INFERENCE ON FUNCTIONAL BIODIVERSITY TOOLS
Tonio Di Battista\textsuperscript{1}, Francesca Fortuna\textsuperscript{1} and Fabrizio Maturo\textsuperscript{2}

\textsuperscript{1} Department of Philosophical, Pedagogical and Quantitative Economic Sciences, University G. D Annunzio, Chieti-Pescara, Italy, (e-mail: dibattis@unich.it, francesca.fortuna@unich.it)
\textsuperscript{2} Department of Management and Business Administration, University G. D Annunzio, Chieti-Pescara, Italy, (e-mail: f.maturo@unich.it)

ABSTRACT:
Biodiversity profiles can be analysed in a functional framework because they are expressed as functions of the unknown abundance vector. This paper aims to construct simultaneous confidence bands for the $\beta$-diversity profile and for its functional mean. Using a suitable sampling design, inference is developed assuming an asymptotic specific distribution of the profile estimator. The use of any probability distribution is hazardous because ecologists generally deals with small sample sizes. For this reason we propose simultaneous confidence bands adopting a non-parametric bootstrap procedure.

KEYWORDS: “$\beta$-diversity profile”, “bootstrap”, “confidence bands”, “functional mean”, “inference in biodiversity”

1 Introduction
Biodiversity measurement and diversity comparison are important aspects of environmental statistics. Although many indices have been proposed in the literature, a single index is not suitable in comparing ecological communities (Gove \textit{et al.}, 1994); a possible solution is the use of diversity profiles (Patil & Taillie, 1979). A diversity profile is a curve depicting several values of diversity indices in a single graph. Since diversity profile can be expressed as a function of the unknown abundance vector, it can be analysed in a functional framework (Gattone & Di Battista, 2009). Functional data analysis (FDA) approach is a useful tool for a deeper analysis of phenomena varying in a fixed domain (Ramsay & Silverman, 2005). In this context, our aim is to identified confidence intervals for the $\beta$-diversity profile and for its functional mean. Assuming that a sample belongs to the family of replicated sampling design, inference is developed assuming an asymptotic specific distribution of the profile estimator, adopting a non-parametric bootstrap procedure. The paper is
organized as follows: Section 2 introduces the \( \beta \)-diversity profile and its unbiased estimator. Section 3 deals with the \( \beta \)-profile and its mean confidence bands.

## 2 \( \beta \)-profile estimator

Let us suppose that an ecological population is composed by \( N \) units and is partitioned into \( s \) species \((i = 1, 2, ..., s)\). Let \( \mathbf{N} = (N_1, ..., N_s)^T \) be the species abundance vector whose generic element \( N_i \) represents the number of individuals belonging to the \( i \)-th species, and let \( \mathbf{p} = (p_1, ..., p_s)^T \) be the relative abundance vector with \( p_i = N_i / \sum_{i=1}^{s} N_i \). To evaluate the community biodiversity, we consider the \( \beta \) diversity profile proposed by (Patil & Taillie, 1979):

\[
\Delta_\beta = \Delta_\beta_{\mathbf{N} \mathbf{p}} = \sum_{i=1}^{s} \left( 1 - \frac{p_i^\beta}{\beta} \right) p_i \quad \beta \geq -1
\]  

(1)

where the value of \( \beta \) denotes the sensibility of the index to rare and common species. The plot of \( \Delta_\beta \) versus \( \beta \) provides the diversity profile, which shows simultaneous values of a large collection of indices. \( \Delta_\beta \) is a function of \( \beta \) in a closed domain, \( \beta \in [-1, 1] \), with the abundance vector \( \mathbf{p} \) as parameter (Gattone & Di Battista, 2009). Since in real surveys species abundances are unknown and a list frame of biological populations are rarely available, diversity indexes are estimated by means of replicated encountered sampling design (Barabesi & Fattorini, 1998). In particular, if \( n \) independent replications of the sampling procedure are performed, the \( n \) samples provide \( n \) i.i.d. unbiased estimators for \( \mathbf{N} \), say \( \hat{\mathbf{N}}_j = (\hat{N}_{1j}, \hat{N}_{2j}, ..., \hat{N}_{sj})^T \). Accordingly, an unbiased, consistent estimator for \( \mathbf{N} \) can be expressed by \( \hat{\mathbf{N}}_n = \frac{1}{n} \sum_{j=1}^{n} \hat{N}_j \), with variance-covariance matrix \( \text{Var}(\hat{\mathbf{N}}_n) = \Sigma / n \), where \( \hat{\mathbf{N}}_j \) be the Horwitz-Thompson estimator of the abundance vector of the \( j \)-th sample. Obviously, owing to the i.i.d. nature of the \( \hat{\mathbf{N}}_j \) random variables, the Central Limit Theorem ensures that, for large \( n \), \( \sqrt{n}(\hat{\mathbf{N}}_n - \mathbf{N}) \to \text{NMV}(\mathbf{0}, \Omega) \), where \( \Omega \) is the variance-covariance matrix of the abundance vector. The joint use of the Sverdrup and the Slutski theorems ensures that the studentized statistic, say \( T_{\Delta_\beta \hat{\mathbf{N}}_n} \), for \( n \to \infty \), is:

\[
T_{\Delta_\beta \hat{\mathbf{N}}_n} = n^{1/2} \frac{\Delta_\beta(\hat{\mathbf{N}}_n) - \Delta_\beta(\mathbf{N})}{[g(\hat{\mathbf{N}}_n)\Omega g(\hat{\mathbf{N}}_n)^T]^{1/2}} \to N(0, 1)
\]  

(2)

where \( g(\hat{\mathbf{N}}_n) \) is is the Jacobian matrix of the function \( \Delta_\beta(\mathbf{N}) \) computed at \( \hat{\mathbf{N}}_n \) and \( \Omega \) is a consistent estimator of \( \Omega \). The expression (2) provides the pivotal
quantity for determining a confidence interval for $\triangle \beta(N)$. However, in real surveys, large sample sizes are rarely available, hence, asymptotic results could not be a practical hypothesis to work with. For this reason, replicated random sampling bootstrap procedure on i.i.d. samples $\hat{N}_1, \hat{N}_2, \ldots, \hat{N}_n$ can be suitably adopted in order to build non-parametric simultaneous confidence region on $\beta$-diversity profiles (Di Battista & Gattone, 2002).

### 3 Confidence intervals for beta diversity profile and its mean

Starting from (2), a confidence set for a single component $\triangle \beta(N)$ can be easily obtained. Let $l_{\beta}$ and $u_{\beta}$ be consistent estimates for the $\gamma$-th and the $1 - \gamma$-th quantiles of the distribution, say $H_{\beta}$, of $T_{\triangle \beta}$. The confidence set with asymptotic coverage probability $1 - \gamma$ for $\triangle \beta(N)$ is $D_{\beta} = \{d \in \triangle \beta(N) : l_{\beta} \leq T(d) \leq u_{\beta}\}$. Simultaneously asserting a confidence region for each component of the $\beta$-diversity profile, we obtain a simultaneous confidence set $D$ for the family of parametric functions $\triangle \beta(N)$. Furthermore, let $\triangle$ be the set of all possible values of $\triangle \beta(N)$ as $N$ varies over the parameter space, we can generate the following $1 - \alpha$ confidence set for $\triangle \beta(N)$ as $D = \{d \in \triangle : \triangle \beta(p) \in D_{\beta} \ \forall \beta \in [-1, 1]\}$. The key factor is how to choose the critical values $\{l_{\beta}, u_{\beta} : \beta \in [-1, 1]\}$. Let $H_{sup}$ and $H_{inf}$ be the left continuous cumulative distribution functions of the transformed root $sup\{H_{\beta} : \beta \in [-1, 1]\}$ and $inf\{H_{\beta} : \beta \in [-1, 1]\}$, respectively. The proposed bootstrap version of $D$ is obtained by taking the critical values $l_{\beta} = H_{\beta}^{-1}(1 - \frac{\alpha}{2})$ and $u_{\beta} = H_{\beta}^{-1}(\frac{\alpha}{2})$ where $H_{sup}, H_{inf}$ and $H_{sup}^*$ are the bootstrap distribution estimates of $H_{\beta}, H_{inf}$ and $H_{sup}$, respectively (Di Battista & Gattone, 2002). Since diversity profile can be expresses as a function of the abundance vector in a fixed domain, it can be analyzed in a functional context (Gattone & Di Battista, 2009). The classical FDA approach (Ramsay & Silverman, 2005) assumes the existence of certain unknown smooth functions $f(\cdot)$ which generate the data. The true form of $f(\cdot)$ is fitted through appropriate techniques such as basis functions expansion. Sometimes these techniques mystify the intrinsic characteristics of the data; for example, when the underlying data process is known in advance (T. Di Battista, 2015). In these cases it is preferable to work directly on the reference functional space, if it is possible. This paper focuses on functional data expressed by a specific function known in advance. In these cases, the functional space $S$ is constituted by a set of functions belonging to the same parametric family that is $S = \{f(\Theta : x)\}$, where $\Theta = (\theta_1, \theta_2, \ldots, \theta_s)^T$ represents a set of parameters taking values in a parameter space $\Theta$, $x$ is the functional domain and $S$ is a convex...
subset of some $L^p$ space. Starting from $n$ parametric functional data, we aim to identify a functional mean that belongs to the same family as the functions. For this purpose, we propose a transformation $T$, that converts the functional space in a new convex space. If $S$ is a parametric set of functions, then $T$ is the unique transformation that associates to each functional datum its parameter; that is $T(f(\theta, x)) = \theta$ and the functional mean is obtained through $T$. In particular, the functional mean is the element of $S$ with the mean of the parameters as associated parameter, that is $f(\bar{\theta}, x)$. In this framework, it is possible to prove, under a suitable hypothesis of monotonic dependence by parameters, the main properties of the mean in the functional case (T. Di Battista, 2015). The $\beta$-profiles in equation (1) belong to a convex subspace $S$ and its functional form is known in advance because it is expressed by a parametric model. For these reasons, we propose to construct confidence intervals for the functional mean working directly on the reference functional space. In this way we can preserve the typical characteristics of the profile.

References


