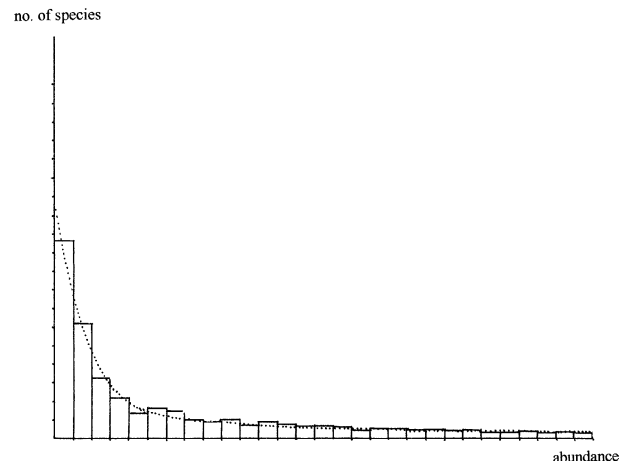


Fig. 4. Average species abundance histogram for MSL herbivores.

stant, every time a primary producer receives enough energy to reproduce (achieved by simply adding another individual to that species), the system removes a recycler (at random) from the system. In this variant, transactions flow both within and through compartments. Although it has not been confirmed by direct experiments, it may be suspected that even when the number of individuals is allowed to drift or fluctuate, the whole system would respond by behaving exactly as it did before, coming to a new equilibrium either sooner or later, depending on the degree of the change.

The histogram (Fig. 4) represents the outcome of a typical experimental run of 1000 display cycles (100,000 basic iterations). The individual bar heights were computed by the MSL program's statistical package as the average number of species in each abundance category for herbivores. Some 50 species had an average of 15.0 individuals and the category widths were all unity, representing a single abundance.

I have superimposed the (general) logistic-J distribution curve on these data to illustrate the very close fit that is typical of such experiments. Such illustrations do not prove that the distribution produced by this version of the MSL is logistic-J, but it seems rather likely, given that species undergo very much the same sort of fluctuations, both theoretically and in the computer model, when compared with the core MSL system.

A third version of the MSL system, in which fractional trophism was allowed, also produced the logistic-J distribution. In all these versions, a mathematical consequence of the basic step in which individuals are transferred is that each species is as likely to increase as it is to decrease. In a fourth version, these probabilities were manipulated directly, being allowed to become unequal for varying periods of time and in a manner that guaranteed a long-term average equality. Again, the logistic-J distribution appears to have emerged. The core model is what computer modelers call “detail hungry”. It is robust and apparently independent of specific details of trophism.

2.2. Testing the distribution on field data

For the purposes of this paper, a “taxocoenosis” is a journal article in which the authors have carried out a systematic sampling procedure for species in a specific taxon or ecological niche within a limited geographic area. Examples included waterbirds along a stretch of northern California coastline, ciliates in a pond in Spain, epiphytes in a plot of Costa Rican rain forest, fungi on fir needles in a Japanese forest, and so on (see the metastudy bibliography in the website referred to at the end of this paper). The usual purpose of such taxocoenoses is sometimes to assess the “biodiversity” of an area, sometimes to compare disturbed with undisturbed areas, sometimes to monitor changes in community structure over time. Most of the groups surveyed in these articles were described by the authors as “communities”. I have assumed, in using the data from these studies, that they are fairly taken samples of their respective communities, with no species, even ones of low abundance, omitted. The actual numbers reported are subject to various sources of variation, including (a) numerical variation in the communities themselves, (b) variation in the sample (luck of the draw), (c) and variation in capture rates for some species in certain communities such as bats or butterflies. With the possible exception of the third factor, standard statistical techniques are adequate to the descriptive task. It is not clear how much one should worry about differential capture rates. It is assumed in this paper that, while this source of variation may bias the counts for certain species in one direction or another, it does so consistently without affecting the stochastic nature of its variation, the essential focus of this paper. In any event, only certain groups of organisms appear to produce the third source of variation.

In the process of examining approximately 200 taxocoenoses at random from the extant literature, I rejected any that failed to meet the following three criteria: They had to (a) report on at least 30 species; (b) not exclude species that were rare or seen infrequently; (c) provide actual counts, percentages, or densities based on counts. I selected a total of 123 papers from which 125 studies emerged, preparing tables of data in which the number of species in each abundance category was added up and recorded. The abundance categories were grouped according to normal curve-fitting practice (Emshoff and Sisson, 1970). In over half the cases, no grouping was required. In about a quarter, abundances were grouped in twos, the remainder being grouped in threes, fours, a few fives, a 10 and one 15. Grouping simplifies the data to an extent but neither enhances nor degrades, on average, the fit one would obtain if no grouping were done at all.

The empirical histograms were fitted, according to the methods described above, to the logistic-J distribution(s). They were also fitted, according to the methods described in Magurran (1988), to the log-series distribution. Thus, each taxocoenosis was compared with each of three candidate theoretical distributions, the logistic-J (special and general)

and the log-series. Each of the resulting 375 matches was assessed by the chi square test (Hays and Winkler, 1971) using custom-written software that computed values for all theoretical distributions according to their respective formulae. The raw scores for the three sets of tests were then normalized, via their corresponding *P*-values, to scores at 10 degrees of freedom. This made it possible to carry out two distinct tests on the chi square scores. First, all three sets of normalized scores were averaged and the results were interpreted according to standard chi square theory. A second test, involving confidence intervals for paired data (Wonnacott and Wonnacott, 1982), made it possible to compare the general logistic-J with the log-series in more telling fashion. This test is more discriminating than ordinary interval tests, exploiting as it does the natural relationship between two test scores for the same taxocoenosis.

A final test involved the logistic-J distributions alone. Since only information about the mean abundance and low-category richness entered the process of determining the parameter values ε and δ , the value of $\Delta (= 1/\delta)$ produced by the solution process amounts to a prediction of the average maximum abundance. Thus, while the Δ -value would be only approximately correct (at best) for individual taxocoenoses, the general theory (inspired by the MSL system) predicts that the average Δ -values should be close to the average maximum abundance.

3. Results

Of the four kingdoms covered, 82 taxocoenoses involved communities of Animalia, 28 of Plantae, eight of Fungi and seven of Protista. The taxocoenoses were conducted in every major climatic zone: boreal, temperate, subtropical, and tropical. Aquatic studies included freshwater and marine (both shallow and abyssal). Terrestrial taxocoenoses were conducted on every continent and included wet and arid climates, savanna, forest, desert, montane, and coastal biomes. The literature selected for the metastudy is cited in the website given at the end of text. The results of the study are summarized in a table on the same website.

Table 1 shows the averages and standard deviations for both versions of the logistic-J distribution as well as for the log-series distribution. The chi square standard deviations are derived from the theoretical mean under the null hypothesis, namely $\sqrt{20.0}$, where 20.0 is the expected variance for 10 degrees of freedom.

The two versions of the logistic-J distribution had remarkably close scores in spite of the fact that their scores on individual tests frequently differed. Each version of the dis-

Table 1
Averages and standard deviations of test scores for the three distributions

Distribution	Average	S.D.	χ^2 S.D.
Special logistic-J	10.84	5.33	4.47
General logistic-J	10.81	5.51	4.47
Log-series	13.56	9.40	4.47

tribution came within one-fifth of a standard deviation of the optimal score of 10, yielding strong evidence in favor of the null hypothesis in this context. The score for the log-series distribution is nearly a full standard deviation away from 10.0, a distinctly poorer performance.

The difference in average scores between the logistic-J distribution and the log-series is no accident, according to the results of a paired difference test (Wonnacott and Wonnacott, 1982). In this test, the differences between the former and latter distributions were computed for each taxocoenosis, producing the same effect as controlling for all variables except the scores of their respective distributions. This test, carried out at a significance level of 99.9%, produced the following confidence interval for the mean difference of 2.72 between scores for the special logistic-J and those for the log-series distribution over all 125 taxocoenoses:

$$[0.640, 4.801](t - \text{value } 0.001)$$

The fact that the interval is bounded away from 0 means that with probability 0.001 we would err in asserting that the population of test scores for the log-series distribution is different from that of the logistic-J distribution. In other words, there is a statistically discernible difference between the two populations, as revealed by their test scores. Since the logistic-J distribution scored lower (i.e., better) on the 125 taxocoenoses, we may assert that it is, with very high probability the superior descriptor of community abundances in the most general sense.

The possibility remains, if this metastudy were extended to, say, 1000 taxocoenoses that the logistic-J distribution would still come in somewhat over 10. This might in fact be the case, given that in all taxocoenoses there are actually two sources of variation. The sampling process is one and the other resides in the communities themselves, already being perturbed versions of whatever theoretical distribution might underlie them. As well, other sampling anomalies such as differential capture play a role in the slight overage. In such a situation one would expect both a somewhat higher standard deviation and mean, as reflected in the table of Fig. 5.

Among the most interesting results to emerge from the metastudy was the distribution of predicted outer limits in relation to the maximum abundances in the taxocoenoses. If the interpretation of delta (Δ) as the average maximum abundance is correct, we would expect that the maximum abundances (Δ') would sometimes exceed and sometimes fall short of the average maximum abundances predicted by the logistic-J distribution. We would also expect the average value of the ratios Δ'/Δ (expressed as percentages) to be close to the 100 mark. The actual average values for this ratio

emerging from the metastudy were surprisingly close: 100.4% for the special logistic-J and 107.8% for the general logistic-J.

4. Discussion

The metastudy presented here makes it clear that the logistic-J distribution outperforms the log-series distribution as a descriptor of abundances in natural communities. The test, carried out in exact conformity with the central interpretation of the chi square distribution, cannot be improved upon in any essential way as a discriminant between various proposals for the “right” abundance pattern for natural communities. To outperform the log-series distribution does not mean that the logistic-J distribution is destined to play such a role, only that the log-series is not. But having outperformed the log-series and having approached so closely to the optimum value, the empirical result nevertheless amounts to strong evidence of that possibility.

The results of the outer limit test, on the other hand, are somewhat astonishing, even to the proposer of the logistic-J distribution. It is highly unlikely that the accuracy of the predicted average maximum abundance is an accident. It can be concluded that the predictive power of the logistic-J distribution is a genuine implication of the underlying stochastic communities hypothesis (see next section) and the theoretical apparatus that emerges from it in a purely mathematical sense. It may now be asked whether the MSL system itself contains the essential elements of a general, pervasive mechanism that produces the logistic-J distribution in nature.

It turns out that there is yet another level of generalization in which details of trophism and reproduction are irrelevant. In a final model, species reproduce (adding an individual) and die (deleting an individual) without reference to any pair interaction at all. Instead, the system simply selects an individual at random and either deletes it or doubles it with equal probability. This system also produces the characteristic J-shaped curve, which is logically interpreted as the logistic-J function.

In the abstract, any system of species that (a) follows an equiprobable birth–death process and (b) imposes a limit on the total number of individuals (alternatively biomass), would appear to produce the logistic-J distribution. In such a process, each individual organism will have more or less equal probabilities of reproducing or dying (before it reproduces). Accordingly, each species also has this property, when it changes abundance, that the probabilities of increase and of decrease are more or less equal. By “more or less”, I mean that the two probabilities may fluctuate, whether in the long or short term, about the same value. The fourth version of the MSL described above, does precisely this. Both assumptions (a) and (b) are adopted by Hubbell (2001) in his analysis of the distribution that results from “zero-sum dynamical rules” and “ecological drift”. Unfortunately, Hubbell is unable to provide a closed formula for his distribution.

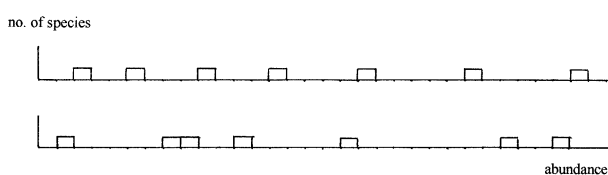


Fig. 5. Predicted (a) and actual (b) abundances in a community.

I hope to provide a more extensive analysis of Hubbell's work in a subsequent publication.

Condition (b) requires some clarification. A system that follows only condition (a) will produce a hyperbolic distribution, just as a system obeying both conditions will. However, the hyperbola so produced will tend to spread out with time, the species of largest abundance becoming steadily larger, without end. The "limit" of condition (b) prevents this from happening. In the MSL system or its variants, constant biomass is not the same thing as constant energy or a constant number of individuals. However, in the distinctly more general setting of the stochastic species hypothesis, only a "limit" is involved. A limit on any one of these factors is automatically a limit on the others, albeit a different one.

Interestingly, any of the first three versions of the MSL system can be interpreted as a purely physical model in one-dimensional particle dynamics. In this model, each particle represents a species and occupies, at any given moment, an integer point on the abundance axis. It "vibrates", in a statistical sense, moving either left or right with equal probability. As a conservative system, the physical model requires that every time one particle moves to the right, another must move to the left. The further out the abundance axis a particle travels, the more rapidly it vibrates. In general, each particle follows what might be called a "stochastic orbit" about the mean population size, occasionally exceeding it, but usually remaining below it. In such a system, the particles would tend to crowd toward the cool end, with only a few particles toward the hot end, vibrating rapidly. The system is *very* loosely coupled, having $n - 1$ degrees of freedom in the dynamical (not statistical) sense. The particles, moreover, would have the logistic-J distribution.

As I pointed out in the predecessor (Dewdney, 2000) of the present paper, populations do not change their abundances in a purely random manner. They only appear to. The combined effect of all influences on the abundance of a given species in a particular community cannot, in general, be known, nor can they be predicted in advance. In a previous paper Dewdney (2000) I listed, as an example, some factors that might prevent a plant from reproducing. The example was sparse, giving only about a dozen such factors, it being possible, with some determination, to list hundreds. In particular, weather and climate have a strong direct or indirect effect on many factors and weather appears to be essentially random, thanks to the presence of chaotic processes.

The central theme of this section is the "stochastic community hypothesis", which relies on the concept of probability: For all natural (non-anthropogenic) communities, the probabilities of increase (p) and decrease (q) peculiar to each component population remain in balance, on average. To the extent that this hypothesis is true, the presence of the logistic-J distribution follows as a deduction.

Apart from the statistical evidence supporting it, the hypothesis would be very difficult to establish. The probabilities p and q cannot be measured directly, although their effect can be observed partially in the short term (say over several

generations). The first counter-intuitive lesson of the MSL system is that p may equal q for a population, yet its abundance is surprisingly likely to drift, in the manner of a random walk, in one direction or another. In the absence of information concerning some obvious and dominating disturbance or influence, there is no way to decide, for a given population, whether its apparent long-term decline (or increase) is due to an imbalance between p and q or whether it is entirely "accidental".

At this point, the only support for the stochastic communities hypothesis outside of statistical testing comes from a general argument: if there is a long-term imbalance between p and q for a given species, its abundance will change over time and in the appropriate direction, with an average rapidity that is proportional to the degree of the imbalance. Depending on the direction of the imbalance, the species will ultimately either disappear from the community or "take over", driving many other species to extirpation. However, the stochastic communities hypothesis would also guarantee that such things will happen in nature, albeit rarely. The question is, therefore, whether the frequency of crashes or booms reflect what the logistic-J distribution would predict. This would be a useful direction for future research. In any event, because the birth of every organism must inevitably be accompanied by its death, the *a posteriori* values of p and q are practically equal for any organism that persists.

In the light of the empirical findings reported in the previous section and in the light of the foregoing remarks, it seems reasonable to make three claims:

- (a) J-curves are ubiquitous in (samples of) natural communities,
- (b) the totality of such J-curves appear to fit the logistic-J distribution very close to optimally in the chi square sense,
- (c) J-curves are expressions of the stochastic communities hypothesis.

Given the truth of propositions (a) and (b), proposition (c) would be the most parsimonious conclusion to draw, given that the stochastic communities hypothesis implies the logistic-J distribution. It has been shown that (a) is almost certainly true in that all the histograms in the metastudy had the shape of a J-curve. It has also been shown that the logistic-J distribution fits the data so closely to optimal that little "room" is left for any of the taxocoenoses to be exceptional. As far as independent support for proposition (c) is concerned, we can only observe that long-term averages for the probabilities p and q for a given population or species must be equal as long as the species persists in the community.

I have used the term "stochastic" in the knowledge that a body of research exploring this view of nature is already under way, although it focuses on species, rather than communities. The use of the term here, however, is broad enough to reconcile to a degree the "stochastic" and density-regulation schools of thought in theoretical ecology (Gaston, 1995; Ricklefs, 1990). A complex determinism may readily

produce stochastic effects that are indistinguishable from pure randomness.

Encountering the J-curve from the sample of a natural community, some biologists might react with horror at all the species apparently about to vanish. However, one must bear in mind the typically great difference between histograms of the sample and of the community from which it came. A glance at the prediction made by the logistic-J about natural communities, where population sizes are orders of magnitude higher than in samples of those communities, clarifies the issue.

Fig. 5a displays the idealized predictions of specific abundances made by the general logistic-J in the case of 50 species with a mean abundance of 15. Due to the relatively high abundances in communities, there is room in this figure for only the first seven species. One can see that there is no crowd of species at the low abundance end as one might find in a sample. Instead, the abundances increase in orderly fashion from three onwards. At first the progression is only slightly faster than geometric, but it increases much faster than geometric at the high abundance end. Each position in Fig. 5a indicates the average abundance for the corresponding species, taken to the nearest integer. Fig. 5b shows how the abundances in a particular community having these parameters might appear at one specific moment of time, as influenced by normal fluctuations in its component populations.

It must be stressed that the orderly appearance of “places” in the upper figure should not be interpreted as implying that each species has a certain, permanent position in the order. Instead, our theory predicts (and the MSL system illustrates) that species will regularly trade places as they pursue their stochastic orbits about the mean abundance. The theory has nothing to say about what species will have what abundances at what times, only that each abundance level will be “occupied” in a statistical sense, by one species or another. The situation is reminiscent of quantum theory (Ponomarev, 1993). When photons pass through a narrow aperture before striking a phosphor screen, we cannot predict where a given photon will manifest, only that the pattern produced by a great many such photons will follow a particular distribution that is characteristic of diffraction.

The present article presents an appropriate test of a proposed theoretical distribution of abundances in natural communities. As a technique it appears to be new in the literature. It is labor-intensive, to be sure, but may be crucial in resolving the problem mentioned by Magurran in Section 1 of this paper.

It has been demonstrated conclusively in a previous publication (Dewdney, 1998) that the lognormal distribution can no longer be used in its veiled or truncated form. It has been observed in the present paper that the untruncated lognormal never appears in natural abundance patterns.

Given these facts, it is hard to see what role, if any, the lognormal distribution has to play in theoretical ecology. The present paper also makes it clear that the log-series distribu-

tion falls short in the only meaningful test that can be made of its powers as a descriptor of natural abundance patterns.

Given the closeness of its average chi square score to the theoretical optimum, it would be extremely difficult to establish the superiority of any other theoretical distribution over the logistic-J by the means used in this paper. It must be added here that the logistic-J distribution also predicts maximum abundances with a somewhat eerie statistical precision.

Finally, I cannot refrain from remarking on one of the most interesting implications of the body of theory arising from the MSL system, the logistic-J distribution and the stochastic community hypothesis. As Williams (1964) had noticed by 1943, taxonomic tables also produce J-curves. For example, for North American birds, the largest number of genera have only one species, the next largest number have two species, each, and so on. The resulting histogram has a typical J-shape. The evolution and extinction of species within a genus may well follow the same general pattern as the birth and death of individuals within a species. Moreover, the logistic-J distribution implies that small populations will be rather common in communities. New mutations are, therefore, not only relatively common in small populations (being numerous), but such mutations are far more likely to spread through small populations than large ones. A major problem for evolutionary theory is the gene cohesion effect, as explained by Mayr (1970): “The real problem of speciation is not how differences are produced but rather what enables populations to escape from the cohesion of the gene complex and establish their independent identity. No one will comprehend how formidable this problem is who does not understand the power of the cohesive forces in a coadapted gene pool”.

Routinely, although by no means always, small populations become large again, sometimes bringing with them new genotypes that are different enough to be considered as a new phenotype, i.e., species. I am currently pursuing a comparison of taxonomic distributions with the logistic-J distribution.

Taxocoenosis bibliography, data, and test results will be found at www.csd.uwo.ca/~akd/ by clicking on “biodiversity research”.

Acknowledgements

I wish to thank the following colleagues for useful discussions on points of ecology or evolution: Bob Bailey, Paul Handford, Laszlo Orloci, and Gary Umphrey. I must also thank (former) graduate students in the Ecology and Evolution Program at my home institution: Michelle Bowman, Monica Hevelka, Dave Muzia, and Kellie White. I also wish to thank the referees for their patience and helpful advice.

References

- Dewdney, A.K., 1997. A dynamical model of abundance in natural communities. *COENOSIS* 12 (2–3), 67–76.
- Dewdney, A.K., 1998. A general theory of the sampling process with applications to the “veil line”. *Theor. Population Biol.* 54, 294–302.
- Dewdney, A.K., 2000. A dynamical model of communities and a new species-abundance distribution. *Biol. Bull.* 198, 152–165.
- Dewdney, A.K., 2001. The forest and the trees: romancing the J-curve. *The Mathematical Intelligencer* 23 (3), 27–34.
- Emshoff, J.R., Sisson, R.L., 1970. Design and Use of Computer Simulation Models. The MacMillan Company, New York 279 p.
- Fisher, R.A., Steven, A., Corbet, S.A., Williams, C.B., 1943. The relation between the number of species and the number of individuals in a random sample of an animal population. *J. Anim. Ecol.* 12, 42–58.
- Garcia-Raso, J.E., 1990. Study of a Crustacea Decapoda taxocenosis of *Poisidonia oceanica* beds from the southeast of Spain. *Mar. Ecol.* 11 (4), 309–326.
- Gaston, K.J., 1995. Introduction: the never-ending debate. In: Cappuccino, N., Price, P.W. (Eds.), *Population Dynamics: New approaches and Synthesis*. Academic Press, San Diego, pp. 1–18.
- Gray, J.S., 1988. Species abundance patterns. In: Gee, J.H.R., Giller, P.S. (Eds.), *Organizations of Communities: Past and Present*. Blackwell Scientific Publications, Oxford, pp. 53–67.
- Hays, W.L., Winkler, R.L., 1971. *Statistics: Probability, Inference, and Decision*. Holt, Rhinehart and Winston, New York 937 p.
- Hubbell, S.P., 2001. *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton University Press, Princeton NJ 375 p.
- Hughes, R.G., 1984. A model of the structure and dynamics of benthic marine invertebrate communities. *Mar. Ecol. Prog. Ser.* 15, 1–11.
- Hughes, R.G., 1986. Theories and models of species abundance. *Am. Nat.* 128, 879–899.
- Judson, O.P., 1994. The rise of the individual-based model in ecology. *Trends Ecol. Evol.* 9 (1), 9–14.
- MacArthur, R.H., 1957. On the relative abundance of bird species. *Proc. Nat. Acad. Sci. USA* 43, 293–295.
- Magurran, A.E., 1988. *Ecological Diversity and its Measurement*. Princeton University Press, Princeton, NJ 179 p.
- May, R.M., 1976. Simple mathematical models with very complicated dynamics. *Nature* 261, 459–467.
- Mayr, E., 1970. *Populations, Species, and Evolution*. The Belknap Press of Harvard University Press, Cambridge, MA 453 p.
- Pielou, E.C., 1975. *Ecological Diversity*. John Wiley and Sons, New York 165 p.
- Ponomarev, L.I., 1993. *The Quantum Dice*. Institute of Physics Publishing, Bristol 255 p.
- Preston, E.F., 1948. The commonness, and rarity of species. *Ecology* 29, 254–283.
- Ricklefs, R.E., 1990. *Ecology*. W.H. Freeman, New York 896 p.
- Sugihara, G., 1980. Minimal community structure: an explanation of species abundance patterns. *Am. Nat.* 116, 770–787.
- Williams, C.B., 1964. *Patterns in the Balance of Nature*. Academic Press, London 321 p.
- Wonnacott, R.J., Wonnacott, T.T., 1982. *Statistics. Discovering its Power*. John Wiley and Sons, New York 378 p.