

Relative Species Abundance:

**Literature review of the theory and model development
behind the use of abundance curves
in the field of Ecology.**

**Kristie L. Allen
Department of Mathematical Sciences
Montana State University**

17 December, 1999

**This writing project submitted in partial fulfillment
of the requirements for the degree**

Master of Science in Statistics

Approval

of a writing project submitted by

Kristie L. Allen

This writing project has been read by the writing project director and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the Statistics faculty.

1/12/00
Date

Robert J. Boik
Robert J. Boik
Writing Project Director

Introduction

The use of the lognormal distribution to describe species abundance has been a key subject of scientific growth in the ecological literature for decades. Each species, whether we are talking about a specific bird, plant, insect, or a fungus, is represented by a certain population size (or number of individuals) in a given area. This population size is the species' abundance. If we then look at a subset of species that occur in a region and determine what percentage of the total number of individuals each species represents, we can determine their relative species abundance. This writing project will focus nearly exclusively on questions concerning the validity of the lognormal distribution applied to species abundance. Before jumping into the particulars of the project focus, I would like to review the beginnings of the theory from the seminal papers of Frank Preston. The specific goals of the project will be addressed at the end of the introduction.

Frank W. Preston, in 1948, first suggested lognormal relative abundance of species in his paper *The Commonness and Rarity of Species*. "Random samples of ecological or taxonomic assemblages indicate that the universes from which they are drawn have, at least approximately, the form of an ordinary Gaussian curve drawn upon a logarithmic base (a "lognormal" curve)." Even though the lognormal distribution cannot be the true underlying model, since frequency of species abundance is discrete, Preston felt it was a good approximation. When Preston plotted his data, he made three decisions that have been carried through or argued in the follow-up literature: (1) Preston used a log 2 base for his plots because to him it was "logical" and "a matter of convenience" for species in one bin to be twice or half as common as those in an adjoining bin, (2) Preston plotted species abundance by grouping species into a range of abundances like bins in a histogram, and he calls the bins "octaves", (3) the octave boundaries were whole integers, for a base 2 logarithm, and he solved the problem by taking the number of species on the boundary, dividing it in half, and assigning one half to the bin on each side. Problems with one or another of these plotting decisions have come up in my own thoughts and in other papers, and will be discussed later in the paper. But, for introductory purposes, understanding what and how he plotted his data is important.

Let Y be the species abundance for a randomly chosen species. Apparently Preston assumed that Y has a lognormal distribution conditional on $Y \geq 1$. That is,

$$f(y | y \leq 1) = e^{-(\log_2 y - \mu)^2 / 2\sigma^2} / y(\ln 2)\sqrt{2\pi\sigma^2} [1 - \Phi(-\mu/\sigma)] I_{(1,\infty)}(y).$$

Preston, however, plotted his data in bins or “octaves.” So, to describe what is found in each bin requires an integration of portions of this pdf,

$$\Pr(c \leq Y \leq d) = \int_c^d f(y | y \leq 1) dy.$$

The empirical estimate of $\Pr(c \leq Y \leq d)$ is obtained by splitting the endpoints into each of the surrounding bins. The equation above can easily be adapted for the discreteness of the data, for plotting purposes, by changing the integral sign to a summation when dealing with abundances that are integers.

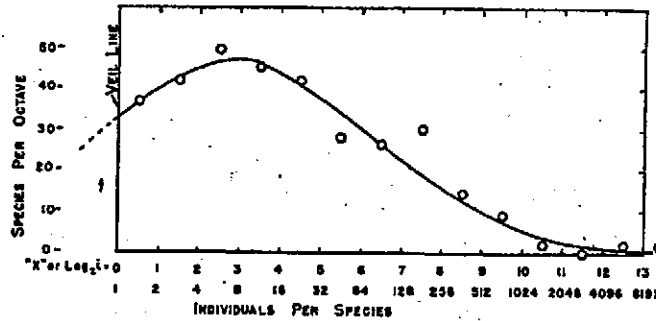


Figure 1 Dirk’s moths of Orono, Maine. The octaves are not equally filled; the curve is humped; we see the right-hand tail, but the left-hand one is hidden behind a veil. From Preston, 1948.

Preston stated that we see the underlying distribution, though the left tail is hidden behind a “veil line” (Fig. 1) or truncated as we described above. The species with population sizes too small to have been captured in the sample are hidden behind this line. If the sample size is doubled, Preston claimed, the species in the original sample would gain representatives at nearly the same rate and wind up with about twice as many specimens and, therefore, be assigned to the next higher “octave.” In the doubling of the sample we may also capture representatives of “rare” species, those not seen in the original sample. The main result being an octave of new species, revealing a step of the lognormal curve that, before the doubling of the sample, was hidden.

Since the Gaussian (or normal) distribution is symmetrical, once we have identified one half of the curve, the other half can be inferred. Preston used this method for predicting the total species diversity of sites being sampled.

Preston expanded on his hypothesis in 1962 by relating the species abundance curve with an individuals per octave curve. The primary result of this paper was to create the "Canonical Lognormal Distribution," one that Preston claimed has some undeniable biological truth. After examining many sets of data, Preston defined a new variable R to be the number of octaves from the mean. Without explaining his reasoning well, Preston saw some basic underlying truth in the variance of R and it was coupled with S , species number, to obtain a fixed variance for most biological communities. With this restriction on the variance of R , and therefore the σ^2 of the underlying lognormal distribution, the dimension of the parameter space is reduced to one (Fig. 2). I found that I had a difficult time determining what Preston's reference to variance and dispersion, which is normally understood as a synonym of variance, referred to. The variances of (1) the underlying lognormal with y as the measure on the x-axis and (2) the re-scaled lognormal with R as the x-axis are very confused in this paper. I believe that Preston referred to the variance in the number of octaves about the mean. In *The Canonical Distribution of Commonness and Rarity, Part I*, Preston discussed the implications of this lognormal distribution on species-area curves and island biogeography, two major areas of research interest in the field of ecology.

Frank Preston founded or developed theoretical support for many of the foundational concepts and exciting research avenues open to ecologists today. However, he was also the first to counsel caution in taking his ideas too far. In Part II of the above-discussed paper, he warns

"The observation that commonness and rarity are often distributed approximately lognormally must be treated as being at present merely an observation. The hypothesis that all possible lognormals are not equally likely, but that the distributions tend to cluster close to that particular lognormal I have ventured to call Canonical, is also without theoretical basis. It would be a great help if we could discover a sound theoretical reason for expecting the results of the present paper. In absence of such a reason, the possibility remains that we may have dealt only with a substantial group of observations that accidentally fitted our surmises and that we may later find others that departed widely from them."

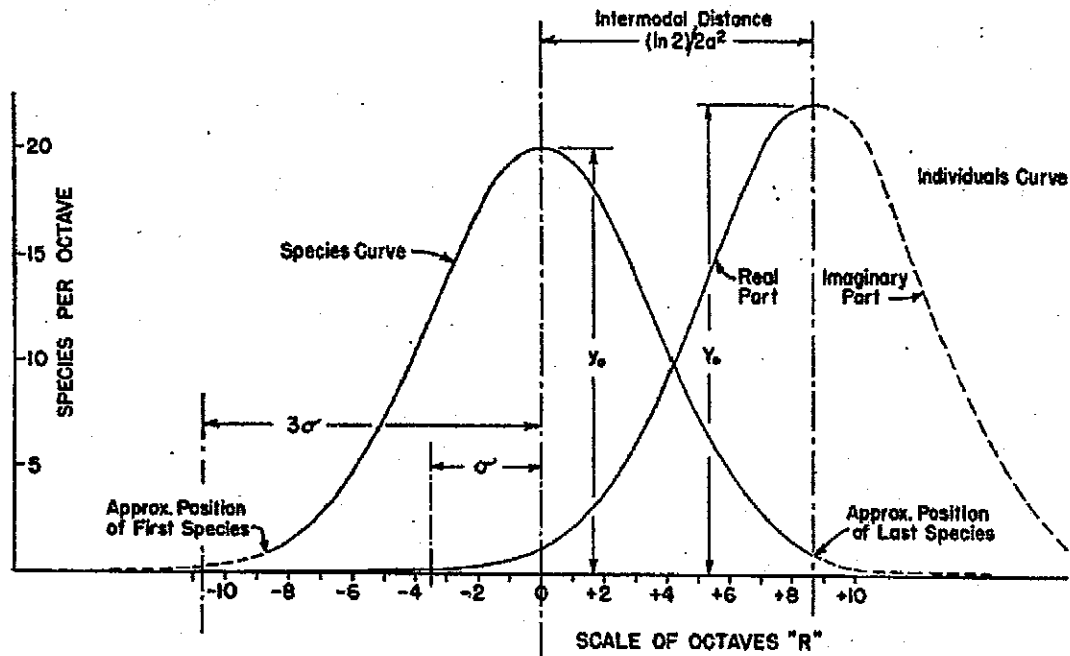


Figure 2 The canonical lognormal distribution for an ensemble of 178 species. By convention, the x-axis is scaled as logarithmic (base 2) abundance classes or “octaves” of individuals/species, adjusted to have a mean of zero. The species curve denotes the number of species in each octave and the individual’s curve shows the number of individuals in each abundance class. This particular distribution is canonical because the mode of the individual’s curve, R_N , coincides with the upper truncation point of the species curve, R_{MAX} (i.e., $\gamma = R_N/R_{MAX} = 1$). Setting $\gamma = 1$ couples species count and variance in a specific way. Figure from Preston, 1962 and Explanation from Sugihara, 1980.

Despite Preston’s humility and warnings about taking his ideas beyond the scope for which they were meant, the lognormal species abundance curve has become accepted as ecological law and appears in a wide range of ecology texts, from the introductory to graduate levels (Brown & Lomolino, 1998; Huston, 1994; Molles, 1999; Rosenzweig, 1995; Stilling, 1999). Acceptance, today, has been based mainly on publications from Robert May (May, 1975) and George Sugihara (Sugihara, 1980).

May published a possible statistical cause for the lognormal distribution that got a great deal of attention. Preston had already suggested this reasoning decades earlier in an obscure journal (Preston, 1957), but the ecological community did not notice it at that time. The explanation of the existence of the lognormal abundance distribution goes roughly as follows. Relative species abundance is regulated by a *very* large number of variables. Environmental and habitat variables control the ranges of individual species (Huston, 1994; MacArthur, 1975; MacArthur & MacArthur, 1961; Pianka, 1967; Platt, Shaw & Lamshead, 1984; Shmida & Wilson, 1985). Variables controlling competition and resource division between species determine niche separation (Cook, 1995; DeVries, Murray & Lande, 1997; Hairston, 1959;

Hughes, 1985; Klopfer & MacArthur, 1961; MacArthur, Recher & Cody, 1966; Paine, 1966; Pianka, 1967; Simpson, 1964; Sugihara, 1980; Sugihara, 1989; Taper & Marquet, 1996; Tokeshi, 1990). Life history variables couple with those of biogeography to influence species composition (Bauer & Sadler, 1993; Brown & Lomolino, 1998; Cotgreave & Harvey, 1994; Diamond & May, 1981; MacArthur & Wilson, 1967; Pielou, 1975; Shmida & Wilson, 1985; Usher, Field & Bedford, 1993). Evolutionary and ecological constraints also contribute to species ranges and densities (Connell & Orias, 1964; Janzen, 1973; MacArthur & MacArthur, 1961; May, 1975; Pianka, 1966; Rosenzweig, 1995; Rosenzweig & Abramsky, 1993; Whittaker, 1972). If these variables affect species abundance additively, the Central Limit Theorem implies that the distribution of these abundances will tend toward a normal. If, instead these variables act on abundance multiplicatively, the distribution will tend toward a lognormal, which is seen in many biological cases.

Sugihara suggested a mechanistic source for the relative abundance distribution. He offered an explanation of how species divide resources in forming their assemblages, and this theory also resulted in the formation of a lognormal distribution of species abundance. Sugihara determined that a hierarchically structured communal niche space will give species abundances that plot lognormally. In other words, when three species will eventually share or compete for resources, the first two species will determine their shares based on the many variables mentioned in the above paragraph, then the third species will carve out its piece of from one of the first two. Mathematically, this simplifies into randomly dividing a unit resource at a point uniformly distributed between 0.5 and 1, as one species will always be dominant. Then, one of the resulting pieces is chosen randomly and it too is considered a unit and split in the same way. Sugihara claimed that this sequential breakage results in a lognormal species abundance curve when applied to a large number of species, though no proof was provided.

So, with this statistical confirmation and theoretical biological base, the lognormal distribution has become entrenched in the ecological literature. Manuel Molles (Molles, 1999) states "The 'distribution of commonness and rarity' among species described by Preston is one of the best documented patterns in natural communities." Huston (Huston, 1994) calls lognormal species abundance "one of the most consistent phenomena in ecology..." and furthers that "The lognormal distribution is ubiquitous in nature...."

The concept of the lognormal species abundance distribution may, however, not be as well supported as it is commonly believed. Debate has raged on both ecological and statistical journals for decades. Data presented in Preston (Preston, 1948; Preston, 1957; Preston, 1958; Preston, 1960) and many other articles discussed in the following sections do show a humped curve in species abundances (see also data from Table 1) that could be described as lognormal. However, a methodological twist with Preston's method causes many such data sets to look humped, when in fact, arithmetically, they are not. Other evidence is from small data sets. Preston's (truncated) lognormal was often cited as the best fit best since it was a versatile shape and was much easier to calculate than other suggested models like the log-series, negative binomial (Fisher, Corbet & Williams, 1943), or Poisson lognormal (Bulmer, 1974).

This project has three goals: (1) to follow the development of the theory over 5 decades, (2) to find realistic parameter ranges for lognormal distributions relating to species abundance for the purpose of future Monte Carlo sampling simulations and (3) to complete a literature review of the support and opposition to the validity of the theory. If the underlying distribution of species abundance is lognormal, and the sampling distribution is then a truncated lognormal as Preston suggested then there must be a range of parameters within which natural species communities arrange themselves. From a review of the literature, I hope to find reported parameters of lognormal abundance distributions and gain insight to the range of the lognormal parameters present in nature. The Monte Carlo simulations I plan on using are included in this project as Appendix I. From this literature review I hope to be able to assemble a complete picture of the theory and data behind the concept. Several questions will be examined. How many data sets support the conclusions of a lognormal species distribution? Do data exist that do not follow the pattern, and if so, can the data be explained? Do the data examined provide us with a concept that is applicable across many or all taxa, or is it specific to a few? I hope to be able to answer some, if not all, of these questions by the end of this paper.

Literature Review – Theory Development across the Decades

The body of literature concerning species diversity is vast, encompassing a wide range of ideas all touching on diversity, but each from different perspectives. The literature pertaining to relative species abundance is mainly developed from two areas of research. The first tract to develop was that of fitting distributions and models to sampled data, with researchers trying to find a model flexible enough to accommodate different communities of different taxa. The other area of research pertaining to relative abundance is that of resource partitioning. How species divide resources will strongly impact their relative abundances. I follow the development of these two perspectives from their beginnings, sequentially charting the literature as it was published.

The first mention I found of species abundance distributions in the ecological literature dates back to 1943, when Fisher, Corbet and Williams looked at entomological collections (Fisher *et al.*, 1943). They determined that when one looked at abundances of the rarer species (they stopped counting after 45 individuals per species), a Fisher-series or log-series fit the data very well. Fisher showed that the log-series was a limiting result of a negative binomial distribution. If the number of each species contained in the sample was a Poisson variate with mean λ_i

$$f(y_i | \lambda_i) = \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!}$$

where $y = 0, 1, 2, \dots, \infty$
and $\lambda > 0$,

and, where the λ_i 's in turn were distributed as a gamma random variable $\{\lambda_i \sim \text{Gamma}(\alpha, \beta)\}$, the resulting joint distribution would be a product of the two,

$$f(y_i, \lambda_i | \alpha, \beta) = \frac{e^{-\lambda_i} \lambda_i^{y_i} \lambda_i^{\alpha-1} e^{-\lambda_i \beta} \beta^\alpha}{y_i! \Gamma(\alpha)}$$

The distribution of y_i , then, without dependence on λ_i , simplifies to a negative binomial as follows

$$f(y_i | \alpha, \beta) = \int_0^\infty f(y_i, \lambda_i | \alpha, \beta) d\lambda_i = \frac{\beta^\alpha}{y_i! \Gamma(\alpha)} \int_0^\infty e^{-\lambda_i(\beta+1)} \lambda_i^{\alpha+y_i-1} d\lambda_i = \frac{\Gamma(\alpha + y_i)}{y_i! \Gamma(\alpha)} (\rho)^\alpha (1-\rho)^{y_i},$$

where $\rho = \beta/(1+\beta)$.

However, since y_i cannot take values less than 1, there exists a truncated negative binomial distribution

$$P(Y_i = y_i | Y_i \geq 1) = \frac{f(y_i | \alpha, \beta)}{P(Y_i \geq 1)} = \frac{f(y_i | \alpha, \beta)}{1 - P(Y_i = 0)} = \frac{f(y_i | \alpha, \beta)}{1 - \rho^\alpha}$$

where. $y = 1, 2, \dots, \infty$
 $\alpha, \beta > 0$
 $0 < \rho < 1.$

When the shape parameter, α , of the gamma approaches zero, the log-series is the result, eventually.

If we rewrite $1/\Gamma(\alpha)$ in the distribution of y using Euler's formula (Abramowitz & Stengun, 1970), we get

$$P(Y_i = y_i | Y_i \geq 1) = \frac{\Gamma(\alpha + y_i)}{y_i! \Gamma(\alpha)} (\rho)^\alpha \frac{(1 - \rho)^{y_i}}{(1 - \rho^\alpha)} = \frac{\Gamma(\alpha + y_i)}{y_i!} (\rho)^\alpha \frac{(1 - \rho)^{y_i}}{(1 - \rho^\alpha)} \alpha e^{-\alpha y} \prod_{n=1}^{\infty} \left[\left(1 + \frac{\alpha}{n} \right) e^{-\alpha/n} \right]$$

Then we take the limit the distribution as $\alpha \rightarrow 0$, finding

$$P(Y_i = y_i | Y_i \geq 1) = \frac{\Gamma(y_i)}{y_i!} (1 - \rho)^{y_i} \left(-\frac{1}{\ln \rho} \right) = \frac{-(1 - \rho)^{y_i}}{y_i \ln \rho} \quad \text{where } y = 1, 2, \dots, \infty$$

by using L'Hopital's rule. The result is the log-series probability mass function.

Though this treatment may seem complex as a method of describing a decreasing species abundance curve, the idea has a firm basis in mathematics. This first attempt at describing a relative species abundance curve remains a viable model, in my opinion, at least accurate in that a discrete process was modeled with a discrete distribution. Ecologists and statisticians alike revisited this model and variations of it in the following decades, and the statisticians at least relied on it heavily in their own species diversity work.

Preston, five years later, published his first paper on the lognormal distribution of relative species abundance (Preston, 1948). I have several problems with his methodology as discussed in the introduction to this paper. Primarily, the plotting of the data into the bins or octaves he used is not only misleading, but also mathematically unwieldy (which as we know is unforgivable). I am going to go into a bit of detail here to exemplify in what way I disagree with Preston's methodology.

Most data shows a humped curve (Table 1), when plotted in \log_2 octaves, giving a location to the mode that is so important in estimating species richness from the lognormal distribution. This data, which

has been construed as strong support for lognormal abundance models, can be very misleading. The existence of a mode in the lognormal plot is *not* indicative of a decrease in the rate at which new species are being added to the sample. This idea is a common misconception among ecologists resulting from Preston's methodology.

I am going to give a step by step description of how the convention of Preston's plotting method can give misleading and highly biased results for estimates of species diversity. At small abundances he would divide the number of species with border abundance (species with abundance equal to 1, 2, 4, 8, 16, etc.) in half and assign one half to the bin below and the other to the bin above. It sounds reasonable, but the effect on whether or not there will be a mode in the distribution is huge. Take, for example, a community that was sampled and the following progression of abundances was found (with the number of species as a superscript): $1^{61}2^{24}3^{13}4^{12}5^57^68^59^210^411^212^313^115^117^118^219^226^130^133^137^140^144^162^264^1$ (Osa peninsula beetles, from (Janzen, 1973)). This sample would have 61 species with only a single specimen of each, 24 species with only two representatives in the sample, etc. To divide these species into abundance bins Preston would proceed as follows. First, the species represented by a single individual would be divided in half with half going in the <1 bin and half assigned to the bin 1-2. Second, species with 2 individuals each would also be halved with one set staying in the 1-2 bin and the other going to the 2-4 bin. Then the species with three specimens would all be assigned to the 2-4 bin, as would half of the species with four representatives, et cetera (see Fig. 3). So, in a community with nearly 40% of its species being represented by a single specimen, Preston's plotting method would still result in a mode being present. In fact, no matter how large the group of species represented by a single individual, there will always be a mode in the 1-2 bin if any species are present in the sample with the abundance of two. If no species are present with an abundance of one, the mode is contained in an octave larger than the second. For this to happen, which is extremely rare, we must have sampled to such an extent in the community to have at least two of each rare species without finding any rarer species with an abundance of only one.

The bias introduced by this methodology would severely underestimate the number of species in the community. And this would be especially true for species-rich or under-sampled sites. In today's world, the effect could be any of the following: underestimation of the amount of additional sampling required, having an area dropped from a conservation priority list because the species diversity estimate

was biased low, or a conservation allotment being too small to support the number of species actually in the habitat.

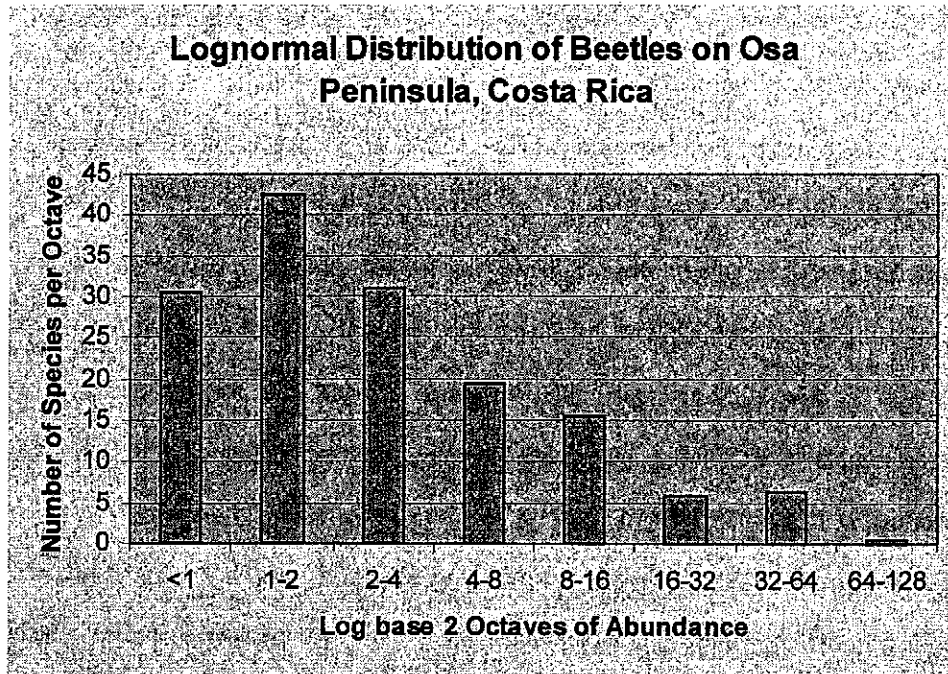


Figure 3. Data from Janzen (1973) plotted using Preston's (1948) method into lognormal abundance octaves. There is a mode using the log₂ base, even when the number of species represented by a single individual is nearly three times the number of species with two individuals.

Mathematically, I must disagree with this plotting method purely on the basis of its messiness. The formation of octaves only needed to be inclusive on the lower end and non-inclusive on the upper to be neat, infinitely easier to work with, and avoid the problems discussed above. However, since it is the splitting of octaves that I have trouble with and not the lognormal distribution itself as a model for species abundance, I would still like to consider the lognormal a candidate model for species abundance. Perhaps Williams' idea of using log₃ sized bins or a non-histogram version of the distribution would be more appropriate. It does bother me that this continuous distribution may give very different estimates of diversity than a discrete one for the same data. However, I would like to examine the differences.

After Preston (Preston, 1948) first introduced the lognormal theory, Simpson (Simpson, 1949), Dobzhansky (Dobzhansky, 1950), Williams (Williams, 1953), Brian (Brian, 1953) and Patrick et al. (Patrick, Hohn & Wallace, 1954) all came out with ideas or data pertaining to species abundance. Simpson

simply clarified the relationships between the log-series, negative binomial, and the Poisson. Dobzhansky related the idea of species abundance to evolutionary and environmental characteristics of the habitat, introducing the idea of niche and the trade-off between being a generalist or specialist species. He did mention that a large number of species represented by only a single individual implies more new species are likely to be found if sampling was to be repeated. The largest contribution to Preston's theory came from Williams. He apparently adopted the lognormal as the appropriate model, after looking at a longer time-series of data from his light-traps. He suggested, however, abandoning the use of log 2 as the base for the species abundance plots and using log 3 instead. When plotting the "octaves" of Preston, we would have to split the number of species between bins having numbers of individuals on an octave boundary. Using log 3 circumvents this problem. Williams also pointed out that whether or not a mode is seen when data is plotted, depends on the log base chosen. The mode can disappear completely as the bins in the histogram change. Brian, on the other hand, proposed a return to the less simplified (than the lognormal) negative binomial distribution suggested by Fisher. He noted that with the zero term removed, the negative binomial is not restricted to the case where the rarest species occur in small numbers. A situation where rare species are the most numerous also could be modeled with the distribution proposed by Fisher. Ruth Patrick applied the models of the day to her large samples of river diatoms. She found that the lognormal fit as well as the others and was the easiest to work with. The debate between different models continued in the decades ahead, but in my opinion, was the purest at the beginning. Theories were being offered up for consideration. How these theories were examined and extrapolated by the ecological community after this introduction of ideas made for interesting reading.

As a follow-up to his original paper, Preston (Preston, 1957) applied the lognormal to two years worth of Maryland State bird counts (averaged), to show the applicability of the theory to avian data. He also introduced the idea of a statistical cause for the distribution the Central Limit Theorem, though he does so without actually naming the theorem and in an obscure journal. Other points brought up in the 1957 paper are possible explanations for deviations commonly seen from the lognormal fitted curves Preston calculated for data. Points tended to fit above the curve on the left side of the distribution and below on the right. Preston after discussion with count coordinators, believed the pattern in the points resulted from the "staking out" of rare species and the underestimation or conservative counts of common species (Fig. 4). It

seems that the data collectors, local bird watchers were involved in an informal competition to determine who could find the most species on the data collection day. Rare species were located at a prior date so they could be found when it was time to collect the data, hence biasing the sample high for rare species.

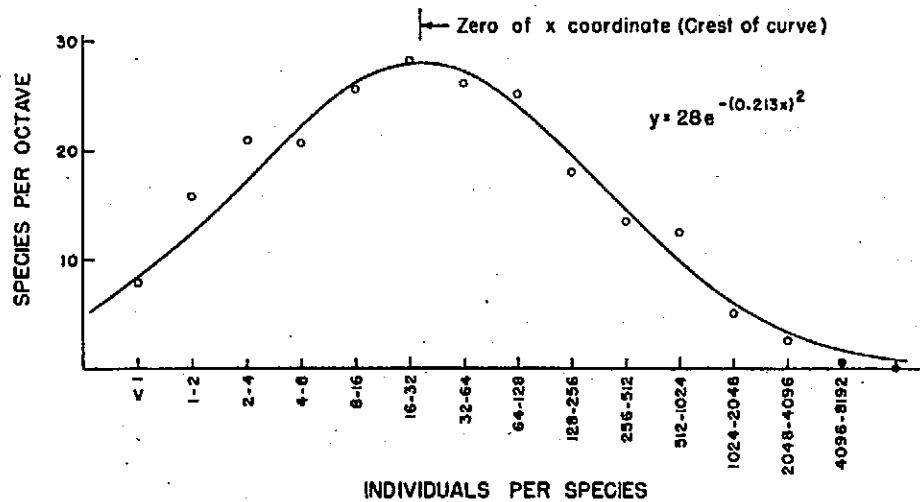


Figure 4. Maryland State-Wide Bird Count in 1956 and 1957 (Spring). Curve is an average of two years. Note the points that fall above the curve on its left side and those that fall below on the right side. Preston believes these points do not fit the curve because of non-random sampling by the collectors. From Preston (1957)

Three more papers came to the surface in the late 1950's on species abundance, though little advance was made theoretically. MacArthur (MacArthur, 1957) continued to develop theory on how species share resources and examined how resources determine the relative abundance of species. He developed three hypotheses of resource splitting, and presented the first as the most compelling. If communities were formed with non-overlapping continuous niches, MacArthur argued, species abundances would conform to what was seen in nature. Of course niches are not discrete or non-overlapping, but an approximation was being made, compromising the truth of the biology with the difficulty of the mathematics. Preston (Preston, 1958) wrote then on the lognormal in regards to the Audubon Christmas Bird Counts. In this paper he reiterated his belief that non-random sampling was taking place in bird counts, citing competition among samplers and their emphasis on rare birds. Hairston (Hairston, 1959) then sides with Preston over Brian's preference of the negative binomial and argued against MacArthur's best hypothesis presented in 1957 by comparing simulations to arthropod data. He showed that only for small sample sizes did MacArthur's theory predict the abundances seen in samples.

Six papers published in the 1960's focused on species abundance, a pair of which advanced theory markedly. For starters, Robert MacArthur published two more papers, one alone (MacArthur, 1960) and another with co-author John MacArthur (MacArthur & MacArthur, 1961), and continued to work on the biological causes of relative abundance of species. In 1960, MacArthur made a distinction between opportunistic and equilibrium species, noting that opportunistic species tend to conform to the lognormal. He insisted that equilibrium species would, if taken from a single habitat or life history, conform to his earlier non-overlapping niche theory. In 1961, the authors questioned more thoroughly the concept of niche and its effect on species abundance. Plant diversity was seen as a correlate with bird diversity, but MacArthur and MacArthur showed that it is due nearly entirely to the addition of niches with layers of canopy.

Chris Pielou (Pielou, 1966) wrote a more statistically oriented paper on biological collections. She outlined how to go about measuring diversity on collections with different sampling properties. In this paper, she also developed the relationship between the Shannon-Weiner index and measures of evenness, effectively kicking off ecology's decades long, and fruitless (in my opinion), love affair with diversity indices.

Preston's work did not falter during this period. In 1960, then again in 1962, he pressed on with new additions to his lognormal species abundance theory. Preston (Preston, 1960) explored the equivalence of space and time on samples of biological populations. He drew parallels between Gleason (Gleason, 1922) and Williams (Fisher *et al.*, 1943) who both said that doubling sampling effort should increase the number of new species by a set amount. Gleason was referring to sampling area, however, while Williams wrote about time. Preston took his lognormal distribution and began to explain the relationship between increasing species richness and increasing area. Preston (Preston, 1962a; Preston, 1962b) introduced the idea of the Canonical Lognormal Distribution. Sugihara (Sugihara, 1980) explained it best, so I will use his words.

By convention, a lognormal species curve is canonical of the parameter γ has a constant value of 1, where γ is the ratio of the position of the individuals curve mode, R_N , to the upper truncation point of the species curve, R_{MAX} (see Fig. 2). This is essentially an empirical relationship deriving from the observation that R_N and R_{MAX} have a tendency to coincide. Notice that when $\gamma = 1$, the independent variables of the general lognormal distribution become coupled to yield a specific one-parameter family of

lognormal curves. For example, the specific coupling between species count, S , and the variance of the lognormal, σ^2 , can be written approximately as

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

...it shows that under the canonical hypothesis, a large variance in the distribution accompanies a large species count. In the general lognormal distribution, on the other hand, these two parameters are completely independent.

So, the canonical lognormal distribution was introduced into the ecological literature, curtailing the relationship between variance of the curve and the number of species sampled. Statisticians ignored this so-called advance, as nowhere has this distribution been analyzed for robustness or appropriateness for the data it is attributed to explaining so well. In the field of ecology, however, the distribution has been truly "canonized." Accepted completely by those who write the texts and model species diversity, the distribution has been the least criticized piece of ecological theory in the last 4 decades.

The book, *Patterns in the Balance of Nature* (Williams, 1964), written during this decade had a great deal to say about species abundance. Williams had the following to say about species diversity:

As a result of ... complex and endless interplay each species establishes temporarily an uneasy balance of numbers among all the others. The pattern of the relative abundance of all the species in a mixed community (and all wild communities are mixed) is thus a synthesis of all the competition and cooperation, and all the difficulties and facilities, that have surrounded all the species of a community in the recent past; and every minute the position of any one species is changing in relation to the others. (page 2)

Williams chose to mention the preferences of both himself and Preston when dealing with the lognormal relationship of relative species abundance. Williams was still convinced of the log3 base being superior, but presented the log2 base as an option as well. He offers numerous examples of sampling data in Chapter 3 and even applies the lognormal frequency distribution to the number of species in genera (Chapter 6) and the abundance of parasites on their hosts (Chapter 8). Clearly he thought the lognormal distribution widely applicable to problems in ecology.

In the 1970's, the ecological literature exploded with papers on species diversity. Though, unfortunately, many of these were focused on finding the ideal index of species diversity (DeJong, 1975; Heck, Belle & Simberloff, 1975; Hill, 1973; Kempton & Taylor, 1974; Patil & Taillie, 1977; Peet, 1974;

Peet, 1975; Smith & Grassle, 1977; Whittaker, 1972). Indices, being a single number, were a combination of the ideas of species richness (number of species in a community) and evenness (how close the relative abundances of species come to being equal). The relative abundance curve, of course, contains both elements, but there was a distinct desire by the ecological community for a single number, instead of a distribution, that could also give the same information. We know that two concepts cannot be explained adequately by a single variable, but this knowledge did not keep ecologists from looking for it. We will ignore this sector of the species diversity literature, and concentrate instead on developments in research on relative species abundance distributions as Hurlbert (Hurlbert, 1971) suggests:

We therefore can muddle along with a plethora of indices, each supported by at least one person's intuition and a few recommended by fashion, or we can sharpen our thoughts and rephrase our questions in terms of biologically meaningful properties which, when calculable on a list of species and their abundances, we might refer to as species composition parameters.

Some authors did research in both areas, trying to fuse the two approaches together. Whittaker (Whittaker, 1972) listed many indices and debated the usefulness of each for different purposes. In the same paper, he hypothesized links between the lognormal distribution and the niche theory that was being explored by MacArthur. He suggested that the geometric series fit communities with a strong dominance of one species over all the others and that MacArthur's non-overlapping niche theory (also known as broken-stick) applied very well to communities of territorial and harem-forming species. Additionally, Whittaker noted that no matter which of the above (or even a combination of both) paradigms were the basis for a community, the sampling distribution of species abundance would appear lognormal.

Dan Janzen did extensive sampling of insects in Costa Rica and on tropical islands (Janzen, 1973). Of 127 samples in different habitats and locations, in only 12 of these samples were singletons, species with only one representative specimen, not the largest class of the abundance curve. The sites where singletons did not prevail tended to be particularly species-poor with an average of only 9.7 species per site. He did not try to fit any models to his data, but if he had the log-series would have fit best for most samples.

Bulmer (Bulmer, 1974) circulated, most probably, the most statistically coherent treatment of species abundance as a sampling problem in the decade of the 1970's. He proposed a similar tactic as Fisher (Fisher *et al.*, 1943), but instead of numbers of species being Poisson random variables with mean, λ_i , and

λ 's distributed gamma, Bulmer suggested the λ 's be distributed lognormally. So, the marginal distribution of y_i is

$$f(y_i | \alpha, \beta) = \int_0^{\infty} f(y_i, \lambda_i | \alpha, \beta) d\lambda_i = \frac{(2\pi\sigma^2)^{-1/2}}{y_i!} \int \lambda_i^{y_i-1} e^{-\lambda_i} e^{-(\log \lambda_i - \mu)^2 / 2\sigma^2} d\lambda,$$

when $\lambda \sim \text{LN}(\mu, \sigma^2)$ and $y_i = 0, 1, 2, \dots, \infty$. His justification for using the lognormal distribution over the gamma is actually a modification of MacArthur's broken stick model, which originally stated that the "stick", or resource, is randomly divided between species. Bulmer makes the suggestion of sequentially breaking the stick, still randomly, resulting in a set of abundances that should be distributed lognormally (he refers us to a discussion of the "theory of breakage" by Aitchison and Brown (Aitchison & Brown, 1957)). Thus, a lognormal distribution of the number of items in a class was expected when the only requirement was independence of breakage on each portion of stick at each stage. Bulmer also developed maximum likelihood estimates for parameters and applied the results to the ubiquitous diversity indices of the time.

Kempton and Taylor (Kempton & Taylor, 1974) used the same distributions as Bulmer, the negative binomial and the Poisson lognormal, and focused on the estimates of the parameters for moth data. Concerned with the biological meaning of the parameters, they fit the models to moth light-trap data and estimated the parameter for the log-series and lognormal. Interestingly, Kempton and Taylor found log-series to fit best in stable communities while the lognormal fit changing communities, and cited greater skewness in the lognormal as the reason. Most research until this point had found the opposite.

In 1974, Steinar Engen (Engen, 1974) advanced even further the theory of Fisher [Fisher, 1942 #50] and Brian (Brian, 1953), that species abundances are distributed with a negative binomial model. He extended the negative binomial distribution to fit more situations (details about what these situations might be are unstated in the paper) by allowing the shape parameter, α , to take values from -1 to 0 , in addition to those greater than 0 . Also provided in this paper are pseudo maximum likelihood and pseudo moment methods for estimating parameters for this "extended negative binomial" model. Later in the '70's, Engen proposed the geometric series as an approximation of relative abundances (Engen, 1975; Engen, 1977). Engen cited a problem with the assumptions of the original models (Brian, 1953; Fisher *et al.*, 1943; Pielou, 1966; Preston, 1948). Authors had assumed that the abundances of different species were independent,

when ecologically, dependence is guaranteed by inter-specific competition. By reformulation of the original models into conditional distributions, Engen derived the geometric distribution as underlying the expected values of the species distributions for both the lognormal and negative binomial distributions. Patil and Taillie (Patil & Taillie, 1977), disagreed with Engen's results in favor of those of Watterson (Watterson, 1974). Watterson had taken models describing population growth of evolving genotypes and applied them to species abundance. He independently found a basis for Fisher's logarithmic series of abundance from three of the genetics-based theories. Engen (Engen, 1978) refuted this challenge in a monograph, *Stochastic Abundance Models*, which was a very thorough review of the models presented in the literature to date. This debate which occurred in the statistical arena, I am sad to say, went nearly unnoticed by the field of ecology. Intelligent discussion of which distribution was most precisely representing the process or which was most elastically descriptive of biological communities was heard by very few ecologists.

While the statisticians debated amongst themselves, trying to decipher sampling assumptions, May (May, 1975) published a chapter in an important ecology text, *Ecology and Evolution of Communities*. He tried to smooth all ruffled feathers by clarifying a statistical reason for the lognormal abundance curve in most situations, while assuring biologists that biological reasons existed for finding other distributions in specific contexts. May reiterated in clear statistical terms Preston's (Preston, 1957) idea that the lognormal was a natural result of complex systems via the Central Limit Theorem. This time the explanation took hold. He also stated that in certain sampling situations, there were biological reasons for other distributions to be found; namely MacArthur's broken stick, geometric, and log-series. May's chapter was a good review but was too conciliatory for any real progress.

Taking a different tact entirely, Pielou in her book *Ecological Diversity* (Pielou, 1975) raised objections concerning the validity of discussing a distribution of abundances. She states

Notice first that to test the fit of a hypothetical model, it is not permissible to fit a theoretical frequency distribution to the "frequencies" of the different species listed. These "frequencies" are not frequencies in the usual sense; they are measures of the "sizes" or abundances of the species encountered, in other words, observations on the variate whose distribution is being investigated.

Another topic is mentioned in the book with which she disagreed with previous authors. She emphasized that when increasing the sample of specimens by either extending the time or area of the sample, one is changing the underlying community being sampled. One is not, she insisted, increasing the intensity of the sampling.

I agree with her completely, but when I look at the problem a different way, I disagree. I agree with her totally on the point that the abundances are not random variables in a statistical sense. For instance, if sampling was doubled in a given place we would not get twice as many data points. We obtain a different number of points (species) with new values associated with them (abundances). This is why, when we increase sample effort we are not increasing the intensity of the sampling, we are actually changing the distribution from which we are sampling. On the other hand, it makes intuitive sense to me that there exists, at a given place at a given time, a true distribution of relative abundances of species. If we sample specimens from the species present, are we not going to get an abundance distribution with which we can estimate the underlying universe? And further, if we sample from the species present more intensively, will we not be able to estimate this distribution with more accuracy? This question is pivotal to the ecological fascination with biological diversity. Are we even sampling in such a way as to make the underlying abundance universe estimable? Is it even estimable in any meaningful way? Pielou certainly is able to state the pivotal questions and question the integral issues in order to force thought about the complexities of the subject, both from a sampling and probability standpoint.

With the advent of the 1980's, we entered an era of computers being commonly applied to tasks in ecological research. There is no surprise to me, then, that several advances were made in this decade resulting from simulation and complex systems analysis.

Sugihara (Sugihara, 1980) disagreed with May's and Preston's explanation that the lognormal distribution came about through a statistical artifact. He wanted to see a mechanistic explanation and started with niche apportionment. Through simulations, Sugihara determined that resource or niche partitioning could explain both the canonical lognormal abundance pattern and the species-area relationship. Bulmer (Bulmer, 1974) first presented this idea, but the simulations showing the same result added credence to the idea. If a resource is divided randomly and sequentially, a lognormal niche size pattern is observed. If we assume that resource size translates directly into population size, niche

apportionment can explain the lognormal-like pattern observed so often. Later in the decade, work was done (Harvey & Godfray, 1987; Sugihara, 1989) on whether or not the assumption that abundance was equal to resource allocation held up under scrutiny.

Preston continued to look at the lognormal abundance distribution and published ideas about when abundance curves are not canonical (Preston, 1980), though he really asked more questions than he answered. After comparing Christmas Bird counts in the tropic and the Nearctic, he was puzzled by the differences in estimated sigma values for the fitted lognormal. He suggested that the canonical lognormal might only describe equilibrium situations, and that there were many causes for ecosystems not to be stable.

May (1981), was the first to speak of the lognormal abundance theory in tones that suggested it was a fact of nature. He stated "Once the community consists of a relatively large assembly of species, ... there is a bell-shaped Gaussian distribution in the logarithms of the species abundances." He did mention, however, that simpler communities, those with low numbers of species, can be described by other distributions and tended to be more stable under human perturbations. The importance of these ideas for global management was his closing statement, which I include here because it was very well-worded, though a bit moralistic:

An important general conclusion is that the large and unprecedented perturbations imposed by man are likely to be more traumatic for complex natural systems than for simple ones. This inverts the naïve, if well intentioned, view that "complexity begets stability," and its accompanying moral that we should preserve, or even create, complex systems as buffers against man's importunities. I would argue that the complex natural ecosystems currently under siege in the tropics and the subtropics are less able to withstand our battering than are the relatively simple temporal and boreal systems.

Benthic marine species assemblages were examined by Hughes (Hughes, 1984; Hughes, 1985; Hughes, 1986), and he discovered that many marine invertebrate communities did not conform to the by now standard log-series and lognormal abundance relationships. When he looked at dominance-diversity plots (species in rank order on the x-axis and number of individuals sampled for that species on the y-axis) of the standard models and his data, he found that there was a concave shape to his data that fit neither the log-series nor the lognormal (truncated to the left of the mode). He did not compare his data to a lognormal distribution truncated to the right of the mode, however. He stated that this method was untenable, since to

predict any of the parameters of the lognormal, the mode is needed. Yet, given his method of choosing between models, which consisted of determining by eye whether or not the predicted model showed the “deeply concave” dominance-diversity plot that his data had, I cannot see that these looking at an approximate lognormal could give any less precise results. I believe that if he had done so, he would have discovered a striking similarity, especially in the shapes of the curves, which he seemed most interested in.

Hughes based his “dynamic” community simulation models on colonization, competition between species, and population dynamics of marine organisms and found that he could duplicate the concave shape of observed marine abundance relationships. He also discovered that stable communities from his models had more similarity to the log-series and disturbed communities looked more lognormal. Please note that this agrees with Kempton and Taylor (Kempton & Taylor, 1974) who also went against popular ecological theory when reporting their simulation findings. In his third paper of this series, Hughes (1986) compared the fit of log-series, lognormal, and his own dynamics models to 222 different datasets of different organisms in different localities. Hughes’ model, based on population growth and colonization, was able to fit 94% of the samples, versus 4% for the log-series and 29% for the lognormal. He also returned to Williams’ notion that the mode of the lognormal abundance distribution was highly dependent on the scale of the logarithm used and suggested the mode had little correspondence to the ecology of the community. It seems to me that the dynamics model, which fits data with mechanistic assembly of species abundances, would be a good thing to use in ecology. However, Hughes neither supplied the model in any of his papers nor did he suggest an approximate form. So, though he claims that his model is better in many ways than those traditionally used by the ecological community, I don’t see the usefulness of Hughes’ model to any but his marine invertebrate studies, as he has not opted to inform the rest of us what the model consists of.

A great book, *Ecological Diversity and Its Measurement*, reviewed the measurement of diversity and appeared in 1988 (Magurran, 1988). Though Magurran primarily focused on diversity indices and how they are used in environmental decision making and monitoring, she prefaced the body of the work with an outstanding review of abundance distributions. She not only discussed the distributions themselves and under what circumstances they could often be found, but also covered methods of fitting them to data and which are the most meaningful measurements to use for abundance. In her opinion, biomass is a much

more meaningful measure and relates more directly to resource use than number of specimens. I tend to agree with her on this point.

Species abundance distributions are still a hot ecological topic in the present decade. Further developments in resource allocation, sampling techniques, and statistical testing have been made. Mutsunori Tokeshi followed in the footsteps of Sugihara's work on resource allocation models. The lognormal abundance theory becomes accepted by the ecological community and is included in several texts (Brown & Lomolino, 1998; Huston, 1994; Molles, 1999; Ricklefs, 1990; Stilling, 1999) and thematic books (Ricklefs & Schluter, 1993; Rosenzweig, 1995). The questions about the use and accuracy of abundance distributions, however, do not disappear and are discussed in several papers seen below.

Tokeshi started off the decade with a paper on niche apportionment (Tokeshi, 1990). He listed a number of mechanisms that, ecologically, were plausible, some of which we had seen before and others completely new and much more complex. His random assortment model was a fresh idea in that it assumed there was no dependence between the species' abundances. Ecologically, this would mean that species are not at high enough population sizes to cause competition impacts on each other. The niches are not full, so there is little relation between the abundance and niche size. His work brought him to the conclusion that most communities contained species under a mixture of regimes: a few dominant species being controlled by niche size and the rest following non-limited random assortment. Obviously a follower of Sugihara's work in the 1980's, Tokeshi updated the model presented by Sugihara in 1980 (Tokeshi, 1996). By changing the way in which a niche was selected for division (see Fig.5) he unified ideas of Sugihara and MacArthur. By using his power fraction method, we can, by changing the weighting procedure for choosing which niche to divide, produce the sequential breakage model of Sugihara, the weighted breakage model of MacArthur, or any gradation between the two. If there is no weighting, the Random Fraction model of Sugihara is the result. If the weighting is proportional to the niche size, MacArthur's Fraction model is created. On the other hand, a more complex weighting is possible with this Power Fraction method. We can weight the probability of choosing a niche by the size of the niche raised to any power. Adding complexity to the model, making it more generalized, Tokeshi hoped to make it more applicable to real biological data.

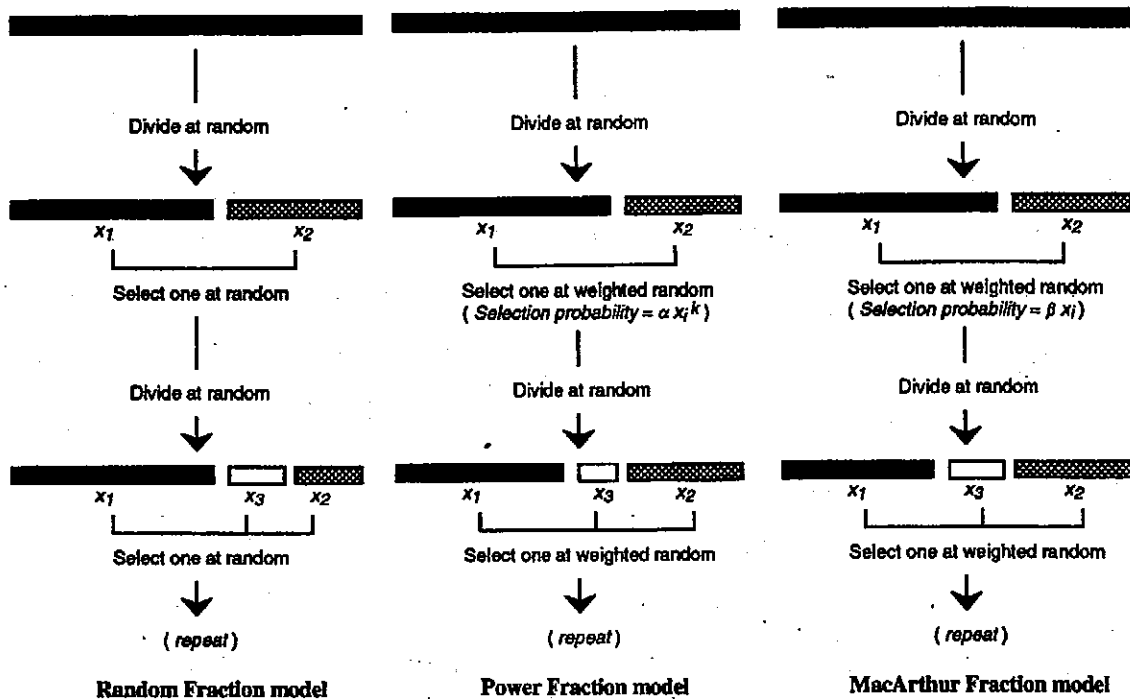


Figure 5. Schematic representation of the Power fraction model (middle), together with the Random fraction and MacArthur fraction models. In the Random fraction model, species with different niche sizes/abundances have the same chance of being selected for subsequent niche division (speciation or invasion by a new species), i.e. probability weighting is zero. In the Power fraction model, the probability of selection is proportional to niche size (or abundance) raised to a power. In the MacArthur fraction model, the probability of selection is proportional only to the size of the niche. From Tokeshi (1996).

Tokeshi tested the fit of the model to high quality data and found that the fit was good. He believed that, compared to Sugihara and MacArthur's models, it had wider theoretical implications. He asked what other niche apportionment models could be encompassed if he expanded the possible range of k to plus and minus infinity.

Colwell and Coddington (Colwell & Coddington, 1994) wrote on a diverse number of subjects pertaining to biodiversity. One section of their paper reviews methods of fitting relative abundance for the purpose of estimating species richness. They review the treatments from the past, admit there are problems with the lognormal model, and encourage development of methods to fit discrete distribution. Since the data are not continuous, discrete distributions give very different estimates of richness, though Colwell and Coddington claimed that they are prohibitively difficult to fit. I am sure this paper will become one of the new "classics" because of Colwell and Coddington's thorough treatment of the subject matter.

Sampling using nested subsets is the topic for two papers with opposing conclusions. Cook (Cook, 1995) insisted that any measurement of diversity that does not include the area sampled is useless for comparison to other sites. Many of the diversity indices of the '70's and '80's were lacking this feature, as are species abundance distributions. On the other hand, it was found in tropical plots where complete census' were performed (Condit *et al.*, 1996) that "species accumulated simply and predictably as a function of individuals counted, but not as a function of area." Condit *et al.* concluded that species-individual curves (relative abundance) were better for estimating diversity than were species-area curves. The second paper also had some very interesting results regarding the shape of sampling plots and its effect on diversity estimates.

There were several papers published in the '90's that contributed small but still important pieces to the puzzle. A paper consisting mostly of diversity indices comparisons (Ghent, 1991) had the two following points in its favor: (1) Ghent tried to put the math surrounding diversity issues into language understandable to most biologists, and (2) he did a good job of determining which indices were meaningful, and mentioning, as if in passing, that even the good indices were no more informative than the variance of the species abundances. The lognormal distribution was seen to also describe the size of species ranges in the United States (Pagel, May & Collie, 1991), a topic linked with species abundance at a given site. Bastow, a botanist, decided to provide standardized methods for fitting the different types of models describing relative abundance (Bastow, 1991). Bastow found that none was consistently better than any other. South African insect abundances were found to follow either lognormal or log-series distributions (Davis, 1993), though the author did not publish the parameters. In an Ecuadorian rainforest, butterflies abundances were also lognormal (DeVries *et al.*, 1997) even at low numbers of specimens per species (6690 individuals of 1130 species). Fagan and Kareiva (Fagan & Kareiva, 1997) looked at species accumulation curves and lognormal species abundance curves of Oregon butterflies, and determined that the lognormal approach gave better estimates of diversity with less chance of overestimation. Keating *et al.* (Keating *et al.*, 1998) used a negative binomial estimator (Efron & Thirsted, 1976) with most success when predicting richness of simulated communities and collected insects. Most of these papers were focused on the organisms in question and not methods of estimation, so the contribution to the theory is small yet important for its support or refute of the existing information.

That concludes the information I have found in the published literature, but there are several developments not yet published, that deserve mention. Maurer and Taper (Maurer & Taper, pers. comm.) have examined long term samples of Breeding Bird Surveys and found that in nearly all instances a gamma distribution fits the species abundances better than the lognormal. Another development occurred pertaining to the negative binomial model, with which Engen took exception due to the assumption that species abundances must be independent. In nature, abundances are clearly not independent due to competition and predation between species. Brown et al. (Brown *et al.*, in prep.) look at correlation matrices of simulated multi-species assemblages, and found that even if species abundances are very dependent on each other in a two-species system, the correlation coefficients between the same species in a larger assemblage are very small. There may exist a moderating effect from all the inter-dependencies between many species such that independence in multi-species communities may not be an insurmountable assumption.

In summary, multiple theories along each of two branches of research converge toward species abundance distributions. Several workable, and equally valid approaches, are presently available to choose from in the literature. I, personally, see no reason from the literature to focus solely on the lognormal. I find the negative binomial to be especially attractive due to its discrete nature matching the form of data and predictions desired.

Table 1. A list of data sets used in literature focused on species diversity and relative abundance theory. Not all entries were originally collected for the type of analysis done in the cited papers, however, each is a sample of the number of organisms, biomass, or pairs of species located in a given place over a given time. Datasets are listed in chronological order of their citation in the species abundance literature, and include the organism type, the original source, total individuals captured, length of sample period (in years), where it was cited in the species abundance literature, and what was found by those who looked at the relative abundance of the data set. Blanks in the table reflect a lack of information in the referring paper. This table is by no means comprehensive. It is merely a list of data, in those papers that I have personally read, that I have seen species abundance theory applied to.

Organisms	Source	Total Ind.	Total Spp.	Yrs.	Cited in	Notes
Moths	Williams, 1939, 1940	15,609	240	4	Fisher et al., 1942	No mode arithmetically
Birds	Saunders, 1936	14,353	80	1	Preston, 1948 Williams, 1953 MacArthur, 1960 Preston, 1962 Williams, 1964 Engen, 1977	Mode log 2 Mode log 3 Mode arithmetically
Birds and Nests	Preston and Norris, 1947		47	2	Preston, 1948	Mode log 2
Moths	Dirks, 1937	56,131	349	4	Preston, 1948 Engen, 1977 Williams, 1964	Mode log 2 No mode arithmetically
Moths	Williams, 1943	15,609	240	4	Preston, 1948	Mode log 2
Moths	King, unpublished	87,110	277	22	Preston, 1948	Mode log 2
Moths	Seamans, unpublished	303,251	291	23	Preston, 1948	Mode log 2
Trees	Black et al. 1950	564	60	1	Dobzhansky, 1950	No mode arithmetically Mode log 2
Trees	Black et al. 1950	423	87	1	Dobzhansky, 1950	No mode arithmetically
Moths		6814	197	1	Williams, 1953	No mode arithmetically Mode log3
Moths		35,428	304	4	Williams, 1953	Mode log 3
Moths		5,972	197	2	Williams, 1953 Bulmer, 1974	Mode log 3 No mode arith
Moths		12,456	249	2	Williams, 1953	Mode log 3
Moths		87,400	346	4	Williams, 1953	Mode log 3
Moths		7,378	197	0.08	Williams, 1953	No mode log 3
Moths		5,232	84	0.08	Williams, 1953	No mode log 3
Moths		26,300	265	0.75	Williams, 1953	Mode at log 3
Beetles	Palmen, 1944	~13,000	393	2/365	Williams, 1953 Williams, 1964	Mode arithmetically
Beetles	Palmen, 1944	~14,000	466	2/365	Williams, 1953 Williams, 1964	Mode arithmetically
Birds	Stewart & Aldrich, 1949	~200	24		Williams, 1953	No mode log 3

Organisms	Source	Total Ind.	Total Spp.	Yrs.	Cited in	Notes
Collembolans	Agrell, 1941	1,744	27		Brian, 1953	Mode neg. bin.
Collembolans	Agrell, 1941	1,832	22		Brian, 1953	Mode neg. bin.
Collembolans	Agrell, 1941	2,565	25		Brian, 1953	Mode neg. bin.
Collembolans	Agrell, 1941	6,414	35		Brian, 1953	Mode neg. bin.
Collembolans	Agrell, 1941	4,189	27		Brian, 1953	No mode log-series
Collembolans	Agrell, 1941	9,033	36		Brian, 1953	No mode log-series
Collembolans	Agrell, 1941	11,126	27		Brian, 1953	Mode neg. bin.
Collembolans	Salt et al., 1948	4,859	24		Brian, 1953	Mode neg. bin.
Birds	Robbins, pers. Comm.	>20,000	223	2	Preston, 1957	Mode log 2
Birds	Scott, 1953	6,273,759		1/365	Preston, 1958	Mode log 2
Birds	Scott, 1956	24,762,840	487	1/365	Preston, 1958	Mode log 2
Birds	Thomas, S. E.		177	10	Preston, 1960 Preston, 1962	Mode log 2 in octave 1-2
Birds	Hicks, 1935		86	10	Preston, 1962	Mode log 2
Birds	Walkinshaw, 1947		31	10	Preston, 1962	Mode log 2
Birds	Williams, 1947		33	15	Preston, 1962	Mode log 2
Birds	Kendeigh, 1946		24	3	Preston, 1962	Neg. contagion
Birds	Kendeigh, 1946		17	3	Preston, 1962	Neg. contagion
Birds	Kendeigh, 1946		18	3	Preston, 1962	Neg. contagion
Gastropods	Kohn, 1959	134	4		Preston, 1962	Neg. contagion
Gastropods	Kohn, 1959	182	9		Preston, 1962	Neg. contagion
Plants	Oosting, 1942		15		Preston, 1962	Pos. contagion
Plants	Oosting, 1942		10		Preston, 1962	Pos. contagion
Herons	Mayfield	1710	3	1	Preston, 1962	Pos. contagion
Herons	Mayfield		2	1	Preston, 1962	Pos. contagion
Birds	Belknap, 1951	>1000	6	1	Preston, 1962	Pos. contagion
Birds	Belknap, 1951	>1700	5	1	Preston, 1962	Pos. contagion
Passerines	Henderson	~55,555	5	1	Preston, 1962	Pos. contagion
Diatoms	Patrick, 1954	~2,000	68		Patrick, 1963 Preston, 1980	Mode log 2
Diatoms	Patrick, 1954	~3500	~150		Patrick, 1963 Preston, 1980	Mode log 2
Diatoms	Patrick, 1954	>40,000	55		Patrick, 1963 Preston, 1980	Mode log 2
Insects	Garthside, 1928	5186	399		Williams, 1964	No mode arith
Insects	Garthside, 1928	5665	48		Williams, 1964	No mode arith
Insects	Graham, 1933		221		Williams, 1964	No mode arith
Insects	Graham, 1933		126		Williams, 1964	No mode arith
Butterflies	Corbet, 1942	9031	620		Williams, 1964 Bulmer, 1974	No mode arith Poss. Zero term
Lepidoptera	Taylor and Carter, 1961	1148	36	3	Williams, 1964	Singletons and doubletons equal
Sphingidae	Williams, 1954	65	19	1/365	Williams, 1964	No mode arith
Sphingidae	Williams, 1954	92	21	1/365	Williams, 1964	No mode arith
Diptera	Haileybury College, 1951	192	28		Williams, 1964	No mode arith

Organisms	Source	Total Ind.	Total Spp.	Yrs.	Cited in	Notes
Coleoptera	Easton, 1947	3102	273	1/365	Williams, 1964	No mode arith
Coleoptera	Easton, 1947	957	150	1/365	Williams, 1964	No mode arith
Carabidae	Den Boer, 1958	8113	63	1	Williams, 1964	No mode arith
Diptera	Laurence, 1955	3035	25	1	Williams, 1964	Mode log 3
Snakes	Dunn and Allandoerfer, 1949	1232 3914 3041 2500	30 52 44 53		Williams, 1964	Mode log 3 Mode log 3 Mode log 3 Mode log 3
Birds	Illinois CBC, 1954-1957	1,145,000	146	4	Williams, 1964	Mode log 3 for each yr. and tot.
Birds	Ennion, 1960	23,934	140	8	Williams, 1964	No mode arith Mode log 3
Algae	Margalef, 1949	2348	355		Williams, 1964	No mode arith
Fungus	Hyde and Williams, 1949	2302	17gen	1	Williams, 1964	No mode arith
Beetles and Bugs		125 data sets		1/365	Janzen, 1973	3 had a mode arithmetically
Insects	Taylor, 1973 Taylor and French, 1973	16 data sets 80-9572	32- 262	1 each	Kempton and Taylor, 1974	None with an arithmetic mode
Insects	Bliss, 1965				Kempton and Taylor, 1974	No mode log 2
Birds	Audubon, 1970- 1978	Several data sets			Preston, 1980	Mode log 2 for most
Birds	D. R. Pfoutz, pers. comm.	~700	11		Preston, 1980	Double octave Complete log 2
Birds	L. Steiner, pers. Comm.	~250	18		Preston, 1980	Double octave Complete log 2
Marine invertebrates		3 sets of his own data			Hughes, 1984	One w/ possible mode log 2
Marine invertebrates	Moore, 1973	2 data sets			Hughes, 1984	No mode log 2
Marine invertebrates	Austin et al., 1980				Hughes, 1984	No mode log 2
Marine invertebrates	Pearson, 1975				Hughes, 1984	Mode log 2
Marine invertebrates	Gage, 1972	2 data sets			Hughes, 1984	Mode log 2
Marine invertebrates	Rosenberg, 1974				Hughes, 1984	No mode log 2
Marine invertebrates	Joserson, 1981	2 data sets			Hughes, 1984	No mode log 2
Marine invertebrates	Buchanan and Warwick, 1975				Hughes, 1984	No mode log 2
Marine invertebrates	Probert, 1981				Hughes, 1984	No mode log 2
Chironomids	Tokeshi, 1985				Tokeshi, 1990	
Dung beetles		32,753	5		Davis, 1993	Log-series
Dung beetles		7189	6		Davis, 1993	Log-series
Dung beetles		44	4		Davis, 1993	Log-series
Dung beetles		48	7		Davis, 1993	Log-series

Organisms	Source	Total Ind.	Total Spp.	Yrs.	Cited in	Notes
Dung beetles		400	16		Davis, 1993	Log-normal
Dung beetles		4894	20		Davis, 1993	Log-series
Dung beetles		5156	34		Davis, 1993	Log-series
Dung beetles		7506	42		Davis, 1993	Log-normal
Dung beetles		11,847	9		Davis, 1993	Log-series
Dung beetles		549	7		Davis, 1993	Log-series
Dung beetles		356	6		Davis, 1993	Log-series
Dung beetles		8571	33		Davis, 1993	Log-series
Dung beetles		15,530	77		Davis, 1993	Log-series
Dung beetles	Doube et al., 1991	131,196	5		Davis, 1993	Log-series
Dung beetles	Doube et al., 1991	133,488	5		Davis, 1993	Log-series
Dung beetles	Lumaret & Kirk, 1991	2946	11		Davis, 1993	Log-series
Butterflies	Hinchliff, 1994 37 Oregon cnys	Tot 11,994 (64 – 1053)	214		Fagan and Kareiva, 1997	Lognormal fit to 28 of 37 cnys
Butterflies		6690	130		DeVries et al., 1997	Mode log 3

Literature Review – Looking for Realistic Lognormal Parameters

Some time ago, before I had accomplished the major literature review above, I attempted to simulate sampling from log-normally-distributed communities. A problem with this plan stalled me at the very beginning. I did not know what the likely lognormal parameters (μ , σ^2) were for biological systems. I predicted it would be difficult to find parameters estimated by other researchers in the literature from large samples so I could get an idea of the range of biologically realistic μ 's and σ 's. What I did not realize, was that there is nearly a complete absence of reported lognormal fit parameter estimates in papers concerning abundance relationships. There are many papers which report the lognormal model as the best fit to at least a subset of their species abundance data, but very few in which the parameter estimates of that lognormal fit are reported. In fact, rarely are the parameter estimates mentioned at all, even in theoretical literature.

Dozens of authors, of course, attempted to fit a variety of models to their data. Many made the determination of which model best fit their data, citing the log-series or lognormal or others as the best fit. Many methods of comparison were used by these authors: mean squared error, AIC values, Chi-squared tests, and Kolmogorov-Smirnov tests. Yet, nearly none of the papers I read included in their write-up the parameters of the distributions that applied to their data. Even the parameters for the model of best fit were excluded. I tried writing to several of these authors to obtain the information, but was unsuccessful in each case and did not pursue this for every author.

The first direct source I found was a small paragraph in a textbook, *Ecology* by Robert E. Ricklefs (Ricklefs, 1990). He mentioned that Preston (Preston, 1948) found dispersion values of 2.3 for birds and 3.1-4.7 for moths; and that Patrick *et al.* (Patrick *et al.*, 1954) calculated values of 2.8-4.7 for diatoms. I had always assumed dispersion and variance were synonyms, but when I reread the papers I discovered that this was not the case. Preston wrote of Gaussian distributions with the variance in terms of the variable a , the dispersion factor mentioned above. In later papers, he supplied the formula $\sigma = 1/\sqrt{2a^2}$ to calculate the variance of these curves though this conversion was fairly evident from the formulae. I calculated σ^2 's from information given in each of the papers. The estimated σ^2 's for the Gaussian distributions in Preston's paper lie between 9.70-21.64 and for Patrick's are between 9.36-20.07. The important thing to note about these variance estimates is that they refer, not to variance of abundance, but variance of octaves

about the mode octave, which is a very loose interpretation of variance, in my opinion. Preston in later papers, estimated Gaussian distributions with variances of 11.02 (Preston, 1957), 30.2 (Preston, 1958) and 13.85 (Preston, 1960), though he felt the middle estimate was biased high from non-random sampling. In the Canonical Lognormal paper, Preston (Preston, 1962a) listed two additional estimates from birds in England & Wales (Fisher, 1952) and Finland (Marikallio, 1958) as 25.5 and 23.4, respectively. We must keep in mind that both of these authors were using a log₂ base for their plots of abundances.

More estimates of lognormal variance appear in Preston's 1962 paper, but are coupled with his Canonical Lognormal theory. Since the number of species determines the variance, in this theory, estimates of the variances increase from 13.8 to 49.8 as the number of species in the sample increase from 100 to one million. Preston called these the parameters of the canonical ensemble.

The best source of parameter estimates I have found has been William's (Williams, 1964) book *Patterns in the Balance of Nature*. Chapter 3 has many references to datasets (Table 1), most of which have an estimated standard deviation. In this text, the octaves are calculated with log base 3, and estimates are in terms of base 3 unless otherwise noted. For moths at Rothamsted the estimated mean and standard deviation is 0.65 ± 0.85 . Maine lepidopterans had estimates 1.25 ± 0.92 . British nesting birds were estimated to have 3.75 ± 1.55 with a log base of 10. The birds of Quaker Run were found to have 1.65 ± 0.8 , and four years worth of the Illinois Christmas bird counts were found to have estimated parameters of 1.30 ± 1.3 .

Kempton and Taylor (Kempton & Taylor, 1974) give a list of estimated μ 's and σ 's of Poisson lognormal distributions fit to moth sampling data. Their estimates of μ varied between -4.38 and 2.16. The negative μ 's must refer to data without a mode (if they do not subscribe to Preston's binning method this is possible). Estimated σ 's ranged from 0.99-2.98, which translates into variances of 0.98-8.88. The authors stated that the discrete Poisson lognormal is more appropriate in cases where the species are clumping on the left side of the curve. This is a very different range of values for μ and σ than we saw from the other two authors who used the continuous lognormal.

The last paper which I found mention of estimates of lognormal parameters was DeVries et al. (DeVries *et al.*, 1997). They reported an estimated mean = 1.852 and variance = 3.278 in log₃ base for their fruit feeding butterflies. I wish all authors had reported as much.

One additional paper mentions estimates of parameters, but not for the lognormal distribution. Brian (Brian, 1953) included a list of estimates for parameters of the negative binomial distribution. He listed six community types with associated species number, number of specimens, and 3 estimated constants of the distribution.

So, I must conclude, that for the purposes of finding appropriate bounds on parameters for Monte Carlo sampling, this literature review was not nearly as productive as I had hoped. What the review did accomplish, however, was to allow me to see the linear progression of theory development. During my first reading of the literature, I had approached it from the opposite end, temporally. Starting with the most recent literature, I worked my way backward through relevant citations. It helped me immensely to see the development of ideas and at the same time have a clear idea on what background those ideas were based. This chronological synthesis also helped me see the direction I would like to take my own research. I discuss some specifics of this in the last section.

Conclusions

In the Sections above, there has been little clear support for the lognormal distribution as a good model for species abundance data. The literature appeared to be divided, with ecologists accepting and modeling mechanisms to support the idea while statisticians discarded it and began looking in several other directions. The parameters published for the lognormal were few and far between, most from nearly a half century ago. The data that has been used over the decades in support of the lognormal and other descriptive distributions tended to be inconclusive (Table 1). Some seemed to support the lognormal, while other samples supported log-series, negative binomial, Poisson lognormal, and geometric distributions. Most of the data showed a humped curve, when plotted in \log_2 octaves, giving a location to the mode that is so important in estimating species richness from the lognormal distribution. It is this data, which could be construed as strong support for lognormal abundances, which is the most misleading.

The mode in the lognormal plot is *not* indicative of a decrease in the rate at which new species are being added to the sample. I will restate the fact that I have a very significant disagreement with Preston's method of determining which species are included in which octave (see Literature Review for details).

I am not going to reject the lognormal as a potentially descriptive model of species abundance, but I have decided while completing this paper to discard it as a valid and true model of abundances. It is simply too contrived and requires too much manipulation to be of use to either ecologists or statisticians. I would like to pursue the fitting of discrete distributions to abundance data. The negative binomial distribution deserves more attention, in my opinion. The Monte Carlo simulations, with which this whole project started, can be looked at again with several candidate models for the universe of the population.

Ultimately, my research goals are to model the parameters of the distributions that look appropriate, using long-term field collection data (i.e. selected Breeding Bird Survey and Audubon Christmas bird counts) as the dependent abundances and environmental covariates as the independent variables. Once a model for the underlying universe is constructed, I should be able to predict the universe at other locations, making estimates of species richness and evenness possible. Additionally, I propose combining the distributions predicted from the environmental variables for a given site with existing data from that site, within an empirical Bayesian framework, to increase the precision of the prediction. I

believe that the predicted model of the relative abundances would make a good prior distribution, and, coupled with the data from the site, a much more precise posterior distribution could result.

Literature Cited

- Abramowitz, M. & Stegun, I. (1970). *Handbook of Mathematical Functions*, Washington, D. C.
- Agrell, I. (1941). Zur Ökologie der Collembolen: Untersuchungen im Schwedischen Lappland. *Opus. Ent.* **3**, 236.
- Aitchison, J. & Brown, J. A. C. (1957). *The Lognormal Distribution*. Cambridge University Press, Cambridge, England.
- Austin, A. D., Austin, S. A. & Sale, P. F. (1980). Community structure of the fauna associated with the coral *Pocillopora damicornis* (L.) on the Great Barrier Reef. *Australian Journal of Marine Freshwater Res.* **31**, 163-174.
- Bastow, W. J. (1991). Methods for fitting dominance/diversity curves. *Journal of Vegetation Science* **2**, 35-46.
- Bauer, A. M. & Sadler, R. A. (1993). Systematics, biogeography and conservation of the lizards of New Caledonia. *Biodiversity Letters* **1**, 107-122.
- Belknap, J. B. (1951). Census of breeding birds on an island in an open lake. *Audubon Field Notes* **5**, 327-328.
- Black, G. A., Dobzhansky, T. & Pavan, C. (1950). Some attempts to estimate the species diversity and population density of trees in Brazilian rainforests. *Bot. Gaz.*
- Bliss, C. I. (1965). An analysis of some insect trap records. In *Classical and Contagious Discrete Distributions* (ed. G. P. Patil), pp. 385-397. Statistical Publishing Society, Calcutta.
- Brian, M. V. (1953). Species frequencies in random samples from animal populations. *Journal of Animal Ecology* **22**, 57-64.
- Brown, J. H., J. Bedrick, E., Ernest, S. K. M. & Cartron, J.-L. E. (in prep.). Constraints on negative relationships: mathematical causes and ecological consequences.
- Brown, J. H. & Lomolino, M. V. (1998). *Biogeography*, Second edition. Sinauer Associates, Inc., Sunderland, Massachusetts.
- Buchanan, J. B. & Warwick, R. M. (1975). An estimate of benthic macrofaunal production in the offshore mud of the Northumberland coast. *Journal of Marine Biological Assessment, U.K.* **54**, 197-222.
- Bulmer, M. G. (1974). On fitting the Poisson lognormal distribution to species-abundance data. *Biometrics* **30**, 101-110.
- Colwell, R. K. & Coddington, J. A. (1994). Estimating terrestrial biodiversity through extrapolation. *Philosophical Transactions of the Royal Society of London, Series B* **345**, 101-118.
- Condit, R., Hubbell, S. P., LaFrankie, J. V., Sukumar, R., Manokaran, N., Foster, R. B. & Ashton, P. S. (1996). Species-area and species-individual relationships for tropical trees: a comparison of three 50-ha plots. *Journal of Ecology* **84**, 549-562.
- Connell, J. H. & Orias, E. (1964). The Ecological Regulation of Species Diversity. *The American Naturalist* **98**, 399-414.
- Cook, R. R. (1995). The relationship between nested subsets, habitat subdivision, and species diversity. *Oecologia* **101**, 204-210.
- Corbet, A. S. (1942). The distribution of butterflies in the Malay Peninsula. *Proceedings of the Royal Entomological Society of London A.* **16**, 101-116.
- Cotgreave, P. & Harvey, P. H. (1994). Associations among biogeography, phylogeny and bird species diversity. *Biodiversity Letters* **2**, 46-55.
- Davis, A. L. V. (1993). Alpha-diversity patterns of dung beetle assemblages (Coleoptera: Scarabaeidae, Aphodiidae, Staphylinidae, Histeridae, Hydrophilidae) in the winter rainfall region of South Africa. *African Entomology* **1**, 67-80.
- DeJong, T. M. (1975). A comparison of three diversity indices based on their components of richness and evenness. *Oikos* **26**, 222-227.
- Den Boer, P. J. (1958). Activiteitsperioden van loopkevers in Meijendal. *Ent. Ber., Amst.* **18**, 80-89.
- DeVries, P. J., Murray, D. & Lande, R. (1997). Species diversity in vertical, horizontal, and temporal dimensions of a fruit-feeding butterfly community in an Ecuadorian rainforest. *Biological Journal of the Linnean Society* **62**, 343-364.
- Diamond, J. M. & May, R. M. (1981). Island Biogeography and the Design of Natural Reserves. In *Theoretical Ecology* (ed. R. M. May), pp. 489. Sinauer Associates, Inc., Massachusetts.

- Dirks, C. O. (1937). Biological studies of Maine moths by light trap methods. *The Maine Agricultural Experiment Station Bulletin* 389.
- Dobzhansky, T. (1950). Evolution in the Tropics. *American Scientist* 38, 208-221.
- Doube, B. M., MacQueen, A., Ridsdill-Smith, T. J. & Weir, T. A. (1991). Native and introduced dung beetles in Australia. In *Dung Beetle Ecology* (ed. I. Hanski and Y. Cambefort), pp. 255-278. Princeton University Press, Princeton.
- Dunn, E. R. & Allendoerfer, C. B. (1949). The application of Fisher's formula to collections of Panamanian snakes. *Ecology* 30, 533-536.
- Easton, A. M. (1947). The Coleoptera of flood refuse: a comparison of samples from Surrey and Oxfordshire. *Ent. mon. Mag.* 83, 113-115.
- Efron, B. & Thisted, R. (1976). Estimating the number of unseen species: How many words did Shakespeare know? *Biometrika* 63, 435-47.
- Engen, S. (1974). On species frequency models. *Biometrika* 61, 263-270.
- Engen, S. (1975). A note on the geometric series as a species frequency model. *Biometrika* 62, 697-699.
- Engen, S. (1977). Comments on Two Different Approaches to the Analysis of Species Frequency Data. *Biometrics* 33, 205-213.
- Engen, S. (1978). *Stochastic Abundance Models with emphasis on biological communities and species diversity*. Chapman and Hall, London.
- Ennion, E. (1960). *The House on the Shore*. Routledge and Kegan Paul, London.
- Fagan, W. F. & Kareiva, P. M. (1997). Using compiled species lists to make biodiversity comparisons among regions: a test case using Oregon butterflies. *Biological Conservation* 80, 249-259.
- Fisher, J. (1952). Bird numbers. *South Eastern Naturalist and Antiquary* 57, 1-10.
- 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* 11, 42-58.
- Gage, J. (1972). Community structure of the benthos on Scottish sea-lochs. 1. Introduction and species diversity. *Marine Biology* 14.
- Garthside, S. (1928). Quantitative studies upon the insect fauna of Jack Pine environments. Unpublished thesis, University of Minnesota.
- Ghent, A. W. (1991). Insights into Diversity and Niche Breadth Analyses from Exact Small-sample Tests of the Equal Abundance Hypothesis. *American Midland Naturalist* 126, 213-255.
- Gleason, H. A. (1922). On the relation between species and area. *Ecology* 3, 158-162.
- Graham, S. A. (1933). The influence of civilisation on the insect fauna of forests. *Annals of the Entomological Society of America* 26, 497-503.
- Hairston, N. G. (1959). Species Abundance and Community Organization. *Ecology* 40, 404-416.
- Harvey, P. A. & Godfray, H. C. J. (1987). How species divide resources. *The American Naturalist* 129, 318-320.
- Heck, K. L., Belle, G. V. & Simberloff, D. (1975). Explicit calculation of the rarefaction diversity measurement and the determination of sufficient sample size. *Ecology* 56, 1459-1461.
- Henderson, N. (1956). A blackbird roost in Oberlin. *Bulletin of the Cleveland Audubon Society* 2, 13-14.
- Hicks, L. E. (1935). A ten year study of a breeding bird study in central Ohio. *American Midland Naturalist* 16, 177-186.
- Hill, M. O. (1973). Diversity and Evenness: a unifying notation and its consequences. *Ecology* 54, 427-432.
- Hinchliff, J. (1994). *An Atlas of Oregon Butterflies*. Oregon State University Press, Corvallis, OR.
- Hughes, R. G. (1984). A model of the structure and dynamics of benthic marine invertebrate communities. *Marine Ecology Progress Series* 15, 1-11.
- Hughes, R. G. (1985). A hypothesis concerning the influence of competition and stress on the structure of marine benthic communities. In *Nineteenth European Marine Biology Symposium*. Cambridge University Press, Plymouth, Devon, U. K.
- Hughes, R. G. (1986). Theories and models of species abundance. *The American Naturalist* 128, 879-899.
- Hurlbert, S. H. (1971). The nonconcept of species diversity: a critique and alternative parameters. *Ecology* 52, 577-586.
- Huston, M. A. (1994). *Biological Diversity, the Coexistence of Species in Changing Landscapes*. Cambridge University Press, Cambridge.
- Hyde, H. A. & Williams, D. A. (1949). A census of mould spores in the atmosphere. *Nature* 164, 668-669.

- Janzen, D. H. (1973). Sweep Samples of Tropical Foliage Insects: description of study sites, with data on species abundances and size distributions. *Ecology* **54**, 659-686.
- Josefson, A. B. (1981). Persistence and structure of two deep macrobenthic communities in the Skaggerak (West coast of Sweden). *J. exp. mar. Biol. Ecol.* **50**, 63-97.
- Keating, K. A., Quinn, J. F., Ivie, M. A. & Ivie, L. L. (1998). Estimating the effectiveness of further sampling in species inventories. *Ecological Applications* **8**, 1239-1249.
- Kempton, R. A. & Taylor, L. R. (1974). Log-series and log-normal parameters as diversity discriminants for the Lepidoptera. *Journal of Animal Ecology* **43**, 381-399.
- Kendeigh, S. C. (1946). Breeding birds of the beech-maple-hemlock community. *Ecology* **27**, 226-245.
- Klopper, P. H. & MacArthur, R. H. (1961). On the causes of tropical species diversity: Niche overlap. *The American Naturalist* **95**, 223-226.
- Kohn, A. J. (1959). The ecology of *Conus* in Hawaii. *Ecological Monographs* **29**, 47-90.
- Laurence, B. R. (1955). The ecology of some British Sphaeroceridae (Boboridae, Diptera). *Journal of Animal Ecology* **24**, 187-199.
- Lobik, P. H. (1954). Illinois Christmas Bird Census. *Audubon (annu.) Bulletin* **93**, 6-12.
- Lobik, P. H. (1955). Illinois Christmas Bird Census. *Audubon (annu.) Bulletin* **97**, 6-12.
- Lobik, P. H. (1956). Illinois Christmas Bird Census. *Audubon (annu.) Bulletin* **101**, 8-9.
- Lobik, P. H. (1957). Illinois Christmas Bird Census. *Audubon (annu.) Bulletin* **105**, 5-13.
- Lumaret, J. P. & Kirk, A. A. (1991). South temperate dung beetles. In *Dung Beetle Ecology* (ed. I. Hanski and Y. Cambefort), pp. 97-115. Princeton University Press, Princeton.
- MacArthur, J. W. (1975). Environmental Fluctuations and Species Diversity. In *Ecology and Evolution of Communities* (ed. M. L. Cody and J. M. Diamond), pp. 545. Belknap Press of Harvard University, Massachusetts.
- MacArthur, R. (1960). On the relative abundance of species. *The American Naturalist* **94**, 25-36.
- MacArthur, R., Recher, H. & Cody, M. (1966). On the relation between habitat selection and species diversity. *The American Naturalist* **100**, 319-325.
- MacArthur, R. H. (1957). On the Relative Abundance of Bird Species. *Proceedings of the National Academy of Sciences* **43**, 293-295.
- MacArthur, R. H. & MacArthur, J. W. (1961). On Bird Species Diversity. *Ecology* **42**, 594-598.
- MacArthur, R. H. & Wilson, E. O. (1967). *The Theory of Island Biogeography*. Princeton University Press, Princeton, New Jersey.
- Magurran, A. E. (1988). *Ecological Diversity and Its Measurement*. Princeton University Press, Princeton, New Jersey.
- Margalef, R. (1949). Les asociaciones de algas en las aguas dulces de pequeno volumen del nor-este de Espana. *Vegetatio* **1**, 258-284.
- Marikallio, E. (1958). *Finnish birds, their distribution and numbers*. Societas pro Fauna et Flora Fennica, Helsinki.
- Maurer, B. & Taper, M. (pers. comm.). Gamma vs. Lognormal fit to BBS species abundances.
- May, R. M. (1975). Patterns of Species Abundance and Diversity. In *Ecology and Evolution of Communities* (ed. M. L. a. J. M. D. Cody), pp. 545. Belknap Press of Harvard University, Massachusetts.
- May, R. M. (1981). Patterns in Multi-Species Communities. In *Theoretical Ecology* (ed. R. M. May), pp. 489. Sinauer Associates, Inc., Massachusetts.
- Molles, M. C. (1999). *Ecology Concepts and Applications*. WBC McGraw-Hill, New York.
- Moore, P. G. (1973). The kelp fauna of Northeast Britain. 2. Multivariate classification, turbidity as an ecological factor. *J. exp. mar. Biol. Ecol.* **5**, 193-200.
- Oosting, H. (1942). An ecological analysis of the plant communities of Piedmont, North Carolina. *American Midland Naturalist* **28**, 1-26.
- Pagel, M. D., May, R. M. & Collie, A. R. (1991). Ecological aspects of the geographical distribution and diversity of mammalian species. *The American Naturalist* **137**, 791-815.
- Paine, R. T. (1966). Food web complexity and species diversity. *The American Naturalist* **100**, 65-75.
- Palmen, E. (1944). Die anemohydrochore Ausbreitung der Insekten als zoogeographischer Faktor. *Ann. Soc. Zool.-Bot. Vanamo* **10**, 1-262.
- Patil, G. P. & Taillie, C. (1977). Diversity as a concept and its implications for random communities. *Bulletin of the International Statistical Institute* **47**, 497-515.

- Patrick, R. (1963). The structure of diatom communities under varying ecological conditions. *Annals New York Academy of Sciences* **108**, 359-365.
- Patrick, R., Hohn, M. H. & Wallace, J. H. (1954). A new method for determining the pattern of the diatom flora. *Notulae Naturae of The Academy of Natural Sciences of Philadelphia* **259**, 1-9.
- Pearson, T. H. (1970). The benthic ecology of Loch Linnhe and Loch Eil, a sea-loch system on the west coast of Scotland. 1. The physical environment and distribution of the macrobenthic fauna. *J. exp. mar. Biol. Ecol.* **20**, 1-34.
- Peet, R. K. (1974). The measurement of species diversity. *Annual Review of Ecology and Systematics* **5**, 285-307.
- Peet, R. K. (1975). Relative Diversity Indices. *Ecology* **56**, 496-498.
- Pfoutz, D. R. (1979?). Bird visits to a feeder in Gibsonia, Pennsylvania.
- Pianka, E. R. (1966). Latitudinal Gradients in Species Diversity: a review of concepts. *The American Naturalist* **100**, 33-46.
- Pianka, E. R. (1967). On Lizard Species Diversity: North American Flatland Deserts. *Ecology* **48**, 333-351.
- Pielou, E. C. (1966). The Measurement of Diversity in Different Types of Biological Collections. *Journal of Theoretical Biology* **13**, 131-144.
- Pielou, E. C. (1975). *Ecological Diversity*. John Wiley & Sons, New York.
- Platt, H. M., Shaw, K. M. & Lamshead, P. J. D. (1984). Nematode species abundance patterns and their use in the detection of environmental perturbations. *Hydrobiologia* **118**, 59-66.
- Preston, F. W. (1948). The commonness, and rarity, of species. *Ecology* **29**, 254-283.
- Preston, F. W. (1957). Analysis of Maryland State-wide bird counts. *Maryland Birdlife* **13**, 63-65.
- Preston, F. W. (1958). Analysis of the Audubon Christmas Bird Counts in terms of the Lognormal Curve. *Ecology* **39**, 620-624.
- Preston, F. W. (1960). Time and Space and the Variation of Species. *Ecology* **41**, 611-627.
- Preston, F. W. (1962a). The canonical distribution of commonness and rarity: Part I. *Ecology* **43**, 185-215.
- Preston, F. W. (1962b). The canonical distribution of commonness and rarity: Part II. *Ecology* **43**, 410-432.
- Preston, F. W. (1980). Noncanonical distributions of commonness and rarity. *Ecology* **61**, 88-97.
- Preston, F. W. & Norris, R. T. (1947). Nesting heights of breeding birds. *Ecology* **28**, 241-273.
- Probert, P. K. (1981). Changes in the benthic community of china clay waste deposits in Mevagissey Bay following a reduction in discharge. *Journal of Marine Biological Assessment, U.K.* **61**, 789-804.
- Ricklefs, R. E. (1990). *Ecology*, Third edition. W. H. Freeman and Company, New York.
- Ricklefs, R. E. & Schluter, D. (1993). Species Diversity in Ecological Communities: Historical and Geographical Perspectives, pp. 414. The University of Chicago Press, Chicago.
- Robbins, M. C. (1948). with Frank Preston about Maryland Statewide bird counts.
- Rosenburg, R. (1974). Spatial dispersion of an estuarine benthic faunal community. *J. exp. mar. Biol. Ecol.* **15**, 69-80.
- Rosenzweig, M. L. (1995). *Species diversity in space and time*. Cambridge University Press, Cambridge.
- Rosenzweig, M. L. & Abramsky, Z. (1993). How are diversity and productivity related? In *Species Diversity in Ecological Communities* (ed. R. E. Ricklefs and D. Schluter), pp. 52-65. The University of Chicago Press, Chicago.
- Salt, G., Hollick, F. S. J., Raw, F. & Brian, M. V. (1948). The Arthropod population of pasture soil. *Journal of Animal Ecology* **17**, 139-150.
- Scott, F. A. (1953). Christmas count totals from North America 1949 - 1952. *Audubon Field Notes* **7**, 327-330.
- Scott, F. A. (1956). Christmas count totals for 1955. *Audubon Field Notes* **10**, 437-440.
- Shmida, A. & Wilson, M. V. (1985). Biological determinants of species diversity. *Journal of Biogeography* **12**, 1-20.
- Simpson, E. H. (1949). Measurement of Diversity. *Nature* **163**, 688.
- Simpson, G. G. (1964). Species Density of North American Recent Mammals. *Systematic Zoology* **13**, 57-73.
- Smith, W. & Grassle, J. F. (1977). Sampling Properties of a Family of Diversity Measures. *Biometrics* **33**, 283-292.
- Society, N. A. (1970). Christmas bird count, 1969. *Audubon Field Notes* **24**.
- Society, N. A. (1971). Christmas bird counts, 1970. *American Birds* **25**.
- Society, N. A. (1972). Christmas bird counts, 1971. *American Birds* **26**.

- Society, N. A. (1973). Christmas bird counts, 1972. *American Birds* 27.
- Society, N. A. (1974). Christmas bird counts, 1972. *American Birds* 28.
- Society, N. A. (1975). Christmas bird counts, 1974. *American Birds* 29.
- Society, N. A. (1976). Christmas bird counts, 1975. *American Birds* 30.
- Society, N. A. (1977). Christmas bird counts, 1976. *American Birds* 31.
- Society, N. A. (1978). Christmas bird counts, 1977. *American Birds* 32.
- Steiner, L. (1979?). Bird visits to a feeder on a farm near Slippery Rock, Pennsylvania.
- Stewart, R. E. & Aldrich, J. W. (1949). Breeding bird populations in the spruce region of the Central Appalachians. *Ecology* 30, 75-82.
- Stilling, P. (1999). *Ecology Theories and Applications*, Third edition. Prentice Hall, Upper Saddle River, New Jersey.
- Sugihara, G. (1980). Minimal Community Structure: an explanation of species abundance patterns. *The American Naturalist* 116, 770-787.
- Sugihara, G. (1989). How Do Species Divide Resources? *The American Naturalist* 133, 458-463.
- Taper, M. L. & Marquet, P. A. (1996). How do species really divide resources? *The American Naturalist* 147, 1072-1086.
- Taylor, L. R. (1973). Monitor surveying for migrant insect pests. *Outl. Agric.* 7, 109-116.
- Taylor, L. R. & Carter, C. I. (1961). The analysis of numbers and distribution in an aerial population of macro-lepidoptera. *Transactions of the Royal Entomological Society, London* 113, 369-386.
- Taylor, L. R. & French, R. A. (1973). Rothamsted light trap. *Report for Rothamsted Experiment Station* 1972, Part 2, 182-211.
- Tokeshi, M. (1985). Life-cycle and production of the burrowing mayfly, *Ephemera danica*: a new method for estimating degree-days required for growth. *Journal of Animal Ecology* 54, 919-930.
- Tokeshi, M. (1990). Niche apportionment or random assortment: species abundance patterns revisited. *Journal of Animal Ecology* 59, 1129-1146.
- Tokeshi, M. (1996). Power fraction: a new explanation of relative abundance patterns in species-rich assemblages. *Oikos* 75, 543-550.
- Usher, M. B., Field, J. P. & Bedford, S. E. (1993). Biogeography and diversity of ground-dwelling arthropods in farm woodlands. *Biodiversity Letters* 1, 54-62.
- Walkinshaw, L. H. (1947). Breeding birds of brushy field, woodlots and pond (10 year summary). *Audubon Field Notes* 1, 214-216.
- Watterson, G. A. (1974). Models for the logarithmic species abundance distributions. *Theoretical Population Biology* 6, 217-250.
- Whittaker, R. H. (1972). Evolution and measurement of species diversity. *Taxon* 21, 213-251.
- Williams, A. B. (1947). Breeding birds of a climax beech-maple forest, 15 year summary. *Audubon Field Notes* 1, 205-210.
- Williams, C. B. (1943). Area and number of species. *Nature* 152, 264-267.
- Williams, C. B. (1953). The relative abundance of different species in a wild animal population. *Journal of Animal Ecology* 22, 14-31.
- Williams, C. B. (1954). Notes on a small collection of Sphingidae from Nigeria. *Nigerian Field* 19, 176-179.
- Williams, C. B. (1964). *Patterns in the Balance of Nature*. Academic Press, New York.