Original Research Article

General mathematical rules regulating the process of species accumulation during progressive sampling: the variations of the numbers of singletons, doubletons,…, x-tons with increasing sampling-size

ABSTRACT

The process of species accumulation during progressive sampling, although looking quite simple, results in fact from a more convoluted underlying process, involving the non-monotonic variations with sampling-size N of the numbers $f_{1(N)}$, $f_{2(N)}$, $f_{3(N)}$, …, $f_{x(N)}$ of those species currently recorded 1-, 2-, 3-, …, x-times. Moreover, although they are partially connected with each other, the $f_{x(N)}$ vary at different paces with increasing sampling-size N, in a relative independence from each other in this respect. Disentangling the respective variations of each of the $f_{x(N)}$ was a first object of the present study. In particular a general expression of the variations of the $f_{x(N)}$ with N is derived, using Taylor expansion procedure. Beyond the speculative aspects of the question (i.e. a more thorough understanding of the underlying mechanism of species accumulation along progressive sampling), more practical aspects are also subsequently addressed. Focusing upon the variations of the *lower*-orders $f_{x(N)}$, (i.e. $f_{1(N)}$, $f_{2(N)}$, $f_{3(N)}$, $f_{4(N)}$) unveils *especially relevant information* regarding the progress of ongoing surveys towards sampling exhaustivity. This is because any further improvement of sampling completeness progressively involves the new records of less and less abundant species within the sampled assemblage. The Taylor expansion of the numbers $f_{1(N)}$, $f_{2(N)}$, $f_{3(N)}$, $f_{4(N)}$, … around the currently reached sampling-size N of an ongoing survey may thus serve as an additional tool to relevantly evaluate the degree of sampling efficiency.

Key-words: 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 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. Progressive sampling thus results, at first, in the definite – and, as will be seen below, partly coordinate – variations of the numbers of singletons, doubletons,… x-tons. 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 with progressive sampling is only *the consequence* 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.

Dealing with the values and variations of the numbers f_1, f_2, f_3,…, 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 expected 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 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 concern the *least abundant* species making 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, with sampling-size N, of each of the $f_{x\ (N)}$ (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, I will concentrate on the determination of the trends of variations of each of the $f_{x\ (N)}$ (in practice f_1, f_2, f_3, f_4) with increasing sampling-size N. Meanwhile, essential general mathematical rules that (i) govern the variations of the $f_{x\ (N)}$ with 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 N

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

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

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, the preceding equation simplifies as:

$$[\partial^x R_{(N)}/\partial N^x] = (-1)^{x-1} (x!/N^x).f_{x\ (N)} \qquad (2)$$

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 Species Accumulation Curves.

From equation (2) it comes:

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

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

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

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:

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

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 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):

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

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:

$$\partial^2 f_{x\,(N)}/\partial N^2 =$$

$$[-\,x.f_{x\,(N)} + (x+1).f_{x+1\,(N)} + x^2.f_{x\,(N)} - (x+1).(2x+1).f_{x+1\,(N)} - (x+1).(x+2).f_{x+2\,(N)}]/N^2$$

and, finally:

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

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^th^ derivative of $f_{x\,(N)}$, $[\partial^i f_{x\,(N)}/\partial N^i]$, involves the recorded values of the (i+1) numbers $f_{x\,(N)}, f_{x+1\,(N)}, f_{x+2\,(N)}, f_{x+3\,(N)}, \ldots f_{x+i\,(N)}$, that is:

$$\partial^i f_{x\,(N)}/\partial N^i = (1/N^i).g_i(f_{x\,(N)}, f_{x+1\,(N)}, f_{x+2\,(N)}, f_{x+3\,(N)}, \ldots f_{x+i\,(N)}) \qquad (6)$$

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

As they result from relationship (2) above, 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 all benefit from the same *general relevance* and, thus, are valid for *all kinds* of 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 practical consequences. This makes possible:

(i) to *characterize quantitatively the main successive stages of variation* of the $f_{x\,(N)}$ along increasing sampling-size N ;

(ii) to unveil a *narrow mathematical linkage* that actually exists between all the successive $f_{x\,(N)}$: indeed, being entirely defined by its successive derivatives $\partial^i f_{x\,(N)}/\partial N^i$, the shape of any $f_{x\,(N)}$, is, thereby, entirely linked to the corresponding values taken by all the following $f_{i\,(N)}$ (i.e. for $i \geq x$).

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:

(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. This successive steps are schematised at Figure 1.

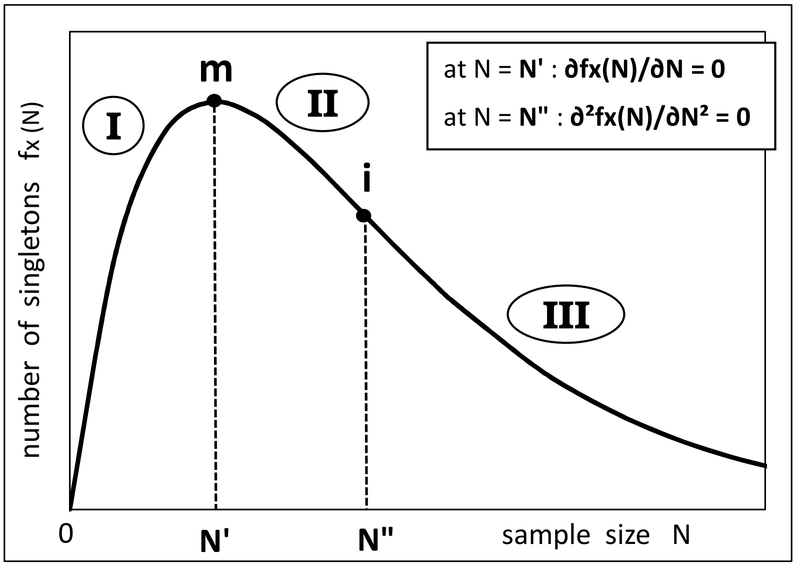


Figure 1 - Typical sketch of variation of the number $f_x(N)$ of x-tons (species recorded x-times) with increasing sampling-size N. The first derivative, $\partial f_x(N)/\partial N$, falls to zero at point 'm' (at N=N') and the second derivative, $\partial^2 f_x(N)/\partial N^2$, falls to zero at point '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 $f_x(N)$.

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:

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

that is:

$$f_{x+1\,(N)} = [x/(x+1)].f_{x\,(N)} \quad when\ \partial f_{x\,(N)}/\partial N = 0 \qquad (7)$$

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, $f_{2\,(N)} = ½\ f_{1\,(N)}$ when $f_{1\,(N)}$ is at its maximum; $f_{3\,(N)} = 2/3\ f_{2\,(N)}$ when $f_{2\,(N)}$ is at its maximum; $f_{4\,(N)} = 3/4\ f_{3\,(N)}$ 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 $f_{x\,(N)}$. Figure 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 my knowledge, such records have only been carried on and published fourth [1 to 4] and for singletons and doubletons only. According to these four references, the recorded co-variations of $f_{1\,(N)}$ and $f_{2\,(N)}$ are, as expected, in fair agreement with the theoretical prediction from equation (7), namely $f_{2\,(N)} = ½\, f_{1\,(N)}$ when $f_{1\,(N)}$ reaches its maximum.

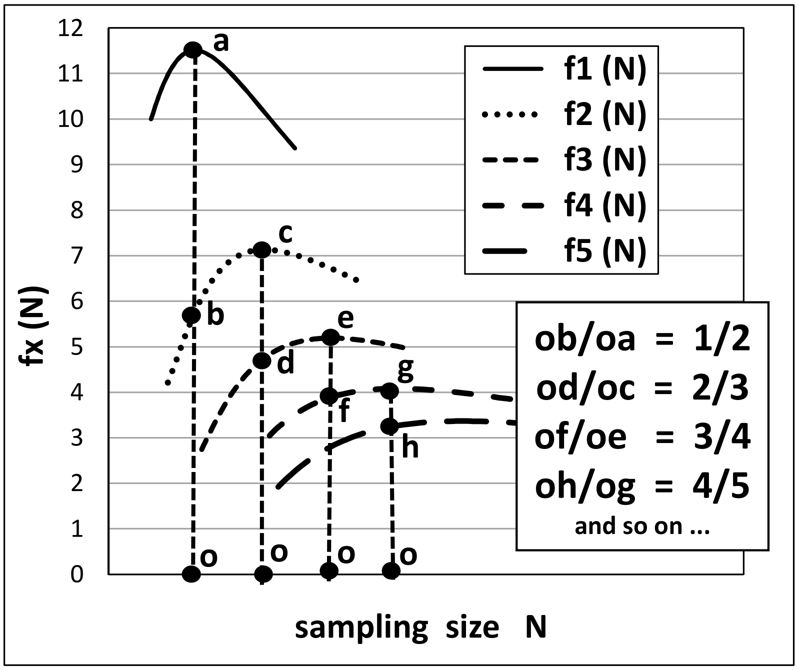


Figure 2 – Graphical representation of the connection between the maximum value of $f_{x\,(N)}$ and the corresponding value taken by $f_{x+1\,(N)}$ at the same sampling-size. Adapted from [5]. This figure highlights the "linkage pattern" between the successive curves $f_{x\,(N)}$ imposed by the constraining 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

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 decreasing rate (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:

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

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').

4. A NARROW MATHEMATICAL CONNECTION UNITES THE SUCCESSIVE NUMBERS OF SPECIES RECORDED x-TIMES (x-TONS)

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 $j > x$. This may be 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 $[N-\delta\ ,\ N+\delta]$ of the sampling-size N, may be written as:

$$f_{x\,(N+\delta)} = f_{x\,(N)} + \Sigma_{i=1\text{ to }\infty}\,(\partial^i f_{x\,(N)}/\partial N^i).(\delta^i/i!).$$

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

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

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 $j > x$:

$$f_{x\,(N+\delta)} = f_{x\,(N)} + \Sigma_{i=1\text{ to }X}\,[(1/N^i).g_i\,(f_{x\,(N)}, f_{x+1\,(N)}, f_{x+2\,(N)}, f_{x+3\,(N)}, \ldots, f_{x+i\,(N)})\,].(\delta^i/i!) \qquad (9)$$

with:

- g_i as a linear function of the $f_{x\,(N)}, f_{x+1\,(N)}, f_{x+2\,(N)}, f_{x+3\,(N)}, \ldots f_{x+i\,(N)}$, 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:

$$\begin{aligned} f_{x\,(N+\delta)} \approx\ & f_{x\,(N)} + [\,x.f_{x\,(N)} - (x+1).f_{x+1\,(N)}\,].(\delta/N) \\ & + [(x^2 - x).f_{x\,(N)} - (2x^2 + 2x).f_{x+1\,(N)} + (x^2 + 3x + 2).f_{x+2\,(N)}].(\tfrac{1}{2}\,\delta^2/N^2) \qquad (10) \end{aligned}$$

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 [6]. The Taylor expansion of R(N) is:

$$R_{(N+\delta)} = R_{(N)} + \Sigma_{i=1\text{ to }\infty}\,[\partial^i R_{(N)}/\partial N^i].(\delta^i/i!)$$

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

$$R_{(N+\delta)} = R_{(N)} + \Sigma_{i=1\text{ to }\infty}\,[(-\,1)^{i-1}\,(i!/N^i).f_{i\,(N)}].(\delta^i/i!)$$

that is:

$$R_{(N+\delta)} = R_{(N)} + \Sigma_{i=1 \text{ to } \infty} \, (-1)^{i-1} \, (\delta/N)^{i} . f_{i\,(N)}$$

In practice:

$$R_{(N+\delta)} = R_{(N)} + \Sigma_{i=1 \text{ to } X} \, (-1)^{i-1} \, (\delta/N)^{i} . f_{i\,(N)} \qquad (11)$$

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:

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

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 Figure 3, with angle I**B**A being *equal* to angle I**A**B (the triangle AIB is *isosceles*) .

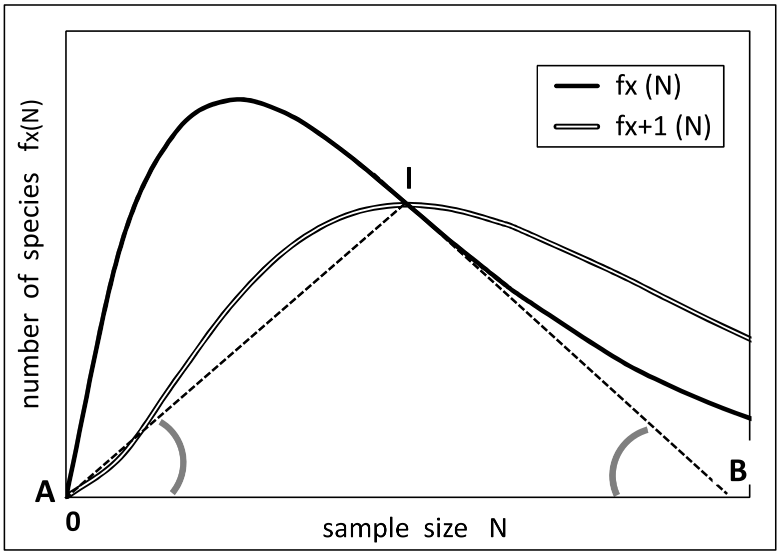


Figure 3 - Typical sketch of variation, with increasing sampling-size N, of the number $f_x(N)$ of x-tons and the number $f_{x+1}(N)$ of (x+1)-tons. Both curves intersect at point I. A remarkable mathematical property at the intersection between $f_x(N)$ and $f_{x+1}(N)$ is that the slope (decreasing rate) of $f_{x(N)}$, there, 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 I**A**B and I**B**A : the triangle AIB is *isosceles*

5. DISCUSSION

The numbers f_1, f_2,..., f_x, of singletons, doubletons, ..., x-tons (species respectively recorded 1-, 2-, ..., 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 (Figure 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* $f_{x\,(N)}$, whatever the value x, that is for any x-ton. In spite of this common general scheme, each number f_x 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 *not* entirely independent from each-other. On the contrary, a remarkable connection has been unveiled between them. This connection appears explicitly by considering the Taylor expansion of the $f_{x\,(N)}$ (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. Moreover, some remarkable consequences of this connection are highlighted graphically at Figures 2 and 3.

All these considerations might seem, at first, of pure speculative interest. However, they also have substantial practical consequences and may answer more pragmatic concerns. Usually, the degree of sampling completeness is, quite naturally, 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 usual approach, the scope is, first of all, focused on what has *already been recorded*. An alternative (complementary) approach would consist to focus upon what is *still to be recorded*. That is to say, to consider first and foremost those species that are, statistically, the *least abundant* in the sampled assemblage. In this second perspective, it is the numbers of species only recorded few (x-tons with low values of x: singletons, doubletons,…) that are more relevantly informative. For example, it is this kind of approach which is implicit in the admittedly common opinion that a survey may be considered virtually complete as soon as the 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), a relevant question to be addressed is: what about (i) the levels and (ii) the trends of variation of the numbers of singletons, doubletons, tripletons, etc … ($f_{1(N)}$, $f_{2(N)}$, $f_{3(N)}$, …) at the currently achieved sampling-size N. Answering this question would reveal highly meaningful, as soon as sampling progress has reached the level where the still unrecorded species mainly concern the *least-abundant* species of the sampled assemblage.

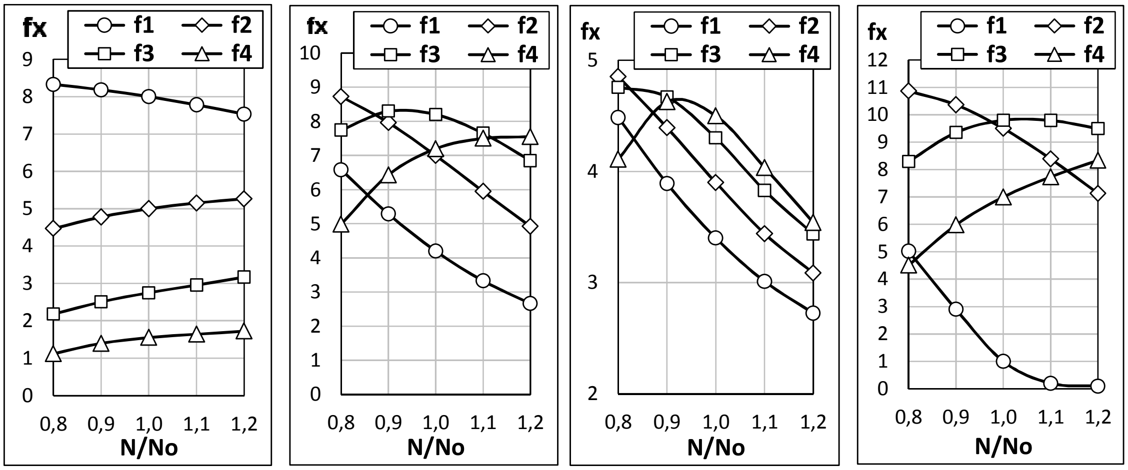
To provide an illustration of the interest and practical significance of this proposition, four examples are considered hereafter, involving four local surveys of butterfly fauna in different suburban localities around Jhansi (India) [7]. For each survey, the variations of the numbers $f_{1(N)}$, $f_{2(N)}$, $f_{3(N)}$, $f_{4(N)}$, of those species respectively recorded 1-, 2-, 3-, 4-times are computed around the achieved sampling-size N_0, using the Taylor expansion of $f_{x\,(N)}$ at order 2 (equation (10)): Figures 4 to 7.

Figure 4 is for butterfly survey at "*Parichha Dam*" (estimated sampling completeness 65%): referring to Figure 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 doubletons, tripletons and quadrupletons are still growing.

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

Figure 6 is for butterfly survey at "*Narayan Bagh*" (estimated sampling completeness 92%): referring to Figure 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, II, respectively. That is, at N_0, the numbers of singletons and doubletons have already begun their last asymptotic decreasing phase while the numbers of tripletons and quadrupletons are in their accelerated decreasing phase.

Figure 7 is for butterfly survey at "*Bundelkhand Institute Engeneering & Technology Campus*" (quasi exhaustive survey): referring to Figure 1, the numbers f_1, f_2, f_3, f_4, at the currently achieved sampling-size N_0, are at stages III, II, ≈ 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 the number of quadrupletons is still increasing.



Figures 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 Taylor expansion around $N = N_0$ (equation (10)).

Four surveys of butterfly fauna in different localities around the city of Jhansi (Béguinot 2017):

- *Fig 4*: "Parichha Dam" : for the achieved sampling-size ($N = N_0$) and, referring to Figure 1, the numbers f_1, f_2, f_3, f_4, are at stages II, I, I, I, respectively

- *Fig 5*: "Jhansi Univ. Campus" : for the achieved sampling-size ($N = N_0$) and, referring to Figure 1, the numbers f_1, f_2, f_3, f_4, are at stages III, II, II, I, respectively

- ***Fig 6***: "Narayan Bagh" : for the achieved sampling-size ($N = N_0$) and referring to Figure 1, the numbers f_1, f_2, f_3, f_4, are at stages III, III, II, II, respectively. Other comments in the text.

- *Fig 7*: "Bundelkhand Institute Eng. & Techn. Campus" : for the achieved sampling-size ($N = N_0$) and referring to Figure 1, the numbers f_1, f_2, f_3, f_4, are at stages III, II, ≈ m, I, respectively

Thus, as expected, there is a global trend for the numbers $f_{x\,(N)}$ passing 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, II, II, respectively. Yet, this correlation is *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, ≈ 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, II, II), for the 92% completeness sampling of "*Narayan Bagh*".

This loose correlation, however, is not surprising. Indeed, the degrees of advancement of the $f_{x(N)}$ along the successive stages of their variations 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.

To close this topic, a more laconic and synthetic presentation of the degree of advancement of the survey of the least abundant species in the sampled assemblage may simply consist in displaying the "score" of those $f_{x(N)}$ that have 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.

CONCLUSION

Although looking quite simple, the monotonic process of species accumulation during progressive sampling is, in fact, far less trivial. Indeed, species accumulation is the cumulated result of a more convoluted underlying process, involving the non-monotonic variations of each of the $f_{x(N)}$ (i.e. the numbers of species recorded x-times at any given sampling-size N). Moreover, although partially connected with each other, the variations of each of the $f_{x(N)}$ progress at different paces, in a relative independence from each other in this respect. Disentangling the respective variations of each of the $f_{x(N)}$ is, thus, quite a non-trivial issue, which has 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, considering 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 towards sampling exhaustivity. This is so because further improvement of sampling completeness progressively involves less and less abundant species of the sampled assemblage, which primarily influence the low-orders $f_{x(N)}$. The Taylor expansion of the numbers $f_{1(N)}$, $f_{2(N)}$, $f_{3(N)}$, $f_{4(N)}$, … 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 tool to accurately evaluate sampling efficiency.

APPENDICES

A.1 - 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 Δ of missed species (unrecorded in the sample) is Δ = 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:

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

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:

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

► $f_3 = C_{N,3}\,\Sigma_i\,[(1-p_i)^{N-3}\,p_i^3]$ 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):

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

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 :

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

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,

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

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 $[i\text{-}1 \rightarrow i]$ 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 $i \geq 1$:

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

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 $[0 \rightarrow 1]$, 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

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

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):

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

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

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

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

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:

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

More generally:

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

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, Tweddale 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. <hal-01516141>.

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 ; <hal-01228695>

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.