

INFERENCE ON ECOLOGICAL DIVERSITY

T. Di Battista

Dipartimento di Metodi Quantitativi e Teoria Economica, Università “G. d’Annunzio”, Italy

Keywords: simple random sampling, abundance vector estimation, diversity index estimation, jackknife, Bayesian diversity index estimation, species-abundance curve model, Poisson distribution, negative binomial distribution, logarithmic series distribution, geometric distribution, broken-stick distribution, lognormal series distribution.

Contents

1. Introduction
 2. Diversity index estimation
 - 2.1. General Results
 - 2.2. Inference for some specific Diversity Indexes under SRS
 - 2.3. Jackknife for Diversity Index Estimation
 - 2.4. Bayesian Approach to Diversity Index Estimation
 3. Species-abundance curve models
 - 3.1. Negative Binomial Distribution
 - 3.2. Logarithmic Series Distribution
 - 3.3. Geometric Distribution and Broken-Stick Distribution
 - 3.4. Lognormal series distribution
- Glossary
Bibliography
Biographical Sketch

Summary

Statistical methods to provide inference on ecological diversity are reviewed. The chapter focuses on the techniques adopted under simple random sampling (SRS) with replacement. Two different issues are developed: the inference on diversity indexes and the inference based on species-abundance curves. As to the first topic, some families of diversity indexes are considered and the properties of the corresponding estimators under SRS are analyzed. As to the second topic, some species-abundance curves belonging to the compound Poisson distribution family and the corresponding estimation of their parameters are considered.

1. Introduction

A complete ecological population is seldom available, since animal and plant censuses are impractical as well as very expensive. Accordingly, field scientists are obviously forced to adopt inferential methods in order to give some insight into the diversity of biological communities. This chapter surveys the basic tools suitable for making inference on ecological diversity when simple random sampling (SRS) with replacement is adopted for drawing the population units. SRS with replacement is used in practice

for its simplicity and many papers on quantitative ecological diversity focus on this approach.

Diversity indexes are fundamental measures adopted to assess the biological diversity of a population (see *e.g.* Pielou, 1977, p.300). These indexes are usually abundance vector functions which satisfy some desirable properties. When the population is not known, the diversity indexes are estimated on the basis of a sample and hence their sampling properties must be determined. Obviously, the diversity index estimators are functions of the observed abundance vector. It is worth noting that the observed abundance vector is a multinomial random vector if the SRS with replacement takes place. This is an appealing setting in which the properties of the diversity index estimators may be easily found, at least in a large-sample framework. Moreover, in this case bias reduction techniques – such as the jackknife – are straightforwardly adapted to diversity index estimation.

Ecological populations usually contain many species and the abundance of each species may vary greatly from the very common to the very rare. Accordingly, in order to analyze the diversity pattern of ecological populations, it may be convenient to consider the frequency of the species with a given number of units – *i.e.* the species-abundance curve – rather than the frequency of units for each species – *i.e.* the abundance vector. The species-abundance curve turns out to be a very useful tool in the study of ecological diversity (see *e.g.* Pielou, 1977, p.269, or Magurran, 1988). The evaluation of ecological diversity on the basis of the species abundance-curve is usually performed in a model-based approach. A parametric probability distribution model – containing a small number of parameters, say one, two or three – is assumed for the species-abundance curve. The model is reasonably chosen according to the characteristics of the population pattern over the study region. Moreover, the model parameters – or some of their functions – may be suitably adopted as diversity indexes. Hence, if a sample of biological units is obtained under SRS with replacement, the inference on diversity reduces to the estimation of the parameters of the proposed distribution model.

These two main approaches to diversity estimation are considered in the following sections. More precisely, Section 2 is devoted to the analysis of diversity index estimation, while the inference based on the species-abundance curves is reviewed in Section 3.

2. Diversity Index Estimation

First, let us suppose that the ecological population is made up of N units and partitioned into s species. Moreover, let N_l be the abundance of l -th species ($l = 1, 2, \dots, s$). Hence, $\mathbf{N} = (N_1, N_2, \dots, N_s)^T$ denotes the abundance vector, while $\mathbf{p} = (p_1, p_2, \dots, p_s)^T$ represents the relative abundance vector with $p_l = N_l/N$, ($l = 1, 2, \dots, s$).

In ecological surveys a number of measures have been proposed to quantify the diversity of the population. The most frequently used diversity indexes may be expressed as a function, say $\Delta(\mathbf{p})$, of the relative abundance vector \mathbf{p} . For example,

Patil and Taillie (1979) (see also Gove *et al.*, 1994) propose the family of average species rarity indexes, *i.e.*

$$\Delta(\mathbf{p}) = \sum_{l=1}^s p_l R_l(\mathbf{p}), \quad (1)$$

where $R_l(\mathbf{p})$ is a suitable function which represents the rarity measure of the l -th species ($l = 1, 2, \dots, s$). Some of the most frequently used diversity indexes – such as Shannon's and Simpson's indexes – are special cases of (1). For example, Shannon's well-known index H is obtained from (1) for $R_l(\mathbf{p}) = \ln p_l$. Good (1953) introduces another general family of diversity indexes, *i.e.*

$$\Delta(\mathbf{p}) = \sum_{l=1}^s p_l^\alpha (-\ln p_l)^\beta, \quad (2)$$

where α and β are positive parameters. It is worth nothing that Shannon's index H is obtained from (2) if $\alpha = \beta = 1$.

We consider the problem of estimating \mathbf{p} – and accordingly $\Delta(\mathbf{p})$ – on the basis of a sample of biological units under SRS with replacement (of size n), a common setup in ecological literature. In this setting, the units are independently selected with the same drawing probability, in such a way that at each drawing, the selected unit is returned to the population.

In Section 2.1 the general theory of diversity index estimation under SRS is considered, while in Section 2.2 some specific results are surveyed. Furthermore, Section 2.3 is devoted to the issue of bias reduction by means of jackknife techniques. Finally, Section 2.4 contains some details on the Bayesian approach to diversity index estimation under SRS.

2.1. General Results

If $\mathbf{n} = (n_1, n_2, \dots, n_s)^T$ represents the observed abundance vector, let $\hat{\mathbf{p}} = (\hat{p}_1, \hat{p}_2, \dots, \hat{p}_s)^T$ be the observed relative abundance vector, where $\hat{p}_l = n_l/n$ ($l = 1, 2, \dots, s$). Therefore, in the present SRS approach, it is at once apparent that \mathbf{n} is a multinomial random vector with parameters n and \mathbf{p} . Accordingly, since $\hat{\mathbf{p}} = n^{-1}\mathbf{n}$, then $\hat{\mathbf{p}}$ turns out to be an unbiased and consistent estimator for \mathbf{p} . It should be remarked that $\hat{\mathbf{p}}$ coincides with the maximum likelihood estimator of \mathbf{p} . The variance-covariance matrix of $\hat{\mathbf{p}}$ is given by $\text{Var}(\hat{\mathbf{p}}) = n^{-1}\mathbf{\Sigma}$, where $\mathbf{\Sigma} = \text{diag}(\mathbf{p}) - \mathbf{p}\mathbf{p}^T$. Moreover, it is at once apparent that a consistent estimator for $\mathbf{\Sigma}$ is given by $\hat{\mathbf{\Sigma}} = \text{diag}(\hat{\mathbf{p}}) - \hat{\mathbf{p}}\hat{\mathbf{p}}^T$. Finally, the use of the Central Limit Theorem straightforwardly provides that

$$n^{1/2}(\hat{\mathbf{p}} - \mathbf{p}) \xrightarrow{d} N_s(\mathbf{0}, \mathbf{\Sigma}), \quad (3)$$

as $n \rightarrow \infty$.

From the previous findings, an estimator for a diversity index may be easily obtained as $\hat{\Delta} = \Delta(\hat{\mathbf{p}})$. On the basis of (3), if

$$\phi_l(\mathbf{x}) = \frac{\partial}{\partial x_l} \Delta(\mathbf{x}), l = 1, 2, \dots, s,$$

are defined in a neighborhood of \mathbf{p} and are continuous and non-null at \mathbf{p} , by using the Delta method Tong (1983) proves that as $n \rightarrow \infty$

$$n^{1/2}(\hat{\Delta} - \Delta) \xrightarrow{d} N(0, \tau^2), \quad (4)$$

where

$$\tau^2 = \tau(\mathbf{p})^2 = \sum_{l=1}^s p_l \phi_l(\mathbf{p})^2 - \left\{ \sum_{l=1}^s p_l \phi_l(\mathbf{p}) \right\}^2.$$

A consistent estimator for τ^2 is given by $\hat{\tau}^2 = \tau(\hat{\mathbf{p}})^2$. Moreover, the joint use of the Sverdrup Theorem and the Slutski Theorem ensures that as $n \rightarrow \infty$

$$n^{1/2} \frac{\hat{\Delta} - \Delta}{\hat{\tau}} \xrightarrow{d} N(0, 1). \quad (5)$$

Hence, expression (5) provides the test statistic for the hypothesis $H_0: \Delta = \Delta_0$, as well as the pivotal quantity for determining a confidence interval for Δ .

2.2. Inference for some specific Diversity Indexes under SRS

First, owing to its widespread use, we consider some results for Shannon's index H . Basharin (1959) obtains (4) in this special case. Accordingly, if

$$\hat{H} = -\sum_{l=1}^s \hat{p}_l \ln \hat{p}_l \quad (6)$$

represents the Shannon's index estimator, then as $n \rightarrow \infty$

$$n^{1/2}(\hat{H} - H) \xrightarrow{d} N(0, \tau^2),$$

where

$$\tau^2 = \sum_{l=1}^s p_l \ln^2 p_l - H^2 .$$

It is worth noting that (6) coincides with the maximum likelihood estimator of H (see e.g. Pielou, 1977, p.301). However, as Pielou (1977, p.301) remarks, estimator (6) is negatively biased for finite samples, i.e. \hat{H} underestimates H – usual behavior for diversity index estimators. More basically, Blyth (1958) proves that there is no unbiased estimator of Shannon’s index. Finally, since a Taylor expansion of $E(\hat{H})$ at \mathbf{p} provides the following expression (Bowman *et al.*, 1971)

$$E(\hat{H}) = H - \frac{s-1}{2n} + O(n^{-2}),$$

it is worth noting that an estimator which removes the $O(n^{-1})$ bias term is given by

$$\tilde{H} = \hat{H} + \frac{s-1}{2n} .$$

Obviously, \tilde{H} displays the same large-sample properties as \hat{H} .

If the family (1) is considered with the dichotomous rarity measure $R_l(\mathbf{p}) = (1 - p_l)^\beta / \beta$ with $\beta = 1, 2, \dots, m$, then the corresponding diversity indexes are polynomials in p_l . A similar remark holds for family (2), if $\beta = 0$ and $\alpha = 1, 2, \dots, m$. In these cases, (1) and (2) may be conveniently expressed as

$$\Delta(\mathbf{p}) = \sum_{l=1}^s \sum_{j=0}^m a_j p_l^j . \tag{7}$$

where a_j ($j=1, 2, \dots, m$) are constant coefficients of the power series expansion (7). It is worth noting that Blyth (1958) and Good (1953) show that only diversity indexes which are polynomials in p_l (of degree $m < n$) have unbiased estimators. Accordingly, by applying the Rao-Blackwell theorem, Smith *et al.* (1979) find the minimum variance estimators for the diversity indexes belonging to (7) for which unbiased estimators exist, *i.e.*

$$\hat{\Delta} = \sum_{l=1}^s \sum_{j=0}^m a_j \binom{n_l}{j} \binom{n}{j}^{-1} . \tag{8}$$

Even if a closed form for the variance of (8) could be obtained, it is cumbersome to evaluate in practice. Hence, Simberloff (1979) suggests obtaining the variance estimator of (8) by using Monte Carlo methods.

Recently, Baczkowski et al. (2000) provided a methodology for deriving the moments of the estimator for a diversity index belonging to family (2). By adopting a Taylor series expansion, Baczkowski *et al.* (2000) obtain the first four moments of the estimator up to the order $O(n^{-3})$. Further results on the Good family estimator, such as large-sample normality and related issues, are contained in Baczkowski *et al.* (1998).

-
-
-

TO ACCESS ALL THE 16 PAGES OF THIS CHAPTER,
Visit: <http://www.eolss.net/Eolss-sampleAllChapter.aspx>

Bibliography

- Anscombe, F.J. (1950). Sampling theory of the negative binomial and logarithmic series distribution, *Biometrika* **37**, 358-382. [This introduces the lognormal series distribution].
- Adams, J.R., McCune, E.D. (1979). Application of the generalized jackknife to Shannon's Measure of information used as an index of diversity, in *Ecological Diversity in Theory and Practice*, J.F. Grassle, G.P. Patil, W. Smith and C. Taillie editors, Fairland (Maryland), International Co-operative Publishing House, 117-132. [This introduces the use of generalized jackknife for diversity index estimation].
- Baczkowski, A.J., Joanes, D.N. and Shamia, G.M. (1998). Range of validity of α and β for a generalized diversity index $H(\alpha, \beta)$ due to Good, *Mathematical Biosciences* **148**, 115-128. [The paper discusses the generalized Good's diversity index deriving its range]
- Baczkowski, A.J., Joanes, D.N. and Shamia, G.M. (2000). The distribution of a generalized diversity index due to Good, *Environmental and Ecological Statistics* **7**, 329-342. [This considers some issues about the estimation of the diversity indexes belonging to Good's family].
- Basharin, G.P. (1959). On a statistical estimate for the entropy of a sequence of independent random variables, *Theory of Probability and its Applications* **4**, 333-336. [This focuses on the estimation of Shannon's index].
- Birch, M.W. (1963). An algorithm for the logarithmic series distribution, *Biometrics* **19**, 651-652. [The paper develops an algorithm to estimate the logarithmic series distribution]
- Bliss, C.I. (1965). An analysis of some insect trap records, in *Classical and Contagious Discrete Distributions*, G.P. Patil editor, Oxford, Pergamon Press, 385-397. [This seminal paper proposes the use of the zero-truncated lognormal series distribution as a model for the observed abundance of the species].
- Blyth, C.R. (1958). Note on estimating information, *Annals of Mathematical Statistics* **30**, 71-79. [This considers some properties of Shannon's index estimator].
- Böhning, D. (1983). Maximum likelihood estimation of the logarithmic series distribution, *Statistische Hefte* **24**, 121-140. [The paper develops the Maximum likelihood method to estimate the logarithmic series distribution]

Bowman, K.O., Hutchenson, K., Odum, E.P. and Shenton, L.R. (1971). Comment on the distribution of indices of diversity, in *Statistical Ecology*, Vol. 3, G.P. Patil, E.C. Pielou and W.E. Waters editors, University Park (Pennsylvania), Pennsylvania State University Press, 315-319. [It contains some remarks on Shannon's index estimation].

Bowman, K.O. and Shenton, L.R. (1970). Properties of the maximum likelihood estimators for the parameter of the logarithmic series distribution, in *Random Counts in Models and Structures*, G.P. Patil editor, University Park (Pennsylvania), Pennsylvania State University Press, 127-150. [It contains estimation aspects of the parameter of the logarithmic series distribution]

Brass, W. (1958). Simplified methods of fitting the truncated negative binomial distribution, *Biometrika* **45**, 59-68. The paper deals with practical aspects to estimate the truncated negative binomial distribution]

Cacoullos, T. and Charalambides, C.A. (1975). On MVUE for truncated binomial and negative binomial distribution, *Annals of the Institute of Statistical Mathematics* **27**, 235-244. [In this paper a minimum variance unbiased estimator for truncated binomial and negative binomial distributions is provide]

Chatfield, C. (1969). On estimating the parameters of the logarithmic series and negative binomial distribution, *Biometrika* **56**, 411-414. [It contains some remarks on estimating the parameters of the logarithmic series and negative binomial distribution]

David, F.N. and Johnson, N.L. (1952). The truncated Poisson distribution, *Biometrics* **8**, 275-285. [This considers the estimation of the parameters of the zero-truncated negative binomial distribution].

Dennis, B. and Patil, G.P. (1988) Applications in ecology, in *Lognormal Distributions: Theory and Applications*, E.L. Crow and K. Shimizu editors, New York, Dekker, 303-330. [This contains a survey of the use of the lognormal series distribution].

Gill, C.A. and Joanes, D.N. (1979). Bayesian estimation of Shannon's index of diversity, *Biometrika* **66**, 81-85. [This introduces the Bayesian approach to diversity index estimation].

Good, I.J. (1953). The population frequencies of species and the estimations of population parameters, *Biometrika* **40**, 237-264. [This proposes a widely used family of diversity indexes].

Good, I.J. (1966). How to estimate probabilities, *Journal of Mathematics and Applications* **2**, 364-383. [The paper concerns about the estimation of the relative abundance vector]

Gove, J.H., Patil, G.P., Swindel, B.F. and Taillie, C. (1994). Ecological diversity and forest management, in *Handbook of Statistics*, Vol. 12 (Environmental Statistics), G.P. Patil and C.R. Rao editors, Amsterdam, Elsevier, 409-462. [This paper surveys the topics relative to the well-known family of average species rarity indexes].

Guttman, I. (1958). A note on a series solution of a problem in estimation, *Biometrika* **45**, 565-567. [This considers the properties of the maximum likelihood estimator of the parameter of the zero-truncated geometric distribution].

Hoaglin, D.C. and Tukey, J.W. (1985). Checking the shape of discrete distribution, in *Exploring Data Tables, Trends and Shapes*, D.C. Hoaglin, F. Mosteller and J.W. Tukey editors, New York, Wiley, 345-416. [This book presents major advances in exploratory data analysis]

Hutcheson, K. and Shenton, L.R. (1974). Some moments of an estimate of Shannon's measure of information, *Communication in Statistics* **3**, 89-94. [In the paper a procedure to derive the moments of the Shannon index is provided]

Kemp, A.W. and Kemp, C.D. (1988). A rapid and efficient estimation procedure for the negative binomial distribution, *Biometrical Journal* **29**, 865-873. [The paper deals with practical aspects to estimate the negative binomial distribution]

MacArthur, R.H. (1957). On the relative abundance of bird species, *Proceedings of the National Academy of Science of Washington* **43**, 293-295. [This introduces the well-known broken-stick model].

Magurran, A.E. (1988). *Ecological diversity and its measurement*, London, Croom Helm. [This is a book which deals with diversity models].

Palmer, M.W. (1991). Estimating species richness: the second-order jackknife reconsidered, *Ecology* **72**, 1512-1513. [This suggests an effective jackknife procedure for diversity index estimation].

Patil, G.P. (1962). Some method of estimation of the logarithmic series distribution, *Annals of the Institute of Statistical Mathematics* **18**, 68-75. [This considers the estimation of the parameters of the logarithmic series distribution].

Patil, G.P. and Taillie, C. (1979). An overview of diversity, in *Ecological Diversity in Theory and Practice*, J.F. Grassle, G.P. Patil, W. Smith and C. Taillie editors, Fairland (Maryland), *Statistical Ecology*, Vol. 3, G.P. Patil, E.C. Pielou and W.E. Waters editors, Fairland (Maryland), International Co-operative Publishing House, 3-27. [This introduces the family of average species rarity indexes].

Pielou, E.C. (1977). *Mathematical Ecology*, New York, Wiley. [A seminal text concerning the commonly-adopted statistical methods in ecological settings].

Preston, F.W. (1948). The commonness, and rarity, of species, *Ecology* **29**, 254-283. [A seminal paper concerning species diversity]

Sampford, M.R. (1955). The truncated negative binomial distribution, *Biometrika* **42**, 58-69. [A seminal paper about the truncated negative binomial distribution]

Schenzle, D. (1979). Fitting the truncated negative binomial distribution without the second sample moment, *Biometrics* **35**, 637-639. [Some considerations to fit the truncated negative binomial distribution are provided]

Shaban, S.A. (1988) Poisson-lognormal distributions, in *Lognormal Distributions: Theory and Applications*, E.L. Crow and K. Shimizu editors, New York, Dekker, 195-210. [A basic text on Poisson lognormal distributions]

Simberloff, D. (1979). Rarefaction as a distribution free method of expressing and estimating diversity, in *Ecological Diversity in Theory and Practice*, J.F. Grassle, G.P. Patil, W. Smith and C.Taillie editors, Fairland (Maryland), International Co-operative Publishing House, 159-176. [Rarefaction is used as a method to estimate diversity]

Smith, W., Grassle, J.F. and Kravitz, D. (1979). Measures of diversity with unbiased estimates, in *Ecological Diversity in Theory and Practice*, J.F. Grassle, G.P. Patil, W. Smith and C. Taillie editors, Fairland (Maryland), International Co-operative Publishing House, 177-192. [This suggests minimum-variance unbiased estimation for a special family of average species rarity indexes].

Tong, Y.L. (1983). Some Distribution Properties of the Sampling Species-Diversity Indices and their Applications, *Biometrics* **39**, 999-1008. [This focuses on the large-sample properties of a general diversity index].

Wyshak, G. (1974). Algorithm AS68: A program for estimating the parameters of the truncated negative distribution, *Applied Statistics* **23**, 87-91. [The paper proposes an algorithm for estimating the parameters of the truncated negative distribution]

Zahl, S. (1977). Jackknifing an index of diversity, *Ecology* **58**, 907-913. [This introduces the use of jackknife for bias reduction in diversity index estimation].

Biographical Sketch

Tonio Di Battista was born on April 8, 1957, in Castiglione a Casauria, Italy. Actually, he lives in Pescara, Italy, with his wife Paola Cappola and his two sons Giorgio and Matteo. He has received his degree in Statistics in 1983 at the “La Sapienza” University of Rome. He is Associated Professor of Statistics at the Faculty of Management Sciences of the “G. D’Annunzio” University, Chieti, Italy since 1998. He is also member of several scientific communities such as the International Environmetrics Society, the Italian Group of the International Federation of Classification Societies and the Italian Statistic Society. He has more than 50 scientific contributions in different research areas. His main scientific interests involve sampling strategy in environmental survey, statistical inference on biological populations and spatial and spatial-temporal models for diffusive phenomena in environmental framework. His main results deals with the bias adjustment of several spatial indexes and diversity indexes by using Delta-method and non-parametric resampling methods such as Bootstrap and Jackknife. In addition, he has also proposed solutions in order to overcome drawbacks in the use of Adaptive Cluster Sampling both in estimating functions of the abundance vector and when large clusters are suspected to be investigated in the biological population. Recently, he has been involved in the study of diversity

profiles. In particular, he has suggested a non-parametric approach in order to build simultaneous confidence regions and hypothesis tests on diversity profiles of biological communities. Finally, in spatial-temporal modelling, his research contributions deal with a conditional simulation in dynamic linear models for spatial and temporal predictions of diffusive phenomena and the construction of a general spatial model based on the diffusive action of some point sources.

UNESCO - EOLSS
SAMPLE CHAPTERS