

Statistical Properties and Performance of Ecological Indices Based On Relative Abundances

Gebriel M. Shamia

Abstract—The Improved Generalized Diversity Index (IGDI) has been proposed as a tool that can be used to identify areas that have high conservation value and measure the ecological condition of an area. *IGDI* is based on the species relative abundances. This paper is concerned with particular attention is given to comparisons involving the MacArthur model of species abundances. The properties and performance of various species indices were assessed. Both *IGDI* and species richness increased with sampling area according to a power function. *IGDI* were also found to be acceptable ecological indicators of conditions and consistently outperformed coefficient of conservatism indices.

Keywords—Statistical ecology, MacArthur model, Functional Diversity.

I. INTRODUCTION

MANY indices have been proposed for measuring diversity of large communities. Among those which incorporate the heterogeneity (dissimilarity) of species abundances, essentially two indices have become widely accepted, Shannon [20] index and Simpson [22] index. Arguments for the use of other forms, especially indices of Hill's [4] family have appeared. Good [3] proposed a general class of diversity index defined for non-negative integer values which included as special cases of both Shannon's index and Simpson's index. Ecological components can also be defined for any of these indices.

Suppose a population consists of s species with π_i being the abundance of species i , $i=1,2,\dots,s$, where the total number of individuals within a community is N , and N_i is the number of individuals belong to a species i , such that $0 \leq \pi_i \leq 1$; $\sum \pi_i = 1$.

In practice, when the abundance data are available, the relative abundance can be estimated by using the maximum likelihood estimator $\hat{\pi}_i = p_i = n_i/n$.

$$H(\alpha, \beta) = \sum_{i=1}^s \pi_i^\alpha (-\ln \pi_i)^\beta; \quad (\alpha, \beta) \in \Re \quad (1)$$

Within this framework, Shamia [19] further generalize Good's diversity index satisfying the two properties due to Pielou [16], (p.7) and allowing to taking values in the real plane \Re , namely:

P1: for fixed s , the index increases as the relative abundances become more equal.

P2: for equal relative abundance, the index is an increasing function of s and should be maximum.

Clearly not all (α, β) are suitable choices of diversity. For example $H(1,0) \equiv 1$ and so is not a useful index, $H(0,0) \equiv s$ measures species richness but does not give information about species abundances (π_i) consequently, they do not satisfy both properties (**P1**) and (**P2**) together. For more details see Shamia [19] and Baczkowski *et al.* [1]-[2].

So, it is sufficient to display the essential features of *IGDI* for real arbitrary α and β for which satisfies the above two requirement properties of a diversity index and determining an acceptable regions of $N^*(\alpha, \beta)$, which based on the third key desirable property of a diversity index proposed by Routledge [18], (p.505) is

P3: the diversity index should equal s in the equiprobable case. In addition, it is a continuous function.

Numerous measures of biodiversity are in use, satisfying different advantageous mathematical properties. Thus, in this paper we deal with transformation of $H(\alpha, \beta)$ and verifying whether such transformation satisfies the requirement properties descriptive above.

We will do that, by comparing the performance of underline index in two different cases which are equiprobable model, and broken-stick model.

This paper is organized as follows. In section II, we will give the theoretical developments of the Improved Generalized Diversity Index (IGDI). Section III discussed how *IGDI* behaves as a diversity measure with respect to species richness and evenness components. Section IV and Section V represent additional properties and performance of such index.

II. THEORETICAL DEVELOPMENTS

An improved generalized diversity index, *IGDI*, $N^*(\alpha, \beta)$, is expressed mathematically as follows.

Gebriel M. Shamia is with the Department of Statistics, Faculty of Science, University of Benghazi, Libya. Tel: +218918733011, e-mail: gmshamia@yahoo.com.

$$N^*(\alpha, \beta) = \begin{cases} \frac{H(\alpha, \beta)}{(\ln s)^\beta} & ; \alpha \neq 1 \\ \text{Exp}\{H(1, \beta)\}^{1/\beta} & ; \alpha = 1 \end{cases} \quad (2)$$

High index scores indicate both high species richness and more equal distribution of individuals among species. In fact, such transformations express the data in terms of number of species and thus are more easily interpreted. The value of a traditional heterogeneity index with equal probabilities can be considered a richness measure. This is because these indices are monotone increasing functions of the number of species s .

To examine the behaviour of $N^*(\alpha, \beta)$, we suppose $H(\alpha, \beta)$ to be a continuous function of s . If $\beta \neq 0$, then $H_{Eq}(\alpha, \beta)$ has a turning point at $s = e^{\beta/(1-\alpha)}$ for equiprobable case. It follows that $H_{Eq}(\alpha, \beta)$, satisfies (P2), being a monotonic increasing function of s for all $s \geq 1$, provided $\alpha \leq 1$ and $\beta \geq 0$, but excluding the case where (α, β) , equals $(0, 0)$ and $(1, 0)$. If $\alpha \geq 1$ and $\beta \leq 0$, but excluding the case where $(\alpha, \beta) = (1, 0)$, then $H_{Eq}(\alpha, \beta)$ a monotonic decreasing function of s for all $s \geq 1$, so that suitable "inverse" of $H_{Eq}(\alpha, \beta)$ would satisfy (P2).

The transformation $N^*(\alpha, 0)$ gives the class of diversity index $N_\alpha = \left[\sum \pi_i^\alpha \right]^{1/(1-\alpha)}$ for some positive integer values of $\alpha > 0$ as Hill [3] index.

MacArthur [10] denoted $N_1 = e^{H'}$ where H' is equivalent to $H(1, 1)$ generalized diversity index. Peet [15] recommends $N^*(1, 1) \equiv N_1$ as the best type-I heterogeneity measure for communities with rare species because the units (number of species) are more clearly understandable to ecologists. Routledge [18] suggested that any index of $N^*(1, 1)$ form measuring the diversity of large communities should be a continuous function of the abundances π_i .

In this paper, we used the transformation of $H(\alpha, \beta)$ with base (e) logs, since it is more common and more convenient, see Magurran [11]. In fact, $N^*(\alpha, \beta)$ gives the same combined acceptable region as for $H(\alpha, \beta)$, particularly for the equiprobable and broken-stick models due to MacArthur [9].

The central role of the exponential quantity gives the measure a privileged place as a measure of complex and diversity in all of the sciences. Since a suitable transformation of $N^*(\alpha, \beta \leq 0)$, such as inverse or others, would not be required to satisfy the properties. It does not need to be borrowed from information theory but arise naturally from this formalism of number equivalents.

The improved index $N^*(\alpha, \beta)$ satisfies the properties P1, P2 and P3 for a finite region $R_1 \subset \{0 < \alpha \leq 1, \beta > 0\}$ and infinite region $R_2 \subset \{\alpha \geq 1, \beta \leq 0\}$. Relaxing the monotonicity requirement from holding for all $s \geq 2$. Region (R_1) is the closed region bounded by the lines $\alpha = 1$ along x-axis and $\beta > 0$ along y-axis and an upper boundary linking the curve line $(\alpha, \beta) = (0, 0)$ and

$(\alpha, \beta) = (1, 1.693)$. This upper boundary of region (R_1) is given by $\beta = 4\alpha(1 - \alpha)$ for $\alpha \leq 0.3267$ and $\beta = 0.1534 + 0.6931\alpha + 0.5\sqrt{0.09416 + 2.7726\alpha}$ for $\alpha > 0.3267$.

In practice, most values of s encountered would lie between 2 and, typically, 200 species. The regions R_1 , for rare species and R_2 , for common species could be extended if we relaxed the restriction that the inequality holds for all $s \geq 2$, see Fig. 1.

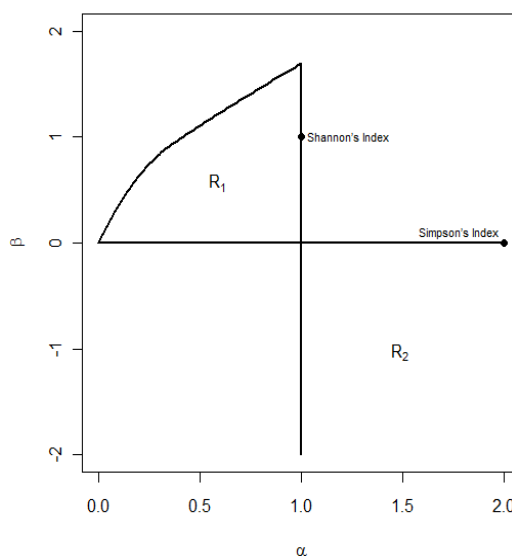


Fig. 1 Acceptable Regions (R_1) and (R_2) for $N^*(\alpha, \beta)$

III. INTERPRETATION OF THE IMPROVED DIVERSITY MEASURE

Furthermore, an interpretation of diversity measures with respect to species evenness (equitability) and richness components could be discussed for different values of (α, β) according to the above optimal acceptable regions R_1 and R_2 .

For this reason, the ratio between the $N^*(\alpha, \beta)$ values for equiprobable model and broken-stick model may be constructed. To assess the extent to which $N^*(\alpha, \beta)$ is sensitive to evenness define the ratio,

$$J = \frac{N_{Br}^*(\alpha, \beta)}{N_{Eq}^*(\alpha, \beta)} \quad (3)$$

where J is an equitability measure and for fixed s . See Sheldon [21].

Sensitivity to evenness is indicated by the J-ratio. If $J \approx 1$ for some (α, β) and given s , then $N^*(\alpha, \beta)$ is not a good discriminator between the two models. Accordingly, small values of J in acceptable regions indicate optimal choices of (α, β) . In practice, of course, the estimated $\hat{N}^*(\alpha, \beta)$ would be evaluated from the data.

Fig. 2 and Fig. 3 show contours of J are plotted for the cases $s=10$ and $s=100$ respectively.

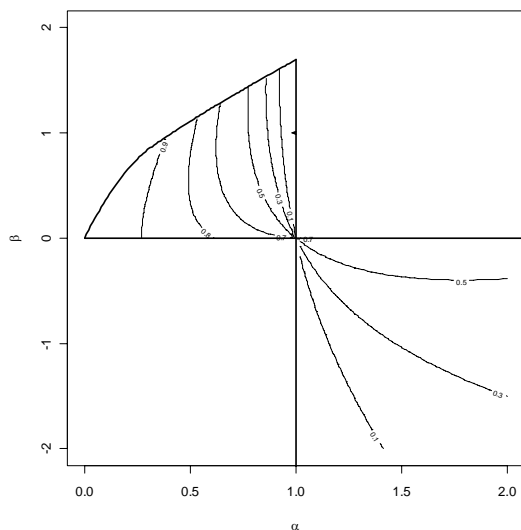


Fig. 2 J -ratio between $N^*(\alpha, \beta)$ for broken- stick model and equiprobable model for $s=10$

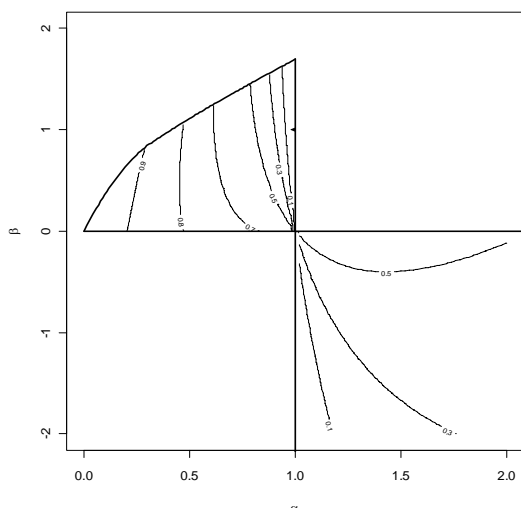


Fig. 3 J -ratio between $N^*(\alpha, \beta)$ for broken stick model and equiprobable model for $s=100$

It can be seen from these figures that the region with $J < 1$ extends beyond the acceptable regions of Fig. 1 and the optimal regions are

- $0 < \alpha \leq 1$ and β small positive in the rare species region (R_1).
- $\alpha > 1$ and β is small negative in the common species region (R_2).

As $\alpha \rightarrow 1$ the index $N^*(\alpha, \beta)$ is continuous along $\beta \rightarrow 0$. If the behaviour of $N^*(\alpha, \beta)$ is considered as a function of β , then it is found that the modified index has an asymptote at $\alpha=1$ for

$\beta \neq 0$. This is not a problem because the acceptable regions are not adjoining except along the line $\beta=0$ for $\alpha > 0$.

In general, the optimal values of α and β are affected by the change of equitability values as s increases. In fact, J decreases as s increases within acceptable region for fixed values of α and β . This is illustrated in Table I which gives J for Shannon's and Simpson's indices for various values of s .

TABLE I
THE J -VALUES FOR SHANNON'S AND SIMPSON'S INDICES

$J(\alpha, \beta)$	Species number			
	$s=10$	$s=25$	$s=50$	$s=100$
Shannon	0.717	0.689	0.670	0.663
Simpson	0.586	0.549	0.524	0.513

The effect of two different values of s for a broken-stick model on the improved generalized diversity index is considered. Define R_{Bs} as the ratio of $N_{Bs}^*(\alpha, \beta)$ with different numbers of species, as examples.

$$R_{Bs} = \frac{N_{Bs}^*(\alpha, \beta; s=50)}{N_{Bs}^*(\alpha, \beta; s=10)} \quad (4)$$

Good choices of (α, β) would give large values of R_{Bs} within the acceptable regions, since these would differentiate best between the two models for $s=50$ and $s=10$, see Fig. 4.

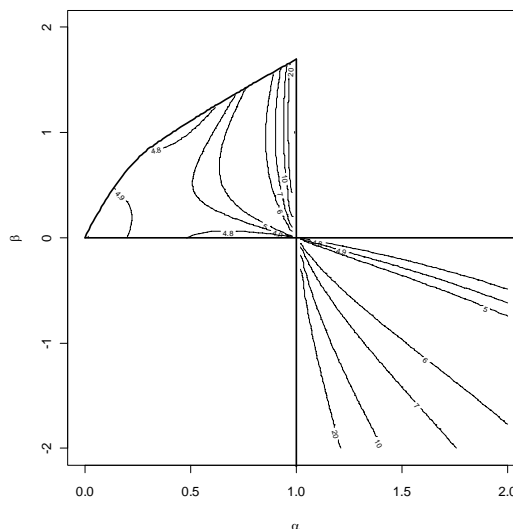


Fig. 4 R -ratio between $N^*(\alpha, \beta)$ for broken-stick model for 50 and 10 species within the acceptable regions

Within the acceptable regions large values of R_{Bs} result if $\alpha \approx 1$, $\beta > 0$ and $\alpha \approx 1$, $\beta < 0$ in the acceptable regions. The richness and evenness components have been discussed separately above. In practice these criteria will interact. In region (R_1) the two criteria are optimized by different values of (α, β) . The preferred choice of (α, β) will depend upon which concept is to be regarded as most important in a particular application.

Such behaviour of J -ratio and R -ratio could be obtained by $IGID$ contours with respect to the species number s .

IV. FURTHER PROPERTIES OF $N^*(\alpha, \beta)$

In this section, we will examine whether an improved generalized diversity index is satisfying another desirable properties due to Marshall and Olkin [12] and Rutledge [18].

A. Concavity of $N^*(\alpha, \beta)$

It is straightforward to prove that $N^*(\alpha, \beta)$ satisfies the monotony "concavity" property (**P4**) for (α, β) in the same acceptable region $(R_1 \cup R_2)$, while the reciprocal power $1/(1-\alpha)$ means there is no need for an "inverting" transformation for $N^*(\alpha, \beta)$. It follows that $H(\alpha, \beta)/[\ln(s)]^\beta$ for given s increases as the π_i 's become more equal. Since $\alpha < 1$ and $\beta > 0$ in R_1 , so $[1/(1-\alpha)] > 0$, it follows that $N^*(\alpha, \beta)$ increases as π_i 's become more equal, thus satisfying (**P1**). Similarly, in region R_2 , for given s , the diversity $N^*(\alpha, \beta)$ is Schur-concave function, thus satisfying (**P1**). This is true for the case $\alpha = 1$; $N^*(\alpha, \beta)$ satisfies (**P1**).

Fig. 5 shows plot of $1/(1-\alpha)$ against α . It is clear that $1/(1-\alpha)$ increases as for $0 < \alpha < 1$ and $\alpha > 1$.

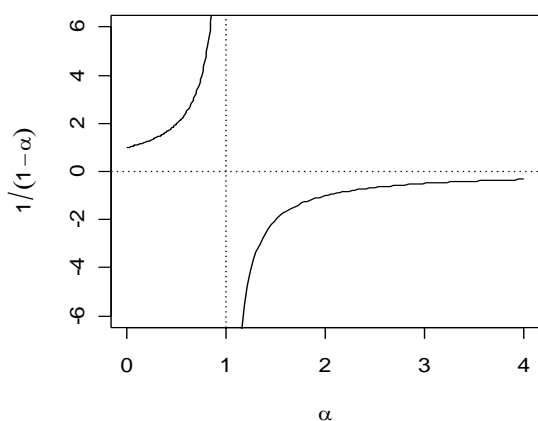


Fig. 5 Shows plot of $1/(1-\alpha)$ against α

Again, consider (**P2**): for the case $\pi_i = 1/s$ for all $i = 1, 2, \dots, s$, the index is an increasing function of s and should be maximum.

Marshall and Olkin [12] specify that the positive power for any concave function results in a concave function too. According to this result and since $N^*(\alpha, \beta)$ is a function power of $H(\alpha, \beta)$. Then $N^*(\alpha, \beta)$ is a Schur-concave function for $\pi = (\pi_1, \pi_2, \dots, \pi_s)' \in \mathfrak{R}$, satisfying (**P4**).

Rutledge [18] suggested a criterion for good diversity measures that any index measuring the diversity of large communities ought to be a continuous function of the proportional "relative" abundance of the species. He observed

that if an index of diversity is to satisfy such property, it must be of Hill's indices. It is convenient to name this property (**P5**).

P5: diversity index is a continuous function in each π_i , ($i = 1, 2, \dots, s$) for $0 \leq \pi_i \leq 1$.

Mathematically from definition of concave function, it is known that a continuous function $f(\cdot)$ with domain and counter domain the real values is called concave (convex) or concave-down (concave-up). This concavity depends on the sign of $f''(x)$ for all $x \in (a, b)$.

As a result of this, it makes sense to consider $N^*(\alpha, \beta)$ is a continuous function since $N^*(\alpha, \beta)$ is a generalized diversity index of Hill's family of indices.

Howard [5], (p.104) defined continuity as "a function f " is said to be continuous on a closed interval $[a, b]$ if the following conditions are satisfied:

1. f is continuous on (a, b) .
2. f is continuous from the right at a .
3. f is continuous from the left at b .

Conversely, William [24] proved that, polynomial and exponential functions are continuous if it is identified. Here, for $\alpha \neq 1$ the index is identified if $[\ln(s)] \neq 0$. It means that at $s > 1$ and in case $\alpha = 1$ the index is identified if $s \neq 0$. Since s is taking values between (2-200) this implies that, the index is continuous function in each π_i for $\pi_i \in [0, 1]$.

V. FUNCTIONAL DIVERSITY

In this context, Patial and Taillie [14] suggested two criteria. The first one refers to increasing of the diversity measure of a community, when a new species is introduced and also assent that the diversity increases each step. The second criterion refers to increasing of the diversity index as evenness increases, when the diversity index has measured by its rarity. The following figure represents the behaviour of $IGDI$ for different values of (α, β) .

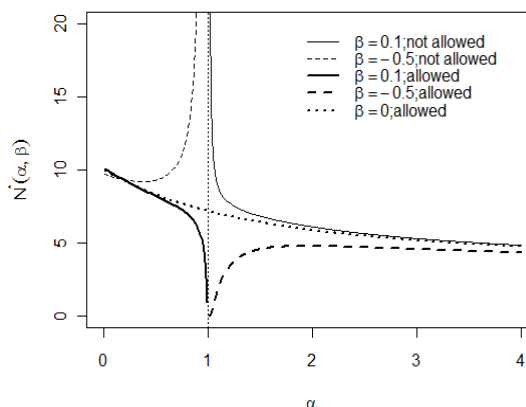


Fig. 6 Shows plots of $N^*(\alpha, \beta)$ against α for $(\beta = 0.1, -0.5, 0.0)$ in broken-stick case with $s = 10$

From Fig. 6 it can be seen that $N^*(\alpha, \beta)$ decreases for fixed $\beta=0.1$ as α increases from zero to one ($0 \rightarrow 1$); $\alpha \in (0.0257, 1)$. For fixed $\beta < 0$ ($\beta = -0.5$) and $\alpha > 1$, the diversity increases up to $\alpha=2$ and then slowly decreases; $N^*(\alpha, \beta)$ reaches a maximum equal to (4.763) before decreasing.

In general, as $\alpha \rightarrow \infty$, it appears that $N^*(\alpha, \beta)$ tends to a constant for all β . For the case $\beta=0$, $N^*(\alpha, \beta)$ decreases allowing $\alpha > 1$. The other cases, when $(\alpha < 1, \beta < 0)$ and $(\alpha > 1, \beta > 0)$ are not allowed since are not in acceptable regions (R_1 , R_2). For (α, β) in the combined acceptable region, $N^*(\alpha, \beta) \rightarrow 0$

as $\alpha \rightarrow 1$. For this to hold it is necessary that, for $(\alpha, \beta) \in R_1; \beta > 0$ and as $\alpha \rightarrow -1, [H(1, \beta)/(\ln(s))^\beta] < 1$. Similarly, for $(\alpha, \beta) \in R_2; \beta < 0$ and $\alpha \rightarrow +1, [H(1, \beta)/(\ln(s))^\beta] > 1$.

The case $\alpha \rightarrow 0$ corresponds to

$$N^*(\alpha, \beta) \rightarrow \sum_{i=1}^s \{ [-\ln(\pi_i)]^\beta / [\ln(s)]^\beta \}. \quad (5)$$

The seemingly nearly common limit for $N^*(\alpha, \beta)$ as $\alpha \rightarrow 0$ is simply a consequence of the choice of β used here. In practice $N^*(\alpha, \beta)$ can take a range of values depending upon β . See Fig. 7, for the broken-stick model with $s = 10$, which the values of $N^*(0, \beta)$ plotted against β .

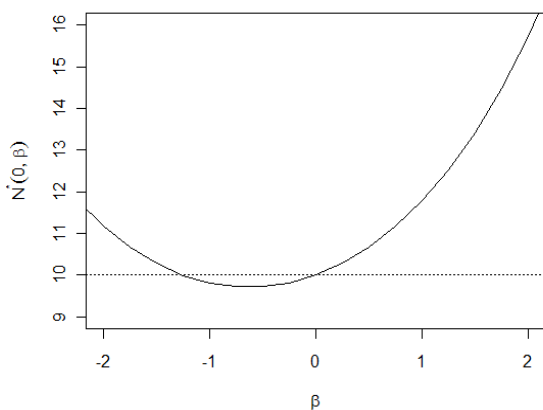


Fig. 7 Shows plots of $N^*(0, \beta)$ against β in broken-stick model for $s = 10$

Now, as $\alpha \rightarrow \infty$, it appears that $N^*(\alpha, \beta)$ tends to a constant for all β . This intuitively obvious because $N^*(\alpha, \beta)$ will be dominated by the α term whatever the value of β .

Recall Fig. 5, it is clear that $1/(1-\alpha)$ increases as for $0 < \alpha < 1$ and for $\alpha > 1$. Note that the case $\alpha=0$ is not in the combined acceptable region (R_1).

When $\alpha=1$, it is necessary for us to show that $N^*(1, \beta) < 1$ for all $0 < \beta < 1.693$, since in the acceptable region (R_1), β is maximized by $\beta = 1 + \ln(2) = 1.693$ at $\alpha=1$. In practice $s \geq 2$.

This suggests that, for $\beta \leq 1.693$ at $\alpha=1$, $N^*(\alpha, \beta)$ is Schur-concave function for all uneven relative abundances. These results yield certain desirable monotonicity properties.

VI. SUMMARY DISCUSSION

It is worth pointing out that $N^*(\alpha, \beta)$ is the universal measure of diversity in biology, physics, chemistry, and in medical. It is the measure of diversity that weights all species proportionality to their frequencies in the sample, rather than favoring common or rare species as others do. This alone is reason enough to select it as the best general-purpose diversity measure. Thus, this form great variety of diversity indices is all united into a single simple formula. This formula has most the mathematical properties expected of a true diversity. When it is applied to s equally-common species, it gives s as richness index.

Since a suitable transformation of $H(\alpha, \beta)$, such as inverse or others, would not be required to satisfy essential properties due to Pielou [16]. The range of validity for an optimal choices of which are allowing the *IGDI* index to satisfy such above properties are determined with broken-stick model; to generate none equally abundance of species.

The sensitivity of richness and evenness have been checked with more valid values of at changeable species relative abundance. This methodology full useful to explain the behaviour of the index under study to the change in the abundance of species in community, if the above procedures could be compared with other model of species abundances like the sequential breakage model due to Sugihara [23] which is related to the canonical lognormal model of Preston [17].

Several authors have proposed diversity ordering to compare communities through the concept of majorization and Schur-concavity properties of a function, see Marshall and Olkin [12], Patil and Taillie [13]-[14], and Lamshead et al. [8]. Shamia [19] proved that $H(\alpha, \beta)$ is Schur-concave function in region R_1 and is Schur-convex function in region R_2 . This result yield certain desirable monotonicity properties, which was applied to investigate the combined acceptable regions obtained.

In summary, high index scores indicate both high species richness and more equally distribution of individuals among species. The transformation expressed, we have introduced, in terms of number of species is more easily interpreted. Specially for the equitability measure $J(\alpha, \beta)$, since it a converted by dividing $N^*(\alpha, \beta)$ over the number of species (s). Here the optimal values of (α, β) within the regions (R_1, R_2) move along $\alpha \approx 1$ for $\beta > 0$ and $\alpha > 1$ for $\beta \leq 0$ as s increases.

Keylock [7] mentioned that the order (α) in the index $N^*(\alpha, 0)$ determines a diversity measure's sensitivity to rare or common species, orders higher than one are disproportional sensitive to the most common species, which orders lower than one are sensitivity to rare species. The critical point that weighs all species by their frequency, without favoring their common or rare species, occurs when $\alpha=1$. Jost [6] showed

that $N^*(\alpha, 0)$ gives the number of equivalents of all standard diversity indices.

We would hope that someday biologist can all agree that the word diversity should properly be applied only to quantities like $N^*(\alpha, \beta)$ which have the mathematical properties we intuitively expect of diversity.

REFERENCES

- [1] A. J. Baczkowski, D. N. Joanes, and G. M. Shamia, "Properties of a generalized diversity index," *Journal of Theoretical Biology*, 188, 1997, pp. 207-213.
- [2] A. J. Baczkowski, D. N. Joanes, and G. M. Shamia, "Range of validity of α and β for a generalized diversity index $H(\alpha, \beta)$ due Good," *Mathematical Bioscience*, 148:2, 1998, pp. 115-128.
- [3] I. J. Good, "The population frequencies of species and the estimation of population parameters," *Biometrika*, 40, 1953, pp. 237-264.
- [4] M. O. Hill, "Diversity and evenness: a unifying notation and its consequences," *Ecology*, 54, 1973, pp. 427-32.
- [5] A. Howard, *Calculus with analytic Geometr*, John Wiley and Sons. Inc., 1995.
- [6] L. Jost, "Entropy and diversity," *Oikos*, 113, 2006, pp. 363-375.
- [7] C. J. Keylock, "Simpson diversity and the Shannon-Wiener index as special cases of a generalized entropy," *Oikos*, 109:1, 2005, pp. 203-207.
- [8] P. J. D. Lamshead, H. M. Platt, and H. M. Shaw, "The detection of differences among assemblages of marine benthic species based on an assessment of dominance and diversity," *Journal of Natural History*, 17, 1983, pp. 859-74.
- [9] R. H. MacArthur, "On the relative abundance of bird species. proceedings of the national academy of science," *USA*, 43, 1957, pp. 293-5."
- [10] R. H. MacArthur, "Patterns of species diversity," *Biological reviews of the cambridge philosophical society*, 40, 1965, pp. 510-533.
- [11] A. E. Magurran, *Ecological Diversity and its Measurement*. Croom Helm, London, 1988.
- [12] A. W. Marshall, and I. Olkin, *Inequalities: Theory of moralization and its applications*, Academic Press, 1979.
- [13] C. P. Patil, and C. Taillie, *An Overview of Diversity: In ecological diversity in theory and practice*, edited by J.F. Grassle, G.P. Patil, W.K. Smith and C. Taillie,. Fairland, Maryland: International Co-operative Publishing House, 1979, pp. 3-27.
- [14] G. P. Patil, and C. Taillie, "Diversity as a concept and its measurement," *Journal of the American Statistical Association*, 77, 1982, pp. 548-67.
- [15] R. K. Peet, "The measurement of species diversity," *Annual Review of Ecology and Systematics*, 5, 1974, pp. 285-307.
- [16] E. C. Pielou, *Ecological Diversity*, Wiley, London, 1975.
- [17] F. W. Preston, "The anonical distribution of commonness and rarity," *Ecology*, 43, 1962, pp. 185-215.
- [18] R. D. Routledge, "Diversity index: which ones are admissible?" *Journal of Theoretical Biology*, 76, 1979, pp. 503-15
- [19] G. M. Shamia, Statistical Properties of the Generalized Diversity Index. PhD, *Dept. of Statistics, University of Leeds*, 1997.
- [20] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, 27, 1948, pp. 379-423.
- [21] A. L. Sheldon, "Equitability indices: dependence on the species count," *Ecology*, 50, 1969, pp. 476-77.
- [22] E. H. Simpson, "Measurement of diversity," *Nature*, 163, 1949, pp. 688.
- [23] G. Sugihara, "Minimal community structure: an explanation of species abundances pattern," *American Nature*, 116, 1980, pp. 770-787.
- [24] H. William, *Durfee Calculus and Analytic Geometry*. Mc Grow Hill, Inc, 1971.