

Basic Theoretical Arguments Advocating Jackknife-2 as Usually being the Most Appropriate Nonparametric Estimator of Total Species Richness

Jean Béguinot^{1*}

¹Biogéosciences, 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/ARRB/2016/25104

Editor(s):

(1) George Perry, Dean and Professor of Biology, University of Texas at San Antonio, USA.

Reviewers:

(1) Loc Nguyen, University of Natural Science, Ho Chi Minh City, Vietnam.

(2) P. Rama Bhat, Mangalore University, Karnataka, India.

Complete Peer review History: <http://sciencedomain.org/review-history/13904>

Method Article

Received 18th February 2016

Accepted 14th March 2016

Published 28th March 2016

ABSTRACT

A lot of nonparametric estimators of the number of unrecorded species after partial sampling of an assemblage of species, have been proposed in the literature. Unfortunately, these different types of estimators provides substantially divergent predictions. While empirical comparisons have failed to consistently select in favour of one among all these estimators, a new approach, based on more theoretical ground, has proven that among three of the most commonly used nonparametric estimators, Chao, Jackknife-1 and Jackknife-2, the latter was the best choice in most cases while Chao or Jackknife-1 should preferably be restricted to samplings approaching completeness. Here, I propose an alternative approach, aiming also at discriminating between the same three estimators on the basis of another theoretical argument: The necessary compliance with the required "rule of additivity", according to which, if an assemblage of species is made of several, distinct groups of species, the estimation of species richness for the whole assemblage should be exactly the sum of the estimations of richness for each group of species. Referring to this rule of additivity, the Jackknife series of estimators (and in particular Jackknife-2 when samples remain far from completeness), proves, once again, being satisfactory in full generality. This strengthens the estimators of the Jackknife series as being particularly appropriate to evaluate, in most cases, the number of unrecorded species of a partially sampled assemblage and the corresponding total species richness of the assemblage.

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

Keywords: Estimator; Chao; Jackknife; less biased estimate; additivity; mathematical constraint; species accumulation; plant galls; Alps; Parc National des Ecrins.

1. INTRODUCTION

Incomplete species samplings are deemed to become increasingly frequent in the future [1]. This is, in particular, a consequence of biodiversity surveys being progressively more and more dedicated to animals (or plants) groups which usually give rise to assemblages of numerous species represented by individuals of small sizes, more or less hard to detect in the field (such as, for example, assemblages of small- or micro-invertebrates).

Incomplete samples (resulting from under sampling) immediately raise the questions of:

- What would be the *total* species richness of the assemblage, that is the expected number of recorded species if the sampling was ideally complete;
- What would be the extrapolated shape of the so called “Species Accumulation Curve”, beyond the currently achieved sample-size, that is the expected kinetic of discovery of new species with sampling growing beyond its present size [2].

Both questions are of importance and thus prompted researchers to propose a series of expressions for both the nonparametric estimation of total species richness (review in [3,4]) and the extrapolation of the Species Accumulation Curve (review in [5]). Accordingly, the issue, now, is rather to select among the varied reported propositions of estimators, since, unfortunately, each of them provides a substantially different result [6]. A considerable amount of work has been devoted to test comparatively these different types of estimators, mainly on an *empirical* basis [7-13]. But, as might be expected, no consensus emerged from these studies. This is because, in fact, each kind of estimator may provide a centered, unbiased prediction in a very specific case only: when the species abundance distribution within the sampled assemblage of species has a particular kind of shape, that is, in turn, a specific shape for the Species Accumulation Curve, specific to each type of estimator [3,13,14-16]. Thus, finally, these empirical approaches hardly help to disclose information of any *general* value.

Accordingly, a more appropriate approach might consist in addressing the issue from a less

empirical, more theoretical point of view, thus providing results of more general applicability and soundness.

A guide of choice was thus derived in this perspective, build on a *purely theoretical basis*, defining which type of nonparametric estimator, among the most commonly used, is able to provide the less biased estimations and under which conditions. Although very simple to use in practice, this guide was developed following a rather long mathematical development [17,18].

Hereafter, I derive quite independently, an alternative theoretical argumentation, finally leading to similar selective guidance, but involving a much simpler and concise mathematical demonstration.

2. REMINDING OF THE MAIN RESULTS OBTAINED ACCORDING TO THE ORIGINAL DERIVATION

Consider the sampling of an assemblage of species with sample size N_0 (N_0 observed individuals) and containing $R(N_0)$ recorded species among which a number f_1 are recorded only once and a number f_2 are recorded only twice. The expected number Δ of species missed by the sampling (and thus the total species richness of the assemblage $R(N_0)+\Delta$) may be conveniently estimated using one of several types of nonparametric estimators, among which the following three are most commonly used:

Chao with $\Delta = f_1^2/(2f_2)$, *Jackknife-1* with $\Delta = f_1$ and *Jackknife-2* with $\Delta = 2f_1 - f_2$.

Now, a specific mathematical relationship (equation A1.1 in Appendix A1) constrains the expression of any theoretically possible Species Accumulation Curve (S.A.C.). As a consequence [18], each type of estimators is associated to a particular type of S.A.C. Accordingly, the expressions of the Species Accumulation Curves $R(N)$ associated to the estimators *Jackknife-1*, *Jackknife-2* and *Chao* are respectively as follows, according to [18]:

$$R(N) = (R(N_0) + f_1) - f_1 \cdot N_0/N \quad (1)$$

$$R(N) = (R(N_0) + 2f_1 - f_2) - (3f_1 - 2f_2) \cdot N_0/N - (f_2 - f_1) \cdot N_0^2/N^2 \quad (2)$$

$$R(N) = (R(N_0) + f_1^2/(2f_2)) \cdot (1 - \exp([\ln((f_1^2/(2f_2)) / (R(N_0) + f_1^2/(2f_2)))](N/N_0))) \quad (3)$$

with $R(N)$ as the number of recorded species when the sample reaches size N ($N > N_0$).

3. AN ALTERNATIVE, INDEPENDENT DERIVATION OF THE (SAME) GUIDE OF CHOICE, BASED ON THE COMPLIANCE WITH THE ‘RULE OF ADDITIVITY’

Consider an assemblage of species that encompass several “mutually exclusive” categories (that is categories that share *no species in common*; for example taxonomic categories such as, genus, families, orders, etc...). A sample ‘A’ of this assemblage gathers the sub-samples ‘a₁’, ‘a₂’, ..., ‘a_x’, corresponding to each of the categories separately. If the sample A is incomplete, an estimation of the number M of missing species may be conveniently obtained using one among the three cited nonparametric estimators, Chao, Jackknife -1 or Jackknife -2. Now, the same procedure may be implemented for the separate estimations of the numbers m_1, m_2, \dots, m_x , of missing species in each of the sub-samples a_1, a_2, \dots, a_x .

All those species that escape recording in one or the other sub-samples a_1, a_2, \dots, a_x are evidently constitutive of the set of species that are missed by the whole sample A, so that the sum $\sum_x m_x$ of the numbers m_1, m_2, \dots, m_x , should coincide exactly with the estimation M of the number of missing species for the whole sample, that is:

$$M = \sum_x m_x \quad (4)$$

Once more, this relationship holds true only for “mutually exclusive” categories that share no species in common.

Now, in fact, all nonparametric estimators of the number of missing species may not equally satisfy this compulsory “rule of additivity”, as shown below.

3.1 Jackknife Estimators JK-1 and JK-2

It may be immediately recognized that both Jackknife estimators considered here (Jackknife-1 with $\Delta = f_1$ and Jackknife-2 with $\Delta = 2f_1 - f_2$) *always satisfy* the required rule of additivity stated by equation (4). This is the obvious consequence of the expressions of these

estimators as a *linear combination* of f_1 and f_2 . Yet, although being a necessary condition, the compliance with the rule of additivity may not be sufficient. A second constraint comes from the obvious fact that the number of missing species is expected to *decrease monotonically* with increasing sample size.

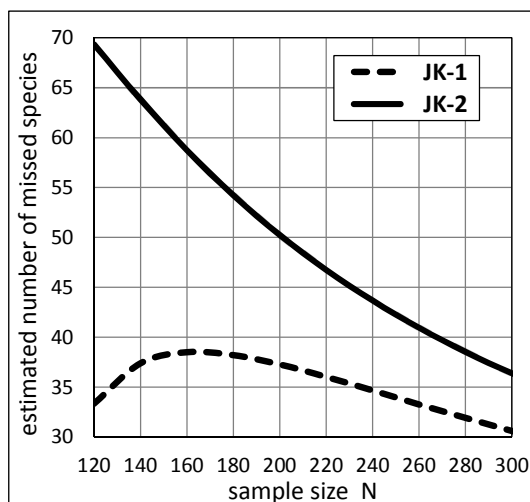


Fig. 1. Evolution, with growing sample size N , of the estimated number of missing species according to Jackknife-1 and to Jackknife-2 (survey of gall-inducing Diptera in “Parc National des Ecrins” (see section 5))

According to Jackknife-2, the estimated number of missing species monotonically decreases with growing sample size, as relevantly expected. By contrast, the estimated number of missing species according to Jackknife-1 ($= f_1$) would begin to increase, pass through a maximum and finally monotonically decrease with growing sampling size. While such an increasing stage, at first, is correct for the number f_1 of species recorded once (to which Jackknife -1 is identified), it is of course quite unsatisfying for an estimation of the number of missing species and Jackknife-1 is thus disqualified for this reason

This is not constantly the case, however, for the number f_1 of species recorded once, all along progressive sampling: with growing sampling size, f_1 begins, at first, to increase, then pass through a maximum before finally decreasing monotonically: illustrative example at Fig. 1. Thus Jackknife-1 ($= f_1$) cannot be satisfactory in all circumstances. In fact, using Jackknife-1 is inadequate for most incomplete samples, that is, as long as samplings still not approach completeness. More precisely, using Jackknife-1 remains inadequate as long as the ratio f_1/f_2 (which is monotonically decreasing with growing sample size and level of sample completeness) remains larger than 1: in such case, Jackknife-2

is to be preferred while Jackknife-1 is to be retained only when f_1/f_2 falls below 1. It should be noted that limiting the use of Jackknife-1 to higher levels of completeness only is an option already suggested by Brose et al. [19], at least for the specific types of species abundance distributions they studied (without indication given, however, relative to the threshold value of the ratio f_1/f_2).

Nota: Alternatively, Chao estimator may also be considered (as is Jackknife-1) for higher levels of completeness [18]. Yet, contrary to Jackknife-1, Chao does not comply with the rule of additivity, as shown below. This should preclude using Chao estimator in any circumstance. But, in fact, the deviation of Chao estimator from the rule of additivity is progressively vanishing when samples approach completeness, as demonstrated in next section. Accordingly, using Chao estimator in the specific case when $f_1/f_2 \leq 0.6$ (and Jackknife-2 when $f_1/f_2 > 0.6$), as proposed previously [18], does not violate appreciably the prescribed rule of additivity. Such a narrowly restricted range of applicability of Chao estimator, was already pointed by several authors [6,16,20-23], on (semi-) empirical basis.

3.2 Chao Estimator

Because its expression is not linearly dependent upon f_1 and f_2 (contrary to the Jackknife estimators), the Chao estimator generally *does not comply* with the prescribed rule of additivity. A particular condition only allows Chao estimator to respect the rule of additivity however: when the ratio f_1/f_2 happens to take a same value – say k – for each of the sub-samples. Then, Chao estimator becomes: $\Delta = f_1^2/(2f_2) = \frac{1}{2} k f_1$ for each of the sub-samples. Under this very specific condition, Chao complies with the additivity rule:

$$\sum_x m_x = \sum_x (\frac{1}{2} k f_{1x}) = \frac{1}{2} k \sum_x (f_{1x}) = M$$

since the number of species recorded once in the whole sample equals the sum $\sum_x (f_{1x})$ of the numbers of species recorded once in each of the sub-samples (the latter being mutually exclusive in terms of species identities, as already pointed). The rule of additivity is then satisfied by Chao estimator *only* when the ratio f_1/f_2 takes the same value for all the sub-samples (and the whole sample).

Now, as the ratio f_1/f_2 provides a relevant appreciation of the degree of sampling completeness (as already mentioned), this identity of the ratios f_1/f_2 between all the sub-

samples means that all of them should show the same degree of completeness. In turn, this common degree of completeness for the whole sample and for all the sub-samples implies an even distribution of species abundances within the whole sampled assemblage (N.B.: when samplings approach completeness, unevenness among abundances of still unrecorded species progressively vanishes).

Interestingly, this constraint of abundance evenness is exactly the same argued independently [24,25] as the necessary condition for Chao expression being able to provide unbiased estimates. Thus, the condition ensuring the compliance of Chao expression with the rule of additivity on the one hand, and the condition for Chao providing unbiased estimations on the other hand, are indeed identical. Satisfactorily enough, both requirements are thus mutually consistent.

Now, what happens in full generality, i.e. when the ratios f_1/f_2 are substantially different between the sub-samples, so that Chao estimator can no longer satisfy the rule of additivity? In this general context, the Chao estimator deviates more or less from what is prescribed by the rule of additivity. More precisely (*see Appendix A2 for a demonstration*), the degree of deviation (i) *increases* with the difference between the ratios f_1/f_2 of each of the sub-samples and (ii) *decreases* with the difference between the values of f_1 for each of the sub-samples.

This is exemplified quantitatively, in Table 1 and Fig. 2, considering, for simplicity, the case when only two categories of species are involved in the whole sample (mathematics behind Table 1 and Fig. 2 is provided at Appendix A2 (i) & (ii)).

Note that the improved version “i-Chao” [24] of Chao estimator, involving f_3, f_4 in addition:

$\Delta = f_1^2/(2f_2) + [f_3/(4f_4)] \times \text{MAX}\{f_1 - (f_2 \cdot f_3)/(2f_4), 0\}$ does not comply substantially better with the required rule of additivity, for the same reason of non-linearity of the expression of i-Chao in terms of f_1, f_2, f_3, f_4 . An illustrative example is provided at section 5.

4. THE RULE OF ADDITIVITY AND THE EXTRAPOLATION OF THE SPECIES ACCUMULATION CURVE

As already mentioned in section 2, a specific shape for the extrapolated Species Accumulation

Curve is associated to each type of nonparametric estimator [18]. The expressions of the Species Accumulation Curves respectively associated to Jackknife-1, Jackknife-2 and Chao estimators are given by equations (1) (2) and (3) respectively. It is immediately apparent that the Species Accumulation Curves associated to Jackknife-1 and Jackknife-2 both respect the rule of additivity in full

generality, while, on the contrary, the Species Accumulation Curve associated to Chao estimator generally does not satisfy the rule of additivity, except in the very particular case when the all the sub-samples happen to have reached the same degree of sampling completeness (as already pointed independently for the estimator itself, see section 3.2).

Table 1. Relative deviation of Chao estimates from the rule of additivity ($M = \sum_x m_x$) for a sample A encompassing two sub-samples {G and H}, with numbers $\{g_1, h_1\}$ of species recorded once and $\{g_2, h_2\}$ of species recorded twice respectively

		$\alpha = \max \text{ of } (g_1/g_2) / (h_1/h_2) \text{ or } (h_1/h_2) / (g_1/g_2)$									
		1	2	3	4	5	6	7	8	9	10
$\beta = \max \text{ of } g_1/h_1 \text{ or } h_1/g_1$	1	0	12	29	44	57	68	78	87	94	101
	2	0	11	26	40	52	63	73	81	88	95
	3	0	9	22	35	46	56	65	73	80	86
	4	0	8	19	31	41	50	58	66	73	79
	5	0	7	17	27	36	45	53	60	66	72
	6	0	6	15	24	33	41	48	55	61	66
	7	0	5	14	22	30	37	44	50	56	61
	8	0	5	12	20	27	34	41	46	52	57
	9	0	4	11	18	25	32	38	43	48	53
	10	0	4	10	17	23	29	35	40	45	50

The relative gap between the Chao estimate M of the number of missing species for the whole sample A and the sum of Chao estimates m and m' for the sub-samples G and H is $D = |M - (m+m')| / \frac{1}{2}(M+(m+m'))$ in %. It is studied against the ratios $\alpha = (g_1/g_2)/(h_1/h_2)$ and $\beta = (g_1/h_1)$

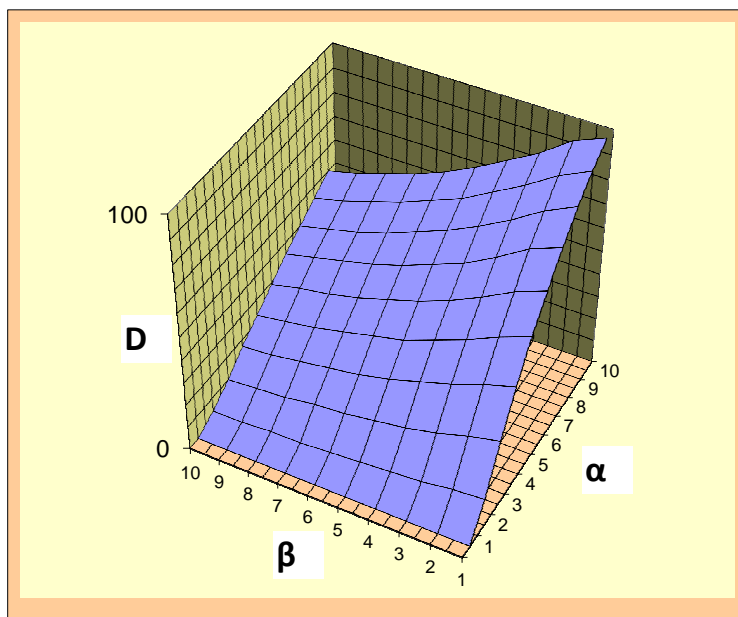


Fig. 2. Relative deviation of Chao estimates from the rule of additivity ($M = \sum_x m_x$) for a sample A encompassing two sub-samples G and H with numbers $\{g_1, g_2\}$ and $\{h_1, h_2\}$ of species recorded once and twice respectively

The relative gap between the Chao estimate M of the number of missing species for A and the sum of Chao estimates m and m' for G and H is $D = |M - (m+m')| / \frac{1}{2}(M+(m+m'))$ in %. It is studied against the ratios $\alpha = (g_1/g_2)/(h_1/h_2)$ and $\beta = (g_1/h_1)$

This leads to the same severe limitations of use of Chao estimator as that already mentioned above for the estimation of the number of missing species in an incomplete sample.

5. AN ILLUSTRATIVE EXAMPLE: THE SAMPLING OF GALL-INDUCING ALPINE FAUNA, JOINTLY INVOLVING FOUR ORDERS OF ARTHROPODS

As an illustrative example, I consider a partial survey of the gall-inducing arthropods fauna performed in the “Parc National des Ecrins”, in French Alps (see [26] for a full taxonomic list of the $R_o = 337$ recorded species and details of localisations of the $N_o = 147$ surveyed sites). Four orders of gall-inducing arthropods are mainly involved in the survey.

Table 2 provides the numbers $R_o (=R(N_o))$ of recorded species and the numbers f_1, f_2, f_3, f_4 of species recorded respectively 1, 2, 3, 4 times at the end of sampling for each of the four orders and for all the four orders considered together. Then, Table 3 provides a comparison between the estimated numbers of missing species according to Chao, i-Chao, Jackknife-1 and Jackknife-2 estimators.

Table 2. The numbers $R_o (=R(N_o))$ of recorded species and the numbers f_1, f_2, f_3, f_4 of species recorded respectively 1, 2, 3, 4 times (i) for each of the four orders Acaria, Hemiptera, Diptera, Hymenoptera, and (ii) for all these four orders considered together

P.N. Ecrins	R_o	f_1	f_2	f_3	f_4
Acaria	128	45	17	11	9
Hemiptera	47	9	9	4	1
Diptera	118	38	14	10	11
Hymenoptera	44	6	8	9	5
all 4 orders	337	98	48	34	26

6. DISCUSSION: THE PRACTICAL IMPORTANCE OF SELECTING THE MORE APPROPRIATE NON-PARAMETRIC ESTIMATOR, ENABLING A RELIABLE EXTRAPOLATION OF THE SPECIES ACCUMULATION CURVE

The practical importance of a reliable extrapolation of total species richness and the additional sampling efforts required to improve the current level of sampling completeness is conveniently highlighted by considering in more details the illustrative example above.

Table 3. Estimating the number of missing species in a partial sampling of gall-inducing arthropods (“Parc National des Ecrins”) according to three nonparametric estimators: Chao, i-Chao [24], Jackknife-1 and Jackknife-2. The sampling encompasses four orders of arthropods: Acaria Thrombidiidae, Hemiptera, Diptera, Hymenoptera

P.N. Ecrins	Chao	iChao	JK-1	JK-2
Acaria	60	70	45	73
Hemiptera	5	5	9	9
Diptera	52	59	38	62
Hymenoptera	2	2	6	4
all 4 orders	100	122	98	148
sum on 4 orders	119	136	98	148
deviation	19	14	0	0

6.1 Estimation of the Total Species Richness

Table 4 provides, for each of the four orders and for the whole survey (all four orders together):

- (i) The estimations of the total species richness S , according to the three nonparametric estimators Jackknife-2, Jackknife-1, Chao;
- (ii) The corresponding degrees of completeness $R(N_o)/S$ (%)

Substantial disagreements occur between the three estimators regarding the prediction of total species richness of the sampled assemblage and the associated evaluation of the level of completeness of samplings.

6.2 Estimation of the Additional Sampling Efforts Required to Reach a Given Increment of Sampling Completeness

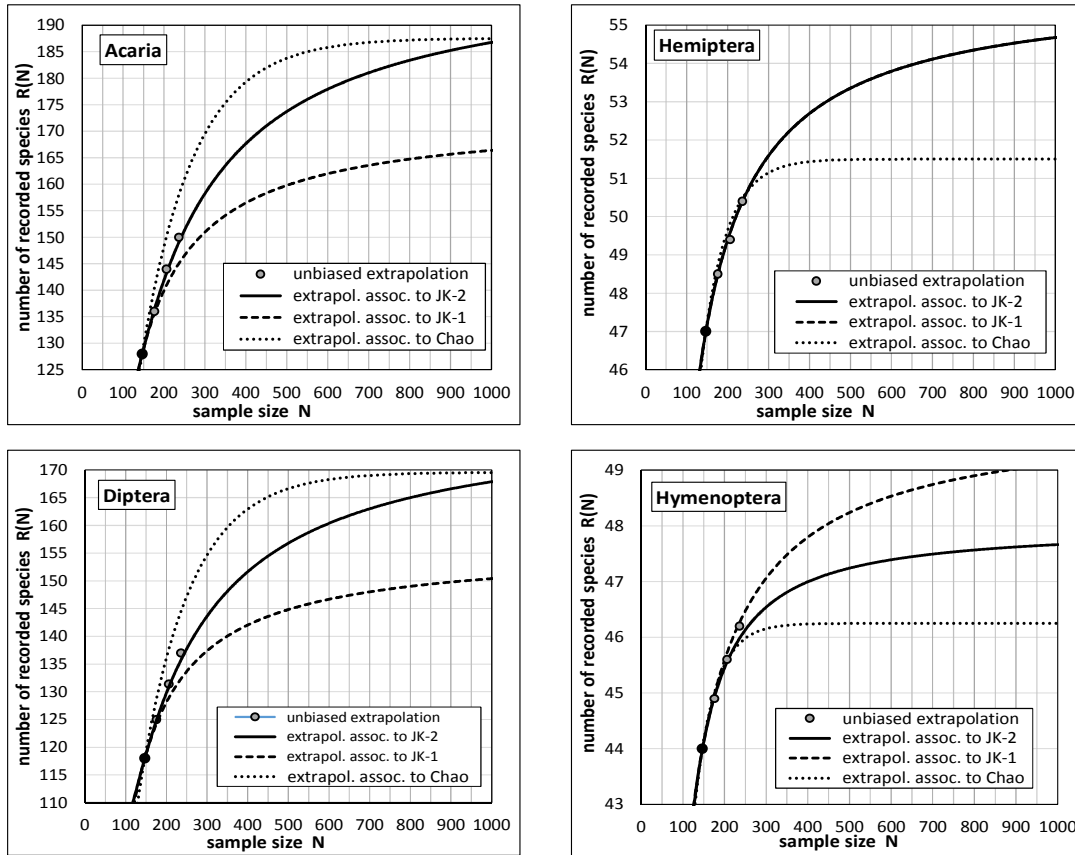
The extrapolations of the Species Accumulation Curve associated to each of the three nonparametric estimators Jackknife-2, Jackknife-1, Chao, are plotted at Figs. 3 to 6, for the four main orders of gall-inducing arthropods. These graphic representations highlight, in each case, the additional sampling effort that would be required to achieve any given increment of the number of recorded species. Extrapolations associated to Jackknife-1, Jackknife-2 and Chao are computed according to equations (1) (2) and (3) respectively. Here, the extrapolation associated to Jackknife-2 is selected as the less biased, since f_1/f_2 remains largely higher than 0.6

in all cases. Besides, a rigorously unbiased extrapolation may also be computed, but limited (no more than 60% of the originally realised effort (procedure of unbiased extrapolation derived in [17])).

Table 4. Estimates of the total species richness S according to the three nonparametric estimators Jackknife-2, Jackknife-1, Chao and the corresponding degrees of completeness Ro/S (%)

P.N. Ecrins	S JK-2	S JK-1	S Chao	Ro/SJK2 according to JK-2	Ro/SJK1 according to JK-1	Ro/SChao according to Chao
Acaria	201	173	188	64%	74%	68%
Hemiptera	56	56	52	84%	84%	90%
Diptera	180	156	170	66%	76%	69%
Hymenoptera	48	50	46	92%	88%	96%
all 4 orders	485	435	437	69%	77%	77%

The numbers, Ro (= R(No)) of currently recorded species : Acaria Ro = 128, Hemiptera Ro = 47, Diptera Ro = 118, Hymenoptera Ro = 44, all four orders together Ro = 337



Figs. 3-6. Extrapolations of the species accumulation curve associated to each of the three nonparametric estimators Jackknife-2, Jackknife-1, Chao and also, a limited, unbiased extrapolation (see text for details)

The black dot corresponds to the realised sampling ($R_0, N_0 = 147$) beyond which the Species Accumulation Curve is extrapolated

Table 5. The predicted additional sampling effort (in % of the already performed sampling effort No = 147 sites) required for expecting 80 % and 90 % sampling completeness, according to the extrapolations of the Species Accumulation Curve respectively associated to each of the three nonparametric estimators Jackknife-2, Jackknife-1, Chao

P.N. Ecrins	80%	80%	80%	90%	90%	90%
	JK-2	JK-1	Chao	JK-2	JK-1	Chao
Acaria	+ 121%	+ 29%	+ 41%	+ 380%	+ 170%	+ 43%
Hemiptera	-	-	-	+ 58%	+ 58%	0%
Diptera	+ 110%	+ 20%	+ 36%	+ 349%	+ 138%	+ 90%
Hymenoptera	-	-	-	-	-	-
all 4 orders	+ 73%	+ 12%	+ 9%	+ 280%	+ 124%	+ 56%

Derived from this graphic data, Table 5 provides the predicted additional sampling efforts (in % of the already performed sampling effort No = 147 sites) required to reach 80% and 90% levels of sampling completeness. Each of the four orders are considered separately and also altogether for the whole inventory. In each case, the required additional sampling effort is computed according to each of the three types of extrapolation (associated to the three estimators Jackknife-2, Jackknife-1, Chao). Predicted extra-sampling efforts differ very strikingly according to the selected estimator, thus demonstrating *how critical is the choice of the appropriate type of estimator (and its associated extrapolation)*.

7. CONCLUSION

Among three of the most commonly used nonparametric estimators of total species richness, Chao, Jackknife-1, Jackknife-2, only the two Jackknife estimators comply with the required "rule of additivity" (Table 3). Among the latter, Jackknife-2 is preferred in general, while Jackknife-1 may be used only when samples are approaching completeness (section 3.1). Interestingly, Jackknife-2 has already been selected, quite independently, according to another strong theoretical argument: having an associated Species Accumulation Curve complying better than the two other estimators with a general mathematical relationship constraining the *theoretical expressions* of all Species Accumulation Curves [18].

Besides, as already pointed above, selecting Jackknife-2 (or Jackknife estimators of higher order) in general (i.e. apart from closely approaching sampling completeness, when extrapolation becomes of lesser interest) is in agreement with the general policy suggested successively by Brose et al. [19] and Reese et al. [27] on semi-empirical basis and also pointed by several authors on empirical basis [6,16,20-23].

Our own argumentations, fully established on theoretical grounds, thus provide a still more solid basis to the propositions of the preceding authors, i.e., the general relevance of Jackknife estimators (and their associated expressions for the extrapolation of the Species Accumulation Curve), while Chao and i-Chao estimators should preferably be restricted to the particular context already mentioned.

The illustrative examples provided above are representative of the importance of the issues involved by the extrapolation of species accumulation beyond incompletely achieved samplings: disagreements are very substantial, indeed, between the estimations of the number of missing species provided by each of the three estimators considered here (Tables 3 and 4). Still more important are the discrepancies between the predicted additional sampling efforts that are required for a given increase of sampling completeness, considering the extrapolations respectively associated to each of the three estimators (Table 5).

Thus, selecting the appropriate (less biased) type of estimator of the number of missing species (and of total species richness) and, by the way, selecting the more reliable extrapolation of the Species Accumulation Curve, are of *prime importance*. Since no reliable procedure of selection in this respect has ever been efficiently developed on an empirical basis, the need for a rational guidance based on a *solid theoretical basis* is obvious, as argued here.

COMPETING INTERESTS

Author has declared that no competing interests exist.

REFERENCES

1. Cam E, Nichols JD, Sauer JR, Hines JE. On the estimation of species richness

- based on the accumulation of previously unrecorded species. *Ecography*. 2002;25: 102–108.
2. Van Rooijen J. Estimating the snake species richness of the Santubong Peninsula (Borneo) in two different ways. *Contributions to Zoology*. 2009;78(4): 141-147.
 3. Gotelli NJ, Chao A. Measuring and estimating species richness, species diversity, and biotic similarity from sampling data. In: Levin SA, (ed.) *Encyclopedia of Biodiversity*, second edition. 2013;5:195-211. Waltham, MA: Academic Press.
 4. Colwell RK, Coddington JA. Estimating terrestrial biodiversity through extrapolation. *Philosophical Transactions of the Royal Society London B*. 1994;345:101-118.
 5. Thompson GG, Withers PC, Pianka ER, Thompson SA. Assessing biodiversity with species accumulation curves; inventories of small reptiles by pit-trapping in Western Australia. *Austral Ecology*. 2003;28:361–383.
 6. Martinez-Sanz C, Garcia-Criado F, Fernandez Alaez C, Fernandez-Alaez M. Assessment of richness estimation methods on macroinvertebrate communities of mountain ponds in Castilla y Leon (Spain). *Ann. Limnol. - Int. J. Limnol.* 2010;46:101–110.
 7. Poulin R. Comparison of three estimators of species richness in parasite component communities. *Journal of Parasitology*. 1998;84(3):485-490.
 8. Herzog SK, Kessler M, Cahill TM. Estimating species richness of tropical bird communities from rapid assessment data. *The Auk*. 2002;119(3):749-769.
 9. Chiarucci A, Enright NJ, Perry GLW, Miller BP, Lamont BB. Performance of nonparametric species richness estimators in a high diversity plant community. *Diversity and Distributions*. 2003;9:283-295.
 10. Fogo A, Attrill MJ, Frost MT, Rowden AA. Estimating marine species richness: An evaluation of six extrapolative techniques. *Marine Ecology Progress Series*. 2003; 248:15-26.
 11. Hortal J, Borges PAV, Gaspar C. Evaluating the performance of species richness estimators: Sensitivity to sample grain size. *Journal of Animal Ecology*. 2006;75:274-287.
 12. Unterseher M, Schnittler M, Dormann C, Sickert A. Application of species richness estimators for the assessment of fungal diversity. *FEMS Microbiology Letters*. 2008;282:205-213.
 13. Basualdo CV. Choosing the best non-parametric richness estimator for benthic macroinvertebrates databases. *Revista Sociedad Entomologica Argentina*. 2011; 70(1-2):27-38.
 14. Soberon MJ, Llorente BJ. The use of species accumulation functions for the prediction of species richness. *Conservation Biology*. 1993;7:480–488.
 15. Walter BA, Morand S. Comparative performance of species richness estimation methods. *Parasitology*. 1998;116:395-405.
 16. Hellmann JJ, Fowler GW. Bias, precision and accuracy of four measures of species richness. *Ecological Applications*. 1999; 9(3):824-834.
 17. 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
 18. Béguinot J. Extrapolation of the species accumulation curve for incomplete species samplings: A new nonparametric approach to estimate the degree of sample completeness and decide when to stop sampling. *Annual Research & Review in Biology*. 2015;8(5):1-9.
DOI: 10.9734/ARRB/2015/22351
 19. Brose U, Martinez ND, Williams RJ. Estimating species richness: Sensitivity to sample coverage and insensitivity to spatial patterns. *Ecology*. 2003;84(9): 2364-2377.
 20. Melo AS, Froehlich CG. Evaluation of methods for estimating macroinvertebrate species richness using individual stones in tropical streams. *Freshwater Biology*. 2001;46:711–721.
 21. Leponce M, Missa O, Delabie JHC. Performance of species richness estimators in a leaf-litter ant assemblage. *Colloque annuel, International Union for the Study of Social Insects (UIEIS)*; 2003. Bruxelles.
 22. Petersen FT, Meier R. Testing species-richness estimation methods on single-

- sample collection data using the Danish Diptera. *Biodiversity & Conservation*. 2003;12:667–686.
23. Marcon E. Mesures de la Biodiversité. Master, Kourou, France, 2015;HAL id:cel-01205813.
 24. Chiu CH, Wang YT, Walther BA, Chao A. An improved nonparametric lower bound of species richness via a modified Good-Turing frequency formula. *Biometrics*. 2014;70:3.
DOI: 10.1111/biom.12200
 25. Béguinot J. An algebraic derivation of Chao's estimator of the number of species in a community highlights the condition allowing Chao to deliver centered estimates. *International Scholarly Research Notices – Ecology*; 2014. Article ID 847328.
DOI: 10.1155/2014/847328
 26. Béguinot J. Investigations sur la faune cécidogène d'altitude dans le Parc National des Ecrins. *Bulletin mensuel de la Société Linnéenne de Lyon*. 2012; 81(5-6):87-116.
 27. Reese GC, Wilson KR, Flather CH. Performance of species richness estimators across assemblage types and survey parameters. *Global Ecology and Biogeography*. 2014;23:585-594.

APPENDIX

A1 – A general mathematical constraint applying to the expression of any *theoretical* Species Accumulation Curve (in practice, to any '*rarefied*' Species Accumulation Curve)

A relationship of general validity (detailed derivation provided in [17, 20]) links:

- on the one hand, the series of numbers, $f_{x(N)}$, of those species respectively recorded x -times in a sample of size N and,
- on the other hand, the successive derivatives, $\partial^x R_{(N)}/\partial N^x$, of the theoretical expression of the Species Accumulation Curve $R_{(N)}$, where $R_{(N)}$ is the number of recorded species in the sample of size N :

$$[\partial^x R_{(N)}/\partial N^x] = (-1)^{(x-1)} f_{x(N)}/C_{N,x} \approx (-1)^{(x-1)} (x!/N^x) f_{x(N)} \quad (A1.1)$$

Note that the curve associated to Jackknife-2 satisfies equation (A1.1) for both $x=1$ and $x=2$ (first and second derivatives), while Jackknife-1 complies only with equation (A1.1) only for $x=1$ (first derivative only), as demonstrated previously [18]. This is the reason why Jackknife-1 may provide a valid estimation within a restricted range only, i.e. when sample has reach a level of completeness not so far from completeness.

A2 - Derivation of the degree of deviation of Chao estimator from what is prescribed by the rule of additivity

The demonstration is conducted for an assemblage of species encompassing two mutually excluding categories of species, so that the whole sampling may relevantly be divided in two, well identified and separated sub-samples. The demonstration may easily be extended to any number of categories.

Let $\{g_1, g_2\}$ and $\{h_1, h_2\}$ be the numbers of species recorded once and twice respectively in each of the two sub-samples a_1 & a_2 under consideration.

- (i) *Deriving the condition ensuring no deviation, so that the estimated number of missing species in the whole sample equals the sum of the numbers of missing species in each of sub-samples*

The estimated numbers m and m' of missing species in each of the two sub-samples a_1 & a_2 are:

$$m = g_1^2/(2g_2) \quad m' = h_1^2/(2h_2)$$

and the estimated number M of missing species in the whole sample A is:

$$M = (g_1+h_1)^2/(2(g_2+h_2))$$

With $k_1 = g_1/h_1$ and $k_2 = g_2/h_2$, it comes:

$$M = (h_1^2/h_2).(k_1+1)^2/(2(k_2+1)) ;$$

$$m = (h_1^2/h_2).(k_1^2/(2 k_2))$$

Accordingly, the additivity rule, $M = m+m'$, leads to:

$$(k_1+1)^2/(2(k_2+1)) = \frac{1}{2} (k_1^2/k_2 + 1), \text{ that is: } k_1^2 + 2 k_1 + 1 = k_1^2 + k_2 + k_1^2/k_2 + 1, \text{ which yields:}$$

$$2 k_1 = k_2 + k_1^2/k_2 \text{ and thus: } (k_1 - k_2)^2 = 0$$

Therefore, the condition for Chao estimator respecting the required additivity rule is:

$$g_1/h_1 = g_2/h_2, \text{ or, as well: } g_1/g_2 = h_1/h_2$$

- (ii) *Evaluating the difference between the estimated number of missing species in the whole sampling and the sum of the numbers of missing species in each of sub-samplings* (for Table 1 and Fig. 2)

Let $\{g_1, g_2\}$ and $\{h_1, h_2\}$ be the numbers of species recorded once and twice, for each of the two considered sub-samples. In full generality, the sub-samples will generally differ in this respect. In the following, these difference are conveniently quantified by the ratios $\alpha = (g_1/g_2)/(h_1/h_2)$ and $\beta = g_1/h_1$. Then:

$$M = (h_1^2/(2 h_2))(1 + \beta)^2/(1 + \beta/\alpha)$$

$$m + m' = (h_1^2/(2 h_2))(1 + \beta.\alpha). \text{ Then:}$$

$$M - (m+m') = (h_1^2/(2 h_2))[(1 + \beta)^2/(1 + \beta/\alpha) - (1 + \beta.\alpha)]$$

$$M - (m+m') = (h_1^2/(2 h_2))[(1 + \beta^2 + 2\beta - 1 - \beta.\alpha - \beta/\alpha - \beta^2)/(1 + \beta/\alpha)]$$

$$M - (m+m') = (h_1^2/(2 h_2))[(2\beta - \beta.\alpha - \beta/\alpha)/(1 + \beta/\alpha)]$$

Likewise:

$$M + (m+m') = (h_1^2/(2 h_2))[(2 + 2\beta^2 + 2\beta + \beta.\alpha + \beta/\alpha)/(1 + \beta/\alpha)]$$

And, accordingly, the relative difference $D = |M - (m+m')|/[1/2 (M + (m+m'))]$

between M and $(m+m')$, considered in Table 1 and Fig. 2, is given by the following expression:

$$D = |M - (m+m')|/[1/2 (M + (m+m'))] = (4\beta - 2\beta.\alpha - 2\beta/\alpha)/(2 + 2\beta^2 + 2\beta + \beta.\alpha + \beta/\alpha).$$

A3 - Estimation of the evolutions of f_1 and f_2 (and thus of Jackknife-1 and Jackknife-2) with sample size N

According to [17,18]:

$$R(N) = (R(N_0) + 2f_1 - f_2) - (3f_1 - 2f_2) N_0/N - (f_2 - f_1) N_0^2/N^2 \quad (\text{A3-1})$$

$$f_1(N) = N.\partial R(N)/\partial N \quad (\text{A3-2})$$

$$f_2(N) = -1/2 N^2.\partial^2 R(N)/\partial N^2 \quad (\text{A3-3})$$

It follows:

$$f_1(N) = N.\partial R(N)/\partial N = (3f_1 - 2f_2) N_0/N + 2(f_2 - f_1) N_0^2/N^2 \quad (\text{A3-4})$$

$$f_2(N) = -1/2 N^2.\partial^2 R(N)/\partial N^2 = (3f_1 - 2f_2) N_0/N + 3(f_2 - f_1) N_0^2/N^2 \quad (\text{A3-5})$$

$$JK-2(N) = 2f_1(N) - f_2(N) = (3f_1 - 2f_2) N_0/N + (f_2 - f_1) N_0^2/N^2 \quad (\text{A3-6})$$

© 2016 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/13904>