Annual Research & Review in Biology

37(4): 30-46, 2022; Article no.ARRB.86099
ISSN: 2347-565X, NLM ID: 101632869

A Universal Key to Rationally Select Which, Among Nonparametric Species Richness Estimators, Performs Best According to Each Particular Incomplete Sampling

Jean Béguinot [a*]

[a] *Université Bourgogne, 6, Boulevard Gabriel, 21000 Dijon, France.*

Author's contribution

The sole author designed, analyzed, interpreted and prepared the manuscript.

Article Information

DOI: 10.9734/ARRB/2022/v37i430500

Open Peer Review History:
This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: https://www.sdiarticle5.com/review-history/86099

Original Research Article

Received 06 February 2022
Accepted 16 April 2022
Published 21 April 2022

ABSTRACT

Since most samplings of local species communities are bound to remain substantially incomplete for practical reasons, a wide variety of nonparametric estimators of the number of unrecorded species have been proposed over the past fifty years. Unfortunately, the distinct formulations of each of these estimators naturally lead to substantially divergent estimates. The will to try to select, in each case, the estimator expected to be the more accurate has long been carried out only on a purely empirical, even arbitrary, basis (as is evident from the extensive consultation of much of the past literature on estimating species richness of incompletely sampled communities). So that extrapolating the true species richness of a community from its incomplete survey has long remained quite unsatisfactory. Indeed, the definition of a truly rational procedure for selecting the most accurate (least-biased) estimator actually requires a solidly established theoretical framework, involving to conform, as best as possible, to the general mathematical characteristics of the Species Accumulation Function. Accordingly, unveiling, first of all, these mathematical characteristics of the Species Accumulation Function was a decisive step forward in this perspective. Thereby making it now possible to propose an objective key to rationally select the one, within the series of various estimators, which, depending on each particular sampling, happens to be the least biased in this particular case, thus providing the most accurate estimate of the number of still unrecorded species. And, consequently, making it possible, now, to deliver the best estimate of the true species richness of a local community, despite its being incompletely surveyed.

**Corresponding author: E-mail: jean-beguinot@orange.fr;*

Keywords: Partial survey; species community; numerical extrapolation; species accumulation equation; Jackknife estimator, Chao estimator

1. INTRODUCTION

Among the most commonly cited descriptors of species communities in the wild, the *species richness* is usually considered as having the "greatest ecological significance" (with the degree of *unevenness* of species abundances coming immediately second). Thus, as already emphasized by BROSE *et al.* [1], "beyond the exhaustive list of their identities, the estimated total number of species in a community or in a given area is, by itself, critical to the development of evolutionary and ecological theories".

Yet, unfortunately, achieving (sub-) exhaustive samplings usually reveals quasi-impossible in practice for most communities in the wild. Community samplings are, thus, often doomed to remain *more or less incomplete*. And this is all the more so when dealing with species-rich communities having rather unevenly distributed species abundances. This is particularly the case for either floras or invertebrate faunas, especially (while not only) under tropical climates.

Hence, the strong incentive to develop *numerical extrapolation* procedures, intended to estimate the number of species that have escaped sampling. Thereby allowing a reliable estimation of the "true" total species richness of communities, despite their being incompletely sampled.

Early attempts in this perspective have led, during the second half of the last century, to the propositions of a series of estimators, formulated as various functions of the numbers f_1, f_2, …f_x of singletons, doubletons, … , x-tons (i.e., species encountered once, twice, …, x-times during sampling). The most often mentioned estimators being the series of 'Jackknife' estimators (JK-1 = f_1, JK-2 = $2f_1 – f_2$, JK-3 = $3f_1 – 3f_2 + f_3$, etc…) and the 'Chao' estimator (Chao1 = $f_1^2/(2.f_2)$) [2].

Now, all these estimators, each of them being formulated in such different ways, can only deliver *substantially distinct estimates* of the number of unrecorded species, for a *same* given sample! Which, naturally, led to the question of how to reliably choose which one of these different estimators might be more appropriate and the most accurate, with respect to the *particular* sample under consideration. Now, in the obvious impossibility of being able to answer properly this question on a rational basis, it has been witnessed, in the literature, a veritable anthology of empirical or even arbitrary proposals, each of them as unsatisfactory as the other: see, for example a critical review in [1]. This, in the end, had regrettably contributed to cast *much doubt on the reliability* of numerical extrapolation from incomplete samples, in order to estimate the true species richness of incompletely sampled communities.

This very unsatisfactory situation has finally led, at the beginning of this century, to the hypothesis that, among all these different estimators, it may be likely that *only one* of them could, *in turn*, prove to be the most appropriate. More specifically, the estimator to be preferred *being, in each case, dependent upon the particular sampling* under consideration. In this new perspective, no estimator could rationally claim to be universally – or, at least, usually – the most appropriate. Hence the necessity to select, *in each case* (i.e. for each particular sampling), *which particular estimator*, from the set of available estimators could really be considered the most appropriate, i.e. the least biased one. With this selection being based upon a strictly *rational* procedure. Indeed, this point of view turned out to be correct and, then, prompted the research and development of such a kind of more or less rigorous selection procedure.

In their seminal paper, BROSE *et al.* [1] deliberately comply with this approach. In particular, they suggest that each of the estimators, within the Jackknife series, could, in turn, be optimal according to the (yet still unknown) degree of completeness of the sampling under consideration. Specifically, these authors argued that the lower the completeness of the sampling, the higher should be the order of the Jackknife estimator to be selected. However, this first attempt, as meritorious as it was, could not lead, yet, to a satisfactory solution in practice (see Discussion section below). And this, in particular, because of the circularity of the procedure which implies that completeness be involved both as a means of selection and then as the result of this selection. Yet, the paper by BROSE *et al.* [1] finally rightly highlighted the possible avenue toward a future, *truly rational* procedure for selecting the desired "optimal nonparametric estimator".

To go further in this direction, a decisive element was still missing to the preceding attempt – namely unveiling the *relevant mathematical relationship which universally constrains the general expression of the so-called "Species Accumulation Function"*. And this is required because the extrapolation of the Species Accumulation Curve (S.A.C.) has the potential to forecast, all-along progressive sampling, the continuous updating of what could be the species richness of the community under consideration. So that a *reliable* estimator of the number of still unrecorded species should correspond to – and thus *comply with* – the (numerical) extrapolation of this S.A.C., continued beyond the currently achieved partial sampling.

Deriving the mathematical relationship universally constraining the Species Accumulation Curve is, thus, intended to play a *decisive role* in enabling the development of the procedure for *rationally selecting* (according to each particular sampling) the particular type of estimator able to deliver the *least-biased estimate* of the number of still unrecorded species. The derivation of this mathematical relationship had been carried out recently [3-5], thus finally opening up the perspective for a *rational estimation of the true species richness* of communities, despite having to rely only upon partial samplings.

Based upon this previously established mathematical relationship [3-5], I describe, hereafter, the procedure allowing to *rationally select*, in each case, which estimator turns out to be the *least-biased* one, among the most commonly referenced nonparametric estimators.

2. METHODS

The so-called Species Accumulation Curve (S.A.C.) accounts for the progressive increase in the number of recorded species along the progressive sampling of a community of species.

Clearly, the shape of the S.A.C. is, in every detail, entirely dependent on the specific distribution of species abundances within the sampled community. Accordingly, there are as many different shapes – and kinds of mathematical expressions – for the S.A.C.s than there are different possibilities of species abundance distributions within species communities. That is, a virtual infinity. This explains that no *general* mathematical expression has ever been derived for the S.A.C.s, at least on a rational basis. Only empirically designed models have been proposed, as pure approximations [6,7], thus irrelevant to our purpose. However, the indefinitely various mathematical shapes that the S.A.C.s could potentially take are yet, in no way, arbitrary. In fact, all of these various mathematical shapes are expected to comply with a *universal*, specific mathematical constraint, *inherent in the very nature* of the process of incremental discovery of new species, during the progress of on-going sampling.

It turns out that this mathematical constraint, framing the virtual infinity of expressions that the S.A.C.s can take, applies to *the series of derivatives of increasing order* of the S.A.C.. With, more specifically, the derivative of order x being related to the observed number, f_x, of x-tons (species which are recorded x-times in the on-going sampling). The existence and formulation of this mathematical relationship, universally constraining the expressions of the S.A.C.s, was demonstrated first in 2014, as reported in reference [3], see also [8]:

$$\partial^x R(N)/\partial N^x = (-1)^{x-1} f_x(N)/C_{N,x} \qquad (1)$$

with:

* N as the sample size, in term of the number encountered individuals,
* R(N) as the number of currently recorded species – namely the "Species Accumulation Function",
* $f_x(N)$ as the number of x-tons,
* $C_{N,x} = N!/x!/(N-x)!$ as the number of combinations of x items among N.

Leaving aside the very beginning of sampling (of no practical relevance here), the sampling-size N rapidly widely exceeds the numbers x of practical concern, so that, in practice, the preceding equation simplifies as:

$$\partial^x R(N)/\partial N^x = (-1)^{x-1}(x!/N^x).f_x(N) \qquad (2)$$

Specifically, these relations (either (1) or (2)) have *general relevance* because their derivation – and thus their validity – *does not require any specific assumption* relative to the particular shape of the distribution of species abundances in the sampled assemblage of species.

Accordingly, the above relations actually constrain the indefinitely diverse theoretical expressions of *all possible kinds* of Species Accumulation Curves.

In addition, it is to be noted that a second, independently established, demonstration of this relation was provided later [4,5] (a summary of these two alternative demonstrations is provided in Appendix). Finally, a third, again independent, demonstration was published recently by LI & LI [9], based upon the specific properties of the so-called Bernstein functions. The coexistence of these three independent demonstrations clearly underlines the *robustness* of this relation, universally constraining the expressions of the S.A.C.s in *whole generality* and, thereby, warrants the reliability of using relations (1) or (2) for practical purposes.

Now, related to our concern of establishing an objective procedure to rationally select the best type of nonparametric estimator, two additional relationships, directly stem from equation (2), are examined below.

2.1 Derivation of the Expression of the First Derivative of the Number of x-tons, $f_x(N)$

It comes from equation (2) (as already shown in references [10, 11]):

$$f_x(N) = (-1)^{x-1}(N^x/x!)\,[\partial^x R(N)/\partial N^x] \qquad (3)$$

The derivation of equation (3), with respect to sample size N, then gives:

$$\partial f_x(N)/\partial N = (-1)^{x-1}/x!\ \{x.\ N^{x-1}.[\partial^x R(N)/\partial N^x] + N^x.[\partial^{x+1} R(N)/\partial N^{x+1}]\}$$

Applying successively equation (2) to the expressions of $[\partial^x R(N)/\partial N^x]$ and of $[\partial^{x+1} R(N)/\partial N^{x+1}]$ finally leads to:

$$\partial f_x(N)/\partial N = [\,x.f_x(N) - (x+1).f_{x+1}(N)]/N \qquad (4)$$

Equation (4) thus provides the expression of the first derivative of the number $f_x(N)$ at any given sample-size N, in terms of the values taken by $f_x(N)$ and $f_{x+1}(N)$, at sampling-size N.

2.2 Derivation of the Expression of the First Derivative of the Number of still Unrecorded Species Δ(N)

Let Δ(N) be the number of unrecorded species (i.e., species having still escape recording by the on-going sampling of a community). Let S_t be the (unknown) true species richness of the sampled community; then $\Delta(N) = S_t - R(N)$. Accordingly, from equation (2), the first derivative of Δ(N) satisfies:

$$\partial\Delta(N)/\partial N = -f_1(N)/N \qquad (5)$$

3. PROCEDURE OF SELECTION OF THE MORE ACCURATE TYPE OF ESTIMATOR OF THE NUMBER OF UNRECORDED SPECIES

The ideal goal of a nonparametric estimator E(N) of the number Δ(N) of still unrecorded species is, of course, to comply, as closely as possible, to Δ(N). In particular, by decreasing with sample size N at the same rate as Δ(N) decreases. Thus, an ideal goal would be:

$$\partial E(N)/\partial N = \partial\Delta(N)/\partial N = -f_1(N)/N \qquad (6)$$

Now, if this ideal goal for E(N) to consistently match Δ(N) cannot be strictly achieved (as is likely), it is at least desirable, for the careful sake of conservatism, that the estimation E(N) be a slight *underestimate* of Δ(N), *rather than* an overestimate. For this purpose, the rate of decrease of E(N) with N – for lack of being able to consistently match the rate of decrease of Δ(N) itself – should be *slightly faster* than this decrease of Δ(N) (rather than slightly slower). So that, in practice, the criterium of selection among the available kinds of estimators should rely on the absolute rate of decrease $|\partial E(N)/\partial N|$ of E(N) with N. With this rate of decrease, $|\partial E(N)/\partial N|$, being thus required to be either *equal* - or if not - *somewhat higher* than is the absolute rate of decrease, $|\partial\Delta(N)/\partial N|$, of Δ(N):

$$|\partial E(N)/\partial N| \geq |\partial\Delta(N)/\partial N|$$

that is (since $\partial\Delta(N)/\partial N$ is, in essence, negative):

$$\partial E(N)/\partial N \leq \partial\Delta(N)/\partial N$$

Then:

$$\partial E(N)/\partial N \leq -f_1(N)/N \qquad (7)$$

3.1 Practical Key of Selection among the Top Five Jackknife Estimators (JK-1 to JK-5)

As already underlined, I shall focus upon the most often mentioned nonparametric estimators, namely the series of 'Jackknife' estimators of increasing orders and the 'Chao' estimator.

Let consider, first, the series of 'Jackknife' estimators of increasing orders. That is: JK-1 = f_1, JK-2 = $2f_1 - f_2$, JK-3 = $3f_1 - 3f_2 + f_3$, … and, more generally, at order 'm' (see reference [5]):

$$JK\text{-}m = \Sigma_{x=1 \text{ to } m} [(-1)^{(x-1)}.C_{(m, x)}.f_x] \qquad (8)$$

where $\Sigma_{x=1 \text{ to } m}$ stands for the summation from x = 1 to x = m and $C_{(m, x)} = m!/x!/(m–x)!$ is the number of combinations of x objects among m.

According to equation (7), it then follows that if the Jackknife estimator at order m, (JK-m), is to be selected, then the first derivative ∂(JK-m)/∂N of 'JK-m' should satisfy:

$$\partial(JK\text{-}m)/\partial N \leq -f_1(N)/N \qquad (9)$$

In particular:

(i) for **Jackknife at order 1**, i.e. **JK-1** = f_1:

$$\partial(JK\text{-}1)/\partial N = \partial f_1(N)/\partial N$$

and from equation (4), it comes:

$$\partial(JK\text{-}1)/\partial N = \partial f_1(N)/\partial N = [f_1(N) - 2f_2(N)]/N$$

Then, from equation (9), it follows:

$$[f_1(N) - 2f_2(N)]/N \leq -f_1(N)/N$$

that is:

$$f_1(N) \leq f_2(N) \qquad (10)$$

(ii) for **Jackknife at order 2**, i.e. **JK-2** = $2f_1 - f_2$:

$$\partial(JK\text{-}2)/\partial N = 2.\partial f_1(N)/\partial N - \partial f_2(N)/\partial N$$

and from equation (4), it comes:

$$\partial(JK\text{-}2)/\partial N = 2.[f_1(N) - 2f_2(N)]/N - [2.f_2(N) - 3f_3(N)]/N$$

Then, from equation (9), it follows:

$$2.[f_1(N) - 2f_2(N)]/N - [2.f_2(N) - 3f_3(N)]/N \leq -f_1(N)/N$$

that is:

$$f_1(N) \leq 2.f_2(N) - f_3(N) \qquad (11)$$

(iii) for **Jackknife at order 3**, i.e. **JK-3** = $3f_1 - 3f_2 + f_3$:

$$\partial(JK\text{-}3)/\partial N = 3.\partial f_1(N)/\partial N - 3.\partial f_2(N)/\partial N + \partial f_3(N)/\partial N$$

and from equation (4), it comes:

$$\partial(JK\text{-}3)/\partial N = 3.[f_1(N) - 2f_2(N)]/N - 3.[2.f_2(N) - 3f_3(N)]/N + [3.f_3(N) - 4f_4(N)]/N$$

Then, from equation (9), it follows:

$$3.[f_1(N) - 2f_2(N)]/N - 3.[2.f_2(N) - 3f_3(N)]/N + [3.f_3(N) - 4f_4(N)]/N \leq -f_1(N)/N$$

that is:

$$f_1(N) \leq 3.f_2(N) - 3.f_3(N) + f_4(N) \qquad (12)$$

(iv) for **Jackknife at order 4**, i.e. **JK-4** = $4f_1 - 6f_2 + 4f_3 - f_4$:

$$\partial(JK\text{-}4)/\partial N = 4.\partial f_1(N)/\partial N - 6.\partial f_2(N)/\partial N + 4\partial f_3(N)/\partial N - \partial f_4(N)/\partial N$$

and, similarly, it comes finally:

$$f_1(N) \leq 4.f_2(N) - 6.f_3(N) + 4.f_4(N) - f_5(N) \qquad (13)$$

(v) more generally, for **Jackknife at order m**, i.e. **JK-m** ($= \Sigma_{x=1 \text{ to } m} [(-1)^{(x-1)}.C_{(m, x)}.f_x]$):

$$\partial(JK\text{-}m)/\partial N = \Sigma_{x=1 \text{ to } m} [(-1)^{(x-1)}.C_{(m, x)}.(\partial f_x(N)/\partial N)]$$

Then, applying similarly equation (4) and equation (9) successively, it comes finally for JK-m:

$$f_1(N) \leq \Sigma_{x=2 \text{ to } m} [(-1)^x.(C_{(m+1, x)} - C_{(m, x)}).f_x(N)] + (-1)^{m+1}.f_{m+1}(N) \qquad (14)$$

The five inequalities (10) to (14) thus define the *respective domains of selection* of Jackknife estimators JK-1 to JK-5, allowing each of them to be the one offering the least-biased estimation of the number of still unrecorded species, depending on the particular sampling.

Let now summarize, combining the series of inequalities above.

It comes the following *key of selection* for the top five Jackknife estimators.

***Select preferentially*:**

JK-1 (= f_1) → when $\mathbf{f_1} \leq f_2$

JK-2 (= $2f_1 - f_2$) → when $f_2 \leq \mathbf{f_1} \leq 2f_2 - f_3$

JK-3 (= $3f_1 - 3f_2 + f_3$) → when $2f_2 - f_3 \leq \mathbf{f_1} \leq 3f_2 - 3f_3 + f_4$

JK-4 (= $4f_1 - 6f_2 + 4f_3 - f_4$) → when $3f_2 - 3f_3 + f_4 \leq \mathbf{f_1} \leq 4f_2 - 6f_3 + 4f_4 - f_5$

JK-5 (= $5f_1 - 10f_2 + 10f_3 - 5f_4 + f_5$) → when $\mathbf{f_1} \geq 4f_2 - 6f_3 + 4f_4 - f_5$

This key of selection highlights the conditions (in terms of the relative values of f_1 as compared to f_2, f_3, f_4, f_5, at the *right* side) that ensure the Jackknife at the corresponding order (at the *left* side) to provide the '*least-biased*' estimation of the number of still unrecorded species (i.e., species having still escape recording by the on-going sampling of a community).

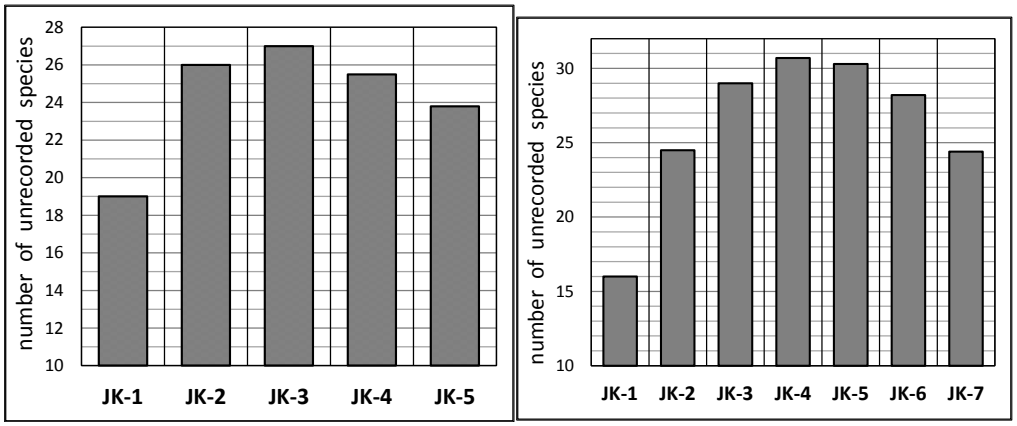
It is worth noting that – as it should be – there is *no discontinuity* in the estimates at both sides of the boundary between the respective domains of two Jackknife of successive orders. Thus, at the boundary between:

* JK-1 and JK-2 (i.e. when $f_1 = f_2$), both JK-1 and JK-2 = f_2 ;
* JK-2 and JK-3 (i.e. when $f_1 = 2f_2 - f_3$), both JK-2 and JK-3 = $3f_2 - 2f_3$;
* JK-3 and JK-4 (i.e. when $f_1 = 3f_2 - 3f_3 + f_4$), both JK-3 and JK-4 = $6f_2 - 8f_3 + 3f_4$;
* JK-4 and JK-5 (i.e. when $f_1 = 4f_2 - 6f_3 + 4f_4 - f_5$), both JK-4 and JK-5 = $10f_2 - 20f_3 + 15f_4 - 4f_5$

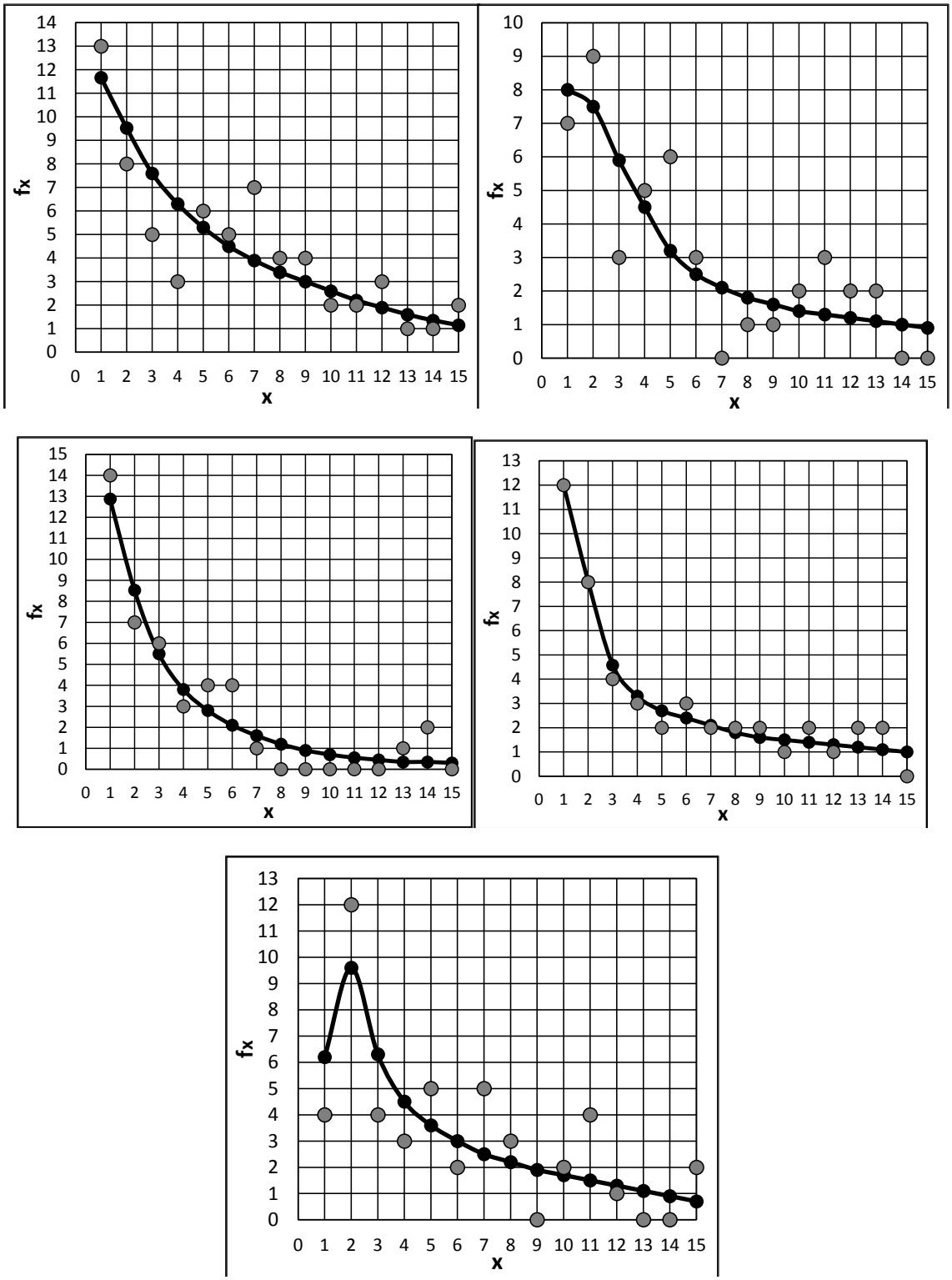
As simple illustrative examples of application, Figures 1 and 2 provide the estimated numbers of unrecorded species obtained by the series of Jackknife estimators, for two Butterfly communities sampled at the same site, in years 1987 and 2013, at Gariwang-san (Korea), as reported in reference [12].

Important Notice

The recorded values of the numbers of x-tons (f_1, f_2, f_3, f_4, f_5) are inevitably subject to a certain dispersion, due to the random draw of individuals during sampling of a community. Accordingly, the resulting risk of bias in the evaluation of the numbers f_x of x-tons should not be overlooked – since these numbers play the determinant role, regarding both the values taken by the nonparametric estimators and by the criteria of selection of the 'least-biased' estimator. Thus, to reduce this risk as far as possible, it is appropriate to regress the distribution of the recorded values of the series f_1, f_2, f_3, f_4, f_5. Practical experience suggests that a simple regression "by eye" is relevant in this respect. For illustrative purposes some examples are provided in Figures 3 to 7.



Figs. 1 and 2. The numbers of unrecorded species estimated by the series of Jackknife estimators, for two Butterfly communities sampled in 1987 (left) and 2013 (right) at Gariwang-san (Korea) [12]. The selected Jackknife estimator is JK-3 (= 27) for year 1987 while it is JK-4 (= 30.7) for year 2013



Figs. 3 to 7. The numbers f_x of species recorded x times for the partial samplings of five reef-associated fish communities investigated off Jakarta Bay [13]. Note that although x is considered from 1 to 5, the regression is continued up to x = 15, for providing a more extended view of the values taken by the f_x, which is best appropriate for the visual regression. As recorded: *grey* discs; visually regressed: *black* discs

3.2 What about the Chao Type of Nonparametric Estimators

In their seminal paper, BROSE *et al.* [1] considered that the series of Jackknife was self-sufficient to account for all the cases of more or less incomplete samplings situations, thus requiring to consider no other types of nonparametric estimators. In particular, it does not appear any need to resort to Chao-type estimator (as Chao 1 = $f_1^2/(2.f_2)$), according to these authors.

Our own theoretical approach also agrees with this statement. Moreover, the Chao type estimators regrettably suffer from a conceptual defect, intrinsically linked to their nonlinear formulation in terms of f_x. As already pointed out previously [14], for this very reason the Chao-type estimators cannot satisfy, as obviously required, the rule of additivity [14]. As a result, when using the Chao-type estimators, the estimate made on a set comprising several subsets unfortunately does not correspond to the sum of the estimates made on each of these subsets (as required – at least for a point estimator).
Thus, in accordance with the previous option of BROSE *et al.* [1], it appears that the series of Jackknife estimators (in practice, aptly limited to the set of the five first Jackknife) is sufficient by itself to offer a relevant panoply of potential estimators, among which to choose for an optimized estimation of the number of still unrecorded species.

4. DISCUSSION

As recalled in Introduction, the possibility of estimating the number of unrecorded species in a presumably incomplete survey began with a rather unsatisfactory situation, up to the end of the preceding century. Namely, the *puzzling dilemma* of having to choose, among a lot of available estimators providing *divergent estimates*. And this, without disposing, however, of any *reliable key to select rationally* the particular type of estimator expected to reliably provide the least-biased estimation. This very uncomfortable era eventually came to an end with the publication of the seminal paper by BROSE *et al.* [1], where is relevantly highlighted that *no unique, universally best* estimator can reasonably exist. Regrettably however, this publication remained too much ignored thereafter, since many authors regrettably persist in choosing, *still rather arbitrarily*, the kind of estimator which they personally consider – or even claim - as being the "best". While BROSE *et al.* put forward that, in fact, the best – *least biased* – estimator might well differ in each practical case, being particular to each given sampling. They further suggested that the sampling criterion to be considered first was the degree of completeness of the sample under consideration. A practical procedure of iterative selection thus arises from this, inviting to determine which estimator among the lot of available ones (in particular the Jackknife series) is intended to perform best, i.e. more accurately.
The *question arguably remained, however*, as to whether this advocated relation between the degree of sampling incompleteness on the one hand and the preferred order of Jackknife estimator on the other hand is:

(i) solidly confirmed from a theoretical point of view,
(ii) the only effective factor to be considered and accounted for, when trying to select the least biased estimator of the number of still unrecorded species.

Thanks to the procedure developed in this work, it now reveals possible to address and to answer (in fact by the negative) each of these two fundamental questions.

Figure 8 summarized the results from a wide series of reported case studies [7, 12, 13, 15-29] involving the numerical extrapolations of 62 incomplete samplings of various animal communities distributed worldwide, both marine and terrestrial. The numerical extrapolations were carried out using the procedure of selection of the “least-biased” estimator of the number of unrecorded species initially proposed in [5] and alternatively argued above in section 3.1. Thanks to its theoretically based establishment, this key of selection of the “lest-biased” estimator may admittedly serve as a reliable *reference* against which to compare other procedures of selection, such as – here – the one previously proposed by BROSE *et al.* [1]. The features highlighted in Figure 8 aptly allow, accordingly, to address the two questions put forward above.

First, as advocated by BROSE *et al.*, it obviously exists a trend for the order of the selected Jackknife estimator to actually increase with decreasing levels of completeness of samplings. Second, there remains, yet, *much scatter* in this relationship.

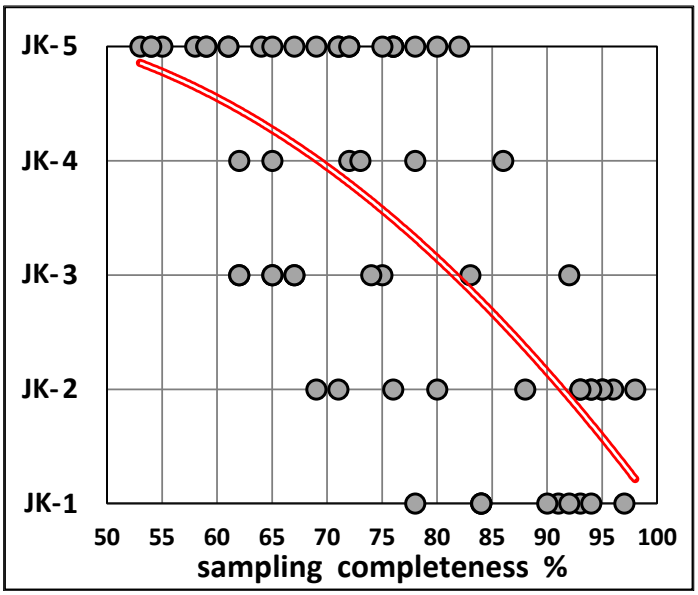


Fig. 8. The selected least-biased Jackknife estimator (among JK-1 to JK-5) of the number of unrecorded species plotted against the corresponding estimated sampling completeness, for a series of 62 samplings [7, 11, 13, 15-29]. Obviously, the order of the selected Jackknife estimator tends to increase with decreasing sampling completeness – as suggested by BROSE *et al.* [1]. *However*, the dispersion turns out to be much too wide to allow to reliably rely upon this criterion *only* to select the best estimator, as too optimistically proposed by BROSE *et al*

Thus, pointing out that *other factors*, besides the degree of sampling-incompleteness, are likely to be also involved. And, accordingly, that these other factors are to be considered, as well, in the rational selection of the least-biased estimator.

More precisely, the coefficient of determination of the correlation between the order of the selected Jackknife estimator on the one hand and the level of sampling incompleteness on the other hand, is $r^2 = 0.48$ *only*. This emphasizes that the degree of sampling incompleteness is *no more than part only* (48%) of the factors involved in determining the selected order of the actually least-biased Jackknife estimator.

This indeed is no surprise. Arguably, the particular distribution of the species abundances in the sampled community, especially the degree of unevenness of species abundance distribution (including the distribution of abundances of the still unrecorded species), is likely to also play a key role in this matter. Indeed, it turns out that the actually recorded numbers f_1, f_2, …f_x of singletons, doubletons, …, x-tons (and, thus, the Jackknife estimators, as any other nonparametric estimators) are sensitive not only to the level of sampling incompleteness, but *also* to the degree of unevenness of the species abundance distribution within the community of species under consideration. And, more generally, that the non-parametric estimators are sensitive to all the factors involved in defining the extrapolation of the Species Accumulation Curve R(N), which ultimately forecasts the number of still unrecorded species – as highlighted by the universal relationship (1) constraining this Curve.

5. CONCLUSION

Incomplete samplings are common cases in most local biodiversity surveys – especially those addressing invertebrate local communities worldwide. Thus, the estimation of the number of species still remaining unrecorded is *key* to evaluate the *true species richness* of these communities – indeed a *major descriptor* of species diversity.

However, estimating the number of species remaining unrecorded due to unavoidable incomplete samplings proves being quite a difficult matter. As proof, the difficulties encountered in this regard throughout the second-half of the 20th century. Difficulties paradoxically resulting from the somewhat "plethoric" creativity of the statisticians, delivering a multiplicity of competing – and unfortunately diverging – nonparametric estimators. The horizon in this matter then began to brighten with

the founding publication of BROSE *et al*. in year 2003. A publication which marked a decisive step ahead, although remaining still partial, due to its still insufficient theoretical foundation. Finally, it is only the unveiling of the fundamental mathematical relation constraining, in all generality, the shape of the Species Accumulation Curve which ultimately made it possible to establish the sound, *theoretical foundation* required to derive a *rational procedure of numerical extrapolation* of the Species Accumulation Curve. And, by doing so, offering the *key* to a procedure making it possible to *rationally select*, *in each case*, which among the series of available estimators actually happens to be the least biased.

ACKNOWLEDGEMENTS

I acknowledge the fruitful comments of three anonymous Reviewers.

COMPETING INTERESTS

Author has declared that no competing interests exist.

REFERENCES

1. Brose U, Martinez ND & Williams RJ Estimating species richness: sensitivity to sample coverage and insensitivity to spatial patterns. Ecology. 2003;84(9):2364-2377.
2. Gotelli NJ & Chao A Measuring and Estimating Species Richness, Species Diversity, and Biotic Similarity from Sampling Data. In: Levin SA (ed.) Encyclopedia of Biodiversity, Second Edition. 2013; 5:195-211.
3. Béguinot J. An algebraic derivation of Chao's estimator of the number of species in a community highlights the condition allowing Chao to deliver centered estimates. ISRN Ecology; 2014. article ID 847328.
DOI:10.1155/2014/847328 ; <hal-01101415>
4. Béguinot J. When reasonably stop sampling? How to estimate the gain in newly recorded species according to the degree of supplementary sampling effort. Annual Research & Review in Biology. 2015;7(5): 300-308.
DOI : 10.9734/ARRB/2015/18809 ; <hal-01228695>
5. Béguinot J. Theoretical derivation of a bias-reduced expression for the extrapolation of the Species Accumulation Curve and the associated estimation of total species richness. Advances in Research. 2016;7(3):1-16.
DOI: 10.9734/AIR/2016/26387; hal-01367803
6. Thompson GG, Withers PC, Pianka ER & Thompson SA. Assessing biodiversity with species accumulation curves; inventories of small reptiles by pit-trapping in Western Australia. Austral Ecology. 2003;28:361–383.
7. Béguinot J, Nidup T. Least-biased extrapolation of a partial inventory of butterfly fauna in Manas Range (Royal Manas National Park, Bhutan). Asian Journal of Environment & Ecology. 2017; 2(2): 1-14. doi: 10.9734/AJEE/2017/32701
8. Marcon E. Mesures de la Biodiversité. 2018; hal : cel-01205813v5.
9. Li CT & Li K-H. Species Abundance Distribution and Species Accumulation Curve: a general framework and results. arXiv. 2020;2011.07270v1 [stat.AP].
10. Béguinot J. On general mathematical constraints applying to the kinetics of species discovery during progressive sampling: consequences on the theoretical expression of the Species Accumulation Curve. Advances in Research. 2016;8(5):1-17.
DOI: 10.9734/AIR/2016/31791. <hal-01516141>.
11. Béguinot J. The variations of the numbers of species recorded 1-, 2-, … x-times (singletons, doubletons, … x-tons) with increasing sampling-size : an analytical approach using Taylor expansion. Advances in Research. 2017;10(6):1-14.
DOI:10.9734/AIR/2017/35223. <hal-01625520>
12. Béguinot J. Inter-annual variations of true species richness in a subtropical butterfly assemblage: anestimation based on least-biased extrapolations of species accumulation curves. Asian Journal of Biology. 2017;2(4):1-16.
DOI: 10.9734/AJOB/2017/33876
13. Béguinot J. Analyzing the role of increasing water pollution on species-richness, interspecific-competition and abundance-unevenness in reef-associated fish communities, off Jakarta Bay (Indonesia). International Journal of

Environment and Climate Change. 2021;11(6):18-43.
DOI: 10.9734/IJECC/2021/v11i630420.

14. Béguinot J. Basic theoretical arguments advocating Jackknife-2 as usually being the most appropriate nonparametric estimator of total species richness. Annual Research & Review in Biology. 2016;10(1):1-12.
DOI: 10.9734/ARRB/2016/25104> ; <hal-01300828>
15. Béguinot J. Least-biased estimations of true species richness of butterfly fauna in sub-urban sites around Jhansi (India) and the range of inter-annual variation of species richness. Asian Journal of Environment & Ecology. 2017;2(1): 1-12.
DOI: 10.9734/AJEE/2017/32040
16. Béguinot J. Extrapolation of total species richness from incomplete inventories: application to the Gastropod fauna associated to coral reefs in 'Mannar Gulf Biosphere Reserve', India. Asian Journal of Environment and Ecology. 2017;4(3): 1-14.
DOI: 109734/AJEE/2017/36831.
17. Béguinot J. Numerical extrapolation of the species abundance distribution unveils the true species richness and the hierarchical structuring of a partially sampled marine gastropod community in the Andaman Islands (India). Asian Journal of Environment and Ecology. 2018;6(4):1–23.
DOI: 10.9734/AJEE/2018/41293 <hal-01807454>
18. Béguinot J. The full hierarchical structuration of species abundances reliably inferred from the numerical extrapolation of still partial samplings: a case study with marine snail communities in Mannar Gulf (India). Asian Journal of Environment and Ecology. 2018;7(3):1-27.
DOI: 109734/AJEE/2018/36831.
19. Béguinot J. Analyzing the role of environmental stresses on species richness and the process of hierarchical structuring of species abundances in marine Gastropods communities at Suva (Fiji Islands). International Journal of Environment and Climate Change. 2018;8(3):200-233.
20. Béguinot J. Inferring total species richness and the exhaustive hierarchical structuring of species abundances in tropical Sea-Stars communities (Asteroidea), using numerical extrapolation of partial inventories. Asian Journal of Environment and Ecology. 2018;8 (2):1-25.
DOI: 109734/AJEE/2018/46272.
21. Béguinot J. Comparing the complete hierarchical structuration of species abundances in reef fish communities according to coral morphology, using the numerical extrapolation of only incomplete inventories. Asian Journal of Environment and Ecology. 2018;8(1):1-20.
DOI: 109734/AJEE/2018/45402.
22. Béguinot J. Influence of Coral Architecture on Species Richness and the Hierarchical Structuration of Species Abundances in Reef Fish Communities: A Case Study in the Eastern Tropical Pacific. Asian Journal of Environment & Ecology. 2018; 8(3):1-21.
Available:https:doi.org/10.9734/ajee/2018/v8i330075
23. Béguinot J. Influence of fishing activity on the total species richness and the abundance unevenness in reef fish communities: a case study in a Brazilian tropical coral complex. International Journal of Environment and Climate Change. 2019; 9(1):58-76.
24. Béguinot J. Influence of Coral complexity on Species Richness and the Hierarchical Structuration of Species Abundances in Reef Fish Communities: A Case Study in south-east Brazil. Asian Journal of Environment & Ecology. 2019;9(3):1-20.
DOI: 10.9734/AJEE/2019/v9i330098.
25. Béguinot J. Influence of environmental heterogeneity on the species composition, species richness and species abundances unevenness in reef-associated Conus communities (Neogastropoda) from Papua New-Guinea. Asian Journal of Environment & Ecology. 2019;10(3):1-21.
DOI: 10.9734/AJEE/2019/v10i330116.
26. Béguinot J. Variations in total species richness and the unevenness of species abundance distribution between two distant Conus communities (Neogastropoda): a case study in Mannar Gulf (India). Asian Journal of Environment & Ecology. 2019;9(4):1-18.
DOI: 10.9734/AJEE/2019/v9i430102.
27. Béguinot J. Inferring true species richness and complete abundance distribution in six reef-fish communities from Red-Sea, using the numerical extrapolation of incomplete samplings. Asian Journal of Environment & Ecology. 2019;11(3):1-21.

DOI: 10.9734/AJEE/2019/v11i330136.

28. Béguinot J. Progressive recovery of a marine Gastropod community following atmospheric nuclear tests in French-Polynesia: a socio-ecological interpretation. Annual Research & Review in Biology. 2021;36(1):77-110
DOI: 10.9734/ARRB/2021/v36i130335.

29. Béguinot J. Interspecific-competition strongly constrains species-richness and species-abundance evenness in a tropical marine molluscan community inhabiting Caulerpa beds, as compared to coral-reefs. Asian Journal of Environment and Ecology. 2021;14(4):26-46.
DOI: 10.9734/AJEE/2021/v14i430214

APPENDIX

A.1 Derivation of a universal mathematical framing of the Species Accumulation Function R(N): the constraining relationship between $\partial^x R_{(N)}/\partial N^x$ and $f_{x(N)}$

The shape of the theoretical Species Accumulation Curve is directly dependent upon the particular Species Abundance Distribution (the "S.A.D.") within the sampled assemblage of species. That means that beyond the common general traits shared by all Species Accumulation Curves, each particular species assemblage give rise to a specific Species Accumulation Curve with its own, unique shape, considered in detail. Now, it turns out that, in spite of this diversity of particular shapes, all the Species Accumulation Curves are, nevertheless, *constrained by a same mathematical relationship* that rules their successive derivatives (and, thereby, rules the details of the curve shape since the successive derivatives altogether define the local shape of the curve in any details). Moreover, it turns out that this general mathematical constraint relates bi-univocally each derivative at order x, $[\partial^x R_{(N)}/\partial N^x]$, to the number, $f_{x(N)}$, of species recorded x-times in the considered sample of size N. And, as the series of the $f_{x(N)}$ are obviously directly dependent upon the particular Distribution of Species Abundance within the sampled assemblage of species, it follows that this mathematical relationship between $\partial^x R_{(N)}/\partial N^x$ and $f_{x(N)}$, ultimately reflects the indirect but strict dependence of the shape of the Species Accumulation Curve upon the particular Distribution of the Species Abundances (the so called S.A.D.) within the assemblage of species under consideration. In this respect, this constraining relationship is central to the process of species accumulation during progressive sampling, and is therefore at the heart of any reasoned approach to the extrapolation of any kind of Species Accumulation Curves.

This fundamental relationship may be derived as follows.

Let consider an assemblage of species containing an unknown total number 'S' of species. Let R be the number of recorded species in a partial sampling of this assemblage comprising N individuals. Let p_i be the probability of occurrence of species 'i' in the sample This probability is assimilated to the relative *abundance* of species 'i' within this assemblage or to the relative *incidence* of species 'i' (its proportion of occurrences) within a set of sampled sites. The number Δ of missed species (unrecorded in the sample) is $\Delta = S - R$.

The estimated number Δ of those species that escape recording during sampling of the assemblage is a decreasing function $\Delta_{(N)}$ of the sample of size N, which depends on the particular distribution of species abundances p_i:

$$\Delta_{(N)} = \Sigma_i (1-p_i)^N \qquad (A1.1)$$

with Σ_i as the operation summation extended to the totality of the 'S' species 'i' in the assemblage (either *recorded* or *not*)

The expected number f_x of species recorded *x times* in the sample, is then, according to the binomial distribution:

$$f_x = [N!/X!/(N-x)!] \Sigma_i [(1-p_i)^{N-x} p_i^x] = C_{N,x} \Sigma_i (1-p_i)^{N-x} p_i^x \qquad (A1.2)$$

with $C_{N,x} = N!/X!/(N-x)!$

We shall now derive the relationship between the successive derivatives of $R_{(N)}$, the theoretical Species Accumulation Curve and the expected values for the series of 'f_x'.
According to equation (A1.2):

► $f_1 = N \Sigma_i [(1-p_i)^{N-1} p_i] = N \Sigma_i [(1-p_i)^{N-1} (1- (1-p_i))] = N \Sigma_i [(1-p_i)^{N-1}] - N \Sigma_i [(1-p_i)^{N-1}(1-p_i))] = N \Sigma_i [(1-p_i)^{N-1}] - N \Sigma_i [(1-p_i)^N]$.

Then, according to equation (A1) it comes: $f_1 = N (\Delta_{(N-1)} - \Delta_{(N)}) = - N (\Delta_{(N)} - \Delta_{(N-1)})$

$= - N (\partial \Delta_{(N)}/\partial N) = - N \Delta'_{(N)}$

where $\Delta'_{(N)}$ is the first derivative of $\Delta_{(N)}$ with respect to N. Thus:

$f_1 = - N \Delta'_{(N)}$ $(= - C_{N,1} \Delta'_{(N)})$ (A1.3)

Similarly:

► $f_2 = C_{N,2} \Sigma_i [(1-p_i)^{N-2} p_i^2]$ according to equation (A1.2)

$= C_{N,2} \Sigma_i [(1-p_i)^{N-2} (1- (1-p_i^2))] = C_{N,2} [\Sigma_i [(1-p_i)^{N-2}] - \Sigma_i [(1-p_i)^{N-2}(1- p_i^2)]]$
$= C_{N,2} [\Sigma_i [(1-p_i)^{N-2}] - \Sigma_i [(1-p_i)^{N-2}(1- p_i)(1+ p_i)]] = C_{N,2} [\Sigma_i [(1-p_i)^{N-2}] - \Sigma_i [(1-p_i)^{N-1}(1+ p_i)]]$
$= C_{N,2} [(\Delta_{(N-2)} - \Delta_{(N-1)}) - f_1/N]$ according to equations (A2.1) and (A1.2)
$= C_{N,2} [- \Delta'_{(N-1)} - f_1/N] = C_{N,2} [- \Delta'_{(N-1)} + \Delta'_{(N)}]$ since $f_1 = - N \Delta'_{(N)}$ (*cf.* equation (A1.3)).
$= C_{N,2} [(\partial \Delta'_{(N)}/\partial N)] = [N(N-1)/2] (\partial^2 \Delta_{(N)}/\partial N^2) = [N(N-1)/2] \Delta''_{(N)}$

where $\Delta''_{(N)}$ is the second derivative of $\Delta_{(N)}$ with respect to N. Thus:

$f_2 = [N(N-1)/2] \Delta''_{(N)} = C_{N,2} \Delta''_{(N)}$ (A1.4)

► $f_3 = C_{N,3} \Sigma_i [(1-p_i)^{N-3} p_i^3]$ which, by the same process, yields:

$= C_{N,3} [\Sigma_i (1-p_i)^{N-3} - \Sigma_i (1-p_i)^{N-2} - \Sigma_i [(1-p_i)^{N-2} p_i] - \Sigma_i [(1-p_i)^{N-2} p_i^2)]]$

$= C_{N,3} [(\Delta_{(N-3)} - \Delta_{(N-2)}) - f_1^*/(N-1) - 2 f_2/(N(N-1))]$ according to equations (A2.1) and (A1.2)

where f_1^* is the number of singletons that would be recorded in a sample of size (N - 1) instead of N.

According to equations (A1.3) & (A1.4):

$f_1^* = - (N-1) \Delta'_{(N-1)} = - C_{N-1,1} \Delta'_{(N-1)}$ and $f_2 = [N(N-1)/2] \Delta''_{(N)} = C_{N-1,2} \Delta''_{(N)}$ (A1.5)
where $\Delta'_{(N-1)}$ is the first derivate of $\Delta_{(N)}$ with respect to N, at point (N-1). Then,
$f_3 = C_{N,3} [(\Delta_{(N-3)} - \Delta_{(N-2)}) + \Delta'_{(N-1)} - \Delta''_{(N)}] = C_{N,3} [-\Delta'_{(N-2)} + \Delta'_{(N-1)} - \Delta''_{(N)}]$
$= C_{N,3} [\Delta''_{(N-1)} - \Delta''_{(N)}] = C_{N,3} [- \partial \Delta''_{(N)}/\partial N] = C_{N,3} [- \partial^3 \Delta_{(N)}/\partial N^3] = C_{N,3} \Delta'''_{(N)}$

where $\Delta'''_{(N)}$ is the third derivative of $\Delta_{(N)}$ with respect to N. Thus :

$f_3 = - C_{N,3} \Delta'''_{(N)}$ (A1.6)

Now, generalising for the number f_x of species recorded *x times* in the sample:

► $f_x = C_{N,x} \Sigma_i [(1-p_i)^{N-x} p_i^x]$ according to equation (A1.2),

$= C_{N,x} \Sigma_i [(1-p_i)^{N-x} (1 - (1 - p_i^x))] = C_{N,x} [\Sigma_i (1-p_i)^{N-x} - \Sigma_i [(1-p_i)^{N-x} (1 - p_i^x)]]$
$= C_{N,x} [\Sigma_i (1-p_i)^{N-x} - \Sigma_i [(1-p_i)^{N-x} (1 - p_i)(\Sigma_j p_i^j)]]$
with Σ_j as the summation from j = 0 to j = x-1. It comes:
$f_x = C_{N,x} [\Sigma_i (1-p_i)^{N-x} - \Sigma_i [(1-p_i)^{N-x+1} (\Sigma_j p_i^j)]]$
$= C_{N,x} [\Sigma_i (1-p_i)^{N-x} - \Sigma_i (1-p_i)^{N-x+1} - \Sigma_k [(\Sigma_i (1-p_i)^{N-x+1} p_i^k)]]$

with Σ_k as the summation from k = 1 to k = x-1 ; that is:

$f_x = C_{N,x} [(\Delta_{(N-x)} - \Delta_{(N-x+1)}) - \Sigma_k (f_k^*/C_{(N-x+1+k),k})]$ according to equations (A1.1) and (A1.2))

where $C_{(N-x+1+k),k} = (N-x+1+k)!/k!/(N-x+1)!$ and f_k^* is the expected number of species recorded k times during a sampling of size (N-x+1+k) (instead of size N).

The same demonstration, which yields previously the expression of f_1^* above (equation (A1.5)), applies for the f_k^* (with k up to x-1) and gives:

$$f_k^* = (-1)^k \, (C_{(N-x+1+k),\,k}) \, \Delta^{(k)}_{(N-x+1+k)} \qquad \text{(A1.7)}$$

where $\Delta^{(k)}_{(N-x+1+k)}$ is the k^{th} derivate of $\Delta_{(N)}$ with respect to N, at point (N-x+1+k). Then,

$$f_x = C_{N,\,x} \, [(\Delta_{(N-x)} - \Delta_{(N-x+1)}) - \Sigma_k \, ((-1)^k \, \Delta^{(k)}_{(N-x+1+k)})] \quad ,$$

which finally yields :

$$f_x = C_{N,\,x} \, [(-1)^x \, (\partial \Delta^{(x-1)}_{(N)} / \partial N)] = C_{N,\,x} \, [(-1)^x \, (\partial^x \Delta_{(N)} / \partial N^x)].$$ That is:

$$f_x = (-1)^x \, C_{N,\,x} \, \Delta^{(x)}_{(N)} = (-1)^x \, C_{N,\,x} \, [\partial^x \Delta_{(N)} / \partial N^x] \qquad \text{(A1.8)}$$

where $[\partial^x \Delta_{(N)} / \partial N^x]$ is the x^{th} derivative of $\Delta_{(N)}$ with respect to N, at point N.

Conversely:

$$[\partial^x \Delta_{(N)} / \partial N^x] = (-1)^x \, f_x / C_{N,\,x} \qquad \text{(A1.9)}$$

Note that, in practice, leaving aside the beginning of sampling, N rapidly increases much greater than x, so that the preceding equation simplifies as:

$$[\partial^x \Delta_{(N)} / \partial N^x] = (-1)^x \, (x!/N^x) \, f_{x(N)} \qquad \text{(A1.10)}$$

In particular:

$$[\partial \Delta_{(N)} / \partial N] = f_{1(N)}/N \qquad \text{(A1.11)}$$

$$[\partial^2 \Delta_{(N)} / \partial N^2] = 2 \, f_{2(N)}/N^2 \qquad \text{(A1.12)}$$

This relation (A1.9) has general relevance since it does not involve any specific assumption relative to either (i) the particular shape of the distribution of species abundances in the sampled assemblage of species or (ii) the particular shape of the species accumulation rate. Accordingly, this relation constrains any theoretical form of species accumulation curves. As already mentioned, the shape of the species accumulation curve is entirely defined (at any value of sample size N) by the series of the successive derivatives $[\partial^x R_{(N)} / \partial N^x]$ of the predicted number R(N) of recorded species for a sample of size N:

$$[\partial^x R_{(N)} / \partial N^x] = (-1)^{(x-1)} \, f_x / C_{N,\,x} \qquad \text{(A1.13)}$$

with $[\partial^x R_{(N)} / \partial N^x]$ as the x^{th} derivative of $R_{(N)}$ with respect to N, at point N and $C_{N,\,x} = N!/(N-x)!/x!$ (since the number of recorded species $R_{(N)}$ is equal to the total species richness S minus the expected number of missed species $\Delta_{(N)}$).

As above, equation (A1.13) simplifies in practice as:

$$\partial^x R_{(N)} / \partial N^x = (-1)^{(x-1)} \, (x!/N^x) \, f_{x(N)} \qquad \text{(A1.14)}$$

Equation (A1.13) makes quantitatively explicit the dependence of the shape of the species accumulation curve (expressed by the series of the successive derivatives $[\partial^x R_{(N)} / \partial N^x]$ of R(N)) upon the shape of the distribution of species abundances in the sampled assemblage of species.

A2 An alternative derivation of the relationship between $\partial^x R_{(N)}/\partial N^x$ and $f_{x(N)}$

Consider a sample of size N (N individuals collected) extracted from an assemblage of S species and let G_i be the group comprising those species collected i-times and $f_{i(N)}$ their number in G_i. The number of collected individuals in group G_i is thus $i.f_{i(N)}$, that is a proportion $i.f_{i(N)}/N$ of all individuals collected in the sample. Now, each newly collected individual will either belong to a new species (probability $1.f_1/N = f_1/N$) or to an already collected species (probability $1- f_1/N$), according to [8]. In the latter case, the proportion $i.f_{i(N)}/N$ of individuals within the group G_i accounts for the probability that the newly collected individual will contribute to increase by one the number of species that belong to the group G_i (that is will generate a transition [i-1 → i] under which the species to which it belongs leaves the group G_{i-1} to join the group G_i). Likewise, the probability that the newly collected individual will contribute to reduce by one the number of species that belong to the group G_i (that is will generate a transition [i → i+1] under which the species leaves the group G_i to join the group G_{i+1}) is $(i+1).f_{i+1(N)}/N$.

Accordingly, for $i \geq 1$:

$$\partial f_{i(N)}/\partial N = [i.f_{i(N)}/N - (i+1).f_{i+1(N)}/N](1 - f_1/N) \qquad (A2.0)$$

Leaving aside the very beginning of sampling, and thus considering values of sample size N substantially higher than f_1, it comes:

$$\partial f_{i(N)}/\partial N = i.f_{i(N)}/N - (i+1).f_{i+1(N)}/N \qquad (A2.1)$$

Let consider now the Species Accumulation Curve R(N), that is the number R(N) of species that have been recorded in a sample of size N. The probability that a newly collected individual belongs to a still unrecorded species corresponds to the probability of the transition [0 → 1], equal to $i.f_{i(N)}/N$ with i = 1, that is: $f_{1(N)}/N$ (as already mentioned).

Accordingly, the first derivative of the Species Accumulation Curve R(N) at point N is

$$\partial R_{(N)}/\partial N = f_{1(N)}/N \qquad (A2.2)$$

In turn, as $f_{1(N)} = N.\partial R_{(N)}/\partial N$ (from equation (A2.2)) it comes:

$$\partial f_{1(N)}/\partial N = \partial[N(\partial R_{(N)}/\partial N)]/\partial N = N(\partial^2 R_{(N)}/\partial N^2) + \partial R_{(N)}/\partial N$$

On the other hand, according to equation (A2.1):

$$\partial f_{1(N)}/\partial N = 1.f_{1(N)}/N - 2.f_{2(N)}/N = f_{1(N)}/N - 2f_{2(N)}/N,$$

and

therefore:

$$N(\partial^2 R_{(N)}/\partial N^2) + \partial R_{(N)}/\partial N = f_{1(N)}/N - 2f_{2(N)}/N$$

And as $\partial R_{(N)}/\partial N = f_{1(N)}/N$ according to equation (A2.2):

$$\partial^2 R_{(N)}/\partial N^2 = -2f_{2(N)}/N^2 \qquad (A2.3)$$

Likewise, as $f_{2(N)} = -N^2/2.(\partial^2 R_{(N)}/\partial N^2)$, it comes:

$$\partial f_{2(N)}/\partial N = \partial[-N^2/2.(\partial^2 R_{(N)}/\partial N^2)]/\partial N = -N(\partial^2 R_{(N)}/\partial N^2) - N^2/2.(\partial^3 R_{(N)}/\partial N^3)$$

As $\partial f_{2(N)}/\partial N = 2f_{2(N)}/N - 3f_{3(N)}/N$, according to equation (A2.1), it comes:

$- N(\partial^2 R_{(N)}/\partial N^2) - N^2/2.(\partial^3 R_{(N)}/\partial N^3) = 2f_{2(N)}/N - 3f_{3(N)}/N$

and as $\partial^2 R_{(N)}/\partial N^2 = - 2f_{2(N)}/N^2$, according to equation (A2.3), it comes:

$$\partial^3 R_{(N)}/\partial N^3 = + 6f_{3(N)}/N^3 \qquad \text{(A2.4)}$$

More generally:

$$\partial^x R_{(N)}/\partial N^x = (-1)^{(x-1)} (x!/N^x) f_{x(N)} \qquad \text{(A2.5)}$$

Peer-recview history:
The peer review history for this paper can be accessed here:
https://www.sdiarticle5.com/review-history/86099