Evolutionary Ecology Research, 2014, **16**: 315–322

An assembly rule could explain the canonical property of the lognormal species-abundance distribution

Julián Simón López-Villalta

Departamento Biología y Geología, IES Pedro Simón Abril, Alcaraz, Albacete, Spain

ABSTRACT

Background: The lognormal distribution of species abundances is a classic model of ecology whose canonical property consists of the fact that the individuals curve tends to peak at the rightmost species-abundance class. Currently ecology lacks any compelling biological or mathematical explanation of the canonical property.

Goal: Propose a hypothesis for the canonical property by comparing the results of a maximum entropy niche-assembly model to the patterns obtained from the canonical relationship in empirical data.

Data: Published data of the numbers of species in lognormal communities, and their standard deviations.

Method: Let a niche dimension be any independent characteristic of the habitat that can be subdivided by species to reduce competition. Obtain relative species abundances as the product of randomly assigned species occupancies along multiple niche dimensions. Plot the standard deviation of the resulting lognormal-like species histogram for different niche dimensions and numbers of species. Estimate the number of species added with each niche dimension by interpolation along the line representing the canonical property and the regression line of empirical data.

Results: Randomly assigning species occupancies along multiple niche dimensions produces lognormal-like distributions. Each added niche dimension roughly doubles the number of species in the community.

Conclusions: The canonical property may originate from the doubling of the number of species with each added niche dimension. This assembly rule is supported by analyses on the relationship between bird species and habitat complexity.

Keywords: assembly rules, biodiversity, community ecology, macroecology, niche assembly, resource partitioning.

INTRODUCTION

A neglected yet interesting ecological question concerns the basis of the canonical property of the lognormal distribution of species abundances (Preston, 1962; Rosenzweig, 1995; May, 1999). This

Correspondence: J.S. López-Villalta, Departamento Biología y Geología, IES Pedro Simón Abril, Paseo de San Francisco 89, 02300 Alcaraz, Albacete, Spain. e-mail: jsimon@edu.jccm.es

classic pattern appears when local species abundances are plotted along doubling abundance classes (octaves, i.e. $\log_2$ bins). The histogram of numbers of species usually resembles a Gaussian curve. The plot of the total number of individuals along the same octaves usually resembles the left half of a Gaussian curve that reaches its peak at the rightmost octave. The canonical property of the plot refers to the relationship of the curves of species and individuals (Fig. 1). We know no mathematical reason for this relationship (Sugihara, 1980; May, 1999). Sugihara (1980) claimed that a sequential niche partitioning of 3:1 results in a canonical lognormal, but Nee *et al.* (1991) showed that not to be the case.

The canonical lognormal was widely known in theoretical ecology for decades, probably for two reasons. First, it provided a reasonably good fit to data from intensely sampled communities. And second, Preston (1962) used it to derive a realistic species–area relationship. However, Leitner and Rosenzweig (1997) noted a critical flaw in Preston's derivation. In addition, Williamson and Gaston (2005) rule out the canonical lognormal as a valid model. However, it has never been convincingly falsified, as have many other ecological models of species abundance (McGill *et al.*, 2007).

In spite of criticisms, the lognormal adequately models the abundances of core species in local communities (Magurran and Henderson, 2003; Ulrich and Ollik, 2004; Ulrich and Zalewski, 2006; Dolan *et al.*, 2009). This ability eliminates one of the main objections to the lognormal, namely that it predicts too few rare species. Yes, the number of rare species observed in nature does exceed the predictions of the lognormal. But the more common species of a community do tend to follow the lognormal curve.

No matter how well it fits data, the canonical property remains a mystery. It is astonishing that no possible explanation for it has been published since that proposed by Sugihara (1980). Here, I examine how the canonical property might arise by comparing the predictions from a maximum entropy model of niche partitioning with expectations obtained from the canonical condition and empirical data.

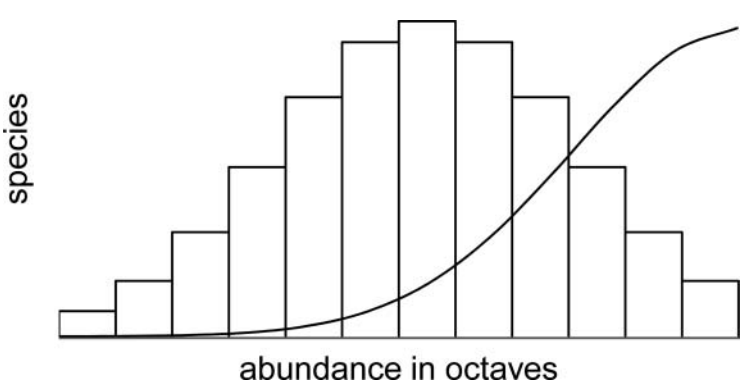

Fig. 1. An example of a canonical lognormal distribution. The x-axis is abundance, plotted in octaves, or $\log_2$ units, which double in magnitude from bin to bin. The columns represent the number of species. (Abundances of species at a bin boundary are counted in the corresponding two octaves.) The curve shows the number of individuals, which is often obtained as the product of the number of species and the mean abundance at each octave. The individuals curve is the left half of a Gaussian distribution that peaks at the rightmost octave. This relationship between abundance and species distributions defines the canonical property.

MODEL DESCRIPTION AND RESULTS

I programmed a computer to build a set of species and their abundances based on a model with three assumptions:

1. Local species abundances, N_i, are proportional to the realized niche breadth of a species, B_i. The realized niche is the subset of the species niche that is available in the local patch (*sensu* Hutchinson, 1957). Thus,

$$N_i = kB_i.$$

   This relationship is not intended to be more than the abstraction of a statistical trend. However, there is support for a positive relationship between niche breadth and abundance in general (Sugihara *et al.*, 2003).
2. The realized niche breadth of each species is the product of real niche occupancies, o_d, of the species along each of the d dimensions of the realized niche hypervolume (*sensu* Hutchinson, 1957):

$$N_i = k\prod_1^d o_d$$

   The occupancy o_d is a continuous variable which might, in principle, have different distributions. Niche dimensions are assumed to be mutually independent.
3. The real niche occupancy of each species in each of the d niche dimensions is obtained by sampling a random number from a uniform distribution ranging from 0 to 1. A uniform distribution implies a maximum entropy approach because it is the distribution with minimal structure. The scaling from 0 to 1 for each niche dimension means each dimension has the potential to contribute the same to realized niche breadth, and thus to species abundance. It also implies a uniform distribution of species abundances if the niche hyperspace has only a single dimension. This pattern is suggested by Whittaker's (1956) finding that abundance peaks of species can be found at random along an ecological gradient.

Computer simulations running this model provide different relative distributions of species abundance depending on the number of niche dimensions (d), the number of species considered (S), and the total number of individuals in the community (N). As expected from the central limit theorem, relative species abundance gradually approaches lognormality as more niche dimensions are added to the simulation (Fig. 2). The distribution is right-skewed even up to 11 niche dimensions.

Figure 3 shows the relationship between the estimated standard deviation of the abundance histogram (in octaves) and the number of species in the community. The figure shows the pattern of standard deviation versus number of species that one expects from the canonical property based on its most straightforward definition (see Sugihara, 1980), i.e. the octave that contains the peak of the individuals curve (R_N) divided by the rightmost octave ($R_{\mathbf{max}}$) is equal to one ($\gamma = R_N/R_{\mathbf{max}} = 1$).

In Fig. 3, the number of species increases with the number of niche dimensions. Both the canonical line and the real data exhibit this pattern. We may understand it in ecological terms: new species are added to a community by means of the subdivision of additional niche dimensions.

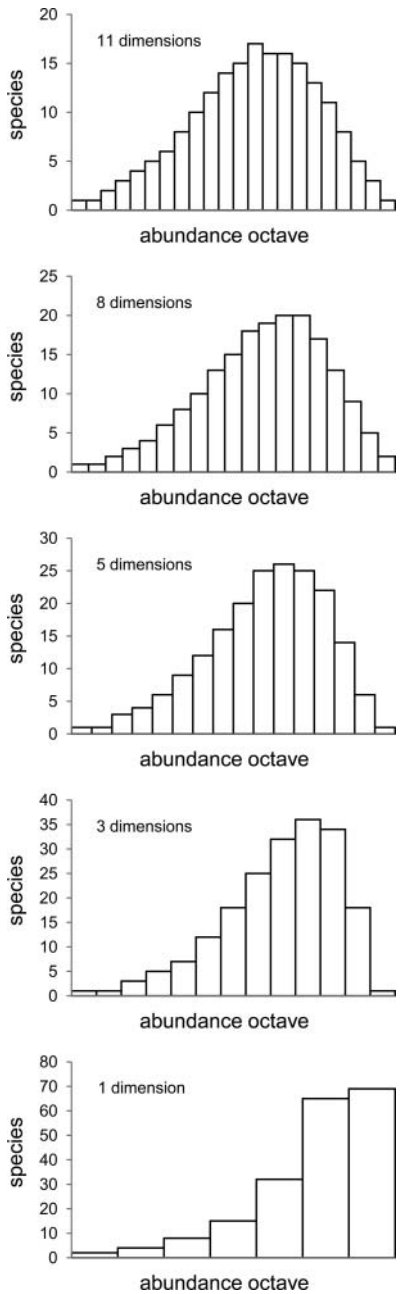

Fig. 2. Relative species-abundance distributions predicted for different numbers of niche dimensions. Mean number of species was obtained for each octave after 100 iterations with the model set at 200 species and 10^6 individuals (10^8 for 11 dimensions in order to show a more complete left tail).

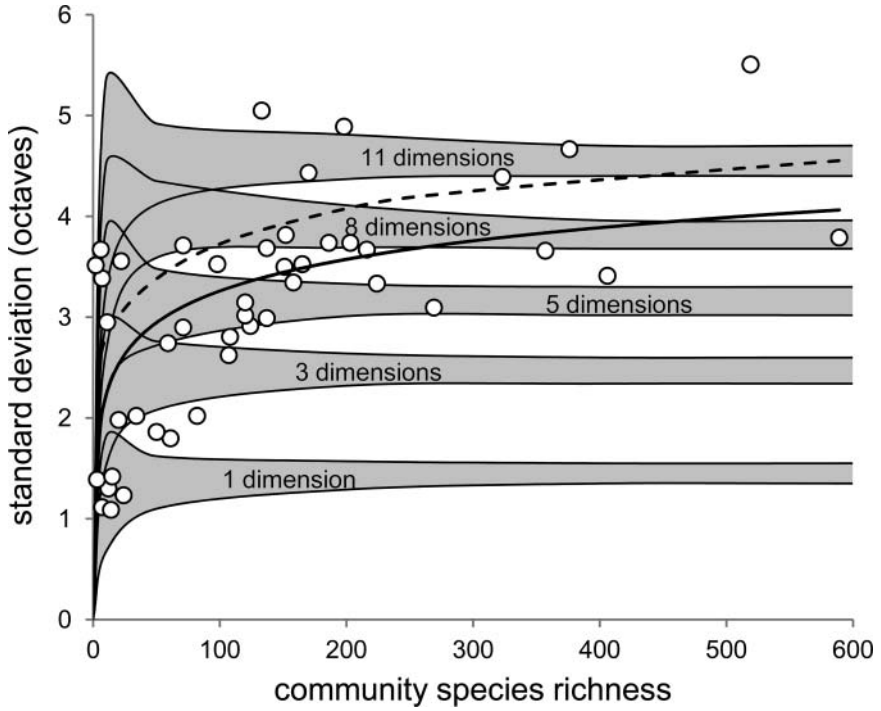

Fig. 3. Standard deviation vs. number of species obtained from the model for different numbers of niche dimensions. Each grey band corresponds to one standard error around the estimate of standard deviation in octaves as obtained from 100 repeated simulations set at d niche dimensions and 10, 50, 200, 400, and 600 species in a virtual community of 10^6 individuals (10^8 for 11 dimensions in order to show the left tail of the distribution as seen in Fig. 2). Open circles: data from real communities taken from Sugihara (1980). Solid line: semi-logarithmic regression of these points ($R^2 = 0.35$). Dashed line: prediction from the canonical property, after Sugihara (1980).

Figure 4 plots the relationship between the number of niche dimensions and the number of species along the line of the canonical property. I fit these data with nonlinear regression, obtaining

$$S = 1.789^d \qquad (R^2 = 0.97) \tag{1}$$

Using nonlinear regression, I also fit the points interpolated from the regression line of data from real communities, and obtained

$$S = 2.053^d \qquad (R^2 = 0.85) \tag{2}$$

The standard errors of the parameters, d, in equations (1) and (2) are 0.008 and 0.026 respectively, and the P-values are both < 0.0001.

These results suggest that the canonical property originates from two causes:

- The number of species increases locally because new niche dimensions are subdivided.
- Each subdivided niche dimension that we add roughly doubles the number of species.

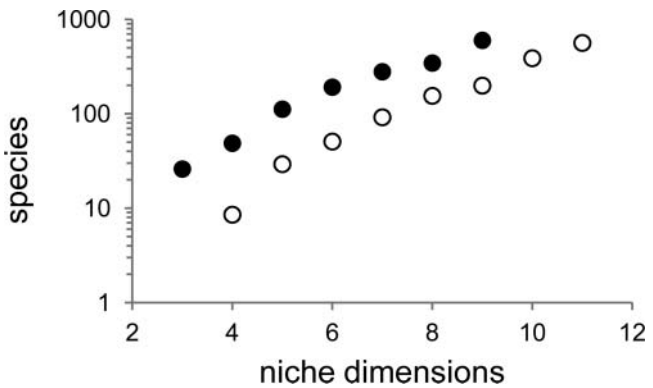

Fig. 4. The number of species rises as a function of the number of niche dimensions. Open circles: points interpolated from the line of the canonical property (Fig. 3). Solid circles: points interpolated from the regression line of data from real communities (Fig. 3). With additional model simulations for 4, 6, 7, 9, and 10 niche dimensions (but with 10^8 individuals for 10 dimensions to show the left tail of the distribution).

DISCUSSION

As May (1999) and Sugihara (1980) argued, there is no mystery at all in obtaining a lognormal distribution by adding random factors of multiplicative effect (because lognormal distributions originate in this way). The turn of the screw in the present paper rests in the study of the canonical property, starting from simulations of a model and progressing to comparison of its predictions with empirical data.

One can debate whether the model I use is a realistic approach to the niche–abundance relationship in natural communities. Its biological assumptions may be too simple but they are general and reasonably based on fundamental ecology. Anyway, the simulations show that the model can serve as a fruitful starting point to propose possible explanations of the canonical property. Furthermore, the model can predict realistic shapes for the relative species-abundance distribution as demonstrated by Fig. 3, in which real data fall within the model predictions.

Compared with real lognormal distributions obtained for core species (e.g. Magurran and Henderson, 2003), the model predicts only a slight right skew in the species histogram. This is evident in Fig. 2 for eight dimensions (a realistic value for 200 species according to Fig. 3). This skew probably arises from the fact that the model does not distinguish between core and satellite species – that is, there is an excess of predicted species in the left tail, species that in nature would be unable to maintain stable populations in the community due to their low population density. Thus one might add to the model a survival filter that depended on species abundance, and used, for example, the coefficient of stability obtained from Taylor's law (Taylor, 1961). However, the assigned survival probability would be arbitrary because for unknown reasons the power in Taylor's law ranges from 1 to 2 (Murdoch and Stewart-Oaten, 1989; Tilman, 1999). Thus, the model with a survival filter would lose generality and gain only a small amount of accuracy. I doubt that the benefits outweigh the costs in this case. I therefore consider the current version of the model sufficient for the purposes of this work, as long as it yields distributions that are reasonably like a lognormal at the niche dimensions and species numbers expected from the canonical property in Fig. 3.

In fact, I have undertaken preliminary simulations implementing an extinction risk as derived from Taylor's law at a power scaling of 1.5 (a commonly observed value). That corrects most of the right skew, thus approaching the lognormal distribution, whereas standard deviations change very little. Thus, the main result holds – that is, each niche dimension approximately doubles the number of species.

Although the uniform distribution of species occupancy of any niche dimension modelled here is suggested by the empirical distribution of species-abundance peaks along ecological gradients (e.g. Whittaker, 1956), other distributions might be possible and they could probably vary depending on the niche dimension. Any other distribution would decrease the randomness of species occupancies along a niche dimension, since modal occupancy values would be more likely. Following the model steps, it is easy to see that the effect of this clustering of niche occupancies will be a species-abundance distribution with a smaller standard deviation for the same number of niche dimensions and a smaller increase in standard deviation with each added niche dimension. Both predictions are confirmed for Gaussian distributions of niche occupancy (data not shown). Thus, the uniform distribution yields the maximum increase of species richness along niche dimensions. According to these arguments, it can be expected that the subdivision of any new niche dimension will multiply species richness by a factor of up to 2.

Independent support for the results shown in equations (1) and (2) comes from the conclusion by MacArthur (1964) that the number of bird species increases with the addition of each vertical habitat layer by a factor of slightly less than 2 (i.e. approximately 1.75). These values are strikingly close to the estimations for the parameters of equations (1) and (2). MacArthur assumed each habitat layer to be an independent niche dimension for the bird community. He then interpreted his data using the standard mathematical reasoning that the total number of distinguishable objects in n independent dimensions, with each dimension having x distinguishable levels, is x^n. Using this argument, MacArthur found that his data suggested that birds are able to discern, on average, 2.5 habitats per layer (x). This implies bird evolution has subdivided each habitat layer into 2.5 distinguishable ecospaces. From this perspective, equations (1) and (2) would mean that evolution subdivides each niche dimension into about two distinguishable ecospaces for species. Again, the value cited by MacArthur is surprisingly close.

Why exactly this subdivision and not any other? There is a reasonable evolutionary answer. In niche-assembled communities, the subdivision of a new niche dimension means a tighter packing of species in the ecospace. If each new dimension roughly doubles species number, then the subdivision allows the co-existence of approximately one more species per each previous species. The new species will probably occupy a niche similar to that of an old species, but be separated by their use of the new niche dimension. For example, the niche 'aerial insectivore' in a small bird community would include the swift and the nightjar, but their niches are entirely separated by their use of the temporal dimension, since one is diurnal and the other nocturnal. In a similar manner, the new species will subdivide the added dimension with a previous species. The point of the argument is that two subdivisions are intrinsically more stable than more subdivisions because the risk of competitive exclusion should increase with the number of subdivisions. For example, three subdivisions (per dimension and previous species) imply three species with similar niches competing along the same resource axis, a situation that would probably end with the competitive exclusion of one or two species from the local patch. On the other hand, with one subdivision (i.e. one species) per dimension, then previous species obviously cannot add any

diversity. The fact that x is sometimes found to be below 2 [see equation (1) and the initial estimation of 1.75 in MacArthur (1964)] suggests that the case of no subdivision happens sometimes but not as a rule. Thus the most stable case compatible with an increase of species number is two subdivisions. This seems to hold for natural communities according not only to the model presented here but also to the conclusions reported by MacArthur (1964). In this context, a niche dimension is simply any independent characteristic of the habitat that is suitable for ecological subdivision as local diversity grows.

These ideas suggest that the canonical property results from an assembly rule. When species accumulate in a community, they tend to occupy separate niches by partitioning a new niche dimension into two subdivisions per species, because competitive exclusion would destabilize any further subdivision. This simple rule, if confirmed by experimental evidence, would provide important insight into how ecological communities are assembled.

REFERENCES

Dolan, J.R., Ritchie, M.E., Tunin-Ley, A. and Pizay, M.D. 2009. Dynamics of core and occasional species in the marine plankton: tintinnid ciliates in the north-west Mediterranean Sea. *J. Biogeogr.*, **36**: 887–895.

Hutchinson, G.E. 1957. Concluding remarks. *Cold Spring Harbor Symp. Quant. Biol.*, **22**: 415–427.

Leitner, W.A. and Rosenzweig, M.L. 1997. Nested species-area curves and stochastic sampling: a new theory. *Oikos*, **79**: 503–512.

MacArthur, R.H. 1964. Environmental factors affecting bird species diversity. *Am. Nat.*, **98**: 387–397.

Magurran, A.E. and Henderson, P.A. 2003. Explaining the excess of rare species in natural species abundance distributions. *Nature*, **422**: 714–716.

May, R.M. 1999. Unanswered questions in ecology. *Phil. Trans. R. Soc. Lond. B*, **354**: 1951–1959.

McGill, B.J., Etienne, R.S., Gray, J.S., Alonso, D., Anderson, M.J., Benecha, H.K. *et al.* 2007. Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecol. Lett.*, **10**: 995–1015.

Murdoch, W.W. and Stewart-Oaten, A. 1989. Aggregation by parasitoids and predators: effects on equilibrium and stability. *Am. Nat.*, **140**: 288–310.

Nee, S., Harvey, P.H. and May, R.M. 1991. Lifting the veil on abundance patterns. *Proc. R. Soc. Lond. B*, **243**: 161–163.

Preston, F.W. 1962. The canonical distribution of commonness and rarity: Part I. *Ecology*, **43**: 185–215.

Rosenzweig, M.L. 1995. *Species Diversity in Space and Time*. Cambridge: Cambridge University Press.

Sugihara, G. 1980. Minimal community structure: an explanation of species abundance patterns. *Am. Nat.*, **116**: 770–787.

Sugihara, G., Bersier, L.F., Southwood, T.R.E., Pimm, S.L. and May, R.M. 2003. Predicted correspondence between species abundances and dendrograms of niche similarities. *Proc. Natl. Acad. Sci. USA*, **100**: 5246–5251.

Taylor, L.R. 1961. Aggregation, variance and the mean. *Nature*, **189**: 732–735.

Tilman, D. 1999. The ecological consequences of changes in biodiversity: a search for general principles. *Ecology*, **80**: 1455–1474.

Ulrich, W. and Ollik, M. 2004. Frequent and occasional species and the shape of relative-abundance distributions. *Divers. Distrib.*, **10**: 263–269.

Ulrich, W. and Zalewski, M. 2006. Abundance and co-occurrence patterns of core and satellite species of ground beetles on small lake islands. *Oikos*, **114**: 338–348.

Whittaker, R.H. 1956. Vegetation of the Great Smoky Mountains. *Ecol. Monogr.*, **26**: 1–80.

Williamson, M. and Gaston, K.J. 2005. The lognormal distribution is not an appropriate null hypothesis for the species–abundance distribution. *J. Anim. Ecol.*, **74**: 409–422.