

## 1. Introduction

**Species diversity** (sometimes called **species heterogeneity**), a characteristic unique to the community level of biological organization, is an expression of community structure. The most useful measures of species diversity incorporate consideration of both the number of species (**richness**) and the distribution of individuals among the species (**evenness**).

A community is said to have a high species diversity if many equally or nearly equally abundant species are present. On the other hand, if a community is composed of a very few species, or if only a few species are abundant, then species diversity is low. For example, if a community had 100 individuals distributed among 10 species, then the maximum possible diversity would occur if there were 10 individuals in each of the 10 species (example A in Table 5B.1). The minimum possible diversity among 100 individuals would occur if there were 91 individuals belonging to one of the species and only one individual in each of the other nine species (example C in Table 5B.1). In the latter case, the typical species in the community is relatively rare, so that Patil and Taillie (1982) refer to species diversity as average rarity of species within a community and relate diversity measures to the probability of interspecific encounters.

High species diversity indicates a highly complex community, for a greater variety of species allows for a larger array of species interactions. Thus, population interactions involving energy transfer (food webs), predation, competition, and niche apportionment are theoretically more complex and varied in a community of high species diversity. This is still the subject of considerable discussion; some ecologists have supported the concept of species diversity as a measure of community stability (the ability of community structure to be unaffected by disturbance of its components), while others have concluded that there is no simple relationship between diversity and stability. Some ecologists have also used diversity as an index of the maturity of a community on the premise that communities become more complex and more stable as they mature. However, this assumption is probably applicable only in certain ecological communities. Diversity in some groups of organisms has been correlated with latitude, climate, productivity, and geography (Schluter and Ricklefs, 1993). The concept of diversity of organisms (biodiversity) is important to the field of conservation biology (see Meffe and Carroll, 1997; Primack, 1997).

On the following pages, we shall discuss *species* diversity, assuming that all individuals in a biological collection can be identified to species. If such identification is not possible or practical (for example, in a class exer-

# 5b

## Species Diversity

cise), then other taxonomic groups may be used. (For example, we may speak of genus or family diversity.) Indeed, specific identification is not needed for most comparative studies; the individuals collected may simply be identified as taxon 1, taxon 2, and so on, as long as such nomenclature is consistent from collection to collection. If you want to compare diversity indices of different communities or subcommunities, try to obtain the same-sized sample from each. This is because all measures of diversity depend to some extent on the number of species collected, which depends in turn on sample size.

Diversity is usually considered for only certain subcommunities at a time rather than for an entire ecological community. Great differences among organism sizes make diversity measures difficult to interpret in large-scale studies. Thus, we speak of the species diversity of birds, insects, or algae, or the species diversity in the soil or on tree trunks. Section 2A.8 discusses quantifying the diversity of habitats, and Section 5B.4 refers to diversity as a measure of niche width.

Section 5A.3 presents the **relative-abundance curve** and the **lognormal curve**, which express the distribution of individuals among species. These plots may be used to show species diversity graphically. Typically, however, it is best to express quantitative measures of diversity as discussed in Section 5B.2.

A large number of measures of diversity have been proposed and many are in contemporary use. Of those mentioned in the next section, we recommend that the student concentrate on the Simpson index ( $D_s$ , in Section 5B.2.2) and the information-theoretic indices ( $H$  and  $H'$ , in Section 5B.2.3).

**Table 5B.1** Various Diversity Indices Computed for Hypothetical Situations of  $N$  Individuals Distributed among  $s$  Species, with  $n_i$  Individuals in the  $i$ th Species. Examples A, B, and C have identical values of  $N$  and  $s$ . Example D has the same  $s$  and species distribution as A, but with a larger  $N$ . Example E has the same  $N$  and evenness as example A, but a smaller  $s$ . (Logarithms used are base 10.)

Species	Abundance	Hypothetical Examples				
		A	B	C	D	E
1	$n_1$	10	29	91	100	20
2	$n_2$	10	19	1	100	20
3	$n_3$	10	14	1	100	20
4	$n_4$	10	11	1	100	20
5	$n_5$	10	9	1	100	20
6	$n_6$	10	7	1	100	20
7	$n_7$	10	5	1	100	20
8	$n_8$	10	3	1	100	20
9	$n_9$	10	2	1	100	20
10	$n_{10}$	10	1	1	100	20
$s$ , number of species		10	10	10	10	5
$N$ , number of individuals		100	100	100	1000	100
$D_a$ , Margalef diversity		4.50	4.50	4.50	3.00	2.00
$D_g$ , Gleason diversity		5.00	5.00	5.00	3.33	2.50
$D_b$ , Menhinick diversity		1.00	1.00	1.00	0.32	0.50
<b>If random sample:</b>						
$l$ , Simpson dominance		0.09	0.16	0.83	0.10	0.19
$D_s$ , Simpson diversity		0.91	0.84	0.17	0.90	0.81
$d_s$ , inverse of $l$		11.00	6.23	1.21	10.09	5.21
<b>If entire community:</b>						
$\lambda$ , Simpson dominance		0.10	0.17	0.83	0.10	0.20
$\Delta_s$ , Simpson diversity		0.90	0.83	0.17	0.90	0.80
$\delta_s$ , inverse of $\lambda$		10.00	5.92	1.21	10.00	5.00
$H$ , Brillouin diversity		0.92	0.79	0.18	0.99	0.66
$H'$ , Shannon diversity		1.00	0.86	0.22	1.00	0.70
$S'$ , equally abundant species		10.0	7.2	1.7	10.0	5.0
$D_{max}$ , maximum $D_s$		0.91	0.91	0.91	0.90	0.81
$E_D$ , evenness, using $D_s$		1.00	0.92	0.19	1.00	1.00
$d_{max}$ , maximum $d_s$		11.00	11.00	11.00	10.09	5.21
$e_d$ , evenness, using $d_s$		1.00	0.57	0.11	1.00	1.00
$\Delta_{max}$ , maximum $\Delta_s$		0.90	0.90	0.90	0.90	0.80
$E_{\Delta}$ , evenness, using $\Delta_s$		1.00	0.92	0.19	1.00	1.00
$\delta_{max}$ , maximum $\delta_s$		10.00	10.00	10.00	10.00	5.00
$e_{\delta}$ , evenness, using $\delta_s$		1.00	0.59	0.12	1.00	1.00
$H_{max}$ , maximum $H$		0.92	0.92	0.92	0.99	0.66
$J$ , evenness, using $H$		1.00	0.86	0.20	1.00	1.00
$1 - J$ , dominance, using $H$		0.0	0.14	0.80	0.0	0.0
$H'_{max}$ , maximum $H'$		1.00	1.00	1.00	1.00	0.70
$J'$ , evenness, using $H'$		1.00	0.86	0.22	1.00	1.00
$1 - J'$ , dominance, using $H'$		0.0	0.14	0.78	0.0	0.0

## 2. Measures of Species Diversity

**2.1 Numbers of Species and Individuals** The simplest measure of species diversity is the number of species ( $s$ ), or the species **richness**. Several indices of diversity have been proposed that incorporate both  $s$  and  $N$ , the total number of individuals in all the species; for example, Margalef's index:

$$D_a = \frac{s - 1}{\log N} \tag{1}$$

(Margalef, 1957), which is very similar to the index of Gleason (1922):

$$D_g = \frac{s}{\log N} \tag{2}$$

and Menhinick's index:

$$D_b = \frac{s}{\sqrt{N}} \tag{3}$$

(Menhinick, 1964). But measures such as  $s$ ,  $D_a$ ,  $D_g$ , and  $D_b$  are inadequate because they do not allow us to differentiate between the diversities of different communities having the same  $s$  and  $N$ . (For instance, examples A, B, and C in Table 5B.1 are declared equally diverse by such indices.) And species richness is directly related to sample size, with larger samples likely to contain more species. A good measure of diversity should take into account both the number of species and the evenness of occurrence of individuals in the various species.

**2.2 Simpson's Index** Simpson<sup>1</sup> (1949) considered not only the number of species ( $s$ ) and the total number of individuals ( $N$ ), but also the proportion of the total that occurs in each species. He showed that if two individuals are taken at random from a community, the probability that the two will belong to the same species is

$$l = \frac{\sum n_i(n_i - 1)}{N(N - 1)} \tag{4}$$

<sup>1</sup> Simpson's approach was to apply to ecology a diversity measure introduced in an econometric context in 1912 by Gini (Bhargava and Uppuluri, 1975; Rao, 1982), so that one sees occasional references to the Gini index or Gini-Simpson index.

**Table 5B.2** A Hypothetical Set of Species Abundance Data, Used in the Text to Illustrate the Calculation of Various Diversity Indices.

Species, <i>i</i>	Abundance, <i>n<sub>i</sub></i>	Relative Abundance, <i>p<sub>i</sub></i>
1	50	50/85 = 0.588
2	25	25/85 = 0.294
3	10	10/85 = 0.118
<i>s</i> = 3	<i>N</i> = 85	

The quantity *l* is, therefore, a measure of **dominance**<sup>2</sup> (the concentration of *N* individuals among *s* species). A collection of species with high diversity will have low dominance, and

$$D_s = 1 - l, \quad (5)$$

namely

$$D_s = 1 - \frac{\sum n_i(n_i - 1)}{N(N - 1)}, \quad (6)$$

is a good measure of diversity<sup>3</sup> (expressing the probability of two randomly selected individuals belonging to different species). For the data of Table 5B.2,

$$\begin{aligned} D_s &= 1 - \frac{50(49) + 25(24) + 10(9)}{85(84)} \\ &= 1 - 3140/7140 \\ &= 1 - 0.44 \\ &= 0.56 \end{aligned}$$

Some ecologists have inverted Simpson's dominance index to arrive at a measure of diversity:

$$d_s = \frac{1}{l} = \frac{N(N - 1)}{\sum n_i(n_i - 1)} \quad (7)$$

<sup>2</sup>The quantity  $\sum n_i(n_i - 1)$  in the numerator of *l* may be computed as  $\sum n_i^2 - N$ , which may prove simpler on some calculators. Morisita's measure of dispersion (Section 4C.3.4) is computationally related to *l*. That is, a large *l* implies an aggregation of individuals in only a few species, whereas a small value of *l* denotes a more uniform distribution of individuals among species.

<sup>3</sup>Hurlburt (1971) computes *D<sub>s</sub>* as  $\sum (n_i/N)[(N - n_i)/(N - 1)]$ ; but Equation 6 is simpler. McIntosh (1967) has proposed  $\sqrt{\sum n_i^2}$  as a dominance measure from which diversity indices may be derived.

This diversity index is an expression of the number of times one would have to take pairs of individuals at random from the entire aggregation to find a pair from the same species. It is also an expression of how many equally abundant species would have a diversity equal to that in the observed collection. The index *d<sub>s</sub>* is preferable to *D<sub>s</sub>* in comparing collections in which the values of *D<sub>s</sub>* are very close to 1.0 and nearly the same. (For example, two collections yielding values of *D<sub>s</sub>* of 0.96 and 0.98 would give us *d<sub>s</sub>* values—with more discrimination—of 25.00 and 50.00, respectively.) Levins (1968) proposed the inverse of Simpson's diversity index as a measure of niche breadth, but others have since presented measures that are preferable for that purpose (see, for example, Feinsinger et al., 1981).

Hurlburt (1971) severely criticized most diversity indices (including those that follow), but praised the characteristics of the above indices as being biologically meaningful, with *D<sub>s</sub>* referring to the **probability of interspecific encounter** (which he calls PIE). This is the probability of an individual in the community encountering a member of another species. He relates this concept to specific kinds of encounters, such as competition and predation.

The above considerations of *l*, *D<sub>s</sub>*, and *d<sub>s</sub>* assume that the data at hand are a random sample from a community or subcommunity. There are occasions when this is not the case, as when we have data from an entire community or subcommunity (e.g., a laboratory culture of animals or scavengers at an animal carcass) rather than from a sample, or when we do have a sample but it is known to be a nonrandom representation of a community or subcommunity. In such cases, the appropriate Simpson measure of dominance is

$$\lambda = \frac{\sum n_i^2}{N^2} \quad (8)$$

or, equivalently,

$$\lambda = \sum p_i^2, \quad (9)$$

where

$$p_i = \frac{n_i}{N}, \quad (10)$$

that is, *p<sub>i</sub>* is the proportion of the total number of individuals occurring in species *i*. And the diversity indices analogous to *D<sub>s</sub>* and *d<sub>s</sub>* are

$$\Delta_s = 1 - \lambda = 1 - \frac{\sum n_i^2}{N^2} \quad (11)$$

and

$$\delta_s = \frac{1}{\lambda} = \frac{N^2}{\sum n_i^2}, \quad (12)$$

respectively, which may also be written as

$$\Delta_s = 1 - \sum p_i^2 \quad (13)$$

and

$$\delta_s = \frac{1}{\sum p_i^2}. \quad (14)$$

**2.3 Information-Theoretic Indices** Measures of species diversity based on information theory (introduced to ecologists by MacArthur, 1955, and Margalef, 1958) are related to the concept of **uncertainty**. In a species aggregation of low diversity (e.g., example C in Table 5B.1), we can be relatively certain of the identity of a species chosen at random. (In this example, it will probably be a member of species 1.) In a highly diverse community, however (e.g., example A in Table 5B.1), it is difficult to predict the identity of a randomly picked individual. Thus, high diversity is associated with high uncertainty and low diversity with low uncertainty. And some authors have equated uncertainty with entropy.

Information-theoretic measures also allow us to consider and calculate measures of hierarchical diversity. Consider example A in Table 5B.1. If the 10 species were each of a different genus, this would intuitively imply a greater diversity than if they were all of the same genus. To learn more about measuring hierarchical diversity, taking into account the distribution of species within genera, genera within families, and so on, consult Pielou (1975, 1977).

Two kinds of ecological collections must be considered. The first is where our species-abundance data compose a sample taken at random from a community or subcommunity. In the second kind of collection (e.g., with a rotting log subcommunity or some laboratory situations), we know the total number of individuals in a collection of species without resorting to samples. Or we may have obtained a nonrandom sample (thus nonrepresentative of its community or subcommunity); in these cases, the sample must also be considered a complete enumeration. For example, trap (Sections 3D.5 and 3E.3.1), artificial substrate (Section 3E.2.5), and seine (Section 3E.3.4) sampling typically favor the collection of certain species. Thus, these collections do not exhibit species compositions and abundances that accurately reflect those of the sampled community.

If our data are a random sample of species abundances from a larger community or subcommunity of

**Table 5B.3** Factors to Convert between Logarithmic Bases 2, *e*, and 10. (For example, a value of 0.86 computed using base 10 is equivalent to a value of  $(0.86)(2.3026) = 1.98$  using base *e*.)

To Convert to	To Convert from		
	2	<i>e</i>	10
2	1.0000	1.4427	3.3219
<i>e</i>	0.6931	1.0000	2.3026
10	0.3010	0.4343	1.0000

interest, then we may appropriately use the Shannon<sup>4</sup> diversity index (Shannon [1948]):

$$H' = -\sum p_i \log p_i \quad (15)$$

where  $p_i$  is as in Equation 10: the proportion of the total number of individuals that belong to species  $i$ . For this calculation, one may use any logarithmic base; bases 10 and *e* are the commonest, although communications engineers (from whom the index has been borrowed) use base 2. The selection of a particular logarithmic base is immaterial as long as it is consistent;  $H'$  computed in one base may be converted to  $H'$  for another base by consulting Table 5B.3. In Appendix D, Table D.1 gives logarithms and Table D.2 gives logarithms of proportions. For the data in Table 5B.2,

$$\begin{aligned} H' &= -[0.588 \log 0.588 + 0.294 \log 0.294 \\ &\quad + 0.118 \log 0.118] \\ &= -[0.588(-0.231) + 0.294(-0.532) \\ &\quad + 0.118(-0.928)] \\ &= -[-0.136 - 0.156 - 0.110] \\ &= 0.40. \end{aligned}$$

A little algebraic manipulation arrives at an equivalent equation:

$$H' = (N \log N - \sum [n_i \log n_i])/N. \quad (16)$$

This equation allows us to compute  $H'$  without first converting abundances ( $n_i$ ) to proportions ( $p_i$ ), both saving time and avoiding rounding errors. Table 5B.4 is very

<sup>4</sup>This often less properly called the Shannon-Weaver or Shannon-Wiener index, for C. E. Shannon's equation received some inspiration from N. Wiener and some clarification from W. W. Weaver (Perkins, 1982). Shannon did not apply this measure to ecological systems.

Table 5B.4 Values of  $n_i \log n_i$  (or  $N \log N$ ) for Use in Equation 16.\*

$n_i$	0	1	2	3	4	5	6	7	8	9	$n_i$
0		0.000	0.602	1.431	2.408	3.495	4.669	5.916	7.225	8.588	0
10	10.000	11.455	12.950	14.481	16.046	17.641	19.266	20.918	22.595	24.296	10
20	26.021	27.767	29.533	31.320	33.125	34.949	36.789	38.647	40.520	42.410	20
30	44.314	46.232	48.165	50.111	52.070	54.042	56.027	58.023	60.032	62.052	30
40	64.082	66.124	68.176	70.239	72.312	74.395	76.487	78.589	80.700	82.820	40
50	84.949	87.086	89.232	91.387	93.549	95.720	97.899	100.085	102.279	104.480	50
60	106.689	108.905	111.128	113.358	115.596	117.839	120.090	122.347	124.611	126.881	60
70	129.157	131.439	133.728	136.023	138.323	140.630	142.942	145.260	147.583	149.913	70
80	152.247	154.587	156.933	159.283	161.639	164.001	166.367	168.738	171.114	173.496	80
90	175.882	178.273	180.668	183.069	185.474	187.884	190.298	192.717	195.140	197.568	90
100	200.000	202.436	204.877	207.322	209.771	212.225	214.682	217.144	219.610	222.079	100
110	224.553	227.031	229.512	231.998	234.487	236.980	239.477	241.978	244.482	246.990	110
120	249.502	252.017	254.536	257.058	259.584	262.114	264.647	267.183	269.723	272.266	120
130	274.813	277.363	279.916	282.472	285.032	287.595	290.161	292.731	295.303	297.879	130
140	300.458	303.040	305.625	308.213	310.804	313.398	315.996	318.596	321.199	323.805	140
150	326.414	329.026	331.640	334.258	336.878	339.501	342.127	344.756	347.388	350.022	150
160	352.659	355.299	357.941	360.587	363.234	365.885	368.538	371.194	373.852	376.513	160
170	379.176	381.842	384.511	387.182	389.856	392.532	395.210	397.891	400.575	403.261	170
180	405.949	408.640	411.333	414.029	416.726	419.427	422.129	424.834	427.542	430.251	180
190	432.963	435.677	438.394	441.113	443.834	446.557	449.282	452.010	454.740	457.472	190
200	460.206	462.942	465.681	468.422	471.165	473.910	476.657	479.406	482.157	484.911	200
210	487.666	490.424	493.183	495.945	498.709	501.474	504.242	507.012	509.784	512.557	210
220	515.333	518.111	520.890	523.672	526.456	529.241	532.029	534.818	537.609	540.402	220
230	543.197	545.994	548.793	551.594	554.397	557.201	560.007	562.815	565.625	568.437	230
240	571.251	574.066	576.883	579.702	582.523	585.346	588.170	590.996	593.824	596.654	240
250	599.485	602.318	605.153	607.989	610.828	613.668	616.509	619.353	622.198	625.045	250
260	627.893	630.743	633.595	636.448	639.303	642.160	645.019	647.879	650.740	653.603	260
270	656.468	659.335	662.203	665.072	667.944	670.816	673.691	676.567	679.444	682.324	270
280	685.204	688.086	690.970	693.856	696.742	699.631	702.521	705.412	708.305	711.199	280
290	714.095	716.993	719.892	722.792	725.694	728.597	731.502	734.409	737.316	740.226	290
300	743.136	746.049	748.962	751.877	754.794	757.711	760.631	763.551	766.474	769.397	300
310	772.322	775.248	778.176	781.105	784.036	786.968	789.901	792.836	795.772	798.709	310
320	801.648	804.588	807.530	810.472	813.417	816.362	819.309	822.257	825.207	828.157	320
330	831.110	834.063	837.018	839.974	842.931	845.890	848.850	851.811	854.774	857.738	330
340	860.703	863.669	866.637	869.606	872.576	875.548	878.520	881.494	884.470	887.446	340
350	890.424	893.403	896.383	899.364	902.347	905.331	908.316	911.303	914.290	917.279	350
360	920.269	923.260	926.253	929.246	932.241	935.237	938.234	941.232	944.232	947.233	360
370	950.235	953.238	956.242	959.247	962.254	965.262	968.271	971.281	974.292	977.304	370
380	980.318	983.332	986.348	989.365	992.383	995.402	998.423	1001.444	1004.467	1007.490	380
390	1010.515	1013.541	1016.568	1019.596	1022.626	1025.656	1028.687	1031.720	1034.753	1037.788	390
400	1040.824	1043.861	1046.899	1049.938	1052.978	1056.019	1059.062	1062.105	1065.149	1068.195	400
410	1071.241	1074.289	1077.338	1080.387	1083.438	1086.490	1089.543	1092.597	1095.652	1098.708	410
420	1101.765	1104.823	1107.882	1110.942	1114.003	1117.065	1120.128	1123.193	1126.258	1129.324	420
430	1132.391	1135.460	1138.529	1141.599	1144.671	1147.743	1150.816	1153.890	1156.966	1160.042	430
440	1163.119	1166.197	1169.277	1172.357	1175.438	1178.520	1181.603	1184.687	1187.773	1190.859	440
450	1193.946	1197.034	1200.123	1203.212	1206.303	1209.395	1212.488	1215.582	1218.676	1221.772	450
460	1224.869	1227.966	1231.065	1234.164	1237.264	1240.366	1243.468	1246.571	1249.675	1252.780	460
470	1255.886	1258.993	1262.101	1265.209	1268.319	1271.429	1274.541	1277.653	1280.767	1283.881	470
480	1286.996	1290.112	1293.229	1296.346	1299.465	1302.585	1305.705	1308.827	1311.949	1315.072	480
490	1318.196	1321.321	1324.447	1327.574	1330.701	1333.830	1336.959	1340.089	1343.220	1346.352	490

\*If values for  $n_i$  (or  $N$ ) larger than 499 are needed, consult Lloyd et al. (1968) or Zar (1974:401-404) or use Appendix D, Table D.1.

conveniently used with this equation. For the above data from Table 5B.2,

$$\begin{aligned}
 H' &= [85 \log 85 - (50 \log 50 + 25 \log 25 + 10 \log 10)]/85 \\
 &= [164.001 - (84.949 + 34.949 + 10.000)]/85 \\
 &= 34.103/85 \\
 &= 0.40.
 \end{aligned}$$

As noted above, the Shannon diversity index,  $H'$ , is appropriate when you have a random sample of species abundances from a larger aggregation, say a random sample of an entire community. Such a sample (unless extremely large) will probably not contain representatives of each species in the entire community. So, typically, our observed value of  $s$  is biased, an underestimate of the number of species in the entire community. However, the lack of data on rare species has little effect on the value of  $H'$  (although it has serious effect on  $H'_{max}$  and  $J'$ , discussed in Section 5B.2.4).

$H'$  may also be calculated for data other than abundances, for example, to express habitat heterogeneity (Section 2A.8) or the diversity of biomass (Section 6A), or of coverage (Sections 3A, 3B, and 3C). Also note that Equations 8, 9, 13, 14, and 15 may be used with relative measures (e.g., relative abundance or relative biomass).

Another way of depicting species diversity is to express the number of equally abundant species that would produce the value of  $H'$  of the observed sample. This measure may be represented as

$$S' = B^{H'}, \quad (17)$$

where  $B$  is the logarithmic base used in computing  $H'$  (e.g., 10,  $e$ , or 2). For example B in Table 5B.1,

$$S' = 10^{0.86} = 7.2.$$

Now let us consider a set of species-abundance data that is considered a nonrandom sample. For such a set, or for collected data that are an entire community or subcommunity, do not use  $H'$  (Pielou, 1966a, 1966b, 1967, 1975); instead use the Brillouin (1962) index:

$$H = \left( \log \frac{N!}{\prod n_i!} \right) / N, \quad (18)$$

where  $\Pi$  (capital Greek pi) means to take the product, just as  $\Sigma$  means to take the sum; thus, we can write Equation 18 as

$$H = \left( \log \frac{N!}{n_1! n_2! \dots n_s!} \right) / N. \quad (19)$$

The computation of factorials—such as  $6! = (6)(5)(4)(3)(2) = 720$ —is tedious, and the numbers typically become unwieldy. Therefore,  $H$  is much more conveniently calculated using logarithms:

$$H = (\log N! - \Sigma[\log n_i!])/N, \quad (20)$$

with the aid of Table 5B.5. For our Table 5B.2 example,

$$\begin{aligned}
 H &= [\log 85! - (\log 50! + \log 25! + \log 10!)]/85 \\
 &= [128.450 - (64.483 + 25.191 + 6.560)]/85 \\
 &= 32.216/85 \\
 &= 0.38.
 \end{aligned}$$

As with  $H'$ , the logarithmic base used is immaterial as long as it is consistent.  $H$  values in one base may be converted to those in another by using Table 5B.3.

The units of  $H$  and  $H'$  are unimportant (and probably meaningless) to the ecologist. These indices are used only in a relative fashion, that is, to determine which species assemblages are more or less diverse than others. Washington (1984) summarizes many uses and interpretations of  $H$  and  $H'$ .

**2.4 Evenness** The diversity indices in Sections 5B.2.2 and 5B.2.3 take into account both the species **richness** (the number of species) and the **evenness** of the individuals' distribution among the species. Separate measures of these two components of diversity are often desirable. Richness can be expressed simply as the number of species. Evenness may be expressed by considering how close a set of observed species abundances are to those from an aggregation of species having maximum possible diversity for a given  $N$  and  $s$ .

The maximum possible diversity for a collection of  $N$  individuals in a total of  $s$  species exists when the  $N$  individuals are distributed as evenly as possible among the  $s$  species, that is, when each  $n_i = N/s$ . The maximum possible values of  $D_s$ ,  $d_s$ ,  $\Delta_s$ ,  $H$ , and  $H'$  are as follows:

$$D_{max} = \left( \frac{s-1}{s} \right) \left( \frac{N}{N-1} \right), \quad (21)$$

$$d_{max} = s \left( \frac{N-1}{N-s} \right), \quad (22)$$

$$\Delta_{max} = 1 - \frac{1}{s}, \quad (23)$$

$$\delta_{max} = s, \quad (24)$$

$$H_{max} = [\log N! - (s-r) \log c! - r \log (c+1)!]/N, \quad (25)$$

Table 5B.5 Values of  $\log n_i!$  (or  $\log N!$ ) for Use in Equation 20.\*

$n_i$	0	1	2	3	4	5	6	7	8	9	$n_i$
0	0.000	0.000	0.301	0.778	1.380	2.079	2.857	3.702	4.606	5.560	0
10	6.560	7.601	8.680	9.794	10.940	12.116	13.321	14.551	15.806	17.085	10
20	18.386	19.708	21.051	22.412	23.793	25.191	26.606	28.037	29.484	30.947	20
30	32.424	33.915	35.420	36.939	38.470	40.014	41.571	43.139	44.719	46.310	30
40	47.912	49.524	51.148	52.781	54.425	56.078	57.741	59.413	61.094	62.784	40
50	64.483	66.191	67.907	69.631	71.363	73.104	74.852	76.608	78.371	80.142	50
60	81.920	83.706	85.498	87.297	89.103	90.916	92.736	94.562	96.394	98.233	60
70	100.078	101.930	103.787	105.650	107.520	109.395	111.275	113.162	115.054	116.952	70
80	118.855	120.763	122.677	124.596	126.520	128.450	130.384	132.324	134.268	136.218	80
90	138.172	140.131	142.095	144.063	146.036	148.014	149.996	151.983	153.974	155.970	90
100	157.970	159.974	161.983	163.996	166.013	168.034	170.059	172.089	174.122	176.160	100
110	178.201	180.246	182.295	184.349	186.405	188.466	190.531	192.599	194.671	196.746	110
120	198.825	200.908	202.995	205.084	207.178	209.275	211.375	213.479	215.586	217.697	120
130	219.811	221.928	224.049	226.172	228.299	230.430	232.563	234.700	236.840	238.983	130
140	241.129	243.278	245.431	247.586	249.744	251.906	254.070	256.237	258.408	260.581	140
150	262.757	264.936	267.118	269.302	271.490	273.680	275.873	278.069	280.268	282.469	150
160	284.673	286.880	289.090	291.302	293.517	295.734	297.954	300.177	302.402	304.630	160
170	306.861	309.094	311.329	313.567	315.808	318.051	320.296	322.544	324.795	327.048	170
180	329.303	331.561	333.821	336.083	338.348	340.615	342.885	345.157	347.431	349.707	180
190	351.986	354.267	356.550	358.836	361.124	363.414	365.706	368.000	370.297	372.596	190
200	374.897	377.200	379.505	381.813	384.123	386.434	388.748	391.064	393.382	395.702	200
210	398.025	400.349	402.675	405.004	407.334	409.666	412.001	414.337	416.676	419.016	210
220	421.359	423.703	426.049	428.398	430.748	433.100	435.454	437.810	440.168	442.528	220
230	444.890	447.253	449.619	451.986	454.355	456.727	459.099	461.474	463.851	466.229	230
240	468.609	470.991	473.375	475.761	478.148	480.537	482.928	485.321	487.715	490.112	240
250	492.510	494.909	497.311	499.714	502.119	504.525	506.933	509.343	511.755	514.168	250
260	516.583	519.000	521.418	523.838	526.260	528.683	531.108	533.534	535.962	538.392	260
270	540.824	543.257	545.691	548.127	550.565	553.004	555.445	557.888	560.332	562.777	270
280	565.225	567.673	570.124	572.575	575.029	577.483	579.940	582.398	584.857	587.318	280
290	589.780	592.244	594.710	597.177	599.645	602.115	604.586	607.059	609.533	612.009	290
300	614.486	616.964	619.444	621.926	624.409	626.893	629.379	631.866	634.354	636.844	300
310	639.336	641.828	644.323	646.818	649.315	651.813	654.313	656.814	659.317	661.820	310
320	664.326	666.832	669.340	671.849	674.360	676.872	679.385	681.899	684.415	686.932	320
330	689.451	691.971	694.492	697.014	699.538	702.063	704.589	707.117	709.646	712.176	330
340	714.708	717.240	719.774	722.310	724.846	727.384	729.923	732.464	735.005	737.548	340
350	740.092	742.637	745.184	747.732	750.281	752.831	755.382	757.935	760.489	763.044	350
360	765.600	768.158	770.716	773.276	775.837	778.400	780.963	783.528	786.094	788.661	360
370	791.229	793.798	796.369	798.941	801.513	804.087	806.663	809.239	811.817	814.395	370
380	816.975	819.556	822.138	824.721	827.305	829.891	832.478	835.065	837.654	840.244	380
390	842.835	845.427	848.021	850.615	853.210	855.807	858.405	861.003	863.603	866.204	390
400	868.806	871.410	874.014	876.619	879.225	881.833	884.441	887.051	889.662	892.273	400
410	894.886	897.500	900.115	902.731	905.348	907.966	910.585	913.205	915.826	918.449	410
420	921.072	923.696	926.321	928.948	931.575	934.204	936.833	939.463	942.095	944.727	420
430	947.361	949.995	952.631	955.267	957.905	960.543	963.183	965.823	968.465	971.107	430
440	973.751	976.395	979.040	981.687	984.334	986.983	989.632	992.282	994.933	997.586	440
450	1000.239	1002.893	1005.548	1008.204	1010.861	1013.519	1016.178	1018.838	1021.499	1024.161	450
460	1026.824	1029.487	1032.152	1034.818	1037.484	1040.152	1042.820	1045.489	1048.160	1050.831	460
470	1053.503	1056.176	1058.850	1061.525	1064.200	1066.877	1069.555	1072.233	1074.913	1077.593	470
480	1080.274	1082.956	1085.639	1088.323	1091.008	1093.694	1096.381	1099.068	1101.757	1104.446	480
490	1107.136	1109.827	1112.519	1115.212	1117.906	1120.600	1123.296	1125.992	1128.689	1131.387	490

\*If values for  $n_i$  (or  $N$ ) larger than 499 are needed, consult Lloyd et al. (1968) or Pearson and Hartley (1966: Table 51). Alternatively, use Appendix D, Table D.1 to compute  $\log n_i!$  (or  $\log N!$ ) by Stirling's approximation:  $\log n_i! = (n_i + 0.5) \log n_i - 0.4343 n_i + 0.3991$ .

where  $c$  and  $r$  are defined below, and

$$H_{max}' = \log s. \quad (26)$$

In Equation 25,  $c$  is the integer portion of  $N/s$  and  $r$  is the remainder. For the Table 5B.2 data, for example,

$$\begin{aligned} N/s &= 85/3 \\ &= 28.33; \end{aligned}$$

that is, the quotient of  $85/3$  is 28 with a remainder of 1. Therefore,  $c = 28$  and  $r = 1$ ; so

$$\begin{aligned} H_{max} &= [\log 85! - (3 - 1) \log 28! - (1) \log 29!]/85 \\ &= [128.450 - 2(29.484) - 1(30.947)]/85 \\ &= 38.535/85 \\ &= 0.45. \end{aligned}$$

The **evenness** of the distribution of  $N$  individuals among the  $s$  species in a set of data is then expressed as the nearness of the diversity index for the observed data to the index of maximum diversity. We use the following expressions for evenness:

$$E_D = D_s/D_{max}, \quad (27)$$

$$e_d = d_s/d_{max}, \quad (28)$$

$$E_\Delta = \Delta_s/\Delta_{max}, \quad (29)$$

$$e_\delta = \delta_s/\delta_{max}, \quad (30)$$

$$J = H/H_{max}, \quad (31)$$

and

$$J' = H'/H_{max}'. \quad (32)$$

Evenness measures are sometimes called measures of **relative diversity**. See Table 5B.1 for some computed examples.

Remember that when you examine only a sample from a community or subcommunity, you typically will underestimate the number of species in the entire community or subcommunity (unless the sample is extremely large). Therefore, expressions of maximum diversity will be underestimates, and the computed evenness measures will be overestimates of the evenness in the actual collection sampled. In addition, we should be aware that these measures of evenness are not independent of the number of species (i.e., species richness) (DeBenedictis, 1973;

Alatalo, 1981), so comparisons of evenness measures are most trustworthy when  $s$  is the same, or at least similar, in those collections being compared.

The quantity  $1 - J$  or  $1 - J'$  (or  $1 - E_D$  or  $1 - e_d$ ) may be used as an expression of dominance; it will have a low value (zero being the minimum) when dominance is low and a high value (with a maximum of 1.0) when dominance is high (see Table 5B.1). The inverse of  $J$  or  $J'$  would also be a measure of dominance, but  $1 - J$  or  $1 - J'$  possesses the appeal of having values between 0 and 1. As  $J$  and  $J'$  are ratios of two quantities with the same units, they are unitless and unaffected by the choice of logarithmic base (as long as both numerator and denominator have the same base). That is,  $J$  (or  $J'$ ) will be the same regardless of the base in which  $H$  and  $H_{max}$  (or  $H'$  and  $H_{max}'$ ) were calculated.

Sheldon (1969) proposed  $B^{H'}/s$  as an evenness index, but Heip (1974) concluded  $(B^{H'} - 1)/(s - 1)$  to be much better;  $B$  is the logarithmic base used in computing  $H'$ . The quantities  $d_s/B^{H'}$  and  $(d_s - 1)/(B^{H'} - 1)$  have also been proposed as evenness measures (Alatalo, 1981). Smith and Wilson (1996) discuss several others. Redundancy indices (e.g., see Zar, 1968; Hurlbert, 1971) are another approach to the expression of evenness.

### 3. Correlation Among Diversity Indices

It has been observed that two major groups of diversity indices are those that are most affected by the occurrence of rare species (i.e., that are based heavily on species richness) and those that are most sensitive to the relative abundances of the species (i.e., that depend heavily on the dominance versus evenness of the species abundances) (Magurran, 1988:73-75; Peet, 1974). In the first group are  $s$ ,  $D_a$ ,  $D_b$ ,  $H_s$  and  $H'$ ; the second group includes  $l$ ,  $D_s$ ,  $d_s$ ,  $\lambda$ ,  $\Delta_s$ ,  $\delta_s$ ,  $E_D$ ,  $e_d$ ,  $E_\Delta$ ,  $e_\delta$ ,  $J$ , and  $J'$ . Measures in the first group tend to discriminate better between communities that have different diversities, but they also tend to be more seriously affected by sample size (Magurran, 1988:80). Although the literature on using species diversity indices has increased impressively, there is no agreement as to which index is best.

### 4. Measuring Niche Width

By width, or breadth, of a niche an ecologist means the diversity of resources used by a species. Typically, the context is the resource utilization of a population in a particular habitat. Aspects of a niche may be measured by diversity indices. For example, the diversity of food items used measures a component of an animal's niche width. This can be quantified by determining the propor-

tion ( $p_i$ ) of each food item in the animals's diet and then calculating the Shannon diversity index ( $H'$ ). Similarly, the diversity of the habitat's structure used by the species is a measure of its niche width. So it might be determined what proportion of a population of plants is in each of several habitat types, or what proportion of time a population of animals is in each of several habitat types; and  $H'$  could then be calculated.

## 5. Comparing Diversity Measurements

Oftentimes we have two samples of data and a diversity index is calculated for each of them. It may then be desired to compare objectively the two indices and ask whether the two sampled taxonomic aggregations are equally diverse. (See Section 1B.3.1 for basic concepts of two-sample testing.)

**5.1 Comparing Simpson Indices** If  $(D_s)_1$  is the Simpson diversity index for one sample and  $(D_s)_2$  is the index for a second sample, then we can test the null hypothesis that these two samples come from aggregations having the same diversity (Keefe and Bergerson, 1977), but first we must calculate  $s^2$ , the variance of  $D_s$ , for each sample, which is, approximately<sup>5</sup>,

$$s^2 = 4[\sum p_i^3 - (\sum p_i^2)^2]/N \quad (34)$$

(where  $p_i$  is as in Equation 10). Then, we compute the following statistic:

$$t = \frac{(D_s)_1 - (D_s)_2}{\sqrt{s_1^2 + s_2^2}} \quad (35)$$

and compare it to the critical value of Student's  $t$  (Table 1B.1) for infinity degrees of freedom ( $DF = \infty$ ).

If we are dealing with data comprising an entire community or subcommunity, then  $(\Delta_s)_1$  and  $(\Delta_s)_2$  may be compared by inspection instead of statistically. If we are using  $\Delta_s$ , because the samples are nonrandom, then several such samples may be collected from each of two communities and the diversities of the two compared by the statistical procedures of Section 1B.3.1, 1B.3.2, or 1B.3.4.

<sup>5</sup>This approximation is good for large  $N$  values; for small  $N$  values, one can use

$$s^2 = \frac{4N(N-1)(N-2)\sum p_i^3 + 2N(N-1)\sum p_i^2 - 2N(N-1)(2N-3)(\sum p_i^2)^2}{[N(N-1)]^2} \quad (33)$$

(Simpson, 1949).

**5.2 Comparing Information-Theoretic Indices** To compare Shannon diversity indices from two collections of data ( $H'_1$  and  $H'_2$ ), we need the variance of  $H'$  for each collection<sup>6</sup>.

$$s^2 = \frac{\sum n_i \log^2 n_i - (\sum n_i \log n_i)^2/N}{N^2} \quad (36)$$

(Basharin, 1959), where  $\log^2 n_i$  is the mathematical notation for  $(\log n_i)^2$ . Then we compute

$$t = \frac{H'_1 - H'_2}{\sqrt{s_1^2 + s_2^2}} \quad (37)$$

and compare it to the critical value of Student's  $t$  (Table 1B.1) for degrees of freedom as follows (rounded down if not a whole number):

$$DF = \frac{(s^2_{H'_1} + s^2_{H'_2})^2}{\frac{(s^2_{H'_1})^2}{n_1} + \frac{(s^2_{H'_2})^2}{n_2}} \quad (38)$$

(Hutcheson, 1970). See Zar (1984:146-148) for examples of this procedure.

If, instead of one  $H'$  for each of the two communities to be compared, we had an  $H'$  for each of several replicate random samples from each community, then we could employ the  $t$  test as in Section 1B.3.1 (i.e., using Equations 10 through 12 in that section). This is because replicated estimates of Shannon diversity from the same community tend to be normally distributed (see Section 1B.3.3). Indeed, if we had replicate  $H'$  values from three or more communities, hypothesis testing could be done via the procedures of analysis of variance and multiple comparisons alluded to in Section 1B.3.2. Alternatively, we could analyze replicate values of  $H'$  by non-parametric testing (see Section 1B.3.4).

If we were using Brillouin's index ( $H$ ) for data that compose an entire community, a statistical procedure such as that above would not be necessary to compare two indices; the comparison would simply be made by inspection. If Brillouin's  $H$  is used because you have a nonrandom sample, then several such samples may be collected from each of two communities and the community diversities may be compared using the procedures of Section 1B.3.1, 1B.3.2, or 1B.3.4.

<sup>6</sup>This approximation is good for large  $N$  values; for small  $N$  values, subtract  $(s-1)/N$  from Equation 36.