



# 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

Jean Béguinot<sup>1\*</sup>

<sup>1</sup>Department of Biogéosciences, Université de Bourgogne, F 21000 – Dijon, France.

## Author's contribution

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

## Article Information

DOI: 10.9734/AIR/2017/35223

### Editor(s):

(1) Alexander Vaninsky, Mathematics Department, Hostos Community College of The City University of New York, USA.

### Reviewers:

(1) G. Y. Sheu, Chang-Jung Christian University, Taiwan.

(2) Süleyman Cengizci, Antalya Bilim University, Turkey.

(3) Kateryna Nesvit, Karazin Kharkiv National University, Ukraine.

Complete Peer review History: <http://www.sciencedomain.org/review-history/20156>

Original Research Article

Received 1<sup>st</sup> July 2017  
Accepted 18<sup>th</sup> July 2017  
Published 22<sup>nd</sup> July 2017

## ABSTRACT

The process of species accumulation, during progressive sampling, results in the regular, *monotonic* increase of the number of recorded species with sampling size. On the contrary, the numbers  $f_{1(N)}$ ,  $f_{2(N)}$ ,  $f_{3(N)}$ , ...,  $f_{x(N)}$  of those species recorded 1-, 2-, 3-, ..., x-times at sampling-size N all show *non-monotonic* variations with N. The major characteristic elements of this *non-monotonic* variations (namely: the maximum reached at  $\partial f_{x(N)}/\partial N = 0$  and the inflexion point at  $\partial^2 f_{x(N)}/\partial N^2 = 0$ ) provide interesting cues regarding the degree of advancement of sampling completeness. Such cues yet remain undetectable however along the regular, monotonic increase of the species accumulation curve itself. Although usually unrecorded, the variations of the  $f_{x(N)}$  may yet be computed and, accordingly, the associated cues above thereby made available in practice. This computation involves the Taylor expansion of the  $f_{x(N)}$ , making use of recently derived mathematical properties of the species accumulation process. *For common practice*, focus is placed upon the variations of the  $f_{x(N)}$  of *lower-orders* (i.e.  $f_{1(N)}$ ,  $f_{2(N)}$ ,  $f_{3(N)}$ ,  $f_{4(N)}$ ), which is sufficient to disclose information of particular relevance in assessing the progress of sampling towards completeness.

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

*Keywords: Species accumulation; survey; biodiversity; completeness; estimate; extrapolation; exhaustivity.*

## 1. INTRODUCTION

The progressive sampling of an assemblage of objects (and, in particular, an assemblage of species) is accounted for numerically by the so-called accumulation (or discovery) curve. The “species accumulation curve” typically shows a very simple shape, monotonically increasing, at a regularly decreasing pace, all along the course of progressive sampling. The process of species accumulation, however, is less simple than would be suggested by this simple shape. In fact, it is upon the *numbers of singletons, doubletons, ... x-tons*, of those species respectively recorded once, twice, ... x-times, that the sampling operation plays a *direct* role. Thus, progressive sampling results, at first, in the definite – and, as will be seen below, partly coordinate – variations of the numbers of singletons, doubletons, ..., x-tons. From this point of view, one may say, metaphorically, that sampling in act plays *directly* on the “keyboard” of the x-tons. In turn, the resulting regular growth of the species accumulation curve along progressive sampling is only *the consequence* (the cumulative result) of these combined variations of the numbers of singletons, doubletons, ..., x-tons.

Thinking this way might appear as a pure verbal or conceptual exercise, unnecessarily focusing upon the underlying details at the origin of the simple shape peculiar to any species accumulation curve. This, however, is not the case.

The priority consideration of the values and variations of the numbers  $f_1, f_2, f_3, \dots, f_x$ , of singletons, doubletons, tripletons, ..., x-tons, indeed has *major practical importance*, especially, as concerns the degree of advancement (completeness) of the sampling procedure. As a well-known example, most nonparametric estimators of the number of still unrecorded species (in particular “Chao” and the “Jackknife” series at different orders) are entirely based upon the values of the numbers  $f_x$  of species currently recorded x-times (considering the smallest values of x). In addition, beyond the *values* taken by the  $f_x$  at a given sampling-size, the *variations* of the  $f_x$  with sampling-size may be *highly informative* about the degree of advancement of the sampling process. And this is of more particular interest when considering the  $f_x$  of lowest order x, which primarily concern

the *least abundant* species, that make the bulk of those species remaining to be recorded.

More specifically, the progress of sampling may be considered either:

- classically and globally, by the estimated ratio of sampling completeness (ratio  $R_0/S_t$  between the number,  $R_0$ , of currently recorded species and the estimated total species richness,  $S_t$ , of the sampled assemblage of species);
- less classically and *more analytically*, by examining the trend of variation of each of the  $f_{x(N)}$  with increasing sampling-size N (focusing of course on the smallest values of x) at the current point of advancement of the considered sampling.

Admittedly, both approaches are *complementary* rather than mutually exclusive and, in current practice, the first, classical approach is likely expected to remain paramount. Yet, departing momentarily from the entire range of species to focus more specifically on the *least abundant* ones – which become progressively decisive for the further improvement of completeness of an ongoing sampling – is also of *substantial interest*.

Hereafter, we will concentrate on the determination of the trends of variations of each of the  $f_{x(N)}$  (in practice  $f_{1(N)}, f_{2(N)}, f_{3(N)}, f_{4(N)}$ ) with increasing sampling-size N. Meanwhile, essential general mathematical rules that (i) govern the variations of the  $f_{x(N)}$  with sampling-size N and (ii) establish a narrow linkage between the successive  $f_{x(N)}$ , will be highlighted.

## 2. PRELIMINARY: THE SUCCESSIVE DERIVATIVES OF THE NUMBER OF X-TONS WITH RESPECT TO SAMPLING-SIZE

The successive derivatives,  $\partial^x R(N)/\partial N^x$ , of the Species Accumulation Curve  $R(N)$  satisfy the following general equation:

$$\partial^x R_{(N)}/\partial N^x = (-1)^{x-1} f_{x(N)}/C_{N,x} \quad (1)$$

with  $R_{(N)}$  as the number of currently recorded species,  $f_{x(N)}$  as the number of x-tons and  $C_{N,x} = N!/x!(N-x)!$  as the number of combinations of x items among N. A detailed proof of this general theorem is provided in Appendix.

Leaving aside the very beginning of sampling (of no practical relevance here), the sampling-size N rapidly exceeds widely the numbers x of practical concern, so that, in practice,  $C_{N, x} \cong x! / N^x$ . Accordingly, the preceding equation may be simplified as:

$$\partial^x R_{(N)}/\partial N^x = (-1)^{x-1} (x!/N^x) \cdot f_{x(N)} \quad (2)$$

This relation has a *general relevance* because its derivation does not require any specific assumption relative to the particular shape of the distribution of species abundances in the sampled assemblage of species. Accordingly, equations (1) and (2) actually constrain the theoretical expressions of *any kind* of theoretical Species Accumulation Curves.

From equation (2) it comes:

$$f_{x(N)} = (-1)^{x-1} (N^x/x!) \cdot [\partial^x R_{(N)}/\partial N^x] \quad (3)$$

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

$$\partial f_{x(N)}/\partial N = (-1)^{x-1} \cdot \{x \cdot N^{x-1} \cdot [\partial^x R_{(N)}/\partial N^x] + N^x \cdot [\partial^{x+1} R_{(N)}/\partial N^{x+1}]\} / x!$$

By applying equation (2) to the expressions of  $[\partial^x R_{(N)}/\partial N^x]$  and  $[\partial^{x+1} R_{(N)}/\partial N^{x+1}]$ , it comes:

$$\partial f_{x(N)}/\partial N = [x \cdot f_{x(N)} - (x+1) \cdot f_{x+1(N)}] / N \quad (4)$$

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 recorded values taken by  $f_{x(N)}$  and  $f_{x+1(N)}$  at sampling-size N.

In turn, the second derivative of  $f_{x(N)}$  is obtained by further operating a new derivation of equation (4):

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

Replacing the derivatives  $\partial f_{x(N)}/\partial N$  and  $\partial f_{x+1(N)}/\partial N$  by their values according to equation (4) yields:

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

and, finally:

$$\partial^2 f_{x(N)}/\partial N^2 = [(x^2 - x) \cdot f_{x(N)} - (2x^2 + 2x) \cdot f_{x+1(N)} + (x^2 + 3x + 2) \cdot f_{x+2(N)}] / N^2 \quad (5)$$

Equation (5) thus provides the expression of the second derivative of the number  $f_{x(N)}$  at any

given sample-size N, in terms of the recorded values taken by  $f_{x(N)}$ ,  $f_{x+1(N)}$  and  $f_{x+2(N)}$  at sampling-size N.

In turn, iteratively operating new derivations of equation (5) would provide successively the derivatives of  $f_{x(N)}$  at any higher order. As a general rule, the expression of the  $i^{\text{th}}$  derivative of  $f_{x(N)}$ ,  $\partial^i f_{x(N)}/\partial N^i$ , involves the recorded values of the  $(i+1)$  numbers  $f_{x(N)}$ ,  $f_{x+1(N)}$ ,  $f_{x+2(N)}$ ,  $f_{x+3(N)}$ , ...  $f_{x+i(N)}$ , that is:

$$\partial^i f_{x(N)}/\partial N^i = (1/N^i) \cdot g_i(f_{x(N)}, f_{x+1(N)}, f_{x+2(N)}, f_{x+3(N)}, \dots, f_{x+i(N)}) \quad (6)$$

where  $g_i(f_{x(N)}, f_{x+1(N)}, f_{x+2(N)}, f_{x+3(N)}, \dots, f_{x+i(N)})$  is a linear function, with integer coefficients, of  $f_{x(N)}$ ,  $f_{x+1(N)}$ ,  $f_{x+2(N)}$ ,  $f_{x+3(N)}$ , ...  $f_{x+i(N)}$ .

As they result from relationship (2) above, the equations (4), (5) and (6) – defining  $\partial f_{x(N)}/\partial N$ ,  $\partial^2 f_{x(N)}/\partial N^2$  and, more generally,  $\partial^i f_{x(N)}/\partial N^i$  – thereby benefit from the same *general relevance* and, thus, are valid for *all kinds* of theoretical Species Accumulation Curves.

As shown below, the possibility of defining the successive derivatives of  $f_{x(N)}$  in terms of the (easily recorded) values of the  $f_{i(N)}$  at sampling-size N has important *theoretical* and *practical* consequences. This makes possible:

- (i) to disclose the *narrow mathematical linkage* that actually exists between all the successive  $f_{x(N)}$ : indeed, the shape of any  $f_{x(N)}$ , being entirely defined by its successive derivatives,  $\partial^i f_{x(N)}/\partial N^i$ , is, thereby, entirely connected to the corresponding values taken by all the following  $f_{i(N)}$  (i.e. for  $i \geq x$ );
- (ii) to *characterize quantitatively the main successive stages of variation* of the  $f_{x(N)}$  (especially those of lowest orders (singletons, doubletons, ...) with increasing sampling-size N, thus providing useful complementary clues to appreciate the level of progress in sampling the less abundant species within the assemblage.

### 3. THE THREE MAIN STAGES OF VARIATION OF THE NUMBER OF SPECIES RECORDED x-TIMES (x-TONS) ALONG PROGRESSIVE SAMPLING

As might be anticipated, the number  $f_{x(N)}$  of species recorded x-times is expected (Fig. 1):

- (i) to continuously *grow*, at first, with increasing sample-size N, then
- (ii) to pass by a *maximum* (at a sampling-size N' such that  $\partial f_x(N)/\partial N = 0$ ) and, finally,
- (iii) to *decrease* and ultimately reach asymptotically the zero level.

Moreover, being finally asymptotic, the decreasing part of the curve is thus expected to pass by a *point of inflection* (at a sampling-size N'' such that  $\partial^2 f_x(N)/\partial N^2 = 0$ ). The variation of any  $f_x(N)$  with N, during progressive sampling, may thus be sequenced according to three successive stages (I, II, III), separated by two threshold values, N' and N'', of sampling-size N.

### 3.1 The Sampling-size Threshold at Which the Number of x-tons Passes by a Maximum and Begins to Decrease

When  $f_x(N)$  reaches its maximum, the first derivative  $\partial f_x(N)/\partial N$  falls to zero and then, according to equation (4), it comes:

$$[x \cdot f_x(N) - (x+1) \cdot f_{x+1}(N)]/N = 0$$

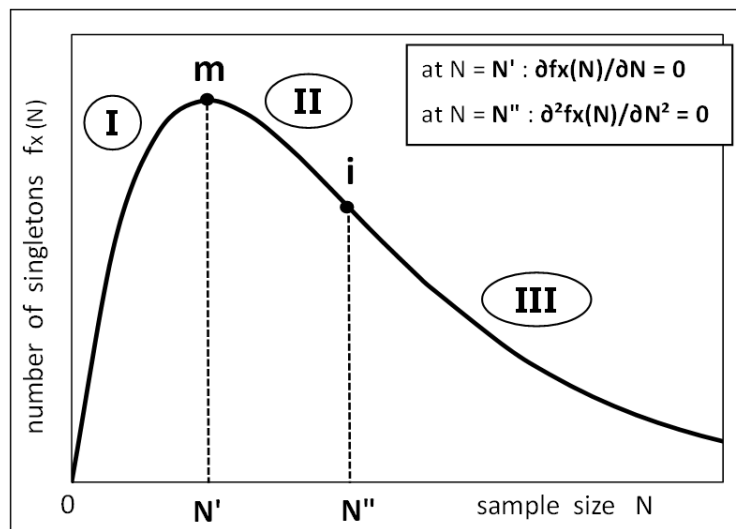
that is:

$$f_{x+1}(N) = [x/(x+1)] \cdot f_x(N) \text{ when } \partial f_x(N)/\partial N = 0 \quad (7)$$

Thus, the number of x-tons,  $f_x(N)$ , reaches its maximum at a sampling-size N' such that the number of (x+1)-tons ( $f_{x+1}(N')$ ) at N' is exactly  $[x/(x+1)]$  times the number of x-tons ( $f_x(N')$ ) at N'. Accordingly,  $f_{2(N)} = 1/2 f_{1(N)}$  when  $f_{1(N)}$  is at its maximum at  $N = N'$ ;  $f_{3(N)} = 2/3 f_{2(N)}$  when  $f_{2(N)}$  is at its maximum, at  $N = N'$ ;  $f_{4(N)} = 3/4 f_{3(N)}$  when  $f_{3(N)}$  is at its maximum and so on...

Equation (7) highlights a first – partial – link between two successive  $f_x(N)$  ( $f_x(N)$  and  $f_{x+1}(N)$ ). This iterative linkage eventually connects, step by step, the whole series of the  $f_x(N)$ . Fig. 2 provides a graphical representation of this iterative connection for the five first x-tons:  $f_{1(N)}$ ,  $f_{2(N)}$ ,  $f_{3(N)}$ ,  $f_{4(N)}$ ,  $f_{5(N)}$ .

In practice, the variation of the  $f_x(N)$  with sampling-size N have rarely been published, which would allow to compare theory and records. To our knowledge, such records have only been carried on and published fourth [1-4] and for singletons and doubletons only. As expected, all these four references confirm the theoretical prediction: the recorded co-variations of  $f_{1(N)}$  and  $f_{2(N)}$  always are in fair agreement with equation (7), namely  $f_{2(N)} = 1/2 f_{1(N)}$  when  $f_{1(N)}$  reaches its maximum.



**Fig. 1. Typical sketch of variation of the number  $f_x(N)$  of x-tons (species recorded x-times) with increasing sampling-size N. The first derivative,  $\partial f_x(N)/\partial N$ , falls to zero at point 'm' (at  $N = N'$ ) and the second derivative,  $\partial^2 f_x(N)/\partial N^2$ , falls to zero at point 'i' (at  $N = N''$ ). Three successive stages of variation of  $f_x(N)$  are thus delimited: at first, a rapid increase (stage I), then a decrease at an accelerating rate (stage II) and, at last, a decrease at a decelerating rate ("asymptotic decrease": stage III). Points 'm' and 'i' correspond respectively to the maximum and the inflection of the curve  $f_x(N)$**

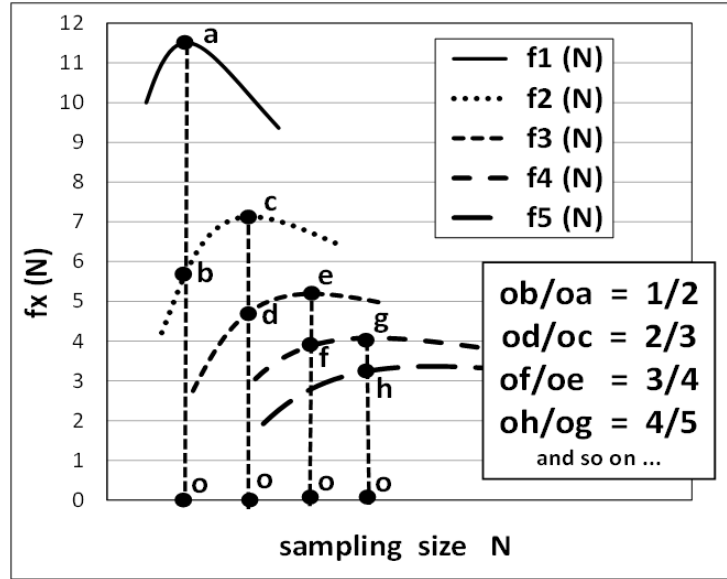


Fig. 2. Graphical representation of the connection between  $f_x(N)$  and  $f_{x+1}(N)$ , i.e. between the maximum value of  $f_x(N)$  and the corresponding value taken by  $f_{x+1}(N)$  at the same sampling-size. Adapted from [5]. This figure highlights the “linkage pattern” between the successive curves  $f_x(N)$  imposed by the constraining relationship (7), according to which  $f_{x+1}(N)$  reaches exactly  $[x/(x+1)]$  times the value of  $f_x(N)$  when the latter reaches its maximum, at  $N = N'$

### 3.2 The Sampling-size Threshold at Which the Number of x-tons Begins Its Decelerating Decrease (“asymptotic decrease”)

After having reach its maximum value at sampling-size  $N'$ ,  $f_x(N)$  then enters a decreasing phase. At first, this decrease is at an *accelerated* pace (stage II) and then at a *decelerating* pace (stage III), in accordance with the final asymptotic vanish. The transition between stages II and III is characterised by an *inflection point*, where  $\partial^2 f_x(N) / \partial N^2$  falls to zero. According to equation (5), it comes:

$$[(x^2 - x).f_x(N) - (2x^2 + 2x).f_{x+1}(N) + (x^2 + 3x + 2).f_{x+2}(N)] = 0 \quad (8)$$

Thus, the number of x-tons,  $f_x(N)$ , begins its second, decelerated, *asymptotic decrease* when the sampling-size reaches a value  $N''$  such that the numbers  $f_x(N)$ ,  $f_{x+1}(N)$  and  $f_{x+2}(N)$  satisfy equation (8).

### 3.3 The Particular Case of the Numbers of Singletons and Doubletons

According to equation (8), the number of singletons,  $f_1(N)$ , begins its asymptotic decrease

when the sampling-size  $N$  reaches a value such that  $-4.f_2(N) + 6.f_3(N) = 0$ . Now, from equation (7), this value of  $N$  also corresponds exactly to the step when  $f_2(N)$  reaches its maximum. Therefore, the number of singletons *always* enters its last, *decelerated* decreasing phase (step ‘i’) *precisely when* the number of doubletons reaches its maximum value (step ‘m’). This is a new remarkable connection between the two first  $f_x(N)$ .

## 4. THE NARROW MATHEMATICAL CONNECTION THAT LINKS THE SUCCESSIVE x-TONS (NUMBERS OF SPECIES RECORDED x-TIMES)

### 4.1 Main Mathematical Linkage

As already stated at section 2, the variations of the number  $f_x(N)$  during progressive sampling are narrowly linked to the variations of *all the*  $f_j(N)$  of *higher order*, i.e. for all  $j > x$ : equation (6). This may be more explicitly highlighted by considering the expression of the Taylor expansion of  $f_x(N)$ . According to the general formulation of Taylor expansion, the variations of  $f_x(N)$  in a range  $[N-\delta, N+\delta]$  around the sampling-size  $N$ , may be written as:

$$f_x(N + \delta) = f_x(N) + \sum_{i=1}^{\infty} [(\partial^i f_x(N) / \partial N^i) \cdot (\delta^i / i!)]$$

In turn, the general relationship (6) allows to express the  $i^{\text{th}}$  derivative,  $\partial^i f_{x(N)} / \partial N^i$ , in terms of the recorded values of the  $(i+1)$  numbers  $f_{x(N)}$ ,  $f_{x+1(N)}$ ,  $f_{x+2(N)}$ ,  $f_{x+3(N)}$ , ...  $f_{x+i(N)}$ :

$$\partial^i f_{x(N)} / \partial N^i = (1/N^i) \cdot g_i(f_{x(N)}, f_{x+1(N)}, f_{x+2(N)}, f_{x+3(N)}, \dots, f_{x+i(N)})$$

Accordingly, the Taylor expansion of  $f_{x(N)}$  may be written as a function of the values taken by the series of the  $f_j(N)$  for  $j > x$ :

$$f_{x(N+\delta)} = f_{x(N)} + \sum_{i=1}^X [(1/N^i) \cdot g_i(f_{x(N)}, f_{x+1(N)}, f_{x+2(N)}, f_{x+3(N)}, \dots, f_{x+i(N)})] \cdot (\delta^i / i!) \quad (9)$$

with:

- $g_i$  as a linear function of the  $f_{x(N)}$ ,  $f_{x+1(N)}$ ,  $f_{x+2(N)}$ ,  $f_{x+3(N)}$ , ...  $f_{x+i(N)}$ , with integer coefficients
- $X$  as the number of individuals of the most abundant species that has been recorded at the currently reached sampling-size  $N$ .

Equation (9) thus highlights the general expression of the *narrow mathematical linkage* that exists between all the successive  $f_{x(N)}$  and, more precisely, between the variation of  $f_{x(N)}$  around the current sampling-size  $N$  and the corresponding values that are taken, at sampling-size  $N$ , by the series of  $f_i(N)$ , for  $i \geq x$ .

In particular, considering the Taylor expansion of  $f_{x(N)}$  at order 2 (which is sufficient, in practice, to encompass the bulk of the local variations of  $f_{x(N)}$ ), it comes:

$$f_{x(N+\delta)} \approx f_{x(N)} + [x \cdot f_{x(N)} - (x+1) \cdot f_{x+1(N)}] \cdot (\delta/N) + [(x^2 - x) \cdot f_{x(N)} - (2x^2 + 2x) \cdot f_{x+1(N)} + (x^2 + 3x + 2) \cdot f_{x+2(N)}] \cdot (\frac{1}{2} \delta^2 / N^2) \quad (10)$$

Note that, similarly, a Taylor expansion of the Species Accumulation Curve,  $R(N)$ , within a neighbourhood  $[N-\delta, N+\delta]$  of the sampling-size  $N$ , has already been derived: see reference [6]. The Taylor expansion of  $R(N)$  is:

$$R_{(N+\delta)} = R_{(N)} + \sum_{i=1}^{\infty} [\partial^i R_{(N)} / \partial N^i] \cdot (\delta^i / i!)$$

Replacing the successive derivatives by their expressions, according to equation (1) leads to:

$$R_{(N+\delta)} = R_{(N)} + \sum_{i=1}^{\infty} [(-1)^{i-1} (i! / N^i) \cdot f_{i(N)}] \cdot (\delta^i / i!)$$

that is:

$$R_{(N+\delta)} = R_{(N)} + \sum_{i=1}^{\infty} (-1)^{i-1} (\delta/N)^i \cdot f_{i(N)}$$

In practice:

$$R_{(N+\delta)} = R_{(N)} + \sum_{i=1}^X (-1)^{i-1} (\delta/N)^i \cdot f_{i(N)} \quad (11)$$

with  $X$  as the number of individuals of the most abundant species that have been recorded at the current sampling-size  $N$ .

## 4.2 An Additional Mathematical Linkage

Still an additional mathematical linkage between the successive  $f_x(N)$  may be unveiled by considering the intersection between  $f_x(N)$  and  $f_{x+1(N)}$ , i.e. when sampling size  $N$  is such that  $f_{x+1(N)}$  becomes equal to  $f_x(N)$ . From equation (4) it comes immediately:

$$\partial f_{x(N)} / \partial N = -f_{x(N)} / N \quad \text{for sampling size } N \text{ such that } f_{x+1(N)} = f_x(N) \quad (12)$$

Accordingly, this demonstrates:

- that  $f_{x+1(N)}$  intersects  $f_{x(N)}$  when the latter has already reached its decreasing phase (since  $f_{x(N)} / N$  is essentially positive);
- that the slope (decreasing rate) of  $f_{x(N)}$  at this intersection point is equal in module and opposite in sign to the ratio  $f_{x(N)} / N$ , thus resulting graphically in a remarkable geometrical property, as shown in Fig. 3, with angle IBA being *equal* to angle IAB (the triangle AIB is *isosceles*).

## 5. DISCUSSION

The numbers  $f_1, f_2, f_3, \dots, f_x$ , of singletons, doubletons, ...,  $x$ -tons (species respectively recorded 1-, 2-, 3-, ...,  $x$ -times) vary, of course, with sampling-size  $N$ . Each number  $f_{x(N)}$  successively shows three phase of variation with  $N$ : a growth period (I), then an accelerated decreasing period (II) and, at last, a decelerated decreasing period (III), eventually ending asymptotically to zero (Fig. 1). The thresholds values  $N'$  and  $N''$ , which delimit these three stages, are dependent on  $x$  (the larger  $x$ , the larger are  $N'$  and  $N''$ ), but these three stages of variation along progressive sampling remain characteristic of the variation of *any* number  $f_{x(N)}$ , whatever the value  $x$ , that is for any  $x$ -ton. In spite of this common general scheme, each number  $f_{x(N)}$  varies, however, at its own pace during progressive sampling. Yet, it has been demonstrated above that the respective variations of the different numbers  $f_x$  are far from being entirely independent from each-other. On the contrary, remarkable *connections* have been unveiled between them. These connections appears explicitly by considering the Taylor

expansion of the number  $f_{x(N)}$  (equations (9) and (10)) which shows that the variations of  $f_{x(N)}$  in a neighbourhood of  $N$  depend on the values taken by the series of the  $f_{i(N)}$ , for  $i \geq x$ . In other words, the variations of the number of species recorded  $x$ -times are *connected* to the numbers of species recorded still more frequently during sampling. Some important consequences of these connections are highlighted graphically at Figs. 2 and 3.

All these considerations might seem, at first, of pure speculative interest. However, they can also address more *pragmatic concerns*.

Thus, for example, the degree of sampling completeness is usually quantified by the ratio between the number of recorded species and the (estimated) total species richness of the sampled assemblage of species. According to this common approach, the scope is, first of all, focused on what has *already been recorded*. Yet, an alternative (complementary) approach, however, would consist to focus upon what is *still to be recorded*. That is to say, focus may relevantly be placed, first and foremost, on those species that are, statistically, the *least abundant* in the sampled assemblage. In this second perspective, it is those numbers of species only recorded few (i.e. the  $x$ -tons with low values of  $x$ : singletons, doubletons, tripletons,...) that feature *more relevantly informative*. For example, it is this kind of approach that is implicitly considered in the generally accepted view that a survey may be considered virtually complete as soon as the recorded number  $f_1$  of singletons has fallen to zero.

In accordance with the preceding point of view (and for the bulk of practical surveys of biodiversity, that have not reached quasi-exhaustivity), the relevant question to be addressed is: what are the values and, more importantly, what are the *trends of variation* of the numbers  $f_{1(N)}$ ,  $f_{2(N)}$ ,  $f_{3(N)}$ ,... of singletons, doubletons, tripletons,... around the currently achieved sampling-size. Answering this question becomes highly meaningful as soon as sampling progress has reached such a level that it is only the *least abundant* species of the sampled assemblage that mainly contribute to  $f_1$ ,  $f_2$ ,  $f_3$ , ...

To provide an illustration of the interest and practical significance of this proposition, four examples are considered hereafter, involving four local surveys of butterfly fauna in different suburban localities around Jhansi (India) [7]. For each survey, the variations of the numbers  $f_{1(N)}$ ,  $f_{2(N)}$ ,  $f_{3(N)}$ ,  $f_{4(N)}$ , of those species respectively

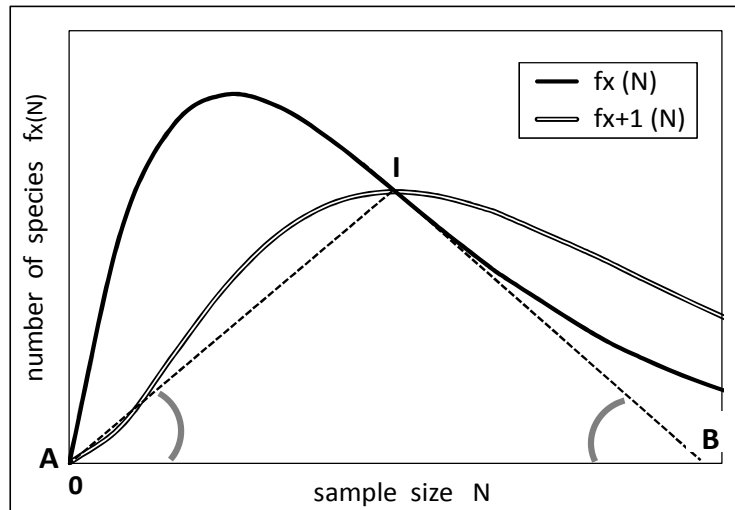
recorded 1-, 2-, 3-, 4- times are computed around the corresponding achieved sampling-size  $N_0$  of each survey, using the Taylor expansion of  $f_{x(N)}$  at order 2 (equation (10)). That is, the Taylor expansion (equation (10)) is implemented with the values of  $f_1$ ,  $f_2$ ,  $f_3$ ,  $f_4$ , recorded at the end of the achieved sampling (i.e. for  $N = N_0$ ): Figs. 4 to 7 (N.B.: more precisely and in order to reduce the influence of drawing stochasticity on the as-recorded values of the numbers  $f_x$ , their distribution is, at first, smoothened by regression of the as-recorded distribution of the  $f_x$  versus  $x$  : see reference [7] for details on this point).

\* Fig. 4 is for butterfly survey at "Parichha Dam" (estimated sampling completeness 65%): referring to Fig. 1, the numbers  $f_1$ ,  $f_2$ ,  $f_3$ ,  $f_4$  at the currently achieved sampling-size  $N_0$ , are at stages II, I, I, I, respectively. That is, at  $N_0$ , the number of singletons begins to decrease while the numbers of doubletons, tripletons and quadrupletons are still growing.

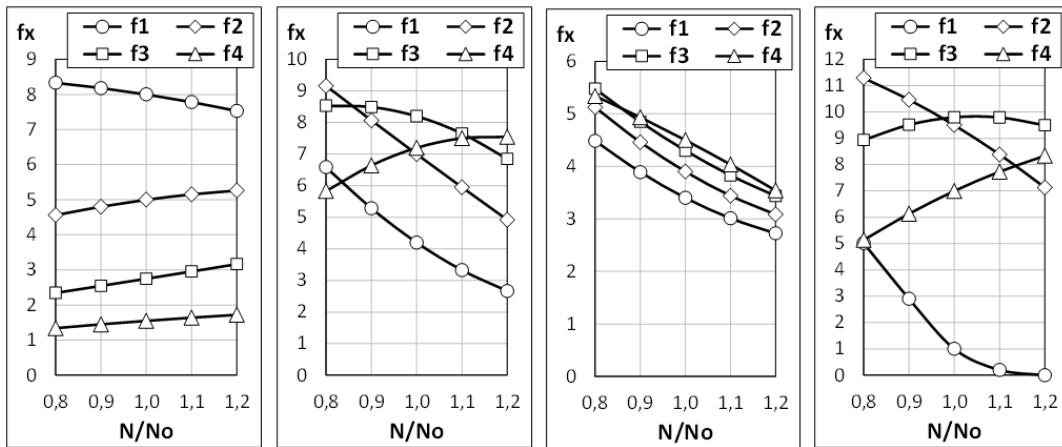
\* Fig. 5 is for butterfly survey at "Jhansi University Campus" (estimated sampling completeness 90%): referring to Fig. 1, the numbers  $f_1$ ,  $f_2$ ,  $f_3$ ,  $f_4$ , at the currently achieved sampling-size  $N_0$ , are at stages III, III, II, I, respectively. That is, at  $N_0$ , the numbers of singletons and doubletons have already begun their last asymptotic decreasing phase, while the number of tripletons has entered its accelerated decreasing phase and the number of quadrupletons is still increasing.

\* Fig. 6 is for butterfly survey at "Narayan Bagh" (estimated sampling completeness 92%): referring to Fig. 1, the numbers  $f_1$ ,  $f_2$ ,  $f_3$ ,  $f_4$ , at the currently achieved sampling-size  $N_0$ , are at stages III, III, III, II, respectively. That is, at  $N_0$ , the numbers of singletons, doubletons and tripletons have already begun their last asymptotic decreasing phase while the number of quadrupletons has entered its accelerated decreasing phase.

\* Fig. 7 is for butterfly survey at "Bundelkhand Institute Engineering & Technology Campus" (quasi exhaustive survey): referring to Fig. 1, the numbers  $f_1$ ,  $f_2$ ,  $f_3$ ,  $f_4$ , at the currently achieved sampling-size  $N_0$ , are at stages III, II,  $\approx m$ , I, respectively. That is, at  $N_0$ , the number of singletons has already begun its last asymptotic decreasing phase, the number of doubletons has entered its accelerated decreasing period, the number of tripletons has just approximately reached its maximum and, at last, the number of quadrupletons is still increasing.



**Fig. 3.** Typical sketch of variation, with increasing sampling-size  $N$ , of the number  $f_x(N)$  of  $x$ -tons and the number  $f_{x+1}(N)$  of  $(x+1)$ -tons. Both curves intersect at point I. A remarkable mathematical property, at the intersection between  $f_x(N)$  and  $f_{x+1}(N)$ , is that, there, the slope (decreasing rate) of  $f_x(N)$  is equal in module and opposite in sign to the ratio  $f_{x(N)}/N$ . This results in a remarkable geometrical property: the equality of angles IAB and IBA; in other words, the triangle AIB is *isosceles*



**Figs. 4, 5, 6, 7.** Variations of the numbers  $f_1, f_2, f_3, f_4$ , of species respectively recorded 1-, 2-, 3-, 4- times according to sampling-size  $N$  around the currently achieved sample-size  $N_0$ . The variations of the  $f_{x(N)}$  are computed using a limited Taylor expansion (equation (10)) around the size,  $N = N_0$ , of the actually achieved sampling. That is, the Taylor expansion (equation (10)) is implemented with the values of  $f_1, f_2, f_3, f_4$ , as recorded at the end of the sampling carried out (i.e. for  $N = N_0$ ) – Fig. 4: “Parichha Dam” ; Fig.5 “Jhansi Univ. Campus” ; Fig. 6: “Narayan Bagh” ; Fig. 7: “Bundelkhand Institute Eng. & Techn. Campus”

Further comments in text.

Thus, as expected, there is a general trend for the numbers  $f_{x(N)}$  to pass the successive steps of their variations (stages I, m, II, i, III) in accordance with increasing level of sampling completeness. For example, at 65% completeness (“Parichha Dam”), the numbers  $f_1$ ,

$f_2, f_3, f_4$ , are at stages II, I, I, I, respectively, while, at 92% completeness (“Narayan Bagh”), the numbers  $f_1, f_2, f_3, f_4$ , have already reached stages III, III, III, II, respectively. Yet, this correlation remains *rather loose*, as is exemplified, by comparing “B.I.E.T. Campus” to “Narayan Bagh”:



$f_1, f_2, f_3, f_4$ , are at stages III, II,  $\approx$  m, I, for the quasi exhaustive sampling of “*B.I.E.T. Campus*” while  $f_1, f_2, f_3, f_4$ , are at more advanced stages (III, III, III, II), for the 92% completeness sampling of “*Narayan Bagh*”.

This loose correlation, however, is not surprising. Indeed, the degrees of advancement of the variations of the  $f_{x(N)}$  along their successive stages are not only related to the global level of sampling completeness ( $R_0/S_t$ ) but are still dependent *also* upon the level of unevenness of species abundance distribution in the sampled assemblage. At any given level of sampling completeness, the more uneven is the species abundance distribution, the slower will be the degree of advancement of the  $f_{x(N)}$  in the passage of the successive steps of their variations.

This highlights the fact that the stages of variations respectively reached by  $f_1, f_2, f_3, f_4$ , deserves consideration alongside the usual sampling-completeness ratio  $R_0/S_t$ . Taking into account the stages of variations respectively reached by  $f_1, f_2, f_3, f_4$ , actually provides non-trivial, truly complementary information regarding the degree of advancement of the sampling process.

To close this topic, a more laconic and synthetic presentation of the degree of advancement in the survey of the least abundant species of the sampled assemblage may simply consist in displaying the “score” of those  $f_{x(N)}$  that have already overstepped their respective maximum ‘m’ and enter their decreasing stages (II or III), at the currently reached sampling-size  $N_0$ . Thus, for the four preceding surveys, the scores are as follows: for “*Parichha Dam*”:  $f_1$  ; for “*Jhansi University Campus*”:  $f_1, f_2, f_3$  ; for “*Narayan Bagh*”:  $f_1, f_2, f_3, f_4$  ; for “*B.I.E.T. Campus*”:  $f_1, f_2$ .

## 6. CONCLUSION

Although looking quite simple, the monotonic process of species accumulation during progressive sampling is, in fact, far less trivial. Indeed, at any given sampling-size, the process of species accumulation is the cumulated result of quite a more convoluted underlying process, involving the non-monotonic variations of each of the numbers  $f_{x(N)}$  of species recorded x-times. Moreover, although partially connected with each other, the variations of each of the  $f_{x(N)}$  progress, however, at different paces, in a relative independence from each other in this respect.

Disclosing the respective variations of each of the  $f_{x(N)}$  is, thus, quite a non-trivial issue, which has yet been successfully addressed above.

In particular, the general expression of the variations of the  $f_{x(N)}$  has been appropriately derived, using a Taylor expansion approach. Beyond the speculative aspects of the question - at the very heart of the detailed understanding of species accumulation rates along progressive sampling - more practical aspects have also been addressed. In particular, accounting for the variations of the low-orders  $f_{x(N)}$ , (such as  $f_{1(N)}, f_{2(N)}, f_{3(N)}, f_{4(N)}$ ) proves *especially significant* regarding the degree of advancement of ongoing surveys. This is so because the further improvements of sampling completeness progressively involve the recording of less and less abundant species, which primarily influence the low-orders  $f_{x(N)}$ . The Taylor expansion of the numbers  $f_{1(N)}, f_{2(N)}, f_{3(N)}, f_{4(N)}, \dots$  around the currently reached sampling-size may thus cast more relevant light upon the effective progress of an ongoing survey and thus provide an additional and complementary tool to accurately evaluate the degree of sampling efficiency.

## ACKNOWLEDGEMENTS

Three anonymous reviewers are acknowledged for their relevant comments on the original manuscript.

## COMPETING INTERESTS

Author has declared that no competing interests exist.

## REFERENCES

1. Scharff N, Coddington JA, Griswold CE, Hormiga G, Per de Place B. When to quit? Estimating spider species richness in a northern European deciduous forest. *The Journal of Arachnology*. 2003;31:246-273.
2. Coddington JA, Agnarsson I, Miller JA, Kuntner M, Hormiga G. Undersampling bias: The null hypothesis for singleton species in tropical arthropod surveys. *Journal of Animal Ecology*. 2009;78:573-584.
3. De Walt RE, Cao Y, Tweddle T, Grubbs SA, Hinz L, Pessiono M, Robinson JL. Ohio USA stoneflies (*Insecta, Plecoptera*): Species richness estimation, distribution of functional niche traits, drainage affiliations and relationships to other states. *ZooKeys*. 2012;178:1-26.

4. Malumbres-Olarte J, Scharff N, Pape T, Coddington JA, Cardoso P. Gauging megadiversity with optimized and standardized sampling protocols: A case for tropical forest spiders. *Ecology and Evolution*. 2017;7:494-506.
5. 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.<hal01516141>
6. 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;<hal01228695>
7. 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
8. Lee SM, Chao A. Estimating population size via sample coverage for closed capture-recapture models. *Biometrics*. 1994;50(1):88-97.

## APPENDICES

### A1 - Derivation of 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  $\Delta$  of missed species (unrecorded in the sample) is  $\Delta = S - R$ .

The estimated number  $\Delta$  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  $p_i$ :

$$\Delta_{(N)} = \sum_i (1-p_i)^N \quad (\text{A1.1})$$

with  $\sum_i$  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:

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

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):

$$\blacktriangleright f_1 = N \sum_i [(1-p_i)^{N-1} p_i] = N \sum_i [(1-p_i)^{N-1} (1 - (1-p_i))] = N \sum_i [(1-p_i)^{N-1}] - N \sum_i [(1-p_i)^{N-1} (1-p_i)] = N \sum_i [(1-p_i)^{N-1}] - N \sum_i [(1-p_i)^N]$$

Then, according to equation (A1.1) 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:

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

Similarly:

$$\begin{aligned} \blacktriangleright f_2 &= C_{N,2} \sum_i [(1-p_i)^{N-2} p_i^2] \quad \text{according to equation (A1.2)} \\ &= C_{N,2} \sum_i [(1-p_i)^{N-2} (1 - (1-p_i^2))] = C_{N,2} [\sum_i [(1-p_i)^{N-2}] - \sum_i [(1-p_i)^{N-2} (1-p_i^2)]] \\ &= C_{N,2} [\sum_i [(1-p_i)^{N-2}] - \sum_i [(1-p_i)^{N-2} (1-p_i)(1+p_i)]] = C_{N,2} [\sum_i [(1-p_i)^{N-2}] - \sum_i [(1-p_i)^{N-1} (1+p_i)]] \\ &= C_{N,2} [(\Delta_{(N-2)} - \Delta_{(N-1)}) - f_1/N] \quad \text{according to equations (A2.1) and (A1.2)} \\ &= C_{N,2} [-\Delta'_{(N-1)} - f_1/N] = C_{N,2} [-\Delta'_{(N-1)} + \Delta'_{(N)}] \quad \text{since } f_1 = -N \Delta'_{(N)} \quad (\text{cf. equation (A1.3)}). \\ &= C_{N,2} [(\partial \Delta'_{(N)}/\partial N)] = [N(N-1)/2] (\partial^2 \Delta_{(N)}/\partial N^2) = [N(N-1)/2] \Delta''_{(N)} \end{aligned}$$

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

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

$$\begin{aligned} \blacktriangleright f_3 &= C_{N,3} \sum_i [(1-p_i)^{N-3} p_i^3] \quad \text{which, by the same process, yields:} \\ &= C_{N,3} [\sum_i (1-p_i)^{N-3} - \sum_i (1-p_i)^{N-2} - \sum_i [(1-p_i)^{N-2} p_i] - \sum_i [(1-p_i)^{N-2} p_i^2]] \\ &= C_{N,3} [(\Delta_{(N-3)} - \Delta_{(N-2)}) - f_1^*/(N-1) - 2 f_2/(N(N-1))] \quad \text{according to equations (A2.1) and (A1.2)} \end{aligned}$$

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):

$$f_1^* = -(N-1) \Delta'_{(N-1)} = -C_{N-1,1} \Delta'_{(N-1)} \quad \text{and} \quad f_2 = [N(N-1)/2] \Delta''_{(N)} = C_{N-1,2} \Delta''_{(N)} \quad (A1.5)$$

where  $\Delta'_{(N-1)}$  is the first derivative of  $\Delta_{(N)}$  with respect to N, at point  $(N-1)$ . Then,

$$\begin{aligned} f_3 &= C_{N,3} [(\Delta_{(N-3)} - \Delta_{(N-2)}) + \Delta'_{(N-1)} - \Delta''_{(N)}] = C_{N,3} [-\Delta'_{(N-2)} + \Delta'_{(N-1)} - \Delta''_{(N)}] \\ &= C_{N,3} [\Delta''_{(N-1)} - \Delta''_{(N)}] = C_{N,3} [-\partial \Delta''_{(N)}/\partial N] = C_{N,3} [-\partial^3 \Delta_{(N)}/\partial N^3] = C_{N,3} \Delta'''_{(N)} \end{aligned}$$

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

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

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

$$\begin{aligned} \blacktriangleright f_x &= C_{N,x} \sum_i [(1-p_i)^{N-x} p_i^x] \quad \text{according to equation (A1.2),} \\ &= C_{N,x} \sum_i [(1-p_i)^{N-x} (1 - (1-p_i^x))] = C_{N,x} [\sum_i (1-p_i)^{N-x} - \sum_i [(1-p_i)^{N-x} (1-p_i^x)]] \\ &= C_{N,x} [\sum_i (1-p_i)^{N-x} - \sum_i [(1-p_i)^{N-x} (1-p_i)(\sum_j p_j^x)]] \end{aligned}$$

with  $\sum_j$  as the summation from  $j=0$  to  $j=x-1$ . It comes:

$$\begin{aligned} f_x &= C_{N,x} [\sum_i (1-p_i)^{N-x} - \sum_i [(1-p_i)^{N-x+1} (\sum_j p_j^x)]] \\ &= C_{N,x} [\sum_i (1-p_i)^{N-x} - \sum_i (1-p_i)^{N-x+1} - \sum_k [(\sum_i (1-p_i)^{N-x+1} p_i^k)]] \end{aligned}$$

with  $\sum_k$  as the summation from  $k=1$  to  $k=x-1$ ; that is:

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

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:

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

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

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

which finally yields:

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

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

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

Conversely:

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

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

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

In particular:

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

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

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  $N$ :

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

with  $[\partial^x R_{(N)}/\partial N^x]$  as the  $x^{\text{th}}$  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:

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

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  $[i-1 \rightarrow i]$  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  $[i \rightarrow i+1]$  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  $i \geq 1$ :

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

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

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

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  $[0 \rightarrow 1]$ , 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

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

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):

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

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

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

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

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:

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

More generally:

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

© 2017 Béguinot; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Peer-review history:  
The peer review history for this paper can be accessed here:  
<http://sciencedomain.org/review-history/20156>