Entropic representation and estimation of diversity indices

Zhiyi Zhang and Michael Grabchak*

Department of Mathematics and Statistics, University of North Carolina at Charlotte, Charlotte, NC, USA

(Received 13 March 2014; accepted 19 March 2016)

This paper serves a twofold purpose. First, a unified perspective on diversity indices is introduced based on an entropic basis. It is shown that the class of all linear combinations of the entropic basis, referred to as the class of linear diversity indices, covers a wide range of diversity indices used in the literature. Second, a class of estimators for linear diversity indices is proposed and it is shown that these estimators have rapidly decaying biases and asymptotic normality.

Keywords: diversity indices; entropy; entropic basis; nonparametric estimation; asymptotic normality

AMS Subject Classifications: 62G05; 62G20

1. Introduction

Diversity is an important notion in a variety of scientific disciplines. Historically, the interest stems from ecological applications, where the diversity of species in an ecosystem is a relevant issue. Other applications include cancer research, where the interest is in the diversity of types of cancer cells in a tumour, and linguistics, where it is in the diversity of an author’s vocabulary. More generally, in information science, one is interested in the diversity of letters drawn from some alphabet. While there is little controversy on the literal meaning of the word ‘diversity’, the best way to quantify this concept is a matter of some dispute. For this reason many indices of diversity have been proposed in the literature. Two of the earliest are Shannon’s entropy introduced in Shannon (1948) and Simpson’s index introduced in Simpson (1949). Many other popular diversity indices have been proposed, including Emlen’s index, the Gini–Simpson index, Hill’s diversity number, Rényi’s entropy, and Tsallis entropy. These are, respectively, named after the authors of Emlen (1973), Gini (1912), Hill (1973), Rényi (1961), and Tsallis (1988). Comprehensive discussions are offered by many, see, e.g. Magurran (1988) and Marcon (2015). It is quite clear that, when it comes to the question of what constitutes a mathematical diversity index, the ideas are diverse. This paper offers a unified perspective for all of the above-mentioned diversity indices based on a re-parameterisation and establishes a general nonparametric estimation procedure for these indices.

In the following section, we give a general discussion of diversity indices. In Section 3 we show that all diversity indices (satisfying very general regularity conditions) are transforms of a relatively simple class of diversity indices, which we call the entropic basis. Furthermore, we
show that most diversity indices used in the literature are either linear combinations (possibly infinite) of the entropic basis or equivalent to such indices. In Section 4 we give a general non-parametric approach for estimating such linear diversity indices. We derive an estimator, which, under general conditions, is consistent, asymptotically normal, and has a bias that decays exponentially fast. In Section 5 we show how to extend these results to Rényi’s entropy, which is not a linear diversity index, but is only equivalent to one.

2. Preliminaries

Let the species in a population be denoted by letters of a countable alphabet $\mathcal{X} = \{\ell_k; k \geq 1\}$ and let the abundances of the species be represented by a probability distribution $p = \{p_k; k \geq 1\}$ associated with these letters. For simplicity of notation, and when there is no risk of ambiguity, let $K = \sum_{k \geq 1} 1[p_k > 0]$, where $1[\cdot]$ is the indicator function, be referred to as the cardinality of $\mathcal{X}$, which may be finite or countably infinite.

An index of diversity $\theta = \theta(\mathcal{X}, p)$ is a function taking the alphabet $\mathcal{X}$ and the probability distribution $p$ into the extended real line $[-\infty, \infty]$. What additional conditions $\theta$ must satisfy is a matter of opinion or need depending on the particular population features of interest. The discussion of this paper starts with the following set of axioms:

$A_{01}$ : $\theta$ is fully determined by $p = \{p_k; k \geq 1\}$.

$A_{02}$ : $\theta$ is invariant under any permutation on the index set $\{k; k \geq 1\}$.

We will refer to these axioms as $\mathcal{A}_0 = \{A_{01}, A_{02}\}$. Note that Axiom $A_{01}$ implies that $\theta = \theta(\mathcal{X}, p) = \theta(p)$ does not depend on $\mathcal{X}$. The axiomatic set $\mathcal{A}_0$ is by no means sufficient for a meaningful diversity index, but it provides a minimal constraint for further discussion. Most, if not all, popular diversity indices studied in the literature satisfy $\mathcal{A}_0$. These include:

1. Simpson’s index $\lambda = \sum_{k \geq 1} p_k^2$,
2. Gini–Simpson index $1 - \lambda = \sum_{k \geq 1} p_k(1 - p_k)$,
3. Shannon’s entropy $H = -\sum_{k \geq 1} p_k \ln(p_k)$,
4. Rényi’s entropy $H_r = (1 - r)^{-1} \ln(\sum_{k \geq 1} p_k^r)$ for any $r > 0, r \neq 1$,
5. Tsallis’ entropy $T_r = (1 - r)^{-1} (\sum_{k \geq 1} p_k^r - 1)$ for any $r > 0, r \neq 1$,
6. Hill’s diversity number $N_r = (\sum_{k \geq 1} p_k^r)^{1/(1-r)}$ for any $r > 0, r \neq 1$,
7. Emlen’s index $D = \sum_{k \geq 1} p_k e^{-p_k}$, and
8. The Richness index $K = \sum_{k \geq 1} 1[p_k > 0]$.

Given the abundance of definitions of various diversity indices, let us define a notion of equivalence between two indices.

**Definition 1** Two diversity indices, $\theta_1$ and $\theta_2$, are said to be equivalent if and only if there exists a strictly increasing function $g$ such that $\theta_1 = g(\theta_2)$. The equivalence is denoted by $\theta_1 \Leftrightarrow \theta_2$.

Noting that if $g$ is strictly increasing then so is its inverse $g^{-1}$, it is clear that the above definition is symmetric with respect to $\theta_1$ and $\theta_2$. Two diversity indices are equivalent if they agree on whether one population is more diverse than another, but they may not agree on what the actual difference is. With this rather trivial notion of equivalence much of the superficial redundancy among various indices in the literature can be erased. For example, Rényi’s entropy
Hr, Tsallis entropy Tr, and Hill’s diversity number Nr are equivalent to each other. Furthermore, for \( r \in (0, 1) \) they are equivalent to a core index

\[
h_r = \sum_{k \geq 1} p_k^r, \quad r > 0, r \neq 1,
\]

which we refer to as Rényi’s equivalent entropy. Although these indices are not equivalent to \( h_r \) for \( r > 1 \), they are nevertheless continuous monotone transformations of \( h_r \). This fact will be useful for statistical estimation, see Section 5.

3. Entropic basis

In this section we give our first main result. We introduce a class of diversity indices and show that they are the building blocks from which all diversity indices satisfying the axioms of \( \mathcal{A}_0 \) are made. For a given \( p = \{p_k; k \geq 1\} \) and an integer pair \((u \geq 1, v \geq 0)\), let \( \zeta_{u,v} = \sum_{k \geq 1} p_k^u(1 - p_k)^v \). We refer to

\[
\zeta = \{\zeta_{u,v}; u \geq 1, v \geq 0\}
\]

as the family of generalised Simpson’s diversity indices. This family was first introduced in Zhang and Zhou (2010). Furthermore, we refer to the sub-family \( \zeta_1 = \{\zeta_{1,v}; v \geq 0\} \) as the entropic basis with regard to \( \mathcal{X} \) and \( p \). Note that \( \zeta_{1,0} = \sum_{k \geq 1} p_k(1 - p_k)^0 \equiv 1 \).

Remark 1 According to Zhang and Zhou (2010), every member of \( \zeta \) can be expressed as \( \zeta_{u,v} = \sum_{k \geq 1}[p_k^{u-1}][p_k(1 - p_k)][(1 - p_k)^{v-1}] \), which may be regarded as a weighted version of the Gini–Simpson diversity index, with weights \( [p_k^{u-1}][(1 - p_k)^{v-1}] \) for various choices of \( u \) and \( v \) at the user’s discretion, and hence the term generalised Simpson’s diversity indices.

Remark 2 Zhang (2012) established an alternative representation of Shannon’s entropy, \( H \), as \( H = \sum_{v=1}^{\infty} v^{-1}\zeta_{1,v} \) provided that \( H < \infty \). This is a linear form of \( \zeta_1 \), and hence the term entropic basis.

We now show that any diversity index satisfying the axioms of \( \mathcal{A}_0 \) must be a function of \( \zeta_{1,v} \), for \( v = 0, 1, \ldots \), i.e. of the members of the entropic basis.

Theorem 1 Given an alphabet \( \mathcal{X} = \{\ell_k; k \geq 1\} \) and an associated probability distribution \( p = \{p_k; k \geq 1\} \), a diversity index \( \theta \) satisfying \( \mathcal{A}_0 \) is fully determined by the entropic basis \( \zeta_1 = \{\zeta_{1,v}; v \geq 0\} \).

A proof of Theorem 1 requires the following lemma due to Zhang and Zhou (2010).

Lemma 1 The distribution \( p = \{p_k; k \geq 1\} \) on \( \mathcal{X} \) and the family of generalised Simpson’s diversity indices \( \zeta = \{\zeta_{u,v}; u \geq 1, v \geq 0\} \) uniquely determine each other up to a permutation of the index set \( \{k; k \geq 1\} \).
Proof of Theorem 1  By Lemma 1, $\zeta$ determines $p$ up to a permutation, and this fully determines a diversity index $\theta$ satisfying $\mathcal{A}_0$. It remains to show that every element of $\zeta_1$ is fully determined by $\zeta_1$. Towards that end, we note that for any pair of fixed integers $u \geq 2$ and $v \geq 0$

$$\zeta_{u,v} = \sum_{k \geq 1} p_k [1 - (1 - p_k)]^{u-1}(1 - p_k)^v$$

$$= \sum_{k \geq 1} p_k \left[ \sum_{i=0}^{u-1} (-1)^i \binom{u-1}{i} (1 - p_k)^i \right] (1 - p_k)^v$$

$$= \sum_{i=0}^{u-1} (-1)^i \binom{u-1}{i} \left[ \sum_{k \geq 1} p_k (1 - p_k)^{v+i} \right]$$

$$= \sum_{i=0}^{u-1} (-1)^i \binom{u-1}{i} \zeta_{1,v+i},$$

which completes the proof. ■

The statement of Theorem 1 holds true for any probability distribution $p$ regardless of whether $K$ is finite or infinite. Theorem 1, essentially, offers a re-parameterisation of $p$ (up to a permutation) in terms of $\zeta_1$. This re-parameterisation is not just an arbitrary one, it has several statistical implications. First, every element of $\zeta_1$ contains information about the entire distribution and not just one frequency $p_k$. This helps to deal with the problem of estimating probabilities of unobserved species. Second, for a random sample of size $n$, there are very good estimators of $\zeta_{1,v}$ for $v = 0, 1, 2, \ldots, n-1$. These are given in Zhang and Zhou (2010) and are discussed below.

While, in general, a diversity index can be any transformation of the entropic basis, in practice, most commonly used indices correspond to transformations of a fairly simple form. Most diversity indices either belong to, or are equivalent to ones that belong to the following class.

**Definition 2** A diversity index $\theta$ is said to be a linear diversity index if it is a linear combination of the elements of the entropic basis, i.e.

$$\theta = \theta(p) = \sum_{v=0}^{\infty} w_v \zeta_{1,v} = \sum_{v=0}^{\infty} w_v \sum_{k \geq 1} p_k (1 - p_k)^v$$

(1)

for any choice of weights $w_v$ such that, for every $p$, the sum either converges or diverges to $\pm \infty$.

Definition 2, essentially, encircles a sub-class of indices among all functions of $\zeta_1$, i.e. all diversity indices satisfying $\mathcal{A}_0$. While there are no fundamental reasons why a search of a good diversity index should be restricted to this sub-class, it happens to cover all of the popular indices that we have come across in the literature, up to the equivalence relationship given in Definition 1. These include:

**Simpson’s index:** $\lambda = \sum_{k \geq 1} p_k^2 = \zeta_{1,0} - \zeta_{1,1},$

**Gini-Simpson index:** $1 - \lambda = \sum_{k \geq 1} p_k (1 - p_k) = \zeta_{1,1},$
Shannon’s entropy: \[ H = - \sum_{k \geq 1} p_k \ln(p_k) = \sum_{v=1}^{\infty} \frac{1}{v} \zeta_{1,v}, \]

Rényi equiv. entropy: \[ h_r = \sum_{k \geq 1} p_k^r = \zeta_{1,0} + \sum_{v=1}^{\infty} \prod_{i=1}^{v} \left( \frac{i-r}{i} \right) \zeta_{1,v}, \]

Emlen’s index: \[ D = \sum_{k \geq 1} p_k e^{-p_k} = \sum_{v=0}^{\infty} e^{-1} \frac{1}{v!} \zeta_{1,v}, \]

Richness index: \[ K = \sum_{k \geq 1} 1[p_k > 0] = \sum_{v=0}^{\infty} \zeta_{1,v}, \]

Gen. Simpson’s index: \[ \zeta_{u,m} = \sum_{k \geq 1} p_k^u (1 - p_k)^m = \sum_{v=0}^{u-1} (-1)^v \frac{(u-1)}{v} \zeta_{1,v}. \]

Note that Tsallis’ entropy is also a linear diversity index. The form of its weights is very similar to that of Rényi’s equivalent entropy. All of the representations above can be verified using Taylor expansions. For example, for the richness index, which is the total number of species in a population, we have

\[
K = \sum_{k \geq 1} 1[p_k > 0] \frac{p_k}{1 - (1 - p_k)}
= \sum_{k \geq 1} 1[p_k > 0] p_k \sum_{v=0}^{\infty} (1 - p_k)^v
= \sum_{v=0}^{\infty} \sum_{k \geq 1} 1[p_k > 0] p_k (1 - p_k)^v
= \sum_{v=0}^{\infty} \zeta_{1,v}.
\]

It is not difficult to see that all linear diversity indices discussed above are of the general form

\[ \theta = \sum_{k \geq 1} p_k h(p_k), \tag{2} \]

where \( h \) has a Taylor expansion around 1 with radius of convergence at least 1. This is not a coincidence. If \( \theta \) satisfies Equation (1) and \( h(t) = \sum_{v=0}^{\infty} w_v (1 - t)^v \) then \( \theta \) is necessarily of the form Equation (2), and, of course, the converse of this statement holds.

**Remark 3** In the literature of diversity indices, it is generally thought that the richness indices, e.g. \( K \), and the evenness indices, e.g. Gini–Simpson’s \( 1 - \lambda \), are two qualitatively different types of indices, see, e.g. Peet (1974) and Heip, Herman, and Soetaert (1998). It may be interesting to note that, in the perspective of the entropic basis, they are both linear diversity indices and merely differ in the weighting scheme \( h(p) \) in Equation (2).

### 4. Estimation of linear diversity indices

In this section we discuss nonparametric estimation of linear diversity indices. Assume that \( X_1, X_2, \ldots, X_n \) are independent and identically distributed (iid) from \( \mathcal{X} \) according to \( p \). We want to estimate

\[ \theta = \theta(p) = \sum_{v=0}^{\infty} w_v \zeta_{1,v} = \sum_{v=0}^{\infty} w_v \sum_{k \geq 1} p_k (1 - p_k)^v. \tag{3} \]

We assume that \( \{w_v : v \geq 0\} \) has been chosen but that \( p = \{p_k : k \geq 1\} \) is unknown. We will make the following assumptions:
There is an $M > 0$ such that $|w_v| \leq M$ for all $v \geq 0$, and

(2) $K < \infty$.

These conditions guarantee that the sum in Equation (3) always converges. Note that the assumption that $|w_v| \leq M$ is satisfied by all of the linear diversity indices discussed in Section 3. Note further that we are not assuming that $K$ is known, only that it is known that $K < \infty$. This is realistic in many applications, including ecology, where there is a finite (even if very large) number of species.

For simplicity of notation assume that the frequencies $p_k$ are ordered such $p_k > 0$ for $k = 1, 2, \ldots, K$ and that $p_k = 0$ for $k > K$. For $t \in [0, 1]$ let

$$h(t) = \begin{cases} \sum_{v=0}^{\infty} w_v (1 - t)^v & \text{if } t \in (0, 1], \\ 0 & \text{if } t = 0. \end{cases}$$

In this case, $w_v = (-1)^v h^{(v)}(1)/v!$ and

$$\theta = \sum_{k=1}^{K} p_k h(p_k).$$

Let $\{x_k = \sum_{i=1}^{n} 1[X_i = \ell_k]\}$ be the observed counts in our sample and let $\{\hat{p}_k = x_k/n\}$ be the sample proportions. Perhaps the most intuitive estimator of $\theta$ is the plug-in estimator given by

$$\hat{\theta}_n = \sum_{v=0}^{\infty} w_v \sum_{k=1}^{K} \hat{p}_k (1 - \hat{p}_k)^v = \sum_{k=1}^{K} \hat{p}_k h(\hat{p}_k). \quad (4)$$

However, it is well known that, in many important situations, the plug-in estimator has a bias that decays very slowly. For instance, in the case of Shannon’s entropy (i.e. when $w_0 = 0$ and $w_v = 1/v$ for $v \geq 1$), the bias decays no faster that $O(1/n)$, see, e.g., Paninski (2003). We now propose another estimator, which has a bias that always decays at least exponentially fast. Our approach is influenced by the estimator of Shannon’s entropy derived in Zhang (2012).

First note that we can write

$$\theta = w_0 + \sum_{v=1}^{n-1} w_v \sum_{k=1}^{K} p_k (1 - p_k)^v + \sum_{v=n}^{\infty} w_v \sum_{k=1}^{K} p_k (1 - p_k)^v =: \eta_n + B_{2,n}. \quad (5)$$

From Zhang and Zhou (2010), we know that an unbiased estimator of $\eta_n$ is given by

$$\hat{\eta}^* = \sum_{v=1}^{n-1} w_v \sum_{k=1}^{K} \hat{p}_k \prod_{j=1}^{v} \left(1 - \frac{x_k - 1}{n - j}\right) = \sum_{k=1}^{K} \hat{\theta}_{n,k}^*,$$

where

$$\hat{\theta}_{n,k}^* = \hat{p}_k \prod_{j=1}^{n-x_k} \left(1 - \frac{x_k - 1}{n - j}\right).$$

It may, at first, appear that one needs to know $K$ in order to evaluate this estimator. However, if a category $k$ is not observed then $\hat{p}_k = 0$ and hence $\hat{\theta}_{n,k}^* = 0$ and does not need to be included in
Applying the delta method to Equation (8) gives the following result.

By construction, the bias of the estimator $\hat{\theta}^p_n$ is given by $B_{2,n}$. Letting $p_\lambda = \min\{p_k : 1 \leq k \leq K\}$ we see that Equation (5) implies that

$$|B_{2,n}| \leq \sum_{v=\min(p_k \wedge v)} m \sum_{k=1}^K p_k (1 - p_k)^v \leq M \sum_{k=1}^K p_k \sum_{v=\min(p_k \wedge v)} (1 - p_k)^v,$$

which decays exponentially fast in $n$. We note that, in the case of Shanon’s entropy, this estimator corresponds to the estimator introduced in Zhang (2012) and Zhang (2013). For that estimator, an approach to further reduce the bias was presented in Zhang and Grabchak (2013). One can modify that approach for our more general situation. This will be dealt with in a future work.

Next, we will establish that $\hat{\theta}^p_n$ is a consistent and asymptotically normal estimator of $\theta$. Along the way, we will show the corresponding results for the plug-in estimator $\hat{\theta}_n$. Our approach is similar to the one used in Zhang (2013) to prove the asymptotic normality of an estimator of Shanon’s entropy. Before proceeding we introduce some notation. We write $\xrightarrow{P}$ to denote convergence in probability and we write $\xrightarrow{L}$ to denote convergence in law.

Let us define the $(K-1)$-dimensional vectors

$$\mathbf{v} = (p_1, \ldots, p_{K-1})^\tau \quad \text{and} \quad \hat{\mathbf{v}}_n = (\hat{p}_1, \ldots, \hat{p}_{K-1})^\tau,$$

and note that $\hat{\mathbf{v}}_n \xrightarrow{P} \mathbf{v}$ as $n \rightarrow \infty$. Moreover, by the multivariate normal approximation to the multinomial distribution

$$\sqrt{n}(\hat{\mathbf{v}}_n - \mathbf{v}) \xrightarrow{L} \MVN(0, \Sigma(\mathbf{v})) \tag{8}$$

where $\Sigma(\mathbf{v})$ is the $(K-1) \times (K-1)$ covariance matrix given by

$$\Sigma(\mathbf{v}) = \begin{pmatrix}
    p_1 (1 - p_1) & -p_1 p_2 & \cdots & -p_1 p_{K-1} \\
    -p_2 p_1 & p_2 (1 - p_2) & \cdots & -p_2 p_{K-1} \\
    \vdots & \vdots & \ddots & \vdots \\
    -p_{K-1} p_1 & -p_{K-1} p_2 & \cdots & p_{K-1} (1 - p_{K-1}) \\
\end{pmatrix}.$$

Let

$$G(\mathbf{v}) = \sum_{k=1}^{K-1} p_k h(p_k) + \left(1 - \sum_{k=1}^{K-1} p_k\right) h\left(1 - \sum_{k=1}^{K-1} p_k\right)$$

and

$$G'(\mathbf{v}) := \nabla G(\mathbf{v}) = \left(\frac{\partial}{\partial p_1} G(\mathbf{v}), \ldots, \frac{\partial}{\partial p_{K-1}} G(\mathbf{v})\right)^\tau.$$

Note that $G(\mathbf{v}) = \theta(\mathbf{p})$. The reason for this change in notation is to emphasise the fact that we are now thinking of this as a function of $\mathbf{v} = \{p_k : k = 1, 2, \ldots, K - 1\}$ and not of $\mathbf{p} = \{p_k : k = 1, 2, \ldots, K\}$. For each $j, j = 1, \ldots, K - 1$, we have

$$\frac{\partial}{\partial p_j} G(\mathbf{v}) = h(p_j) + p_j h'(p_j) - h\left(1 - \sum_{k=1}^{K-1} p_k\right) - \left(1 - \sum_{k=1}^{K-1} p_k\right) h'\left(1 - \sum_{k=1}^{K-1} p_k\right).$$

Applying the delta method to Equation (8) gives the following result.
Proposition 1  If \( \hat{\theta}_n \) is the plug-in estimator given by Equation (4) and \( g^T(\mathbf{v})\Sigma(\mathbf{v})g(\mathbf{v}) > 0 \), then
\[
\sqrt{n}(\hat{\theta}_n - \theta)[g^T(\mathbf{v})\Sigma(\mathbf{v})g(\mathbf{v})]^{-1/2} \xrightarrow{L} N(0, 1).
\]

Remark 4  It is well known that \( \Sigma(\mathbf{v}) \) is a positive-definite matrix, see, e.g. Tanabe and Sagae (1992). For this reason, the condition \( g(\mathbf{v}) \neq 0 \) is equivalent to the condition that \( g(\mathbf{v}) \neq 0 \). The question of when this holds depends on the function \( h \). In the case of entropy (when \( h(t) = -\log t \)) and Rényi’s equivalent entropy (when \( h(t) = t^{r-1} \)), it is easy to verify that \( g(\mathbf{v}) = 0 \) if and only if \( p_k = 1/K \) for \( k = 1, 2, \ldots, K \).

In order to use Proposition 1 in applications, we need to be able to estimate \( g^T(\mathbf{v})\Sigma(\mathbf{v})g(\mathbf{v}) \). By the continuous mapping theorem, we can estimate \( \Sigma(\mathbf{v}) \) by \( \hat{\Sigma}(\hat{\mathbf{v}}) \). However, \( g(\hat{\mathbf{v}}) \) may not be defined when there are species that have not been observed in the sample. To deal with this, for \( \mathbf{x} = (x_1, x_2, \ldots, x_{K-1})^T \in [0, 1]^{K-1} \) with \( \sum_{i=1}^{K-1} x_i = 0 \) define
\[
\bar{g}(\mathbf{x}) = (\bar{g}_1(\mathbf{x}), \bar{g}_2(\mathbf{x}), \ldots, \bar{g}_{K-1}(\mathbf{x}))^T,
\]
where for \( j = 1, 2, \ldots, K - 1 \)
\[
\bar{g}_j(\mathbf{x}) = \begin{cases} \frac{\partial}{\partial x_j} G(\mathbf{x}) & \text{if } x_j > 0, \\ 0 & \text{otherwise}. \end{cases}
\]

Since \( \hat{\mathbf{v}}_n \xrightarrow{P} \mathbf{v} \), \( \bar{g} \) is continuous for all \( \mathbf{x} \in (0, 1)^{K-1} \), and \( \mathbf{v} \in (0, 1)^{K-1} \), the continuous mapping theorem implies that \( \bar{g}^T(\hat{\mathbf{v}}_n) \) is a consistent estimator of \( g^T(\mathbf{v}) \). From this and Slutsky’s theorem, we get the following.

Corollary 1  If \( \hat{\theta}_n \) is the plug-in estimator given by Equation (4) and \( g^T(\mathbf{v})\Sigma(\mathbf{v})g(\mathbf{v}) > 0 \), then
\[
\sqrt{n}(\hat{\theta}_n - \theta)[\bar{g}^T(\hat{\mathbf{v}}_n)\Sigma(\hat{\mathbf{v}}_n)\bar{g}(\hat{\mathbf{v}}_n)]^{-1/2} \xrightarrow{L} N(0, 1).
\]

Since both \( \Sigma(\hat{\mathbf{v}}_n) \) and \( \bar{g}(\hat{\mathbf{v}}_n) \) have zeros in locations that correspond to unobserved species, we can pretend that these species do not exist for the purposes of estimating \( \bar{g}^T(\hat{\mathbf{v}}_n)\Sigma(\hat{\mathbf{v}}_n)\bar{g}(\hat{\mathbf{v}}_n) \). For this reason, we do not actually need to know the value of \( K \) to evaluate this quantity. We now extend our results to the estimator defined in Equation (6).

Theorem 2  If \( \hat{\theta}_n^\tau \) is the estimator given by Equation (6) and \( g^T(\mathbf{v})\Sigma(\mathbf{v})g(\mathbf{v}) > 0 \), then
\[
\sqrt{n}(\hat{\theta}_n^\tau - \theta)[g^T(\mathbf{v})\Sigma(\mathbf{v})g(\mathbf{v})]^{-1/2} \xrightarrow{L} N(0, 1).
\]

Before giving the proof, we state the following corollary. Its proof is similar to that of Corollary 1.

Corollary 2  If \( \hat{\theta}_n^\tau \) is the estimator given by Equation (6) and \( g^T(\mathbf{v})\Sigma(\mathbf{v})g(\mathbf{v}) > 0 \), then
\[
\sqrt{n}(\hat{\theta}_n^\tau - \theta)[\bar{g}^T(\hat{\mathbf{v}}_n)\Sigma(\hat{\mathbf{v}}_n)\bar{g}(\hat{\mathbf{v}}_n)]^{-1/2} \xrightarrow{L} N(0, 1).
\]

As before, note that we do not need to know the value of \( K \) to evaluate \( \bar{g}^T(\hat{\mathbf{v}}_n)\Sigma(\hat{\mathbf{v}}_n)\bar{g}(\hat{\mathbf{v}}_n) \). The proof of Theorem 2 will be based on the following.
Lemma 2. For $p \in [0, 1]$ and $n \in \mathbb{N}$ let

$$g_n(p) = p \sum_{v=1}^{\lfloor n(1-p)\rfloor+1} w_v \prod_{j=1}^{v} \left(1 - \frac{np - 1}{n - j}\right)$$

and let

$$g(p) = p \sum_{v=1}^{\infty} w_v (1 - p)^v.$$

1. If $0 < c < d < 1$ then

$$\lim_{n \to \infty} \sup_{p \in [c,d]} \sqrt{n}|g_n(p) - g(p)| = 0.$$  

2. Let $p_n \in [0, 1]$ such that $np_n \in \{0, 1, 2, \ldots, n\}$ then

$$\sqrt{n}|g_n(p_n) - g(p_n)| \leq M \sqrt{n}(n+2) \leq 2Mn^{3/2}.$$

Proof. Note that

$$\sqrt{n}|g_n(p) - g(p)| \leq M\sqrt{np} \sum_{v=1}^{\lfloor n(1-p)\rfloor+1} \prod_{j=1}^{v} \left(1 - \frac{np - 1}{n - j}\right) - (1 - p)^v$$

$$+ M\sqrt{np} \sum_{v=\lfloor n(1-p)\rfloor+2}^{\infty} (1 - p)^v =: M(\Delta_1 + \Delta_2).$$

We begin by showing Part 1. Throughout the proof of this part, we assume that $n > 2/c$; this ensures that $n(1 - p) + 1 < n - 1$ for all $p \in [c,d]$. Fix $v \in \mathbb{N}$ such that $v \leq n(1 - p) + 1$. Note that

$$\prod_{j=1}^{v} \left(1 - \frac{np - 1}{n - j}\right) = \prod_{j=0}^{v-1} \left(1 - \frac{p - (j/n)}{1 - (j+1)/n}\right)$$

and thus

$$\left|\prod_{j=1}^{v} \left(1 - \frac{np - 1}{n - j}\right) - (1 - p)^v\right| = (1 - p)^v \left|\prod_{j=0}^{v-1} \left(1 - \frac{(j/(n(1-p)))}{1 - (j+1)/n}\right) - 1\right|$$

$$\leq (1 - p)^v \frac{v^2}{n - v},$$

where the inequality follows by the proof of Part 1 of Lemma 2 in Zhang (2013). Let $V_n = \lfloor n^{1/8} \rfloor$. For large enough $n$, $V_n < \lfloor n(1 - d) + 1 \rfloor$. For such $n$, we have

$$\Delta_1 \leq \sqrt{np} \sum_{v=1}^{\lfloor n(1-p)\rfloor+1} (1 - p)^{v-1} \frac{v^2}{n - v}$$

$$= \sqrt{np} \sum_{v=1}^{V_n} (1 - p)^{v-1} \frac{v^2}{n - v} + \sqrt{np} \sum_{v=V_n+1}^{\lfloor n(1-p)\rfloor} (1 - p)^{v-1} \frac{v^2}{n - v} =: \Delta_{11} + \Delta_{12}.$$
We have

$$
\Delta_{11} \leq \sqrt{np} \frac{V_n^2}{n - V_n} \sum_{i=1}^{\infty} (1 - p)^{i-1} = \sqrt{n} \frac{V_n^2}{n - V_n} \leq \frac{n^{3/4}}{n - n^{1/8}} \to 0,
$$

$$
\Delta_{12} \leq \sqrt{np} (1 - p)^{V_n} \frac{(n(1 - p) + 1)^2}{n - (n(1 - p) + 1)} \sum_{i=1}^{\infty} (1 - p)^{i-1}
= \sqrt{n} (1 - p)^{V_n} \frac{[n(1 - p) + 1]^2}{np - 1} \leq \sqrt{n} (1 - c)^{\lfloor n/\Delta_1 \rfloor} \frac{[n(1 - c) + 1]^2}{nc - 1} \to 0,
$$

and

$$
\Delta_2 \leq \sqrt{np} (1 - p)^{\lfloor n(1 - p) + 2 \rfloor} \sum_{i=0}^{\infty} (1 - p)^{i} \leq \sqrt{n} (1 - c)^{\lfloor n(1 - d) + 2 \rfloor} \to 0.
$$

Now to show Part 2. Note that $p_n \neq 0$ implies $p_n \geq 1/n$, which means that $(np_n - 1)/(n - j) \in [0, 1]$ where $j \leq n(1 - p_n) + 1$. Thus, either $p_n = 0$ or $\prod_{j=1}^{V_n} (1 - (np_n - 1)/(n - j)) \in [0, 1]$ when $v \leq n(1 - p_n) + 1$. This implies that

$$
\Delta_1 = \sqrt{np_n} \sum_{v=1}^{\lfloor n(1 - p_n) + 1 \rfloor} \left| \prod_{j=1}^{V_n} \left( 1 - \frac{np_n - 1}{n - j} \right) - (1 - p_n)^V \right| 1_{p_n \neq 0} \leq \sqrt{n}(n + 1).
$$

and

$$
\Delta_2 \leq \sqrt{np_n} \sum_{i=0}^{\infty} (1 - p_n)^i 1_{p_n \neq 0} \leq \sqrt{n}.
$$

This completes the proof. ■

Proof of Theorem 2 Since $\sqrt{n}(\hat{\theta}_{n,k}^z - \theta) = \sqrt{n}(\hat{\theta}_{n,k}^z - \hat{\theta}_n) + \sqrt{n}(\hat{\theta}_n - \theta)$, by Proposition 1 and Slutsky’s theorem it suffices to show that $\sqrt{n}(\hat{\theta}_{n,k}^z - \hat{\theta}_n) \overset{p}{\to} 0$. Furthermore, by Slutsky’s theorem it suffices to show that $\sqrt{n}(\hat{\theta}_{n,k}^z - \hat{\theta}_{n,k}) \overset{p}{\to} 0$, where $\hat{\theta}_{n,k}^z$ is given by Equation (7) and

$$
\hat{\theta}_{n,k} = \hat{p}_k \sum_{v=1}^{\infty} w_v (1 - \hat{p}_n)^v.
$$

We can write

$$
\sqrt{n}(\hat{\theta}_{n,k}^z - \hat{\theta}_{n,k}) = \sqrt{n}(\hat{\theta}_{n,k}^z - \hat{\theta}_{n,k}) 1_{\hat{p}_k \leq \hat{p}_k/2} + \sqrt{n}(\hat{\theta}_{n,k}^z - \hat{\theta}_{n,k}) 1_{\hat{p}_k \geq (1 + \hat{p}_k)/2}
+ \sqrt{n}(\hat{\theta}_{n,k}^z - \hat{\theta}_{n,k}) 1_{\hat{p}_k/2 < \hat{p}_k < (\hat{p}_k + 1)/2} =: A_1 + A_2.
$$

By Part 2 of Lemma 2, it follows that

$$
E|A_1| \leq 2Mn^{3/2} |P(\hat{p}_k > (1 + \hat{p}_k)/2) + P(\hat{p}_k \leq \hat{p}_k/2)|
\leq 2Mn^{3/2} |P(|\hat{p}_k - p_k| \geq (1 - \hat{p}_k)/2) + P(|\hat{p}_k - p_k| \geq \hat{p}_k/2)|
\leq 4Mn^{3/2} [e^{-n(1 - p_k)^2} + e^{-np_k^2/2}] \to 0,
$$

where the third line follows by Hoeffding’s inequality, see Hoeffding (1963). Thus it follows that $A_1 \overset{p}{\to} 0$. By Part 1 of Lemma 2, it follows that $A_2 \overset{p}{\to} 0$. ■
5. Estimation of Rényi’s entropy

The only diversity indices that we have discussed that do not belong to the class of linear diversity indices are Rényi’s entropy and Hill’s diversity number. However, both are transformation of Rényi’s equivalent entropy, $h_r$. In this section we extend Theorem 2 to Rényi’s entropy. We can use a similar approach to extend it to Hill’s diversity number.

Fix $r > 0$ such that $r \neq 1$ and let $\psi_r(t) = (1 - r)^{-1} \ln t$. Note that

$$H_r = \frac{\ln h_r}{1 - r} = \psi_r(h_r),$$

where $H_r$ is Rényi’s entropy and $h_r$ is Rényi equivalent entropy. Let

$$\hat{h}_r^x(n) = 1 + \sum_{k=1}^{K} p_k \sum_{v=1}^{n-x_k} w_v \prod_{j=1}^{v} \left(1 - \frac{x_k}{n - j}\right),$$

where for $v \geq 1$

$$w_v = \prod_{i=1}^{v} \left(\frac{i - r}{i}\right).$$

This is the estimator of $h_r$ given by Equation (6). Let

$$\hat{H}_r^x(n) = \frac{\ln \hat{h}_r^x(n)}{1 - r} = \psi_r(\hat{h}_r^x(n)).$$

Since $\psi_r'(t) = t^{-1}(1 - r)^{-1}$ the delta method together with Theorem 2, Remark 4, and the fact that $\psi_r'(h_r) > 0$ implies the following.

**Theorem 3** Provided that there exists a $k \in \{1, 2, \ldots, K\}$ with $p_k \neq 1/K$

$$\sqrt{n}(\hat{H}_r^x(n) - H_r)h_r(1 - r)[g^r(\bar{v}) \Sigma(\bar{v}) g(\bar{v})]^{-1/2} \xrightarrow{L} N(0, 1).$$

We note that asymptotic normality for other estimators of Rényi’s entropy are derived in Leonenko and Seleznjev (2010) and Kälberg, Leonenko, and Seleznjev (2012). By arguments similar to the proof of Corollary 1, we get the following.

**Corollary 3** Provided that there exists a $k \in \{1, 2, \ldots, K\}$ with $p_k \neq 1/K$

$$\sqrt{n}(\hat{H}_r^x(n) - \theta)\hat{h}_r^x(n)(1 - r)[\bar{g}^r(\hat{v}_n) \Sigma(\hat{v}_n) \bar{g}(\hat{v}_n)]^{-1/2} \xrightarrow{L} N(0, 1).$$

6. Discussion

In this paper we showed that every diversity index satisfying axioms $A_{01}$ and $A_{02}$ is a function of the entropic basis. Furthermore, for the important class of linear diversity indices, we developed consistent estimators with exponentially decaying biases and asymptotic normality. We then showed how to extend these estimators to diversity indices that are not linear but only equivalent to linear ones.

Our results are very general encompassing an infinite number of possible indices. While we can present our theoretical results in this generality, it is not possible to verify the results through
simulation without focusing of only a few special cases. For several important cases, this has already been done. Simulations and data analysis verifying the performance of our estimators for the case of Shannon’s entropy are given in Zhang (2012) and Zhang and Grabchak (2013), and for the case of generalised Simpson’s indices they are given in Zhang and Zhou (2010).

While the form of our estimators and their standard deviations may appear complicated, for many important diversity indices they are implemented in the EntropyEstimation package (Cao and Grabchak 2015) for the statistical software R. In particular, methods for the estimation of Shannon’s entropy, Rényi’s entropy, Tsallis entropy, Hill’s diversity number, and the entropic basis are given there. The estimators of the entropic basis are particularly useful since estimators of all other linear diversity indices can be built from these.

To conclude, we mention that, while our discussion of diversity indices is in the context of ecological applications they are commonly used in many other situations. For applications to statistics, machine learning, and bioinformatics, see Kälberg et al. (2012). For applications to authorship attribution studies, see Grabchak, Zhang, and Zhang (2013), Grabchak, Cao, and Zhang (2015), and the references therein. For applications to information theory, see Cover and Thomas (2006). For these and other application areas our estimators can be used and will have all of the properties given here.

Acknowledgments

The authors wish to thank Dr Eric Marcon for correcting a mistake in the formula for $h_r^x(n)$ and the two anonymous referees whose comments led to improvements in the presentation of this paper.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Cao, L., and Grabchak, M. (2015), ‘Entropy Estimation: Estimation of Entropy and Related Quantities’, Ver 1.2, http://cran.r-project.org/package = EntropyEstimation.

Cover, T.M., and Thomas, J.A. (2006), Elements of Information Theory (2nd ed.), Hoboken, NJ: John Wiley and Sons.

Emlen, J.M. (1973), Ecology: An Evolutionary Approach, Reading, MA: Addison-Wesley Publishing Co.

Gini, C. (1912), ‘Variabilità e mutabilità’, reprinted in Memorie di metodologica statistica, ed. E. Pizetti E and T. Salvemini, Rome: Libreria Eredi Virgilio Veschi (1955).

Grabchak, M., Cao, L., and Zhang, Z. (2015), ‘Authorship Attribution Using Diversity Profiles’, Journal of Quantitative Linguistics, Submitted.

Grabchak, M., Zhang, Z., and Zhang, D.T. (2013), ‘Authorship Attribution Using Entropy’, Journal of Quantitative Linguistics, 20(4):301–313.

Heip, C.H.R., Herman, P.M.J., and Soetaert, K. (1998), ‘Indices of Diversity and Evenness’, Océanis, 24(4):61–87.

Hill, M.O. (1973), ‘Diversity and Evenness: A Unifying Notation and Its Consequences’, Ecology, 54(2):427–431.

Hoeffding, W. (1963), ‘Probability Inequalities for Sums of Bounded Random Variables’, Journal of the American Statistical Association, 58(301):13–30.

Kälberg, D., Leonenko, N., and Seleznjev, O. (2012), ‘Statistical inference for Rényi entropy functionals’, in Conceptual Modelling and Its Theoretical Foundations, eds. A. Düsterhöft, M. Klettke, and K. Schewe, Heidelberg: Springer, pp. 36–51.

Leonenko, N., and Seleznjev, O. (2010), ‘Statistical Inference for the $\epsilon$-entropy and the Quadratic Rényi Entropy’, Journal of Multivariate Analysis, 101(9):1981–1994.

Magurran, A.E. (1988), Ecological Diversity and its Measurement, Princeton, NJ: Princeton University Press.

Marcon, E. (2015), Mesures de la Biodiversité. Available at http://www.ecofog.gf/spip.php?article427.

Paninski, L. (2003), ‘Estimation of Entropy and Mutual Information’, Neural Computation, 15(6):1191–1253.

Peet, R.K. (1974), ‘The Measurements of Species Diversity’, Annual Review of Ecology and Systematics, 5, 285–307.

Rényi, A. (1961), ‘On Measures of Entropy and Information’, Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1, ed. J. Neyman, Berkeley, CA: University of California Press, pp. 547–561.
Shannon, C.E. (1948), ‘A Mathematical Theory of Communication’, *The Bell System Technical Journal*, 27, 379–423, 623–656.

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

Tanabe, K., and Sagae, M (1992), ‘An Exact Cholesky Decomposition and the Generalized Inverse of the Variance–Covariance Matrix of the Multinomial Distribution, with Applications’, *Journal of the Royal Statistical Society Series B (Methodological)*, 54(1):211–219.

Tsallis, C (1988), ‘Possible Generalization of Boltzmann–Gibbs Statistics’, *Journal of Statistical Physics*, 52(1–2): 479–487.

Zhang, Z (2012), ‘Entropy Estimation in Turing’s Perspective’, *Neural Computation*, 24(5):1368–1389.

Zhang, Z (2013), ‘Asymptotic Normality of an Entropy Estimator with Exponentially Decaying Bias’, *IEEE Transactions on Information Theory*, 59(1):504–508.

Zhang, Z., and Grabchak, M (2013), ‘Bias Adjustment for a Nonparametric Entropy Estimator’, *Entropy*, 15(6): 1999–2011.

Zhang, Z., and Zhou, J (2010), ‘Re-parameterization of Multinomial Distribution and Diversity Indices’, *Journal of Statistical Planning and Inference*, 140(7):1731–1738.