OLM11.6. Bestiary of species-abundance distributions

Applying species-abundance distributions in the description of community structure has a long tradition (Fisher et al. 1943; Preston 1948). We have seen a new surge of interest in them after the turn of the millennium, because the key prediction of Hubbel's neutral theory regards species-abundance distributions and, indeed, the distributions observed do not significantly differ from the ones predicted by the theory. We have shown in Ch11Finit that this is not proof for the neutral theory, but a corollary of the fact that species-abundance distributions are weak statistical tests (McGill 2003). Note 11.1 has briefly introduced the species-abundance distributions touched upon in the book, but the literature has several different distributions as well. This OLM aims at reviewing the distributions used for modelling species-abundance patterns, the relations between them, and some of their mathematically correct applications.

The species-abundance distribution is a function specifying the probability (P_n) that a species randomly chosen from among those of a community will have n individuals. n can only be a natural number, so modelling species-abundance distributions requires the use of discrete distributions. Still it is not unusual to see applications in the literature illegitimately fitting continuous distributions – like the lognormal (Preston 1948) – on discrete data.

The number of individuals in a randomly chosen species is the result of two different random process (species choice and sampling). Therefore, these two processes are best modelled separately, and then the results combined. We start with modelling the sampling process, assuming that we have chosen a species. We are interested in specifying the distribution of the numbers of individuals of the focal species in samples of the same community.

If the individuals of a species are independently included in the samples, then the distribution of their numbers is Poissonian. The probability of having exactly n individuals of the species in the sample is

$$e^{-\lambda}\frac{\lambda^n}{n!} \tag{11.6.1}$$

The only parameter of the distribution is the expected number of individuals λ, which is a constant determined by the species and the sampling effort. If the inclusions of the individuals in the sample is not independent of one another (because of their spatial aggregation, for example), then using more complicated distributions with more parameters is unavoidable. Here we constrain the discussion to the simplest case (in which the Poisson distribution is applicable), which is sufficiently realistic for a large fraction of species.

For a single species the expected number of individuals (i.e., the λ parameter of the Poisson distribution) is a constant. It is a random variable for a randomly chosen species of the community, however. If the probability density function ($f(\lambda)$) of λ is known, then the species-individual distribution is given by

$$P_n = \int_0^\infty e^{-\lambda} \frac{\lambda^n}{n!} f(\lambda)\, d\lambda \,. \tag{11.6.2}$$

Integration starts from zero, because the average number of individuals (λ) cannot be negative.

Unlike the number of observed individuals within the sample – which is a discrete number – the distribution of the expected number of individuals in a randomly chosen species (λ) is continuous. It is important to make this distinction between the two distributions; it is only the former that we call species-abundance distribution.

Since λ>0, $f(\lambda)$ can be specified as distributions of random variables taking positive values only. From among the best known ones the gamma and the lognormal distribution are such, for example. Both are frequently encountered in the literature of ecology related to species-abundance distributions.

Let us first turn to the case where λ is a gamma-distributed random variable. The probability density function (PDF) of the gamma distribution is

$$f(\lambda) = \lambda^{k-1} \frac{e^{-\lambda/\Theta}}{\Theta^k \Gamma(k)} \tag{11.6.3}$$

In which $\Gamma(k)$ is the gamma function, $\Gamma(k) = \int_0^\infty t^{k-1} e^{-t} dt$, and for positive integer k values, $\Gamma(k)$=(k-1)!

The parameters of the distribution are Θ>0 (scale) and k>0 (shape). Its expected value is $k\Theta$, its variance is $k\Theta^2$. The Θ parameter depends on the sampling effort, k is a community specific constant.

Substituting the PDF of the gamma distribution (11.6.3) into Eq. (11.6.2) we get the following species-abundance distribution

$$P_n = \int_0^\infty \lambda^{k-1} \frac{e^{-\lambda/\Theta}}{\Theta^k \Gamma(k)} e^{-\lambda} \frac{\lambda^n}{n!} d\lambda = \frac{\Gamma(n+k)}{\Gamma(k) n!} \frac{\Theta^n}{(\Theta+1)^{n+k}} = \frac{\Gamma(n+k)}{\Gamma(k) n!} \left(\frac{\Theta}{\Theta+1}\right)^n \left(\frac{1}{\Theta+1}\right)^k \tag{11.6.4}$$

which is the negative binomial distribution. Upon the generic parameter setting of this distribution we obtain

$$P_n = \frac{\Gamma(n+k)}{\Gamma(k)n!} p^n (1-p)^k \qquad (11.6.5)$$

where $p = \Theta/(1+\Theta)$, that is, $0<p<1$.

According to this formula the probability that a species occurring in the community is not included in the sample is

$$P_0 = (1-p)^k \qquad (11.6.6)$$

Larger k implies more species missing from the sample. For $k \to 0$, $P_0 \to 1$. Writing the distribution only for the species found in the sample, i.e.,

$$P_n^* = \frac{P_n}{\sum_{i=1}^{\infty} P_i} = \frac{P_n}{1-P_0} \qquad (11.6.7)$$

the $k \to 0$ limit leads to Fisher's log-series distribution. Fisher (Fisher et al. 1943) had introduced the distribution that was later named after him the same way, as a "technical" distribution, but – as we have shown in OLM 11.5 – neutral dynamics leads to this same distribution.

The gamma distribution can be replaced by the lognormal for modelling the distribution of the expected values:

$$f(\lambda) = \frac{1}{\sqrt{2\pi}\sigma} exp\left(-\frac{(ln\lambda-\mu)^2}{\sigma^2}\right) \qquad (11.6.8)$$

The parameters of this one are the expectation (μ) and the standard deviation (σ) of $\ln\lambda$. In this case the species-abundance distribution of the sample is the Poisson-lognormal:

$$P_n = \frac{1}{\sqrt{2\pi}\sigma n!} \int_0^{\infty} e^{-\lambda} \lambda^{n-1} exp\left(-\frac{(ln\lambda-\mu)^2}{\sigma^2}\right) d\lambda \qquad (11.6.9)$$

μ depends on the sampling effort, whereas σ is a community specific constant. Unfortunately with this distribution P_n cannot be written in closed form like the negative binomial.

Surprisingly, the shapes of the PDF-s of the lognormal and the gamma distribution are very similar (Figure 11.6.1), provided that their expected values and standard deviations are the same and their squared variation coefficients (c.v.^{2}) are less than 1 (Silva & Lisboa 2007). For the lognormal distribution, $c.v.^2 = e^{\sigma^2} - 1$, thus the condition is met if $\sigma < \sqrt{ln2} \approx 0.832$. For the gamma distribution $c.v.^2 = 1/k$, so the condition is met if $k>1$. The similarity of the two distributions explains the good fit of both to empirical data (Figure 11.16).

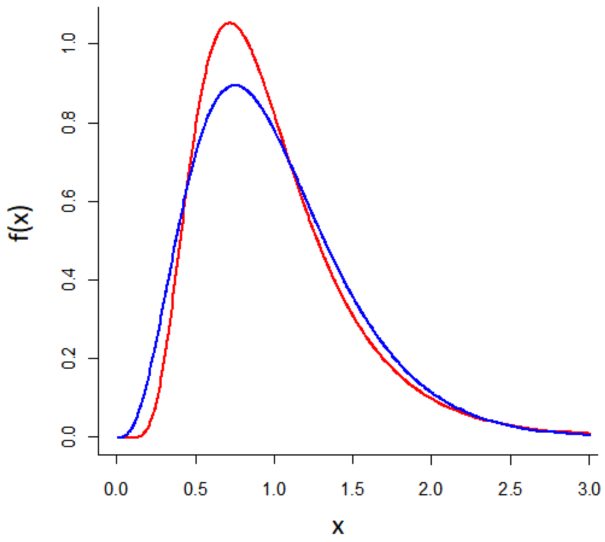

Figure 11.6.1: The PDF of the lognormal (red lines) and the gamma (blue lines) distribution for expectation 1 and variation coefficient ¼.

Species-abundance distributions can be visualized by histograms as is often done in statistics (the use of log2-transformed abundance histograms is common), but the so-called rank-abundance curves are also often used (Figure 11.6.2). These have the decreasing ranks of the species on the abscissa (rank 1 assigned to the most abundant species) and the logarithms of abundance on the ordinate.

McGill and co-authors (2007) have shown that the shape of rank-abundance curves strongly depends on the number of species occurring; therefore, conclusions based on the differences of such curves must be drawn with caution. It is safer to compare empirical distribution functions. Rank-abundance curves still have to be dealt with, because there are two mechanistic models which are often referred to and based on rank-abundance curves. These models have in common that both are primarily applied to resource partitioning, assuming that the relative abundances of the species are proportional to the resources they use.

The oldest of these models is the niche pre-emption model of Motomura (1932), which assumes that the dominant species pre-empts as its share a fraction *p* of the limiting resource, the second strongest species pre-empts a fraction *p* of what remains and so on. Accordingly putting the species in decreasing order of abundance the abundance of the *i*-th species is expected to be

$$n_{(i)} = p(1-p)^{i-1} \tag{11.6.10}$$

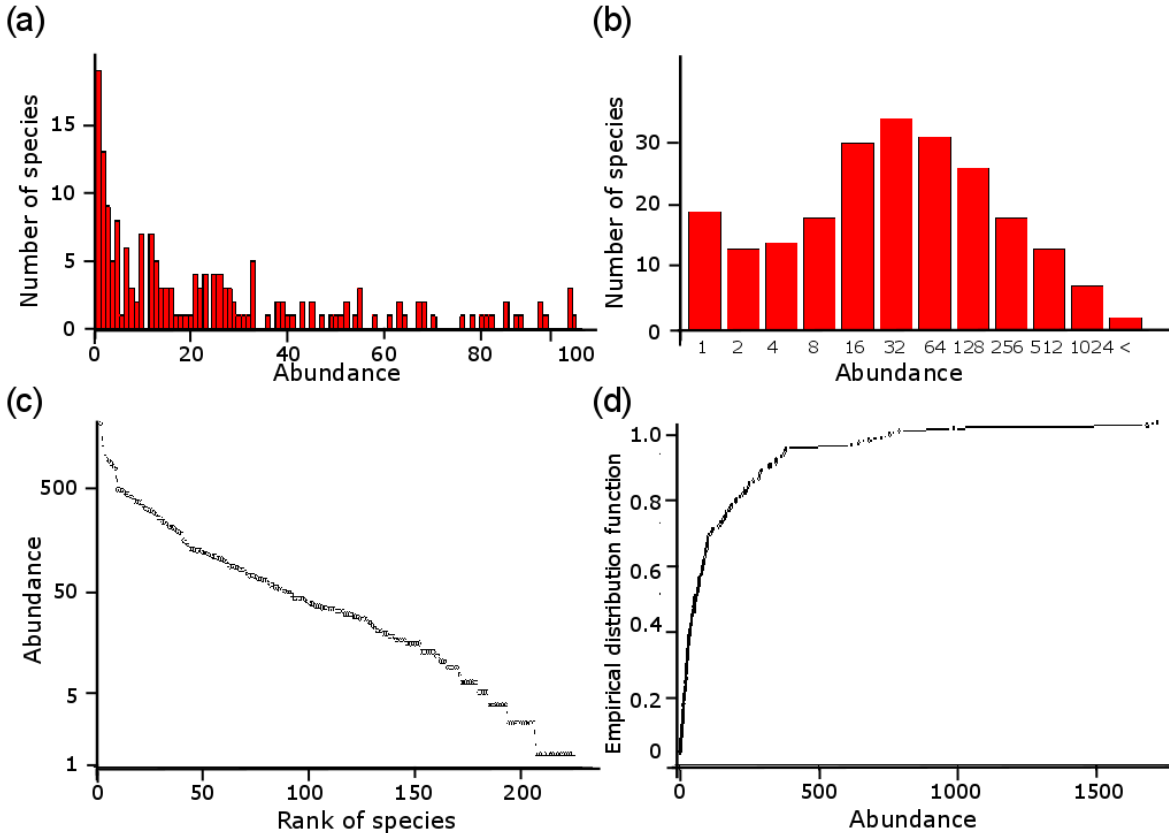

Figure 11.6.2: Different visualizations of species-abundance distribution data.

a) histogram (values over 100 not shown). (b) histogram with logarithmic abundance categories, numbers show the upper limit of categories. c) rank-abundance curve. d) empirical distribution function. Data are numbers of tree individuals over DBH=10 cm within a 50 ha section of rain forest on Baro Colorado island (data from Condit et al. 2002).

The rank-abundance curve predicted by this model is linear. The distribution of species abundances conforming to the niche pre-emption model is very uneven: species of intermediate abundance are rare.

The other model of rank-abundance curves is the broken stick model proposed by MacArthur (1957), in which the resources are represented by a stick of unit length. The stick is broken at S-1 random positions to yield a number of fragments equal to the number of species. The lengths of the fragments are assumed to be proportional to the resources allocated to the species, and thus also to the relative abundances of the species. Arranging the species in decreasing order of abundance the expectation for the relative frequency of the *i*-th species is

$$n_{(i)} = \frac{1}{S}\sum_{j=i}^{S}\frac{1}{S} \qquad (11.6.11)$$

The same length distribution can be arrived at by hierarchical breaking, i.e., by breaking stick fragments in two at random positions until we get the required number of fragments, provided that the probability of choosing the next stick fragment to break is proportional to its length. Choosing the next stick fragment at random (i.e. without considering the fragment's length) results in a lognormal distribution of lengths.

Both the niche pre-emption and the broken stick model predict expectation for the relative abundances of the species; therefore, they correspond to the distribution of the expected abundances (which is often represented by the lognormal or the gamma distribution) in species-abundance models. This is another reason for doubts about the fitting of empirical rank-abundance curves to these theoretical models. For a discussion of the practical issues regarding fitting (Wilson et al. 1998).

The theoretical distributions corresponding to the rank-abundance curves expected from the models can be found in two steps (Pielou 1975): first the empirical cumulative distribution function expected from the rank-abundance curve is drawn; then we look for the distribution which fits to that (Figure 11.6.3).

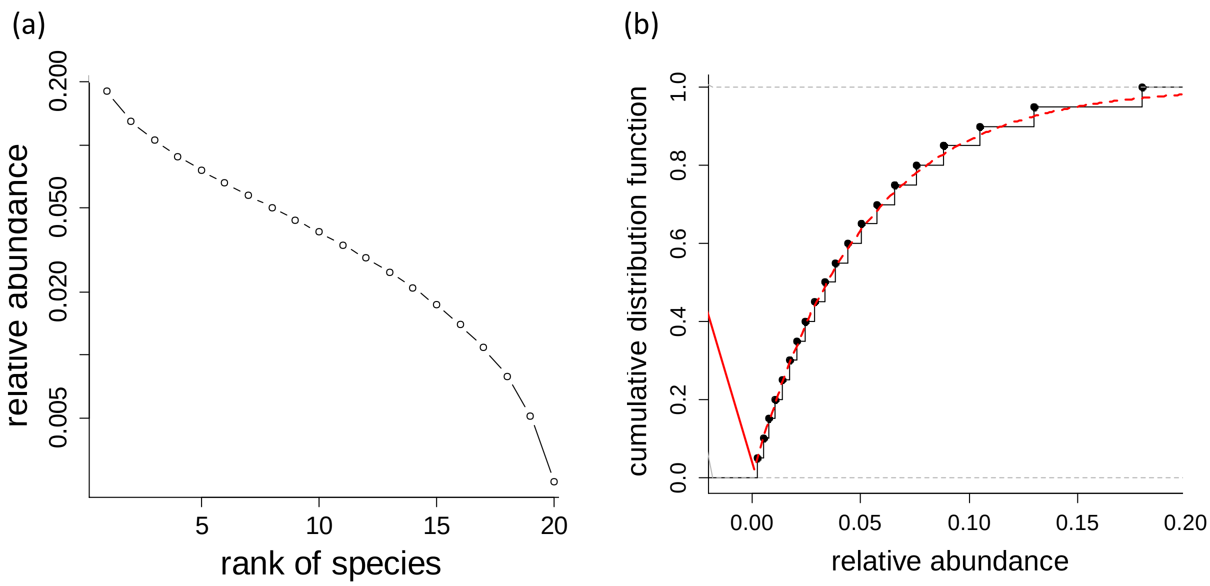

Figure 11.6.3: Broken stick model of a 20-species community.

a) expected rank-abundance curve of relative abundances and b) the corresponding empirical cumulative distribution function. The red dashed line is the distribution function of the best fitting exponential distribution.

The distribution thus determined for the broken stick model is the gamma, with parameters k=1 and $\Theta=1/s$. Note that the gamma distribution becomes the exponential distribution for k=1; this is why the literature often discusses this distribution in relation to the broken stick model. The niche pre-emption model cannot be associated with any special distribution of expected abundances.

References

Condit, R, Pitman, N, Leigh, E.G., Chave, J., Terborgh, J., Foster, R.B., Nuñez, P., Aguilar, S., Valencia, R., Villa, G., Muller-Landau, H.C., Losos, E. & Hubbell, S.P. (2002). Beta-diversity in tropical forest trees. *Science* 295, 666–669.

Fisher, R.A., Corbet, A.S. & Williams, C.B., 1943. The Relation Between the Number of Species and the Number of Individuals in a Random Sample of an Animal Population. *Journal of Animal Ecology*, 12(1), pp.42–58.

MacArthur, R.H., 1957. On the Relative Abundance of Bird Species. *Proceedings of the National Academy of Sciences of the United States of America*, 43(3), pp.293–295.

McGill, B., 2003. Strong and weak tests of macroecological theory. *Oikos*, 102(3), pp.679–685.

McGill, B.J. et al., 2007. Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecology Letters*, 10(10), pp.995–1015.

Motomura, I., 1932. A statistical treatment of associations [in Japanese]. *Japanese Journal of Zoology*, 44, pp.379–383.

Pielou, E.C., 1975. *Ecological Diversity*, New York: John Wiley & Sons.

Preston, F.W., 1948. The Commonness, And Rarity, of Species. *Ecology*, 29(3), pp.254–283.

Silva, E. L., & Lisboa, P. (2007). Analysis of the characteristic features of the density functions for gamma, Weibull and log-normal distributions through RBF network pruning with QLP. *Proc. 6th WSEAS Int*, 223-228.

Wilson, J.B. et al., 1998. Relative abundance distributions in plant communities: effects of species richness and of spatial scale. *Journal of Vegetation Science*, 9(2), pp.213–220.