COENOSES 5 (1): 33-35, 1990

FISH SPECIES DIVERSITY AND ITS RELATIONSHIP TO DISTANCE FROM THE COAST LINE

R.K. Jain, Department of Mathematics and Statistics, Memorial University of Newfoundland, St. John's, Newfoundland, Canada A1C 5S7

Keywords: Fish, Diversity, Distance, Coast line

Abstract. This paper investigates the characteristics of a fish community and the effect of the distance from the coast line on the measure of fish species. Fluctuation in species diversity is noted as the distance from the coast line is increased. The effect of the timing of the day is examined.

Introduction

Fish community structure and its form are changing rapidly due to the use of modern technology in fishing industries. It is essential to monitor these changes continuously, using mathematical and statistical models. Specifically, measurement of fish species diversity plays an important role in describing the structure and behaviour of the fish community. The most common measures are due to Shannon and Simpson (1949). Hill (1973) discussed the use of Simpson and Shannon's index and argued that they both provide a good estimate for the species diversity and differ only with respect to common and rare species. The present study explores the characteristics of a fish community in Brazil and the effect of the distance from the coast line on the measure of species as well as the timing of the day.

Basic measures of diversity

The basic principle in the measurement of diversity is closely related to the theory of information introduced by Shannon (1949) to measure the information content of the code. Rényi (1961) discussed a general class of functions used in communication known as entropy of order α of an s-symbol code, of which a proportion p_i of the symbols are of the i^{th} kind. This is defined as follows:

$$H_\alpha = \frac{\log_e\left(\sum_{i=1}^{s} p_i^\alpha\right)}{1-\alpha} \qquad (1)$$

It can be shown that $\lim_{\alpha\to 1} H_\alpha = -\sum_{i=1}^{s} p_i \log_e p_i$. Let

$$H_1 = -\sum_{i=1}^{s} p_i \log_e p_i = H' \qquad (2)$$

Clearly, H' is identical with entropy of order 1 which was originally proposed by Shannon and Weaver (1949). Margalef (1958) discussed the use of H', which measures uncertainty, to measure diversity in ecology. Uncertainty increases as diversity increases (Pielou (1969)). If we denote $\alpha=2$, then

$$H_2 = -\log_e\left(\sum_{i=1}^{s} p_i^2\right) \qquad (3)$$

Denote $\sum_{i=1}^{s} p_i^2 = C$, thus

$$H_2 = -\log_e C \qquad (4)$$

If p_i ($i=1, 2, \ldots, s$) is the true proportion of the i^{th} species in the community, then p_i^2 is the probability that two independently selected individuals are from the same i^{th} species by chance. Hence, C is the probability that two independently selected individuals come from identical species. The function C measures 'concentration' or 'dominance' of a many-species community. Since p_i^2 is a probability measure and C lies between 0 and 1, it makes sense to consider the following quantity D as a measure of diversity:

$$D=1-C \qquad (5)$$

D is also the probability that two independently selected individuals belong to different species. D is called Simpson's index of diversity. H_2 is also a good candidate for an index of diversity. However, H_2 is not a popular measure in literature (Pielou (1975)).

Estimation procedures

Since ecological populations are, in general, indefinitely large, it is not possible to count them all and identify all the individuals in the community. However, it may be possible to take a sample from the population, and based on the sample, one may estimate the index of diversity. The maximum likelihood method of estimation of diversity is often used in the ecological literature. The maximum likelihood estimate of H' is given by

$$\hat{H}' = -\sum_{i=1}^{s} [(n_1/n) \log_e (n_i/n)] \quad (6)$$

where $n_1, n_2, ..., n_s$ are the frequencies of the species; n is the total number of frequencies and s is the number of species present in the sample. Any logarithmic base may be used to compute $\hat{H}'$. In the present paper, natural logarithms are used. The estimator $\hat{H}'$ underestimates H' by an approximate magnitude of (s/2n), where s is the true number of species in the community [Basharin (1959)]. Unfortunatily, no adjustment can be made unless one knows the true value of s, and in general, the true values of s cannot be obtained from a random sample taken. However, if one has prior information of an approximate magnitude of the number of species s of a given community, then taking a sufficiently large sample size in comparison of s would eliminate the bias term of (s/2n).

The variance of the maximum likelihood estimator, $\hat{H}'$, of the Shannon index is given by [Lloyd et al. (1968)]

$$\mathrm{Var}(\hat{H}') = [\, n \sum_{i=1}^{s} n_i (\log_e n_i)^2 - (\sum_{i=1}^{s} n_i \log_e n_i)^2] / (n^3) + (s-1)/(2n^2) \quad (7)$$

The second term in equation (7) is often ignored if it is negligible. If two samples are selected independently from one another, then Hutcheson (1970) showed that the statistic

$$t = \frac{\hat{H}'_1 - \hat{H}'_2}{[\mathrm{Var}(\hat{H}'_1) + \mathrm{Var}(\hat{H}'_2)]^{1/2}} \quad (8)$$

has a student t-distribution with degrees of freedom given by

$$\nu = \frac{[\mathrm{Var}(\hat{H}'_1) + \mathrm{Var}(\hat{H}'_2)]^2}{\frac{(\mathrm{Var}(\hat{H}'_1))^2}{n_1} + \frac{(\mathrm{Var}(\hat{H}'_2))^2}{n_2}} \quad (9)$$

where n_1 and n_2 are the number of individuals in the first and the second sample, respectively.

The unbiased estimator of Simpson's index of diversity is given by

$$\hat{D} = (n/(n-1)) \, [1 - \sum_{i=1}^{s} (n_i/n)^2] \quad (10)$$

The variance of $\hat{D}$ is given by Lyons and Hutcheson (1978) as follows:

$$\mathrm{Var}(\hat{D}) = (2/n)(n-1)^{-1} [\, 2(n-2) \sum_{i=1}^{s} (n_i/n)^3 + \sum_{i=1}^{s} (n_i/n)^2 - (2n-3) (\sum_{i=1}^{s} (n_i/n)^2)^2] \quad (11)$$

Lyons and Hutcheson (1978) proposed the following statistic to compare estimates of D from two independent random samples:

$$Z = \frac{\hat{D}_1 - \hat{D}_2}{[\mathrm{Var}(\hat{D}_1) + \mathrm{Var}(\hat{D}_2)]^{1/2}} \quad (12)$$

This Z has an approximate standard normal distribution, provided n_1 and n_2 are sufficiently large.

Pielou (1975) noted that Simpson's index is more sensitive to changes in the more common species, whereas Shannon's index is more sensitive to changes in the rarer species.

Table 1. $\hat{H}'$ and Var ($\hat{H}'$) with increasing distance from the coast and time of day.

Distance (km)	Number of species	Time of the day	$\hat{H}'$	Var ($\hat{H}'$)
1.3	15	Morning	0.72942	0.0000862
	16	Afternoon	0.85079	0.0000653
1.5	22	Morning	0.70791	0.0000292
	17	Afternoon	0.80937	0.0000228
1.9	16	Morning	0.53084	0.0000469
	18	Afternoon	0.93830	0.0001503
2.1	17	Morning	0.67136	0.0001171
	20	Afternoon	0.82407	0.0000498
2.3	24	Morning	0.66999	0.0001648
	23	Afternoon	0.76643	0.0000672
3.5	17	Morning	1.02113	0.0001892
	17	Afternoon	0.77518	0.0000317

Table 2. $\hat{D}$ **and Var** ($\hat{D}$) **with increasing distance from the coast and time of the day.**

Distance (km)	Number of species	Time of the day	$\hat{D}$'	Var ($\hat{D}$')
1.3	15	Morning	0.38401	0.0000209
	16	Afternoon	0.48184	0.0000092
1.5	22	Morning	0.40943	0.0000068
	17	Afternoon	0.50874	0.0000033
1.9	16	Morning	0.31631	0.0000200
	18	Afternoon	0.53130	0.0000174
2.1	17	Morning	0.33494	0.0000292
	20	Afternoon	0.48293	0.0000110
2.3	24	Morning	0.33189	0.0000374
	23	Afternoon	0.41433	0.0000153
3.5	17	Morning	0.56608	0.0000124
	17	Afternoon	0.49202	0.0000042

Study area

The data were collected at Almofala on the coast of Brazil to study the fish community structure during June 1980. The location, Almofala beach, is in Acarac county (Ceara, Brazil), latitude 02°50's, longitude 40°09'w. The samples were taken with increasing distance from the coast at six sites (1.3, 1.5, 1.9, 2.1, 2.3 and 3.5 km). The detailed descriptions of the sampling procedures are given in Rocha (1980). A total of 44 species with a total of 161173 individuals were collected. Total number of individuals (n), number of individuals per species (n_i) and number of species (s) present were used to calculate $\hat{H}$', Var ($\hat{H}$'), $\hat{D}$ and Var ($\hat{D}$) according to equations (6), (7), (10) and (11).

Results and discussion

The Tables 1 and 2 present Shannon and Simpson's indices of fish species diversity and its variance with increasing distance from the coast line and timing of the day, respectively.

The difference in the quantitative characteristic of the fish species diversity between morning and afternoon is highly significant based upon the t-statistic for Shannon's index and the Z-statistic for Simpson's index. Both diversity measures gave the same test results, probably owing to the very large sample sizes. Both $\hat{H}$' and $\hat{D}$ suggest that the change in diversity is statistically significant as the distance from the coast line is increased.

No attempt was made in the present paper either to confirm or disprove that the diversity of fish species decreases with increasing distance from the coast line (Beumer 1980). However, the pattern of increasing fish species diversity with distance from the source has also been reported by Harrell, Davis and Dorris (1967). The present study did not reveal any trend or correction of diversity with increasing distance. It appears that the diversity fluctuations with changing distance from the coast depend on the reproductive strategies of the species and the environmental conditions.

Acknowledgements. The author appreciates the referee and Professor L. Orlóci for comments and suggestions which has improved the presentation. The author wishes to thank Memorial University for granting him sabbatical leave for research work.

REFERENCES

BASHARIN, G.P. 1959. On a statistical estimate for the entropy of a sequence of independent random variables. Theory of Probability and its Application 4: 333-336.

BEUMER, J.P. 1980. Hydrology and fish diversity of a north Queensland tropical stream. Aust. J. Ecology 5: 159-186.

HARRELL, R.C., B.J. DAVIS and T.C. DARRIS. 1967. 1967. Stream order and species diversity of fishes in an intermittent stream. Am. Midl. Natur. 78: 428-436.

HILL, M.O. 1973. Diversity and evenness: A unifying notation and its consequences. Ecology 54: 427-432.

HUTCHESON, K. 1970. A test for comparing diversities based on the Shannon's formula. J. Theoret. Biol. 29: 151-154.

LLOYD, M., J.H. ZAR and J.R. KARR. 1968. On the calculation of information theoretical measures of diversity. Am. Midl. Natur. 79: 257-272.

LYONS, N.I. and K. HUCHESON. 1978. Comparing diversities: Gini's index. J. Statist. Comput. Simul. 8: 75-78.

MARGALEF, D.R. 1958. Information theory in ecology. Gen. Syst. 3: 36-71.

PIELOU, E.C. 1969. An Introduction to Mathematical Ecology. Wiley.

PIELOU, E.C. 1975. Ecological Diversity. Wiley.

RÉNYI, A. 1961. On measures of entropy and information. Proc. 4th Berkeley Symp. on Math. Stat. and Prob. 1: 547-561.

ROCHA, C.A.S. 1980. Statistical analysis and diversity with special reference to Brazilian fish. M. Sc. Thesis, Memorial University of Newfoundland.

SHANNON, C.E. and W. WEAVER. 1949. The Mathematical Theory of Communication. Univ. of Illinois Press, Urbana.

SIMPSON, E.H. 1949. Measurement of diversity. Nature 163: 688.

Manuscript received: December 1988