



## A critique of the use of jackknife and related non-parametric techniques to estimate species richness

A. S. Melo

*Programa de Pós-Graduação em Ecologia, Museu de História Natural, IB, Universidade Estadual de Campinas, 13083-970 Campinas, SP, Brazil and Departamento de Biologia Geral, ICB, Universidade Federal de Goiás, 74001-970 Goiânia, GO, Brazil. Present address: Departamento de Ecologia, IB, Universidade Federal do Rio Grande do Sul, Porto Alegre 91540-000, RS, Brazil. Fax: 55 51 3316-7626, E-mail: adrimelo@ecologia.ufrgs.br*

**Keywords:** Community ecology, Diversity, Log-series, Rare species, Sample size, Species accumulation curve.

**Abstract.** Species richness in an assemblage is frequently used as a measure of biological diversity. However, observed species richness is strongly dependent on sample size. If more samples are collected, then more species are observed. Non-parametric species richness estimators, such as the jackknife 1 and 2 and the Chao 1 and 2, are indicated in the literature as potential solutions to the problem of dependence of observed species richness on sampling effort. These methods are intended to estimate the total species richness in an area or assemblage with small sampling effort. Non-parametric estimators are based on the number of species observed, and the number of rare species in a sample, i.e., that occurred in one and/or two sampling units, or with one and/or two individuals. High estimates are produced when samples contain large proportions of rare species. Using a range of real datasets, I show that estimates produced by non-parametric methods are generally dependent on observed species richness. An implicit assumption of these non-parametric techniques is that the rare species curve should present high values at small sample sizes and decreasing values as sampling effort is increased. This assumption was observed in only one out of eight datasets presented. Instead, the rare species curve generally flattens off around a constant value as sampling effort increases. I conclude that non-parametric estimators are not reliable to estimate species richness in an assemblage when the rare species curve does not show a decreasing trend. Comments are made on the possibilities of using non-parametric estimators in the comparison of species assemblages.

### Introduction

Researchers are increasingly looking for new tools in order to understand ecological systems and to improve conservation efforts to save biological diversity. Included in these tools are the protocols for rapid assessment of biodiversity (Coddington et al. 1991), the use of morphospecies and higher taxa as a surrogate for species diversity (Oliver and Beattie 1996), the selection of indicator taxa (Brown and Freitas 2000), and improved techniques to estimate species richness in a given area or assemblage (Palmer 1990, Colwell and Coddington 1994, Patil and Taillie 2001). All these techniques are intended to guide the selection of reserves, by optimizing the use of scarce funds to save the greatest number of species, endemic or threatened taxa, key taxa in the ecosystem functioning, and unique ecosystems.

In this sense, species richness estimators are thought to be a valuable technique, as they would estimate diversity in a given assemblage with small sampling effort (Palmer 1990, Colwell and Coddington 1994). Differ-

ently from the number of observed species in a survey, which is generally dependent on sample size, estimated species richness is expected to be independent of sample size. Fewer samples than what would be required by counting observed species would produce good estimates of species richness in an area (Gotelli and Colwell 2001), thus saving time and money. Figure 1 depicts an estimated curve that would be produced by an ideal method as well as the corresponding observed species accumulation curve in function of increasing sample size. While the curve of observed accumulated species richness increases slowly with sampling effort, the ideal estimator would produce values close to the actual species richness in the area using small sample sizes and then flatten off.

Species richness estimators were reviewed by Colwell and Coddington (1994), who present the several techniques currently available in an easy way to biologists. They distinguished three classes of estimators, namely Extrapolations of Species Accumulation Curves, Parametric Estimators, and Non-parametric Estimators. Included in the first class are several asymptotic functions,

**Table 1.** Non-parametric estimators of species richness in an area.  $S_{obs}$  = observed species richness.  $Q_1$  = number of species which occurred in exactly one sampling unit.  $Q_2$  = number of species which occurred in exactly two sampling units.  $n$  = number of sampling units.  $F_1$  = number of species which occurred with exactly one individual.  $F_2$  = number of species which occurred with exactly two individuals.

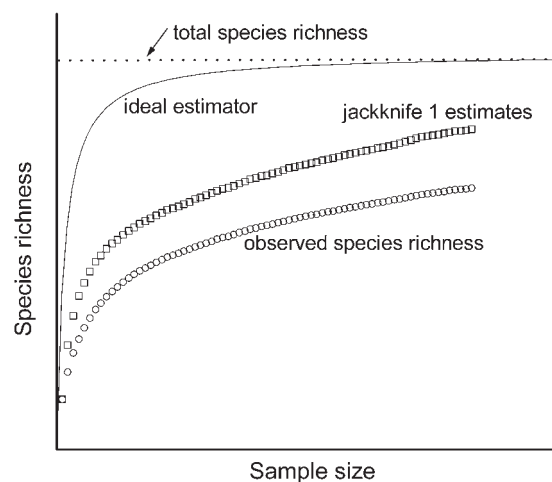
| Estimators  | Formulae   | References   |
|-------------|--|--|
| Jackknife 1 | $S_{Jack1} = S_{obs} + Q_1 \left( \frac{n-1}{n} \right)$                               | Burnham and Overton 1978,<br>Colwell and Coddington 1994 |
| Jackknife 2 | $S_{Jack2} = S_{obs} + \left( \frac{Q_1(2n-3)}{n} - \frac{Q_2(n-2)^2}{n(n-1)} \right)$ | Burnham and Overton 1978,<br>Colwell and Coddington 1994 |
| Chao 1      | $S_{Chao1} = S_{obs} + \frac{F_1^2}{2F_2}$   | Chao 1984,<br>Colwell and Coddington 1994                |
| Chao 2      | $S_{Chao2} = S_{obs} + \frac{Q_1^2}{2Q_2}$   | Chao 1987,<br>Colwell and Coddington 1994                |

such as the Michaelis-Menten hyperbole of enzyme kinetics used by biochemists (Keating and Quinn 1998). Parametric estimators are based on the assumption that biological data follow a specific distribution, such as the lognormal. Besides having no assumption regarding data distribution, the non-parametric methods including jackknife 1, jackknife 2, Chao 1, and Chao 2 (Table 1) are easily computed. They involve summing the number of species already observed in a sample and a second term

related to the proportion of observed species that were rare in the sample (Table 1). High species richness estimates are obtained when non-parametric techniques are employed on samples with high proportion of rare species.

As previously outlined, a fundamental requirement of species richness estimators is relative independence of sample size. Thus, they should be able to produce reliable values even when using small sampling efforts. However, Colwell and Coddington (1994) discuss limitations of using non-parametric estimators on small sample sizes. They point out that jackknife estimators attain their plateau values at approximately twice the observed richness, while Chao estimators at about half the square of the observed number of species. As a consequence, Colwell and Coddington (1994, p. 111) predict that "...these estimators should correlate strongly with sample size until half (or the square root of twice) the total fauna is observed and thereafter become gradually independent of sample size until finally the observed richness and the estimate converge." In fact, in a recent evaluation of species richness estimators, we show that estimated richness curves do not attain an asymptote early (Melo and Froehlich 2001a). Instead, they follow the observed species accumulation curve in a quite regular way, estimating values in a fixed proportion above the observed richness along most of the increasing sampling effort (Fig. 1). Similar results showing the dependence of estimates on the observed species curve are provided by Condit et al. (1996) and Fisher (1999a).

Here, I empirically explore the dependence of estimators on observed species richness. Comments are made on the usefulness of using non-parametric estimators to pre-



**Figure 1.** Observed species accumulation curve, the respective jackknife 1 curve, and a hypothetical ideal estimate curve. The jackknife 1 curve follows the observed species curve in a regular way, increasing slowly with sample size. On the other hand, the hypothetical ideal estimator produce richness estimates around the total number of species in the area using small sample size and then flatten off.

**Table 2.** Summary of the datasets used to investigate the dependence of non-parametric richness estimates on observed species richness. All localities are in Brazil.

|                             | Stream<br>macroinvertebrates<br>local         | Stream<br>macroinvertebrates<br>regional      | Spiders                     | Trees                              | <i>Drosophila</i> spp.             | Harvestmen<br>(Opiliones)  | Frogs                               |
|-----------------------------|---|---|-----------------------------|------------------------------------|------------------------------------|----------------------------|-------------------------------------|
| Locality                    | Jundiaí,<br>São Paulo                         | Iporanga,<br>São Paulo                        | Linhares,<br>Espírito Santo | Campinas,<br>São Paulo             | Barreiro Rico,<br>São Paulo        | Ubatuba,<br>São Paulo      | Ilha de São Sebastião,<br>São Paulo |
| Geographical<br>Coordinates | 23° 14'S,<br>46° 56'W                         | 24° 18'S,<br>48° 25'W                         | 19° 10'S,<br>40° 05'W       | 22° 49'S,<br>47° 07'W              | 22° 40'S,<br>48° 10'W              | 23° 26'S,<br>45° 04'W      | 23° 47'S,<br>45° 24'W               |
| Vegetation                  | Tropical Semi-Deciduous<br>Montane Forest     | Atlantic Rain<br>Forest                       | Atlantic Rain<br>Forest     | Tropical Semi-<br>Deciduous Forest | Tropical Semi-<br>Deciduous Forest | Atlantic Rain<br>Forest    | Atlantic Rain<br>Forest             |
| Sampling units              | stones (15-20 cm. diam.)<br>in stream riffles | stones (15-20 cm.<br>diam.) in stream riffles | time intervals              | 10 x 10 m<br>contiguous plots      | traps using<br>fermented bananas   | 8 x 8 m plots on<br>litter | 8 x 8 m plots on<br>litter          |
| Sample size                 | 75 (1 stream site)                            | 475 (10 streams inside<br>a catchment)        | 243                         | 100                                | 180                                | 63                         | 92                                  |
| Species richness            | 66  | 162   | 287                         | 101                                | 57                                 | 40                         | 15                                  |
| Individuals                 | 3759  | 17,598  | 1982                        | 1465                               | 8166                               | 764                        | 846                                 |

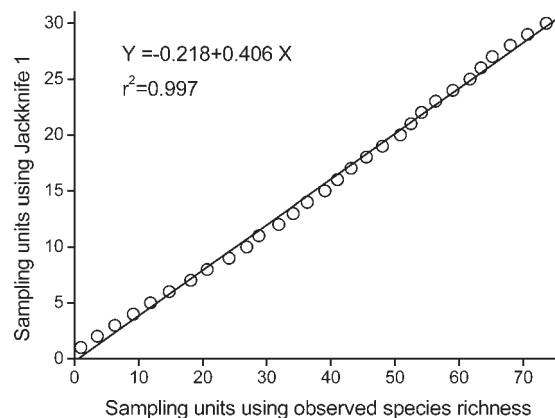
dict species richness in an area and in the comparisons of species assemblages.

**An empirical relationship among sample sizes needed to observe and to estimate a given richness value**

A striking relationship of dependence of richness estimates on observed values was obtained by plotting the number of sampling units needed to observe a given richness value, when constructing a species accumulation curve, and the number of sampling units needed to estimate the same value using a non-parametric estimator. Figure 2 shows such a relationship using the jackknife 1 estimator and a dataset of macroinvertebrates living on stream stones (Table 2, local dataset, Melo and Froehlich 2001a). Notice that dots in Fig. 2 do not depict richness values. Richness values were used only to match the corresponding number of sampling units in which the same number of species can be obtained from the cumulative observed list and from estimates of the jackknife 1. As the jackknife estimates may be non-integer values, a precise match between the two metrics was achieved by using the species richness estimated by the first order jackknife for each cumulative number of sampling units (y-axis) and the corresponding interpolated x-value (Fig. 2).

The coefficient of determination obtained from the relationship depicted in Fig. 2 is very high ( $r^2 = 0.997$ ), demonstrating that the sample size needed to estimate a given richness value is a linear function of the sample size needed to observe the same value from a species accumulation curve. The relationship is so strongly linear that the

extrapolation of the fitted linear regression in Fig. 2 can be used as richness estimator in a larger sample size. In fact, such estimates of species richness for extrapolated sample sizes are very reliable (Melo et al. 2003).



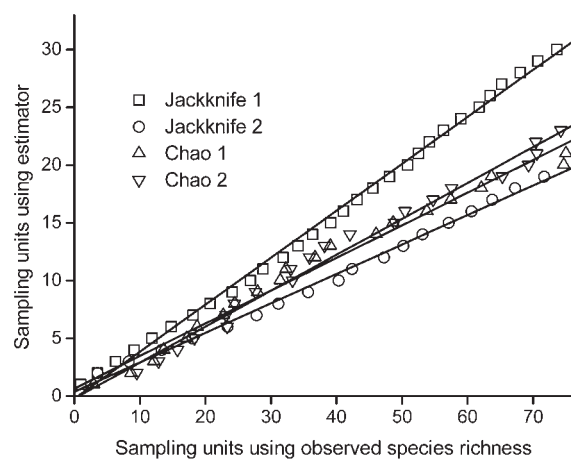
**Figure 2.** Relationship between the number of sampling units required to observe a given species richness and the number of sampling units required to estimate the same species richness using the jackknife 1 estimator. A precise match was obtained by using jackknife 1 species richness estimates for each cumulative number of sampling units and the corresponding interpolated value needed to observe the same richness in a species accumulation curve. Observed species richness in a given sample size can be obtained using the jackknife 1 estimator over a nearly fixed proportion of the same sample size, represented by the inclination of the linear fit. For the dataset used, this proportion is around 40%. Data are from stream macroinvertebrates occurring in 75 sampling units (stones) collected in a stream site (local dataset in Table 2).

**Table 3.** Parameters of the linear model and the coefficient of determination obtained by fitting the relationship of the number of sampling units needed to observe a given richness value and the number of sampling units needed to estimate the same value using a non-parametric estimator.

| Dataset                             | Jackknife1 |       |                | Jackknife 2 |       |                | Chao 1 |       |                | Chao 2  |       |                |
|-------------------------------------|------------|-------|----------------|-------------|-------|----------------|--------|-------|----------------|---------|-------|----------------|
|                                     | a          | b     | r <sup>2</sup> | a           | b     | r <sup>2</sup> | a      | b     | r <sup>2</sup> | a       | b     | r <sup>2</sup> |
| Stream macroinvertebrates, regional | -2.876     | 0.388 | 0.998          | -2.891      | 0.244 | 0.995          | 6.970  | 0.305 | 0.994          | -1.038  | 0.338 | 0.993          |
| Spiders                             | 1.899      | 0.393 | 0.999          | 2.168       | 0.239 | 0.999          | -4.008 | 0.342 | 0.998          | -19.698 | 0.378 | 0.899          |
| Trees                               | 2.400      | 0.329 | 0.992          | 2.391       | 0.200 | 0.990          | -3.103 | 0.338 | 0.981          | -6.200  | 0.310 | 0.931          |
| Stream macroinvertebrates, local    | -0.218     | 0.406 | 0.997          | 0.374       | 0.255 | 0.994          | 0.643  | 0.283 | 0.984          | -0.203  | 0.311 | 0.985          |
| <i>Drosophila</i> spp.              | 0.833      | 0.400 | 0.999          | 1.658       | 0.244 | 0.996          | 3.885  | 0.163 | 0.967          | 3.457   | 0.168 | 0.969          |
| Harvestmen ( <i>Opiliones</i> )     | 1.505      | 0.299 | 0.993          | 2.226       | 0.154 | 0.983          | 1.200  | 0.212 | 0.981          | 0.912   | 0.089 | 0.929          |
| Frogs                               | -1.134     | 0.430 | 0.987          | -1.043      | 0.288 | 0.949          | 0.051  | 0.436 | 0.989          | -4.400  | 0.390 | 0.932          |

The relationship shown in Fig. 2 is easily extended to other non-parametric estimators. Figure 3 shows the relationship using jackknife 1, jackknife 2, Chao 1, and Chao 2 for the same stream macroinvertebrates dataset. Notice that, when using different non-parametric estimators, differences are greatly restricted to the inclination of the linear relationship. For the dataset used, jackknife 2 estimator is able to produce a given observed richness value using the smallest sample size, while the jackknife 1 provides the same when using the largest sample size.

I further constructed the relationship depicted in Fig. 2 to a range of other datasets (Table 2). The datasets com-



**Figure 3.** The same relationship shown in Fig. 2 extended to other non-parametric estimators. Jackknife 2 estimates a given richness value using the smallest proportion of the sample size needed to observe the same given richness value, while the jackknife 1 uses the largest proportion. Data used are from stream macroinvertebrates occurring in 75 sampling units (stones) collected in a stream site (local dataset in Table 2). Parameters of the linear fit and the determination coefficient are shown in Table 3.

prise different taxa, species richness, sampling effort, and data structure. As seen using the stream invertebrate dataset, there were strong correlations between the number of sampling units required to estimate and to observe a given species richness value (Table 3). Better correlations were observed for jackknife 1 and 2 estimators rather than for Chao 1 and 2 estimators.

A practical observation of the relationship depicted in Fig. 2 is obtained by comparing the results of Hellmann and Fowler (1999) and Melo and Froehlich (2001a). Hellmann and Fowler (1999) used five datasets of plants with different species richness and proportions of rare species. They found that for the jackknife1 and the jackknife 2, the sub-sample sizes needed to estimate actual species richness in the total samples were respectively, 36.8-43.9% and 22.6-29.1% of total samples. Melo and Froehlich (2001a) used six datasets of stream macroinvertebrates comprising different locations and spatial scales (local = one stream site, and regional = several sites inside a same catchment basin). They observed that jackknife 1 and jackknife 2 estimated actual richness in the total samples using subsamples of sizes 35.6-41.3% and 22.4-26.7% of the total samples, respectively.

### Non-parametric estimates and the number of rare species in a sample

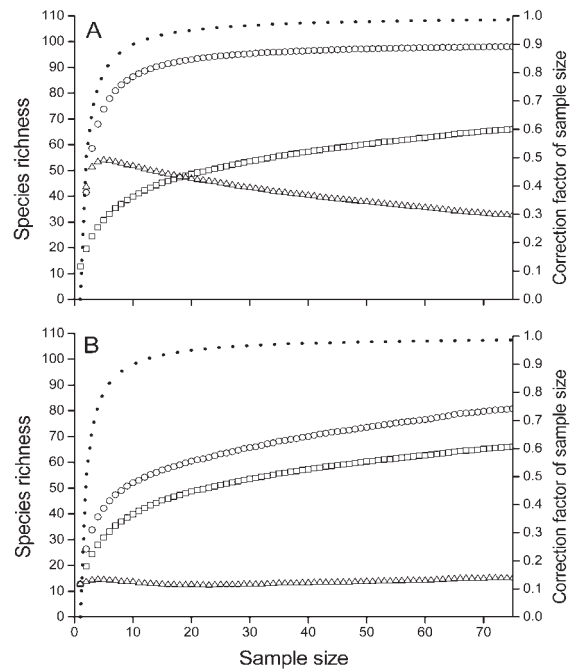
An implicit assumption to obtain the ideal estimated curve in Fig. 1 is that the number of rare species (i.e., those occurring in/with 1 and or 2 sampling units/individuals) is high in small samples and decreases as sampling effort increases. The assumption is similar to the process to obtain a lognormal distribution from a truncated lognormal distribution (Magurran 1988, p. 25). As sample size is increased, the octave class containing rare species is decreased until the situation where all species

once rare become common – the non-truncated lognormal distribution. Figure 4A illustrates a hypothetical example using a jackknife 1 estimate curve and its three components, (1) the observed species richness curve, (2) the rare species curve (i.e., the number of species observed in 1 sampling unit), and (3) the correction factor for sample size (i.e.  $[n-1]/n$ , where  $n$  is the number of sampling units). For each sample size, jackknife 1 estimate is obtained by multiplying the number of rare species by the correction factor and then summing the obtained value with the observed species richness (Table 1). Observed species richness curve in Fig. 4A was obtained from the stream macroinvertebrates dataset (Table 2, local dataset), while the rare species curve was obtained mathematically in order to produce the hypothetical jackknife 1 curve. Notice that this hypothetical rare species curve required in order to produce the ideal estimator is unreal, as at small sample sizes the number of rare species is higher than the number of observed species. The asymptote of the hypothetical jackknife 1 curve was chosen arbitrarily, but it is in the range of species richness commonly found in other similar streams in the region (Melo and Froehlich 2001a,b).

For the same dataset from which the observed species richness curve in Fig. 4A was obtained, Fig. 4B shows the jackknife 1 curve using the actual rare species curve. The actual rare species curve differs from the ideal rare species curve in two ways. The actual rare species curve does not decrease linearly as sample size increases and the absolute number of rare species is low when compared to the ideal rare species curve. Thus, after around 10 sampling units are collected, the correction factor of sample size tends to flatten off close to the unity and the jackknife 1 curve becomes mostly the sum of the observed species richness with a practically constant number of rare species. Figure 5 shows curves for number of rare species (occurring in only one sampling unit), observed species richness, and the corresponding jackknife 1 estimates for other six datasets. Despite the range of assemblage types, sample effort, and taxa, the curves of the numbers of rare species in all six datasets do not present a clear trend of decrease as sample size increase. Instead, the curves tend to flatten off around a constant number of rare species. A possible exception could be that from forest trees, which presented a slight decreasing trend.

#### Situations where non-parametric methods will likely produce reliable estimates

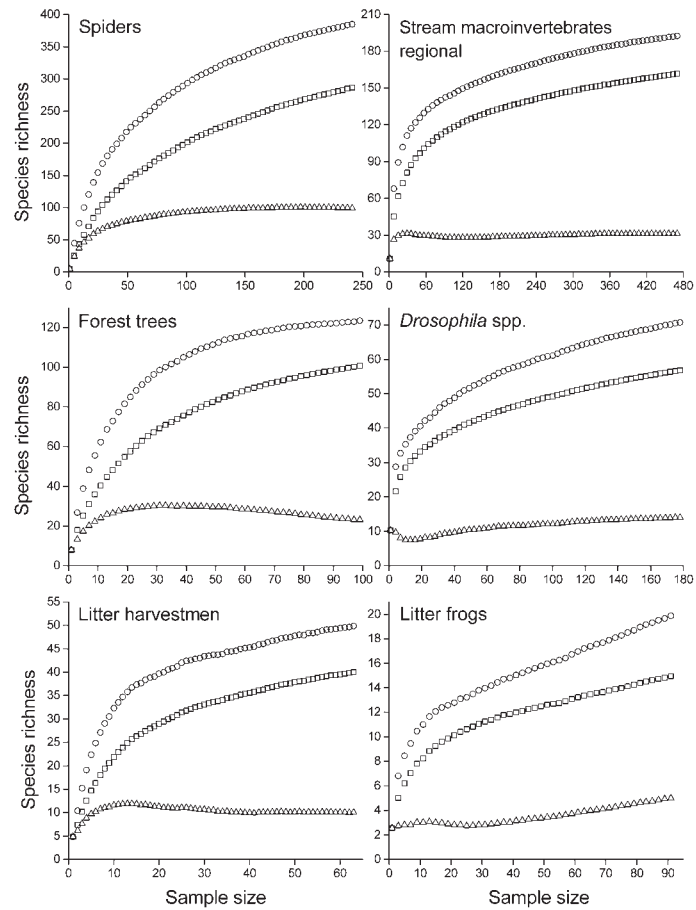
I have shown above empirical datasets for which non-parametric methods failed to produce reliable estimates. However, we may wonder if non-parametric estimators will always produce poor estimates. I argued that poor



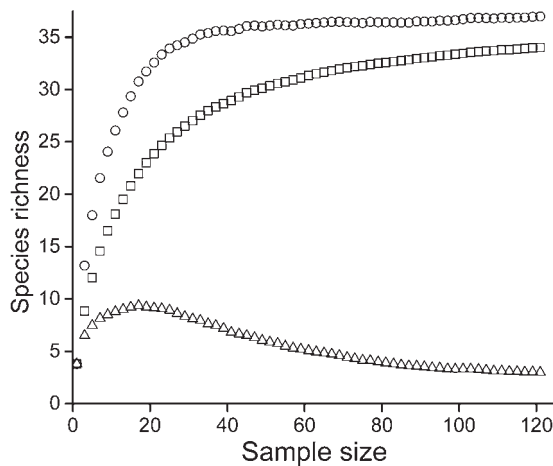
**Figure 4.** Jackknife 1 curve and its three components in desired hypothetical (A) and real (B) situations.  $\circ$  = Jackknife 1 curve.  $\square$  = Observed species accumulation curve.  $\triangle$  = Rare species curve represented by the number of species which occurred in exactly one sampling unit.  $\dots$  = Correction factor of sample size ( $[n-1]/n$ , where  $n$  is sample size). Jackknife 1 estimates are obtained by multiplying the number of rare species by the correction factor and then summing the obtained value with the observed species richness. Observed species richness curves were obtained from the stream macroinvertebrates dataset (Table 2, local dataset). Curve of rare species in (A) was obtained mathematically in order to produce the hypothetical jackknife 1 curve. Curve of rare species in (B) was obtained from the stream macroinvertebrates dataset. The real rare species curve (B) is low and does not decrease with sample size, as would be expected in order to produce the jackknife 1 estimates in the hypothetical situation (A). Notice that the ideal rare species curve in (A) is unreal at small sample sizes, as the number of rare species is higher than the number of observed species.

performance of non-parametric estimators is a consequence of the non-decreasing behavior of the rare species curves as sampling effort increases. We may thus look for datasets in which the rare species curve does show a decreasing behavior along with increased sample sizes. A good example of this situation is the seed bank dataset presented by Colwell and Coddington (1994) and available by downloading the EstimateS software (Colwell 1997). The dataset contains 34 species and 121 sample units and is derived from a diversity study of the soil seed bank in a tropical forest (Butler and Chazdon 1998). In contrast to the rare species curves (uniques) presented in

**Figure 5.** Jackknife 1 curve and two of its component curves for six real datasets described in Table 2. ○ = Jackknife 1 curve. □ = Observed species accumulation curve. △ = Rare species curve represented by the number of species which occurred in exactly one sampling unit. Rare species curves do not decrease with sample size, as would be expected in order to produce the ideal situation depicted in Fig. 4A.

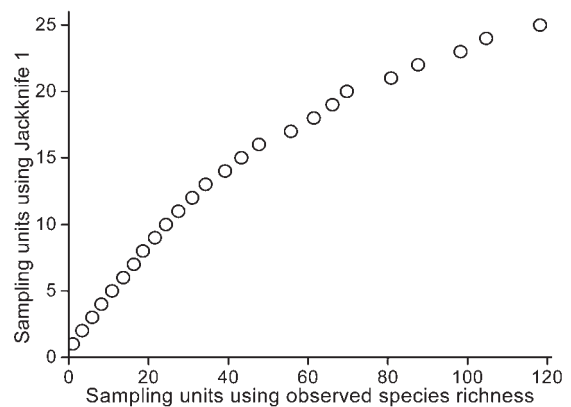


Figures 4A and 5, the curve for rare species in the seed bank dataset shows a distinct decreasing trend after 20 sampling units are pooled (Fig. 6). As predicted above, the jackknife 1 curve behaves well for the dataset, increasing steeply until 45 sample units are pooled and then stabilizing around the value of 36 species. I further con-



**Figure 6.** Observed species accumulation curve (□), rare (unique) species curve (△), and the respective jackknife 1 curve (○) obtained from the seedbank dataset, composed of 34 species distributed in 121 sampling units.

structed the relationship depicted in Fig. 2 to the seed bank dataset (Fig. 7). Instead of the linear relationship observed in Fig. 3 and Table 3, the seed bank dataset produced a linear relationship until ca. 45 sampling units are



**Figure 7.** Relationship between the number of sampling units required to observe a given species richness and the number of sampling units required to estimate the same species richness using the jackknife 1 estimator. The construction of the relationship is explained in the legend to Fig. 2. Data used are from the seedbank dataset, composed of 34 species distributed in 121 sampling units.

pooled, and then the slope decreases. For sample sizes larger than 45 sampling units, the increase in richness estimate produced by the jackknife 1 when one sampling unit is added is matched by the species observed curve using a larger number of sampling units.

## Discussion

The assumption that the number of rare species decreases as sampling size increases was observed in the intentionally selected seed bank dataset, weakly supported in the forest tree dataset, and clearly not supported in the remaining six datasets presented. For the last six datasets, the number of rare species increased at small sample sizes and then tended to flatten off. This finding is not completely unexpected, as it is predicted in the Log-Series distribution of Fisher et al. (1943). In this distribution, the number of species is infinite and the expected number of species with one individual is given by multiplying the parameters  $x$  and  $\alpha$ , where  $x$  varies from 0 to 1 and  $\alpha$  is known as the diversity index of the distribution (Magurran 1988, p. 133). When the ratio number of individuals by number of species ( $N/S$ ) is large,  $x$  tends to the unity and if for instance the ratio is 20,  $x$  is around 0.99. Thus, for large  $N/S$  ratios the expected number of species with one individual is nearly equal to the  $\alpha$  value. It has been found in a number of studies that the  $\alpha$  value is constant as sample size increases (Taylor et al. 1976, Condit et al. 1996), and this independence of sample size is considered a good property of this diversity index (Southwood 1978, Magurran 1988). Further support to the constancy of the number of rare species as sample size increases is observed in the extensive collections of deep-sea invertebrates (Grassle and Maciolek 1992), leaf-litter ants (Fisher 1999b), herbivorous insects in host plants (Novotný and Basset 2000), and spiders (Toti et al. 2000). It is noteworthy to observe that sample sizes in most datasets in Table 2 are not small. For example, the sample size of the stream macroinvertebrates dataset (local) is 2-3 times larger than the sample size usually employed to assess diversity in a stream site (Stout and Vandermeer 1975, Minshall et al. 1985, Melo and Froehlich 2001b).

The above results about the constant number of rare species in a sample imply that non-parametric estimates will be simply the sum of observed species richness and a nearly constant value. Indeed, recent evaluations of species richness estimators have shown that their accuracy is strongly dependent on sample size (Condit et al. 1996, Wagner and Wildi 2002). The goal of estimating the number of species *accurately* in an assemblage is illusive unless sample size is so large that the rare species curve starts to decrease. In this situation, the usefulness of non-

parametric estimators becomes doubtful, as the researcher may have a good estimate of species richness in the area by simply using the number of species already sampled. This might be the case for the seed bank dataset explored above. The jackknife 1 curve for the seed bank dataset stabilized at 36 species after ca. 45 sampling units were pooled. For this sample size, the observed species richness curve produced values around 30 species. Whether the increase from 30 to 36 species is relevant will depend on the question being evaluated and on the user's judgment.

In the same line of reasoning, Chazdon et al. (1998, p. 305) state that estimators will fail when the number of rare species continues to remain high as new quadrats are sampled. Implicit in their argument is the idea that in many cases the number of rare species decreases as sample sizes are increased. Indeed, this has been observed in intensively sampled datasets, as the seed bank used above (Colwell and Coddington 1994, Butler and Chazdon 1998) and others available in the literature (Bini et al. 2001, Walther and Martin 2001). As pointed out above, in these cases researchers should judge whether the improvement produced by a non-parametric method is counterbalanced by the use of a simpler estimate, the number of species observed.

It is likely that many rare species in a sample are actually vagrant species, collected accidentally, or with different habitat requirements and that only occasionally occur in the habitat sampled (Magurran and Henderson 2003). After removing these species from datasets it is possible that, as sampling effort is increased, the rare species curve would tend to decrease, improving non-parametric richness estimates. Longino et al. (2002) provide a good example in which it was possible to examine rare species carefully and to judge whether they belonged to the studied assemblage. However, given the current poor knowledge of the natural history of most species in rich assemblages, especially in the tropics, this would be arbitrary in most cases.

An additional problem in the estimation of species richness using non-parametric estimator based on incidence (presence or absence in sample units), such as the jackknife 1 and 2 and Chao 2, is the effect of sample unit size on the curve of rare species. For the same total area sampled, the size of the sample units used will affect the curve of rare species. A species occurring with several individuals in a single patch may be fully sampled by a large sample unit and in this case would be considered a rare species. In contrast, if the size of the sample unit is small, individuals might be sampled in three sample units, and in this case would not be considered a rare species. Con-

sequently, for the same assemblage and total sample effort, different results will be produced depending on the size of the sample units (Chiarucci et al. 2003).

A more balanced interpretation of values produced by current methods is that they represent an estimate of the minimum number of species in the community (Longino et al. 2002). In this sense, estimates would be preferable to observed richness as they would be less negative biased. A potential usefulness of such interpretation, yet to be evaluated, is in comparisons of different assemblages (Walther and Martin 2001). Stout and Vandermeer (1975) showed that previous beliefs that temperate streams were richer in insect species than their tropical counterparts were due to insufficient sampling in the latter. Extrapolating a species accumulation curve with an asymptotic function, they showed that tropical streams are in fact richer than temperate streams. However, this becomes clear only after a large sample has been collected. This is because species rich assemblages generally comprise a large number of species with patchy distribution and a great number of rare species. Stout and Vandermeer (1975) argue that methods for estimating the species pool could potentially provide better comparative grounds, as they would produce high values for assemblages with high numbers of rare species. In this sense, despite the strong dependency of non-parametric estimators on observed species richness, it is likely that estimated richness would provide a better comparative basis as they are less negatively biased than the observed species richness. Further studies addressing specifically this question are necessary to confirm this speculation.

While this work is restricted to non-parametric estimators, it is possible that other currently available estimators of species richness in assemblages are also strongly dependent on the observed richness. This speculation is based on the similarity in behavior of non-parametric and other estimators in recent evaluations (Melo and Froehlich 2001a). If this speculation is found to be true in future evaluations, the question remains on the feasibility of estimating species richness in an area or assemblage using small samples.

**Acknowledgements.** I am deeply grateful to Ricardo J. Sawaya, Adalberto J. Santos, George J. Shepherd, Glauco Machado, and Hermes F. Medeiros for gently allowing the use of the datasets that strengthened this work and for insightful discussions on the subject. Ellen Wang reviewed the first version of the manuscript. Claudio G. Froehlich, Glauco Machado, Jorge L. Nessimian, José A. F. Diniz-Filho, Luis M. Bini, Robert K. Colwell, Rodrigo A. S. Pereira, Thomas M. Lewinsohn, Valerio P. Pillar, Virginia S. Uieda, and one anonymous referee made useful comments on the manuscript. I received fellowships and a grant from Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP) (proc. no. 98/00590-0, 02/12538-0, and 03/10936-1) and a fellowship from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (proc. no. 302175/2002-4).

This work is part of the BIOTA/FAPESP - The Biodiversity Virtual Institute Program ([www.biota.org.br](http://www.biota.org.br)).

## References

- Bini, L.M., S.M. Thomaz and D.C. Souza. 2001. Species richness and  $\beta$ -diversity of aquatic macrophytes in the Upper Paraná River floodplain. *Arch. Hydrobiol.* 151:511-525.
- Brown, K.S.Jr. and A.V.L. Freitas. 2000. Atlantic forest butterflies: indicators for landscape conservation. *Biotropica* 32:934-956.
- Burnham, K.P. and W.S. Overton. 1978. Estimation of the size of a closed population when capture probabilities vary among animals. *Biometrika* 65:625-633.
- Butler, B.J. and R.L. Chazdon. 1998. Species richness, spatial variation, and abundance of the soil seed bank of a secondary tropical rain forest. *Biotropica* 30:214-222.
- Chao, A. 1984. Nonparametric estimation of the number of classes in a population. *Scand. J. Stat.* 11:265-270.
- Chao, A. 1987. Estimating the population size for capture-recapture data with unequal catchability. *Biometrics* 43:783-791.
- Chazdon, R.L., R.K. Colwell, J.S. Denslow and M.R. Guariguata. 1998. Statistical methods for estimating species richness of woody regeneration in primary and secondary rain forest of northeastern Costa Rica. In: F. Dallmeier and J. A. Comiskey (eds.), *Forest Biodiversity Research, Monitoring and Modeling: Conceptual Background and Old World Case Studies*. Parthenon Publishing, Paris. pp. 285-309.
- Chiarucci, A., N.J. Enright, G.L.W. Perry, B.P. Miller and B.B. Lamont. 2003. Performance of nonparametric species richness estimators in a high diversity plant community. *Div. Distrib.* 9: 283-295.
- Coddington, J.A., C.E. Griswold, D.S. Dávila, E. Peñaranda and S.F. Larcher. 1991. Designing and testing sampling protocols to estimate biodiversity in tropical ecosystems. In: E. C. Dudley (ed.), *The Unity of Evolutionary Biology: Proceedings of the Fourth International Congress of Systematics and Evolutionary Biology*. Dioscorides Press, Portland. pp. 44-60.
- Colwell, R. K. 1997. EstimateS: Statistical estimation of species richness and shared species from samples. Version 5. User's Guide and application published at: <http://viceroy.eeb.uconn.edu/estimates>
- Colwell, R.K. and J.A. Coddington. 1994. Estimating terrestrial biodiversity through extrapolation. *Philos. Trans. R. Soc. Lond. B* 345:101-118.
- Condit, R., S.P. Hubbell, J.V. Lafrankie, R. Sukumar, N. Manokaran, R.B. Foster and P.S. Ashton. 1996. Species-area and species-individual relationships for tropical trees: a comparison of three 50-ha plots. *J. Ecol.* 84:549-562.
- Fisher, R.A., A.S. Corbet and C.B. Williams. 1943. The relation between the number of species and the number of individuals in a random sample of an animal population. *J. Anim. Ecol.* 12:42-58.
- Fisher, B.L. 1999a. Ant diversity patterns along an elevational gradient in the Réserve Naturelle Intégrale d'Andohahela, Madagascar. *Fieldiana Zoology (n.s.)* 94:129-147.
- Fisher, B.L. 1999b. Improving inventory efficiency: a case study of leaf-litter ant diversity in Madagascar. *Ecol. Appl.* 9:714-731.
- Gotelli, N.J. and R.K. Colwell. 2001. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecol. Lett.* 4:379-391.



- Grassle, J.F. and N.J. Maciolek. 1992. Deep-sea species richness: regional and local diversity estimates from quantitative bottom samples. *Am. Nat.* 139:313-341.
- Hellmann, J.J. and G.W. Fowler. 1999. Bias, precision, and accuracy of four measures of species richness. *Ecol. Appl.* 9:824-834.
- Keating, K.A. and J.F. Quinn. 1998. Estimating species richness: the Michaelis-Menten model revisited. *Oikos* 81:411-416.
- Longino, J.T., J. Coddington and R.K. Colwell. 2002. The ant fauna of a tropical rain forest: estimating species richness three different ways. *Ecology* 83:689-702.
- Magurran, A.E. 1988. *Ecological Diversity and its Measurement*. Croom Helm, London.
- Magurran, A.E. and P.A. Henderson. 2003. Explaining the excess of rare species in natural species abundance distributions. *Nature* 422:714-716.
- Melo, A.S. and C.G. Froehlich. 2001a. Evaluation of methods for estimating macroinvertebrates species richness using individual stones in tropical streams. *Freshwater Biol.* 46:711-721.
- Melo, A.S. and C.G. Froehlich. 2001b. Macroinvertebrates in neotropical streams: richness patterns along a catchment and assemblage structure between 2 seasons. *J. North Am. Benthol. Soc.* 20:1-16.
- Melo, A.S., R.A.S. Pereira, A.J. Santos, G.J. Shepherd, G. Machado, H.F. Medeiros and R.J. Sawaya. 2003. Comparing species richness among assemblages using sample units: why not use extrapolation methods to standardize different sample sizes? *Oikos* 101:398-410.
- Minshall, G.W., R.C.J. Petersen and C.F. Nimz. 1985. Species richness in streams of different size from the same drainage basin. *Am. Nat.* 125:16-38.
- Novotný, V. and Y. Basset. 2000. Rare species in communities of tropical insect herbivores: pondering the mystery of singletons. *Oikos* 89:564-572.
- Oliver, I. and A.J. Beattie. 1996. Invertebrates morphospecies as surrogates for species: a case study. *Conserv. Biol.* 10:99-109.
- Palmer, M.W. 1990. The estimation of species richness by extrapolation. *Ecology* 71:1195-1198.
- Patil, G.P. and C. Taillie. 2001. Estimation of species richness based on species range. *Community Ecol.* 2:209-211.
- Southwood, T.R.E. 1978. *Ecological Methods*. Chapman and Hall, New York.
- Stout, J. and J. Vandermeer. 1975. Comparison of species richness for stream-inhabiting insects in tropical and mid-latitude streams. *Am. Nat.* 109:263-280.
- Taylor, L.R., R.A. Kempton and I.P. Woiwod. 1976. Diversity statistics and the Log-Series model. *J. Anim. Ecol.* 45:255-272.
- Toti, D.S., F.A. Coyle and J.A. Miller. 2000. A structured inventory of Appalachian grass bald and heath bald spider assemblages and a test species richness estimator performance. *J. Arachnol.* 28:329-345.
- Wagner, H.H. and O. Wildi. 2002. Realistic simulation of the effects of abundance distribution and spatial heterogeneity on non-parametric estimators of species richness. *Ecoscience* 9:241-250.
- Walther, B.A. and J.-L. Martin. 2001. Species richness estimation of bird communities: how to control for sampling effort? *Ibis* 143:413-419.