Listy Biometryczne – Biometrical Letters
Vol. 29 (1992), No. 1, 41-52

The widening range of diversity investigations

János Izsák

Berzsenyi College, Department of Zoology
Szombathely H-9701, Károlyi Gáspár tér 4, Hungary

Summary

Diversity investigations are time-honoured methods of statistical ecology. But it is worth at times gathering information about the diversity studies in other research fields. In this article the following types of investigations are dealt with: epidemiological, genetical, demographical and clinical diversity studies, the question of entropy in death cause tables and linguistic diversity studies.

It can be said of these disciplines that the diversity indices used are almost exclusively dominance indices and the methodology of the ecological diversity studies is the most developed one. At the same time the various conditions of the enumerated diversity investigations are very instructive for ecological statisticians as well. The cited articles aid exploration of literature of the reviewed fields.

1. Introduction

Diversity investigation is a classical method of statistical ecology (Williams 1964, Pielou 1975, Grassle et al. 1979, Washington 1984). Besides this there are other applications of diversity studies as well, some of which are less known by the ecologists, being more peripheral from the point of view of ecology, or which have only recently emerged.

Earlier Patil and Taillie (1982) have given a brief account of several disciplines of this kind. We think that it is certainly useful to enumerate some related topics, according to the recent development of diversity investigations. Namely, arguing in style, the diverse fields constitute different opportunities to develop diversity analysis methods, increasing by that the diversity concepts and methodical particulars. This can affect profitably the classical diversity investigations too. In addition, this article covers some literary background to epidemiological diversity investigations. Accordingly, in the following we refer often to possible links with this field.

Key words: diagnoses, diversity, entropy, genotypes, Zipf's law

2. A brief review of diversity investigations

Although the first aim of the paper is to point out possible links between diversity ideas and measuring methods in different research fields, it will be useful recalling some basic aspects of that investigations. Excellent articles and books treat both the conceptional and methodical questions in more details. As mentioned, diversity measures are used firstly in the statistical ecology. An illustrative example on this field help to follow the basic concepts. Consider a floristic collection or sample of plants. The sample frequencies of the s species represented in the collection are $n_1,...,n_s$. According to the well-known Shannon-Wiener index, the (floral) diversity, in other words the polynomial entropy relating to the categories of species is given by definition

$$H' = -\sum_{i=1}^{s} p_i \ln p_i ,$$

where p_i $(i=1,...,s, \sum p_i = 1)$ are the probabilities of the species. Substituting the relative frequency $n_i / \sum n_i = n_i / n$ for p_i, the estimate of H' is given by the formula

$$\hat{H}' = -\sum_{i=1}^{s} \frac{n_i}{n} \ln \frac{n_i}{n} .$$

Choosing a plant randomly, H' expresses the uncertainty concerning the represented species. If one calculates the H' values for different areas, information is obtainable about the departures of floral diversity. This often indicates differences in the general environmental conditions, such as in water quality. Based on the approximately normal distribution of the H' sample statistic, a single significance test serves to judge the significance of diversity differences (Hutcheson 1970). But the formulae of standard errors of other diversity indices (see below) are in general rather complicate. Thus significance tests are often omitted. Jackknifing is a suitable method to estimate the sample variance and confidence limits. In its course one constructs smaller samples from the original sample and estimates the diversity and the sample variance (Magurran 1988).

If s is fixed, H' reaches its maximum when the *evenness component* is maximal, that is the p_i 's are equal. Besides the maximal H' increases with an increasing category number. (Unfortunately, the latter, so called *richness component* and the evenness are non additive components of the diversity.)

The calculation of H' is not the single way to measure the diversity. There is a plethora of diversity measures. For example, a well known diversity index family is given by the formula (Hill 1973)

$$N_a = \left(\sum_i p_i^a \right)^{\frac{1}{1-a}} , \quad N_1 = \lim_{1} N_a .$$

This family of indices is related to Renyi's generalized entropy. It is worthy of note that $N_1 = \exp(H')$. The well known (reciprocal) Simpson index with the formula

$$N_2 = (\sum_i p_i^2)^{-1}$$

is also a member of this index family. A sample statistic arises also here by substituting relative frequencies for the p_i's. The sensitivity of the N_a indices to small frequencies (i.e. to rare species) changes inversely to a.

Another widely used index family is introduced by Hurlbert (1971). The formula is

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

Minimum variance unbiased estimate of $s(m)$ is

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

$\hat{\mathrm{V}}\mathrm{ar}(\hat{s}(m))$ is given in explicit form by Smith and Grassle (1977). $s(m)$ is the expectation of the number of categories represented in a sample obtained by choosing m elements (individuals) without replacement from the population. For example, in a population dynamical model an individual encounters on the average $s(m)$ species by encountering m individuals. Sensitivity to small frequencies increases with an increasing m. According to the above, $s(2)$ is the expectation of the category number in a sample with two elements. That is, $s(2) - 1 = 1 - \sum p_i^2$ (the well known Gini-Simpson index) is the probability of occuring different categories in the sample.

One more concept leading to a group of diversity indices is the mean rarity of the categories. Let $r(i)$ the rarity of the i-th category or species. One can define $r(i)$ in different ways. If the p_i probabilities are arranged in declining order ($p_1 > p_2 > \ldots > p_s$), then in the simplest case $r(i) = i$ and the mean rarity, R, is

$$\sum_{i=1}^{s} r(i)\, p_i = \sum_{i=1}^{s} i\, p_i\,.$$

A group of diversity indices is based on model distributions. Let f_r the sample frequency of species, which are represented by r individuals ($r = 1,2,\ldots$). Numerous models are known concerning the distribution of the ensamble or the f_r values. A well fit is often attainable by the logarythmic series distribution, the formula of which is (Fisher et al. 1943)

$$\alpha\, x^r / r \quad (r = 1,2,\ldots).$$

x is here the fitting parameter. Denoting with n the sample size, alpha equals to $n(1 - x)/x$.

The r-th member of the series approximates the f_r number of such species, which are represented with r individuals in the sample. α is an often used measure of diversity.

In the end we mention the quadratic entropy measure (Rao 1982). It makes possible to take into account the distance between the categories. Its formula is

$$\Pi' \Delta \Pi ,$$

where Π is the $(p_1,...,p_s)$ probability vector and Δ is an optional symmetrical matrix, the (i,j)-th element of which expresses taxonomical, genetic or other distance between the i-th and j-th categories.

To demonstrate concrete calculations, consider the following frequencies of species: 7, 3, 2, 1, 2, 15, 4, 1, 20, 3, 1. The number of categories is 11, the total sample number is 59. In this case

$$\hat{H}' = -(\frac{7}{59}\ln\frac{7}{59} + 2\cdot\frac{3}{59}\ln\frac{3}{59} + \ldots + \frac{20}{59}\ln\frac{20}{59}) = 1.890,$$

$$\hat{N}_{1/3} = ((\frac{7}{59})^{1/3} + 2\cdot(\frac{3}{59})^{1/3} + \ldots + (\frac{20}{59})^{1/3})^{3/2} = 9.194,$$

$$\hat{N}_1 = \exp(\hat{H}') = 6.619.$$

With the often used estimation formula

$$\hat{N}_2 = \sum_{i=1}^{s} \frac{n\,(n-1)}{n_i\,(n_i-1)}$$

we obtain

$$\hat{N}_2 = 59\cdot 58(\frac{1}{7\cdot 6} + \frac{2}{3\cdot 2} + \ldots + \frac{1}{20\cdot 19}) = 5.185.$$

$$\hat{s}(2) = 11 - \frac{2}{59\cdot 58}\left(\binom{59-7}{2} + \binom{59-3}{2} + \ldots + \binom{59-1}{2}\right) = 1.807$$

$$\hat{s}(50) = 11 - \frac{50!9!}{59!}\left(\binom{59-7}{50} + \binom{59-3}{50} + \cdots + \binom{59-1}{50}\right) = 10.495.$$

Substituting n_i/n for p_i and arranging the relative frequencies in decreasing order :

$$\begin{array}{ccc} 1. & 2. & 11. \\ \frac{20}{59}, & \frac{15}{59}, \ldots, & \frac{1}{59} \end{array}$$

we obtain

$$R=\frac{20}{59}+2\cdot\frac{15}{59}+\ldots+11\cdot\frac{1}{59}=2.746.$$

For an example of calculation of the α index, see Magurran (1988).

As it appears, the diversity values are quite different. On the other hand there exists a considerable correlation between them (Magurran 1988).

It is interesting, that diversity investigations have few links with other multivariate methods. ANOVA-like analyses are based on the equation

$$\mathrm{Div(total)} = \mathrm{Div(within)} + \mathrm{Div(between)}.$$

The latter component is the formal difference of Div(total) and Div(within). Mathematical aspects of diversity partitioning are treated by Nayak (1986). As for concrete studies, Rao and Boudreau (1984) dealt with the decomposition of blood group diversity in populations. According to the results, major part (60 %) of the diversity is explainable by the denomination of populations rather than by the countries (20 %).

Clustering and making a dendrogram on the basis of diversity values is possible when different ways of categorization are given. Rao and Boudreau (1984) clustered populations according to the diversities relating to the HLA-A, HLA-B, ABO, MNS and Rh blood group systems. Denoting by $d_1,\ldots,d_5$ and $d'_1,\ldots,d'_5$ the diversity values of two populations, a possible compound measure of the diversity difference is

$$-\sum_{i=1}^{5} \ln 2(d_i \cdot d'_i)^{1/2} \,/\, (d_i + d'_i\,).$$

Discriminating ability of diversity indices was analysed for example by Robinson and Sandgren (1984). They found H' the best discriminator between artificial cultures of algae. Conversely, discriminating power of typing methods in bacteriology was analysed by the Gini-Simpson index (Hunter and Gaston 1988).

The sparse links between diversity investigations and other multivariate methods is due to a number of causes. One of these is that both the functional idea of diversity and the practice of its measure are elaborated almost independently in the different research fields. Therefore a concise survey of non-ecological applications may prove to be useful.

3. Epidemiological diversity investigations

Herdan (1957) was the first who published a precursory article on this field. Being a statistical linguist, he treated diversity characters of morbidity diagnoses as a special kind of word variety of a text. He found that frequency distribution of registered diseases in hospitals fits satisfactorily to the logarithmic series. The α diversity index proved greater with the regional hospital group than with the teaching hospital group.

Our first publication relating to this topic appeared about ten years ago (Izsák and Juhász-Nagy 1980). At that time we were not aware of works of other authors in this field.

This is the reason why we were looking for investigations, the philosophy of which is similar to that of this topic.

The basic concept of epidemiological diversity investigations (EDI) is the following. Categorized data of morbidity or mortality statistics offer opportunity to study new diversity and concentration phenomena. For example, by studying the changes with age or sexual differences of epidemiological diversity, one can observe interesting facts. Concretely, the diversity of death causes, relating to the International Classification of Diseases (ICD) has a characteristic maximum about the 20th year of life. This is followed by a moderate decline. If the universe is not the whole ICD, but only sections, such as Neoplasms or Circulatory Diseases, the same observations can be made. With some ICD sections the maximum is followed by a characteristic hollow. This is more expressed with males, and is due probably to socioeconomical effects. Making a comparison between corresponding age groups, the diversity of death causes is in general greater in the female group (Izsák and Juhász-Nagy 1982, Izsák 1988, 1989). Investigations of this nature contribute to the adaptation of multivariate statistical methods in the analysis of the relatively underutilized morbidity and mortality data. For a wider exposition see Izsák and Juhász-Nagy (1984).

The topics, outlined below, show similar traits to these investigations. By reviewing them we point to the possible links between those and EDI.

4. Genetic diversity investigations

The term genetic or genotypic diversity is used for a type of genetic variability. All diversity indices used generally in the practice are adequate to measure it. Still the (reciprocal) Simpson index was used almost exclusively in this field. This index was introduced originally as the *effective allelic number*, which is similar to the *equivalent number of species* (Kimura and Crow 1964). The Gini-Simpson index, which can be also written in the form

$$\sum_{i \neq j}^{s} p_i p_j ,$$

is also used to measure genetic diversity (Zhang and Allard 1986). Here p_i is the probability of the i-th genetic category and s denotes the number of categories taken into consideration.

Occasionally, the Shannon-Wiener index (polynomial entropy) is also used (Lewontin 1972, Silander 1979). Taillie and Patil (1979) compared in a very informative paper the behavior of the Shannon-Wiener and the Gini-Simpson index, relating to their changes near the Hardy-Weinberg equilibrium. The fitness entropy, introduced by Desharmais (1986) is similar to the Shannon-Wiener index. Further diversity indices, suitable to the measuring of genetic diversity, are discussed by Gregorius (1978).

All these indices are really dominance indices in the sense that they overemphasize the dominant frequencies. It would be worth-while to study the genetic diversity by other diversity indices, for example by $s(50)$ or by $s(100)$, too. An interesting question is, for example, how the genetic diversity changes near the Hardy-Weinberg equilibrium, measured by indices, which are sensitive also to the rarer genotypes.

If the genetic categories are the alleles of a simple locus (in this case k is the number of alleles), the matter is the distribution and diversity of those alleles in the population (allelic or locus diversity). For a locus ensamble we can define the *mean allelic diversity* (Zhang and Allard 1986, Stoddart and Taylor 1988). This is the arithmetic mean of the allelic diversity values belonging to the loci. The ensemble of the considered loci can be regarded as a sample from a greater locus ensemble, occasionally from the whole ensemble.

In a more general treatment the genetic categories are genotypes regarding more loci with their gene pools. In this case p_i denotes the probability of the i-th genotype and we can speak about genotype or multilocus diversity (Stoddart 1983, Innes et al. 1986, Zhang and Allard 1986, Stoddart and Taylor 1988). The usual estimation of the Simpson index is named in this case as "*effective genotype number*". In certain cases this coincides whith the clonal diversity (Stoddart 1983), and the connection with the species diversity is clear. The relationship between the genetic diversity and the classical species diversity is theoretically obvious and it is impossible to draw a line between them.

From a theoretical point of view the connection between the genotype and epidemiological diversity is also clear. But in reality the transition is very entangled and complicated. In addition, the epidemiological diversity is strongly influenced by environmental factors. Concrete results relating to such correlations are to be expected perhaps in the diversity studies of diseases caused by parasites. Here one can assume a possible correlation between the presence of a genetically specified enzyme and the susceptibility to a specific infection. In the future it will be presumably possible to connect human genetical diversity investigations (cf. O'Brien et al. 1980, Lewontin 1982) with EDI.

5. Demographical and sociological diversity investigations

The basis of categorization in the demographic diversity studies are e.g. language, race, social status, political conviction, and religion (Lieberson 1969, Sharma 1986). As follows from the problems emerging in this field, the combination of categories is also frequent. Even the notation of those combinations resembles often the notation of genom types. For example, in the 4th table of the quoted work of Lieberson, AbcD is a response combination or compound attitude, where the great (small) letters denote positive (negative) response relating to categories of attitude towards races. The calculated Gini-Simpson index shows the probability, that two randomly selected persons have different compound attitude. Similar symbols and methods could be used in the future in the diversity studies of compound diagnoses.

It is interesting that the essential paper of Lieberson is unknown to many ecologists. These studies throw new light also upon the possible links between the classical demographic investigations and EDI, bringing forth new methodical aspects. From the methodical point of view both disciplines can be traced back to the concentration studies initiated by Gini (1912). Nowadays, the specialists of these fields know scarcely anything about each other. At the same time it is essential for EDI to take into account precedents of demographical diversity analyses. The reason is the similar (human) study material and the accompanying similarity of methods.

Recently, an attempt has been made to apply further diversity indices in demography. For example, White (1986) proposed diversity indices used in ecology for human populations. The author analysed the ethnic diversity of several cities in the U.S. relating to 15 ethnic categories. He found, among others, a positive correlation between metropolitan ethnic diversity and metropolitan population size. The indices used are the Shannon- Wiener index and the Simpson index.

Homicide is a relatively peripheral domain of mortality, from a biological point of view. Its relationship to ethnic diversity is treated in a suggestive article of Avison and Loring (1986). It is stated, that the effect of income inequality on homicide rates is exacerbated by increasing ethnic differentiation in the societies. A more remote research field of demography from the point of view of EDI is the analysis of areal concentration and segregation. An example of this is, for instance, Lichter's article from 1985. This discipline has also important methodical background and is worth to follow from the methodical aspects of EDI.

6. Diversity investigations on laboratory data

Although one could classify these investigations as genotype diversity studies or EDI, the tendency of literature is to indicate the outlining and spreading of this field. The relationship regarding genetic diversity is often obvious. This is the case, for example, for studying the diversity of red cell markers (Barbujani and Milani 1986). Similar questions have been discussed by Lewontin (1982). In an above quoted paper Rao and Boudreau (1984) treat blood group diversity patterns of different populations. This issue also has laboratory aspects. In other cases, the subject is specifically the diversity of clinical diagnoses. This stands naturally closer to EDI in relation to genetic diversity studies.

The use of Simpson index in analysing microbiological laboratory data (Hunter and Gaston 1988) is also mentioned above. Diversity analysis of Salmonella infections in domestic livestocks (Hunter and Izsák 1990) is related to laboratory activity, too.

7. Analysing the entropy of cause-of-death tables

Death cause tables, frequently used in EDI, give case numbers of death by age groups and by causes of death. The columns detail the distribution by cause in a given age group, the rows inform about the age distribution relating to a given cause of death. That is, the a_{ij} element of the $n \times m$ matrix $\mathbf{A} = \{a_{ij}\}$ $i = 1,\ldots,n;$ $j = 1,\ldots,m$ equals the number of deceased belonging to the i-th cause and to the j-th age group. So, the $\sum_i a_{ij}$ marginals (column sums) relate to the (empirical) age group distribution, while the marginals $\sum_j a_{ij}$ (row sums) give polynomial death cause distribution, regardless of age.

On the basis of such a matrix, one can define several sorts of entropy. Namely, a one-dimensional density function and entropy belong to the vector of column sums, while a polynomial entropy value belongs to the vector of row sums. One ought to note, that Theil, in his work on information theory, already long ago touched upon polynomial entropy as a measuring number of the concentration by qualitative categories (Theil 1967, Chap. 8).

Conditional entropy values belong to the individual rows and columns. We analysed with EDI even the entropy values belonging to the columns, investigating their age dependence and sexual differences (see at the References). It would be also interesting to investigate the joint entropy and the average conditioned entropies of mortality tables. Further, one could parallel the secular changes of death causes entropy (Izsák 1986, 1989) or other types of diversity indices with the well known rectangularization of the human survival curves. As regards similar (not identical) entropy concepts in demographical models, Keyfitz (1977) introduced the entropy defined by the formula

$$\left(-\int_0^w l_a \ln l_a d_a\right)\left(\int_0^w l_a d_a\right)^{-1},$$

where w is the oldest age of life and l_a is the survivorship in age a. Demetrius (1984) uses in his population dynamical models entropy functions regarding to the age of reproducing individuals, as well. For more details, see also in Tuljapurkar (1982). These entropy concepts will probably spread in the growing life history studies of ecology. Analogous extension of EDI can lead to new connection possibilities with commonly used demographical analyses. This could reduce the relative isolation of EDI.

8. Diversity investigations in linguistics

In linguistics diversity indices were used in the past to measure the variety of words occuring in a text. In the preceding decades the diversity investigations in this field and in ecology showed numerous common features. The connection is today of minor importance. As mentioned above, in the first paper on diagnostic diversity the matter was presented as a special case of statistical linguistics. But Herdan's related work (1957) remained for a long time unnoticed. Much later Höpker (1975) had used the well-known Zipf law of statistical linguistics to analyse the rank statistics of diagnoses.

An important question is the optimal construction of thesauri in diagnostics (Kayser et al. 1980). This is related to the mean gain in information for each search step. The best solution is derived in part from the polynomial entropy of the diagnoses. In this sense the diagnostic diversity correlates to the optimal construction of a thesaurus. To form further connection, it would be worthy to extend EDI over the individual and dialectical characterization of the working diagnostic vocabulary of a physician. In connecting EDI to the study of diagnostic strings, once again we use methods of formal linguistics.

We have reached, herewith, the end of the enumeration of some disciplines which can be related to EDI. As the literature is rather sparse, new scopes can at any time come into light. A good example for remote, peripheral fields is the diversity investigation of firms (Varadarajan and Ramanujam 1987).

9. Conclusions

Diversity investigations are widely used in statistical ecology. There are also other disciplines, where the idea of diversity and some methods of its measurement are known. But the applied methodical apparatus is often rather scanty, compared with that in ecology. A brief account of various applications of diversity indices certainly contribute to reduce the relative isolation of diversity studies in very different research fields, such as epidemiology, genetics, demography, laboratory work, analysis of death tables and linguistics. In these briefly discussed topics the diversity investigations will potentially be in more general use. Further, up to now almost exclusively the Shannon-Wiener index, the Simpson and Gini-Simpson indices are used. Confidence intervals and significance tests are in general not applied.

In a brief article a more detailed report and discussion is unrealizable. We tried compensate this by offering a rather wide bibliography.

Acknowledgement

The author wishes to thank a referee for helpful comments and for suggestions in organizing this paper.

REFERENCES

Avison, W.R., Loring, P.L. (1986). Population diversity and cross-national homicide-the effects of inequality and heterogeneity. *Criminology* **24**, 733-749.

Barbujani, G., Milani, F. (1986). Diversity of some gene-frequencies in European and Asian populations. Effect of longitude. *J. Hum. Evol.* **15**, 61-69.

Demetrius, L., Ziehe, M. (1984). The measurement of Darwinian fitness in human populations. *Proc. of the Royal Soc. of London*, **B 222**, 33-50

Desharmais, R.A. (1986). Natural selection, fitness entropy and the dynamics of coevolution. *Theor. Pop. Biol.* **30**, 309-340.

Fisher, R.A., Corbet, A.S., Williams, C.B. (1943). The relation between the number of species and the number of individuals in a random sample of an animal population. *J. Anim. Ecol.* **12**, 42-58.

Gini, C. (1912). Variabilité e mutabilitá. *Studi Economico-Giuridici Fac. Giurisprudenza Univ. Cagliari*, A. III., Parte II., 3-159.

Grassle, J.F., Patil, G.P., Smith, W.K., Taillie, C. (eds.) (1979). *Ecological Diversity in Theory and Practice*. International Co-operative Publishing House, Fairland, Maryland.

Gregorius, H.R. (1978). The concept of genetic diversity and its formal relationship to heterozygosity and genetic distance. *Math. Biosci.* **41**, 253-271.

Herdan, G. (1957). The mathematical relation between the number of diseases and the number of patients in a community. *J. Royal Stat. Soc.*, Ser. A, **120**, 320-330.

Hill, M.O. (1973). Diversity and evenness: a unifying notation and its consequences. *Ecology* **54**, 427-432.

Höpker, W.W. (1975). Zur Quantifizierung selections-bedingter Gruppenunterschiede in der Epidemiologie. *Meth. Inform. Med.* **14**, 144-149.

Hunter, P.R., Gaston, M.A. (1988). Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. *J. Clin. Microbiol.* **26**, 2465- 2466.

Hunter, P.R., Izsák, J. (1990). Diversity studies of Salmonella incidents in some domestic livestock and their potential relevance as indicators of niche width. *Epidemiol. Infect.* **105**, 501-510.

Hurlbert, S.M. (1971). The nonconcept of species diversity: a critique and alternative parameters. *Ecology* **52**, 577-586.

Hutcheson, K. (1970). A test for comparing diversities based on the Shannon formula. *J. Theor. Biol.* **29**, 151-154.

Innes, D.J., Schwartz, S.S., Hebert, P.D. (1986). Genotypic diversity and variation in mode of reproduction among populations in the Daphnia pulex group. *Heredity* **57**, 345- 355.

Izsák, J. (1986). Measuring the secular changes of the concentration of death causes. *Genus* **42**, 197-208.

Izsák, J. (1988). Secular changes of the concentration of neoplasm death causes in the male population of some countries. *Genus* **44**, 119-130.

Izsák, J. (1989). Numerical experimentation on diversity curves applied in epidemiology. *Biométrie-Praximétrie* **29**, 97-113.

Izsák, J., Juhász-Nagy, P. (1980). Investigation on diversity of Hungarian mortality statistics. *Annales Univ. Sci. Budapestinensis, Sectio Biologica* **22-23**, 35-45.

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

Izsák, J., Juhász-Nagy, P. (1984). Diversity studies on mortality data. *Annales Univ. Sci. Budapestinensis, Sectio Biologia*, **24-26**, 29-46.

Kayser, K., Ramisch, W., von Kenne, H., Höpker, W.W. (1980). Thesaurus and Information. *Med. Inform.* **5**, 155-166.

Keyfitz, N. (1977). *Applied Mathematical Demography*. New York, Wiley.

Kimura, M., Crow, J.F. (1964). The number of alleles that can be maintained in a finite population. *Genetics* **49**, 725-738.

Lewontin, R.C. (1972). The apportionment of human diversity. *Evolutionary Biology* **6**, 381-398.

Lewontin, R.C. (1982). *Human Diversity*. Scientific American Books, Inc. Freeman and Company. Chap. 8.

Lichter, D.T. (1985). Racial concentration and segregation across U.S. Counties, 1950-1980. *Demography*, **22**, 603-609.

Lieberson, S. (1969). Measuring population diversity. *Am. Soc. Rev.* **34**, 850-862.

Magurran, A.E. (1988). *Ecological Diversity and its Measurement*.Croom Helm, London.

Nayak, T.K. (1986). An analysis of diversity using Rao's quadratic entropy. *Sankhya* **48**, Ser. B, 315-330.

O'Brien, S.J., Gail, M.H., Levin, D.L. (1980). Correlative genetic variation in natural populations of cats, mice, and man. *Nature* **288**, 580-583.

Patil, G.P., Taillie, C. (1982). Diversity as a concept and its measurement. *J. Am. Stat. Assoc.* **77**, 548-561.

Pielou, E.C. (1975). *Ecological Diversity*. Wiley, New York.

Rao, C.R. (1982). Diversity and dissimilarity coefficients: a unified approach. *Theor. Popul. Biol.* **21**, 24-43.

Rao, C.R., Boudreau, R. (1984). Diversity and cluster analyses of blood group data on some human populations. In: Chakravarti A. (ed.) *Human Population Genetics: Pittsburgh Symposium*. Van Nostrand Company, N.Y. pp. 331-362.

Robinson, J.V., Sandgren, C.D. (1984). An experimental evaluation of diversity indices as environmental discriminators. *Hydrobiologia* **108**, 187-196.

Sharma, B.D. (1986). Social Entropy. *Man in India* **66**, 101-108.

Silander, J.A. (1979). Microevolution and clone structure in Spartina patens. *Science* **203**, 658-660.

Smith, W., Grassle, F. (1977). Sampling properties of a family of diversity measures. *Biometrics* 33, 283-292.

Stoddart, J.A. (1983). A genotypic diversity measure. *J. of Heredity* 74, 489-490.

Stoddart, J.A., Taylor, J.F. (1988). Genotypic diversity: Estimation and prediction in samples. *Genetics* 118, 705-711.

Taillie, C., Patil, G.P. (1979). Diversity, entropy, and Hardy-Weinberg equilibrium. In: Grassle, J.F., Patil, G.P., Smith, W. and Taillie, C. (eds.) *Ecological Diversity in Theory and Practice*, 1979, Internat. Co-operative Publishing House, pp. 63-72.

Theil, H. (1967). *Economics and Information Theory*. North-Holland Publishing Company, Amsterdam.

Tuljapurkar, S. (1982). Why use population entropy? It determines the rate of convergence. *J. Math. Biol.* 13, 325-337.

Varadarajan, P.R., Ramanujam, V. (1987). Diversification and performance: A reexamination using a new two-dimensional conceptualization of diversity in firms. *Acad. Management Journal* 30, 380-393.

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

White, M.J. (1986). Segregation and diversity measures in population-distribution. *Pop. Index* 52, 198-221.

Williams, C.B. (1964). *Patterns in the Balance of Nature*. Academic Press, London - New York.

Zhang, Q., Allard, R.W. (1986). Sampling variance of the genotypic diversity index. *J. of Heredity* 77, 54-55.

Received 13 March 1991; revised 26 June 1991