



Fitting abundance distribution models in tropical arboreal communities of SE Brazil

R. Cielo Filho¹, F. R. Martins^{1,3} and M. A. Gneri²

¹*Department of Botany, Institute of Biology, Campinas State University. Caixa postal 6109, Campinas 13083-970, SP, Brasil. E-mail: fmartins@obelix.unicamp.br*

²*Department of Statistics, Institute of Mathematics, Statistics and Computer Science, Campinas State University, Caixa postal 6065, Campinas 13083-970, SP, Brasil. E-mail: gneri@ime.unicamp.br. Author of Appendices 1-2*

³*Corresponding author*

Keywords: Cohens's table, Community structure, Disturbance, Diversity, Simpsons's index variance, Stress, Tropical forests.

Abstract: The species abundance distribution of ecological communities has been represented through several mathematical models, of which the most common are: geometric series, logseries, lognormal, and a type of broken stick, this latter found only in animal communities. There is no consensus on the underlying biological processes, but initial observations on plant communities related these models to equilibrium and high richness (lognormal), stress or disturbance and low richness (logseries and geometric series). Recently the value of these relationships was challenged, and other descriptors were considered better predictors of richness, disturbance and stress. We aimed at investigating how these models and their parameters, as well as dominance and evenness are related with species richness, stress and disturbance in six tropical forest communities, SE Brazil: two well-conserved fragments, two disturbed by fire, and two swampy forests (anoxic stress). The models did not show consistent relationships with richness, disturbance or stress. The parameters and indices of diversity α (logseries) and λ (lognormal) varied closely with richness, and the dominance was larger in the communities submitted to stress or disturbance. Our results indicate the need of further studies in order to validate (or refute) the use of abundance distribution models for detection of patterns related to richness, stress or disturbance in tropical arboreal communities. On the other hand, richness and dominance did respond to disturbance and stress.

Introduction

The investigation of the species abundance distribution of a community is one of many possible ways to study biodiversity, and it allows establishing a mathematical model to represent the relationship between the number of species and their abundance. If the data adjust to the distribution generated by the model, then a parameter of the distribution may be used as an estimator of the diversity (Greig-Smith 1983). The abundance distribution among the species in an ecological community has been represented by mathematical and graphic formulations and explained verbally through biological arguments generally associated to a mathematical model (MacArthur 1957, Whittaker 1972, May 1975, Sugihara 1980).

Mathematical models

The basic mathematical models of abundance distribution are the geometric series (Motomura 1932 apud Whittaker 1965), logseries (Fisher et al. 1943), lognormal (Preston 1948) and a type of broken stick (MacArthur

1957). If the observed abundance distribution fits the distribution generated by the geometric series model, the parameter k may be used to express the inverse of the diversity (Martins and Santos 1999). In the logseries model, the parameter α is used as diversity index (Fisher et al. 1943). In the lognormal model the number λ , the ratio between the parameters S^* (number of species in the community) and σ (standard deviation of the distribution), is used as diversity measure (Magurran 1988). In the broken stick model, since the species abundance distribution is highly even (high equitability), the parameter S (observed number of species) can be used as a diversity measure (Magurran 1988).

Assuming the canonical form for the lognormal model (Preston 1948), $\alpha = 5$ for the logseries, and $k = 0.4$ for the geometric series, May (1975) demonstrated that the equitability of the distributions generated by the lognormal and broken stick models increases with the number of species. From approximately 10 species on, the broken stick model always predicts the largest equitability, the geometric series and logseries the smallest, and

the lognormal generates intermediary equitability. This gradient of equitability, described mathematically by May (1975), has been observed in nature (Magurran 1988), once the communities fit to the respective models with the values of α and k close to 5 and 0.4, respectively.

Biological models

These models are used to make inference on community organization, that is, on the ecological or evolutionary processes that result in a given abundance distribution. The abundance distributions observed in natural communities may be explained through biological hypotheses (Magurran 1988). The biological hypotheses assume that if the abundance of a species were proportional to the amount of resources it uses, then the abundance distribution would reflect the niche structure in the community (Sugihara 1980). The niche pre-emption hypothesis (Whittaker 1972) is associated to the geometric series model: the niche hyperspace is emptied as a few species arrive one by one at constant time intervals in a habitat dominated by a strong ecological factor, and each one takes the same fraction of the remaining hyperspace. This hypothesis is also associated to the logseries model, but the species would arrive at aleatory intervals (May 1975). Sugihara (1980) proposed a biological hypothesis associated to the lognormal model: the distribution of abundance among the species would come from an aleatory sequential breakage of the niche hyperspace, until all the species have been accommodated. In the biological model associated to the broken stick, the niche space is considered unidimensional and represented as a stick that is aleatory and simultaneously broken in as many pieces as the number of species (MacArthur 1957).

Reasoning this way could lead to a biological explanation whenever a mathematical model represents the abundance distribution observed in a natural community. The biological processes acting in the generation of such a model could happen at an ecological or evolutionary scale of time. At the ecological time scale, the species abundance would be the outcome of competition and other biotic interactions; at the evolutionary time scale, the abundance would be an innate property of the species as a result of coevolution (Wilson 1991).

On the other hand, purely statistical explanations have also been applied to abundance distributions represented by certain mathematical models (May 1975), and critics have been made to the traditional biological hypotheses (Ugland and Gray 1982). Cohen (1968) demonstrated that different and even contradictory hypotheses could generate the same abundance distribution. Therefore, the study of the abundance distribution is not enough to interpret

and explain the processes acting in an ecological community (Whittaker 1972). If these limitations are kept in mind and no biological explanation is evoked, the mathematical models can be used as tools for detecting consistent patterns among natural communities, and thus could provide good assistance in the understanding of the organization of ecological communities. From now on, we will deal only with the mathematical models.

The broken stick model proved satisfactory for small communities of taxonomically related organisms (Pielou 1975, Magurran 1988), but plant communities in general do not fit this model (Whittaker 1965, Wilson 1991, Wilson and Gitay 1995, Watkins and Wilson 1994). The lognormal model is frequently found in mature communities with many species (Whittaker 1965, May 1975, Hughes 1986, Gray 1987), while the geometric and logarithmic series occur in poor, disturbed communities in restrictive environments (MacArthur 1960, Whittaker 1965, 1972, Gray 1987). According to Hill and Hamer (1998), as the lognormal and the logseries models are frequently found in undisturbed and disturbed habitats, respectively, they have been used as disturbance indicators in natural communities.

The empirical relationships between abundance distribution models and the community characteristics described above have been recently challenged (Watkins and Wilson 1994, Wilson and Gitay 1995, Wilson et al. 1996, Wilson et al. 1998). According to Wilson (1991), limitations in the methods of fitting observed data to the different mathematical models blur the value of the initial attempts to relate the observed abundance distribution of communities to the ecological circumstances. The recurrent failure in finding consistent patterns involving abundance distribution has been attributed to the individualistic character of the communities studied (Watkins and Wilson 1994, Wilson and Gitay 1995). However, the mathematical models of abundance distribution may not be good indicators of the ecological mechanisms governing community organization (Watkins and Wilson 1994). In this case, other characteristics, like equitability may be more efficient in revealing patterns. As discussed previously, it is expected that, in general, communities with similar number of species and different types of abundance distribution have different degrees of equitability. On the other hand, communities with different number of species fitting the same model may also have different degrees of equitability (May 1975). Moreover, using abundance classes, such as the lognormal octaves (Preston 1948), implies grouping in the same class species with quite different abundance in different communities. For example, if the most abundant class interval ranges from

Table 1. Characteristics of the six forests considered for the fitting to the models of abundance distribution. MRC = Ribeirão Cachoeira, SVI = São Vicente, PED = Pedreira, POM = Pomar, BIA = Campinas Agronomic Institute, and BSG = Santa Genebra. Climatic classification according to Koeppen.

Abbreviation of the community	Local	Climate	Reference
MRC	Campinas/SP	Cwa	Cielo Filho 2001
SVI	Campinas/SP	Cfa	Bernacci 1992
PED	Piracicaba/SP	Cwa	Catharino 1989
POM	Piracicaba/SP	Cwa	Tabanez et al. 1997
BIA	Campinas/SP	Cwa	Torres et al. 1994
BSG	Campinas/SP	Cwa	Toniato et al. 1998

128 to 256 individuals, species of different communities will be put into this class, since their abundance falls in this range. The consequence of using wide abundance classes is that different communities with great differences in the abundance magnitudes of the most abundant species may present similar frequency of the most abundant classes, thus fitting the same model, despite their different degrees of equitability.

Our objective is to investigate whether these mathematical models, their parameters, and equitability can discriminate between different plant communities. Following the literature, we expect that well-conserved, species-rich communities fit the lognormal model, but communities disturbed by fire, with few species or under stressing conditions would fit the logseries or geometric series models. We expect no community data to fit the broken stick model.

Materials and methods

Data collection

We used data from four surveys in the seasonal semideciduous forest and two surveys in swampy forests, all in the state of São Paulo, SE Brazil (Table 1). Ribeirão Cachoeira (MRC) and São Vicente (SVI) are well-conserved semideciduous forest fragments, without notice of recent anthropic disturbance. Pedreira (PED) and Pomar (POM) have been disturbed by fire until the decade of 1980. Campinas Agronomic Institute (BIA) and Santa Genebra (BSG) are swampy forests, in which the environment is very restrictive due to anaerobic stress (Joly 1986, Joly 1991), caused by overflowing and uprising of the water table (Toniato et al. 1998).

Fitting and goodness of fit tests

We considered the number of individuals in each species as abundance measure, since, according to Magurran (1988), in large samples with frequent species, the

number of individuals behaves as a continuous variable. Generally, the rarest species of a community are not included in a sample, and then a lognormal distribution can be veiled, that is, the classes on the left side (the rarest species) of the distribution will be absent. In this case, the distribution is truncated at a point known as the veil line (Preston 1948). Pielou's (1975) method was used to fit the truncated lognormal model.

We used the abundance distribution histogram to fit the models of the broken stick, truncated lognormal, and logseries, and the log-transformed abundance rank diagram to fit the geometric series (Whittaker 1965, Pielou 1975, May 1975, Magurran 1988). When using abundance distribution histogram, the abundance classes are generally plotted on the x-axis on a logarithm to the base of 2, while the y-axis shows the number of species in each class. When using rank diagram, the x-axis presents the position of each species in a rank order of decreasing abundance, and the y-axis shows its \log_2 -transformed abundance. In the abundance distribution histograms we used the following upper limits of the abundance classes for the truncated lognormal: 0.5; 1.5; 2.5; 4.5; 8.5; 16.5; ...; until the upper limit of the most abundant species class. For the other models we used 1.5; 2.5; 4.5; 8.5; 16.5; ...; until the upper limit of the most abundant species class.

The goodness of fit was tested by the one-sample Kolmogorov-Smirnov test (Sokal and Rohlf 1995, pp. 711-714). This test is recommended for testing agreement to models of abundance distribution (Hill and Hamer 1998, Basset et al. 1998) since it is more powerful than the chi-square statistic. However, using the Kolmogorov-Smirnov test with grouped data and theorized distribution with parameters estimated from the sample results in a conservative test (Conover 1999). Nevertheless, following literature recommendation, we used this test since it is more appropriate to deal with distributions and did not present different results from the chi-square.

In the goodness of fit test for the truncated lognormal the first class must be disregarded. To test the goodness of fit to the geometric series, we transformed the theoretical rank abundance diagram generated by this model into an abundance distribution histogram by taking from the y-axis of rank abundance diagram the same abundance intervals as specified above and counting the number of species in the x-axis falling in each interval.

The following equations describe the distributions tested here:

Broken stick

$$S_{(n)} = [S(S-1)/N] (1 - n/N)^{S-2},$$

where $S_{(n)}$ = number of species with n individuals; N = total number of individuals; and S = observed total number of species.

Truncated lognormal

$$S_e = p_e S^*,$$

where S_e = cumulative number of species in the e^{th} abundance class; p_e = cumulative proportion of the e^{th} abundance class in the normal curve; e = abundance class (0 = 0-0.5, 1 = >0.5-1.5, 2 = >1.5-2.5, 3 = >2.5-4.5, 4 = >4.5-8.5, 5 = >8.5-16.5, ..., until the most abundant class); and S^* = total estimated number of species.

The following procedure to fit a truncated lognormal distribution was described by Pielou (1975). In a lognormal distribution, the logarithms of the distribution values have normal distribution (Zar 1999). Therefore, to obtain p_e it is necessary to

1. Calculate the mean (\bar{x}) and variance (s^2) of the \log_{10} -transformed abundance of all the species (x_i):

$$\bar{x} = \sum_{i=1}^S x_i / S$$

and

$$s^2 = \sum_{i=1}^S (x_i - \bar{x})^2 / S$$

2. Calculate $\gamma = s^2 / (\bar{x} - ls_0)^2$, where $ls_0 = -0.30103$, that is, $\log_{10}(0.5)$, the upper limit of the non-sampled species class of abundance;

3. Obtain the value of θ for the calculated γ . This value can be found in the table of Cohen (1961, Magurran 1988, see Appendix 1 for a description of generating Cohen's table). The value θ allows to estimate μ_x and V_x , respectively the population mean and variance of x_i :

$$\mu_x = \bar{x} - \theta (\bar{x} - ls_0)$$

and

$$V_x = s^2 + \theta (\bar{x} - ls_0)^2$$

4. Determine p_e by substituting ls in the equation $z = (ls_e - \mu_x) / \sqrt{V_x}$ for the \log_{10} of the upper limit of the e^{th} abundance class and entering with the value of z obtained in a table of proportions of the normal curve (Zar 1999). This table gives directly the value of p_e only when z is negative. If z is positive, make $p_e = 1$ minus the value given in the table. The upper limit of the class 0 is 0.5, that is, the point where it is admitted that the observed distribution is truncated. To determine S^* the value of p_0 should be entered in the expression $S^* = S/(1-p_0)$. The non-cumulative number of species in the e^{th} abundance class is obtained by making $S_e - S_{e-1}$. The diversity index of the lognormal distribution is then calculated as $\lambda = S^*/\sigma$, being $\sigma = \sqrt{V_x}$.

Logarithmic series

$$S_{(n)} = \alpha x^n / n,$$

where $S_{(n)}$ = number of species with n individuals; x = constant ($x < 1$); and $\alpha = [N(1-x)]/x$, which is a constant and a diversity index.

To obtain x , the expression $[(1-x)/x][-\ln(1-x)]$, should be iterated until the result approximates S/N .

Geometric series

$$n_j = NC_k k(1-k)^{j-1},$$

where n_j = number of individuals of the species in the j^{th} position of decreasing order of abundance; $C_k = [1 - (1-k)^S]^{-1}$ is a normalization constant, which assures that $\sum n_j = N$; and k = proportion of the remaining abundance allocated for each species. To obtain k , the expression $[k/(1-k)][(1-k)^S] / [1 - (1-k)^S]$ should be iterated until it approximates n_{min}/N , where n_{min} is the number of individuals of the least abundant species. The diversity index of the geometric series is then calculated as $1/k$.

We estimated the species richness $S_{(m)}$ by the rarefaction method (Hurlbert 1971), the diversity by the parameters of the abundance distribution models that fitted the data, and the equitability by the evenness index E (Bulla 1994) and dominance index I (Simpson 1949):

$$I = [N(N-1)]^{-1} \sum_{i=1}^S n_i(n_i-1),$$

where n_i = number of individuals of species i .

The rarefaction method eliminates the influence of the sample size variation by furnishing an expected richness value for a standard sample size m . Thus, it allows comparison between the communities (Hurlbert 1971, Magur-

Table 2. Species richness ($S_{(380)}$), diversity indices (α , λ), evenness (E) and dominance (I) in different forests: Ribeirão Cachoeira (MRC), São Vicente (SVI), Pedreira (PED), Pomar (POM), Campinas Agronomic Institute (BIA), and Santa Genebra (BSG). Values of ($S_{(380)}$), α , I and E are accompanied by the 95% confidence interval.

Parameters	MRC	SVI	PED	POM	BSG	BIA
$S_{(380)}$	81.1 (7.83)	71.3 (5.52)	62.7 (4.70)	44.4 (3.92)	32.9 (2.91)	24.5 (3.92)
α	34.1 (6.14)	26.7 (5.70)	22.0 (5.12)	13.6 (3.88)	8.7 (2.83)	6.7 (2.23)
λ	211	156	137	88	56	46
E	0.36 (0.029)	0.39 (0.043)	0.35 (0.045)	0.33 (0.039)	0.40 (0.038)	0.34 (0.018)
I	0.042 (0.0192)	0.057 (0.0262)	0.101 (0.0279)	0.108 (0.0282)	0.130 (0.0275)	0.117 (0.0208)

ran 1988), although the result of such a comparison depends on m . The standard sample size cannot be greater than the smallest sample size of the communities compared. The smallest sample size in the communities we took for comparison was 522. However, due to computing problems, we adopted 380 individuals as the standard sample size ($m = 380$). The confidence interval of $S_{(380)}$ was calculated using its large-sample variance $\left(S_{S(m)}^2\right)$, obtained through the formula provided by Heck et al. (1975), in the following equation:

$$IC_{S_{(380)}} = 1.96 \sqrt{S_{S_{(380)}}^2}.$$

The evenness index E allows to calculate the confidence interval (Bulla 1994), thus increasing the consistency of the comparison among samples of different sizes and species number, which are factors strongly affecting the evenness (Hurlbert 1971, Hill 1973). The dominance, estimated by I , suffers little or no influence of the sample size (Simpson 1949, Magurran 1988). The variance of I was approximated using a formula derived from the Simpson's formula for its variance (Simpson 1949, Appendix 2):

$$\text{Variance}(I) \leq 4/N \cdot P_{MAX} + 4/[N(N-1)],$$

where: P_{MAX} was obtained from a set of 64 forests in the state of São Paulo compiled in the data bank system FI-TOGEO (Scudeller and Martins 2002), updated by R. J. de Oliveira (personal communication). In that set of forests, P_{MAX} takes the value of 0.025. The confidence interval of I was calculated through the formula:

$$IC_I = 1.96 \sqrt{S_I^2}.$$

The parameters λ and α are correlated and have high discriminating power, but α is preferable for being easier to calculate and for allowing the determination of its confidence interval (Magurran 1988).

Results

Species richness

The estimated richness varied from 24.5 (± 3.92) to 81.1 (± 7.83) species for BIA and MRC, respectively. The richness values were highest for the well-conserved communities, smallest in the swampy forests, and intermediate in the disturbed forests (Table 2, Fig. 1).

Broken stick

The D_{max} values, which measure the deviation between the observed and the expected distributions, were very high to be considered as achieved by chance, since the probability of finding such a so high value is very small. Therefore, the null hypothesis (observed distribution equal to the expected by the model) was rejected for all the communities studied (Table 3). The use of the parameter S as diversity estimate only makes sense if the observed distribution fits the model well, with high equitability and samples of similar size (Magurran 1988).

Lognormal

All the communities studied fitted the lognormal model (Table 3). The diversity estimated by λ was small-

Table 3. Values of D_{max} and p for the goodness of fit test for the models of the broken stick (VQ), lognormal (LN), logseries (LS) and geometric series (SG) in different forests: Ribeirão Cachoeira (MRC), São Vicente (SVI), Pedreira (PED), Pomar (POM), Campinas Agronomic Institute (BIA) and Santa Genebra. Significantly values are in boldtype, S = number of species. The tested hypothesis is H_0 : the observed distribution is equal to the expected for the respective model.

Communities	S	VQ	LN	LS	SG
MRC	119	0.31756	0.05974	0.10888	0.23529
		$p < 0.001$	$p > 0.5$	$p > 0.1$	$p < 0.001$
SVI	84	0.25012	0.03929	0.06991	0.21429
		$p < 0.001$	$p > 0.5$	$p > 0.5$	$p < 0.001$
PED	71	0.31734	0.05789	0.10866	0.26761
		$p < 0.001$	$p > 0.5$	$p > 0.2$	$p < 0.001$
POM	50	0.34838	0.04791	0.12433	0.26000
		$p < 0.001$	$p > 0.5$	$p > 0.2$	$p < 0.002$
BSG	36	0.27291	0.05676	0.04729	0.16667
		$p < 0.01$	$p > 0.5$	$p > 0.5$	$p > 0.2$
BIA	33	0.39471	0.07054	0.15023	0.27273
		$p < 0.001$	$p > 0.5$	$p > 0.2$	$p < 0.02$

est for BIA (46) and largest for MRC (211), accompanying the variation of richness (Table 2, Fig. 2).

Logseries

All the communities studied fitted the logseries model (Table 3). The values of α varied between 6.7 (± 2.23) and 34.1 (± 6.14) for BIA and MRC, respectively, accompanying the variation in species richness (Table 2, Fig. 3).

Geometric series

Only one community (BSG) fitted the geometric series model (Table 3), and we did not calculate the diversity index $1/k$.

Evenness (E) and Dominance (l)

POM and BSG had the smallest and largest evenness values, respectively, $E = 0.33$ (± 0.039) and $E = 0.40$ (± 0.038) (Table 2, Fig. 4). The dominance showed an inverse variation in relation to richness, α and λ , with BSG and BIA exchanging their positions (Table 2, Fig. 5). The largest value of l was found for BSG (0.130 ± 0.0275), the smallest for MRC (0.042 ± 0.0192).

Discussion

Broken stick

Our results agreed with the statement that plant communities in general do not follow the broken stick model

(Whittaker 1965, Wilson 1991, Watkins and Wilson 1994, Wilson and Gitay 1995). According to Greig-Smith (1983), the model is inappropriate for plant communities because it does not imply high dominance, which may be strong in many plant communities. We observed that the model underestimates not only the number of very abundant species but also the number of very rare ones. In fact, the generalised disagreement between theoretical and empirical distributions occurs mainly because the broken stick model fails to predict the relatively large number of singleton species found in the communities.

Lognormal

Since all the communities fitted this model, it showed no relationship with richness, environmental stress or disturbance. Hill and Hamer (1998) claimed that the lognormal model could be used as an indicator of no disturbance in natural communities. However, Nummelin (1998) demonstrated that the goodness of fit to this model depends on the time elapsed after the disturbance and on the community type and that the disagreement with the lognormal model cannot be considered a simple and universal indicator of disturbance. On the other hand, our results showed that even disturbed communities fit the lognormal model, thus discouraging its use as a disturbance indicator. The same may be said of its use as an indicator of stress or richness. The visual analysis of the abundance distributions suggests that the lognormal model does not give the best fit to all the communities we studied. This

means that the application of better analysis methods, which allow distinguishing the model with the best fit, could improve the conclusions (Wilson 1991). However, such analyses have also failed in detecting consistent patterns between the lognormal model and species richness or environmental stress or disturbance (Watkins and Wil-

son 1994, Wilson and Gitay 1995, Wilson et al. 1996, Wilson et al. 1998).

Logseries

All communities fitted the logseries model, but they also fitted the lognormal model. The agreement of the same data set to both models is frequent when the lognormal distribution is truncated close to the modal abundance class (Magurran 1988). The veil line hides the rare species, but as sampling effort increases, it moves to the left and a clear distinction between the lognormal and the logseries models is then possible (Magurran 1988). This happens because the logseries does not give good fitting for the portion of the curve lying on the left of the modal abundance class (Magurran 1988). Studies in tropical forests have confirmed this trend (Hubbell and Foster 1983, Leigh 1999). Hence, the goodness of fit to the logseries model may be just a consequence of sample size (May 1975). This could explain the agreement of MRC and SVI, well-conserved forests, in following the logseries model. On the other hand, the agreement of the disturbed and stressed forests to the truncated lognormal model could disappear if we had used the non-truncated lognor-

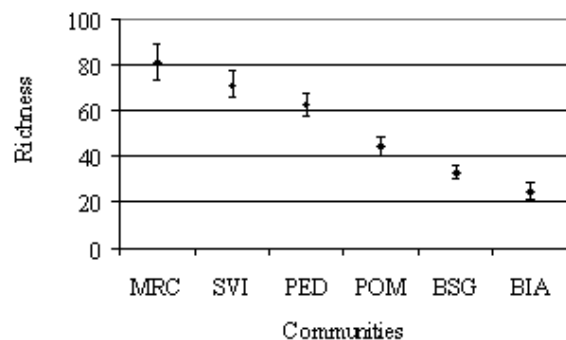


Figure 1. Species richness ($S_{(380)}$) estimated by the rarefaction method and the respective 95% confidence interval for different forests: Ribeirão Cachoeira (MRC), São Vicente (SVI), Pedreira (PED), Pomar (POM), Campinas Agro-nomic Institute (BIA), and Santa Genebra (BSG).

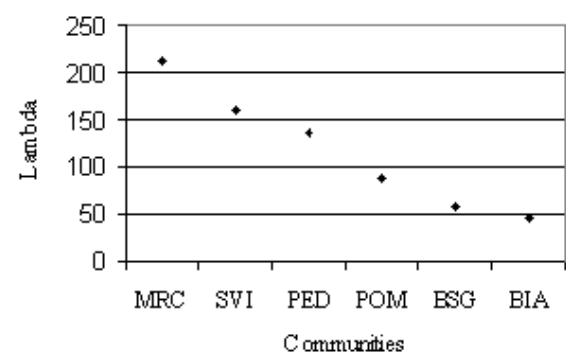


Figure 2. Diversity estimated by the index λ of the lognormal model for the forests studied (see caption to Fig. 1, for abbreviations).

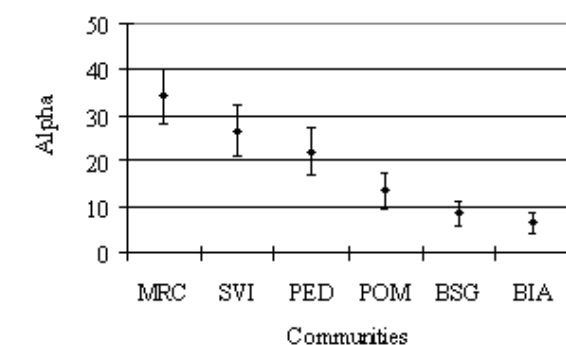


Figure 3. Diversity estimated by the α index of the logseries model and the respective 95% confidence interval. (See caption to Fig. 1, for abbreviations).

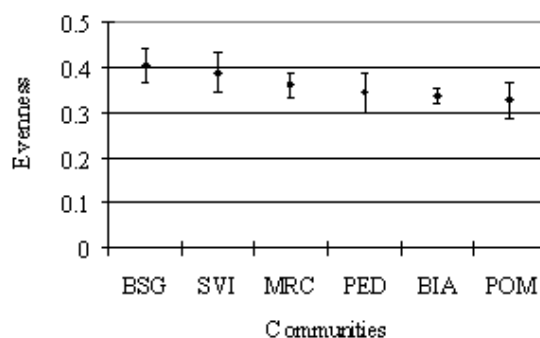


Figure 4. Evenness estimated by the index E and the respective 95% confidence interval for the forests studied (see caption to Fig. 1, for abbreviations).

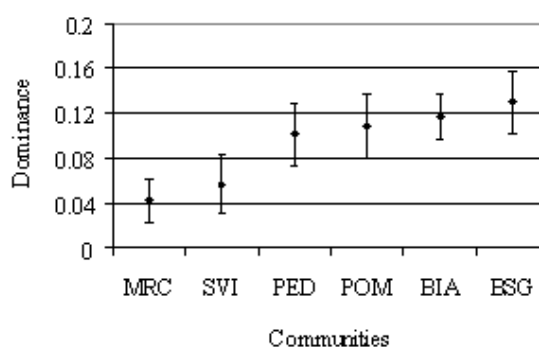


Figure 5. Simpson's $1/l$ index of dominance and the respective 95% confidence interval for the forests studied (see caption to Fig. 1, for abbreviations).

mal ignoring the singletons.

Geometric series

The disturbed and one of the swampy forests did not fit the geometric series model, indicating that the empirical relationships between this model and stress, disturbance and low species richness were not confirmed. Watkins and Wilson (1994) and Wilson et al. (1996) found similar results considering variation in environmental adversity and species richness. The same was found by Wilson et al. (1998) in relation to disturbance. As in the broken stick model, the poor fit to the geometric series comes from the underestimate of the singleton species. The ratio expected/observed singletons was of 0.5 for the unique forest that fitted the model (BSG), and below 0.38 for the others forests. The geometric series also underestimated the number of very abundant species. This is due to the relatively large number of species in the communities studied here. If S is large, then the result of the expression $[k/(1-k)][(1-k)^S] / [1 - (1-k)^S]$ is very small when compared to n_{\min}/N . In this case, the value of k should be low in order to approximate the value of the two expressions (see Materials and methods). Since k is the proportion of the remaining abundance taken by each species, the difference between the abundances of the species in the theoretical rank generated by the model is small when S is large; in other words, the dominance is small. Our results show that, for the species-poor communities, $k = 0.14$ (BIA, 33 species) and $k = 0.11$ (BSG, 36 species), but the theoretical distributions generated smaller dominance than observed: for BIA $l_{adjusted} = 0.076$ and $l_{observed} = 0.117$; for BSG $l_{adjusted} = 0.059$ and $l_{observed} = 0.130$. Our results suggest, therefore, that in communities as species-rich as those we studied here, the geometric series, as the broken stick model, does not imply high dominance. Moreover, the gradient of equitability described by May (1975) holds only when $k \approx 0.4$, a value that is typical of species-poor communities (May 1975, Magurran 1988).

Our results show that further studies are needed in order to validate (or reject) the abundance distribution models as predictors of species richness, stress and disturbance. However, other community descriptors gave better answers to stress and disturbance.

Lambda (λ), alpha (α) and species richness

Both λ and α can be used as diversity indices and presented similar results, showing a diversity gradient similar to the observed for the species richness estimated by rarefaction. Well-conserved communities had the largest diversity and species richness, swampy forests the smallest, and disturbed forests had intermediate positions. This

coincidence of results is expected, since λ and α are quite sensitive to the species number (Magurran 1988). The confidence intervals of $S_{(380)}$ and α for POM, BSG and BIA did not overlap those for MRC and SVI, thus showing that these two groups of communities had quite different species richness and diversity.

Evenness (E) and dominance (l)

Unlike richness, λ and α , the evenness did not show any relationship with disturbance or stress. The only communities in which the evenness confidence intervals did not overlap were BIA and BSG, both swampy forests, in which the overlap was expected. However, the dominance showed an inverse variation in relation to species richness, λ and α , being larger in the swampy and disturbed communities. The confidence intervals of l for BSG and BIA did not overlap those for MRC and SVI, thus showing that these two groups of communities had quite different dominances. According to Bulla (1994), the E index of evenness gives the same weight for rare or very abundant species, but Simpson's l index of dominance is determined predominantly by the most abundant species. By comparing our results for E and l , it can be seen that disturbance or stress alters the community abundance distribution mainly on the tail of the distribution corresponding to the most abundant species. As the number of the most abundant species was small and the amplitude of the most abundant classes was very large, the increase in frequency of these classes in the disturbed and stressed communities was also small. This, together with the difference in species number between the communities, may be the cause of the insensitivity of the log-normal and logseries models to stress and disturbance.

If we accept that the variation of λ and α is related to variations in species richness, we may consider that the community attributes influenced by stress and disturbance were species richness and dominance. According to Grime (1983), high levels of stress and disturbance decrease species richness, since few species can survive under these conditions. In the swampy forests, tree survival depends on metabolic or morphologic adaptations to anoxic conditions imposed by flooding and water table up-rising (Joly and Crawford 1982), developed by few species (Joly 1991). Metzger et al. (1998) found that Shannon's index was significantly smaller in swampy or frequently flooded forests than in the seasonal forests that are very rarely flooded. Hence, we expected that the diversity estimated by richness-sensitive indices, such as Shannon's (Magurran 1988), were smaller in the swampy forests. The larger dominance we found in swampy or disturbed communities may have no ecological meaning, be-

cause it is expected that species-poor communities have smaller equitability for purely statistical reasons (Cotgreave and Harvey 1994). However, Simpson's I index is not much influenced by species richness (Magurran 1988), and high dominance is expected in communities under stress or disturbance (May 1975). Communities under stress or disturbance may be regulated by only one or a few ecological factors, and species with different competitive ability in relation to the prevailing factors could divide the niche space in a quite hierarchical manner (May 1975). On the other hand, if several factors play a similar role in the determination of the abundance of species with different competitive abilities, in relation to different factors, the dominance would be smaller. If so, in communities under stress or disturbance, a smaller number of species would compete in a more hierarchical way than in communities where these factors are less intense. However, given the small number of communities under comparison, this explanation should be regarded carefully.

Comparing a larger number of communities could provide a more precise evaluation of the usefulness of abundance models for detecting of alteration of the community abundance distribution caused by stress or disturbance. Nevertheless, our results agree with those of Wilson et al. (1996), who compared 84 communities, and concluded that the adjustment to abundance models is not the best procedure for detecting patterns related to species richness, stress or disturbance in natural communities.

Acknowledgements: We thank to the Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq for the grants for the first author; to John Du Vall Hay, Miguel Petreire Junior and Kikyo Yamamoto for the critical reading of the manuscript; to Steven M. Holland for providing a software for the calculation of the variance of $S_{(m)}$ (available from www.uga.edu/~strata/software).

References

- Basset, Y., V. Novotny, S.E. Miller and N.D. Springate. 1998. Assessing the impact of forest disturbance on tropical invertebrates: some comments. *J. Appl. Ecol.* 35: 461-466.
- Bulla, L. 1994. An index of evenness and its associated diversity measure. *Oikos* 70: 167-171.
- Catharino, E.L.M. 1989. *Estudos fisionômico, florístico e fitossociológico em matas residuais secundárias do município de Piracicaba, SP*. M.Sc. Thesis. Campinas State University, Campinas, Brazil.
- Cielo Filho, R. 2001. *Estrutura de abundância de um trecho da Floresta Estacional Semidecídua no município de Campinas, estado de São Paulo: Mata Ribeirão Cachoeira*. M.Sc. Thesis. Campinas State University, Campinas, Brazil.
- Cohen, A.C. Jr. 1959. Simplified estimators for the normal distribution when samples are singly censored or truncated. *Technometrics* 1: 217-237.
- Cohen, A.C. Jr. 1961. Tables of maximum likelihood estimates: singly truncated and singly censored samples. *Technometrics* 3: 535-541.
- Cohen, J.E. 1968. Alternative derivations of a species abundance relation. *Am. Nat.* 102: 165-172.
- Conover, W.J. 1999. *Practical Nonparametric Statistics*. 3rd ed. John Wiley, New York.
- Cotgreave, P. and P. Harvey. 1994. Evenness of abundance in bird communities. *J. Anim. Ecol.* 63: 365-374.
- Fisher, R.A., A.S. Corbet and C.B. Williams. 1943. The relation between the number of species and the number of individuals in a random sample of an animal population. *J. Anim. Ecol.* 12: 42-58.
- Gray, J.S. 1987. Species-abundance patterns. In: J.H.R. Gee and P.S. Giller (eds.), *Organization of Communities: Past and Present*. Blackwell Science, Oxford. pp. 53-67.
- Greig-Smith, P. 1983. *Quantitative Plant Ecology*. 3rd ed. Blackwell, Oxford.
- Grime, J.P. 1983. *Plant Strategies and Vegetation Processes*. Wiley, New York.
- Heck, K.L.J., G. Van Belle and D. Simberloff. 1975. Explicit calculation of the rarefaction diversity measurement and the determination of sufficient sample size. *Ecology* 56: 1459-1461.
- Hill, J.K. and K.C. Hamer. 1998. Using species abundance models as indicators of habitat disturbance in tropical forests. *J. Appl. Ecol.* 35: 458-460.
- Hill, M.O. 1973. Diversity and evenness: a unifying notation and its consequences. *Ecology* 54: 427-432.
- Hubbell, S.P. and R.B. Foster. 1983. Diversity of canopy trees in a neotropical forest and implications for conservation. In: S.L. Sutton, T.C. Whitmore and A.C. Chadwick (eds.), *Tropical Rain Forest: Ecology and Management*. Blackwell Scientific Publications, Oxford. pp. 25-41.
- Hughes, R.G. 1986. Theories and models of species abundance. *Am. Nat.* 128: 879-899.
- Hurlbert, S.H. 1971. The nonconcept of species diversity: a critique and alternative parameters. *Ecology* 52: 577-586.
- Joly, C.A. 1986. Heterogeneidade ambiental e diversidade de estratégias adaptativas de espécies arbóreas de mata de galeria. *Anais do X simpósio da Academia de Ciências do Estado de São Paulo* 1: 19-38.
- Joly, C.A. 1991. Flooding tolerance in tropical trees. In: M.B. Jacksons, D.D. Daves and H. Lambers (eds.), *Plant Life under Oxygen Deprivation*. SPB Academic Publishing, The Hague. pp. 23-34.
- Joly, C.A. and R.M.M. Crawford. 1982. Variation in tolerance and metabolic responses to flooding in some tropical trees. *J. Exp. Bot.* 33: 799-809.
- Leigh, E.G.Jr. 1999. *Tropical Forest Ecology: A View from Barro Colorado Island*. Oxford University Press, New York.
- MacArthur, R.H. 1957. On the relative abundance of bird species. *Proceedings of the National Academy of Science* 43: 293-295.
- MacArthur, R.H. 1960. On the relative abundance of species. *Am. Nat.* 94: 25-36.
- Magurran, A.E. 1988. *Ecological Diversity and its Measurement*. Princeton University Press, Princeton.
- Martins, F.R. and F.A.M. Santos. 1999. Técnicas usuais de estimativa da biodiversidade. *Revista Holos*, edição especial: 236-267.
- May, R.M. 1975. Patterns of species abundance and diversity. In: M.L. Cody and J.M. Diamond (eds.), *Ecology and Evolution of*

- Communities*. Belknap Press of the Harvard University, Cambridge. pp. 81-120.
- Metzger, J.P., R. Goldenberg and L.C. Bernacci. 1998. Diversidade e estrutura de fragmentos de mata de várzea e de mata mesófila semidecídua submontana do Rio Jacaré-Pepira (SP). *Revista Brasileira de Botânica* 21: 321-330.
- Motomura, I. 1932. A statistical treatment of associations. *Jap. J. Zool.* 44: 379-383.
- Nummelin, M. 1998. Log-normal distribution of species abundances is not a universal indicator of rain forest disturbance. *J. Appl. Ecol.* 35: 454-457.
- Pielou, E.C. 1975. *Ecological Diversity*. Wiley, New York.
- Preston, F.W. 1948. The commonness and rarity of species. *Ecology* 29: 254-283.
- Scudeller, V.V. and F.R. Martins. 2002. FITOGEO – Um sistema de banco de dados aplicado à fitogeografia. *Acta Amazonica* (in press).
- Simpson, E.H. 1949. Measurement of diversity. *Nature* 163: 688.
- Sokal, R.R. and F.J. Rohlf. 1995. *Biometry: the Principles and Practice of Statistics in Biological Research*. 3rd ed. W.H. Freeman and Company, New York.
- Sugihara, G. 1980. Minimal community structure: an explanation of species abundance patterns. *Am. Nat.* 116: 770-787.
- Tabanez, A.J.A., V.M. Viana and A.D.S. Dias. 1997. Consequências da fragmentação e do efeito de borda sobre a estrutura, diversidade e sustentabilidade de um fragmento de floresta de planalto de Piracicaba, SP. *Revista Brasileira de Biologia* 57: 47-60.
- Toniato, M.T.Z., H.F. Leitão Filho and R.R. Rodrigues. 1998. Fitosociologia de um remanescente de floresta higrófila (mata de brejo) em Campinas, SP. *Revista Brasileira de Botânica* 21: 197-210.
- Torres, R.B., L.A.F. Matthes and R.R. Rodrigues. 1994. Florística e estrutura do componente arbóreo de mata de brejo em Campinas, SP. *Revista Brasileira de Botânica* 17: 189-194.
- Ugland, K.I. and J.S. Gray. 1982. Lognormal distributions and the concept of community equilibrium. *Oikos* 39: 171-178.
- Watkins, A.J. and J.B. Wilson. 1994. Plant community structure, and its relation to the vertical complexity of communities: dominance/diversity and spatial rank consistency. *Oikos* 70: 91-98.
- Whittaker, R.H. 1965. Dominance and diversity in land plant communities. *Science* 147: 250-260.
- Whittaker, R.H. 1972. Evolution and measurement of species diversity. *Taxon* 21: 213-251.
- Wilson, J.B. 1991. Methods of fitting dominance/diversity curves. *J. Veg. Sci.* 2: 35-46.
- Wilson, J.B. and H. Gitay. 1995. Community structure and assembly rules in a dune slack: variance in richness, guild proportionality, biomass constancy and dominance/diversity relations. *Vegetatio* 116: 93-106.
- Wilson, J.B., H. Gitay, J.B. Steel and W.McG. King. 1998. Relative abundance distributions in plant communities: effects of species richness and of spatial scale. *J. Veg. Sci.* 9: 213-220.
- Wilson, J.B., T.C.E. Wells, I.C. Trueman, G. Jones, M.D. Atkinson, M.J. Crawley, M.E. Dodd and J. Silvertown. 1996. Are there assembly rules for plant species abundance? An investigation in relation to soil resources and successional trends. *J. Ecol.* 84: 527-538.
- Zar, J.H. 1999. *Biostatistical Analysis*. 4th ed. Prentice Hall, New Jersey.

Appendix 1. Generating Cohen's table

Let X be a random variable whose distribution is normal with mean μ and variance σ^2 truncated on the left with truncation point x_0 (this means that observation is possible only if $x \geq x_0$).

Let ϕ and F denote, respectively, the density function and the cumulative distribution function of the normal standard distribution, that is:

$$\phi(t) = \frac{1}{\sqrt{2\pi}} \cdot \exp(-x^2/2)$$

and

$$F(t) = \int_{-\infty}^t \phi(s) ds.$$

The likelihood function of a sample $\{x_1, x_2, \dots, x_n\}$ of X is

$$L(x_1, x_2, \dots, x_n | \mu, \sigma, \eta) = (1-F(\eta))^n (\sigma\sqrt{2\pi})^{-n} \exp \left(-\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2} \right),$$

where

$$\eta = (x_0 - \mu) / \sigma.$$

Remark 1. $F(\eta)$ is the mass of the distribution being truncated; η may be estimated together with μ and σ (in terms of \bar{x} , x_0 and s) as the simultaneous solutions of the following likelihood equations:

$$x_0 - \mu = \sigma \eta$$

$$\bar{x} - \mu = \sigma Z$$

$$s^2 + (\bar{x} - \mu)^2 = \sigma^2 (1 + \eta Z)$$

where $Z(\eta) = \phi(\eta) / (1 - F(\eta))$ and \bar{x} and s^2 are the sample mean and variance respectively, that is:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad \text{and} \quad s^2 = \frac{\sum_{i=1}^n (x_i - \mu)^2}{n}$$

Remark 2. We assume in the following that $\xi < 0$ or equivalently $x_0 < \mu$. This means that the truncated mass is smaller than 1/2.

Remark 3. Observe that $\phi'(\eta) > 0$ (because $\eta < 0$), then:

$$Z'(\eta) = [(\phi(\eta))^2 + \phi'(\eta)(1 - F(\eta))] / (1 - F(\eta))^2$$

is strictly positive; this implies that the mapping

$$\eta \rightarrow Z(\eta)$$

is strictly increasing and then one to one.

Remark 4. Following Cohen (1959) the likelihood equations may be reformulated as follows:

$$\sigma^2 = s^2 + \theta(\bar{x} - x_0)^2$$

$$\mu = \bar{x} - \theta(\bar{x} - x_0)$$

$$[1 - Z(Z - \eta)] / (Z - \eta)^2 = s^2 / (\bar{x} - x_0)^2$$

where $\theta(\eta) = Z(\eta) / (Z(\eta) - \eta)$.

We have $Z(\eta) > 0$ (by definition) and $Z(\eta) > \eta$ (because $\eta < 0$); hence $\theta > 0$.

Remark 5. We have $Z'(\eta) > 0$ because $\eta < 0$ (see Remark 3); then:

$$\theta'(\eta) = [Z(\eta) - (Z'(\eta)\eta)] / (Z(\eta) - \eta)^2 > 0;$$

this in turn implies that the mapping

$$\eta \rightarrow \theta(\eta)$$

is one to one and then θ is determined by η and vice-versa.

Cohen (1959) gives a table for the estimated values of θ as a function of $s^2 / (\bar{x} - x_0)^2$ which depends only on sample values.

Hence we may calculate the estimators of σ and μ after substitution of the estimated value of θ in the two first equations of the second version of the likelihood equations (see Remark 4).

Remark 6. The computation of the estimators is based in the numerical solution of the equation

$$[1 - Z(Z - \eta)] / (Z - \eta)^2 = s^2 / (\bar{x} - x_0)^2.$$

Since Cohen's work was done in 1959, it would be interesting to recalculate Cohen's tables in light of the present numerical calculation techniques.

Appendix 2. An upper bound of the variance of an estimator of the measure of diversity λ

Following Simpson (1949): "Consider an infinite population such that each individual belongs to one of Z groups, and let

$$\pi_1, \pi_2, \dots, \pi_Z \quad \left(\sum_{j=1}^Z \pi_j = 1 \right)$$

be the proportion of the individual in the various groups. Then, λ defined as

$$\sum_{j=1}^Z \pi_j$$

is a measure of the concentration of the classification. It can take any value between $1/n$ and 1, the former representing the smallest concentration or largest diversity possible with Z groups, and the latter complete concentration, all the individuals being in a single group".

"Now suppose a sample of N individuals to be chosen at random of a population of this kind, and let

$$n_1, n_2, \dots, n_Z \quad \left(\sum_{j=1}^Z n_j \right)$$

be the number of individuals falling into the various groups. It is easily shown that:

$$l = (N(N-1))^{-1} \cdot \sum_{j=1}^Z n_j(n_j-1)$$

is an unbiased estimator of λ ".

If repeated samples of size N are drawn for the same population, the values of l obtained will be distributed about λ with variance:

$$\frac{4N(N-1)(N-2) \sum_{j=1}^Z \pi_j^3 + 2N(N-1) \sum_{j=1}^Z \pi_j^2 - 2N(N-1)(2N-3) \left(\sum_{j=1}^Z \pi_j^2 \right)^2}{[N(N-1)]^2}$$

It is easy to see that:

$$\begin{aligned} \text{Variance}(l) &= \frac{4(N-2) \sum_{j=1}^Z \pi_j^3}{N(N-1)} + \frac{2 \sum_{j=1}^Z \pi_j^2}{N(N-1)} - \frac{2(2N-3) \left(\sum_{j=1}^Z \pi_j^2 \right)^2}{N(N-1)} = \\ &= \frac{4(N-2) \sum_{j=1}^Z \pi_j^3}{N(N-1)} + \frac{2 \sum_{j=1}^Z \pi_j^2}{N(N-1)} - \frac{2[2(N-1)-1] \left(\sum_{j=1}^Z \pi_j^2 \right)^2}{N(N-1)} \leq \\ &\leq \frac{4 \sum_{j=1}^Z \pi_j^3}{N} + \frac{2 \sum_{j=1}^Z \pi_j^2}{N(N-1)} - \frac{4 \left(\sum_{j=1}^Z \pi_j^2 \right)^2}{N} + \frac{2 \left(\sum_{j=1}^Z \pi_j^2 \right)^2}{N(N-1)} = \\ &= \frac{4 \sum_{j=1}^Z \pi_j^3}{N} - \frac{4 \left(\sum_{j=1}^Z \pi_j^2 \right)^2}{N} + \frac{2 \sum_{j=1}^Z \pi_j^2}{N(N-1)} + \frac{2 \left(\sum_{j=1}^Z \pi_j^2 \right)^2}{N(N-1)} \leq \\ &\text{(because } \lambda = \sum_{j=1}^Z \pi_j^2 \leq 1) \end{aligned}$$

$$\leq \frac{4 \sum_{j=1}^Z \pi_j^3}{N} - \frac{4 \left(\sum_{j=1}^Z \pi_j^2 \right)^2}{N} + \frac{4}{N(N-1)} = \frac{4}{N} \left(\sum_{j=1}^Z \pi_j^3 - \left(\sum_{j=1}^Z \pi_j^2 \right)^2 \right) + \frac{4}{N(N-1)};$$

then we have proved that:

$$\text{Variance}(l) \leq \frac{4}{N} \left(\sum_{j=1}^Z \pi_j^3 - \left(\sum_{j=1}^Z \pi_j^2 \right)^2 \right) + \frac{4}{N(N-1)}.$$

Now, observing that $0 \leq \pi_j \leq 1$ implies $\pi_j^3 \leq \pi_j^2$ and then that

$$\sum_{j=1}^Z \pi_j^3 \leq \sum_{j=1}^Z \pi_j^2$$

we have

$$\text{Variance}(l) \leq \frac{4}{N} \left(\sum_{j=1}^Z \pi_j^2 - \left(\sum_{j=1}^Z \pi_j^2 \right)^2 \right) + \frac{4}{N(N-1)} = \frac{4}{N} \lambda(1-\lambda) + \frac{4}{N(N-1)}.$$

Observing that the quadratic function:

$$f(\lambda) = \lambda(1-\lambda)$$

reaches its maximum value at $\lambda=1/2$ and then that:

$$f(\lambda) \leq f(1/2) = 1/4 \text{ for all } \lambda,$$

we finally have

$$\text{Variance}(I) \leq 1 / N + 4 / [N(N-1)].$$

A more realistic approach

Let P be defined by

$$P = \sum_{j=1}^Z \pi_j^3 - \left(\sum_{j=1}^Z \pi_j^2 \right)^2$$

In the preceding section we proved that if $0 \leq \pi_j \leq 1$ for $1 \leq j \leq Z$ and $\sum \pi_j = 1$, then:

$$P \leq 1/4$$

and as a consequence of this inequality we obtained that

$$\text{Variance}(I) \leq 1 / N + 4 / [N(N-1)].$$

In this line of argument we considered all (mathematical) possibilities, but it is possible that most of them cannot occur in any natural (biological) situation. Then, instead of using $1/4$ as a lower bound of P , we can use the maximum value of P we know. Let us denote this maximum value by P_{MAX} .

We proved in the preceding section that:

$$\text{Variance}(I) \leq \frac{4}{N} \left(\sum_{j=1}^Z \pi_j^3 - \left(\sum_{j=1}^Z \pi_j^2 \right)^2 \right) + \frac{4}{N(N-1)}$$

then (in any situation we know) we have:

$$\text{Variance}(I) \leq 4 / N \cdot P_{MAX} + 4 / [N(N-1)].$$