

Non Parametric Inference on Diversity Profiles of Biological Populations

Inferenza non parametrica sui profili della diversità di popolazioni biologiche

Stefano A. Gattone, Tonio Di Battista

Dipartimento di Metodi Quantitativi e Teoria Economica

Università "G. D'Annunzio"

www.dmqte.unich.it

Riassunto: Quando gli ecologisti parlano della struttura di una comunità biologica essi fanno riferimento al numero degli individui presenti, al numero delle specie interagenti e alla proporzione di ciascuna specie. Gli indici di diversità incorporano questi parametri in un singolo valore numerico. Un'analisi più completa viene eseguita attraverso una funzione che fornisce uno spettro delle diverse misure di diversità: i profili della diversità. Essi possono essere espressi come funzione dell'ignoto vettore delle abbondanze. In questo articolo si sviluppa un approccio non parametrico basato sul bootstrap al fine di poter fare inferenza sui profili della diversità. La procedura proposta viene applicata su dati relativi a quattro parchi di Milano.

Keywords: Diversity profile, Replicated Sampling, Bootstrap Root Estimate, Balanced Confidence Regions.

1. Introduction

In many surveys in environmental and natural phenomena the aim is to evaluate the diversity of biological communities. Community diversity, diversity measurement and diversity comparisons have been important issues in ecological studies for more than four decades (for a review, Gove *et al.* (1994)). Many diversity indices have been developed that incorporate both species richness and the degree of evenness with which individuals are distributed among species. Generally, the main indices can be obtained as function of the unknown relative abundance vector $\mathbf{p} = (p_1, p_2, \dots, p_s)$, where the i -th element of $\mathbf{p}$ represents the proportion of units belonging to the i -th species,

$$g(\mathbf{p}) = \sum_{i=1}^s R_i(\mathbf{p}) p_i \quad (1)$$

which is the family of average species rarity indices, where $R_i(\mathbf{p})$ represents the rarity measures of the i -th species. It is important to point out it is not possible to recommend a single index as superior to all the others and choosing the appropriate index depends on what sort of question is being asked. Furthermore, studies dealing with the topic of comparison communities by means of diversity indices reached the overwhelming conclusion that no single diversity index adequately summarizes community structure (i.e. see Hurlbert (1971)). Instead of a single index, one can use a parametric family of indices whose members have varying sensitivities to the rare and abundant species: the diversity profiles.

More specifically, due to Patil and Taillie (1982), we may develop a set of profiles that incorporates indices from (1) as the following:

$$\Delta_\beta(\mathbf{p}) = \frac{1 - \sum_{i=1}^s p_i^{\beta+1}}{\beta} \quad (2)$$

For $\beta \geq -1$ we have the one-parameter family of Δ_β of diversity indices. We note that the Species Count, the Shannon index and the Simpson index are special cases of the Δ_β diversity profile with $\beta = -1, 0$ and 1 , respectively. The parameter β represents the sensibility of the index to rare or common species. Thus, a plot of Δ_β versus β provides simultaneous values of a large collection of indices showing different aspects of diversity in a single picture. As a matter of fact, changes, or differences, in community diversity can be studied by comparing profiles. Let $\Delta_\beta(C_1)$ and $\Delta_\beta(C_2)$ be the diversity profiles of community C_1 and C_2 , respectively. If $\Delta_\beta(C_1) \geq \Delta_\beta(C_2)$, that is if the profile of community C_1 lies uniformly above that of community C_2 , then C_1 is more intrinsically diverse than C_2 . Conversely, when the profiles intersect, then the communities are not intrinsically comparable. Finally, if $\Delta_\beta(C_1) = \Delta_\beta(C_2)$, then there is no difference in diversity between C_1 and C_2 .

In the literature do not seem to be the concern of inferring communities diversity profiles. The main purpose of this article is to test hypotheses in order to compare communities according to their diversity. In section 2, using the first-order Taylor expansion of (2), we express the diversity profile as a linear combination of the unknown relative abundance vector $\mathbf{p}$. Then, a non parametric methodology based on bootstrap is pursued in order to test the presence of an intrinsic diversity ordering between two communities under study. Finally, in section 3, the developed procedure is applied to a real data set.

2. Hypothesis Testing

Let C be a community with relative abundance vector $\mathbf{p}$. Let $\bar{\mathbf{p}}$ be a consistent and unbiased estimates of $\mathbf{p}$. Then, assuming Δ_β differentiable in a neighbouring of $\mathbf{p}$, we can write

$$\Delta_\beta(\bar{\mathbf{p}}) = \Delta_\beta(\mathbf{p}) + (\bar{\mathbf{p}} - \mathbf{p})\Delta'_\beta(\mathbf{p}) + o_p\|(\bar{\mathbf{p}} - \mathbf{p})\| \quad (3)$$

as $\bar{\mathbf{p}} \rightarrow \mathbf{p}$.

From the consistency of $\bar{\mathbf{p}}$, we have for large n

$$\Delta_\beta(\bar{\mathbf{p}}) = W + \Delta'_\beta(\mathbf{p})\bar{\mathbf{p}} + o_p(n^{1/2}) \quad (4)$$

where $W = \Delta_\beta(\mathbf{p}) + \Delta'_\beta(\mathbf{p})\mathbf{p}$ is a constant.

In order to evaluate and monitor biodiversity at different spatial (or temporal) scales, suitable hypotheses for comparing diversity have to be assessed. Let C_1 and C_2 two communities under study with relative abundance vector $\mathbf{p}_1$ and $\mathbf{p}_2$, respectively. Using (4) we express the diversity profiles $\Delta_\beta(\mathbf{p}_1)$ and $\Delta_\beta(\mathbf{p}_2)$ with their first-order Taylor expansion about $\mathbf{p}_1$ and $\mathbf{p}_2$:

$$\Delta_\beta(\bar{\mathbf{p}}_1) = W_1 + \Delta'_\beta(\mathbf{p}_1)\bar{\mathbf{p}}_1 + o_p(n_1^{1/2}) \quad (5)$$

$$\Delta_\beta(\bar{\mathbf{p}}_2) = W_2 + \Delta'_\beta(\mathbf{p}_2)\bar{\mathbf{p}}_2 + o_p(n_2^{1/2}) \quad (6)$$

Furthermore, let $\boldsymbol{\theta} = \Delta_\beta(\mathbf{p}_1) - \Delta_\beta(\mathbf{p}_2)$ be the vector of the difference in the diversity profile of the two communities. The equivalence hypothesis of the two profiles may be written as

$$H_0 = \boldsymbol{\theta} = \boldsymbol{\theta}_0 = \mathbf{O}$$

so that a suitable statistic to evaluate departure from the null hypothesis is

$$G_\beta(\boldsymbol{\theta}_0) = \frac{\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}_0}{\left[\frac{\Delta'_\beta(\mathbf{p}_1)^T \mathbf{V}_1 \Delta'_\beta(\mathbf{p}_1)}{n_1} \right]^{1/2} + \left[\frac{\Delta'_\beta(\mathbf{p}_2)^T \mathbf{V}_2 \Delta'_\beta(\mathbf{p}_2)}{n_2} \right]^{1/2}} \quad (7)$$

for every value of β where $\bar{\boldsymbol{\theta}} = \Delta'_\beta(\mathbf{p}_1)(\bar{\mathbf{p}}_1 - \mathbf{p}_1) - \Delta'_\beta(\mathbf{p}_2)(\bar{\mathbf{p}}_2 - \mathbf{p}_2)$ is an estimate of $\boldsymbol{\theta}$, $\bar{\mathbf{p}}_1$ and $\bar{\mathbf{p}}_2$ are the sample estimates of $\mathbf{p}_1$ and $\mathbf{p}_2$ with covariance matrices $\mathbf{V}_1$ and $\mathbf{V}_2$, respectively.

In this framework, the procedure of Beran (1988) may be applied to the roots G_β in order to obtain a simultaneous confidence set for $\boldsymbol{\theta}$ that, under H_0 , is asymptotically balanced and has a correct overall coverage probability. More explicitly, let $Q_\beta(\boldsymbol{\theta})$, $Q_{sup}(\boldsymbol{\theta})$ and $Q_{inf}(\boldsymbol{\theta})$ be the cdf's of G_β , $\sup\{Q_\beta(\boldsymbol{\theta})\}$ and $\inf\{Q_\beta(\boldsymbol{\theta})\}$ respectively. Let $Q_\beta^*(\boldsymbol{\theta})$, $Q_{sup}^*(\boldsymbol{\theta})$ and $Q_{inf}^*(\boldsymbol{\theta})$ denote the corresponding bootstrap estimates of these cdf's obtained by taking bootstrap data from the empirical distributions of the two communities separately. We define a $1 - \alpha$ simultaneous confidence set for $\boldsymbol{\theta}$ as

$$D_{\boldsymbol{\theta}, 1-\alpha} = \{\boldsymbol{\theta} : l_\beta \leq G_\beta \leq u_\beta\} \quad (8)$$

where the critical values are obtained by

$$l_\beta = Q_\beta^{*-1}[Q_{inf}^{*-1}(\frac{\alpha}{2})] \quad (9)$$

and

$$u_\beta = Q_\beta^{*-1}[Q_{sup}^{*-1}(\frac{\alpha}{2})] \quad (10)$$

On the basis of (8) we can derive a decision rule as follows:

- accept $H_0 : \Delta_\beta(\mathbf{p}_1) = \Delta_\beta(\mathbf{p}_2)$ if $\bar{\boldsymbol{\theta}} \leq D_{\boldsymbol{\theta}, 1-\alpha}$
- reject H_0 if $\bar{\boldsymbol{\theta}} \geq D_{\boldsymbol{\theta}, 1-\alpha}$ or, alternatively, if $\bar{\boldsymbol{\theta}} \leq D_{\boldsymbol{\theta}, 1-\alpha}$
- finally, if both $\bar{\boldsymbol{\theta}} \geq D_{\boldsymbol{\theta}, 1-\alpha}$ and $\bar{\boldsymbol{\theta}} \leq D_{\boldsymbol{\theta}, 1-\alpha}$ hold, then reject H_0 and the implication is that the profiles intersect and the communities C_1 and C_2 are not intrinsically comparable.

3. An application

The proposed procedures were carried out on the avian communities of 4 parks in Milan (Italy): Groane, Lambro, Forlanini and Trenno park. A detailed description of the sampling design is given in Fattorini and Marcheselli (1999). We remark that independent estimates $\hat{\mathbf{p}}_i$ ($i = 1, \dots, n$) were available from the replicated use of a suitable encounter design. Accordingly, the bootstrap procedures described in section 2 can be applied. 9 evenly spaced β values in the range $-1 \leq \beta \leq 1$ were generated to construct the diversity

profile for each park. In order to rank the parks according to their diversity we evaluate an estimate of θ , the difference in the diversity profile for each of the possible couples of parks. Under the null hypothesis of no difference in diversity we derive a 0.95 simultaneous confidence sets for θ . Results show that Groane park and Trenno park turn out to be the most intrinsically diverse and the least intrinsically diverse, respectively. Moreover, Lambro and Forlanini parks can be located at an intermediate diversity level being these parks equivalent in terms of intrinsic diversity profiles.

References

- Beran R. (1988) Balanced simultaneous confidence sets, *Journal of the American Statistical Association*, 83, 679–686.
- Fattorini L. and Marcheselli M. (1999) Inference on intrinsic diversity profiles of biological populations, *Environmetrics*, 10, 589–599.
- Gove H.J., Patil J.P. and Swindel B.F. (1994) Ecological diversity and forest management, *Handbook of Statistics*, 12, 409–462.
- Hurlbert S.H. (1971) The non concept of species diversity: a critique and alternative parameters, *Ecology*, 52, 577–586.
- Patil G.P. and Taillie C. (1982) Diversity as a concept and its measurement, *Journal of the American Statistical Association*, 77, 548–567.