COENOSES 9(2):59-67, 1994
© C.E.T.A., Gorizia, Italy

NUMERICAL PROPERTIES OF JACKKNIFED DIVERSITY INDICES TESTED ON LOOSE SETS OF COENOLOGICAL SAMPLES (DIPTERA, DROSOPHILIDAE)[1]

J. Izsák[2] & L. Papp[3]

2 Department of Zoology, Berzsenyi College, Szombathely, Károlyi G. tér 4, H-9701 Hungary
3 Department of Zoology, Hungarian Natural History Museum, Budapest, Baross u. 13, H-1088 Hungary

Keywords: Confidence intervals, Diversity indices, Drosophilid assemblages, Index sensitivity, Jackknife estimation, Significance tests.

Abstract. In a former paper by one of us (L. Papp) diversity measures of frequency lists of Drosophilidae assemblages in four Hungarian low mountain valleys were compared. The present paper elaborates on the diversity measures themselves with the intentions to find how closely conclusions based on the Shannon-Wiener index can be corroborated by other indices. The Shannon-Wiener index, Fisher's index, as well as some indices of Hurlbert's and Hill's index families were calculated. Confidence intervals of the jackknifed diversity values as well as the significance of diversity differences were also determined. Through the high number of original and jackknifed diversity estimates, an opportunity presented itself to study the statistical properties of the original estimates. The bias of estimation and the radii of confidence intervals of jackknife estimates as functions of sample size, were analysed. Strong correlations were found between the diversity indices. A relationship between frequency size and index response was also demonstrated. It is illuminating that when two or three elements of a species frequency vector are changed by an appropriate algorithm for those in a fixed vector, the modified diversity value approximates the diversity belonging to the fixed vector very closely.

Introduction

There are widely used analytical methods for diversity investigations in ecology. For a review of diversity indices see Washington (1984). Diversity or a diversity index can be regarded as a functional or structural parameter of the community. Because of this, diversity analysis has become an important methodological tool in ecology on both community and species levels. The focus of the present paper is on the methodical aspects in the study of alpha diversity (Whittaker 1972). Several methodical innovations described earlier by one of us (J. Izsák) are tested here in the context of a faunistical-coenological study. The earlier observations (Papp 1992) were based on the Shannon-Wiener index. We now pose the question: how closely can we corroborate past findings by using other indices and by *jackknife* estimates and confidence intervals that makes it possible to test the significance of diversity differences? The high number (cca. 400) of original and jackknife estimates makes it possible to perform numerical experiments.

In most coenological and ecological analyses the diversity investigation is reduced to a mere calculation of a *single* diversity index. This is mostly the Shannon-Wiener index or the Simpson index. These calculations do not give a satisfactory and reliable picture of diversity relations. It is also important that some species-population assemblages do not form a separable and functionally connected system, other than at most a guild (cf. Papp 1992). In such cases "true" diversity as an inherent functional parameter does not exist. Yet even in these cases, diversity indices may express structural and functional properties of the community. Therefore, it is generally worthwhile to amplify the methodological arsenal of diversity investigations.

Materials

In the cited paper, studies are reported on apple-bait collections of Drosophilidae in four sites. The collection from 1988-1990 yielded more than 9000 specimens. Individuals of 40 species were captured (31, 31 and 36 spp. per year). Some of these were new to or found rarely in Hungary.

As a usual phenomenon, the population frequencies in the assemblages changed profoundly from one year to the next. The population size ratio of the rare and the dominant species is approximately 1 to 10^3or to 10^4 or even higher. The Shannon- Wiener diversity index values are not very high. Evenness varies from medium to low values. The differences in the assemblages of the collecting sites were analysed by simple parameters, including the Berger-Parker index, Czekanowski and Renkonen similarity indices and

---

[1] Supported by the National Scientific Research Fund (OTKA No. 5266 for J. I. and OTKA No. 3188 for L. P.)

dendrograms based on the latters. It was found that a good many of the drosophilid species is so rare that only a small portion of the extant rare species is detectable at all in a given area. However, the ratios of dominant-subdominant, constant-subconstant species populations (at least theoretically, and so they were hypothesized) indicate more non-detected species than do population frequencies deformed by human activity.

Baiting with fermenting fruits is one of the classical methods in drosophilid studies. The literature on the results is vast. For a bibliographical review see Bächli and Rocha Pite (1982). Most of the studies were in Europe, presenting extensive and intensive collectings and quantitative analyses of populations (Bächli 1979, Burla and Bächli 1991, etc.). Our four collecting sites were:

A. *Aggtelek National Park*: Ménes Valley, 150 m downstream from the Ménes source.

B. *Bükk National Park:* Garadna Valley, 200-300 m above the Hámori Lake (Miskolc).

M. *Börzsöny Landscape Protection Area:* the valley of the Keskenybükki Creek, upstream from the settlement Magyarkút (Verőce).

V. *Pilis Landscape Protection Area:* Apátkúti Valley, cca. 2 km upstream from Visegrád.

The four sites are all affected by tourism to some extent; least at the Ménes source and most at Magyarkút.

The sample codes consist of a letter for the site (A,B, M,V) and five numbers (day, month and year). For example, the code A11090 identifies the sample: Aggtelek, 11 September, 1990. Bait is made from smashed apple with some sugar added and fermented in lukewarm water for more than a week. The bait is placed on 40x40 cm square plastic sheets within a 10 m range of the creeks. The distance between collecting points was at least 10 m. To obtain a simple sample, baits were placed down at 15 to 18 sites. The size of the baits was 4 to 5 deciliters. Exposure time was usually 30 min. Drosophilids were sweep-netted for a period of three hours, in some cases two hours. In cases, when less than 100 specimens were captured, collection time was prolonged. For other details see Papp (1992).

Methods

The primary data are those listed in Papp (1992, Tables 1 and 2). The 1992 article used the Shannon-Wiener index (H'). Four indices remain to be discussed. These were formerly used by one of us (J. Izsák) to analyze the diversity of death causes and other epidemiological conditions (Izsák and Juhász-Nagy 1984, Hunter and Izsák 1990). Two of these indices belong to the Hurlbert diversity index family,

$$s(m) = \sum (1-(1-p_i)^m) \quad (m=1,2,\ldots) \ .$$

In this we define p_i as the probability of the i-th category, s the number of species in the sample, n_i species abundance and n the sum of the n_i-s. The estimation is based on:

$$\hat{s}(m)=\sum_{i=1}^{s} (1-(1-\binom{n-n_i}{m}/\binom{n}{m}) \quad (m \leq n) \ ,$$

This is the minimum variance unbiased estimator of s(m) (Smith and Grassle 1977). We note that $s(2) = 1 - \Sigma p_i^2$, which apart from an additive constant is identical to the Simpson diversity index.

s(2) emphasizes the dominance of the leading frequencies. By increasing parameter m, the indices become more and more sensitive to small frequencies. Therefore, when we apply different s(m) indices, different characteristics of the frequency structure emerge. However, the increaseability of m is limited by the fact that an increased weight of the low frequencies destabilizes the numerical results. In addition, the condition $m \leq n$ also limits the increase of m. Taking these points into account, we chose s(50) as our measure, which stresses the low frequencies to a higher degree.

Another index family (Hill 1973) is defined by

$$N_a = (\sum p_i^a)^{1/1-a}, \quad N_1 = \lim N_a.$$

The natural logarithm of this quantity is Rényi's (1961) entropy of order a.

An estimation formula of N_a is

$$N_a = (\sum (n_i/n)^a)^{1/1-a}.$$

We calculated this index with a = 0.5 parameter value. This puts relatively large stress on the smaller frequencies.

Fisher's α diversity index was also calculated. This is widely used in diversity studies in entomology. Its estimation can be based on

$$\alpha=n/(x^* - 1) \ ,$$

where x^* is an approximate solution of the equation

$$x-1-(n/s) \ln x = 0 \ .$$

x^* can be obtained by iteration (Birch 1963, Izsák and Juhász- Nagy 1982).

The *jackknifed estimation* and determination of the associated confidence intervals are an integral part of our investigations. The essence of the jackknife method is an algorithm, by which further samples are formed from the original sample by omission of certain elements (Miller 1974, Frangos 1987). In the simplest case applied also here, new samples are formed by omitting a single element of the original sample. Let the original estimate and the estimate after omission of the i-th element be denoted by $\hat{\Theta}$ and $\hat{\Theta}_{-i}$, respectively. The formula of the jackknife estimate is

$$J(\hat{\Theta}) = n \hat{\Theta} - ((n - 1)/n) \sum_{i=1}^{n} \hat{\Theta}_{-i} \ .$$

It must be noted that in ecological diversity jackknife calculations not a fixed number of elements but mostly entire relevés are omitted (cf. Zahl 1977, Heltshe and Forrester 1987). The theoretical basis of this modified method is not clear (Izsák 1994). Interestingly, the original jackknife method has seldom been applied for the estimation of diversity indices (Adams and McCune 1979).

Calculations were made on a PC AT computer with programmes written in BASIC and in PASCAL. Original and jackknifed estimates with 95% confidence intervals were calculated for all drosophilid imago samples. It is worth to mention the algebraic identity of s(m) and its jackknifed value.

The following statistic serves to calculate the *significance of diversity differences*:

$$t = \left| \frac{J(\hat{\Theta}_1) - J(\hat{\Theta}_2)}{s_d} \right|$$

with

$$s_d = \left[\frac{(SQ_1 + SQ_2)\,(n_1 + n_2)}{n_1 \cdot n_2\,(n_1 + n_2 - 2)} \right]^{1/2}$$

SQ is the sum of squares of differences for the pseudovalues $n\hat{\Theta} - (n-1)\hat{\Theta}_{-i}$. The t statistic is approximately t-distributed. The critical t-value is to be determined on the required probability level at degree of freedom $\nu = n_1 + n_2 - 2$.

An essential question is *how sensitive are the diversity indices to changes in a single frequency.* The formulation of the problem is not new. Rajczy and Padisák (1983) attempted to treat the question. They registrated the effect of dropping single frequencies. To study the infinitesimal sensitivity, another method was developed. Denoting by D the actual diversity function, the infinitesimal sensitivity of the index relating to changes in the l-th frequency can be expressed by

$$\Delta_l = \frac{\partial D(x)}{\partial x_l} \cdot \frac{n_l}{D(x)} .$$

The partial derivatives are in concrete form (Izsák 1991, 1992):

$$\frac{\partial s(m)}{\partial x_l} = \frac{m}{n} \left[\left(1 - \frac{n_l}{n}\right)^{m-1} - \sum_{i=1}^{s} \frac{n_i}{n} \left(1 - \frac{n_i}{n}\right)^{m-1} \right]$$

$$\frac{\partial H'}{\partial x_l} = -\frac{1}{\sum n_j} \ln \frac{n_l}{\sum n_j} - \frac{1}{(\sum n_j)^2} \sum_{i=1}^{s} n_i \ln \frac{n_i}{\sum n_j}$$

The partial derivative for the Hill indices is:

$$\frac{\partial N_a}{\partial x_l} = \frac{a}{1-a} \left[n^{a/(a-1)} \cdot n_l^{a-1} \left(\sum n_i^a\right)^{a/(a-1)} - n^{1/(1-a)} \left(\sum n_i^a\right)^{1/(1-a)} \right].$$

Plotting Δ_l as a function of n_l, characteristic graphs can be obtained. These show numerous aspects of sensitivity. Actual results are presented for a combined species abundance list. It is based on collection at Aggtelek in 1988, 1989 and 1990 (cf. Papp 1992, Table 2A).

A simple algorithm was constructed (Izsák 1992) to demonstrate a *surprising reduction of diversity differences by the substitution of some frequencies.* The essentials of the methods are the following. Consider the frequency vectors **a** = $(a_1, \ldots, a_n)$ and **b** = $(b_1, \ldots, b_n)$, based on a common set of categories. The algorithm searches in every step such an i index that the substitution of a_i by b_i maximally reduces the diversity difference between **b** and the new vector. When such a b_i does not exist, we substitute b_i determined by the equation

$$| a_i - b_i | = \max_j | a_j - b_j |.$$

The algorithm is suboptimal as regards the speed of approximation. Nevertheless, after a few steps the difference of the diversity values practically disappears. The calculations were made on a PC AT computer in BASIC.

Results and discussion

The original and the jackknifed estimates as well as confidence interval radii are summarized in Table 1. (When calculating H', decimal logarithm and not natural logarithm was used. That is why one must multiply the values given in our previous paper by 0.43429 in order to get the values in Table 1.) The high number of numerical results can be regarded as a statistical multitude, in which stochastic relationships exist. One of the important aspects of the study is the generalization about statistical properties of specific diversity indices.

The relationship between the original and jackknifed estimates

In the case of H', the original estimate is always lower than the jackknifed one. This is not unexpected, since H' is a downwards biased estimate (Basharin 1959) and jackknifing reduces the bias in most cases. Incidentally, H' can also be corrected with the formula given by Hutcheson (1970). However, the jackknife estimation is regarded more favourably, since it can be applied in a similar way to the majority of the diversity indices. The jackknife estimation of N_a gives higher values in every case than the original estimation. This is again in connection with the fact that N_a is downward biased. The situation is different with α, as the $\hat{\alpha}$ estimates can be lower or higher than $J(\hat{\alpha})$.

Increase in sample size reduces the bias. As is to be expected, with an increasing sample size the difference between the original and jackknifed estimate decreases. For indices H' and $N_{0.5}$ we can recognize an intensive reduction in difference up to the 200 value of sample size (Figs. 1-2). The increase is moderate above that value. One can relate this fact to independent findings. In a study of death cause diversity it was found that extreme fluctuations of the diversity index values disappear at about a sample size 100 (Izsák 1984). Based on drosophilid statistics published by Papp (1992) one can infer about rare species in higher number when the sample size is above 400. The relative difference of the original and the jackknife estimates varies by the diversity index. It is particularly high with the α index and even much higher with N_a than with H' (Figs. 1-2). As for α, the background of the high differences is that the above estimation is not "smooth". This condition generally limits the application of jackknifing.

Confidence intervals

The relative confidence interval radius is defined as the quotient: confidence interval radius/estimated index value.

Table 1. Original and jackknifed diversity index estimates with the radius of confidence intervals (CI). Probability level: 0.95. For s(2) and s(50) the original and jackknifed estimates are identical.

Sample code	Total number of cases	s(2) original and jack-knifed	Rad. of CI	s(50) original and jack-knifed	Rad. of CI	α		Rad. of CI	$N_{0.5}$		Rad. of CI	H'		Rad. of CI
						original	jackknifed		original	jackknifed		original	jackknifed	
A10058	104	1.7834	0.0471	8.4546	1.5532	2.7270	2.4622	1.9440	7.3257	8.0617	1.4688	0.7633	0.7876	0.0835
A19078	403	1.5863	0.0504	7.0591	0.8405	2.8167	3.1066	0.9258	6.4663	7.0983	1.0923	0.5789	0.5885	0.0562
A14098	20	1.5158	0.3095			2.9062	3.8131	1.5859	4.1915	5.9354	2.8159	0.4686	0.5689	0.3088
A12108	357	1.5118	0.0516	5.9517	0.9593	2.9049	3.6989	0.9878	5.4099	6.2450	1.2227	0.4728	0.4848	0.0603
A13108	115	1.7606	0.0351	8.3086	2.1137	3.3725	3.9910	2.2602	7.2827	8.7109	2.0600	0.7168	0.7484	0.0931
A02089	107	1.5495	0.1075	7.2849	2.0291	2.6992	3.1591	3.9735	5.5697	6.6073	1.7212	0.5386	0.5662	0.1152
A07099	148	1.7865	0.0355	8.5263	1.3489	2.7463	2.1785	1.7012	7.6458	8.1497	1.1701	0.7703	0.7880	0.0696
A15080	366	1.8405	0.0185	11.2087	1.0924	4.8400	4.8913	2.8136	12.4264	13.6613	1.7300	0.9363	0.9520	0.0509
A11090	61	1.7016	0.0650	5.7596	1.7088	1.6497	1.4046	2.0456	4.5360	4.9514	1.1142	0.5798	0.6030	0.0951
A25090	10	1.7556	0.3452			3.9795	3.1627	1.8534	4.4226	6.2998	3.0282	0.5903	0.7481	0.3552
M21088	395	1.6638	0.0398	6.4617	0.9038	2.8309	3.1191	0.9628	6.5141	7.1764	1.1142	0.6211	0.6310	0.0503
M22088	268	1.6852	0.0496	8.6237	1.2233	3.7313	4.1101	1.3389	8.2198	9.3053	1.6057	0.6961	0.7131	0.0701
M17098	111	1.6975	0.0708	7.4829	1.4254	2.3126	2.2020	1.7395	6.1370	6.7520	1.3208	0.6574	0.6780	0.0900
M15108	186	1.7121	0.0449	7.5592	1.1943	2.5580	2.6838	1.3875	6.6506	7.3469	1.3116	0.6722	0.6880	0.0703
M01079	104	1.7909	0.0554	9.0203	1.7913	3.1073	3.0691	2.1347	8.0186	9.0450	1.7642	0.7981	0.8264	0.0881
M28079	287	1.7991	0.0197	8.4507	1.1341	3.3649	3.5433	1.2700	8.6690	9.5525	1.4316	0.7946	0.8089	0.0527
M29079	333	1.7270	0.0296	7.8049	0.9431	2.6943	2.5116	1.1032	7.3431	7.8039	1.0049	0.6993	0.7089	0.0524
M30079	353	1.7367	0.0180	5.5866	0.7035	1.6818	1.5233	0.8228	5.4400	5.6727	0.6559	0.6379	0.6439	0.0361
M01109	112	1.8156	0.0347	9.7486	1.6961	3.4064	2.8756	2.1319	8.6658	9.5606	1.6361	0.8296	0.8568	0.0828
M18080	363	1.7523	0.0305	9.0141	1.0612	3.9784	4.2637	3.4538	9.3383	10.3632	1.5118	0.7723	0.7861	0.0559
M19080	358	1.7995	0.0244	9.1812	1.0089	3.7125	3.7903	1.0808	9.6739	10.5602	1.4102	0.8298	0.8426	0.0498
M23090	32	1.8347	0.0860			4.9935	5.3202	1.6167	8.1879	10.8713	3.3144	0.8327	0.9309	0.1798
B21078	1070	1.7575	0.0209	9.3522	0.4228	3.0742	3.2604	5.1013	9.4156	9.8906	0.8352	0.8013	0.8057	0.0299
B14108	1200	1.3738	0.0344	5.8489	0.4076	2.2250	2.5546	1.8409	4.8428	5.1427	0.5717	0.3929	0.3961	0.0336
B03089	276	1.5319	0.0685	7.7483	1.0745	2.8326	2.8745	1.2206	6.4618	7.0667	1.1696	0.5557	0.5680	0.0736
B08099	195	1.8417	0.0256	11.0824	1.4246	4.1277	3.4121	1.7906	11.0270	11.9071	1.5585	0.9187	0.9391	0.0644
B11109	121	1.7576	0.0489	8.2960	1.3490	2.5861	2.3842	1.7209	7.0962	7.7532	1.3648	0.7348	0.7555	0.0816
B16080	294	1.7972	0.0377	11.2290	1.1332	4.5386	4.9940	8.0642	11.5942	12.9943	1.8818	0.8958	0.9140	0.0621
B13090	216	1.6290	0.0598	5.9381	0.9486	1.8973	2.1787	1.0747	5.2100	5.7355	1.0542	0.5705	0.5819	0.0634
B27090	196	1.4642	0.0830	6.3371	1.4063	2.8197	3.7698	1.4522	5.2586	6.3199	1.5340	0.4625	0.4819	0.0876
V20088	75	1.6760	0.0848	7.7725	2.3784	2.6704	2.8013	2.6418	5.8483	6.8566	1.7982	0.6208	0.6545	0.1206
V14079	213	1.7492	0.0386	9.3018	1.3425	3.6807	3.7836	1.5361	8.7028	9.7476	1.6516	0.7593	0.7786	0.0727
V27079	216	1.8259	0.0266	10.6011	1.2971	3.9903	3.9973	1.5715	10.5143	11.6396	1.7383	0.8873	0.9072	0.0638
V03099	72	1.6776	0.0750	6.5693	1.4925	1.9165	1.5782	1.9018	5.0276	5.4607	1.1427	0.5910	0.6138	0.1033
V07090	158	1.6690	0.0767	8.8962	1.1247	2.6894	2.1537	1.5493	7.3042	7.7510	1.1186	0.6929	0.7091	0.0862

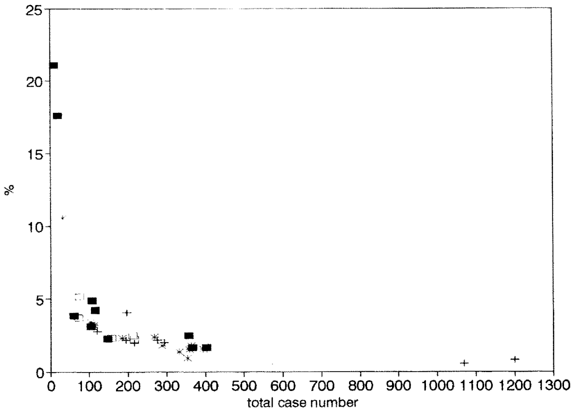

Figure 1. Dependence of the difference of original and jackknifed H' estimates. The source of data is Table 1. Symbols ■, +, *, and ❑ signify Aggtelek, Bükk, Börzsöny and Visegrád.

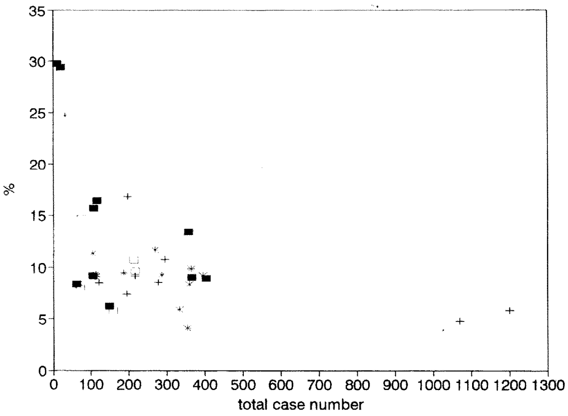

Figure 2. Dependence of the difference of original and jackknifed $N_{0.5}$ estimates. The source is Table 1. Symbols as in Figure 1.

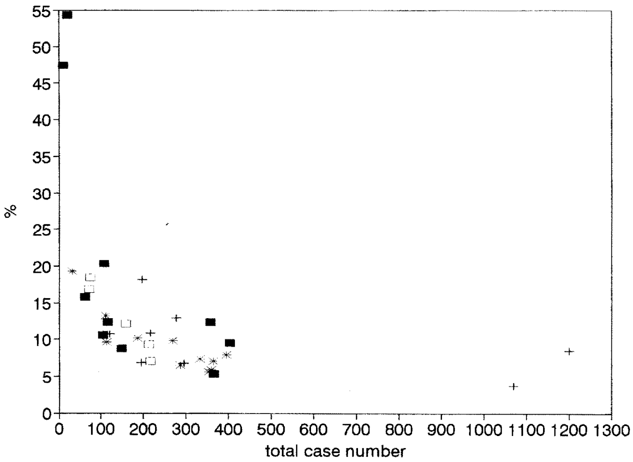

Figure 3. Dependence of the CI radius*100/$J(\hat{H}')$ ratio on the case numbers. Probability level: 0.95. Symbols as in Figure 1.

This is also very high with the α index. One can expect that in general the relative radius is a decreasing function of the sample size. This can be seen for H' in Fig. 3. The steep decrease of the relative radius ceases at about sample size 200. The same is true for indices s(2), s(50) and $N_{0.5}$.

We can conclude, as an obvious tendency, that the cost of increasing sensitivity to low frequencies is the decrease in the confidence of the estimation.

Correlations between the indices

A level of positive correlation between the various indices has since long been known (Magurran 1988). Spearman's rank correlation coefficient is appropriate for analysis. The reason is that one cannot postulate the joint normal distribution of the indices and in practice in most cases only the ranks of diversity values are relevant. Results on the A and M data are given in Table 2. The strong correlation of indices is obvious, albeit α is peculiar in this respect. The 1-d diversity index, derived from the Berger- Parker index d, was examined. This index is not normally regarded as a true diversity index, since its value is determined by a single (dominant) frequency. However, regarding the strong correlation of this index with all the others, we cannot exclude it. This raises the question: why do the usual diversity indices give similar information about the structure of frequencies as does the 1-d index? We think, there are two reasons for this. One is that the highest frequency plays a special role in determining the value of these indices. Furthermore, there are subdominant frequencies that have very little influence. We shall return to this subject when we discuss the index sensitivity. Some diversity indices are correlated with each other to a lesser extent than with the 1-d index. This is explained by their different sensitivity to medium size frequencies. We must not forget that in the example only the differences in mathematical properties can affect the variability of index values, since the frequency vector is fixed.

Table 2. Spearman's correlation coefficients of the diversity index values. Upper rows: site at Aggtelek, lower rows: site at Börzsöny.

	s(2)	s(50)	α	$N_{0.5}$	H'	1-d
s(2)		0.833	-0.030	0.769	0.988	0.990
		0.718	-0.343	0.741	0.881	0.980
s(50)			0.333	0.905	0.833	0.963
			0.555	0.891	0.964	0.900
α				0.321	0.006	0.840
				0.825	0.741	0.837
$N_{0.5}$					0.733	0.964
					0.895	0.920
H'						0.986
						0.942

Testing significance of diversity differences

One of the advantages of the jackknife estimation is that it facilitates significance tests on differences. However, when several indices are applied, the significance test can give a positive result for one and a negative result for another. A pattern emerges in this way in the significance of diversity differences, which puts in doubt the advisability of categorical statements. This dubious situation will not arise if a single index is used. In the course of our significance studies diversity values of the species abundance lists combined yearly and combined for the three years are compared. The t values are given in Table 3. The jackknifed estimates and significances at the 95 % level are summarized in Table 4.

It is to be observed that the ranks of diversity values in consecutive years differ, even regarding a single index. Similar deviations are found comparing the numerical values of the different indices for the same year. It is conspicuous that there are very few significant differences in the case of the α index. We can conclude that even if we use significance tests, only a tendency can be established rather than a clear-cut diversity relationship. One of the tendencies is that the diversity at site A can be regarded as highest, followed by that at site M. Sites B and V go alternately to the last place. The diversity ranks differ again with the different indices (cf. Papp 1992, Table 3).

As it appears from our numerical studies, the results cannot be stabilized even by calculation of several diversity measures. The diversity of samples taken at the same site also shows considerable temporal variation. To illustrate this, the diversity of species abundance values from consecutive days were compared. In Table 5, significance levels are given for diversity differences of five pairs of days. The estimates themselves are given in Table 1. Significant differences on 0.95 % level are rather common.

Infinitesimal index sensitivity

In the course of sensitivity investigations, a number of characteristic properties were observed on the Δ_l graphs. For convenience, the Δ_l values are plotted against log n_l. A continuous curve was drawn through the graph points by hand.

Calculations were made on the basis of the three years combined abundance values. Results are given in Fig. 4 for

Table 3. Comparison of diversity values. t statistics are calculated according to the formula in text. Asterisk indicates significant differences. Probability level 0.95 is used.

Sites compared	1988	1989	1990	1988-1990 combined
		s(2)		
A-B	13.41*	-0.61	7.85*	12.87*
A-M	4.04*	1.13	5.96*	3.03*
A-V	6.22*	0.66	7.92*	3.18*
B-M	-10.07*	2.41*	-3.72*	-15.02*
B-V	-1.27	1.50	3.12*	-7.25*
M-V	2.83*	-0.36	5.74*	1.54
		s(50)		
A-B	6.85*	-4.94*	4.14*	9.25*
A-M	3.35*	0.99	4.28*	9.17*
A-V	3.13*	-0.78	3.77*	5.97*
B-M	-2.53*	8.08*	0.27	1.12
B-V	1.33	4.80*	0.95	0.13
M-V	1.91	-2.29*	0.74	-0.53
		α		
A-B	0.86	-8.75*	0.31	0.11
A-M	0.63	-0.31	-0.03	-0.10
A-V	0.47	-8.87*	3.35*	0.25
B-M	-0.06	-0.24	-0.05	-0.21
B-V	0.10	-2.33*	2.59*	0.05
M-V	0.10	0.17	0.64	0.59
		$N_{0.5}$		
A-B	7.44*	-4.41*	3.32*	7.33*
A-M	3.14*	-1.35	2.59*	3.49*
A-V	2.82*	-1.54	5.01*	4.41*
B-M	-3.88*	2.98*	-0.72	-3.98*
B-V	1.14	3.02*	2.64*	-0.74
M-V	2.46*	0.04	2.92*	1.74
		H'		
A-B	12.85*	-3.17*	6.89*	40.18*
A-M	3.71*	0.87	5.59*	6.23*
A-V	4.91*	-0.13	7.22*	5.16*
B-M	-8.98*	5.78*	-1.94	-9.58*
B-V	0.02	3.63*	2.10*	-4.73*
M-V	3.44*	-1.28	3.49*	1.00

Table 4. The sequence of sites in descending order of diversity values. Symbols >> indicate significance at the 0.95 level.

s(2)		$N_{0.5}$	
1988	A>>M>>V>B	1988	A>>M>>B>>V
1989	B>A>V>M	1989	B>>M>V>A
1990	A>>M>>B>>V	1990	A>>M>B>>V
combined	A>>M>V>>B	combined	A>>M>V>B
s(50)		H'	
1988	A>>M>>B>V	1988	A>>M>>B>V
1989	B>>V>A>M	1989	B>>V>A>M
1990	A>>B>M>V	1990	A>>M>B>>V
combined	A>>B>V>M	combined	A>>M>V>>B
α			
1988	A>M>B>V		
1989	M>V>>B>>A		
1990	M>A>B>>V		
combined	M>A>B>V		

Table 5. Significant differences (+) between the diversity values of two consecutive colletions.

Codes of the consecutive days	s(2)	s(50)	α	$N_{0.5}$	H'
A12108-A13108	+	+	-	+	+
A21088-A22088	-	+	-	+	-
M28079-M29079	+	-	-	+	+
M29079-M30079	-	+	-	+	+
M18080-M19080	+	-	-	-	-

the Aggtelek site. The graphs, relating to the other sites are similar. The sensitivity of the index is positive in the domain of low frequencies. This means that a small increase in frequency increases the index value. In this domain the measure of sensitivity increases with the frequency itself. After reaching a maximum, sensitivity decreases, then it becomes negative. The highest absolute valued (negative) sensitivity belongs to the highest frequency. The sensitivity in the low frequency domain decreases in the order of indices $N_{0.5}$, H', s(50) and s(2). It is very informative that sensitivity is very low with some high frequencies. For instance, sensitivity of the index s(2) is not higher at the abundance value of 237 (the third highest frequency) than at the abundance value of 7. Sensitivity to the leading frequencies is predominating even in the case of the indices that put more stress on the low frequencies. This and several other properties of the indices are expressed by the graphs. For instance, the a parameter of index N_a was chosen as small as needed so that instability of the index value approximates the limit of tolerance. However, the relative change in index value as a response to change in leading frequency (427) is 8.4 times higher than to change in the frequency value of 29, the point of local maximum of sensitivity.

According to the foregoing numerical properties of the diversity indices, the cause of differences in diversity in the drosophilid samples is attributed to their differences in high frequencies. However, since the sensitivity may be unexpectedly small in some of the high frequency cases (see above), it seems advisable to state this feature more carefully: *differences in the diversity index values can be attributed to changes in a surprisingly low number of frequencies (of species).*

Sensitivity studies by substituting frequencies

One can verify the foregoing statement from another respect, by exchanging one by one frequencies of the first

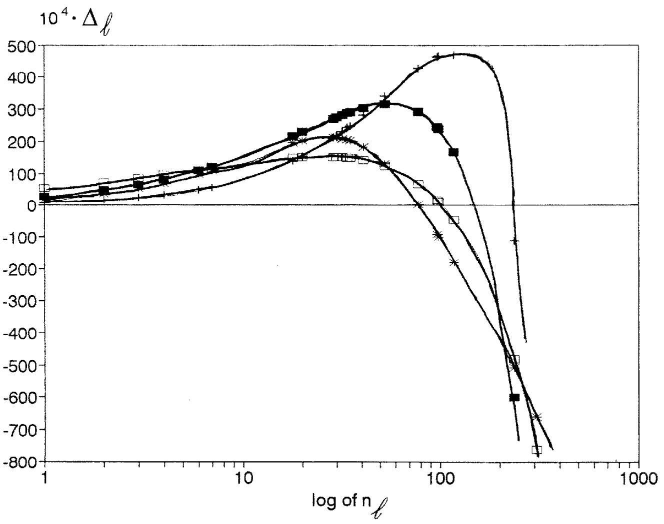

Figure 4. Relationship between frequency and index response. Calculations are based on the combined 1988-90 data for Aggtelek site. Symbols: ■: $10^4 \cdot H'$, *: $10^4 \cdot s(50)$, ❑: $10^4 \cdot N_{0.5}$, +: $10^5 \cdot s(2)$.

substitutions for s(2)

B	A	A(1)	A(2)	A(3)
1	0	0	0	0
1	0	0	0	0
1	7	7	7	7
2	91	91	91	91
10	2	2	2	2
0	1	1	1	1
8	25	25	25	25
17	25	25	25	25
0	1	1	1	1
12	2	2	2	2
21	6	6	6	6
1	0	0	0	0
0	3	3	3	3
4	0	0	0	0
33	85	85 →	33	33
3	4	4	4	4
3	18	18	18	18
0	3	3	3	3
1	0	0	0	0
1	1	1	1	1
1	0	0	0	0
63	0	0	0	0
56	30	30	30	30
4	0	0	0	0
0	1	1	1	1
263	42 →	263	263	263
52	65	65	65 →	52
0	3	3	3	3
208	1	1	1	1
0	20	20	20	20
0	1	1	1	1
index value 1.7594	1.8736	1.7889	1.7677	1.7617

substitutions for s(50)

B	A	A(1)	A(2)	A(3)
1	0	0	0	0
1	0	0	0	0
1	7	7	7	7
2	91	91	91	91
10	2	2	2	2
0	1	1	1	1
8	25	25	25 →	8
17	25	25	25	25
0	1	1	1	1
12	2	2	2	2
21	6	6	6	6
1	0	0	0	0
0	3	3	3	3
4	0	0	0	0
33	33	85	85	85
3	4	4	4	4
3	18	18	18	18
0	3	3	3	3
1	0	0	0	0
1	1	1	1	1
1	0	0	0	0
63	0	0	0	0
56	30	30	30	30
4	0	0	0	0
0	1	1	1	1
263	42 →	263	263	263
52	65	65	65	65
0	3	3	3	3
208	1	1	1	1
0	20	20 →	0	0
0	1	1	1	1
9.6865	12.2128	10.661	39.967	9.6684

Figure 5. The process of consecutive substitutions of frequency values combined for 1990 between sites A and B. Rows of the original table were omitted where the given species were absent from both sites.

vector of vector pairs A-B, A-M and B-M for those of the second vector. The algorithm is detailed under Methods. In the course of this process one can trace how the diversity in the modified vector approximates the diversity in the second vector. The procedure was performed for indices s(2) and s(50) which are sensitive to low frequencies to a smaller and a larger extent, respectively. The process of approximation is illustrated on the abundance vector pair from the A and B sites (see Fig. 5). Notwithstanding that the algorithm is suboptimal, the difference in diversity index values practically disappears after a few substitutions. In the case of the index s(2), when the abundance values 42, 85 and 65 are exchanged for 263, 33 and 52 of vector B, the difference in diversity values of vector B and of the slightly modified vector A will be as low as 0.0023 (2.01%). In the case of s(50), if the abundance values of 42, 20 and 25 are exchanged for abundances 263, 0 and 8, the difference will be 0.018 (0.72%). These abundance values belong to the species of *Drosophila phalerata, D. (Sophophora) obscura* and *Drosophila testacea* in the first case, to the species of *Drosophila phalerata, D. (Hirtodrosophila) oldenbergi* and *Scaptomyza graminum* in the second. Diversity index values of the modified vector A and of vector B are almost the same, although the two abundance lists actually *remain very different.* It is a matter of course that this algorithm does not provide the "essential proof" for the explanation of differences in diversity. The aim is the illustration of unexplored possibilities.

Summary

Diversity investigations can make use of methods often applied by statistical ecology. However, the methodology of the application of diversity indices seems to lag. This is owing the scarcity of knowledge of the mathematical properties of the sample statistics. As a consequence, determination of confidence intervals and the examination of significance are cumbersome. To overcome the difficulties we applied the jackknife technique. It gave us a realistic and reliable basis for spatial and temporal comparisons of *alpha* diversity.

Our analysis was performed on a Diptera data collection. The frequency lists that contain more than 9,000 specimens of 40 drosophilid species were published in a previous paper. Estimations of five diversity indices and their jackknife estimations were calculated for 35 samples collected in four sites over a period of three years. The radius of confidence intervals of the jackknifed values was also calculated. These large sets of results allowed us to study the numerical properties of jackknife estimation empirically.

It was found that the difference of the strongly biased original estimation and the jackknife estimation becomes minor if the number of all cases is greater than 200. In addition, the width of the relative confidence interval is also decreased to 5-10 % when the number of all cases is over 200. If the number of all cases is as high as 400, its further increase results only in slight improvements.

Jackknife estimation serves as a tool of investigations of the significance of differences in diversity values. These calculations were made for three years with all the diversity indices used. It was found that the different diversity indices suggest somewhat different notions about diversity relations. We also analysed how far the statements published in our previous paper agree with the results gained by the jackknife method.

Sensitivity studies were also performed. By means of these we gained considerable insight of index responses. Attention is called to the existence of some subdominant categories, to the changes of which the diversity index value is surprisingly insensitive. We used sensitivity graphs for better interpretation of diversity differences.

Finally, an object-lesson is demonstrated in connection with interpretation of differences in diversity index values. By means of an appropriate algorithm it is possible to show that by changing a surprisingly low number of the elements of a species abundance list for those of another list, the difference of the two index values practically diminishes.

Acknowledgement. It is with grateful memory that we mention the late Professor Pál Juhász-Nagy (Eötvös Loránd University, Dept. of Plant Taxonomy and Ecology, Budapest) who gave invaluable advice during the initial steps in this study.

REFERENCES

Adams, J. E. and McCune, E. D. 1979. Application of the generalized jackknife to Shannon's measure of information used as an index of diversity. In: Grassle et al. (eds): Ecological Diversity in Theory and Practice, International Publishing House, Fairland, Maryland, pp. 117-131.

Bächli, G. 1979. Quantitative methods for population analysis of a Drosophila (Diptera, Drosophilidae) collection. Aquilo Ser. Zool. 20: 33-40.

Bächli, G. and Rocha Pite, M.T. 1982. Annotated bibliography of Palaearctic species of Drosophilidae (Diptera). Beitr. Ent. 32: 303-392.

Basharin, G. P. 1959. On a statistical estimate for the entropy of a sequence of independent random variables. Theory of Probability and Its Applications 4: 333-336.

Birch, M. W. 1963. An algorithm for the logarithmic series distribution. Biometrics 19: 651-652.

Burla, H. and Bächli, G. 1991. A search for pattern in faunistical records of drosophilid species in Switzerland. Z. Zool. Syst. Evolut.-forsch. 29: 176-200.

Frangos, C. C. 1987. An updated bibliography on the jackknife method. Comm. Statist. A 16: 1543-1584.

Heltshe, J. F. and Forrester, N. E. 1985. Statistical evaluation of the jackknife estimate of diversity when using quadrat samples. Ecology 66: 107-111.

Hunter, P. R. and Izsák, J. 1990. Diversity studies of Salmonella incidents in some domestic livestock and their potential relevance as indicators of niche width. Epidemiology and Infection 105: 501-510.

Hutcheson, K. 1970. A test for comparing diversities based on the Shannon formula. Journal of theoretical Biology 29: 151- 154.

Izsák, J. and Juhász-Nagy, P. 1982. Studies of lognormality on mortality statistics. Biometrical Journal 24: 731-741.

Izsák, J. and Juhász-Nagy, P. 1984. Studies of diversity indices based on mortality statistics. Annales Univ. Sci. Budapestinensis (Sectio Biologica) 24-26: 11-27.

Izsák, J. 1991. New aspects of sensitivity investigations on diversity indices. A case study. Biometrie und Informatik in Medizin and Biologie 22: 107-115.

Izsák, J. 1989. Numerical experimentation on diversity curves applied in epidemiology. Biom. Praxim. 29: 97-113.

Izsák, J. 1992. Sensitivity studies on Hurlbert's indices of diversity. Biom. Praxim. 32: 101-114.

Izsák, J. 1994. Applying the jackknife method to significance tests of diagnostical diversity. Meth. Inf. Med. 33:214-219.

Magurran, A. E. 1988. Ecological diversity and its measurement. Croom Helm Ltd., London, chap. 4.

Miller, R. G. 1974. The jackknife - a review. Biometrika 61: 1-15.

Papp, L. 1992. Drosophilid assemblages in mountain creek valleys in Hungary (Diptera: Drosophilidae) I. Folia Ent. Hung. 53: 139-153.

Rajczy, M. and Padisák, J. 1983. DIVDROP analysis. A new method for interpretation of species importance in diversity changes. Annls hist.-nat. Mus. Natn Hung. 75: 97-105.

Smith, W. and Grassle, F. 1977. Sampling properties of a family of diversity measures. Biometrics 33: 283-292.

Washington, H. G. 1984. Diversity, biotic and similarity indices. A review with special relevance to aquatic ecosystems. Water Research 18: 653-694.

Whittaker, R. H. 1972. Evolution and measurement of species diversity. Taxon 21: 213-251.

Zahl, S. 1977. Jackknifing an index of diversity. Ecology 58: 907-913.

Manuscript received: December 1993.