Athens Institute for Education and Research
ATINER

ATINER's Conference Paper Series
ENV2012-0106

The Robustness of Ecological Measure Generalizing the Hill Index and its Application to Bacterial Data

Gebriel M. Shamia
Associate Professor
Department of Statistics, Faculty of Science,
University of Benghazi

Athens Institute for Education and Research
8 Valaoritou Street, Kolonaki, 10671 Athens, Greece
Tel: + 30 210 3634210 Fax: + 30 210 3634209
Email: info@atiner.gr URL: www.atiner.gr
URL Conference Papers Series: www.atiner.gr/papers.htm

Printed in Athens, Greece by the Athens Institute for Education and Research.

ISSN **2241-2891**
6/09/2012

An Introduction to ATINER's Conference Paper Series

ATINER started to publish this conference papers series in 2012. It includes only the papers submitted for publication after they were presented at one of the conferences organized by our Institute every year. The papers published in the series have not been refereed and are published as they were submitted by the author. The series serves two purposes. First, we want to disseminate the information as fast as possible. Second, by doing so, the authors can receive comments useful to revise their papers before they are considered for publication in one of ATINER's books, following our standard procedures of a blind review.

Dr. Gregory T. Papanikos
President
Athens Institute for Education and Research

This paper should be cited as follows:

Shamia, G.M. (2012) "**The Robustness of Ecological Measure Generalizing the Hill Index and its Application to Bacterial Data**" Athens: ATINER'S Conference Paper Series, No: ENV2012-0106.

The Robustness of Ecological Measure Generalizing the Hill Index and its Application to Bacterial Data

Gebriel M. Shamia
Associate Professor
Department of Statistics, Faculty of Science, University of Benghazi

Abstract

In this paper, we defined ecological measure generalizing the Hill index. The Improved Generalized Diversity Index (*IGDI*) has been proposed as a tool that can be used to identify data including that it contains hierarchical components and measure the ecological condition of an area. It generalizes in a natural way the Hill's family to incorporate species relatedness.

The Case study using our method will be shown. The purposes of this investigation are to study the diversity within species of bacterial data using analysis was carried out for phylogenetic entropy populates all genera. Such a decomposition can be made precise if the total species diversity is decomposed into the genus diversity and the average species diversity within genera.

The results of the analysis showed that there are close results between the diversity for whole data and by decomposition property. Using this terminology, shows that, generalized index is one of the best approaches to quantify community diversity in ecology.

Keywords: Hill's Index; Phylogenetic entropy; Segregation property.

Contact Information of Corresponding author:

Introduction

Diversity indices or measures are among the best defined elements within descriptive statistical indices. The application of these indices is common in ecological analysis to summarize community structure. Several studies appeared concerning the general properties of some classes of descriptive statistical indices.

Numerous measures of biodiversity are in use, satisfying different advantageous mathematical properties. Routledge (1979) proved that, if the diversity index satisfies some sensible properties, it must be one of Hill's (1973) indices and he suggested that, of the various indices proposed for measuring species diversity, the admissible forms are those that belong to the Hill family of indices $\{N_\alpha ; \quad for\, \alpha > 0\}$.

Hill (1973) proposed a unifying formulation of diversity and provided with the generalized entropies had many desirable properties as diversity indices

$$H_\alpha = \left(\frac{1}{1-\alpha}\right)\log \sum_{i=1}^{s} \pi_i{}^{\alpha} .$$

H_α makes mathematical sense for $-\infty \le \alpha \le \infty$. In other hand, he defined another diversity index family of order α which is:

$$N_\alpha = \left(\sum_{i=1}^{s} \pi_i{}^{\alpha}\right)^{1/(1-\alpha)} .$$

Where N_α represents the effective number of species of H_α, that is the number of species that would give the same generalized entropy measure if they were all equal. Also this family includes two common indices of diversity; $N_1 = \exp(H')$, where H' is the Shannon’s (1948) index, and $N_2 = 1/\lambda$, where $\lambda = \sum_{i=1}^{s} \pi_i^2$ is Simpson's (1949) index.

Many indices have been proposed for measuring the diversity of large communities. Among those which incorporate the heterogeneity (dissimilarity) of species abundances. Ecological components can also be defined for any of these indices. Thus, in this paper we deal with transformation of $H(\alpha,\beta)$ and verifying whether such transformation satisfies the requirement properties.

Suppose a population consists of s species with π_i being the abundance of species i, i=1,2,…,s. Baczkowski et al. (1997, 1998) proposed a generalized diversity index $H(\alpha,\beta)$ satisfying

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

This index generalized an index due to Good (1953) for positive integer values of (α,β).

Suppose a population consists of s species with π_i being the abundance of species i, i=1,2,…,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 $0 \le \pi_i \le 1$ and $\sum_{i=1}^{s} \pi_i = 1$. In practice, when the abundance data are available, the relative abundance can be estimated by using the maximum likelihood estimator $\hat{\pi}_i = p_i = n_i / n$.

Within this framework, Shamia (1997,2008) proposed a further generalization of $H(\alpha,\beta)$.

The diversity index, IGDI, is expressed mathematically as follows.

$$N^*(\alpha,\beta) = \left\{ \frac{H(\alpha,\beta)}{(\ln s)^{\beta}} \right\}^{1/(1-\alpha)} \qquad ;\alpha \neq 1$$

$$N^*(1,\beta) = Exp\{H(1,\beta)\}^{1/\beta} \qquad ;\alpha = 1$$

This family of indices has satisfied additional properties such as consistency and monotonicity. For more details, see Al-Shareef (2010).

The centeral 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 $N^*(\alpha,\beta \le 0)$, 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 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 (1973) index.

In this paper, we have used the transformation of $H(\alpha,\beta)$ with base (e) logs, since it is more common and more convenient, see Magurran (1988). In fact, $N^*(\alpha,\beta)$ gives the same combined acceptable region as for $H(\alpha,\beta)$, particularly for the equiprobable model and broken-stick model due to MacArthur (1957). So, it is sufficient to display the essential features of $N^*(\alpha,\beta)$, IGDI, for real arbitrary α and β which satisfied the requirement properties of a diversity index and determined an acceptable regions: Rare Species Region $\mathbf{R_1}$ contained in the region $\alpha \in (0,1]$ and $\beta \ge 0$ and Common Species Region $\mathbf{R_2}$, given by $\alpha > 1$ and $\beta \le 0$.

So, in this article we are dealing with an Improved Generalized Diversity Index, IGDI, is satisfying the key properties due to Pielou (1975, p.7) and another desirable properties due to Rutledge (1979, p. 511).

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 is more easily interpreted. The value of a traditional heterogeneity index with equal probabilities can be considered as a richness measure. This is because these indices are monotone increasing functions of the number of species s.

Decomposition Methodology of $N^*(\alpha,\beta)$

Ecological diversity relates to the different forms of life which are present in a particular site; in a more precise since, it concern the different species of a particular genus which are present in an ecological community. The measures or indices of ecological diversity are statistical summaries of the abundance vector, that is, the frequencies or proportions of each species in the community. For more details, see Rao (1982).

As a concept, diversity relates both to the number of species (richness) and to their apportionment within the community. There is grate diversity when the number of species grows, and when all the species are fairly represented. According to the aims pursued with the employment of diversity indices, additional formal property may be recognized. In words: the overall species diversity is equal to the genus diversity multiplied by the weighted average of variety within each partition.

In order to clarify this property, suppose that the relative abundance of genus i is π_i for $\pi_i \in [0,1]$. Within genus i, let the relative abundance of species j be θ_{ij}. If ρ_{ij} represents the proportion abundance of this species relative to the whole community, then

$$\rho_{ij} = \theta_{ij} * \pi_i \quad \text{and} \quad \pi_i = \sum_k \rho_{ik} \,,$$

let $\pi = (\pi_1, \pi_2, ...)$; $\theta_i = (\theta_{i1}, \theta_{i2}, ...)$; $\rho = (\rho_{11}, \rho_{12}, ..., \rho_{21}, \rho_{22}, ...)$, then

$$N(\rho) = N(\pi) * \{\text{some average of the } N(\theta_i)\text{'s}\}.$$

That is the species diversity, in general, is equal to genus diversity multiplied by the average species diversity within genera. The average should satisfy the following sub-properties:

(a) Consistency: if all the genera were equal diverse, then their average diversity should be $N^*(\alpha,\beta)$.

(b) Strict monotonicity: if the diversity of any genus is increased, then the overall diversity should be increased.

(c) Homogeneity: if the diversity within each genus is increased by a factor "C", the overall diversity should be increased by the same factor.

(d) Symmetry: the average should be independent of the ordering of the genera.

Let M is some average that is (a) Consistency, (b) Strict monotonicity, (c) Homogeneity, and (d) Symmetry. Then

$$M[N(\theta_1),...,N(\theta_g)] = \begin{cases} \{[\sum_i^g \pi_i^\alpha (-\ln \pi_i)^\beta . N(\theta_i)^{(1-\alpha)}] \big/ [\sum_i^g \pi_i^\alpha (-\ln \pi_i)^\beta]\}^{1/(1-\alpha)} \,; & \alpha > 0, \ \alpha \neq 1. \\ \text{Exp}\,.\{\{\sum_i^g \pi_i (-\ln \pi_i)^\beta . [\ln N(\theta_i)]^\beta \big/ H(1,\beta)\}^{1/\beta}\} \ ; & \alpha = 1, \ \beta \neq 0. \end{cases}$$

This result has conformed the discussion due to Marshall and Olkin (1979) and Hardy et al. (1934). Work in progress indicates that although the estimation of any $N^*(\alpha,\beta)$ is necessarily tedious, it may slightly easier to obtain $N^*(\alpha,\beta)$ for $\alpha \neq 1$ than for $\alpha = 1$ from data as pointed out above. In what follows it is assumed for convenience that species quantities are expressed as numbers of individuals. The arguments would be unaltered if some other measure of quantity were used instead.

Case Study: An application of Bacterial data

In order to evaluate the performance on real data sets we apply the proposed procedure on a Bacterial data used by Dr. Salha F. Ben-Jwirif for M.Sc study in biology about the degree of bacterial contaminate and assess the effectiveness of disinfectant against isolated bacteria due to Al-Abeidy (2010).

This application investigates the property of decomposition of IGDI by using bacterial data from Al-Jumhuria hospital in Benghazi–Libya which represents 200 samples collected from 1.6.2009 to 31.12.2009.

In this section we will apply the decomposition property in which the species diversity of some region is equal to genus diversity multiplied by the average species diversity within genera. To investigate such property, we will use $N^*(\alpha,\beta)$ to calculate $N(\rho)$ according to the methodology in Section-2 and compare it with the diversity value obtained from the relative abundance of each species to the whole community data.

A valuable data represents species of bacterial department into non-pathogenic bacteria and pathogenic bacteria which is branching into g-groups. Here, we will apply this property for each type of bacteria, in addition for all types as a pooled set.

Table 1 explains species of bacteria with relative abundance data for each species as:

Genus: dented by $g=2$ and refers to bacterial branch according to ecological conditions.

Abundance of species: denoted by N_i and refers to individuals present of each species, when $N = \sum_i N_i$.

π_i : refers to the relative abundance of species for the same group, since $\pi_1 = \frac{4+14+...+31}{4+14+...+132} = 0.407$, then we have relative abundance vector: $\pi_i = (0.407, 0.593)$.

θ_{ij}: refers to the relative abundance of species jth of ith group since $\theta_{11} = \frac{4}{212} = 0.019.$

So, $\theta_{ij} = (0.019, 0.066, ..., 0.427).$

ρ_{ij} : refers to the relative abundance of each species to the whole community.

Since, $\rho_{11} = \frac{4}{521} = 0.008.$ So, $\rho_{ij} = (0.008, 0.027, ..., 0.253).$

Table 1. *Bacterial data obtained from Al-Jamahiriya hospital with relative abundance data* π_i *,* θ_{ij} *and* ρ_{ij} *.*

g Genus-	Bacteria Species of	n_i	π_i	θ_{ij}	ρ_{ij}
	Non-pathogenic bacteria:				
1	Bacillus circulan	4	*0.407*	0.019	0.008
	Bacillus pumilus				
	Bacillus sphaericus				
	CDC group-Eo-2	14		0.066	0.027
	Chromobacterium violaceum				
	Corynebaterium bovis				
	Dermacoccus luteus	2		0.009	0.004
	Dermacoccus nishinomiyaensis				
	Methylobacterium extorquens	11		0.052	0.021
	Micrococcus luteus				
	Moraxella catarrhalis				
	Staphylococcus capitis	150		0.708	0.288
	Staphylococcus epidermidis				
	Staphylococcus equorum				
	Staphylococcus haemolyticus				
	Staphylococcus homimis				
	Staphylococcus kloosii				
	Staphylococcus lentus				
	Staphylococcus saprophyticus				
	Staphylococcus sciuri				
	Kocuria rosea	31		0.146	0.060
	Pantoea agglomerans				
	Tatumella ptyseas				
	Pathogenic bacteria:				
2	Acinetobacter baumannii	30	*0.593*	0.097	0.058
	Acinetobacter haemoly				
	Acinetobacter faecalis				
	Enterobacter cloacae	27		0.087	0.052
	Enterobacter faecalis				
	Enterobacter faecium				
	Pseudomonas aeurginosa	96		0.311	0.184
	Pseudomonas fluoresens				
	Pseudomonas luteola				
	Pseudomonas species				
	Pseudomonas stutzeri				
	Staphylococcucs aureus	24		0.078	0.046
	Shigella flexneri				
	Bacillus ccereus	132		0.427	0.253
	E-coli				
	Klebsiella pneumonia spp				

We begin the calculation by prove some aim properties of the pooled data which is, for example,

i) $\pi_i = \sum_j \rho_{ij}$

At $i=1 \Rightarrow \sum_j \rho_{1j} = 0.008+0.027+0.004+0.021+0.288+0.059 = 0.407 = \pi_1.$

At $i=2 \Rightarrow \sum_j \rho_{2j} = 0.058+0.052+0.184+0.046+0.253 = 0.593 = \pi_2.$

ii) $\rho_{ij} = \theta_{ij}.\pi_i$

At $i=j=1 \Rightarrow \rho_{11} = \theta_{11}.\pi_1$, here $\theta_{11} = 0.019$, $\pi_1 = 0.407$.

$\therefore \theta_{11}.\pi_{11} = (0.019)(0.407) = 0.008 = \rho_{11}$.

Now, we will calculate the species diversity, $N(\rho)$, which leads to obtain the average of species within the genera, $N(\theta_i)$'s, and so obtain the genus diversity, $N(\pi)$, where the relative abundances are π_i's as two groups. We will do that for non-pathogenic bacteria, pathogenic bacteria, and the pooled set at different choice of (α,β): (0.5,0.25), (1.0,1.0), (1.0,0.5); for rare species region (**R$_1$**) and (1.25,-0.75), (2.0,0.0); for common species region (**R$_2$**). Then make comparison with the values of $N^*(\rho)$ to check the decomposition property of IGDI.

The following are the results, as examples, for different (α,β) in the rare species region (**R$_1$**)

a) At $\alpha = 0.5$ and $\beta = 0.25$

The index, for whole (pooled) community, is given by

$$H(0.5,0.25) = \{(0.088)(1.485)+(0.164)(1.379)+...+(0.503)(1.082)\} = 3.467.$$

$$\Rightarrow N^*(\rho) = \{3.467/2.398\}^2 = 7.763.$$

The average species diversity, for both non-pathogenic and pathogenic bacteria, is given by

$$N(\theta_1) = \{2.061/(\ln(6))^{0.25}\}^2 = 3.172.$$

$$N(\theta_2) = \{2.314/(\ln(5))^{0.25}\}^2 = 4.221.$$

$$M[N(\theta_1), N(\theta_2)] = \{2.451/1.276\}^2 = 3.692.$$

$$\Rightarrow N(\rho) = N(\pi).M[N(\theta_1), N(\theta_2)] = (1.955)(3.692) = 7.219.$$

b) At $\alpha = 1$ and $\beta = 1$

$$H(1,1) = \{(0.008)(4.869)+(0.027)(3.617)+...+(0.253)(1.373)\} = 1.883.$$

$$\Rightarrow N^*(\rho) = Exp.\{1.883\} = 6.572.$$

The average species diversity is given by

$$N(\theta_1) = \{0.978\} = 2.659.$$

$$N(\theta_2)=\{0.978\}\ =3.913.$$

$$M[N(\theta_1),N(\theta_2)]=Exp.\{(0.780)/(0.676\}=3.174.$$

$$\Rightarrow N(\rho)=(1.966)(3.174)=6.239.$$

The overall results are illustrated in the **Table 2**, gives $N^*(\rho)$ for the pooled data and $N(\rho)$ by decomposition at different choices of α and β within acceptable regions.

Table 2. Values of $N^*(\rho)$ and $N(\rho)$ of bacterial data at different choices of α and β.

Region	α	β	$N^*(\rho)$	$N(\rho)$
Rare Region (**R**$_1$)	*0.50*	*0.25*	*7.763*	*7.219*
	1.00	*1.00*	*6.572*	*6.239*
--------------------------	*1.00*	*0.50*	*6.081*	*5.788*
Common Region (**R**$_2$)	*1.25*	*-0.75*	*1.033*	*0.945*
	2.00	*0.00*	*5.156*	*5.156*

It can be seen from the above results that, the values of the index are feasible especially at (α=2.0,β=0.0) by Simpson index in the common species region (**R**$_2$), while they give small deviations at different (α,β) from (2.0,0.0). Generally, the results support that, the IGDI index satisfies the decomposition property within the combined acceptance regions. In fact, differences in species diversity reflect real differences in the distribution of rare and common species as types of bacteria.

Discussion

It is worth pointing out that $N^*(\alpha,\beta)$ as an ecological measure generalizing the Hill index is the universal measure of diversity in biology, physics, chemistry, and in medical. It is the only 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 the most mathematical properties expected of a true diversity.

This methodology completely useful to explain the behaviour of the index under study to the change in the abundance of species in community. A measure of diversity should also contain ecological components. These components satisfy the property of decomposition of $N^*(\alpha,\beta)$. A diversity index should be decomposable in two ways; in the case of α≠1 and in the case of α=1. $N^*(\alpha,\beta)$ in case of α≠1 and β=0, possesses a major advantage in measuring diversity within decomposition. As seen in **Table 2**.

In summary, the most important point to be made with regard to this application of bacterial data is that the concept of the community has no local validity. High index scores indicate both components of diversity, the evenness and the richness, 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 $\beta \leq 0$. In particular, the common region showed more close results.

The illustration of a spatial IGDI index was constrained to yield particular values for particular spatial arrangements for rare species region (**R$_1$**) and for common species region (**R$_2$**).

We would hope that someday biologists can reach agreement that the word diversity should properly be applied only to quantities like $N^*(\alpha,\beta)$ which have the mathematical properties. It is especially useful to convert to true diversities when dealing with multiple kinds of indices.

All indices can be transformed into true diversities, which possess a uniform set of mathematical properties that accurately capture the diversity concept. Further work could be done in order to extend the procedure to compare communities according to their $N^*(\alpha,\beta)$ profile. Converting raw indices to true diversities (Hill numbers) makes possible the construction of meaningful index-independent general equations, measures, or formulas involving diversity.

References

Al-Abeidy N. (2010). 'Study of bacterial contamination of computer keyboard and mouse used in some Benghazi hospitals.' MSc. Theses, Academy of Graduate Studies. Benghazi-Libya.

Al-Shareef, A.I. (2010). 'A case study in the concept of biodiversity measurement: statistical properties of an improved diversity index.' MSc. Theses, Academy of Graduate Studies. Benghazi-Libya.

Baczkowski, A.J., Joanes, D.N. & Shamia, G.M. (1997). 'Properties of a Generalized Diversity Index.' Journal of Theoretical Biology, **188**, 207-213.

Baczkowski, A.J., Joanes, D.N. & Shamia, G.M. (1998). 'Range of validity of α and β for a generalized diversity index H(α,β) due Good.' Mathematical Bioscience, **148(2)**, 115-128.

Good, I.J. (1953). 'The Population Frequencies of Species and the Estimation of Population Parameters.' Biometrika, **40**, 237-64.

Hardy, G.H., Littlewood, J. E. & Pólta, G. (1934). Inequalities. Cambridge, Cambridge University Press.

Hill, M.O. (1973). 'Diversity and Evenness: A Unifying Notation and Its Consequences.' Ecology, **54**, 427-32.

MacArthur, R.H. (1957). 'On The Relative Abundance Of Bird Species.' Proceedings of the National Academy of Science, USA, **43**, 293-5.

Magurran, A.E. (1988). Ecological Diversity and Its Measurement. Croom Helm, London.

Marshall, A.W. & Olkin, I. (1979). Inequalities: Theory of Moralization and its Applications. Academic Press.

Pielou, E.C. (1975). Ecological Diversity. Wiley, London.

Rao, C.R. (1982). 'Diversity: Its Measurement, Decomposition, Apportionment and Analysis.' Sankhya, **A44**, 1-22.

Routledge, R.D. (1979). 'Diversity Index: Which Ones Are Admissible?.' Journal of Theoretical Biology, **76**, 503-15.

Shamia, G.M. (1997). 'Statistical Properties of the Generalized Diversity Index.' PhD, Dept. of Statistics, University of Leeds.

Shamia, G.M. (2008). 'Properties of Modified Generalized Diversity Index.' Journal of 1st Mathematical Statistics Conference. Aleppo University, Faculty of Science, 241-251.

Shannon, C.E. (1948). 'A Mathematical Theory of Communication.' The Bell System Technical Journal, **27**, 379-423.

Simpson, E.H. (1949). ‘Measurement of Diversity.’ Nature, **163**, 688.