

2

SPECIES DIVERSITY

Studies of species diversity are a major component of community ecology and biogeography. Patterns of species diversity in time and space form the basis of many important ecological models, including mechanisms of succession (Connell and Slatyer 1977), explanations for latitudinal diversity gradients (Pianka 1966; Stevens 1989), hypotheses of mass extinction in the fossil record (Raup et al. 1973), and relationships between diversity and stability (May 1973; Goodman 1975). Measures of species diversity are often used to evaluate the success of nature reserves in preventing extinction (Soulé and Wilcox 1980), and to assay the effects of environmental pollution on the “well-being” of natural ecosystems (Tomascik and Sander 1987).

SPECIES DIVERSITY

Unfortunately, “species diversity” has taken on a variety of confusing meanings (Hurlbert 1971). Here, we define diversity as a measure of both the number of species in the community and their relative abundances (Margalef 1968; Washington 1984). Diverse communities are characterized by a large number of species and/or a relatively even distribution of species abundances (Lambshhead et al. 1983). Considerable effort has been expended to devise a single numerical index that measures these two properties (see reviews by Peet 1974; Washington 1984; Magurran 1988). Regardless of the mathematics of the index, all diversity indices incorporate three important assumptions (Peet 1974):

1. *All individuals assigned to a specific class are assumed to be equal.* Most diversity indices do not recognize intraspecific differences among individuals, even though age- and size-structured populations may have important community-level effects (Polis et al. 1989). In theory, ages, sexes, or size classes could be

treated as different “species” (Hendrickson and Ehrlich 1971), but their numerical dependence on one another is different from their dependence on other species.

2. *All species or classes are assumed to be equally different from one another.* Thus, diversity indices will not permit us to directly identify keystone species (Paine 1966), which have an inordinate influence on local community structure. If communities are organized into guilds of species that share similar resources (Root 1967), removal of species within a guild will have different effects on diversity than removal of species not in a guild.
3. *Community structure is assumed to be measured in appropriate units.* For most animal communities, individuals are assigned to species or “operational taxonomic units” (OTUs; Sneath and Sokal 1973) and counted. For assemblages of plants or sessile invertebrates, measures of biomass or percent cover may be more appropriate. The “correct” units are important in interpreting diversity measures (Dickman 1968). For example, many ecologists have interpreted the relative abundance of a species in a community as a reflection of its relative use of ecological resources (MacArthur 1960; Sugihara 1980). Thus, in plant communities, numerically dominant species sequester more water, light, and nutrients than relatively rare species (Greig-Smith 1964; but see Rabinowitz et al. 1984). In animal communities, however, resource use may be more or less even than suggested by relative abundance distributions, because rare species tend to be large-bodied and comprise more biomass than their numbers would suggest (Harvey and Godfray 1987; Pagel et al. 1991).

PROBLEMS WITH DIVERSITY INDICES

Rather than addressing these important assumptions, most descriptions of diversity indices focus on the algebra of the index and the theoretical value of the index in measuring diversity. Washington (1984) reviewed a plethora of diversity, similarity, and biotic indices used in both theoretical and applied contexts. Most of these indices are highly correlated with one another (Ghent 1991), and many families of indices can be derived from a single algebraic expression (Hill 1973).

Perhaps the most celebrated index is the Shannon-Wiener diversity index: $H' = -\sum p_i \ln(p_i)$, where p_i is the relative abundance of the i th species ($\sum p_i = 1.0$).

The index has been used since the 1950s (Good 1953; MacArthur 1955) and became a “magic bullet” among ecologists (Washington 1984). The tenuous theoretical justification for H' came from information theory (Margalef 1958), but the idea that H' is a measure of entropy (reviewed by Goodman 1975) is no longer warranted (Hurlbert 1971). Nevertheless, the Shannon-Wiener index continues to be used and until recently, was thoroughly entrenched in the ecological literature. With the exception of metrics such as Hurlbert's (1971) PIE (probability of an interspecific encounter), most diversity indices cannot be interpreted in biologically meaningful ways.

Another problem with diversity indices is that they lack a probabilistic basis. Because the indices are not sampled from a known distribution, it is impossible to assign a probability value to them (Ghent 1991). Consequently, there is no way to evaluate the statistical and biological differences between two communities with H' values of, say, 1.7 and 2.0. Some ecologists have repeatedly sampled communities to estimate the mean and variance of a diversity index, or have used a jackknife of an individual sample to obtain confidence limits about a diversity index (Adams and McCune 1979). However, these approaches are unsatisfactory because all diversity indices are highly sensitive to both the number of species and the number of individuals in the collection; repeated sampling does not eliminate those biases. Another solution is to rescale the index algebraically so that it is bounded between 0.0 and 1.0. However, scaling procedures do not remove the sampling bias caused by differences in the size of two samples (Peet 1975).

In spite of these serious conceptual and statistical problems, there is a continued interest in measures of species diversity (Magurran 1988), and a need to quantify patterns statistically in natural communities. We offer two suggestions. The first is to abandon the idea of incorporating both evenness and species richness into a single index. Although there is obviously a relationship between species richness and relative abundance, the two components of diversity are distinct and should be analyzed separately (James and Rathbun 1981). The second suggestion is to use explicit null models to estimate the nonbiological effects of sample size on species richness and species evenness. Null models for species richness have been well developed, but relatively little work has been done with the evenness problem.

SPECIES RICHNESS

Species richness, as measured by a direct count of species, is the simplest and easiest diversity index to interpret (Peet 1974). However, species counts are

influenced not only by the species richness of the surrounding community, but also by the number of individuals counted and the amount of area sampled (Williams 1964). The areas sampled might represent natural, insular communities, such as oceanic islands, or arbitrary sampling units, such as vegetation quadrats. In either case, the more area sampled, the more species will be detected (e.g., Connor and McCoy 1979; Anderson and Marcus 1993). For insular communities, island area typically accounts for 50% of the variation in species number (Boecklen and Gotelli 1984). Even within a well-defined area, the number of species recorded depends on the thoroughness of sampling. For example, the number of botanical collecting trips to the Galápagos Islands was a better predictor of plant species number than area, elevation, or isolation (Connor and Simberloff 1978).

Consequently, unless two communities are censused exhaustively and identically, it is inappropriate to compare simple species counts. The problem is illustrated in Table 2.1, which shows the number of individuals of different carabid beetle species collected from pitfall traps in young (<20 years old) and old (20–60 years old) pine plantations in northern Europe (Niemelä et al. 1988). A total of 243 individuals of 31 species was collected from the young plantations, compared to 63 individuals of 9 species collected from the old plantations.

Which community exhibits greater species richness? We cannot automatically assume that richness was greater for the young plantations. Almost four times as many individuals were collected in the young plantations, so it is not surprising that more species accumulated. Nor can we simply rescale the smaller collection by a constant multiplier, because the relationship between the number of individuals in the collection and the number of species is seldom linear.

RAREFACTION

Sanders (1968) addressed the problem of comparing the species richness of different habitats in a study of marine benthic communities. He devised an algorithm for “rarefying” the large samples, that is, calculating the expected species richness based on random subsamples of individuals. This rarefied sample can then be compared directly with the smaller collection, because the species richness of both collections is now based on an identical number of individuals.

If the rarefaction procedure is carried out for a number of different abundances, a rarefaction curve can be plotted on a graph that has the number of

Table 2.1

Pitfall trap collections of carabid beetles from young and old pine plantations in northern Europe

Species	No. individuals collected	
	Young plantations	Old plantations
<i>Calathus micropterus</i>	48	29
<i>Pterostichus oblongopunctatus</i>	23	9
<i>Notiophilus biguttatus</i>	3	1
<i>Carabus hortensis</i>	2	0
<i>Carabus glabratus</i>	15	1
<i>Cychrus caraboides</i>	6	1
<i>Amara brunnea</i>	2	0
<i>Trechus secalis</i>	30	15
<i>Leistus terminatus</i>	3	1
<i>Amara familiaris</i>	1	0
<i>Amara lunicollis</i>	7	0
<i>Bembidion gilvipes</i>	2	0
<i>Bradycellus caucasicus</i>	1	0
<i>Calathus melanocephalus</i>	3	0
<i>Carabus nitens</i>	1	0
<i>Carabus violaceus</i>	1	0
<i>Cicindela sylvatica</i>	10	0
<i>Cymindis vaporariorum</i>	3	0
<i>Harpalus quadripunctatus</i>	7	0
<i>Harpalus</i> sp.	1	0
<i>Leistus ferrugineus</i>	1	0
<i>Miscodera arctica</i>	13	0
<i>Notiophilus aestuans</i>	2	0
<i>Notiophilus germinyi</i>	9	0
<i>Notiophilus palustris</i>	9	0
<i>Pterostichus adstrictus</i>	23	0
<i>Pterostichus cupreus</i>	1	0
<i>Pterostichus diligens</i>	1	2
<i>Pterostichus niger</i>	7	0
<i>Pterostichus strenuus</i>	4	4
<i>Synuchus vivalis</i>	4	0
Total abundance	243	63
Total species richness	31	9

From Niemelä et al. (1988).

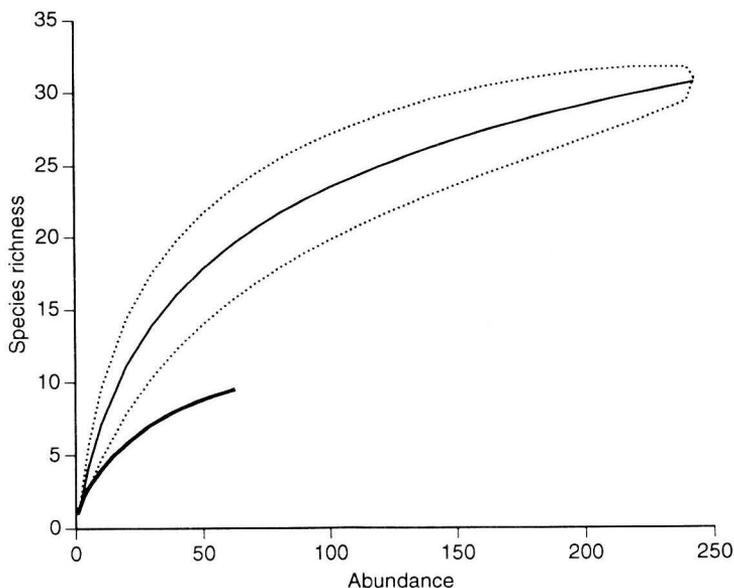


Figure 2.1. Rarefaction curves for carabid beetle assemblages in old and young pine plantations. The dotted line represents the 95% confidence interval for the young plantation curve and the thick line represents the old plantation curve. Data from Table 2.1.

individuals on the abscissa and the number of species on the ordinate. The rarefaction curve increases monotonically from a minimum point at (1,1) (a random sample of one individual will always contain exactly one species) to a maximum at (N,S) , where N and S are the number of individuals and species, respectively, in the original sample. Figure 2.1 shows the two rarefaction curves (and their 95% confidence limits) for the data in Table 2.1. The rarefaction curve for the old pine plantations lies significantly below that of the young pine plantations. Thus, the higher species richness of the young pine plantations was not simply an artifact of the greater number of individuals collected. For example, 63 individuals of 9 species were collected from the old pine plantations, whereas a random sample of 63 individuals from the young pine plantations would be expected to contain approximately 20 species.

Sanders's (1968) rarefaction was correct in principle, but his algorithm overestimated the expected number of species for a random subsample (Hurlbert 1971; Fager 1972; Simberloff 1972). The only case in which Sanders's (1968) formula is correct is when the individuals in the collection are uniformly distributed in space (Kobayashi 1982). Although competition in some communities may generate a uniform spatial distribution (e.g., Ryti

and Case 1986; Chapin et al. 1989), random or clumped spatial patterns are much more typical. Thus, Sanders's (1968) algorithm should not be used as a general null model.

The correct rarefaction model is based on the hypergeometric distribution, sampling without replacement from some parent distribution (Feller 1968), which in this case is the observed collection. Although rarefaction usually is applied to collections of individuals classified as species, the technique is appropriate for any hierarchical classification, such as species/genus (S/G) ratios, discussed in Chapter 1.

When individuals are sampled equiprobably and independently of one another, the expected species richness ($E(S_n)$) in a small sample is

$$E(S_n) = \sum_{i=1}^S \left[1 - \frac{\binom{N-m_i}{n}}{\binom{N}{n}} \right] \quad (2.1)$$

where N is the total number of individuals in the collection, S is the total number of species in the collection, m_i is the number of individuals of species i in the collection, and n is the number of individuals in the subsample (Hurlbert 1971). The term inside the summation sign is the probability that a sample of n individuals will contain species i . Summing over all the species in the collection gives the expected species richness. The variance is

$$\begin{aligned} \sigma^2(S_n) = & \binom{N}{n}^{-1} \left[\sum_{i=1}^S \binom{N-m_i}{n} \left(1 - \frac{\binom{N-m_i}{n}}{\binom{N}{n}} \right) \right. \\ & \left. + 2 \sum_{\substack{j=2 \\ i < j}}^S \left(\binom{N-m_i-m_j}{n} - \frac{\binom{N-m_i}{n} \binom{N-m_j}{n}}{\binom{N}{n}} \right) \right] \quad (2.2) \end{aligned}$$

(Heck et al. 1975). Both equations are formidable to calculate by hand, but fortunately, published computer algorithms (Simberloff 1978b) and ecological software packages (Ludwig and Reynolds 1988) are available. In spite of the repeated publication of Equations 2.1 and 2.2, Sanders's (1968) original algorithm continues to be incorrectly used, particularly by marine ecologists (e.g., Beukema 1991). We suspect this error persists because of the important influence of Sanders's (1968) paper, and because his equation can be solved on a hand calculator.

ASSUMPTIONS OF THE RAREFACTION MODEL

The rarefaction model entails a number of assumptions and restrictions:

1. *Sampling has been sufficient to guarantee an adequate characterization of the parent distribution* (Tipper 1979). If the samples are too small, rarefaction curves will always appear similar, because all rarefaction curves converge at the (1,1) coordinate.
2. *The spatial distribution of individuals is random*. If the distribution is not random, Equation 2.1 will give a biased estimate of the expected species richness. Specifically, the more spatially clumped individuals are, the more the rarefaction curve will overestimate the expected species richness (Simberloff 1986). Different rarefaction algorithms are available for different spatial distributions (Kobayashi 1982, 1983; Smith et al. 1985), but unless a great deal of information is available on the spatial organization of the assemblage, Equation 2.1 should be used.
3. *The samples to be compared are taxonomically “similar” and are drawn from the “same” community type*. For example, two collections of lepidoptera might exhibit similar rarefaction curves, but if one collection contains only moths and the other only butterflies, it would be incorrect to conclude they were drawn from the same distribution (Simberloff 1978b). Likewise, assemblages should be sampled from similar habitats, although this is often a matter of judgment (Raup 1975); many studies use rarefaction for the purpose of comparing species richness in different habitats (e.g., James and Wamer 1982; Mehlhop and Lynch 1986). What is important is that the sampling efficiency not change in different habitats. For example, a comparison between samples of birds mist-netted in a prairie and in a forest might be invalid if some species avoided the net disproportionately in the open prairie.
4. *Standardized sampling techniques are used for all collections*. Different sampling methods have different biases, so collections should not be compared by rarefaction (or any other technique) unless they were gathered with similar methods. For example, marine benthic collections gathered with cores and dredges should not be compared by rarefaction (Abele and Walters 1979a), because these two methods have very different sampling properties. Similarly, it would not be appropriate to use rarefac-

tion to compare the diversity of extant and fossil echinoderms, because the fossil record is biased toward taxa with skeletonized hard parts (Raup 1975). Even subtle differences in sampling area (Engstrom and James 1981) and trap design (Butman 1989) can affect the size and diversity of the collection. A standardized method must be used for meaningful comparisons.

5. *Rarefaction can be used for interpolation to a smaller sample size, but not for extrapolation to a larger sample size.* Because the rarefaction curve makes no assumptions about the shape of the underlying parent distribution, it cannot be properly extrapolated to larger sample sizes (e.g., Calef and Hancock 1974). If one is willing to assume, a priori, a particular species abundance distribution, such as the log normal (Preston 1962), then it is possible to derive the expected species richness for samples of any size (e.g., Kempton and Taylor 1974). As described later in this chapter, there are several estimators of total species richness based on species accumulation curves that are more appropriate for extrapolation (Colwell and Coddington 1994).

STATISTICAL ISSUES IN RAREFACTION

Rarefaction can be used to ask two slightly different questions (Simberloff 1978b): (1) For a collection of N individuals of S species, what is the expected number of species in a small sample of n individuals ($n < N$)? (2) What is the likelihood that two collections of sizes N and n were both drawn from the same parent distribution (Figure 2.2)? Equations 2.1 and 2.2 are appropriate for answering the first question. However, ecologists are usually more interested in answering the second question, and in this case, Equation 2.2 is a biased estimator of the variance. Smith and Grassle (1977) gave approximate confidence intervals for the expected variance in this case, but the calculation is complex.

However, Equation 2.2 is a minimum variance estimator, so if two samples do not differ in expected species richness based on this variance, they certainly will not by other measures. For several large collections of diatoms (Patrick 1968), Equation 2.2 gave similar results to other variance estimators (Simberloff 1979a); thus it seems a reasonable choice for most studies. The important point is to use *some* estimator of variance so that the null hypothesis that two collections were drawn from the same parent distribution can be tested.

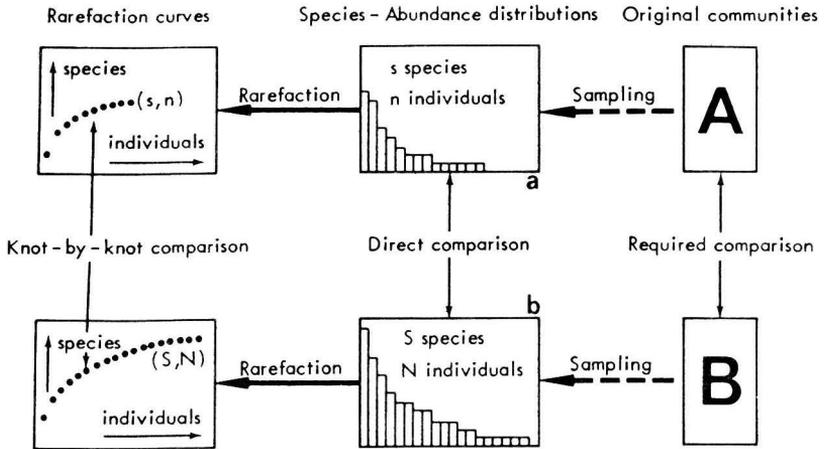


Figure 2.2. Samples from two communities are compared by rarefaction curves to generate expected species richness. From Tipper (1979), with permission.

It is also important to calculate rarefaction curves along their full length, that is, for several different values of n . If two rarefaction curves should cross, then the difference in expected species richness of the two collections will depend on the abundance being compared (Simberloff 1979a). However, most groups of published curves do not cross (Simberloff 1978b). Another reason for calculating curves along their full length is that they are based on discrete “knots” of abundance. If too few knots are used for comparison, pairs of rarefaction curves can appear more or less similar, depending on the placement of the knots. For a collection of S species and N individuals, a minimum of $S + 1$ subsamples of n , evenly spaced along the x axis, should be used to construct the curve (Tipper 1979). Once the curves have been constructed, conventional statistical tests can be used to compare expected species richness at different knots. In many cases, simply plotting rarefaction curves and their 95% confidence intervals is sufficient to reveal differences in expected species richness.

One final issue is how to deal with multiple collections from the same habitat or location. Tipper (1979) recommended calculating separate rarefaction curves for each, then using means and variances of the expected species richness to estimate the composite rarefaction curve. However, if the individual samples are small, the composite curve will be too short to allow for a powerful test. In such cases, pooling samples will provide a more accurate estimate of the relative abundances of rare species in the collection. In a winter bird-population study (Engstrom and James 1981), a pooled rarefaction curve

accurately predicted observed species richness in nine 6.5-ha subplots, even though the subplots varied somewhat in habitat structure.

SPECIES RICHNESS VERSUS SPECIES DENSITY

Rarefaction allows for a comparison of the number of species in samples that have been standardized for abundance. A closely related measure of diversity is species density (Simpson 1964), which is the expected number of species per unit area. Species density allows for comparison of the number of species in samples that have been standardized for the area censused. Species density depends on the species abundance relationship, the density of individuals, and, to a lesser extent, their spatial distribution (James and Wamer 1982). The spatial scale and the boundaries of the community from which the sample was drawn influence both species richness and species density. The two measures need not give similar results.

A study of North American breeding birds (James and Wamer 1982) illustrates the distinction between species richness and species density. Using census data from a variety of forest habitats, James and Wamer (1982) compared species number for samples with an equal number of individuals (species richness) and for samples with an equal area (species density). Rarefaction curves generated for each forest type did not cross, indicating that differences among habitats were consistent. The rate of species accumulation also differed among habitats. For example, at an abundance of 50 territorial pairs, the accumulation curve for the wax myrtle forest was leveling off, whereas the curve for second-growth forest was still rising rapidly.

Determining species density on equal-sized areas requires an estimate of the density of territorial pairs. James and Wamer (1982) used point estimates of density in each habitat, and then extrapolated these estimates to larger and smaller areas. Their analysis assumed that density was constant across all areas, yielding a set of linear abundance curves (Figure 2.3b). These curves did not fall out in exactly the same order as the original rarefaction curves. For a plot of a given size, James and Wamer (1982) used these density curves to estimate the number of territorial pairs, and then used the rarefaction curves to get the expected species richness (Figure 2.3a). The interaction between density and expected species richness determined species density in different habitats. The resulting species density (Figure 2.3c) was highest in the tulip tree-maple-oak forest, in part because this habitat supported the greatest density of individual territories.

James and Wamer's (1982) study nicely illustrates the interplay of area, density, and species abundance in producing patterns of species richness. Their

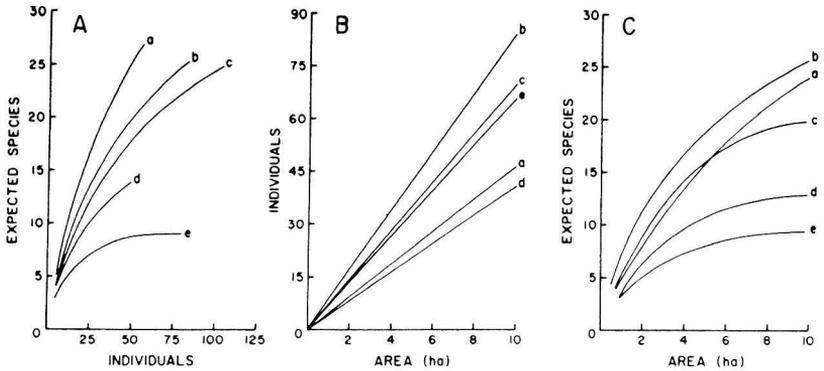


Figure 2.3. Avian species richness in five forest habitats. (A) Rarefaction curves, which give the expected species richness for a sample of individuals. (B) Density relationships for each habitat, which were assumed to be linear across a range of plot sizes. (C) Species density relationships, which incorporate both species richness and density relationships. Note the different orderings of the habitats based on these measures. (a) Maple-pine-oak second-growth forest; (b) tulip tree-maple-oak forest; (c) cottonwood floodplain forest; (d) mature jack pine with birch forest; (e) wax myrtle forest. From James and Wamer (1982), with permission.

method should be used whenever sampling is based on plots or quadrats of known area. However, not all sampling methods are area-based. Point counts, baits, or stationary traps may sample individuals from an unknown area. For these sampling schemes, species richness can still be compared by plotting the expected abundance per sample on the x axis of the rarefaction curve (Engstrom et al. 1984).

Figure 2.4 illustrates two rarefaction curves based on pitfall trap samples of ants in an Oklahoma grassland (Gotelli in press). Each curve is a pooled sample of the contents of six pitfall traps. The “ant lion zone” traps were placed beneath a cliff ledge in a dense aggregation of ant lion larvae, whereas the other traps were placed approximately 1 m beyond the ant lion zone in a “predator-free zone” (Gotelli 1993). The two rarefaction curves were very similar to one another: the ant lion zone curve fell completely within the 95% confidence interval for the predator-free zone curve, a pattern that was consistent across three different sites and five consecutive years (Gotelli in press).

Although rarefaction curves were similar in the two habitats, abundance per trap was not. Ant abundance was significantly greater in the predator-free zone, and this translated into a greater species number on the rarefaction curve. In James and Wamer’s (1982) analysis, differences in bird species density were caused by differences in abundance and expected species richness. In the

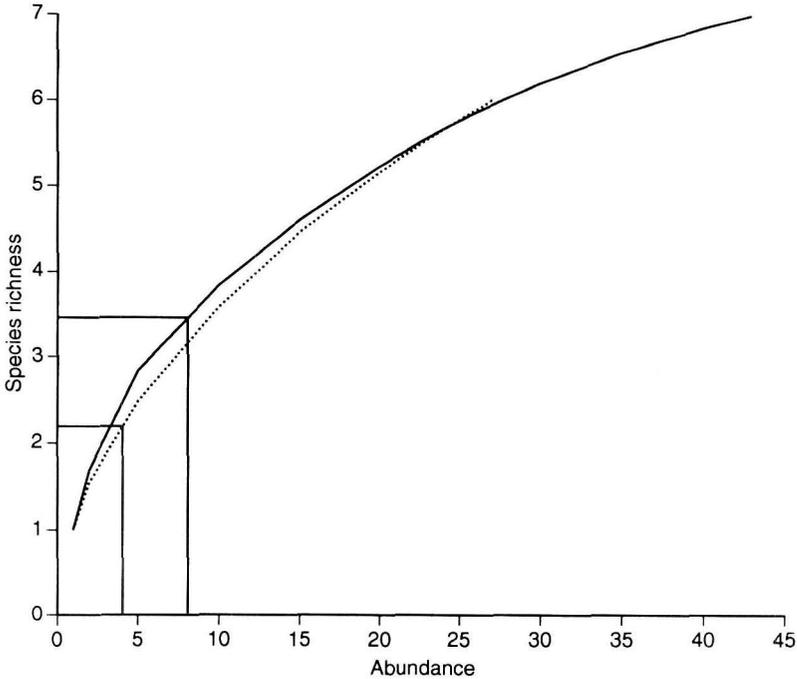


Figure 2.4. Rarefaction curves for ant assemblages collected in pitfall traps in the presence (dotted curve) and absence (solid curve) of ant lion predators. The two rarefaction curves are indistinguishable, but abundance per trap (which affects species richness) is greater in the absence of predators. Vertical lines indicate average abundance in the two microhabitats. See Gotelli (in press).

Oklahoma ant assemblage, differences in species density were primarily due to differences in abundance. Indicating the expected or average sampling abundance on a rarefaction curve (e.g., Engstrom et al. 1984) gives important information about both species richness and species density of a sample.

USES OF RAREFACTION

In this section, we review some of the diverse applications of rarefaction to problems in ecology and evolutionary biology.

Deep-Sea Diversity

Sanders's (1968) paper was noteworthy not only for introducing rarefaction, but for popularizing the "stability-time" hypothesis in the ecological literature.

Sanders (1968) compared rarefaction curves for marine invertebrates (polychaetes and bivalves) in different benthic habitats and found the highest species richness for shallow tropical reefs and the deep ocean floor. Sanders (1968; Sanders and Hessler 1969) argued that these habitats were relatively constant and experienced little abiotic fluctuation through time. Under these circumstances, biotic interactions were thought to promote species coexistence and lead to complex, stable assemblages with high species richness. In contrast, habitats with low species richness, such as estuaries, were subject to variable abiotic conditions. Physiological stress controls the assemblage and low species richness results, because few species can tolerate highly variable conditions. The relatively high deep-sea diversity was a surprising finding, and Sanders's (1968) explanation reflected conventional wisdom that species-rich, biologically complex communities such as tropical rain forests evolved over long periods of physical stability (Fischer 1960).

Abele and Walters (1979a,b) pointed out the potential circularity in this argument—stressful environments are often defined or recognized as those with low species richness (Peters 1976)—and also attacked Sanders's (1968) methods. They noted that in some cases Sanders (1968) compared dredge and core samples and that the original rarefaction algorithm was biased. Abele and Walters (1979a,b) calculated corrected rarefaction curves (and their 95% confidence intervals) and found that most did not differ significantly from one another. Although estuaries were relatively depauperate, deep-sea samples did not appear to be unusually species-rich.

Abele and Walters (1979a) suggested that deep-sea diversity could be effectively explained by the large area of the habitat and by small-scale variation in habitat quality. Recent intensive sampling of the continental slope has again generated high estimates of species richness (Grassle and Maciolek 1992), and these have been attributed to microhabitat heterogeneity and small-scale biogenic disturbances. Whatever the “correct” explanation for deep-sea diversity, rarefaction was an important tool for revealing the basic patterns of diversity that precipitated this debate.

Evolutionary Ecology

As discussed in Chapter 1, the rarefaction methodology was developed independently by A. Mailliefer (1929), C. B. Williams (1947a), and others to examine species/genus (S/G) ratios and other indices of taxonomic diversity. Whether one is examining species richness in a benthic core sample, or the number of coexisting congeneric species on an island, rarefaction is the appropriate method for studying the diversity of a hierarchically classified sample.

Although the questions were very different from those addressed by Sanders (1968), the statistics of rarefaction are identical for comparisons of taxonomic diversity. The result from many independent studies is that taxonomic diversity indices rarely deviate significantly from an appropriate rarefaction curve (Simberloff 1970; Harvey et al. 1983). The confidence bands for most rarefaction curves are broad, so that small assemblages would have to be most unusual taxonomically to generate a statistically significant result.

However, this does not mean that taxonomic diversity indices have no value. Although individual communities may not deviate significantly, samples of multiple communities may reveal nonrandom patterns. For example, Simberloff (1970) found that approximately 70% of the examples of S/G ratios deviated in a positive direction: in most genera, there were slightly more coexisting species in a genus than expected by chance. Gotelli and Abele (1982) found that the number of bird families represented on islands of the West Indies did not differ statistically from the rarefaction curve of the pooled species list. However, deviations from the rarefaction curve were positively correlated with island area, perhaps reflecting increased habitat diversity on large islands. Similarly, Järvinen and Sammalisto (1976) discovered a latitudinal gradient in S/G ratios of breeding birds on Finnish peatlands. After correction for the number of species in each sample, the S/G ratio still increased with latitude. These results suggest that the (standardized) residual from a rarefaction curve is a useful diversity index that controls for differences in the size of the assemblage. For multiple samples, this residual can then be correlated with other physical or biological attributes of the community.

Sampling Efficiency

Because most communities cannot be fully enumerated, their properties will always have to be inferred from those of a sample. Common species will usually be encountered first when sampling, and there will be diminishing returns on effort as the rare species in an assemblage are captured only with more intensive effort. What is the correct stopping point? If the individuals are sampled randomly, the rarefaction curve shows the expected accumulation of species and can be used to plan effective sampling strategies.

For breeding birds of North American forests, some rarefaction curves did not begin to reach an asymptote until 75 pairs had been censused, corresponding to a minimum plot area of approximately 10 ha (James and Wamer 1982). Using rarefaction, Haila and Kuusela (1982) determined that single-census visits to small Baltic islands would reveal 70% or more of the resident bird species. Heck et al. (1975) rarefied a large collection of benthic invertebrates

and found that 90% of the species could be expected from a collection that was 70% as large as the original. In summary, rarefaction can be used to efficiently minimize the collection effort required for a single sample, allowing for more spatial or temporal replication of samples.

Paleobiology

Rarefaction is an ideal technique for comparing the diversity of fossil assemblages, because differences in the “abundance” of fossil collections are likely to reflect both taphonomic biases and collection effort. A simple “balls in buckets” null model illustrated such sampling effects in the measurement of higher levels of taxonomic diversity (Raup 1972). Raup (1975) discussed the biological considerations in using rarefaction in paleobiology, and Tipper (1979) reviewed the use (and abuse) of the technique in paleobiology.

Rarefaction was used to show that an apparent increase in echinoid families after the Paleozoic was real, whereas there has not been a significant increase in familial diversity since the mid-Cretaceous (Raup 1975). In paleontological studies, rarefaction is frequently carried out at high taxonomic levels (orders, families), because it is often difficult to reliably classify fossils at the generic and specific levels.

For example, Raup (1979) used rarefaction to estimate species losses at the Permo-Triassic boundary. He first constructed rarefaction curves for orders, families, and genera of living echinoid species as a reasonable estimate of a generalized rarefaction curve for a clade. In addition to the other restrictions imposed by rarefaction, Raup’s (1979) analysis assumed that the rarefaction curve for living echinoids was comparable in shape to that of other fossil clades. Next, he estimated the percent extinction of orders, families, and genera of well-skeletonized marine organisms at the Permo-Triassic boundary. Finally, he back-calculated from these percentages on the living rarefaction curve to estimate extinction at the species level. The resulting estimate of 88 to 96% extinction at the species level agreed with rates estimated by other methods (Valentine et al. 1978).

Antia (1977) used rarefaction to explore the effects of taphonomic biases on the estimation of diversity and the characterization of fossil assemblages. He thoroughly sampled living mollusc assemblages and dead molluscan shell cumulates from seven marine habitats in southeastern England. Although rarefaction of the dead shell cumulates consistently over-estimated diversity compared to the living assemblages, expected species richness in different habitats was consistent for fossil and living assemblages. This fact suggests that rarefaction can be effectively used to compare the diversity of different fossil assem-

blages, even though they may be subject to preservational biases. Because rarefaction estimates diversity for an entire assemblage, it is relatively robust to preservational biases that may distort the abundances of individual species.

Pollution Studies

Pollution effects can be monitored by the performance of individual indicator species (Shubert 1984) and by changes in indices of community structure (Washington 1984). Rarefaction has been used in both field and laboratory studies as a simple index of community response to pollution. For example, Dean-Ross (1990) found that the expected number of bacterial strains in laboratory streams decreased in response to high (but not low) concentrations of zinc. Berger (1990) examined the recolonization by benthic macrofauna of oiled and unoled sediment boxes in two Norwegian fjords. In a eutrophic fjord, oil had no effect on expected species richness as measured by rarefaction; abundance and species richness of both control and oiled sediments returned to ambient levels in 3 months. In contrast, oil significantly reduced expected species richness in a noneutrophic fjord: recovery time was approximately 9 months for the oiled treatment and 16 months for the control. In pollution studies, it is especially important to compare both expected species richness and species density. If a pollutant is equally toxic to all resident species, it will effectively “rarefy” an assemblage to a reduced abundance. The rarefaction curves for a control and treatment assemblage would be identical in this case, even though abundance and species density are reduced by the treatment.

Successional Gradients

Rarefaction of breeding bird assemblages in successional gradients has revealed temporal trends in species richness and/or species density. Using mist-net captures, Mehlhop and Lynch (1986) compared avian assemblages near Chesapeake Bay, in habitats ranging from recently abandoned pasture to mature hardwood forest. The highest expected species richness occurred in forests of intermediate age. However, capture rates increased with forest age, so patterns of species density may be different, as James and Wamer (1982) found. In seral stages of beech forest in Hungary, expected bird species density increased with stand age (Moskat and Szekely 1989). For standardized plots of 28 ha, species density increased from a minimum of 12 species in 1- to 2-year-old clear-cuts to a maximum of 26 species in century-old beech forests.

Engstrom et al. (1984) is one of the few successional studies to track temporal changes in species richness directly. In the upland coastal plains of the

southeastern United States, frequent lightning fires discourage hardwood growth and maintain a virtual monoculture of long-leaf pine. If fire is suppressed, a structurally complex beech-magnolia forest develops. The species density was 21 bird species per 8.6-ha plot for long-leaf pine forest compared to 16 species for beech-magnolia forest. Engstrom et al. (1984) generated rarefaction curves for census data collected during 15 successive years on a single plot of long-leaf pine from which fire had been excluded since 1967. Expected bird species density declined during the first 10 years of the study, then began increasing again in 1978. Changes in species density were accompanied by dramatic changes in species composition; open-habitat species dropped out during the first five years and were replaced by species associated with mesic vegetation.

RELATED QUESTIONS

Although rarefaction is used to study the total number of species in an assemblage, investigators frequently are curious about the abundance of a particular species (or higher taxon) in a sample. For example, in many tropical communities, certain bird families present on the mainland are consistently absent from nearby islands (MacArthur et al. 1972; Terborgh and Winter 1978). The apparent “disharmony” of island faunas (Carlquist 1974) has often been attributed to the effects of interspecific competition (Case et al. 1979) and reduced habitat availability on islands (D. L. Lack 1976). Before such explanations can be invoked, it is necessary to reject the null hypothesis that the groups are missing from islands by chance, and are not present simply because the total number of species on the island is small.

Haila and Järvinen (1983) studied missing species in the avifaunas of islands in the Baltic Sea. They compared the avifauna of the island of Ulversö (68 species) with the avifauna of Main Åland (121 species). Main Åland is the largest island in the archipelago, and the Ulversö fauna is a proper subset of the Main Åland fauna. Why were 53 species (121 – 68) absent from Ulversö? On Main Åland, 45 species occurred on fewer than six line transects and were defined as “rare.” Only six of these rare species occurred on Ulversö, compared to the expected number of 5.2. Thus, 39 of the 53 absences (74%) could be accounted for on the basis of “rarity” on Main Åland. The remaining absences could be attributed to habitat impoverishment of Ulversö (12%) or poor dispersal potential (3%). Rarefaction curves for assemblages censused in different island habitats on Main Åland were good predictors of species richness in those same habitats on Ulversö. The results suggest that the small area of Ulversö and

differences in habitat availability were the most important determinants of the composition and size of insular avifaunas.

We can modify the rarefaction algorithm to evaluate the frequency of any particular species in a collection of individuals. From the hypergeometric distribution, the expected abundance of any particular species s_i in a small random sample is

$$E(s_i) = \frac{nm_i}{N} \quad (2.3)$$

which is simply its proportional frequency in the parent distribution. The variance is

$$\sigma^2(s_i) = \left(\frac{m_i n}{N}\right) \left(1 - \frac{m_i}{N}\right) \left(\frac{N-n}{N-1}\right) \quad (2.4)$$

and the probability of obtaining the observed number of occurrences or fewer is

$$P(x \leq c) = \sum_{x=0}^c \frac{\binom{m_i}{x} \binom{N-m_i}{n-x}}{\binom{N}{n}} \quad (2.5)$$

Gotelli and Abele (1982) used the hypergeometric model to examine the number of species in different bird families that occurred on islands of the West Indies. Although rarefaction showed that each island supported the expected number of families, there were some significant deviations: on most islands there were more species of Columbidae and Mimidae than expected by chance, whereas Psittacidae were weakly underrepresented. Graves and Gotelli (1983) also found little evidence for extinction-prone taxa on land-bridge islands off the coast of South America: only one family (Bucconidae) was significantly underrepresented, whereas three families (Columbidae, Tyrannidae, and Parulidae) were overrepresented. However, species with small mainland geographic ranges were consistently underrepresented on land-bridge islands (Figure 2.5). The results of these null model studies contrast with earlier regression analyses (Faaborg 1979), which suggested that species in certain Neotropical bird families were unusually extinction-prone on islands.

Equation 2.5 also can be used to design effective sampling strategies. When the focus of a study is a particular species, rather than the entire community, an important question is how many individuals must be sampled to ensure capture of the target species. Returning to the carabid beetle data of Niemelä et al.

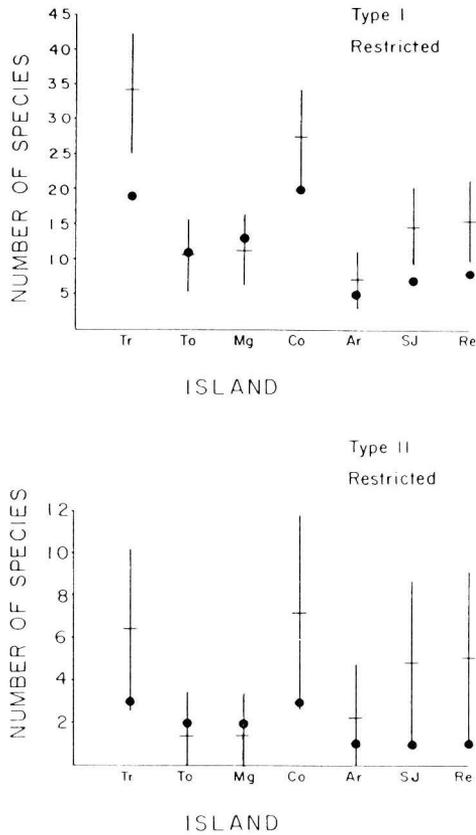


Figure 2.5. Observed and expected avian species richness on Neotropical land-bridge islands. The horizontal bar represents the expected number of species with restricted (=small) mainland geographic ranges, and the vertical line indicates two standard deviations. Expectation and variance were calculated from Equations 2.3 and 2.4. The observed number is shown by a filled circle. “Type I restricted” are species whose mainland ranges total less than $100 1^\circ \times 1^\circ$ blocks. “Type II restricted” are species whose entire range fits within a contiguous $10^\circ \times 10^\circ$ block. For both measures, species with restricted mainland geographic ranges were underrepresented on land-bridge islands, compared to the null model of random sampling from adjacent mainland source pools. Island codes: Tr = Trinidad; To = Tobago; Mg = Margarita; Co = Coiba; Ar = Aruba; SJ = San José; Re = Rey. San José and Rey are oceanic islands. From Graves and Gotelli (1983), with permission.

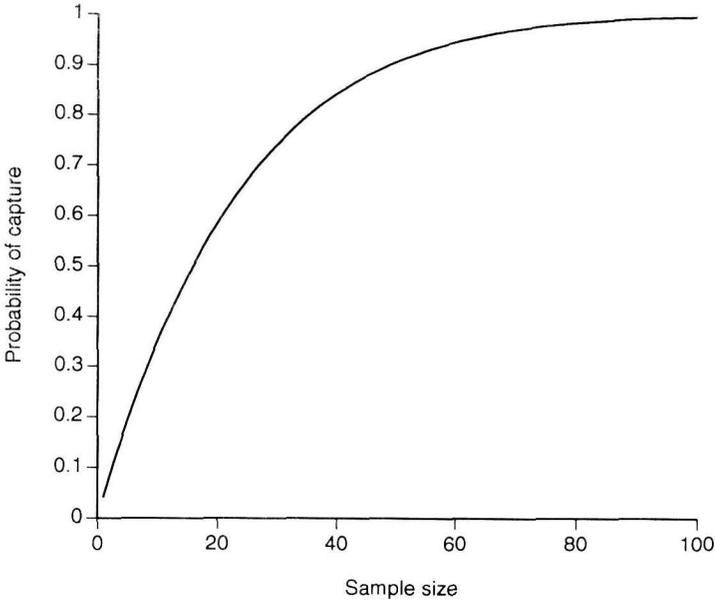


Figure 2.6. The probability of finding at least one individual of the carabid *Cicindela sylvatica* in pitfall samples from young pine plantations. Data from Table 2.1. Probabilities calculated using Equation 2.5.

(1988), Figure 2.6 illustrates the probability of capturing at least one individual of *Cicindela sylvatica* for a given sample size. A minimum of 31 individuals is needed in a sample to ensure with a 75% probability that *C. sylvatica* will be represented. Twice this number is needed to ensure a 95% probability.

CRITICISMS OF RAREFACTION

Rarefaction has been criticized for a number of reasons: it assumes a random spatial distribution of individuals (Kobayashi 1982, 1983), it loses information about species identity and relative abundance (Williamson 1973), and it is difficult to calculate by hand (Magurran 1988). However, all but the last criticism apply to other diversity indices as well. The potential disadvantages of rarefaction are outweighed by the facts that species richness is easy to interpret and that the rarefaction curve is based on an appropriate statistical model.

Furthermore, the shape of the rarefaction curve does reflect the shape of the relative abundance distribution, albeit indirectly. If the relative abundances in

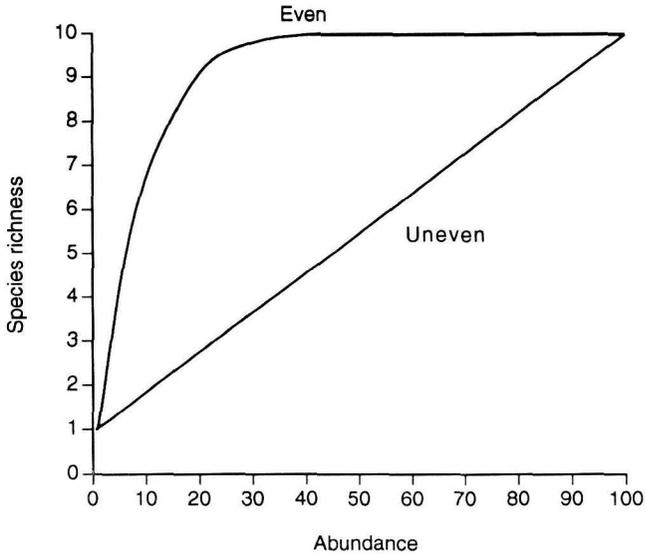


Figure 2.7. Rarefaction curves reflect the shape of the underlying species abundance distribution. Both rarefaction curves were drawn from samples of 100 individuals of 10 species. In the “even” sample, each species is represented by 10 individuals. In the “uneven” sample, 9 species are represented by 1 individual, and 1 species is represented by 91 individuals.

the collection are at maximum evenness (all species have equal abundance), then the rarefaction curve rises rapidly to an asymptote, because few samples are required to accumulate quickly all of the species. In contrast, if the relative abundances are maximally uneven (all species except the dominant represented by a single individual), then the rarefaction curve will increase in a slow linear fashion, because it takes many samples to accumulate additional species, all of which are rare (Figure 2.7). If density is constant with respect to island area or sample quadrat size, then similar comments apply to the shape of the species-area relationship under a null model of passive sampling (see Chapter 8).

Lambshhead et al. (1983) noted that the rarefaction curve is related to the k -dominance plot, which graphs percentage cumulative abundance against species rank. They recommended using k -dominance plots, because they are easier to calculate and to interpret than rarefaction curves. However, k -dominance plots are based on percentages, so they are distorted by small sample sizes. The rarefaction curve is the only diversity measure that is sensitive to rare species and is unbiased by sample size (Smith and Grassle 1977).

EXTRAPOLATION AND ESTIMATION OF TOTAL SPECIES RICHNESS

What is the asymptotic species number in an assemblage? Although rarefaction is used to interpolate species richness, extrapolation would be necessary to determine the total number of species present in an assemblage (Bunge and Fitzpatrick 1993). Species accumulation curves (Palmer 1990) can be used to estimate the asymptote of species richness as sample size increases. Colwell and Coddington (1994) summarized the diverse literature on this problem and compared many of the proposed theoretical solutions to a large data set on plant species richness. They found that the most promising nonparametric estimators have been adapted from mark-and-recapture statistics. For a single collection of species, the best estimator of total species richness was (Chao 1984)

$$S_{tot} = S_{obs} + (a^2 / 2b) \quad (2.6)$$

S_{tot} is the estimate of the total number of species, S_{obs} is the observed number of species, a is the number of species represented by only a single individual ("singletons"), and b is the number of species in the sample represented by exactly two individuals ("doubletons"). The variance for this estimate is

$$\sigma^2_{S_{tot}} = b \left[\left(\frac{a/b}{4} \right)^4 + (a/b)^3 + \left(\frac{a/b}{2} \right)^2 \right] \quad (2.7)$$

For the carabid beetle data in Table 2.1, a is 4 and b is 8. The estimated total species richness for the young plantations is 39 species, with a variance of 40.0. Extrapolation is obviously much riskier than interpolation. Nevertheless, Equations 2.6 and 2.7 performed well in an empirical test. Further tests are needed against other data sets, but these equations seem especially promising for estimating total species richness in applied conservation problems (Colwell and Coddington 1994).

SPECIES EVENNESS

In spite of the extensive literature on evenness indices, there have been few null model studies of evenness. The essential problem is that rare species are missed disproportionately in small collections, so that any measure of evenness will decrease with sample size. The most extreme case is a collection of one individual, which represents minimum evenness for any index. For this reason, evenness cannot be evaluated by visual inspection of species abundance distri-

butions (James and Rathbun 1981). Differences in the appearance of species abundance distributions may reflect differences in abundance or species richness, rather than differences in evenness. Even if total abundance and species richness are identical (which they never are), species abundance distributions may differ by chance.

Ghent (1991) used an explicit null model to examine the behavior of a number of diversity and niche breadth indices. His null model was that individuals are assigned to species equiprobably. He then used a multinomial exact test to assess the probability of finding the observed distribution by chance. An important result of his analysis was that nearly all the common diversity indices (e.g., Shannon's H' , Brillouin's H , and Simpson's C') yielded similar rank orderings of probabilities based on the equiprobable model. Consequently, the choice of the particular index to use is much less important than is the comparison with an appropriate null model. Ghent (1991) also argued that use of any of these indices implied an underlying null model of an equiprobable distribution of individuals. He recommended using Simpson's C' as an index, because its rank orderings were identical to those based on the variance of species counts.

Ghent's (1991) approach is promising for studies of niche breadth (see Chapter 4) but seems less suitable for diversity indices. Rather than assigning individuals equiprobably to different species, we prefer a null model that is based on rarefaction—a random sample of individuals is drawn from a given species abundance distribution to estimate sampling effects for the index. For such an analysis, we suggest Hurlbert's (1971) probability of an interspecific encounter (PIE), which measures the chance that two individuals drawn randomly from the collection represent different species:

$$\text{PIE} = \left(\frac{N}{N-1} \right) \left(1 - \sum_{i=1}^S \left(\frac{m_i}{N} \right)^2 \right) \quad (2.8)$$

This index is closely related to Simpson's (1949) "measure of concentration." Hurlbert's (1971) PIE also has an important analog in population genetics. It is equivalent to the calculation of heterozygosity (H), the probability that two alleles are not identical by descent (Charles J. Goodnight, personal communication).

Returning again to the beetle data in Table 2.1, we used a Monte Carlo simulation to draw, without replacement, samples of individuals from the two assemblages. For each subsample, we calculated PIE, and repeated the procedure 100 times to estimate the variance and the 95% confidence interval at several abundance levels (Figure 2.8). Although the variance decreased with

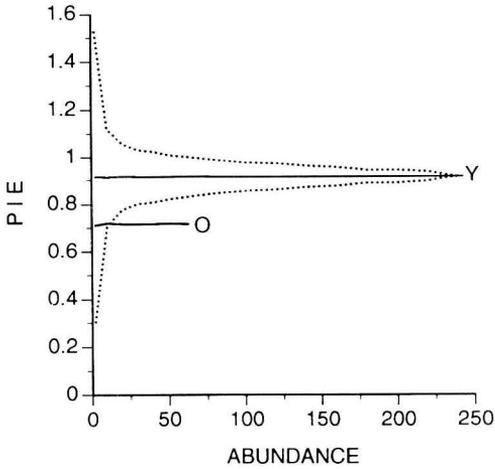


Figure 2.8. An unbiased evenness index. Hurlbert's (1971) PIE (Probability of an Interspecific Encounter) is graphed as a function of abundance in the sample. Samples were drawn randomly from the young (Y) and old (O) plantation data in Table 2.1. The curves and 95% confidence interval for the young plantation data (dotted line) were constructed from 100 random samples at each level of abundance. Note that the index remains unbiased, even at small sample sizes.

increasing sample size, PIE was correctly estimated even for small samples. Because PIE is an unbiased estimator and has a straightforward statistical interpretation, we prefer it as a simple diversity index.

The same restrictions and assumptions for the rarefaction model also apply to the analysis of species diversity. However, there is an important conceptual difference between the rarefaction curve and the sampling curves for diversity indices. In the rarefaction curve, the response variable is the number of species. In the diversity curves, the response variable is a metric that depends on both species richness and the number of individuals. Consequently, the variance in Figure 2.8 has two components: (1) variation due to the number of species in the sample, and (2) variation due to the evenness of the sample for a given number of species. Thus, the evenness curve still depends on species richness for a given level of abundance. The problem is that there is no way to control species number and abundance simultaneously in a nonarbitrary fashion. Nevertheless, the simulations at least control for the gross effects of abundance differences.

Finally, we note an interesting diversity index that is based directly on rarefaction. Siegel and German (1982) constructed a scaled index that measures the position of the observed rarefaction curve relative to boundaries set by

curves of maximum evenness and maximum unevenness for the collection. The drawback to their approach is that no probability value can be assigned to the index. Tagatz et al. (1983) used the index in a study of benthic colonization of creosote-contaminated sediment. Among a set of eight experimental treatments, the index ranged only from 0.42 to 0.52, and there is currently no way to assess the significance of this variation. Nevertheless, the approach of using bounded rarefaction curves is promising and deserves further study with null models.

RECOMMENDATIONS

We recommend the rarefaction curve (and its variance) as a useful measure of species richness. If collections are based on standardized sampling effort, then *it will be possible to measure density, species density, and species richness with rarefaction.* The rarefaction curve can also be used to estimate sampling efficiency, and the associated hypergeometric model can be used to measure the expected abundance of a particular species. For questions of species evenness, we recommend a Monte Carlo simulation of Hurlbert's (1971) PIE. This diversity index is unbiased at small sample sizes and has a simple interpretation as the probability that two individuals drawn randomly from an assemblage represent different species. Finally, to extrapolate from a sample and estimate total species richness, Colwell and Coddington (1994) should be consulted.