Ecological Methodology

Charles J. Krebs
University of British Columbia

HARPER & ROW, PUBLISHERS, New York
Cambridge, Philadelphia, San Francisco
London, Mexico City, São Paulo, Singapore, Sydney

Sponsoring Editor: Claudia M. Wilson
Project Editor: Steven Pisano
Text Design Adaptation: Barbara Bert/North 7 Atelier Ltd.
Cover Design: Wanda Lubelska Design
Text Art: Fineline Illustrations, Inc.
Production Manager: Kewal Sharma
Compositor: TAPSCO, Inc.
Printer and Binder: R. R. Donnelley & Sons Company

Ecological Methodology

Library of Congress Cataloging in Publication Data

Krebs, Charles J.
Ecological methodology/Charles J. Krebs.
p. cm.
Includes index.
ISBN 0-06-043784-7
1. Ecology—Statistical methods. I. Title.
QH541.15.S72K74 1989 88-19838
574.5'072—dc19 CIP

89 90 91 9 8 7 6 5 4 3 2

most cases ecologists worry about *species diversity*, but there is no reason why *generic* diversity or *subspecific* diversity could not be analyzed as well. Within the classification system, all the individuals assigned to a particular class are assumed to be identical. This can cause problems. For example, males may be smaller in size than females—should they be grouped together or kept as two groups? Should larval stages count the same as an adult stage? This sort of variation is usually ignored in species diversity studies.

Most measures of diversity assume that the classes (species) are all equally different. There seems to be no easy way around this limitation.

Diversity measures require an estimate of species importance in the community. The simple choices are numbers, biomass, cover, or productivity. The decision will depend in part on the question being asked. Numbers are used in most cases, although Dickman (1968) found that with lake samples of plankton the best measure was productivity.

A related question is how much of the community we should include in our sampling. We must define precisely the collection of species we are trying to describe. Most authors pick one segment—bird species diversity or tree species diversity. Rarely do diversity measures cross trophic levels, and only rarely are they applied to whole communities. Colwell (1979) argues convincingly that ecologists should concentrate their analyses on parts of the community that are functionally interacting, the guilds of Root (1973). These guilds, or networks, often cross trophic levels and include taxonomically unrelated species in them. The choice of what to include in a "community" is critical to achieving ecological understanding, yet no rules are available to help you make this decision. The functionally interacting networks can be determined only by detailed natural history studies of the species in a community.

10.2 CONCEPTS OF SPECIES DIVERSITY

Early naturalists very quickly observed that tropical areas contained more species of plants and animals than did temperate areas. But as ecological ideas matured and ideas of quantitative measurement were introduced, it became clear that the idea of species diversity contains two quite distinct concepts.

10.2.1 Species Richness

This is the oldest and the simplest concept of species diversity—the number of species in the community. McIntosh (1967) coined the name *species richness* to describe this concept. The basic measurement problem is that it is often not possible to enumerate all of the species in a natural community.

10.2.2 Heterogeneity

If a community has 10 equally abundant species, should it have the same diversity as another community with 10 species, one of which makes up 99% of the total individ-

uals? No, answered Simpson (1949), who proposed a second concept of diversity which combines two separate ideas, species richness and evenness. In a forest with 10 equally abundant tree species, two trees picked at random are likely to be different species. But in a forest with 10 species, one of which is dominant and contains 99% of all the individuals, two trees picked at random are unlikely to be different species. Figure 10.1 illustrates this concept.

The term *heterogeneity* was first applied to this concept by Good (1953), and for many ecologists this concept is synonymous with *diversity* (Hurlbert, 1971). The popularity of the heterogeneity concept in ecology is due partly to the fact that it is relatively easily measured.

10.2.3 Evenness

Since heterogeneity contains two separate ideas—species richness and evenness—it was only natural to try to measure the evenness component separately. Lloyd and Ghelardi (1964) were the first to suggest this concept. For many decades field ecologists had known that most communities of plants and animals contain a few dominant species and many species that are relatively uncommon. Evenness measures attempt to quantify this unequal representation against a hypothetical community in which all species are equally common. Figure 10.1 illustrates this idea.

10.3 SPECIES RICHNESS MEASURES

Some communities are simple enough to permit a complete count of the number of species present, and this is the oldest and most simple measure of species richness. Complete counts can often be done on bird communities in small habitat blocks, mammal communities, and temperate and polar communities of higher plants, reptiles, amphibians, and fish. But it is often impossible to enumerate every species in communities of insects, intertidal invertebrates, soil invertebrates, or tropical plants, fish, or amphibians. How can we measure species richness when we have only a sample of the community's total richness? The larger the sample, the greater the expected number of species. Three approaches have been used in an attempt to solve this sampling problem.

10.3.1 Rarefaction Method

One problem that frequently arises in comparing community samples is that they are based on different sample sizes. One way to overcome this problem is to standardize all samples to a common size. Sanders (1968) proposed the rarefaction method for achieving this goal. Rarefaction is a statistical method for estimating the number of species expected in a random sample of individuals taken from a collection. Rarefaction answers this question: *if the sample had consisted of n individuals ($n < N$), what number of species (s) would likely have been seen?* Note that if the total sample has S species and N individuals, the rarefied sample must always have $n < N$ and $s < S$.

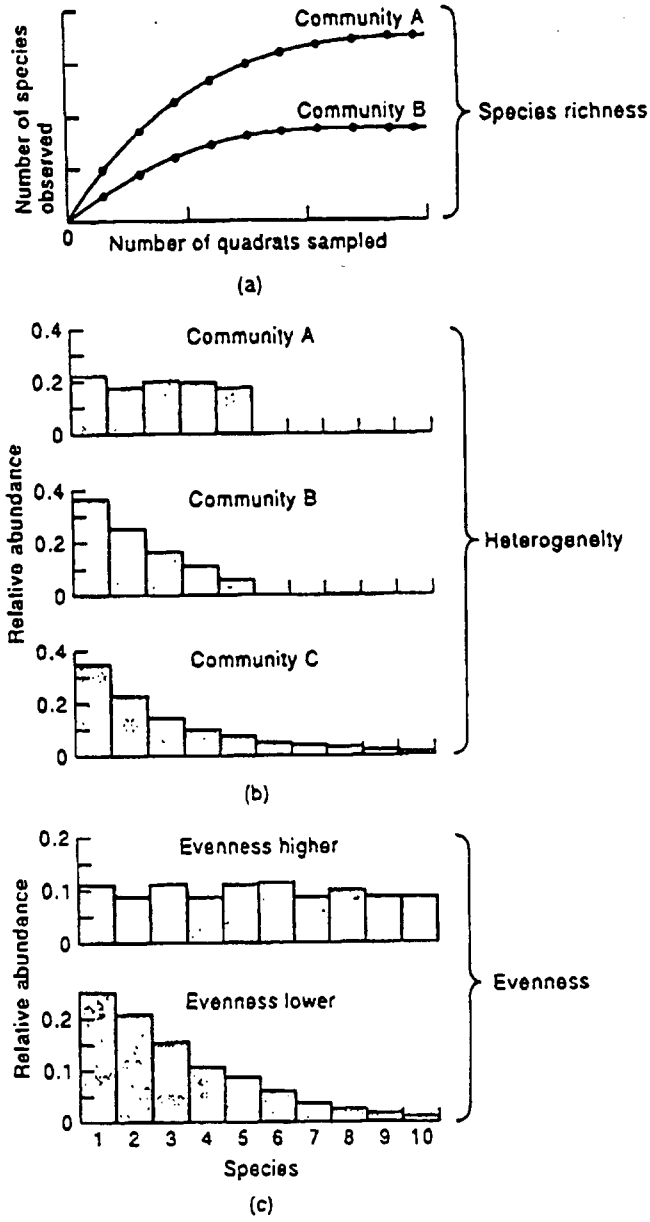


Figure 10.1 Concepts of species diversity. (a) *Species richness:* community A has more species than community B and thus higher species richness. (b) *Heterogeneity:* community A has the same number of species as community B but the relative abundances are more even, so by a heterogeneity measure A is more diverse than B. Community C has the same abundance pattern as B but has more species, so it is more diverse than B. (c) *Evenness:* when all species have equal abundances in the community, evenness is maximal.

Sanders' (1968) rarefaction algorithm was wrong, and it was corrected independently by Hurlbert (1971) and Simberloff (1972) as follows:

$$E(\hat{S}_n) = \sum_{i=1}^{s} \left[1 - \frac{\binom{N - N_i}{n}}{\binom{N}{n}} \right] \tag{10.1}$$

where $E(\hat{S}_n)$ = Expected number of species in a random sample of n individuals
S = Total number of species in the entire collection
N_i = Number of individuals in species i
N = Total number of individuals in collection = $\sum N_i$
n = Value of sample size (number of individuals) chosen for standardization ($n \leq N$)
$\binom{N}{n}$ = Number of combinations of n individuals that can be chosen from a set of N individuals = $N!/n!(N-n)!$

The large-sample variance of this estimate was given by Heck et al. (1975) as

$$\text{var}(\hat{S}_n) = \binom{N}{n}^{-1} \left[\sum_{i=1}^{s} \binom{N - N_i}{n} \left[1 - \frac{\binom{N - N_i}{n}}{\binom{N}{n}} \right] + 2 \sum_{i=1}^{s-1} \sum_{j=i+1}^{s} \left[\binom{N - N_i - N_j}{n} - \frac{\binom{N - N_i}{n}\binom{N - N_j}{n}}{\binom{N}{n}} \right] \right] \tag{10.2}$$

where $\text{var}(\hat{S}_n)$ = Variance of the expected number of species in a random sample of n individuals

and all other terms are defined above.

Box 10.1 illustrates the calculation of the rarefaction method for some rodent data. Because these calculations are so tedious, a computer program should normally be used for the rarefaction method. Program RAREFACT in Appendix 10.1 will do these calculations. It is modified from the program given by Simberloff (1978).

There are important ecological restrictions on the use of the rarefaction method. Since rarefaction is not concerned with species names, the samples to be compared by rarefaction should be taxonomically similar. As Simberloff (1979) points out, if the larger sample is primarily butterflies and the smaller sample is mostly moths, no calculations are necessary to tell you that the smaller sample might not be a random sample of the larger set.

Sampling methods must also be similar for two samples to be compared by rarefaction (Sanders, 1968). For example, you should not compare insect light trap samples with insect sweep net samples, since whole groups of species are amenable to capture in one technique but not available to the other.

BOX 10.1 CALCULATION OF EXPECTED NUMBER OF SPECIES BY THE RAREFACTION METHOD

A sample of Yukon rodents produced four species in a collection of 42 individuals. The species abundances were 21, 16, 3, and 2 individuals. We wish to calculate the expected species richness for samples of 30 individuals.

Expected Number of Species

From equation (10.1)

$$E(\hat{S}_n) = \sum_{i=1}^{s} \left[1 - \frac{\binom{N - N_i}{n}}{\binom{N}{n}} \right]$$

$$E(\hat{S}_{30}) = \left[1 - \frac{\binom{42 - 21}{30}}{\binom{42}{30}} \right] + \left[1 - \frac{\binom{42 - 16}{30}}{\binom{42}{30}} \right] + \left[1 - \frac{\binom{42 - 3}{30}}{\binom{42}{30}} \right] + \left[1 - \frac{\binom{42 - 2}{30}}{\binom{42}{30}} \right]$$

$$\binom{42 - 21}{30} = \binom{21}{30} = 0 \qquad \text{(by definition)}$$

$$\binom{42}{30} = \frac{42!}{30!(42 - 30)!} = 1.1058 \times 10^{10}$$

$$\binom{42 - 16}{30} = \binom{26}{30} = 0 \qquad \text{(by definition)}$$

$$\binom{42 - 3}{30} = \frac{39!}{30!(39 - 30)!} = 2.1192 \times 10^{8}$$

$$\binom{42 - 2}{30} = \frac{40!}{30!(40 - 30)!} = 8.4766 \times 10^{8}$$

$$E(\hat{S}_{30}) = 1 + 1 + \left(1 - \frac{2.1192 \times 10^{8}}{1.1058 \times 10^{10}}\right) + \left(1 - \frac{8.4766 \times 10^{8}}{1.1058 \times 10^{10}}\right)$$

$$= 1 + 1 + 0.981 + 0.923$$

$$= 3.90 \text{ species}$$

Large Sample Variance of the Expected Number of Species

From equation (10.2)

$$\operatorname{var}(\hat{S}_n) = \binom{N}{n}^{-1}\left[\sum_{i=1}^{s}\binom{N-N_i}{n}1-\frac{\binom{N-N_i}{n}}{\binom{N}{n}}+2\sum_{i=1}^{s-1}\sum_{j=i+1}^{s}\binom{N-N_i-N_j}{n}-\frac{\binom{N-N_i}{n}\binom{N-N_j}{n}}{\binom{N}{n}}\right]$$

$$\begin{aligned}\operatorname{var}(\hat{S}_{30}) = \binom{42}{30}^{-1}\Bigg\{&\binom{21}{30}\left(1-\frac{\binom{21}{30}}{\binom{42}{30}}\right)+\binom{26}{30}\left(1-\frac{\binom{26}{30}}{\binom{42}{30}}\right)\\&+\binom{39}{30}\left(1-\frac{\binom{39}{30}}{\binom{42}{30}}\right)+\binom{40}{30}\left(1-\frac{\binom{40}{30}}{\binom{42}{30}}\right)\\&+2\left[\left(\binom{42-21-16}{30}-\frac{\binom{42-21}{30}\binom{42-16}{30}}{\binom{42}{30}}\right)\right.\\&+\left(\binom{42-21-3}{30}-\frac{\binom{42-21}{30}\binom{42-3}{30}}{\binom{42}{30}}\right)\\&+\left(\binom{42-21-2}{30}-\frac{\binom{42-21}{30}\binom{42-2}{30}}{\binom{42}{30}}\right)\\&\left.+\left(\binom{42-16-3}{30}-\frac{\binom{42-16}{30}\binom{42-3}{30}}{\binom{42}{30}}\right)\right.\end{aligned}$$

$$+\left(\binom{42-16-2}{30}-\frac{\binom{42-16}{30}\binom{42-2}{30}}{\binom{42}{30}}\right)$$

$$+\left(\binom{42-3-2}{30}-\frac{\binom{42-3}{30}\binom{42-2}{30}}{\binom{42}{30}}\right)\Bigg]\Bigg\}$$

Note that for this particular example almost all of the terms are zero.

$$\text{var}(\hat{S}_{30}) = (1.1058 \times 10^{-10})[2.0785 \times 10^{8} + 7.8268 \times 10^{8} + (2)(-5.9499 \times 10^{6})]$$

$$= 0.0885$$

$$\text{Standard deviation of } (\hat{S}_{30}) = \sqrt{\text{var}(\hat{S}_{30})} = \sqrt{0.0885} = 0.297$$

These tedious calculations are done by Program RAREFACT (see Appendix 10.1).

Sanders (1968) argued that rarefaction should be used only on samples from the same or similar habitats, since everyone knows that different habitats—coniferous forest versus deciduous forest, for example—have different species diversities.

Rarefaction curves cannot be extrapolated beyond the number of individuals in the large sample. The only way one can extrapolate beyond the limits of the samples is by assuming an underlying statistical distribution, like the log-series or the lognormal.

One assumption that rarefaction does make is that all individuals in the community are randomly dispersed with respect to other individuals of their own or of different species. In practice, most distributions are clumped (see Chapter 3) within a species and there may be positive or negative association between species. Fager (1972) used computer simulations to investigate the effect of clumping on rarefaction estimates and observed that the more clumped the populations are, the greater the overestimation of the number of species by the rarefaction method. The only way to reduce this bias in practice is to use large samples spread widely throughout the community.

The variance of the expected number of species (equation 10.2) is appropriate only with reference to the sample under consideration. Suppose you wish to ask a related question: *given a sample of N individuals from a community, how many species would you expect to find in a second, independent sample of n ($n < N$) individuals?* Smith and Grassle (1977) give the variance estimate appropriate for this more general question and have a computer program for generating these variances. Simberloff (1979) showed that the variance given in equation (10.2) provides estimates only slightly smaller than the Smith and Grassle (1977) estimator.

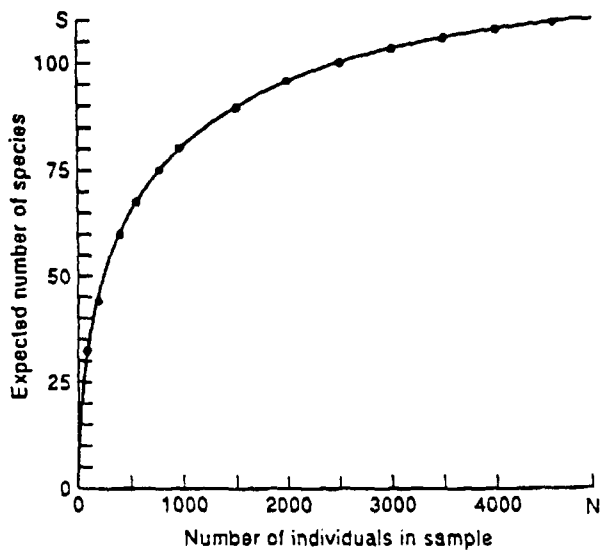


Figure 10.2 Rarefaction curve for the diatom community data from Patrick (1968). There were 4874 individuals in 112 species in this sample ("Box 8"). Original data in Table 10.1.

Figure 10.2 illustrates a rarefaction curve for the diatom community data in Table 10.1. James and Rathbun (1981) provide examples from bird communities.

10.3.2 Jackknife Estimate

When quadrat sampling is used to sample the community, it is possible to use another nonparametric approach, the jackknife,* to estimate species richness. This estimate is based on the observed frequency of rare species in the community and is obtained as follows (Heltshe and Forrester, 1983a). Data from a series of random quadrats are tabulated in the form shown in Table 10.2, recording only the presence (1) or absence (0) of the species in each quadrat. Tally the number of *unique species* in the quadrats sampled. A *unique species* is defined as a species that occurs in one and only one quadrat. Unique species are spatially rare species and are not necessarily numerically rare, since they could be highly clumped. From Heltshe and Forrester (1983b) the jackknife estimate of the number of species is

$$\hat{S} = s + \left(\frac{n-1}{n}\right)^{k} \tag{10.3}$$

where $\hat{S}$ = Jackknife estimate of species richness
s = Observed total number of species present in n quadrats
n = Total number of quadrats sampled
k = Number of unique species

* For a general discussion of jackknife estimates, see Chapter 13.

TABLE 10.1 TWO SAMPLES OF A DIATOM COMMUNITY OF A SMALL CREEK IN PENNSYLVANIA IN 1965[a]

Species	Numbers of individuals Box 8	Box 7	Species	Numbers of individuals Box 8	Box 7
Nitzschia frustulum v. *perminuta*	1446	1570	*Melosira italica* v. *valida*	6	15
Synedra parasitica v. *subconstricta*	456	455	*Navicula cryptocephala* v. *veneta*	6	6
Navicula cryptocephala	450	455	*Cymbella turgida*	5	8
Cyclotella stelligera	330	295	*Fragilaria intermedia*	5	5
Navicula minima	318	305	*Gomphonema angustatum* v. *obesa*	5	16
N. secreta v. *apiculata*	306	206	*G. angustatum* v. *producta*	5	4
Nitzschia palea	270	225	*G. longiceps* v. *subclavata*	5	9
N. frustulum	162	325	*Meridion circulare*	5	4
Navicula luzonensis	132	78	*Melosira ambigua*	5	—
Nitzschia frustulum v. *indica*	126	180	*Nitzschia acicularis*	5	—
Melosira varians	118	140	*Synedra rumpens* v. *familiaris*	5	37
Nitzschia amphibia	93	95	*Cyclotella meneghiniana*	4	8
Achnanthes lanceolata	75	275	*Gyrosigma spencerii*	4	2
Stephanodiscus hantzschii	74	59	*Fragilaria construens* v. *venter*	3	—
Navicula lanceolata	69	245	*Gomphonema gracile*	3	10
N. viridula	68	72	*Navicula cincta*	3	2
Rhoicosphenia curvata v. *minor*	61	121	*N. gracilis* fo. *minor*	3	—
Navicula minima v. *atomoides*	59	47	*Navicula decussis*	3	2
N. pelliculosa	54	19	*N. pupula* v. *capitata*	3	10
Melosira granulata v. *angustissima*	54	73	*N. symmetrica*	3	—
Navicula seminulum	52	36	*Nitzschia dissipata* v. *media*	3	4
N. gregaria	40	34	*N. tryblionella* v. *debilis*	3	1
Nitzschia capitellata	40	16	*N. sigmoidea*	3	—
Achnanthes subhudsonis v. *kraeuselii*	39	51	*Anomoeoneis exilis*	2	—
A. minutissima	35	61	*Caloneis hyalina*	2	2
Nitzschia dissipata	35	53	*Diatoma vulgare*	2	—
Amphora ovalis v. *pediculus*	33	53	*Eunotia pectinalis* v. *minor*	2	1
Cymbella tumida	29	95	*Fragilaria leptostauron*	2	3
Synedra parasitica	24	42	*Gomphonema constrictum*	2	—
Cymbella ventricosa	21	27	*G. intricatum* v. *pumila*	2	10
Navicula paucivisitata	20	12	*Navicula hungarica* v. *capitata*	2	5
Nitzschia kutzingiana	19	70	*N. protracta*	2	3
Gomphonema parvulum	18	66	*Synedra acus* v. *angustissima*	2	—
Rhoicosphenia curvata	18	22	*Bacillaria paradoxa*	1	—
Synedra ulna	18	36	*Cyclotella kutzingiana*	1	—
Surirella angustata	17	11	*Cymbella triangulum*	1	—
Synedra ulna v. *danica*	17	37	*Cocconeis* sp.	1	—
Navicula pupula	17	27	*Caloneis bacillum*	1	3
Achnanthes biporoma	16	32	*Fragilaria bicapitata*	1	—
Stephanodiscus astraea v. *minutula*	16	21	*Frustulia vulgaris*	1	—
Navicula germainii	13	19	*Gomphonema carolinense*	1	1
Denticula elegans	12	4	*G.* sp. [MH IV Ridley]	1	—
Gomphonema sphaerophorum	11	40	*Navicula capitata* v. *hungarica*	1	1
Synedra rumpens	11	13	*N. contenta* f. *biceps*	1	1
S. vaucheriae	11	14	*N. cincta* v. *rostrata*	1	—

(continued overleaf)

TABLE 10.1 (Continued)

Species	Numbers of individuals: Box 8	Box 7	Species	Numbers of individuals: Box 8	Box 7
Cocconeis placentula v. *euglypta*	10	5	*N. americana*	1	—
Navicula menisculus	10	5	*Nitzschia hungarica*	1	—
Nitzschia linearis	10	18	*N. sinuata* v. *tabularia*	1	—
Stephanodiscus invisitatus	10	22	*N. confinis*	1	5
Amphora ovalis	9	16	*Synedra pulchella* v. *lacerata*	1	1
Cymbella sinuata	9	5	*Surirella ovata*	1	3
Gyrosigma wormleyii	9	5	*Achnanthes cleveii*	—	2
Nitzschia fonticola	9	6	*Amphora submontana*	—	1
N. bacata	9	7	*Caloneis silicula* v. *ventricosa*	—	3
Synedra rumpens v. *meneghiniana*	9	17	*Eunotia lunaris*	—	2
Cyclotella meneghiniana small	8	4	*E. tenella*	—	1
Nitzschia gracilis v. *minor*	8	10	*Fragilaria pinnata*	—	3
N. frustulum v. *subsalina*	7	10	*Gyrosigma scalproides*	—	1
N. subtilis	7	16	*Gomphonema sparsistriata*	—	—
Cymbella affinis	6	3	*Meridion circulare* v. *constricta*	—	3
Cocconeis placentula v. *lineata*	6	13	*Navicula tenera*	—	3
			N. omissa	—	1
			N. ventralis	—	1
			N. mutica	—	1
			N. sp. [LL 30]	—	1
			N. mutica v. *cohnii*	—	1
			Nitzschia brevissima	—	1
			N. frequens	—	1

[a] The numbers of individuals settling on glass slides were counted. Data from Patrick (1968).

TABLE 10.2 QUADRAT SAMPLING DATA SUMMARIZED IN A FORM NEEDED FOR THE JACKKNIFE ESTIMATE OF SPECIES RICHNESS[a]

Species	Quadrat A	B	C	D	E	F	Row sum
1	1	0	0	1	1	0	3
2	0	1	0	0	0	0	1
3	1	1	1	1	1	0	5
4	0	1	0	0	1	0	2
5	1	1	1	1	1	1	6
6	0	0	0	0	1	0	1
7	0	0	1	1	1	1	4
8	1	1	0	0	1	1	4

[a] Only presence-absence data are required. *Unique species* are those whose row sums are 1 (species 2 and 6 in this example). 0 = absent; 1 = present.

The variance of this jackknife estimate of species richness is given by

$$\text{var}(\hat{S}) = \left(\frac{n-1}{n}\right)\left[\sum_{j=1}^{s}(j^2 f_j) - \frac{k^2}{n}\right] \tag{10.4}$$

where $\text{var}(\hat{S})$ = Variance of jackknife estimate of species richness
f_j = Number of quadrats containing j unique species (j = 1, 2, 3, . . . , s)
k = Number of unique species
n = Total number of quadrats sampled

This variance can be used to obtain confidence limits for the jackknife estimator as follows:

$$\hat{S} \pm t_\alpha \sqrt{\text{var}(\hat{S})} \tag{10.5}$$

where $\hat{S}$ = Jackknife estimator of species richness (equation 10.3)
t_α = Student's t value for $n - 1$ degrees of freedom for the appropriate value of α
$\text{var}(\hat{S})$ = Variance of $\hat{S}$ from equation (10.4)

Box 10.2 gives an example of these calculations.

The jackknife estimator of species richness tends to have a positive bias, that is, it tends to *overestimate* the number of species in a community (Heltshe and Forrester, 1983b). This bias is usually less than the negative bias of the observed number of species (S), which as a rule is always less than the true value of species richness in the community.

Note from equation (10.3) that the maximum value of the jackknife estimate of species richness is twice the observed number of species. Thus this approach cannot be used for communities with exceptionally large numbers of rare species or communities that have been sampled too little (so that S is low).

10.3.3 Bootstrap Procedure

An alternative method of estimating species richness from quadrat samples is to use the bootstrap procedure (Smith and van Belle, 1984). The bootstrap method* is related to the jackknife, but it requires simulation on a computer to obtain estimates. The essence of the bootstrap procedure is as follows: given a set of data on species presence-absence in a series of q quadrats (like Table 10.2):

1. Draw a random sample of size n from the q quadrats within the computer, using sampling *with* replacement; this is the "bootstrap sample."
2. Calculate the estimate of species richness from the equation (Smith and van Belle, 1984)

$$B(\hat{S}) = S + \sum (1 - p_i)^n \tag{10.6}$$

* See Chapter 13 for more discussion of the bootstrap method.

Box 10.2 JACKKNIFE ESTIMATE OF SPECIES RICHNESS FROM QUADRAT SAMPLES

Ten quadrats from the benthos of a coastal creek were analyzed for the abundance of 14 species (Heltshe and Forrester, 1983a). For a sample taken from a subtidal marsh creek, Pettaquamscutt River, Rhode Island, April 1978.

	Quadrat									
Species	1	2	3	4	5	6	7	8	9	10
Streblospio benedicti		13	21	14	5	22	13	4	4	27
Nereis succinea	2	2	4	4	1	1	1		1	6
Polydora ligni		1						1		
Scoloplos robustus	1		1	2		6			1	2
Eteone heteropoda			1	2			1			1
Heteromastus filiformis	1	1	2	1		1			1	5
*Capitella capitata**	1									
*Scolecolepides viridis**	2									
*Hypaniola grayi**		1								
*Branis clavata**			1							
Macoma balthica			3							2
Ampelisca abdita			5	1		2				3
*Neopanope texana**								1		
Tubificocodies sp.	8	36	14	19	3	22	6	8	5	41

NOTE: Blank entries in table are absent from quadrat.

Five species (marked with *) occur in only one quadrat and are thus defined as *unique species*. Thus, from equation (10.3),

$$\hat{S} = S + \left(\frac{n-1}{n}\right)k$$

$$\hat{S} = 14 + \left(\frac{9}{10}\right)(5)$$

$$= 18.5 \text{ species}$$

The variance, from equation (10.4), is

$$\text{var}(\hat{S}) = \left(\frac{n-1}{n}\right)\left[\sum_{j=1}^{s} (j^2 f_j) - \frac{k^2}{n}\right]$$

From the table we tally:

No. of unique spp., j	No. of quadrats containing j unique species, f_j
1	3 (i.e., quadrats 2, 3, and 8)
2	1 (i.e., quadrat 1)
3	0
4	0
5	0

10.4

Thus,

$$\text{var}(\hat{S}) = \left(\frac{9}{10}\right)\left[(1)^2(3) + 2^2(1) - \frac{5^2}{10}\right]$$

$$= 4.05$$

For this small sample, for 95% confidence $t_\alpha = 2.26$, and thus the 95% confidence interval would be approximately

$$18.5 \pm (2.26)(\sqrt{4.05})$$

or 13.9 to 23.1 species.

Program RICHNESS (Appendix 10.2), does these calculations for quadrat data.

where $B(\hat{S})$ = Bootstrap estimate of species richness
S = Observed number of species in original data
p_i = Proportion of the n bootstrap quadrats that have species i present

3. Repeat steps 1 and 2 N times in the computer, where N is between 50 and 200.

The variance of this bootstrap estimate is given by

$$\text{var}[B(\hat{S})] = \sum_i (1 - p_i)^n[1 - (1 - p_i)^n] + \sum\sum_{i \neq j} \{q_{ij}^n - [(1 - p_i)^n(1 - p_j)^n]\} \quad (10.7)$$

where $\text{var}[B(\hat{S})]$ = Variance of the bootstrap estimate of species richness
n, p_i, p_j = As defined above
q_{ij} = Proportion of the n bootstrap quadrats that have both species i and species j absent

Smith and van Belle (1984) recommend the jackknife estimator when the number of quadrats is small and the bootstrap estimator when the number of quadrats is large. The empirical meaning of "large" and "small" for natural communities remains unclear; perhaps $n = 20$ quadrats is a rough division for many community samples, but this is only a guess.

10.4 HETEROGENEITY MEASURES

The measurement of diversity by means of heterogeneity indices has proceeded along two relatively distinct paths. The first approach is to use statistical sampling theory to investigate how communities are structured. The logarithmic series was first applied by Fisher et al. (1943) to a variety of community samples. Preston (1948, 1962) applied

the lognormal distribution to community samples. Because of the empirical nature of these statistical distributions, other workers looked to information theory for appropriate measures of diversity. Arguments continue about the utility of both of these approaches since they are not theoretically justified (Washington, 1984; Hughes, 1986). But both approaches are widely used in diversity studies, and it would be premature to dismiss any index because it lacks comprehensive theoretical justification.

10.4.1 Logarithmic Series

One very characteristic feature of communities is that they contain comparatively few species that are common and comparatively large numbers of species that are rare. Since it is relatively easy to determine for any given area the *number of species* on the area and the *number of individuals* in each of these species, a great deal of information of this type has accumulated (Williams, 1964). The first attempt to analyze these data was made by Fisher et al. (1943).

In many faunal samples the number of species represented by a single specimen is very large, species represented by two specimens are less numerous, and so on until only a few species are represented by many specimens. Fisher et al. (1943) plotted the data and found that they fitted a "hollow curve" (Figure 10.3). They concluded that the data available were best fitted by the logarithmic series, which is a series with a finite sum whose terms can be written as a function of two parameters:

$$\alpha x, \frac{\alpha x^2}{2}, \frac{\alpha x^3}{3}, \frac{\alpha x^4}{4}, \ldots \tag{10.8}$$

where αx = Number of species in the total catch represented by *one* individual

$\frac{\alpha x^2}{2}$ = Number of species represented by two individuals, and so on

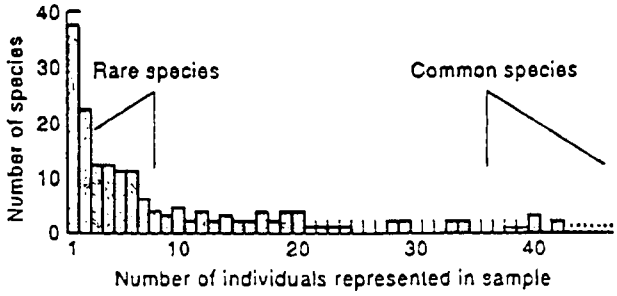


Figure 10.3 Relative abundance of Lepidoptera (butterflies and moths) captured in a light trap at Rothamsted, England, in 1935. Not all of the abundant species are shown. There were 37 species represented in the catch by only a single specimen (rare species); one very common species was represented by 1799 individuals in the catch. A total of 6814 individuals were caught, representing 197 species. Six common species made up 50 percent of the total catch. (*Source:* Williams, 1964.)

The sum of the terms in the series is equal to $-\alpha \log_e(1 - x)$, which is the total number of species in the catch. The logarithmic series for a set of data is fixed by two variables, the *number of species* in the sample and the *number of individuals* in the sample. The relationship between these is

$$S = \alpha \log_e\left(1 + \frac{N}{\alpha}\right) \tag{10.9}$$

where S = Number of species in sample
N = Number of individuals in sample
α = Index of diversity

The constant α is an expression of species diversity in the community. It is low when the number of species is low and high when the number of species is high.

There are several methods of fitting a logarithmic series to a set of species abundance data (Williams, 1964, Appendix A). Only two variables are needed to fit a logarithmic series: the total number of species in the sample (S) and the total number of individuals (N). Williams (1964, p. 311) and Southwood (1978, p. 431) provide nomograms from which α may be read directly from values of N and S. A more accurate procedure is to estimate an approximate value of x from Table 10.3 and then solve the following equation iteratively for a more accurate value of x:

$$\frac{S}{N} = \frac{1 - x}{x}[-\log_e(1 - x)] \tag{10.10}$$

where S = Total number of species in the sample
N = Total number of individuals in the sample
x = Parameter of logarithmic series (equation 10.8)

TABLE 10.3 RELATION BETWEEN VALUES OF x AND THE AVERAGE NUMBER OF UNITS PER GROUP (N/S) IN SAMPLES FROM POPULATIONS DISTRIBUTED ACCORDING TO THE LOGARITHMIC SERIES

x	N/S	x	N/S	x	N/S
0.50	1.443	0.97	9.214	0.9990	144.6
0.60	1.637	0.980	12.53	0.9992	175.1
0.70	1.938	0.985	15.63	0.9994	224.5
0.80	2.483	0.990	21.47	0.9996	319.4
0.85	2.987	0.991	23.38	0.9998	586.9
0.90	3.909	0.992	25.68	0.99990	1086.0
0.91	4.198	0.993	28.58	0.99995	2020
0.92	4.551	0.994	32.38	0.999990	8696
0.93	4.995	0.995	37.48	0.999995	16.390
0.94	5.567	0.996	45.11	0.9999990	71.430
0.95	6.340	0.997	57.21	—	—
0.96	7.458	0.998	80.33	—	—

Source: Williams, 1964, p. 308.

Trial values of x are used until this equation balances. Given this estimate of x, we obtain $\hat{\alpha}$ from

$$\hat{\alpha} = \frac{N(1-x)}{x} \tag{10.11}$$

where $\hat{\alpha}$ = Index of diversity from logarithmic series
N = Total number of individuals in sample

Program LOGSERIE (Appendix 10.3) does these calculations. Given α and x, the theoretical values of the entire logarithmic series can be calculated from equation (10.8).

The large-sample variance of the diversity index α was given by Anscombe (1950) as the equation

$$\text{var}(\hat{\alpha}) = \frac{\alpha}{-\log_e(1-x)} \tag{10.12}$$

where all terms are defined as above. Taylor et al. (1976) pointed out that many authors (including Williams, 1964) have used the wrong formula to calculate the variance of alpha.

To analyze any set of empirical community data, the first thing you should do is to plot a *species abundance curve*. Species abundance curves can be plotted in three different ways (May, 1975); on arithmetic or log scales:

y axis: Relative abundance, density, cover, or some measure of the importance of a species

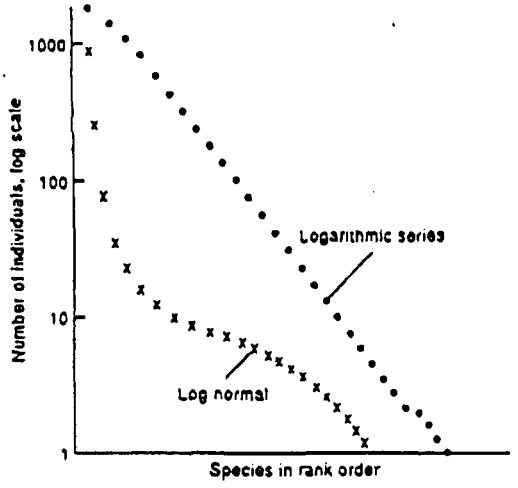
x axis: Rank of the n species from 1 (most abundant species) to n (most rare species)

By taking logs of the y or the x axis you can vary the shape of the resulting curves. Figure 10.4 illustrates a standard plot of species abundances, after Whittaker (1965). I call these *Whittaker plots* and recommend that the standard species abundance plot utilize log relative abundance-species ranks. The expected form of this curve for the logarithmic series is nearly a straight line and is shown in Figure 10.4.

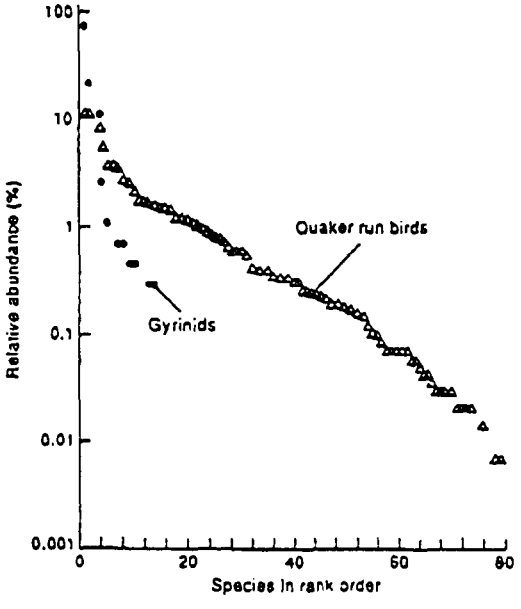
The theoretical Whittaker plot for a logarithmic series (e.g., Figure 10.4a) can be calculated as indicated in May (1975) by solving the following equation for n:

$$R = \alpha E_1\left[n \log_e\left(1 + \frac{\alpha}{N}\right)\right] \tag{10.13}$$

where R = Species rank (x axis, Figure 10.4) (i.e., 1, 2, 3, 4, . . . , s)
α = Index of diversity calculated in equation (10.11)
n = Number of individuals expected for specified value of R (y axis of Figure 10.4)
N = Total number of individuals in sample
E_1 = Standard exponential integral (Abramowitz and Stegun, 1964, Chapter 5)



(a)



(b)

Figure 10.4 Whittaker plots of species abundance data. (a) Theoretical plots. The logarithmic series produces a nearly straight line, while the lognormal distribution predicts a reverse S-shaped curve. (b) Observed data. The relative abundances of 11 species of Gyrinids from Mount Tremblant Park, Quebec (Lake Des Fammes), is quite well described by the logarithmic series (data from Williams, 1964, p. 271). The relative abundances of 79 species of birds from Quaker Run Valley, New York, is better described by the lognormal distribution (data from Williams, 1964, p. 49).

By solving this equation for n using integer values of R you can reconstruct the expected Whittaker plot and compare it to the original data. Program LOGSERIE (Appendix 10.3) has an option to calculate these theoretical values for a Whittaker plot.

There is considerable disagreement in the ecological literature about the usefulness of the logarithmic series as a measure of heterogeneity. Taylor et al. (1976) analyzed light trap catches of Macrolepidoptera from 13 sites in Britain, each site with 6 to 10 years of replicates. They showed that the logarithmic series parameter α was the best measure of species diversity for these collections. Hughes (1986), by contrast, examined 222 samples from many taxonomic groups and argued that the logarithmic series was a good fit for only 4% of these samples. May (1975) attempted to provide some theoretical justification for the logarithmic series as a description of species abundance patterns, but in most cases the logarithmic series is treated only as an empirical description of a sample from a community. Wolda (1983) concluded that alpha of the logarithmic series was the best measure of species diversity available.

Box 10.3 illustrates the calculation of the logarithmic series for a community of rodents.

The goodness of fit of the logarithmic series to a set of community data can be tested by the usual chi-square goodness-of-fit test (Taylor et al., 1976). But this chi-square test is of low power, and thus many samples are accepted as fitting the logarithmic series when it is in fact not a good fit (Routledge, 1980b). Thus, in most cases the decision on whether to use the logarithmic series to describe the diversity of a data set must be made on ecological grounds (Taylor et al., 1976; Hughes, 1986) rather than statistical goodness-of-fit criteria.

Koch (1987) used the logarithmic series to answer a critical methodological question in paleoecology: If two samples are taken from the same community, how many species will be found in both data sets and how many species will appear to be unique to one data set? Sample size effects may be critical in paleoecological studies, since absent species are typically classed as extinct. Koch (1987) used the logarithmic series and simple probability theory to predict the expected number of unique species in large samples from paleocommunities. These predictions can serve as a null model to compare with observed differences between samples. Figure 10.5 illustrates that the percentage of "unique species" can be very large when samples differ in size, even when the samples are taken from the same community. Rare species are inherently difficult to study in ecological communities, and sample size effects should always be evaluated before differences are assumed between two collections.

10.4.2 Lognormal Distribution

The logarithmic series implies that the greatest number of species has minimal abundance, that the number of species represented by a single specimen is always maximal. This is not the case in all communities. Figure 10.6 shows the relative abundance of breeding birds in Quaker Run Valley, New York. The greatest number of bird species is represented by 10 breeding pairs, and the relative abundance pattern does not fit the hollow-curve pattern of Figure 10.3. Preston (1948) suggested expressing the X axis (number of individuals represented in sample) on a geometric (logarithmic) scale

Box 10.3 FITTING A LOGARITHMIC SERIES TO SPECIES ABUNDANCE DATA

Krebs and Wingate (1976) sampled the small-mammal community in the Kluane region of the southern Yukon and obtained these results:

	No. of Individuals
Deer mouse	498
Northern red-backed vole	495
Meadow vole	111
Tundra vole	61
Long-tailed vole	45
Singing vole	40
Heather vole	23
Northern bog lemming	5
Meadow jumping mouse	5
Brown lemming	4
	$N = 1287$

$$S = 10$$

$$N/S = 128.7$$

From Table 10.3, an approximate estimate of x is 0.999. From equation (10.10) using this provisional estimate of x:

$$\frac{S}{N} = \left(\frac{1-x}{x}\right)[-\log_e(1-x)]$$

$$\frac{10}{1287} = \left(\frac{1-0.999}{0.999}\right)[-\log_e(1-0.999)]$$

$$0.007770 = 0.006915$$

Since the term on the right side is too small, we reduce the estimate of x. Try 0.99898:

$$0.007770 = \left(\frac{1-0.99898}{0.99898}\right)[-\log_e(1-0.99898)]$$

$$= 0.007033$$

The right side of the equation is still too small, so we reduce x to 0.99888:

$$0.007770 = \left(\frac{1-0.99888}{0.99888}\right)[-\log_e(1-0.99888)]$$

$$= 0.0076183$$

The right side of the equation is still slightly too small, so we repeat this calculation with $x = 0.998854$ to obtain, using equation (10.10):

$$0.007770 = 0.007769$$

We accept 0.998854 as an estimate of the parameter x of the logarithmic series.

From equation (10.11):

$$\hat{\alpha} = \frac{N(1-x)}{x}$$

$$\hat{\alpha} = \frac{1287(1-0.998854)}{0.998854}$$

$$= 1.4766$$

The variance of this estimate of α is, from equation (10.12),

$$\text{var}(\hat{\alpha}) = \frac{\hat{\alpha}}{-\log_e(1-x)}$$

$$= \frac{1.4766}{-\log_e(1-0.998854)}$$

$$= 0.2181$$

The individual terms of the logarithmic series are given by equation (9.8):

$$\alpha x, \frac{\alpha x^2}{2}, \frac{\alpha x^3}{3}, \ldots$$

i	No. of species represented by i individuals
1	1.475
2	0.737
3	0.491
4	0.367
5	0.294
6	0.244
7	0.209
⋮	⋮

The sum of the terms of this series is the number of species in the sample ($S = 10$).

Program LOGSERIE (Appendix 10.3) does these calculations.

rather than an arithmetic scale. One of several geometric scales can be used, since they differ only by a constant multiplier; a few scales are indicated in Table 10.4.

When this conversion of scale is done, relative abundance data take the form of a bell-shaped, normal distribution, and because the X axis is expressed on a geometric or logarithmic scale, this distribution is called lognormal. The lognormal distribution has been analyzed comprehensively by May (1975). It is completely specified by two parameters, although, as May (1975) shows, there are several ways of expressing the equation:

$$S_T = \frac{1.772454}{a} S_0 \qquad (10.14)$$

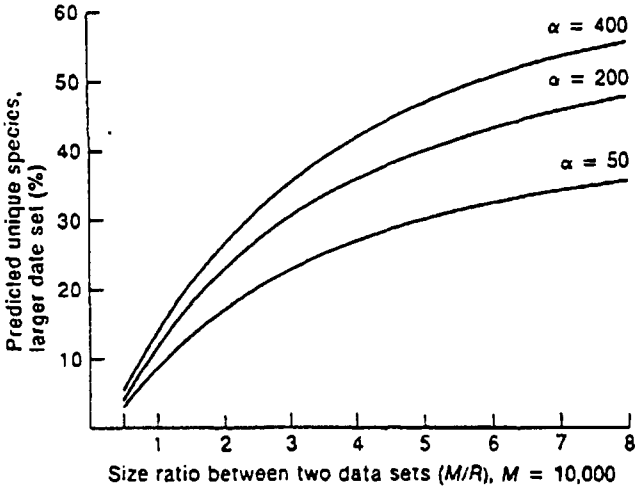


Figure 10.5 Use of the logarithmic series to predict the percentage of unique species in the larger of two data sets from the same biological community. Three values of α, the diversity parameter of the logarithmic series, are plotted. The larger data set is assumed to have a sample size of 10,000 individuals. These curves illustrate how difficult it is to sample the rare species in a diverse biological community. (*Source:* Koch, 1987.)

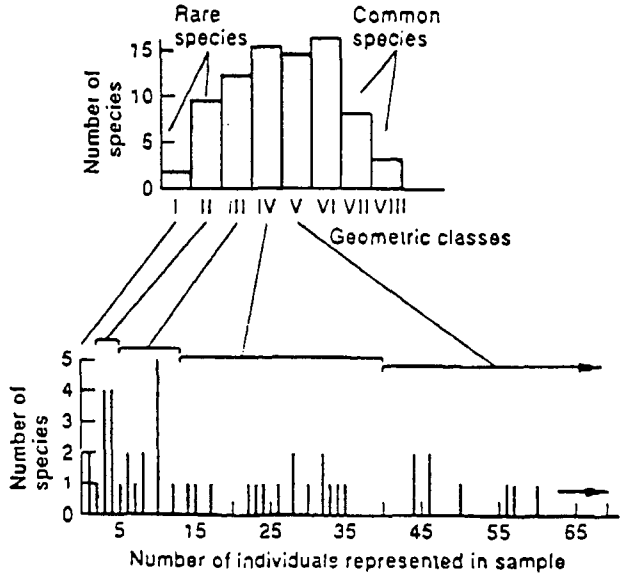


Figure 10.6 Relative abundance of nesting bird species in Quaker Run Valley, New York. The lower figure shows the distribution on an arithmetic scale, and the upper figure shows the same data on a geometric scale with ×3 size groupings (1, 2–4, 5–13, 14–40, 41–121, etc.). (*Source:* Williams, 1964.)

TABLE 10.4 GROUPINGS OF ARITHMETIC SCALE UNITS INTO GEOMETRIC SCALE UNITS FOR THREE TYPES OF GEOMETRIC SCALES[a]

Geometric scale no.	Arithmetic numbers grouped according to:		
	×2 Scale[b]	×3 Scale[c]	×10 Scale[d]
1	1	1–2	1–9
2	2–3	3–8	10–99
3	4–7	9–26	100–999
4	8–15	27–80	1,000–9,999
5	16–31	81–242	10,000–99,999
6	32–63	243–728	100,000–999,999
7	64–127	729–2,186	—
8	128–255	2,187–6,560	—
9	256–511	6,561–19,682	—

[a] This type of grouping is used in Figure 10.5.

[b] Octave scale of Preston (1948), equivalent to $\log_2$ scale.

[c] Equivalent to $\log_3$ scale.

[d] Equivalent to $\log_{10}$ scale.

where $\hat{S}_T$ = Total number of species in the community
a = Parameter measuring the spread of the lognormal distribution
S_0 = Number of species in the largest class

The lognormal distribution fits a variety of data from surprisingly diverse communities (Preston, 1948, 1962).

The shape of the lognormal curve is supposed to be characteristic for any particular community. Additional sampling of a community should move the lognormal curve to the right along the abscissa but not change its shape. Few communities have been sampled enough to test this idea. Figure 10.7 shows some data from moths caught in light traps which suggest that additional sampling moves the curve out toward the right. Since we cannot collect one-half or one-quarter of an animal, there will always be some rare species that are not represented in the catch. These rare species appear only when very large samples are taken.

Preston (1962) showed that data from lognormal distributions from biological communities commonly took on a particular configuration that he called the *canonical distribution*. Preston showed that for many cases $a \approx 0.2$ so that the entire lognormal distribution could be specified by *one* parameter:

$$\hat{S}_T = 5.11422S_0 \tag{10.15}$$

where $\hat{S}_T$ = Total number of species in the community
S_0 = Parameter measuring number of species in the modal (largest) class of the lognormal as defined above

Note that when the species abundance distribution is lognormal, it is possible to estimate the total number of species in the community, including rare species not yet collected. This is done by extrapolating the bell-shaped curve below the class of minimal abundance and measuring the area. Figure 10.8 illustrates how this can be

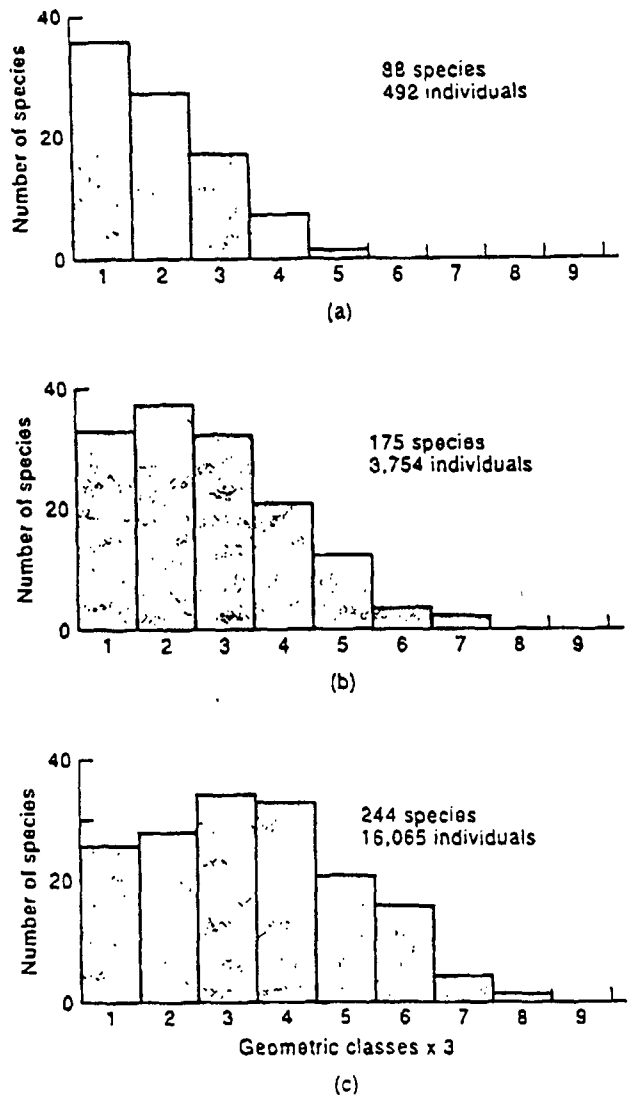


Figure 10.7 Lognormal distributions of the relative abundances of Lepidoteran insects captured in light traps at Rothamsted Experimental Station, England, in periods ranging from (a) 1/8 years to (b) 1 year to (c) 4 years. Note that the lognormal distribution slides to the right as the sample size is increased. (*Source:* Williams, 1964.)

done. This can be a useful property for communities where all the species cannot readily be seen and tabulated.

Although the lognormal distribution is an attractive model for species abundance relationships, in practice it is a very difficult distribution to fit to ecological data (Hughes,

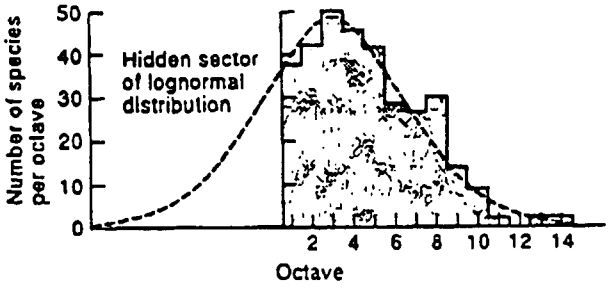


Figure 10.8 Species abundances in a collection of moths caught in a light trap. Data from Preston (1948). The lognormal distribution is truncated at the point where species are represented by a single individual. More intensive sampling would cause the distribution to move to the right and to unveil the hidden sector of rare species. (*Source:* Preston, 1948.)

1986). A sample should in practice be described by a truncated lognormal only if there is some evidence of a mode or maximum in the species abundance curve (e.g., Figures 10.6 and 10.8). Many authors have calculated a lognormal distribution from data like those in Figure 10.6a, which have no mode, but this should not be done. Hughes (1986) showed that parameter estimates from artificial lognormal distributions that did not include the mode were wildly inaccurate. The shape of the "true" lognormal distribution cannot be calculated from small samples, unless you have independent evidence that the first octave of your sample is close to the true mode for that community.

The lognormal distribution is a continuous statistical distribution, but species abundance data are discrete in terms of individuals. Strictly speaking, the species abundance data should be treated as Poisson variates, and one should fit the *Poisson lognormal* (= discrete lognormal) to most community data (Pielou, 1975, p. 49; Bulmer, 1974). The Poisson lognormal is difficult to compute, and Bulmer (1974) has discussed methods of evaluating it. For practical purposes the ordinary lognormal is usually fitted to species abundance data, using the maximum likelihood methods devised by Cohen (1959, 1961) and described by Pielou (1975, pp. 50–53). Gauch and Chase (1974) discussed a nonlinear regression method for fitting the lognormal distribution to species abundance data, but Hansen and Zeger (1978) showed that this regression method was not appropriate for species abundance data and recommended the method of Cohen.

To fit a lognormal distribution to species abundance data by the methods of Cohen (1959, 1961), proceed as follows:

1. Transform all the observed data (number of individuals, biomass, or other measure of species importance) logarithmically:

$$x_i = \log n_i \qquad (10.16)$$

where n_i = Observed number of individuals of species i in sample
i = Species counter ($i = 1, 2, 3, \ldots, S_0$)*
x_i = Transformed value for lognormal distribution

Any base of logarithms can be used as long as you are consistent. Here, I will use log base 10.

2. Calculate the observed mean and variance of x_i by the usual statistical formulas (see Appendix 1). Sample size is S_0, the observed number of species.

3. Calculate the parameter y:

$$y = \frac{s^2}{(\bar{x} - x_0)^2} \tag{10.17}$$

where y = Parameter of lognormal distribution
s^2 = Observed variance (calculated in step 2)
$\bar{x}$ = Observed mean (calculated in step 2)
$x_0 = \log(0.5) = -0.30103$ if using $\log_{10}$

4. From Table 10.5 obtain the estimate of θ corresponding to this estimate of y.

5. Obtain corrected estimates of the mean and variance of the lognormal distribution from the equations

$$\hat{\mu} = \bar{x} - \theta(\bar{x} - x_0) \tag{10.18}$$

$$\hat{\sigma}^2 = s^2 + \theta(\bar{x} - x_0)^2 \tag{10.19}$$

where $\hat{\mu}$ = Estimate of true mean of the lognormal
$\bar{x}$, s^2 = Observed mean and variance from step 2
θ = Correction factor from Table 10.5 (step 4)
x_0 = Truncation point of observed data = log(0.5)
$\hat{\sigma}^2$ = Estimate of true variance of the lognormal

6. Calculate the standardized normal deviate corresponding to the truncation point:

$$z_0 = \frac{x_0 - \mu}{\sigma} \tag{10.20}$$

7. From tables of the standard normal distribution (e.g., Rohlf and Sokal, 1981, p. 78), find the area (p_0) under the tail of the normal curve to the left of z_0. Then:

$$\hat{S}_T = \frac{S_0}{1 - p_0} \tag{10.21}$$

where $\hat{S}_T$ = Estimated number of species in the community (including those to the left of the veil line, e.g., Figure 10.8)
S_0 = Observed number of species in sample
p_0 = Area of standard normal curve to left of z_0

* To avoid confusion but maintain traditional symbols, I use S_0 for the number of species observed and s for the standard deviation.

TABLE 10.5 VALUES OF THE ESTIMATION FUNCTION θ CORRESPONDING TO VALUES OF y OBTAINED IN EQUATION (10.17)

y	.000	.001	.002	.003	.004	.005	.006	.007	.008	.009	y
0.05	.00000	.00000	.00000	.00001	.00001	.00001	.00001	.00001	.00002	.00002	0.05
0.06	.00002	.00003	.00003	.00003	.00004	.00004	.00005	.00006	.00007	.00007	0.06
0.07	.00008	.00009	.00010	.00011	.00013	.00014	.00016	.00017	.00019	.00020	0.07
0.08	.00022	.00024	.00026	.00028	.00031	.00033	.00036	.00039	.00042	.00045	0.08
0.09	.00048	.00051	.00055	.00059	.00063	.00067	.00071	.00075	.00080	.00085	0.09
0.10	.00090	.00095	.00101	.00106	.00112	.00118	.00125	.00131	.00138	.00145	0.10
0.11	.00153	.00160	.00168	.00176	.00184	.00193	.00202	.00211	.00220	.00230	0.11
0.12	.00240	.00250	.00261	.00272	.00283	.00294	.00305	.00317	.00330	.00342	0.12
0.13	.00355	.00369	.00382	.00396	.00410	.00425	.00440	.00455	.00470	.00486	0.13
0.14	.00503	.00519	.00536	.00553	.00571	.00589	.00608	.00627	.00646	.00665	0.14
0.15	.00685	.00705	.00726	.00747	.00769	.00791	.00813	.00835	.00858	.00882	0.15
0.16	.00906	.00930	.00955	.00980	.01006	.01032	.01058	.01085	.01112	.01140	0.16
0.17	.01168	.01197	.01226	.01256	.01286	.01316	.01347	.01378	.01410	.01443	0.17
0.18	.01476	.01509	.01543	.01577	.01611	.01646	.01682	.01718	.01755	.01792	0.18
0.19	.01830	.01868	.01907	.01946	.01986	.02026	.02067	.02108	.02150	.02193	0.19
0.20	.02236	.02279	.02323	.02368	.02413	.02458	.02504	.02551	.02599	.02647	0.20
0.21	.02695	.02744	.02794	.02844	.02895	.02946	.02998	.03050	.03103	.03157	0.21
0.22	.03211	.03266	.03322	.03378	.03435	.03492	.03550	.03609	.03668	.03728	0.22
0.23	.03788	.03849	.03911	.03973	.04036	.04100	.04165	.04230	.04296	.04362	0.23
0.24	.04429	.04497	.04565	.04634	.04704	.04774	.04845	.04917	.04989	.05062	0.24
0.25	.05136	.05211	.05286	.05362	.05439	.05516	.05594	.05673	.05753	.05834	0.25
0.26	.05915	.05997	.06080	.06163	.06247	.06332	.06418	.06504	.06591	.06679	0.26
0.27	.06768	.06858	.06948	.07039	.07131	.07224	.07317	.07412	.07507	.07603	0.27
0.28	.07700	.07797	.07896	.07995	.08095	.08196	.08298	.08401	.08504	.08609	0.28
0.29	.08714	.08820	.08927	.09035	.09144	.09254	.09364	.09476	.09588	.09701	0.29
0.30	.09815	.09930	.10046	.10163	.10281	.10400	.10520	.10641	.10762	.10885	0.30
0.31	.1101	.1113	.1126	.1138	.1151	.1164	.1177	.1190	.1203	.1216	0.31
0.32	.1230	.1243	.1257	.1270	.1284	.1298	.1312	.1326	.1340	.1355	0.32
0.33	.1369	.1383	.1398	.1413	.1428	.1443	.1458	.1473	.1488	.1503	0.33
0.34	.1519	.1534	.1550	.1566	.1582	.1598	.1614	.1630	.1647	.1663	0.34
0.35	.1680	.1697	.1714	.1731	.1748	.1765	.1782	.1800	.1817	.1835	0.35
0.36	.1853	.1871	.1889	.1907	.1926	.1944	.1963	.1982	.2001	.2020	0.36
0.37	.2039	.2058	.2077	.2097	.2117	.2136	.2156	.2176	.2197	.2217	0.37
0.38	.2238	.2258	.2279	.2300	.2321	.2342	.2364	.2385	.2407	.2429	0.38
0.39	.2451	.2473	.2495	.2517	.2540	.2562	.2585	.2608	.2631	.2655	0.39
0.40	.2678	.2702	.2726	.2750	.2774	.2798	.2822	.2847	.2871	.2896	0.40
0.41	.2921	.2947	.2972	.2998	.3023	.3049	.3075	.3102	.3128	.3155	0.41
0.42	.3181	.3208	.3235	.3263	.3290	.3318	.3346	.3374	.3402	.3430	0.42
0.43	.3459	.3487	.3516	.3545	.3575	.3604	.3634	.3664	.3694	.3724	0.43
0.44	.3755	.3785	.3816	.3847	.3878	.3910	.3941	.3973	.4005	.4038	0.44
0.45	.4070	.4103	.4136	.4169	.4202	.4236	.4269	.4303	.4338	.4372	0.45
0.46	.4407	.4442	.4477	.4512	.4547	.4583	.4619	.4655	.4692	.4728	0.46
0.47	.4765	.4802	.4840	.4877	.4915	.4953	.4992	.5030	.5069	.5108	0.47
0.48	.5148	.5187	.5227	.5267	.5307	.5348	.5389	.5430	.5471	.5513	0.48
0.49	.5555	.5597	.5639	.5682	.5725	.5768	.5812	.5856	.5900	.5944	0.49
0.50	.5989	.6034	.6079	.6124	.6170	.6216	.6263	.6309	.6356	.6404	0.50
0.51	.6451	.6499	.6547	.6596	.6645	.6694	.6743	.6793	.6843	.6893	0.51
0.52	.6944	.6995	.7046	.7098	.7150	.7202	.7255	.7308	.7361	.7415	0.52
0.53	.7469	.7524	.7578	.7633	.7689	.7745	.7801	.7857	.7914	.7972	0.53
0.54	.8029	.8087	.8146	.8204	.8263	.8323	.8383	.8443	.8504	.8565	0.54
0.55	.8627	.8689	.8751	.8813	.8876	.8940	.9004	.9068	.9133	.9198	0.55
0.56	.9264	.9330	.9396	.9463	.9530	.9598	.9666	.9735	.9804	.9874	0.56
0.57	.9944	1.001	1.009	1.016	1.023	1.030	1.037	1.045	1.052	1.060	0.57

TABLE 10.5 (*Continued*)

y	.000	.001	.002	.003	.004	.005	.006	.007	.008	.009	y
0.58	1.067	1.075	1.082	1.090	1.097	1.105	1.113	1.121	1.129	1.137	0.58
0.59	1.145	1.153	1.161	1.169	1.177	1.185	1.194	1.202	1.211	1.219	0.59
0.60	1.228	1.236	1.245	1.254	1.262	1.271	1.280	1.289	1.298	1.307	0.60
0.61	1.316	1.326	1.335	1.344	1.353	1.363	1.373	1.382	1.392	1.402	0.61
0.62	1.411	1.421	1.431	1.441	1.451	1.461	1.472	1.482	1.492	1.503	0.62
0.63	1.513	1.524	1.534	1.545	1.556	1.567	1.578	1.589	1.600	1.611	0.63
0.64	1.622	1.634	1.645	1.657	1.668	1.680	1.692	1.704	1.716	1.728	0.64
0.65	1.740	1.752	1.764	1.777	1.789	1.802	1.814	1.827	1.840	1.853	0.65
0.66	1.866	1.879	1.892	1.905	1.919	1.932	1.946	1.960	1.974	1.988	0.66
0.67	2.002	2.016	2.030	2.044	2.059	2.073	2.088	2.103	2.118	2.133	0.67
0.68	2.148	2.163	2.179	2.194	2.210	2.225	2.241	2.257	2.273	2.290	0.68
0.69	2.306	2.322	2.339	2.356	2.373	2.390	2.407	2.424	2.441	2.459	0.69
0.70	2.477	2.495	2.512	2.531	2.549	2.567	2.586	2.605	2.623	2.643	0.70
0.71	2.662	2.681	2.701	2.720	2.740	2.760	2.780	2.800	2.821	2.842	0.71
0.72	2.863	2.884	2.905	2.926	2.948	2.969	2.991	3.013	3.036	3.058	0.72
0.73	3.081	3.104	3.127	3.150	3.173	3.197	3.221	3.245	3.270	3.294	0.73
0.74	3.319	3.344	3.369	3.394	3.420	3.446	3.472	3.498	3.525	3.552	0.74
0.75	3.579	3.606	3.634	3.662	3.690	3.718	3.747	3.776	3.805	3.834	0.75
0.76	3.864	3.894	3.924	3.955	3.986	4.017	4.048	4.080	4.112	4.144	0.76
0.77	4.177	4.210	4.243	4.277	4.311	4.345	4.380	4.415	4.450	4.486	0.77
0.78	4.52	4.56	4.60	4.63	4.67	4.71	4.75	4.79	4.82	4.86	0.78
0.79	4.90	4.94	4.99	5.03	5.07	5.11	5.15	5.20	5.24	5.28	0.79
0.80	5.33	5.37	5.42	5.46	5.51	5.56	5.61	5.65	5.70	5.75	0.80
0.81	5.80	5.85	5.90	5.95	6.01	6.06	6.11	6.17	6.22	6.28	0.81
0.82	6.33	6.39	6.45	6.50	6.56	6.62	6.68	6.74	6.81	6.87	0.82
0.83	6.93	7.00	7.06	7.13	7.19	7.26	7.33	7.40	7.47	7.54	0.83
0.84	7.61	7.68	7.76	7.83	7.91	7.98	8.06	8.14	8.22	8.30	0.84
0.85	8.39	8.47	8.55	8.64	8.73	8.82	8.91	9.00	9.09	9.18	0.85

* These values are used to fit the lognormal distribution to species abundance data.

Source: Cohen, 1961.

In the notation of equation (10.14), note that

$$\hat{a} = \frac{1}{\sqrt{2\hat{\sigma}^2}} \tag{10.22}$$

where $\hat{a}$ = Parameter measuring the spread of the lognormal distribution
$\hat{\sigma}^2$ = True variance of the lognormal (equation 10.19)

The variance of these estimates of the parameters of the lognormal distribution can be estimated, following Cohen (1961), as

$$\text{var}(\hat{\mu}) = \frac{\mu_{11}\hat{\sigma}^2}{S_0} \tag{10.23}$$

where $\text{var}(\hat{\mu})$ = Estimated variance of mean of the lognormal
μ_{11} = Constant from Table 10.6
$\hat{\sigma}^2$ = Estimate of true variance of lognormal (equation 10.19)
S_0 = Observed number of species in sample

TABLE 10.6 FACTORS FOR ESTIMATING THE VARIANCE OF THE MEAN AND STANDARD DEVIATION OF A LOGNORMAL DISTRIBUTION*

	For truncated samples			For truncated samples	
z_0	μ_{11}	μ_{22}	z_0	μ_{11}	μ_{22}
−4.0	1.00054	.502287	0.0	22.1875	4.03126
−3.5	1.00313	.510366	0.1	27.1403	4.46517
−3.0	1.01460	.536283	0.2	33.1573	4.94678
−2.5	1.05738	.602029	0.3	40.4428	5.48068
−2.4	1.07437	.622786	0.4	49.2342	6.07169
−2.3	1.09604	.646862	0.5	59.8081	6.72512
−2.2	1.12365	.674663	0.6	72.4834	7.44658
−2.1	1.15880	.706637	0.7	87.6276	8.24204
−2.0	1.20350	.743283	0.8	105.66	9.1178
−1.9	1.26030	.785158	0.9	127.07	10.081
−1.8	1.33246	.832880	1.0	152.40	11.138
−1.7	1.42405	.887141	1.1	182.29	12.298
−1.6	1.54024	.948713	1.2	217.42	13.567
−1.5	1.68750	1.01846	1.3	258.61	14.954
−1.4	1.87398	1.09734	1.4	306.78	16.471
−1.3	2.10982	1.18642	1.5	362.91	18.124
−1.2	2.40764	1.28690	1.6	428.11	19.922
−1.1	2.78311	1.40009	1.7	503.57	21.874
−1.0	3.25557	1.52746	1.8	591.03	24.003
−0.9	3.84879	1.67064	1.9	691.78	26.311
−0.8	4.59189	1.83140	2.0	807.71	28.813
−0.7	5.52036	2.01172	2.1	940.38	31.511
−0.6	6.67730	2.21376	2.2	1091.4	34.405
−0.5	8.11482	2.43990	2.3	1265.4	37.575
−0.4	9.89562	2.69271	2.4	1458.6	40.858
−0.3	12.0949	2.97504	2.5	1677.8	44.392
−0.2	14.8023	3.28997			
−0.1	18.1244	3.64083			

* The table is entered with a value of z_0 as calculated in equation (10.20).

Source: Cohen, 1961.

The variance of the standard deviation of the lognormal is given by

$$\text{var}(\hat{\sigma}) = \frac{\mu_{22}\hat{\sigma}^2}{s_0} \tag{10.24}$$

where $\text{var}(\hat{\sigma})$ = Variance of estimated standard deviation of the lognormal
μ_{22} = Constant from Table 10.6
$\hat{\sigma}^2$ = True variance of lognormal
s_0 = Observed number of species in the sample

These two variances may be used to set confidence limits for the mean and standard deviation in the usual way. The goodness of fit of the calculated lognormal distribution can be determined by a chi-square test (example in Pielou, 1975, p. 51) or by a non-parametric Kolmogorov-Smirnov test.

Unfortunately, no estimate is available of the precision of S_T (equation 10.21), and this is the parameter of the lognormal we are most interested in (Pielou, 1975; Slocomb and Dickson, 1978). Simulation work on artificial diatom communities by Slocomb and Dickson (1978) showed that unreliable estimates of S_T were a serious problem unless sample sizes were very large (>1000 individuals) and the number of species in the sample was 80% or more of the total species in the community. Such large-scale sampling is rare in the most species-rich communities that we might wish to fit with the lognormal distribution.

Program LOGNORM (Appendix 10.4) fits a truncated lognormal distribution to species abundance data and calculates an expected distribution, using the approach outlined in Pielou (1975).

Box 10.4 illustrates these calculations for a lognormal distribution.

10.4.3 Simpson's Index

Partly because of the complexity of the logarithmic series and the lognormal distribution and the lack of a theoretical justification for these statistical approaches, ecologists have turned to a variety of nonparametric measures of heterogeneity that make no assumptions about the shape of species abundance curves. The first nonparametric measure of diversity was proposed by Simpson (1949). Simpson suggested that diversity is inversely related to the probability that two individuals picked at random belong to the same species. For an infinite population this is given by

$$D = \sum p_i^2 \tag{10.25}$$

where D = Simpson's index
p_i = Proportion of species i in the community

To convert this probability to a measure of diversity, most workers have suggested using the complement of Simpson's original measure:

$$\text{Simpson's index of diversity} = \left\{\begin{array}{l}\text{Probability of picking two}\\ \text{organisms at random that}\\ \text{are different species}\end{array}\right\}$$

$$= 1 - \left\{\begin{array}{l}\text{Probability of picking two}\\ \text{organisms that are the}\\ \text{same species)}\end{array}\right\}$$

Thus,

$$1 - D = 1 - \sum (p_i)^2 \tag{10.26}$$

where $1 - D$ = Simpson's index of diversity
p_i = Proportion of individuals of species i in the community

Strictly speaking, this formula can be used to estimate Simpson's index only for an infinite population. Pielou (1969) showed that for a finite population the appropriate estimator is

$$1 - \hat{D} = 1 - \sum_{i=1}^{s} \left[\frac{n_i(n_i - 1)}{N(N - 1)}\right] \tag{10.27}$$

Box 10.4 FITTING A TRUNCATED LOGNORMAL TO SPECIES ABUNDANCE DATA

Kempton and Taylor (1974) provided moth data for site 49, Fort Augustus, Scotland, in 1969; 4534 individuals were collected in 165 species:

Individuals per species	Midpoint of interval, x	Observed no. of species, f_x
1	1	24
2–3	2.5	22
4–7	5.5	30
8–15	11.5	22
16–31	23.5	30
32–63	47.5	21
64–127	99.5	9
128–255	191.5	7

1. Calculate the mean and variance of the transformed data (log base 10) using the formulas for grouped data:

$$\bar{x} = \frac{\sum xf_x}{\sum f_x}$$

$$= \frac{(\log 1)(24) + (\log 2.5)(22) + (\log 5.5)(30) + \cdots}{24 + 22 + 30 + 22 + 30 + 21 + 9 + 7}$$

$$= \frac{164.5991}{165}$$

$$= 0.99757$$

$$s^2 = \frac{\sum x^2 f_x - (\sum xf_x)^2/n}{n - 1}$$

$$= 0.41642$$

2. Estimate the parameter y from equation (10.17):

$$y = \frac{s^2}{(\bar{x} - x_0)^2}$$

$$= \frac{0.41642}{[0.99757 - (-0.30103)]^2}$$

$$= 0.24693$$

3. From Table 10.5, interpolating between y of 0.246 and 0.247,

$$\theta = 0.04912$$

4. Correct the observed mean and variance for the effects of truncation from equations (10.18) and (10.19):

$$\hat{\mu} = \bar{x} - \theta(\bar{x} - x_0)$$
$$= 0.99757 - (0.04912)[0.99757 - (-0.30103)]$$
$$= 0.93378$$
$$\hat{\sigma}^2 = s^2 + \theta(\bar{x} - x_0)^2$$
$$= 0.41642 + (0.04912)[0.99757 - (-0.30103)]^2$$
$$= 0.49925$$

5. Calculate the standard normal deviate corresponding to the truncation point from equation (10.20):

$$z_0 = \frac{x_0 - \hat{\mu}}{\hat{\sigma}}$$
$$= \frac{-0.30103 - 0.93378}{\sqrt{0.49925}}$$
$$= -1.7476$$

6. From Table 11 of Rohlf and Sokal (1981) obtain the area under the normal curve to the left of z_0:

$$\hat{p}_0 = 0.02005$$

7. From equation (10.21) calculate the estimated number of species in the whole community:

$$S_T = \frac{S_0}{1 - \hat{p}_0}$$
$$= \frac{165}{1 - 0.02005}$$
$$= 168.4 \text{ species}$$

Kempton and Taylor (1974) cautioned that this fitting procedure may give inexact parameter estimates compared with Bulmer's (1974) procedure.

where n_i = Number of individuals of species i in the sample
N = Total number of individuals in the sample = $\Sigma\, n_i$
s = Number of species in the sample

Note that this formula (10.27) can be used only when there are counts of individuals in the samples. When cover, biomass, or productivity is used as the measure of species importance, the previous equation (10.26) must be used. In practice, with a large sample there is almost no difference between these two equations.

There is some confusion in the literature over what should be called "Simpson's index." Washington (1984) argues strongly for maintaining Simpson's original for-

mulation, in which case equations (10.26) and (10.27) are the *complement* of Simpson's diversity. To confuse matters further, Williams (1964) and MacArthur (1972) used the reciprocal of Simpson's original formulation:

$$\frac{1}{D} = \frac{1}{\sum p_i^2} \tag{10.28}$$

where $1/D$ = Simpson's reciprocal index (= Hill's N_2)
p_i = Proportion of species i in the community

Hill (1973) called this reciprocal N_2.

Simpson's index ($1 - D$) ranges from 0 (low diversity) to almost 1 ($1 - 1/s$). The reciprocal of Simpson's original formulation ($1/D$) varies from 1 to s, the number of species in the sample. In this form, Simpson's diversity can be most easily interpreted as the number of equally common species required to generate the observed heterogeneity of the sample.

Diversity is almost always measured by a sample from a community, and it is virtually impossible for an ecologist to obtain a simple random sample (Pielou, 1969; Routledge, 1980a, 1980b). One way around this problem is to treat the community sample as a *collection*, or a complete statistical "universe," and make inferences about this finite collection (Pielou, 1966). Another approach is to use sampling units such as quadrats for plants or nets for insects and estimate diversity using a jackknife procedure. Zahl (1977) was the first to propose using this procedure to provide confidence estimates for Simpson's diversity measure. Routledge (1980a) showed that small samples (<30 quadrats) could give biased estimates for Simpson's diversity ($1 - D$ is underestimated), especially when fewer than 10 quadrats were counted. Heltshe and Forrester (1985) suggested that the jackknife estimate of confidence limits for Simpson's diversity ($1 - D$) was too large when applied to clumped populations, causing excessively wide confidence limits when more than 40 quadrats were sampled in their artificial populations. This overestimation depended on the exact shape of the species abundance curves.

Jackknife procedures for estimating Simpson's index of diversity and its confidence limits from quadrat samples are outlined clearly in Routledge (1980a). Lyons and Hutcheson (1986) proposed an alternative method for estimating confidence limits for Simpson's diversity using Pearson curves, but there was little improvement over the jackknife procedure.

Peet (1974) recognized two categories of diversity indices. *Type I* indices are most sensitive to changes in the rare species in the community sample. The Shannon-Wiener index is an example of a type I index. *Type II* indices are most sensitive to changes in the more abundant species. Simpson's index is an example of a type II index. The choice of heterogeneity measure to use on your data should be made on this basis—are you more interested in emphasizing the dominant or the rare species in your community?

Program DIVERS (Appendix 10.5) calculates Simpson's index of diversity from species abundance data.

10.4.4 Shannon-Wiener Function

The most popular measures of species diversity are based on information theory. The main objective of information theory is to try to measure the amount of *order* (or disorder) contained in a system (Margalef, 1958). Four types of information might be collected regarding *order* in the community: (1) the number of species, (2) the number of individuals in each species, (3) the places occupied by individuals of each species, and (4) the places occupied by individuals as separate individuals. In most community work only data of types 1 and 2 are obtained.

Information theory, Margalef suggested, provides one way to escape some of the difficulties of the lognormal curve and the logarithmic series. We ask the question: How difficult would it be to predict correctly the species of the next individual collected? This is the same as the problem faced by communication engineers interested in predicting correctly the name of the next letter in a message. This uncertainty can be measured by the Shannon-Wiener function:*

$$H' = \sum_{i=1}^{s} (p_i)(\log_2 p_i) \tag{10.29}$$

where H' = Information content of sample (bits/individual)
= Index of species diversity
s = Number of species
p_i = Proportion of total sample belonging to ith species

Information content is a measure of the amount of uncertainty, so the larger the value of H', the greater the uncertainty. A message such as bbbbbbb (or a community with only one species in it) has no uncertainty in it, and $H' = 0$. Any base of logarithms can be used for this index, since they are all convertible to one another by a constant multiplier:

$$H' \text{ (base 2 logs)} = 3.321928\ H' \text{ (base 10 logs)}$$

$$H' \text{ (base } e \text{ logs)} = 2.302585\ H' \text{ (base 10 logs)}$$

If base 2 logs are used, the units of H' are *bits per individual;* if base e logs, *nits;* and if base 10 logs, *decits.*

Strictly speaking, the Shannon-Wiener measure of information content should be used only on random samples drawn from a large community in which the total number of species is known (Pielou, 1966). For most community samples this is not the case, and Pielou (1966) thus recommends using the more appropriate Brillouin index.

The Shannon-Wiener measure H' increases with the number of species in the community and in theory can reach very large values. In practice, for biological communities H' does not seem to exceed 5.0 (Washington, 1984). The theoretical maximum value is $\log(S)$, and the minimum value (when $N \gg S$) is $\log[N/(N - S)]$ (Fager, 1972).

* This function was derived independently by Shannon and Wiener and is sometimes mislabeled the Shannon-Weaver function.

Many workers have used H' as a measure of species diversity, but the information theoretic approach has been heavily criticized by Hurlbert (1971) and by Washington (1984). The decision to use H' as a measure of species diversity should be made more on empirical grounds than on theoretical grounds. For example, Taylor et al. (1976) showed that α of the logarithmic series was a better diversity statistic than H' because α varied less in replicate samples of moths taken at the same site over several years.

Sampling distributions for the Shannon-Wiener index H' have been determined by Good (1953) and Basharin (1959), but these standard errors of H' are valid only if you have a simple random sample from the community. This is never the case in field data when nets, traps, quadrats, or transects are used for sampling (Kempton, 1979). Adams and McCune (1979) showed that estimates of H' from field data are usually biased, with the observed H' less than the true H', and that the jackknife technique could be used to reduce this bias and to estimate standard errors for H' so that confidence limits might be calculated. Zahl (1977) and Routledge (1980a) presented jackknife estimators for the Shannon-Wiener function when data are collected by quadrat sampling. Adams and McCune (1979) have prepared a computer program for jackknifing the Shannon-Wiener function.

The Shannon-Wiener index may be expressed in another form (MacArthur, 1965):

$$N_1 = e^{H'} \tag{10.30}$$

where $e = 2.71828$
H' = Shannon-Wiener function (calculated with base e logs)
N_1 = Number of equally common species which would produce the same diversity as H'

If a different base of logarithms is used, replace e with the base used. Hill (1973) recommends using N_1 rather than H' because the units (number of species) are more clearly understandable to ecologists. Peet (1974) recommends N_1 as the best type I heterogeneity measure.

Program DIVERS (Appendix 10.5) calculates the Shannon-Wiener function for species abundance data.

10.4.5 Brillouin Index

Many community samples should be treated as collections rather than as a random sample from a large biological community, according to Pielou (1966). In any case in which the data can be assumed to be a finite collection and sampling is done without replacement, the appropriate information theoretic measure of diversity is Brillouin's formula:

$$H = \frac{1}{N} \log\left(\frac{N!}{n_1!n_2!n_3!\cdots}\right) \tag{10.31}$$

where H = Brillouin's index
N = Total number of individuals in entire collection
n_1 = Number of individuals belonging to species 1
n_2 = Number of individuals belonging to species 2 (etc.)

Any base of logarithms may be used, as with the Shannon function. If base 2 logs are used, the units of H are *bits* per individual. Margalef (1958) was the first to propose using Brillouin's index as a measure of diversity.

There is much argument in the literature about whether the Brillouin index or the Shannon-Wiener function is a better measure of species diversity (Peet, 1974; Washington, 1984). In practice, this argument is irrelevant to field ecologists because H and H' are nearly identical for most ecological samples (when N is large). Legendre and Legendre (1983) also point out that Brillouin's index cannot be used when biomass, cover, or productivity is used as a measure of species importance in a community. Only the *number* of individuals can be used in equation (10.31).

If the Brillouin index is applied to quadrats, the mean and standard error of the Brillouin index can be estimated by the jackknife procedure (Heltshe and Forrester, 1985).

The Brillouin index is like the Shannon function in being most sensitive to the abundances of the rare species in the community. It is thus a type I index (Peet, 1974).

Program DIVERS (Appendix 10.5) calculates the Brillouin index from species abundance data (counts of individuals).

Box 10.5 illustrates the calculation of Simpson's index, the Shannon-Wiener function, and Brillouin's index for a forest community.

10.5 EVENNESS MEASURES

Many different measures of evenness (or *equitability*) have been proposed. The most common approach has been to scale one of the heterogeneity measures relative to its maximal value when each species in the sample is represented by the same number of individuals. Two formulations are possible:

$$\text{Evenness} = \frac{D}{D_{MAX}}$$

$$\text{Evenness} = \frac{D - D_{MIN}}{D_{MAX} - D_{MIN}}$$

where D = Observed index of species diversity
D_{MAX} = Maximum possible index of diversity, given S species and N individuals
D_{MIN} = Minimum possible index of diversity, given S and N

These two measures (labeled V' and V by Hurlbert, 1971) are convergent for large samples, and evenness measures of the first type (V') are most commonly used in the literature. All evenness measures range from 0 to 1.

For Simpson's measure of heterogeneity, maximum diversity is obtained when all abundances are equal ($p = 1/S$), so in a very large population:

$$\hat{D}_{MAX} = \frac{1}{S} \qquad (10.32)$$

Box 10.5 CALCULATION OF SIMPSON'S INDEX, THE SHANNON-WIENER FUNCTION, AND BRILLOUIN'S INDEX OF SPECIES DIVERSITY

Hough (1936) tallied the abundance of large trees in a virgin forest in Pennsylvania:

Tree species	No. of individuals, n_i	Proportional abundance, p_i
Hemlock	1940	0.521
Beech	1207	0.324
Yellow birch	171	0.046
Sugar maple	134	0.036
Black birch	97	0.026
Red maple	93	0.025
Black cherry	34	0.009
White ash	22	0.006
Basswood	15	0.004
Yellow poplar	7	0.002
Magnolia	4	0.001
Total	3724	1.000

Simpson's Index

From equation (10.25):

$$\hat{D} = \sum \hat{p}_i^2$$
$$= 0.521^2 + 0.324^2 + 0.046^2 + 0.036^2 + \cdots$$
$$= 0.381$$

The two indices of diversity follow from equations (10.26) and (10.28):

$$1 - \hat{D} = 1 - 0.381$$
$$= 0.619$$

This measure is the probability that two individuals chosen at random will be different species.

$$\frac{1}{\hat{D}} = \frac{1}{0.381}$$
$$= 2.623 \text{ species}$$

This is the number of equally common species required to produce the observed value of D.

Note that with this large sample the finite-population formula (equation 10.27) gives results identical to equation (10.26).

Shannon-Wiener Function

From equation (10.29):

$$\hat{H}' = -\sum \hat{p}_i \log_2 \hat{p}_i$$
$$= (0.521)(\log_2 0.521) + (0.324)(\log_2 0.324)$$
$$+ (0.046)(\log_2 0.046) + \cdots$$
$$= 1.829 \text{ bits per individual}$$

From equation (10.30):

$$\hat{N}_1 = e^{\hat{H}'} \quad \text{(base } e \text{ logs)}$$
$$= 2^{\hat{H}'} \quad \text{(base 2 logs)}$$
$$= 2^{1.829}$$
$$= 3.55 \text{ species}$$

Brillouin's Index

From equation (10.31)

$$\hat{H} = \frac{1}{N} \log\left(\frac{N!}{n_1!n_2!n_3!\cdots}\right)$$
$$= \frac{1}{3724} \log_2\left(\frac{3724!}{1940!1207!171!134!97!\cdots}\right)$$
$$= 1.818 \text{ bits per individual}$$

This is virtually identical to $\hat{H}'$.

Note that $\log_2(x) = 3.321981 \log_{10}(x)$.

Program DIVERS (Appendix 10.5) does these calculations.

where $\hat{D}_{MAX}$ = Maximum possible value for Simpson's index (equation 10.25)
S = Number of species in the sample

Given this value, the maximum values of the complement (equations 10.26 and 10.27) and the reciprocal (equation 10.28) of Simpson's index can be obtained. Note that the maximum possible value of the reciprocal $1/D$ is always equal to the number of species observed in the sample.

When dealing with a finite population it may be desirable to use equation (10.27) to estimate Simpson's diversity. In this case, calculations are slightly more complex:

1. Calculate $N/S = I + J/S$ where I and J are two integers and J is less than S.
2. Then:

$$\hat{D}_{MAX} = \frac{I[2J + S(I-1)]}{N(N-1)} \tag{10.33}$$