

Species Abundance Distribution

Evans Otieno Ochiaga (evansochiaga@aims.ac.za)
African Institute for Mathematical Sciences (AIMS)

Supervised by: Dr Cang Hui
Stellenbosch University, South Africa

22 May 2014

Submitted in partial fulfillment of a structured masters degree at AIMS South Africa



Abstract

Species abundance distribution (SAD) is one of the major ecological metrics. SAD curves take a hollow curve or hyperbolic shape on a histogram plot with many rare species and just a few common species. In general, the shape of SAD is largely lognormally distributed. The forces that drive SAD to take this shape still remains elusive. However, this project is aimed at deriving some parametric and mechanistic approaches that best describe species abundance distribution. In addition, the theories derived are Maximum Entropy Theory of Ecology, Neutral theory and briefly discussed my theory of proportionate effect which also describes SAD even though it is still being developed for better modelling of SAD. Parametric approaches discussed in this project include; log series, negative binomial, lognormal and geometric distributions.

Declaration

I, the undersigned, hereby declare that the work contained in this research project is my original work, and that any work done by others or by myself previously has been acknowledged and referenced accordingly.



Evans Otieno Ochiaga, 22 May 2014

Contents

Abstract	i
1 Introduction	1
1.1 Ecological background	1
1.2 Species Abundance Distribution (SAD)	1
1.3 Factors affecting species abundance distribution	3
1.4 Application of SAD	3
1.5 Importance of this project	3
2 Parametric Approach to Species Abundance Distribution	4
2.1 The Logarithmic Series or Log series Distribution	5
2.2 Lognormal distribution	7
2.3 Negative binomial distribution	10
2.4 Geometric distribution	11
2.5 Evaluating parametric approach	12
3 Mechanistic approach to Species Abundance Distribution	14
3.1 Maximum Entropy Theory of Ecology (METE)	14
3.2 Neutral theory	21
3.3 The Theory of Proportionate effect	28
4 Conclusion and Future Work	31
4.1 Conclusion	31
4.2 Future work	31
References	34

1. Introduction

1.1 Ecological background

Ecology is one of the major branches of biology which plays a central role as far as the study of living organisms and their relationship with environment is concerned [Odum and Barrett \(1953\)](#). The study of Ecology is characterised by the study of a defined ecological community which is habitat for species of different kinds. What is more in a given ecological community is that, species that constitute the community varies as far as their abundance is concerned, generally there are many rare species with relatively few common species [Bowler and Kelly \(2010\)](#). For clarification, rare species are those species represented by few individuals in the community i.e. singleton species, on the other hand common species are those species that are represented by many individuals in the community. Furthermore, factors affecting population size of a given species in the community combine themselves as a product and not as a sum and this imply that absence of one factor may negatively impact on the population size of the species [Harte \(2011\)](#).

A clear picture of species abundance in the ecological communities is understandable by considering the following example. Lets assume that you are walking through a forest, you will observe that there are different species of different kinds of trees growing in groups with different population sizes. It is clear that between these different population sizes of tree species that is commonness and rarity, there lies a probability distribution called Species Abundance Distribution (SAD), which is one of the most important metrics in ecology to date. SAD helps in understanding and describing how species are distributed in our communities [Harte \(2011\)](#).

Furthermore, the following example can also give insight as far as species abundance distribution is concerned. Consider a basket full of red and yellow balls and assume that we can't tell the abundance of each ball in the basket. If a ball is picked at random without replacement from the basket and the color of the ball is recorded, then after several picking you will end up with abundance of each ball based on their colors. Then the question is, what is the distribution of the abundance of these balls in the basket based on their colors. When the balls are replaced by species in the community and the question of the abundance is answered then this will give a clear picture of what SAD is all about [Harte \(2011\)](#). The next section discusses into details what species abundance distribution is all about.

1.2 Species Abundance Distribution (SAD)

The pattern in rarity and commonness of species in the ecological community has been center for discussion for several years by ecologist [Volkov et al. \(2005\)](#). Before further discussion lets first understand what we mean by 'species'. Species is defined as a group of living organism consisting of similar individuals who are capable of exchanging genes and interbreeding in a given environment, on the other hand abundance is the total number of individuals of a given species in ecological community. As explained before, in a given community of species, it is observed that species varies in terms of their abundances. Therefore the question we need to ask ourselves is, what is species abundance distribution (SAD)?

SAD is defined as a description of abundance of species in a given community [Harte \(2011\)](#). It can also be described as a community-level metric denoted by $\langle \Phi(n) \rangle$ that tells us the probability of a given species having n abundance [Sognnæs \(2011\)](#). In addition, SAD is referred to us as a community level

metric since it deals with the collection of individuals belonging to different species in the community. In general, definition of species abundance distribution is based on the pattern of commonness and rarity of species in the community and this pattern is best explained using a vector of comparable abundances representing a given community [McGill and Magurran \(2011\)](#), where a vector is used in SAD to imply that there exist some abundance measures for multiple species.

Species abundance is characterised by variation with changes in the size of a given defined piece of land in which different species inhabit, this variation is due to the fact that a large piece of land contains adequate environmental conditions that contribute positively to species abundance, for example; good slope, soil moisture and migration in and out of those their habitats [Harte \(2011\)](#). Before further discussion lets understand what community means, community is a group of all organism that belong to a given taxonomic group in a chosen area, where taxonomic group can be described as a family, order or class of species in which ecologist is interested in studying [Pielou \(1977\)](#). A typical SAD shape is right skewed or it can be described as truncated normal distribution with truncated veil line being at one since for any ecological community it is clear that there is no species with less than one individual. Furthermore, in order to clearly bring the picture of SAD shape for easy understanding we are going to have a look at some simple plot for species abundance distribution data as illustrated in the next section.

1.2.1 Illustrative example of SAD. In illustrating SAD, using R a [BCI \(2006\)](#) dataset was used for different species of trees growing in a forest to plot a SAD histogram as shown below.

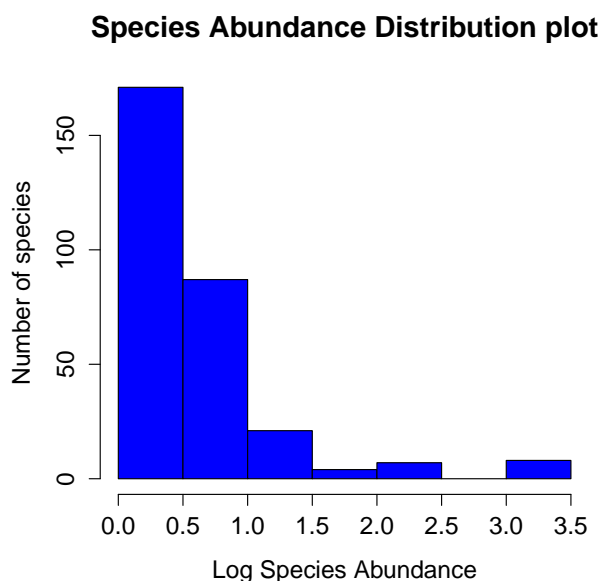


Figure 1.1: Species Abundance Distribution histogram plot for free-standing woody trees in a 50 ha plot of Barro Colorado Island (BCI) for the year 1985 [BCI \(2006\)](#).

From figure (1.1), we can observe that it is right skewed plot and this gives a clear picture of SAD, where rare species are many as compared to common species in the community. This histogram takes the exact shape of log normal distribution when the scaling of lognormal plot is done from $x = 1$ in x-axis and this implies that SAD follows lognormal distribution. Indeed several approaches have been put in place to model SAD, for example; log-series, geometric, negative binomial and log-normal distributions [Pielou](#)

(1977). This project is aimed at examining and evaluating a few parametric approaches that is used in describing SAD shape and further give insight in some theories that are fundamental in describing SAD.

1.3 Factors affecting species abundance distribution

Species Abundance Distribution mainly depend on the adaptation of a given species to its surrounding. These adaptations are based on the conditions of both abiotic and biotic environments in which the species inhabit. The two major factors that affect species abundance distribution is explained below.

1.3.1 Biotic factors. Biotic factor is defined as a factor created by living thing or any component of living organism within an environment in which the action of one organism affects the life of the other organism, for example predator consuming its prey. If this is the case then as time increases the population of preys will become extinct where as the population of species belonging to predator group will increase upto to some boundary limit and in the long run this will lead to high abundance of predators as compared to their preys.

1.3.2 Abiotic factors. Abiotic factor is defined as a non-living physical and chemical attributes of a given system, for example; climatic condition, soil PH and pressure in the environment. There are many abiotic factor that affects SAD, i.e. when we consider climate and soil we find that both affects the type of plant to grow in a given area. In addition, since plants acts as a source and shelter for other organism then they will also dictate the species of organism to be found in a given ecological community.

1.4 Application of SAD

SAD plays a major role in understanding different ecological phenomena as far as ecosystem is concerned. The following are some areas where study of SAD is of important application as explained by McGill and Magurran (2011).

- It plays a key role in testing models or theories of how communities are assembled.
- It helps in pinpointing rare and common species in the community for management purposes and justifying why some species are rare while others are common.
- It is used in comparing two or more locations, for example how do the SAD of two or more communities along a productivity gradient or along along a human disturbance gradient changes.

1.5 Importance of this project

Many people do not understand how species are organised in our ecosystem, upto date there is a believe in most of the people that common species are the most abundant. This project will shade some light to the entire community as far as their understanding of species abundance is concerned by making people understand that rare species in a community are not few as we think. The project also discusses various approaches describing SAD apart from log-normal distribution. This project is organised as follows; chapter two discusses parametric approaches that has been put in place in describing SAD with chapter three discussing and deriving some theories that can as well describe the skewed shape of SAD. Finally conclusion of the whole project is presented in chapter four.

2. Parametric Approach to Species Abundance Distribution

Parametric approach is one of the major approaches used in describing species abundance distribution. This approach is based on the use of some parametric measures of SAD. Parametric measures are defined as parameters of mathematically described probability distribution that can be interpreted based on some assumptions of a given probability distribution that is applied to a given data [McGill and Magurran \(2011\)](#). This chapter discusses some parametric measures that has been put in place as far as SAD is concerned. In order to understand parametric measures of SAD, then lets consider a given ecological community of species having N individuals belonging to S species, in this community there are different species having different number of individuals, that is each distinct species has N_1, N_2, \dots, N_i individuals. The following is the diagrammatic representation of the ecological community having three species.

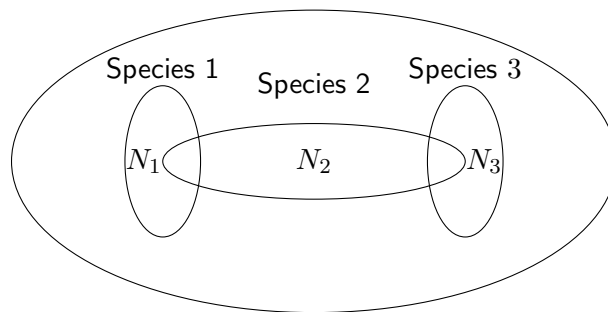


Figure 2.1: Diagrammatic representation of ecological community having three species. This community can be generalize for all the species in the environment, that is S_1, S_2, \dots, S_i species with N_1, N_2, \dots, N_i individuals.

The main interest of ecologist in studying species is to understand species abundance distribution and this is only possible if a sample of species is collected from the parent population and their abundance distribution is studied after which we can generalise about the species abundance distribution of the entire community. Therefore this calls for random sampling of some species from community (2.1). This sampling is based on the assumption that species are randomly distributed in community (2.1). In addition, when we randomly collect a sample from community (2.1), then the number of individuals that will represent j^{th} species in the collection will follow poisson distribution with parameter λ_j , where parameter λ_j is the mean abundance of the j^{th} species. It now follows that the probability of j^{th} species being represented by r individuals in the sample is defined as follows [Pielou \(1977\)](#).

$$\text{Pr}(j^{\text{th}} \text{ species is represented by } r \text{ individuals}) = e^{-\lambda_j} \frac{\lambda_j^r}{r!} \quad (2.0.1)$$

The probability in (2.0.1) is specific for j^{th} species. Now suppose that we are interested in studying species abundance distribution for the whole community, then the probability shown in (2.0.1) will be modified. The modification of the probability is brought about by variation of densities from species to species. Besides those, the total number of species in the community is assumed to be S^* . Therefore, several values of λ is considered as constituting a sample of size S^* from some continuous distribution

of λ having a pdf of $f(\lambda)$. Then the modified probability of j^{th} species represented by r individuals in the sample is the expectation of the probability of j^{th} species being represented by r individual in the collected sample defined as follows Pielou (1977).

$$P_r = \int_0^{\infty} \frac{\lambda^r e^{-\lambda}}{r!} f(\lambda) d\lambda, \quad r = 0, 1, 2, \dots \quad (2.0.2)$$

The probability distribution presented in (2.0.2) is the distribution of the species frequencies n_0, n_1, n_2, \dots , where n_r is the expected frequency of species represented by r individual in the sample defined as follows Pielou (1977).

$$n_r = S^* P_r \quad (2.0.3)$$

Furthermore, observed distribution of species abundance distribution is a truncated type of theoretical distribution since zero class is not represented in the collection, this is due to the fact that little information is known about them. In calculation if S is the observed number of species in the samples collected then the missing class species is calculated as shown below Pielou (1977).

$$S^* - S = n_0 \quad (2.0.4)$$

The next section discusses some families of compound poisson distributions that is used in describing species abundance distribution as discussed by Pielou (1977).

2.1 The Logarithmic Series or Log series Distribution

Logarithm distribution is one of the parametric approaches put in place in describing species abundance distribution defined as follows McGill and Magurran (2011).

$$P_r(X = x) = K \frac{C^x}{x}, \quad K = -1/\log(1 - C) \quad (2.1.1)$$

When we assume that several values of λ for different species collected from community (2.1) follows Pearson Type III distribution (Gamma distribution), that is $f(\lambda)$ is defined using the equation below.

$$f(\lambda) = \frac{P^{-k} \lambda^{k-1} e^{-\frac{\lambda}{P}}}{\Gamma(k)}, \quad \lambda \geq 0 \text{ and } k, P > 0 \quad (2.1.2)$$

Using (2.1.2) and (2.0.2), then probability that a species will be represented by r individuals in the collection is as follows.

$$P_r = \int_0^{\infty} \lambda^r \frac{e^{-\lambda}}{r!} \frac{P^{-k} \lambda^{k-1} e^{-\frac{\lambda}{P}}}{\Gamma(k)} d\lambda \quad (2.1.3)$$

The solution of equation (2.1.3) is a negative binomial distribution with probability P_r defined as shown below.

$$P_r = \frac{\Gamma(k+r)}{r! \Gamma(k)} \left(\frac{P}{1+P} \right)^r \left(\frac{1}{1+P} \right)^k, \quad \text{where } r = 0, 1, 2, \dots \quad (2.1.4)$$

Equation (2.1.4) can further be simplified by letting $P/(1+P) = X$, we now have the following equation Pielou (1977).

$$P_r = \frac{\Gamma(k+r)}{r!\Gamma(k)}(1-X)^k X^r, \text{ where } 0 < X < 1 \quad (2.1.5)$$

Equation (2.1.5) represents the probability that a given species will be represented by r individuals in the collection without ignoring the zero class species. We now need to define a new probability P'_r which is the probability that a given species will be represented by r individuals in the collection when zero class species is ignored. This probability distribution P_r is a truncated negative binomials distribution and is calculated as shown below.

$$P'_r = \frac{P_r}{1-P_0} = \frac{\Gamma(k+r)}{r!\Gamma(k)} \frac{X^r(1-X)^k}{[1-(1-X)^k]}, \quad r = 1, 2, \dots \quad (2.1.6)$$

where P_0 is the probability of the zero class species and calculated from (2.1.4) as follows.

$$P_0 = \frac{\Gamma(k)}{0!\Gamma(k)} X^0(1-X)^k = (1-X)^k \quad (2.1.7)$$

In equation (2.1.6), if we collect terms together that are independent of r and substitute them by C then we have the following.

$$P'_r = C \frac{\Gamma(k+r)}{r!} X^r, \quad \text{and } C = \frac{(1-X)^k}{[1-(1-X)^k]} \frac{1}{\Gamma(k)} \quad (2.1.8)$$

In equation (2.1.8) parameter k measures variability in the densities of different species. A large value of k implies that the variability in the densities of species is very small and small value of k implies that their is large variability in the densities of species. In a given defined ecological community it is clear that we have different species and this implies that their is high variation in the mean abundance of the species, therefore it is very important to let $k \rightarrow 0$ in (2.1.8). It follows that.

$$\lim_{k \rightarrow 0} P'_r = \lim_{k \rightarrow 0} C \frac{\Gamma(k+r)}{r!} X^r = \gamma \frac{X^r}{r}, \quad \text{where } \gamma = \lim_{k \rightarrow 0} C \quad (2.1.9)$$

Using equation (2.1.9), then expected frequency of species with r individuals is as follows Pielou (1977).

$$n_r = S P'_r = S \gamma \frac{X^r}{r} = \alpha \frac{X^r}{r} \quad (2.1.10)$$

where S is the total number of observed species in the collected sample from community (2.1). In addition using equation (2.1.10), the total number of species observed in the collection, S and the total abundance of all the species in the collection, N are defined as follows.

$$S = \sum_{r=1}^{\infty} n_r = \sum_{r=1}^{\infty} \frac{\alpha X^r}{r} = -\alpha \log(1-X) \quad (2.1.11)$$

$$N = \sum_{r=1}^{\infty} r n_r = \sum_{r=1}^{\infty} \frac{r \alpha X^r}{r} = \frac{\alpha X}{1-X} \quad (2.1.12)$$

Equation (2.1.10) is a log series distribution as shown in (2.1.1) with parameters α and X . When the expected frequency in (2.1.10) is plotted then we end up with a plot describing a clear picture of SAD as shown below.

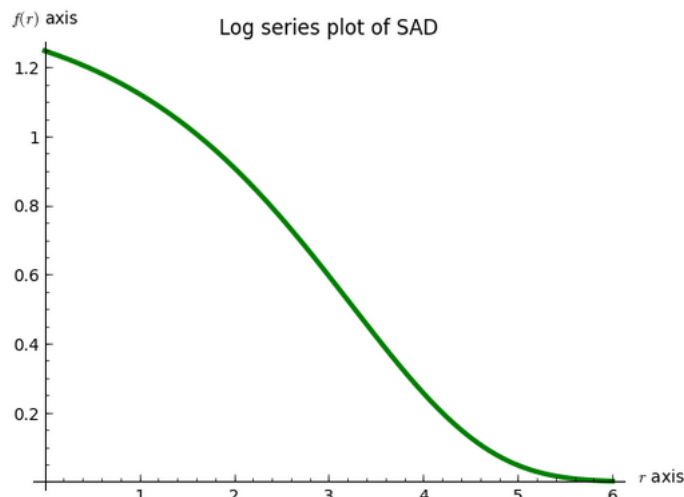


Figure 2.2: Log series distribution illustrating species abundance distribution with $X = 0.2$ and $\alpha = 2$

From figure (2.1), we observe that it is right skewed. In relation to ecological community this plot implies that rare species are more abundant as compared to common species. This is due to the decreasing nature of the plot with increase in r values. Using log series distribution in describing species abundance distribution there is one main disadvantage encountered, that is in deriving expression for probability that any species will be represented by r individual in the collection, we let $k \rightarrow 0$ in (2.1.9) and this implies that there is another uncollected class with $k = 0$ that has not been catered for in the collection, therefore using this method it is very hard to figure out the actual number of species S^* in the parent population of species. Furthermore, fitting this distribution to species abundance data is presented in section 2.5 alongside other parametric approaches to SAD presented in this chapter.

2.2 Lognormal distribution

Lognormal distribution is another important parametric approach that is used in describing SAD. Lognormal distribution is defined as a distribution whose variate conforms to the normal laws of probability [Aitchison and Brown \(1957\)](#). Probability density function of lognormal distribution is defined as shown below.

$$f(\lambda) = \frac{1}{\lambda\sigma\sqrt{2\pi}} \exp \left[-\frac{1}{2\sigma^2} \left(\ln \frac{\lambda}{m} \right)^2 \right], \text{ where } \lambda > 0 \quad (2.2.1)$$

Where $\ln \lambda$ follows normal distribution with mean $\ln m$ and variance σ^2 [Aitchison and Brown \(1957\)](#). Suppose that several values of λ for species in the sample collected from community (2.1) follows lognormal distribution, then using equation (2.2.1), probability that a given species will be represented by r individuals in the collection is as follows [Pielou \(1977\)](#).

$$P_r = \int_0^\infty \lambda^r \frac{e^{-\lambda}}{r!} \frac{1}{\lambda\sigma\sqrt{2\pi}} \exp \left[-\frac{1}{2\sigma^2} \left(\ln \frac{\lambda}{m} \right)^2 \right] d\lambda \quad (2.2.2)$$

$$P_r = \frac{1}{r!\sigma\sqrt{2\pi}} \int_0^\infty \exp \left[-\lambda + r \ln \lambda - \frac{1}{2\sigma^2} (\ln \lambda - \ln m)^2 \right] \frac{d\lambda}{\lambda} \quad (2.2.3)$$

According to Pielou (1969), equation (2.2.3) can further be simplified by letting $\ln \lambda = x$ to give us equation below.

$$P_r = \frac{1}{r! \sigma \sqrt{2\pi}} \int_0^{\infty} \exp \left[e^{-x} + rx - \frac{1}{2\sigma^2} (x - \ln m)^2 \right] dx \quad (2.2.4)$$

The probability distribution in (2.2.4) is a poisson lognormal distribution. This distribution is dependent on two parameters, medium species abundance m and variation of species abundance σ^2 . In addition based on Pielou (1977), this probability distribution has no explicit expression for the integral. This distribution was first tested by Preston by using theoretical lognormal frequencies. According to Preston he assumes that a given species is represented by its expected number of individuals in the collection and this is not affected by variation caused during sampling. In addition, Preston grouped the values of r into some groups called octaves such that the midpoint of each group is double that of proceeding one as shown below Pielou (1977).

$$r = 1, 2, 4, 8, 16, 32, \dots \quad (2.2.5)$$

where r is group boundaries and the midpoint of the octaves are as follows.

$$r = 1\frac{1}{2}, 3, 6, 12, \dots \quad (2.2.6)$$

The groupings of r implies that if a species falls on the boundary of the group, for example 2^x individuals, then it is considered to contribute half a species to the octave (2^{x-1} to 2^x) and a half a species to (2^x to 2^{x+1}) as explained by Pielou (1977). This is similar to transforming species abundance with respect to \log_2 and the plot of the transformed abundance is as follows.

Species Abundance Distribution plot

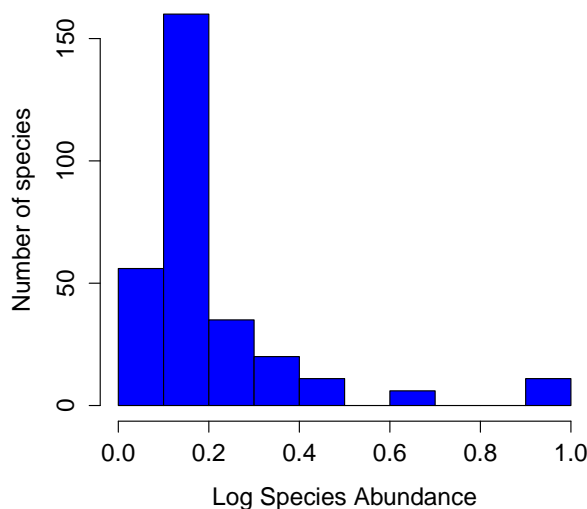


Figure 2.3: Histogram plot for BCI dataset for trees in the forest illustrating Preston octaves by transforming species abundance using \log_2 .

In order to bring a clear picture of species abundance distribution as described by lognormal distribution, then it is very important to do a plot for lognormal distribution function and by using Sage the following plots were obtained.

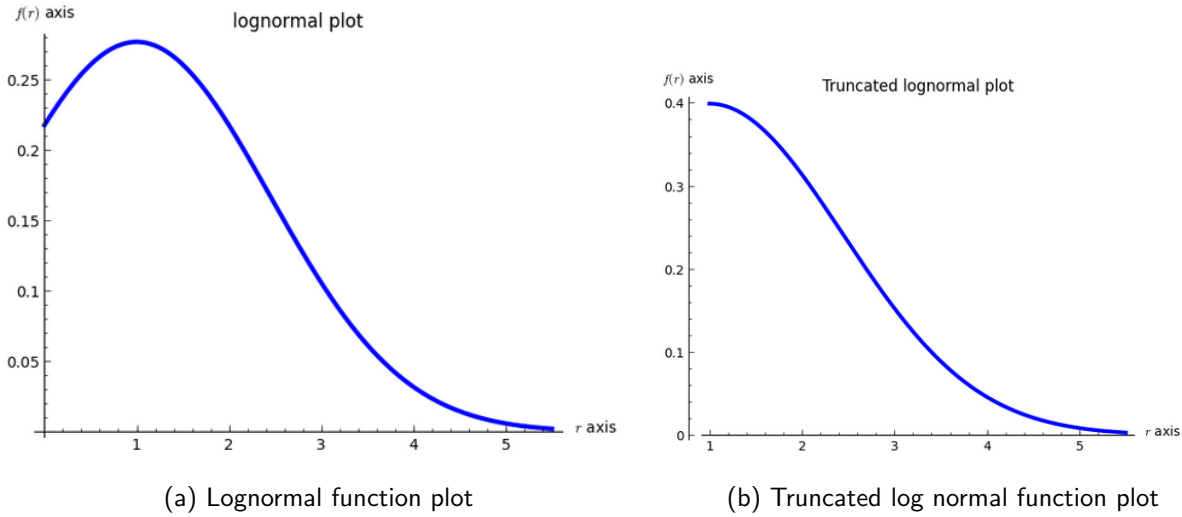


Figure 2.4: Illustration of species abundance distribution based on Log normal distribution with mean $\mu = 3$ and variance $\sigma^2 = 3$.

2.2.1 Interpretation of the plots. In figure (2.4), the two plots for lognormal distribution that is truncated and non-truncated lognormal distribution, we observe that for the non-truncated one the curve rises to some maximum modal value and then starts decreasing, this implies that more abundance species will be somewhere to the right of the veil line which shouldn't be the case in real ecological community, on the other hand for truncated one, that is when we let the curve starts at some value of $r = 1$ since we assume that no species can be represented by less than one individual in the community, it is clear that the plot of truncated lognormal distribution is a decreasing function from $r = 1$ and this gives a clear picture of how species are distributed in our ecosystem. In addition, plot illustrating octaves in (2.3) gives a clear picture of lognormal shape of species abundance distribution. In general when lognormal distribution is fitted to species abundance data and the plot for the histogram of SAD and the density curve are done on the same axis we end up with the following plot.

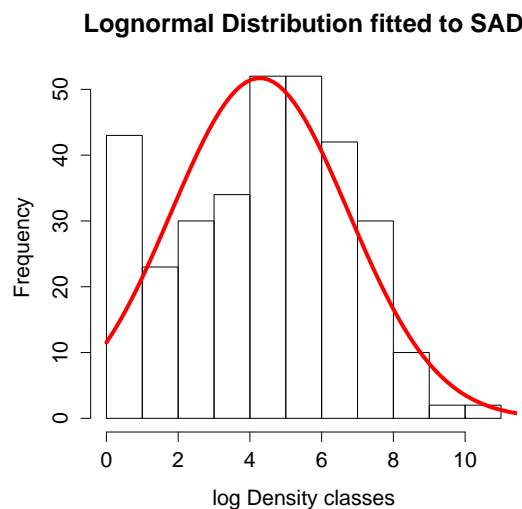


Figure 2.5: Lognormal distribution fitted to BCI (2006) dataset for the year 1982.

From plot (2.5), it is clear that lognormal distribution gives a good fit to SAD data, but we can see that the curve rises to some modal value is when it starts decreasing and this implies that the most abundance species are somewhere to the right of the veil line which should not be the case due to the fact that in real ecological community rare species are more as compared to common species and this is similar to interpretation for plots (2.4a). However, unlike log series distribution, when species abundance distribution is fitted using lognormal curve it is possible to estimate the total number of species in the population including both collected and uncollected species Pielou (1977).

2.3 Negative binomial distribution

Negative binomial distribution is another fundamental distribution used by ecologist in describing species abundance. Negative binomial distribution is defined as follows Furbes et al. (1957) .

$$p(y) = \frac{\Gamma(y)}{\Gamma(r)\Gamma(y-r)!} p^r q^{y-r}, \quad y = r, r+1, r+2, \dots, \quad 0 \leq p \leq 1 \quad (2.3.1)$$

From the previous section, it is clear that if $\ln \lambda$ follows lognormal distribution, then logarithmic plot of species abundance distribution will first rise to some modal values before starts decreasing, this is only possible if the sample is large enough for the veil line to fall to the left of the modal value. This implies that λ has some modal value which is greater than zero and this is a clear indication that rare and common species are few as compared to some species having some intermediate value of λ . If we now assume that densities of several species follows Pearson Type III distribution then we have the following pdf for λ Pielou (1977).

$$f(\lambda) = \frac{1}{\Gamma(k)} P^{-k} \lambda^{k-1} e^{-\frac{\lambda}{P}} \quad (2.3.2)$$

Equation (2.3.2), implies that the expected frequency of species with r individuals without letting $k \rightarrow 0$ will follow negative binomial distribution. We are interested in analysing whether $f(\lambda)$ is a monotonically decreasing function at some value of $\lambda = 0$ or if it has some modal value at $\lambda > 0$. To check for this property then in equation (2.3.2), we let P to be a constant and we differentiate (2.3.2) with respect to λ . In doing so we have the following equation Pielou (1977).

$$\frac{df(\lambda)}{d\lambda} = \frac{1}{\Gamma(k)} P^{-k} (k-1) \lambda^{k-2} e^{-\frac{\lambda}{P}} - \frac{1}{\Gamma(k)} P^{-k} \lambda^{k-1} \frac{1}{P} e^{-\frac{\lambda}{P}} \quad (2.3.3)$$

$$\frac{df(\lambda)}{d\lambda} = \frac{1}{\Gamma(k)} P^{-k} \lambda^{k-2} e^{-\frac{\lambda}{P}} \left\{ k-1 - \frac{\lambda}{P} \right\} \quad (2.3.4)$$

In (2.3.4), if $k > 1$ then $f(\lambda)$ will have its maximum if $\lambda = P(k-1)$. Therefore this implies that if the species abundance distribution is fitted using truncated negative binomial distribution with $k > 1$, then some intermediate species will be more as compared to rare and common species. In addition, if $0 \leq k \leq 1$, then $\frac{df(\lambda)}{d\lambda}$ is negative for all values of λ and this implies $f(\lambda)$ will have its maximum in the range of $0 \leq k \leq 1$, hence rare species will be more as compared to abundant species Pielou (1977). The two conditions for negative binomial distribution that describes species abundance is summarized in the plots below.

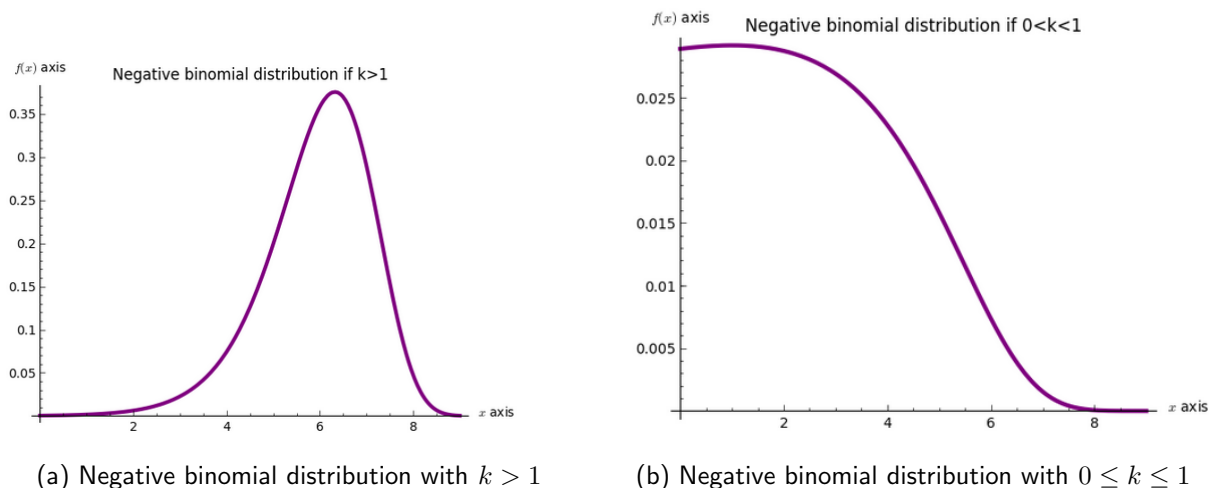


Figure 2.6: Two plots of negative binomial distribution conditions showing description of species abundance with $p = 40$ and values of k are; $k > 1$ and $0 \leq k \leq 1$ for (a) and (b) respectively.

Figure (2.6) gives a clear picture when negative binomial distribution is used in describing species abundance distribution using different conditions of k as explained in equation (2.3.4). From (2.6a) it is clear that some intermediate species will be more as compared to rare and common species also from (2.6b) we can observe that rare species will be more abundant as compared to common species and this is the same interpretation given using equation (2.3.4).

2.4 Geometric distribution

Geometric distribution is a special case of negative binomial distribution defined as follows Wackerly et al. (2008).

$$p(y) = q^{y-1}p, \text{ where } y = 1, 2, \dots \text{ and } 0 \leq p \leq 1 \tag{2.4.1}$$

From the previous section it is clear that when several values of λ follows Pearson Type III distribution and we let $k \rightarrow 0$ this leads to log series distribution of species abundance, in addition if we let k to be defined by the data then this yields a negative binomial distribution explaining species abundance distribution. If we now let $k = 1$ in (2.3.2), this will yield for us the following probability density function for λ Pielou (1977).

$$f(\lambda) = \frac{1}{\Gamma(1)} P^{-1} \lambda^0 e^{-\frac{\lambda}{P}} \tag{2.4.2}$$

$$f(\lambda) = \frac{1}{P} e^{-\frac{\lambda}{P}}, \text{ where } \lambda \geq 0 \tag{2.4.3}$$

Equation (2.4.3) is probability density function of exponential distribution. In addition, when we put $k = 1$ in (2.1.4), this will yield for us geometric distribution which is a discrete case of exponential distribution in (2.4.3). We now have the following equation Pielou (1977).

$$P'_r = \frac{\Gamma(1+r)}{r! \Gamma(1)} \frac{P^r}{(1+P)^{r+1} [1 - (1+P)^{-1}]} \tag{2.4.4}$$

$$P'_r = \left(\frac{P}{1+P} \right)^{r-1} \left(\frac{1}{1+P} \right), \text{ where } r = 1, 2, \dots \quad (2.4.5)$$

When we now plot geometric distribution in (2.4.5) that describes species abundance distribution as a function of some r values, then we end up with the following plot that is used in describing species abundance distribution in the community.

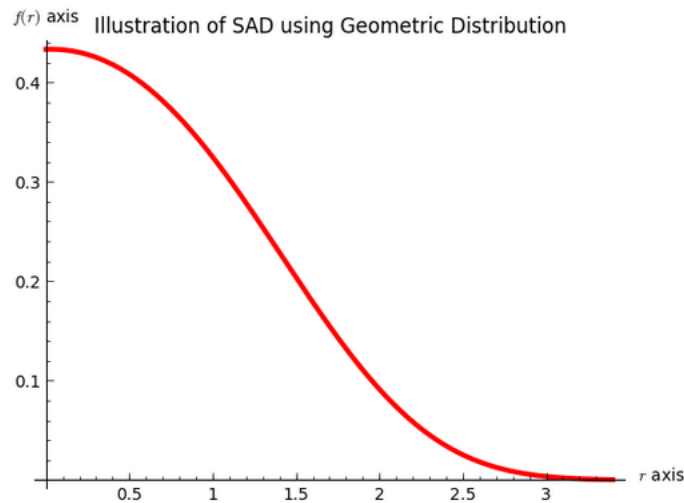
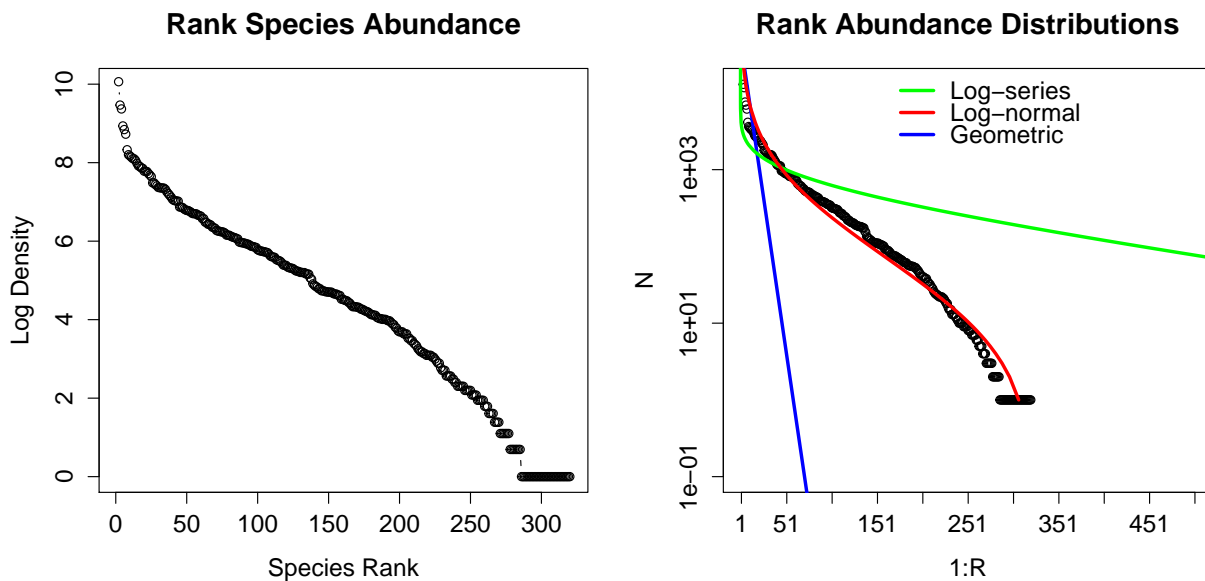


Figure 2.7: Illustration of species abundance distribution using geometric distribution with $p = 0.3$

From figure (2.7), we observe that it is a decreasing function with increase in the value of r . In relation to species abundance distribution, this plot is interpreted as rare species are more as compared to common species which is the case for any ecological community. Even though this distribution has been tested in many different ways apart from the one presented in this section, it still emerge the best as compared to negative binomial and log series distribution as we shall see in the next section.

2.5 Evaluating parametric approach

In evaluating which distribution best fit the species abundance data, rank species abundance was plotted together with rank abundance distributions. Rank species abundance is used in this section to mean that most abundant species i.e. singleton species is represented by rank one, then the second abundance species is represented by rank two and the system continues for the rest of the species in the community. The following plots were obtained.



(a) Rank Species Abundance.

(b) Fitting Distributions to Species Abundance data.

Figure 2.8: A few common rank abundance distributions fitted using BCI (2006) data for the year 1982. The log normal curve best fits the data.

From figure (2.8b), it is clear that log normal distribution is the best in fitting and describing species abundance distribution since the log normal density curve matches with the rank species abundance curve. In addition we can observe that all the plots takes the shape of decreasing function and this implies that rare species are more as compared to common species and this the same interpretation presented in the previous sections of this chapter. The next chapter present some mechanistic approaches that are also used in describing species abundance distribution alongside the parametric approaches.

3. Mechanistic approach to Species Abundance Distribution

This chapter discusses some theories that has been put in place in describing species abundance distribution.

3.1 Maximum Entropy Theory of Ecology (METE)

This is one of the most fundamental theory applied in Physics and Ecology to date. This theory is based on the use of available information entropy and maximize it to infer on what is in line with the available entropy information. Information entropy is defined as the measure of confusion about a given state of system. In addition, from thermodynamics point of view, entropy is described as the number of specific ways in which a thermodynamic system can be arranged. In ecology this theory is used to infer to the SAD based on the available information regarding; area occupied by species, number of species, number of individuals per species and total rate of metabolic energy required by all the individual belonging to different species. Maximum entropy is the process of maximizing information entropy subjected to some constraints of the system under study. In this section we are going to apply Lagrange multiplier method in maximizing entropy information which will helps us in deriving species abundance distributing [Harte \(2011\)](#). In addition, maximum entropy theory of ecology discussed here is based on [Harte \(2011\)](#) and [Sognnæs \(2011\)](#) arguments.

3.1.1 Background information of METE. In thermodynamics Maximum Entropy theory is defined based on some state variables and the same principle is also applied in METE. State variables are properties of a given system that need to be specified in order to help in implementing a given theory but whose determination doesn't make part of the content of the theory itself. In METE we base our calculations and arguments on the following state variables [Sognnæs \(2011\)](#).

- (i) A_0 : The total area of ecological community.
- (ii) S_0 : The total number of species from any specified taxonomic category (i.e plants or animals) within this area A_0 .
- (iii) N_0 : The total population of individuals belonging to different species in the area.
- (iv) E_0 : The total rate of metabolic energy consumed by individuals belonging to various species in the area A_0 .

The above state variables has been chosen from the ecosystem based on some criteria. For example, A_0 has been chosen due to the fact that it is the tool used in measuring physical scale of a given system, S_0 is chosen due to the fundamental role played by species richness in an ecological community and macro ecological metrics. Finally total species abundance N_0 and total metabolic rate of individuals in a given species E_0 are chosen since they play a major role in describing a given species [Sognnæs \(2011\)](#).

3.1.2 Deriving METE. This theory is widely used in ecology. It is applicable in deriving different ecological metrics such as species abundance distribution (SAD), species area relation (SAR), energy distribution among species and intra-specific metabolic energy probability density [Harte \(2011\)](#). As discussed before this theory is based on applying the principle of Maximum entropy in ecology making it to give us a lot of insight information based on some few selected state variable [Sognnæs \(2011\)](#). The

derivation of this theory is based on some constraints defined by the state variables, A_0 , S_0 , N_0 and E_0 . In addition the choice of these state variable is based on the fact that their is no prior information that can justify these state variable or the choice of any other state variable. However, these chosen state variable can only be justified based on their failure or success when the model is fitted using real species abundance distribution data. [Sognnæs \(2011\)](#).

3.1.3 Ecosystem structure function, $R(n, \varepsilon)$. Ecosystem structure function ($R(n, \varepsilon)$) is a joint conditional probability distribution function that is defined over species and individuals of a given species in a given defined ecological community area A_0 with respect to metabolic rate of each individual, whereby metabolic rate is the rate at which metabolism occurs in a living organism [Sognnæs \(2011\)](#). In this function n represent species abundance and ε is the metabolic energy rate of each individuals belonging to various species in a community. In addition this function is discrete over species abundance n and continuous over metabolic energy rate ε of individuals. In simple terms $R(n, \varepsilon)$ is defined as joint conditional probability distribution function describing that when a given species is picked randomly from a given pool of species then it has abundance n and if again an individual is picked at random from the group of already picked species then that individual has metabolic energy requirement in the interval $(\varepsilon, \varepsilon + d\varepsilon)$ [Harte \(2011\)](#).

In deriving this theory we set minimum metabolic rate to be 1, i.e. $\varepsilon_{min} = 1$. This minimum metabolic rate simply imply that no matter how small you chose metabolic rate from a pool of individuals belonging to different species then it can't be less than one. The normalization condition on $R(n, \varepsilon)$ is as shown below [Harte \(2011\)](#).

$$\sum_{n=1}^{N_0} \int_{\varepsilon=1}^{E_0} R(n, \varepsilon | A_0, S_0, N_0, E_0) d\varepsilon = 1 \quad (3.1.1)$$

From (3.1.1), the three state variable S_0 , N_0 and E_0 define two new constraints; the average species abundance, $\frac{N_0}{S_0}$ and the average total metabolic rate of individuals defined over species given by $\frac{E_0}{S_0}$. The new defined constraints are represented as shown below [Harte \(2011\)](#).

$$\sum_{n=1}^{N_0} \int_{\varepsilon=1}^{E_0} n \cdot R(n, \varepsilon | A_0, S_0, N_0, E_0) d\varepsilon = \frac{N_0}{S_0} \quad (3.1.2)$$

$$\sum_{n=1}^{N_0} \int_{\varepsilon=1}^{E_0} n \cdot \varepsilon \cdot R(n, \varepsilon | A_0, S_0, N_0, E_0) d\varepsilon = \frac{E_0}{S_0} \quad (3.1.3)$$

From (3.1.2) and (3.1.3), we can see that the state variable A_0 has not played any role in defining new constraints and this imply that it doesn't give us much insight in derivation of SAD, so it is realistic to ignore its effect in the derivation and drop it at this stage since we will not use it in the rest of our calculations. Also from now henceforth we are going to denote our ecosystem structure function by $R(n, \varepsilon)$ instead of $R(n, \varepsilon | A_0, S_0, N_0, E_0)$. At this stage we can't assume anything about the species abundance, that is we cannot claim that the species under study have the same abundance or if they have different abundance. In addition, we cannot also say anything about the distribution of energy among individuals of the species. The only information we have is about the state variable N_0 , S_0 , and E_0 . So we are going to use this limited information to helps us infer and describe the shape of species abundance distribution. By now applying maximum entropy principle in maximizing the information entropy of distribution $R(n, \varepsilon)$, we now have the following. [Sognnæs \(2011\)](#).

$$I_R = - \sum_{n=1}^{N_0} \int_{\varepsilon=1}^{E_0} R(n, \varepsilon) \log(R(n, \varepsilon)) d\varepsilon \quad (3.1.4)$$

By maximizing (3.1.4) subjected to constraints (3.1.2) and (3.1.3), taking into consideration the normalization condition in (3.1.1), then we obtain expression representing $R(n, \varepsilon)$ after which species abundance distribution is obtained from the expression of $R(n, \varepsilon)$ by integrating out energy requirements of individuals belonging to different species as shown below Harte (2011).

$$\langle \Phi(n) \rangle = \int_{\varepsilon=1}^{E_0} R(n, \varepsilon) d\varepsilon \quad (3.1.5)$$

3.1.4 Solving for $R(n, \varepsilon)$ and metric derived from it. Using maximum entropy method by subjecting (3.1.4) to constraint (3.1.2) and (3.1.3) together with normalized condition (3.1.1), then the expression for $R(n, \varepsilon)$ is obtained as follows.

Lets define Lagrange multiplier function ($L(R, \lambda, \mu)$) as follows.

$$L(R, \lambda, \mu) = \left\{ I_R - \lambda_1 \left(\sum_{n=1}^{N_0} \int_{\varepsilon=1}^{E_0} n \cdot R(n, \varepsilon) d\varepsilon - \frac{N_0}{S_0} \right) - \lambda_2 \left(\sum_{n=1}^{N_0} \int_{\varepsilon=1}^{E_0} n \cdot \varepsilon \cdot R(n, \varepsilon) d\varepsilon - \frac{E_0}{S_0} \right) - \mu \left(\sum_{n=1}^{N_0} \int_{\varepsilon=1}^{E_0} R(n, \varepsilon) d\varepsilon - 1 \right) \right\} \quad (3.1.6)$$

Where λ_1 and λ_2 are Lagrange multipliers and μ is some arbitrary constant ($\lambda_0 - 1$). By substituting for I_R in (3.1.6), it follows that.

$$L(R, \lambda, \mu) = \left\{ \left(- \sum_{n=1}^{N_0} \int_{\varepsilon=1}^{E_0} R(n, \varepsilon) \cdot \log(R(n, \varepsilon)) d\varepsilon \right) - \lambda_1 \left(\sum_{n=1}^{N_0} \int_{\varepsilon=1}^{E_0} n \cdot R(n, \varepsilon) d\varepsilon - \frac{N_0}{S_0} \right) - \lambda_2 \left(\sum_{n=1}^{N_0} \int_{\varepsilon=1}^{E_0} n \cdot \varepsilon \cdot R(n, \varepsilon) d\varepsilon - \frac{E_0}{S_0} \right) - \mu \left(\sum_{n=1}^{N_0} \int_{\varepsilon=1}^{E_0} R(n, \varepsilon) d\varepsilon - 1 \right) \right\} \quad (3.1.7)$$

Then by finding derivative of (3.1.7) with respect to $R(n, \varepsilon)$ and setting derivative to zero, we have the following.

$$\frac{dL(R, \lambda, \mu)}{dR(n, \varepsilon)} = \left\{ \left(- \sum_{n=1}^{N_0} \int_{\varepsilon=1}^{E_0} 1 + \log(R(n, \varepsilon)) d\varepsilon \right) - \lambda_1 \left(\sum_{n=1}^{N_0} \int_{\varepsilon=1}^{E_0} n d\varepsilon \right) - \lambda_2 \left(\sum_{n=1}^{N_0} \int_{\varepsilon=1}^{E_0} n \cdot \varepsilon d\varepsilon \right) - (\lambda_0 - 1) \left(\sum_{n=1}^{N_0} \int_{\varepsilon=1}^{E_0} 1 d\varepsilon \right) \right\} = 0 \quad (3.1.8)$$

It now follows that.

$$\frac{dL(R, \lambda, \mu)}{dR(n, \varepsilon)} = \sum_{n=1}^{N_0} \left\{ \int_{\varepsilon=1}^{E_0} (-1 - \log(R(n, \varepsilon)) - \lambda_1 n - \lambda_2 n\varepsilon - (\lambda_0 - 1)) d\varepsilon \right\} = 0 \quad (3.1.9)$$

Since (3.1.9) is zero for all $dL(R, \lambda, \mu)$, then it follows that.

$$(-1 - \log(R(n, \varepsilon)) - \lambda_1 n - \lambda_2 n\varepsilon - \lambda_0 + 1) = 0 \quad (3.1.10)$$

Then by solving for $R(n, \varepsilon)$ from (3.1.10), we end up with the following expression

$$R(n, \varepsilon) = \exp(-\lambda_0 - \lambda_1 n - \lambda_2 n\varepsilon) \quad (3.1.11)$$

Then from (3.1.11) all Lagrange multipliers $(\lambda_0, \lambda_1, \lambda_2)$ are fixed by doing substitution in normalized condition (3.1.1). Then we now have the following.

$$\sum_{n=1}^{N_0} \int_{\varepsilon=1}^{E_0} \exp(-\lambda_0 - \lambda_1 n - \lambda_2 n \varepsilon) d\varepsilon = 1 \quad (3.1.12)$$

$$e^{-\lambda_0} \sum_{n=1}^{N_0} \int_{\varepsilon=1}^{E_0} \exp(-\lambda_1 n - \lambda_2 n \varepsilon) d\varepsilon = 1 \quad (3.1.13)$$

Since in (3.1.13) we can see that $e^{-\lambda_0}$ is a common factor therefore we need to consider it separately, we now have the following.

$$Z(\lambda_1, \lambda_2) = \sum_{n=1}^{N_0} \int_{\varepsilon=1}^{E_0} \exp(-\lambda_1 n - \lambda_2 n \varepsilon) d\varepsilon \quad (3.1.14)$$

With

$$e^{-\lambda_0} = \frac{1}{Z(\lambda_1, \lambda_2)}, \quad \lambda_0 = \log Z(\lambda_1, \lambda_2) \quad (3.1.15)$$

Using (3.1.15), the ecosystem structure function in (3.1.11) is fully defined as shown below.

$$R(n, \varepsilon) = \frac{1}{Z(\lambda_1, \lambda_2)} \exp(-\lambda_1 n - \lambda_2 n \varepsilon) \quad (3.1.16)$$

where $Z(\lambda_1, \lambda_2)$ is a partition function defined in (3.1.14). The two Lagrange multipliers λ_1 and λ_2 in (3.1.16) can be solved by substituting (3.1.16) in the two constraint (3.1.2) and (3.1.3) and this calculation are lengthy and there is no explicit solution for the two Lagrange multipliers, so according to Harte (2011) he defined expressions for the multipliers as follows.

$$\beta = \lambda_1 + \lambda_2 \quad (3.1.17)$$

$$\sigma = \lambda_1 + E_0 \lambda_2 \quad (3.1.18)$$

3.1.5 Simplification of the ecosystem structure function $R(n, \varepsilon)$. According to Harte (2011), the expression of $R(n, \varepsilon)$ can further be simplified by simplifying the equations that are used in its calculations as shown in the workings that follows. In addition for easy calculations we denote our partition function with Z instead of $Z(\lambda_1, \lambda_2)$. The exact solutions of the integrals of the constraints in (3.1.2) and (3.1.3) are as follows.

$$\sum_{n=1}^{N_0} \int_{\varepsilon=1}^{E_0} \frac{1}{Z} n e^{-\lambda_1 n} e^{-\lambda_2 n \varepsilon} d\varepsilon = \frac{N_0}{S_0} \quad (3.1.19)$$

$$\frac{\sum_{n=1}^{N_0} (e^{-\beta n} - e^{-\sigma n})}{Z \cdot \lambda_2} = \frac{N_0}{S_0} \quad (3.1.20)$$

and

$$\sum_{n=1}^{N_0} \int_{\varepsilon=1}^{E_0} \frac{1}{Z} \cdot n \cdot \varepsilon \cdot e^{-\lambda_1 n} e^{-\lambda_2 n \varepsilon} d\varepsilon = \frac{E_0}{S_0} \quad (3.1.21)$$

When we integrate (3.1.21) by parts then we end up with the following solution

$$\sum_{n=1}^{N_0} \frac{(e^{-\beta n} - E_0 e^{-\sigma n})}{Z \cdot \lambda_2} + \sum_{n=1}^{N_0} \frac{(e^{-\beta n} - e^{-\sigma n})}{Z \cdot \lambda_2^2} = \frac{E_0}{S_0} \quad (3.1.22)$$

Again the integral of the partition function Z is as follows

$$Z = \sum_{n=1}^{N_0} \int_{\varepsilon=1}^{E_0} \exp(-\lambda_1 n - \lambda_2 n \varepsilon) d\varepsilon = \sum_{n=1}^{N_0} \frac{(e^{-\beta n} - e^{-\sigma n})}{\lambda_2 \cdot n} \quad (3.1.23)$$

When we now substitute $Z\lambda_2$ from (3.1.23) in (3.1.20) and (3.1.22) then we end up with the below two equations for the constraints.

$$\frac{\sum_{n=1}^{N_0} (e^{-\beta n} - e^{-\sigma n})}{\sum_{n=1}^{N_0} \frac{(e^{-\beta n} - e^{-\sigma n})}{n}} = \frac{N_0}{S_0} \quad (3.1.24)$$

and

$$\frac{\sum_{n=1}^{N_0} (e^{-\beta n} - E_0 e^{-\sigma n})}{\sum_{n=1}^{N_0} \frac{(e^{-\beta n} - e^{-\sigma n})}{n}} + \frac{1}{\lambda_2} = \frac{E_0}{S_0} \quad (3.1.25)$$

From equation (3.1.25) and (3.1.24) we can observe that they have unknown variables β and σ , therefore by applying (3.1.17) and (3.1.18), then they can best approximate Lagrange multipliers as a function of state variables S_0 , N_0 and E_0 Harte (2011). In mathematical derivation simplicity matters a lot, so we can still further simplify the expression for partition function Z in order for us to obtain a more simplified species abundance distribution. Since

$$\sum_{n=1}^{N_0} e^{-\beta n} = \frac{e^{-\beta} - e^{-\beta(N_0+1)}}{1 - e^{-\beta}} \quad (3.1.26)$$

then by using (3.1.26), we can now join (3.1.23) and (3.1.24) to give us the expression for partition function as follows Harte (2011).

$$Z = \frac{S_0}{\lambda_2 N_0} \sum_{n=1}^{N_0} (e^{-\beta n} - e^{-\sigma n}) = \left(\frac{S_0}{\lambda_2 N_0} \right) \left(\frac{e^{-\beta} - e^{-\beta(N_0+1)}}{1 - e^{-\beta}} - \frac{e^{-\sigma} - e^{-\sigma(N_0+1)}}{1 - e^{-\sigma}} \right) \quad (3.1.27)$$

Using (3.1.5) we can now obtain species abundance distribution as follows.

$$\langle \Phi(n) \rangle = \int_{\varepsilon=1}^{E_0} R(n, \varepsilon) d\varepsilon = \int_{\varepsilon=1}^{E_0} \frac{1}{Z} e^{-\lambda_1 n} e^{-\lambda_2 n \varepsilon} d\varepsilon \quad (3.1.28)$$

$$\langle \Phi(n) \rangle = \frac{e^{-\lambda_1 n}}{\lambda_2 Z n} \left(e^{-\lambda_2 n} - e^{-\lambda_2 n E_0} \right) = \left(\frac{e^{-\beta n} - e^{-\sigma n}}{\lambda_2 Z n} \right) \quad (3.1.29)$$

Equation (3.1.29) is the expression for species abundance distribution and the expression for $Z\lambda_2$ can be obtained from (3.1.27). The expression of SAD in (3.1.29) can further be simplified to give us a meaningful species abundance distribution as shown in the next section.

3.1.6 Good approximation to species abundance distribution. From Harte (2011), he claimed that in deriving expression for λ_2 we need to drop all terms in (3.1.24) and (3.1.25) having $e^{-\sigma n}$, this is due to the fact that for realistic values of state variables $e^{-\sigma}$ is very small as compared to $e^{-\beta}$ since if $S_0 > 4$ then $\beta < 1$ and $\sigma \geq S_0$ implying that $e^{-\sigma} \ll e^{-\beta}$ (see Table 7.2) in Harte (2011)). Therefore by dropping $e^{-\sigma n}$ in (3.1.24) we have the following derivation for λ_2 .

$$\frac{\sum_{n=1}^{N_0} e^{-\beta n}}{\sum_{n=1}^{N_0} \frac{e^{-\beta n}}{n}} = \frac{N_0}{S_0} \quad (3.1.30)$$

and in (3.1.25) is as follows

$$\frac{\sum_{n=1}^{N_0} e^{-\beta n}}{\sum_{n=1}^{N_0} \frac{e^{-\beta n}}{n}} + \frac{1}{\lambda_2} = \frac{E_0}{S_0} \quad (3.1.31)$$

By now combining (3.1.31) and (3.1.30), then we have the following simplified expression for λ_2 .

$$\lambda_2 = \frac{S_0}{E_0 - N_0} \quad (3.1.32)$$

According to Harte (2011) the denominator of (3.1.30) can further be simplified. In particular if $\beta N_0 \gg 1$ and $\beta \ll 1$, then we have the following simplification of the denominator of (3.1.30).

$$\sum_{n=1}^{N_0} \frac{e^{-\beta n}}{n} \approx \log\left(\frac{1}{\beta}\right) \quad (3.1.33)$$

Using (3.1.33) we can simplify equation (3.1.30) as follows.

$$\frac{N_0}{S_0} \approx \frac{e^{-\beta} - e^{-\beta(N_0+1)}}{1 - e^{-\beta}} \cdot \frac{1}{\log\left(\frac{1}{\beta}\right)} \quad (3.1.34)$$

Equation 3.1.34 implies that.

$$\frac{S_0}{N_0} \approx \frac{1 - e^{-\beta}}{e^{-\beta} - e^{-\beta(N_0+1)}} \cdot \log\left(\frac{1}{\beta}\right) \quad (3.1.35)$$

From (3.1.35) if $\beta \ll 1$, then equation (3.1.35) can further be simplified as shown below.

$$\frac{S_0}{N_0} = \beta \log\left(\frac{1}{\beta}\right) \quad (3.1.36)$$

In addition when we now use (3.1.36), then the simplified expression for Z in (3.1.27) by dropping all terms having $e^{-\sigma}$ is as shown below.

$$Z \approx \left(\frac{S_0}{\lambda_2 N_0}\right) \left(\frac{e^{-\beta} - e^{-\beta(N_0+1)}}{1 - e^{-\beta}}\right) \approx \frac{\log\left(\frac{1}{\beta}\right)}{\lambda_2} \quad (3.1.37)$$

By now substituting equation for Z from (3.1.37) in (3.1.29) and by dropping $e^{-\sigma n}$ in (3.1.29), we end up with the following simplified species abundance distribution.

$$\langle \Phi(n) \rangle \approx \frac{1}{\log\left(\frac{1}{\beta}\right)} \cdot \frac{e^{-\beta n}}{n} \quad (3.1.38)$$

From equation (3.1.38), when we let $\left(\frac{1}{\log(\beta^{-1})}\right)$ to be some constant α and we let $e^{-\beta} = X$, then the equation reduces to log series distribution as shown below.

$$\langle \Phi(n) \rangle = \alpha \frac{X^n}{n} \quad (3.1.39)$$

Derived METE equation in (3.1.38) best describe species abundance distribution as explained in the next section.

3.1.7 Evaluation of the model. In evaluating the fit of this model, a plot was done in Sage using some arbitrary values of the parameters and the following plot that depicts the SAD shape was obtained.

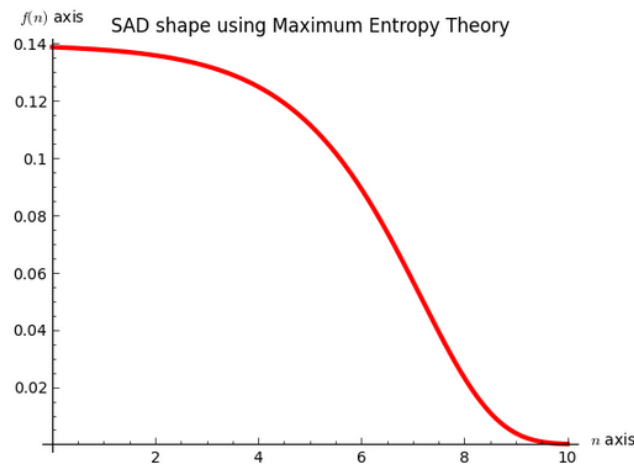


Figure 3.1: A plot of Log series distribution derived from METE showing Species Abundance Distribution with $\beta = 0.007$.

Interpretation of figure (3.1) is as follows; rare species are more abundant as compared to common species since this is a decreasing function from $n = 1$ to $n = 10$ and this plot gives a clear picture of a given ecological community as far as species abundance distribution is concerned.

METE is derived based on the choice of some state variables from the ecological community. As it has been shown earlier we realize that by maximizing information regarding this state variables, we can as well derive other ecological metrics like Spatial-Abundance Distribution $\Pi(n)$ and Species Area Relationship $S(A)$. Therefore we can conclude that this method is good when deriving ecological metrics by taking into consideration the role played by state variables of a community in ecology. However since METE is a null theory since it is derived based on some hypothesized state variables, this implies that the failure of the model in predicting species abundance distribution is due to changes in the system that causes relative change in the state variable over a short period of time. In addition this model gives a best fit of species abundance data as claimed by Sognnæs (2011). The next section discusses Neutral theory.

3.2 Neutral theory

Neutral theory as explained by most of the ecologists is one of the most fundamental theories in ecology used to describe species abundance distribution of the neutral local and meta communities He and Hu (2005). From an ecological point of view, a meta community is defined as the self-contained evolutionary biogeographical unit within which most members of a given species originate, live, grow and die He (2005), this community is characterised by speciation which is similar to immigration in a local community, whereby speciation is defined as an evolutionary biological process giving rise to new species in the community. A local community is a community embedded in a meta community which is characterised by exchange of migrants with the meta community or other local communities via emigration or immigration He (2005). Despite the fact that a local community is embedded in the meta community there is no clear distinction rather a sharp line defining the boundary of a local community from a meta community however on the actual landscape there is a continuum from a local community to a meta community based on variation of the degree in which their dynamics are affected by migration He (2005). Before continuing it is very important to clearly understand some major pillars in which neutral theory is based on. In neutral theory the concept of neutrality has been put in place to imply that species that share the same food-web level in the community are similar as far as their population dynamics are concerned that is, death, birth, speciation and dispersal rates Chave (2004). Furthermore, this theory is based on four major population dynamics; birth, death, immigration and emigration. However, speciation also plays a major role in this theory since it governs the process of emergence of new species generation in the meta community. What is more about the theory is that it is purely based on the interaction between local and meta communities via emigration and immigration processes.

3.2.1 Developments in Neutral theory. With no point of doubt, the neutral theory family is made up of a quite number of neutral theories proposed by different scholars, however for the purpose of this project I am going to discuss unified neutral theory for species abundance that unifies the meta and local communities based on the fundamental mechanism of population dynamics discussed by He (2005). In addition below are some recent developments of neutral theory described by some scholars.

3.2.2 The Unified Neutral theory of Biodiversity and Biogeography. This theory was developed by Hubbell. In this theory Hubbell described processes and dynamics that are associated with local communities. This theory describes migration processes across the local community to another local community or meta community. To clearly understand Hubbell's theory, let's consider a given defined ecological community made up of local communities and meta communities, whereby a meta community is defined by the total number of individuals denoted by J_M which is the size of the meta community with relative species abundance P_i . In addition, meta community individual dynamics is governed by a speciation-extinction process at a rate of V , so according to Hubbell he explained that species i can move from the meta community to a local community of size J having relative species abundance N_i at some rate m . So the point that Hubbell wanted to drive home based on his argument of species emigrating from meta community to local community is that, he wanted to make it clear that local communities and meta communities are related through migration at a rate m . Furthermore, he explained that migration rate m plays a major role in understanding species abundance that is, very small values of m implies that the rate at which species emigrate from meta to local community is low, therefore in simple terms small values of m implies that the role played by immigration in local community doesn't merge with speciation processes in meta community. However, immigration process is dependent on the species abundance since the most abundant species is likely to emigrate from meta community to local community. The diagrammatic representation of this theory is as follows Hubbell (2001).

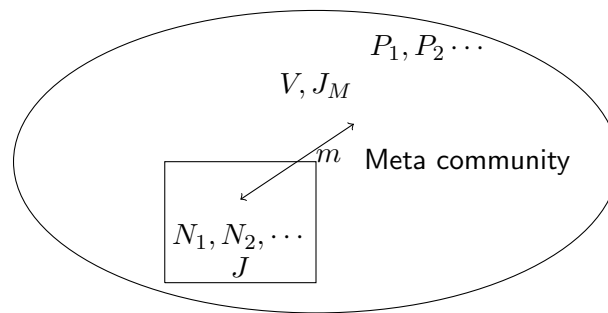


Figure 3.2: Diagrammatic representation of Hubbell's neutral argument where square inside meta community is the local community and V is the meta community speciation rate, m is migration rate, J_M is the total population of individuals in the meta community, P_i is the relative species abundance in the meta community, N_i is the number of individuals belonging to different species in local community and J in the size of the local community.

3.2.3 The Distribution of Abundance in Neutral Communities. Apart from Hubbell, Bell also contributed in developing a theory that boosted family of Neutral theory. He developed a model that describes species abundance of a community based on birth and death rates of individuals of different species. Just to make points clear he simulated a community where all its individuals have birth rates b and death rates d , with birth rates being higher than death rates, that is $b \geq d$ Chave (2004). In this theory since $b \geq d$, then there is a clear indication that with increase in time the community will be saturated, in order to avoid this disaster this calls for Bell to modify his theory by setting some fixed ceiling condition k individuals, that is if any individual exceeds this ceiling condition then that individual will be removed at random from the community. In addition, in this theory Bell explained that the meta-community is characterised by some fixed maximal number of species S and any S species can migrate freely into the meta community regardless of its abundance in the local community. He summarized the theory by explaining that the assumption that each species can immigrate at rate m may lead to species being distributed based on their different abundances Bell (2000). These immigration parameter m can also be interpreted as the probability that the death that has occurred in the local community is being replaced by the offspring of an individual outside the local community Chisholm and Lichstein (2009). In summary this theory is represented in the diagram below.

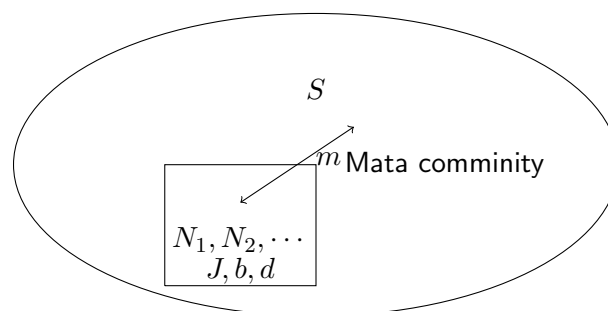


Figure 3.3: Diagrammatic representation of Bell's neutral argument with square in the meta community representing local community and N_i is the number of individuals belonging to different species in the local community, J is the size of local community, b and d are birth and death rates in the local community respectively, m is the migration rate and S is the fixed maximal number of species in the meta community

3.2.4 Spatially structured neutral model. This is another very important development in neutral theory. It is based on the relationship between offspring with their parents as far as their habitat is concerned. This theory is based on the fact that offspring are dispersed in some limited neighbourhood around their parents. In simulation of this model a space is considered as the grid in which one cell of the grid represent either a local community or a single individual such that if an individual gives birth in the cell then the offspring will remain in the cell for their entire life time. Therefore in the long run this will lead to creation of a clear distinction between local community and meta community as explained by [Chave \(2004\)](#).

3.2.5 The model. The derivation of species abundance distribution model is based on unified local and meta communities Markov chain for the stochastic dynamics of population growth shown below [He \(2005\)](#).

$$\frac{dP_{n,k}(t)}{dt} = P_{n+1,k}(t)d_{n+1,k} + P_{n-1,k}(t)b_{n-1,k} - P_{n,k}(t)(b_{n,k} + d_{n,k}) \quad (3.2.1)$$

where the parameters of (3.2.1) are defined as follows [Volkov et al. \(2003\)](#).

- (i) $P_{n,k}$ is the probability that k^{th} species will have n individuals at some time t .
- (ii) $b_{n,k}$ is the probability of birth for k^{th} species having n individuals defined as shown below.

$$b_{n,k} = (1 - \mu) \frac{n}{N} \left(\frac{N - n}{N - 1} \right) + \frac{\mu}{S} \left(1 - \frac{n}{N} \right) \quad (3.2.2)$$

- (iii) $d_{n,k}$ is the probability of death for k^{th} species having n individuals expressed as follows.

$$d_{n,k} = (1 - \mu) \frac{n}{N} \left(\frac{N - n}{N - 1} \right) + \frac{\mu}{S} \left((S - 1) \frac{n}{N} \right) \quad (3.2.3)$$

Derivations of equations (3.2.1), (3.2.2) and (3.2.3) are contained in [McKane and Sole \(2000\)](#). Population growth in (3.2.1) is subjected to the following boundary condition [McKane and Sole \(2000\)](#).

$$P_{-1,k}(t) = d_{0,k} = 0 \quad (3.2.4)$$

The boundary condition in (3.2.4) implies that the probability of death of k^{th} species with no individual is zero and this is equal to the probability that k^{th} species will contain -1 individuals. If we now assume that the community defined in (3.2.1) has total number of species as S , then the species abundance distribution of this community is defined as follows [He \(2005\)](#)

$$\langle \Phi(n) \rangle = \sum_{k=1}^S P_{n,k} \quad (3.2.5)$$

Equation (3.2.5) can be interpreted as the average number of species with n individuals [He \(2005\)](#). Equation (3.2.5) will only make sense if we find expression for stationary probability distribution $P_{n,k}$, based on [McKane and Sole \(2000\)](#) this is done by setting $\frac{dP_{n,k}(t)}{dt}$ in (3.2.1) to zero then we calculate for $P_{n,k}(t)$ as shown below.

$$P_{n+1,k}(t)d_{n+1,k} + P_{n-1,k}(t)b_{n-1,k} - P_{n,k}(t)b_{n,k} - P_{n,k}d_{n,k} = 0 \quad (3.2.6)$$

$$P_{n+1,k}(t)d_{n+1,k} - P_{n,k}(t)b_{n,k} = P_{n,k}(t)d_{n,k} - P_{n-1,k}(t)b_{n-1,k} \quad (3.2.7)$$

Equation (3.2.7) is true for all n , this implies that McKane and Sole (2000).

$$P_{n,k}(t)d_{n,k} - P_{n-1,k}(t)b_{n-1,k} = J \quad (3.2.8)$$

By now imposing the boundary condition if $n = 0$ in (3.2.8), we will have the following.

$$d_{0,k} = P_{-1,k} = 0 \quad (3.2.9)$$

Equation (3.2.9) implies that $J = 0$. Therefore equation (3.2.8) will reduce as shown below.

$$P_{n,k}(t)d_{n,k} - P_{n-1,k}(t)b_{n-1,k} = 0 \quad (3.2.10)$$

$$P_{n,k}(t)d_{n,k} = P_{n-1,k}(t)b_{n-1,k}; \quad n = 0, 1, \dots, N \quad (3.2.11)$$

If $\mu \neq 0$, then $d_{n,k} \neq 0$ for all n such that $0 < n \leq N$. From (3.2.11) then we have the following.

$$P_{n,k}(t) = \frac{b_{n-1,k}}{d_{n,k}} P_{n-1,k}(t); \quad n = 0, 1, \dots, N \quad (3.2.12)$$

Based on Moran (2008) argument relation in (3.2.12) leads to the following equation.

$$P_{n,k}(t) = \frac{b_{n-1,k}b_{n-2,k} \cdots b_{0,k}}{d_{n,k}d_{n-1,k} \cdots d_{1,k}} P_{0,k}(t); \quad n = 1, \dots, N \quad (3.2.13)$$

From (3.2.13) we now have the following.

$$P_{n,k} = P_{0,k} \prod_{i=0}^{n-1} \frac{b_{i,k}}{d_{i+1,k}} \quad (3.2.14)$$

When we now substitute (3.2.14) in (3.2.5) then we have species abundance distribution as follows.

$$\langle \Phi(n) \rangle = \sum_{k=1}^S P_{0,k} \prod_{i=0}^{n-1} \frac{b_{i,k}}{d_{i+1,k}} \quad (3.2.15)$$

$$\langle \Phi(n) \rangle = SP_{0,k} \prod_{i=0}^{n-1} \frac{b_{i,k}}{d_{i+1,k}} \quad (3.2.16)$$

In deriving local and meta community models, several assumptions has been made about the birth and death rates, by now considering birth and death as a linear process together with the immigration and emigration rates for every k^{th} species, then we have the following equation for birth and death rates He (2005).

$$b_{n,k} = b_k n + \lambda_k \quad \text{and} \quad d_{n,k} = d_k n + \mu_k \quad (3.2.17)$$

Where λ_k and μ_k are immigration and emigration rates respectively of k^{th} species. In some population birth and death are constants, however in this community under the study they are assumed to vary from species to species as defined by the subscript k . Based on the assumption of neutrality, whereby all species are assumed to have the same birth, death, immigration, emigration and speciation rates,

then from (3.2.17), we now ignore the subscript k in our calculations and we now have the following equations for birth and death rates He (2005).

$$b_n = bn + \lambda \text{ and } d_n = dn + \mu \quad (3.2.18)$$

By now substituting (3.2.18) in (3.2.16), then we have the following equation for species abundance distribution.

$$\langle \Phi(n) \rangle = \frac{SP_0(V + \lambda)(b + \lambda)(2b + \lambda) \cdots ((n - 1)b + \lambda)}{(d + \mu)(2d + \mu) \cdots (nd + \mu)} \quad (3.2.19)$$

In equation (3.2.19) we can observe that $b_0 = V + \lambda$, this implies that when the species become extinct that is $n = 0$, then it can only be replaced through speciation with the rate of V or via immigration at some rate λ . Since this is the case then the model is no longer meta community model but local community model coupled with meta community since immigration has been catered for in the equation (3.2.19) which is not the case for meta community He (2005). Equation (3.2.19) can further be simplified to give us a meaningful equation as follows.

$$\langle \Phi(n) \rangle = \frac{SP_0 b \left(\frac{V}{b} + \frac{\lambda}{b}\right) b \left(1 + \frac{\lambda}{b}\right) \cdots b \left((n - 1) + \frac{\lambda}{b}\right)}{d \left(1 + \frac{\mu}{d}\right) d \left(2 + \frac{\mu}{d}\right) \cdots d \left(n + \frac{\mu}{d}\right)} \quad (3.2.20)$$

$$\langle \Phi(n) \rangle = \frac{SP_0 \left(\frac{V}{\lambda} + 1\right) \frac{\lambda}{b} \left(1 + \frac{\lambda}{b}\right) \cdots \left((n - 1) + \frac{\lambda}{b}\right) \left(\frac{b}{d}\right)^n}{\left(1 + \frac{\mu}{d}\right) \left(2 + \frac{\mu}{d}\right) \cdots \left(n + \frac{\mu}{d}\right)} \quad (3.2.21)$$

From (3.2.21) we let

$$\alpha = \frac{\lambda}{b}, X = \frac{b}{d}, \theta = SP_0 \left(\frac{V}{\lambda} + 1\right) \text{ and } \beta = \frac{\mu}{d} \quad (3.2.22)$$

We now have the following

$$\langle \Phi(n) \rangle = \theta \frac{\alpha(1 + \alpha)(2 + \alpha) \cdots ((n - 1) + \alpha)}{(1 + \beta)(2 + \beta) \cdots (n + \beta)} (X)^n \quad (3.2.23)$$

Equation (3.2.23) is simplified using gamma function as follows.

$$\langle \Phi(n) \rangle = \theta \frac{\Gamma(n + \alpha)\Gamma(1 + \beta)}{\Gamma(\alpha)\Gamma(n + 1 + \beta)} X^n; \quad n = 1, 2, \dots \quad (3.2.24)$$

3.2.6 Interpretation and evaluation of the model. From equation (3.2.24), we can observe that it is equation representing local community model since migration processes i.e immigration and emigration has been included in the model. If we assume that there is no migration in the community under study, that is $\lambda = \mu = 0$ or $\beta = \alpha = 0$, then equation (3.2.24) will now reduce to log series distribution of meta community as illustrated below.

$$\langle \Phi(n) \rangle = \theta \frac{\Gamma(n)}{\Gamma(0)} \cdot \frac{\Gamma(1)}{\Gamma(n + 1)} X^n \quad (3.2.25)$$

Though $\Gamma(0)$ is undefined, for our calculation we will bend this rule so that we let.

$$\theta \frac{\Gamma(1)}{\Gamma(0)} = \gamma \quad (3.2.26)$$

When we now substitute (3.2.26) in (3.2.25), then we end up with the following log series distribution, which is one of the distributions that is used by different ecologists in describing species abundance distribution.

$$\langle \Phi(n) \rangle = \gamma \frac{X^n}{n} \quad (3.2.27)$$

In addition, if we ignore the effect of emigration in the model in (3.2.24) i.e. $\mu = 0$ or $\beta = 0$ then we end up with the following equation He (2005).

$$\langle \Phi(n) \rangle = \theta \frac{\Gamma(n + \alpha)}{\Gamma(\alpha)} \frac{X^n}{n!} \quad (3.2.28)$$

Distribution presented in (3.2.28), is truncated negative binomial distribution and this distribution is also one of the fundamental distributions that give log normal like shape of species abundance distribution.

From the model in equation (3.2.24), we can observe that it has two parameters α and β . Where α is used in measuring relative strength of the community based on immigration in relation to birth, therefore high values of α implies that there is high immigration relative to birth rate in the community. This will lead to left skewed species abundance distribution plot, implying that there are more common species as compared to rare species on the other hand small values of α implies that there is small immigration rate relative to birth rate and this leads to a right skewed plot of species abundance distribution hence there are more abundant rare species as compared to common species as shown in the plots below He (2005).

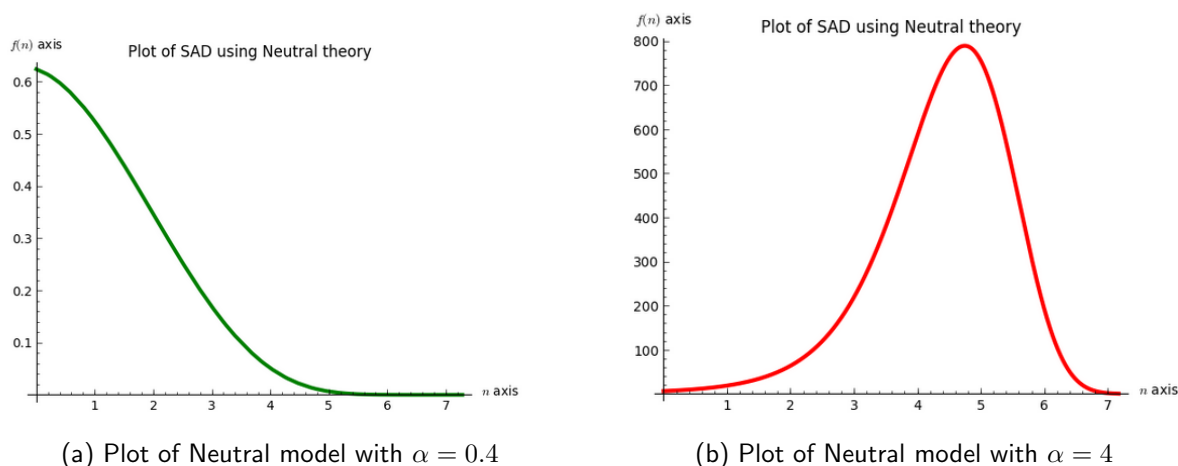


Figure 3.4: Illustration of effect of α in species abundance distribution model in (3.2.24) given $\theta = 5$, $\beta = 1$ and $X = 0.9$

Unlike α , β is used in measuring strength of community with respect to emigration in relation to death rate. Where irrespective of value of β in equation (3.2.24), this will lead to right skewed plot of species abundance distribution implying that rare species are more abundant as compared to common species as illustrated in the plot that follows.

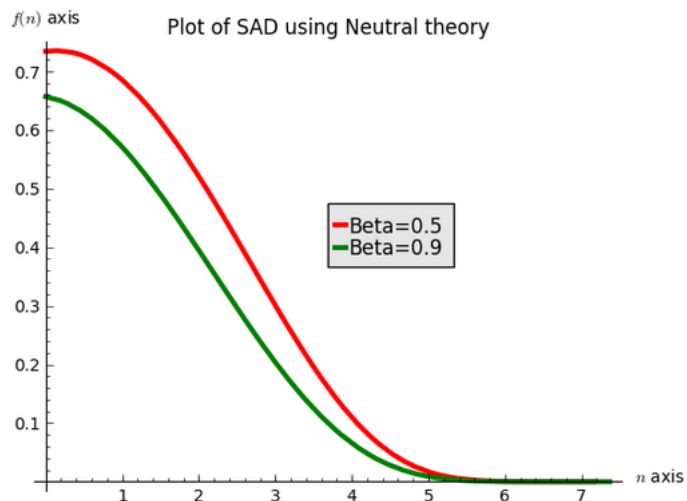


Figure 3.5: Plot of Neutral model with $\beta = 0.7$ and $\beta = 0.9$ with $\theta = 5$, $X = 0.9$ and $\alpha = 0.4$

This theory unifies both the local and meta communities therefore it is more advantageous than other neutral models due to its simplicity, generality and reduced number of assumption. Unlike other neutral models that assumes the conservation of both total number of species and community size, this models takes into consideration the assumption of conservation of total number of species and not the conservation of community size, hence reduced assumption.

One major problem with neutral theory is that during its derivation it is assumed that the immigration and emigration rates of the species are constant but this is not the case since emigration of species from meta community to local community will depend on abundance of a given species in the meta community, whereby more abundance species will immigrate faster as compared to rare species. On the other hand during formation of local community there is no competition involved, therefore immigration rate is assumed to be the same in all the species hence the assumption of constant immigration rate can only be applied in the case of local community but not in meta community emigration [He \(2005\)](#).

3.3 The Theory of Proportionate effect

This section discusses a theory that I developed that can also be used in describing species abundance distribution alongside other already existing theories. This theory is a modified version of the theory in [Brown and Sanders \(1981\)](#) which was used in describing distribution of income among the population of Sydney. This theory is based on partitioning an ecological community into groups of species with respect to different characteristics and then applying the law of proportionate effect in analysing distribution of species abundance. Consider a given ecological community having total species abundance N . The main interest is to determine species abundance distribution in this community. This will call for partitioning of the community into species groups S_i with N_i abundance based on the distinct characteristics of each species. Then we determine the distribution of abundance of each distinct species with respect to some change in time as far as evolution process in a given ecosystem is concerned, thereafter we will generalize the species abundance distribution of the entire community.

3.3.1 Assumptions of the model.

- (i) We assume that there is no predator-prey interaction in the community.
- (ii) We assume that there are other factors in the ecosystem that contributes to the abundance of a given species. i.e. biotic and abiotic factors.
- (iii) We assume that species in the community of interest are capable of giving birth and the offspring are able to give birth with the same proportion to other new offspring. This will lead to exponential growth of species abundance with time.

Let us assume that the abundance of species j is X_j . This abundance is a joint effect of a large number of mutually independent factors that contribute positively to its increase with respect to some point in time. It follows that after some step in time changes in j^{th} species abundance is proportion to a function $\Phi(X_{j-1})$, which is a function of already obtained species abundance. This proportionality is due to the fact that we assume that species abundance will depend on the already obtained species abundance. Mathematically this is represented as follows [Aitchison and Brown \(1957\)](#).

$$X_j - X_{j-1} \propto \Phi(X_{j-1}) \quad (3.3.1)$$

From (3.3.1), we now have the following.

$$X_j - X_{j-1} = \varepsilon_j \Phi(X_{j-1}) \quad (3.3.2)$$

where we assume that the set of values ε_j is mutually independent and independent of the set X_j . Now we need to apply the law of proportionate effect which is defined as follows [Aitchison and Brown \(1957\)](#).

3.3.2 Theorem (The law of proportionate effect). *A variate subjected to a process of change is said to obey the law of proportionate effect if the change in the variate at any step of the process is a random proportion of the previous value of the variate.*

From the law of proportionate effect in (3.3.2), Equation (3.3.2) reduces as shown below.

$$X_j - X_{j-1} = \varepsilon_j X_{j-1} \quad (3.3.3)$$

From (3.3.3), we have the following.

$$X_j = (1 + \varepsilon_j) X_{j-1} \quad (3.3.4)$$

From the relation in (3.3.4), we now have the following Crow and Shimizu (1988).

$$X_n = X_0 \prod_{j=1}^n (1 + \varepsilon_j) \quad (3.3.5)$$

When we now introduce natural logarithm in equation (3.3.5), we end up with the following equation.

$$\log X_n = \log X_0 + \sum_{j=1}^n \log(1 + \varepsilon_j) \quad (3.3.6)$$

By assuming that the absolute value of ε_j is less than one, from Taylor's Expansion of $\log(1 + x)$, Equation (3.3.6) will reduce to the following approximation.

$$\log X_n = \log X_0 + \sum_{j=1}^n \varepsilon_j \quad (3.3.7)$$

Then by an additive form of the central limit theorem $\log(X_n)$ is asymptotically normally distributed and this implies X_n is asymptotically log normally distributed Aitchison and Brown (1957). So far we have proved that j^{th} species abundance distribution with respect to some time step is asymptotically lognormally distributed. The species abundance distribution of the entire community is as follows. Assume the total abundance of all the species in the ecological community to be N and defined as follows.

$$N = \sum_{i=1}^j N_i, \quad i = 1, 2, \dots, j \quad (3.3.8)$$

where N_i is the abundance of species i . Since we have shown that the species abundance for one species is lognormally distributed, then abundance for species S_i , where $i = 1, 2, \dots, j$ is lognormally distributed. We now need to introduce a theorem that was used by Beaulie and Xie (2004) in optimal lognormal approximation to lognormal sum distribution.

3.3.3 Theorem (An Optimal Lognormal Approximation to Lognormal Sum Distribution). *The sum of identically independent lognormally distributed random variable is lognormal.*

The proof of the theorem is contained in Beaulie and Xie (2004). It follows that by summing the abundance of all the species in the community as shown in (3.3.8), then the distribution of this sum will be lognormal based on theorem (3.3.3) which is the distribution of SAD. Hence we have proved that species abundance distribution in a given ecological community is lognormal.

3.3.4 Evaluating the model. In evaluating this model approximated mean $\bar{\mu}$ and variance $\bar{\sigma}$ for the total species abundance for all the years in BCI dataset were used to fit density curve over histogram plot for a single year SAD and in doing so the following plot was obtained.

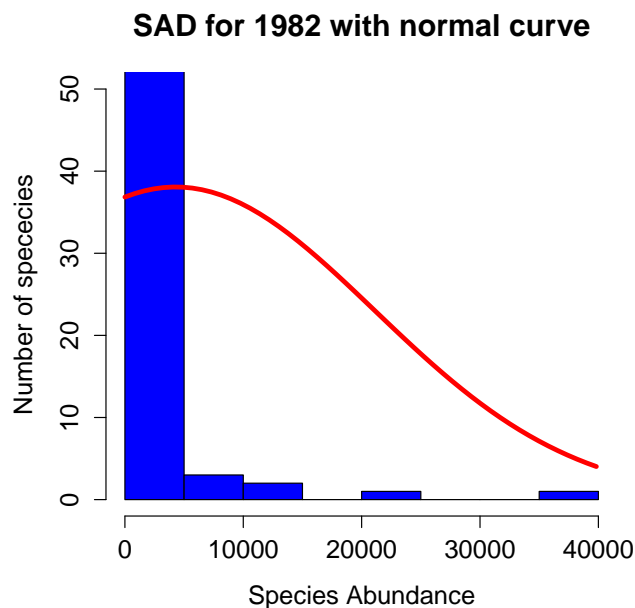


Figure 3.6: Evaluation of theory of proportionate effect by fitting normal curve over histogram SAD for 1982 obtained from BCI (2006), using approximated mean and standard deviation for the total SAD as $\bar{\mu} = 4289.69375$ and $\bar{\sigma} = 16766.03597$

From figure (3.6), we can see that there is a mismatch between the histogram SAD plot for the year 1982 and normal curve fitted over it, this implies that even though from theoretical point of view my theory is convincing and describes SAD's lognormal shape but it doesn't give exact fit to SAD data, therefore we can conclude that there is still more to be done in this theory to evaluate it's weakness and validate whether it is applicable or not in describing SAD so that it can be introduced in ecology or ignored as far as species abundance distribution is concerned. However, for now the two discussed theories, Maximum Entropy Theory of Ecology and Neutral theory are the best in describing species abundance distribution.

4. Conclusion and Future Work

4.1 Conclusion

This project was aimed at describing the forces behind lognormal shape of species abundance distribution or the right skewed shape of species abundance distribution. Different parametric and mechanistic approaches describing SAD shape is presented. However, for the interest of time no simulation was done using distributions and methods discussed in this project. Furthermore, based on the two major approaches discussed it was observed that apart from lognormal distribution, SAD is also explained using fisher's log series distribution as well as negative binomial distribution and geometric distribution. In addition, Neutral theory from its meta and local communities point of view it is clear that SAD is explained using log series distribution in the meta community and truncated negative binomial distribution in local community. Moreover, I proposed a theory that is based on the law of proportionate effect that can also be used in describing species abundance distribution and this theory is not yet recognized by ecologists but I hope that in the near future after a thorough evaluation and scrutiny the theory will be introduced into ecology. In general from the whole project we conclude that ecological community is characterised by rarity and abundance of species with rare species i.e. singleton and doubleton species being more abundance as compared to common species.

4.2 Future work

To clearly understand species abundance distribution it is important to do some simulations using the real data for species abundance that is collected from different ecological community. Therefore it is important to use the theoretical model derived in this project for simulations. In addition, I developed a model that has never been presented anywhere in ecology and therefore it is very important for more research to be done about this model to validate and analysing its strength and weaknesses as far as study of species abundance distribution is concerned.

Acknowledgements

First and foremost I would like to thank The Almighty God for this far He has brought me. Indeed He is "Ebeneza". Secondly I acknowledge my supervisor Dr. Cang Hui for his assistance and guidance to ensure that this project is a success. I would also like to thank the entire AIMS family for their support for the time I have stayed at AIMS. I also acknowledge my dear friend Irene for helping me edit my grammer. Last but not least I give special thanks to my family for their support in my education, specifically to my loving parents Mr. and Mrs. Edward Ochiaga and not forgetting my guardian Mrs. Mary Akunja. May God bless you all. Shalom, Shalom, Shalom!!!!

References

- Bci species abundance. Harvard University, 2006. URL <https://ctfs.arnarb.harvard.edu/webatlas/datasets/bci/abundance/bciN10.html>. Accessed May 2014.
- J. Aitchison and J. A. Brown. *The Lognormal Distribution*. Cambridge University Press, 1957.
- N. C. Beaulie and Q. Xie. An optimal lognormal approximation to lognormal sum distribution. 53: 479–489, 2004.
- G. Bell. The distribution of abundance in neutral communities. *The American Naturalist*, 155:606–617, 2000.
- M. G. Bowler and C. K. Kelly. The general theory of species abundance distributions. 2010.
- G. Brown and J. W. Sanders. Lognormal genesis. *Applied Probability*, 18:542–547, 1981.
- J. Chave. Neutral theory and community ecology. *Ecology*, 7:241–253, 2004.
- K. Cherukumilli. Testing the maximum entropy theory of ecology for predicting species abundance distribution and species area relationship in control and warmed plots at the rocky mountain biological laboratory. 2012.
- R. A. Chisholm and J. W. Lichstein. Linking dispersal, immigration and scale in the neutral theory of biodiversity. *Ecology*, 12:1385–1393, 2009.
- E. L. Crow and K. Shimizu. *Lognormal Distribution, Theory and Applications*. Marcel Dekker, Inc, 1988.
- C. Furber, M. Evans, N. Hastings, and B. Peacock. *Statistical distribution*. John Wiley and Sons, Inc., 1957.
- J. Harte. *Maximum Entropy and Ecology, A theory of Abundance, Distribution and Energetics*. American Mathematical Society, 2011.
- F. He. Deriving a neutral model of species abundance from fundamental mechanism of population dynamics. *Functional Ecology*, 19:187–193, 2005.
- F. He and X.-S. Hu. Hubbell's fundamental biodiversity parameter and simpson diversity index. *Ecology*, 8:386–390, 2005.
- S. P. Hubbell. *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton University Press, 2001.
- B. J. McGill and A. E. Magurran. *Biological Diversity, Frontiers in measurements and assessment*. Oxford University Press, 2011.
- A. McKane and D. A. R. V. Sole. A mean field stochastic theory for species-rich assembled communities. 2000.
- P. A. P. Moran. Random process in genetics. *Mathematical Proceedings of the Cambridge Philosophical Society*, 54:60–71, 2008.

-
- J. C. Nekola, A. L. Sizing, A. G. Boyer, and D. Storch. Artifacts in the log-transformation of species abundance distribution. *Folia Geobot*, 43:259–268, 2008.
- E. P. Odum and G. W. Barrett. *Fundamental of Ecology 5th edition*. Thomsom Books, 1953.
- E. . C. Pielou. *Mathematical Ecology*. A Wiley-Interscience publication, 1977.
- E. C. Pielou. *An Introduction to Mathematical Ecology*. Wiley-Interscience, 1969.
- I. A. B. Sognnæs. Maximum entropy and maximum entropy production in macroecology, 2011. Master of Science in Physics and Mathematics.
- I. Volkov, J. R. Banavar, S. P. Hubbell, and A. Maritan. Neutral theory and relative species abundance in ecology. *Nature*, 424:1035–1037, 2003.
- I. Volkov, J. R. Banavar, F. He, S. P. Hubbell, and A. Maritan. Density dependence explains tree species abundance and diversity in tropical forest. *Nature*, 438:658–661, 2005.
- D. D. Wackerly, W. M. III, and R. L. Scheaffer. *Mathematical Statistics with Applications*. Thomas Learning, Inc, 2008.