

DIVERSITY : ITS MEASUREMENT, DECOMPOSITION, APPORTIONMENT AND ANALYSIS*

By C. RADHAKRISHNA RAO

Indian Statistical Institute and University of Pittsburgh

SUMMARY. Two general methods of obtaining measures of diversity within a population are discussed. One is based on an intrinsic notion of dissimilarity between individuals and the other makes use of the concepts of entropy. Some examples are given of the decomposition of diversity within a population in terms of given or conceptual factors. Methods for apportionment of diversity in a hierarchically classified set of populations are discussed. The concept of analysis of diversity as a generalization of analysis of variance is developed for populations classified by combinations of different levels of chosen factors.

1. INTRODUCTION

There is an extensive literature on measures of diversity within populations and analyses based on them. They have been used in a wide variety of studies in anthropology (Rao, 1948; Mahalanobis, Majumdar and Rao, 1949; Majumdar and Rao, 1958; Rao, 1971a, 1971b, 1977), in genetics (Cavalli-Sforza, 1969; Karlin *et al.*, 1979; Morton and Lalovel, 1973; Nei, 1978; Sanghvi, 1953; Sanghvi and Balakrishnan, 1972), in economics (Gini, 1912; Sen, 1973), in sociology (Agresti and Agresti, 1978; Rao, 1981a) and in biology (Sokal and Sneath, 1963; Pielou, 1975; Patil and Taille, 1979). A complete bibliography of papers on measures of diversity and their applications is compiled by Dennis *et al.* (1979).

A diversity measure can be used to decompose the total diversity within a population as due to a number of factors. Thus we might ask as to how much of the diversity between individuals of a population is due to size and how much due to shape. We refer to such problems as decomposition of diversity or DEDIV for short.

If we have a mixture of several populations, it would be of interest to know how much of the diversity in a composite population is due to diversity within populations and how much due to between populations. We refer to this problem as apportionment of diversity or APDIV for short.

In analysis of variance we partition the variability in a given set of quantitative data into a number of additive components, each of which is

* Paper presented at the Conference on Mathematical Statistics and Probability, Indian Statistical Institute, New Delhi, December 1980.

used to test a certain null hypothesis or to estimate a component of variance. Of particular practical interest is the analysis of data classified by the levels of a number of factors, where the total variability is partitioned as due to main effects and interactions of the factors and their significance tested. The corresponding analysis when the data are categorical (qualitative) in nature is not straight forward, although a variety of techniques are available (see Bishop, Feinberg and Holland, 1975 and Haberman, 1974). Further, even in the case of quantitative data, the usual analysis of variance technique is not applicable when measures of variability other than the sum of the squared deviations from the mean are considered. We introduce a general measure of diversity (variability) applicable to both quantitative and qualitative data and extend the concept of analysis of variance (ANOVA) to the more general analysis of diversity (ANODIV).

We note that such an approach to the analysis of one-way classified categorical data (equivalent to two-dimensional contingency table), called CATANOVA, was developed by Light and Margolin (1971) using the Gini-Simpson index of diversity for a multinomial population. Further contributions in this direction have been made recently by Anderson and Landis (1980). The present paper provides a general framework for extending the analysis of the above authors using a general diversity measure to more complex situations where the data are classified by levels of two or more factors, and the main effects and interactions of these factors have to be examined.

The present study is largely motivated by problems posed by research workers in applied fields concerning the measurement of diversity of individuals in a population and apportionment of diversity as between and within populations (see Lewontin, 1972 and Nei, 1973 for problems in genetics; Sen, 1973 in economics; and Agresti and Agresti, 1978 in sociology). It is hoped that ANODIV (analysis of diversity) as outlined in the present paper will be of use in carrying out a wider analysis of data in more complex situations.

2. MEASURES OF DIVERSITY

2.1 Postulates and notations. We consider a measurable space (Ω, \mathcal{B}) and a convex set \mathcal{P} of probability measures defined on it. A function $H(\cdot)$ mapping \mathcal{P} into the real line is said to be a measure of diversity if it satisfies the following conditions:

$$C_1: H(P) > 0 \quad \forall P \in \mathcal{P} \text{ and } H(P) = 0 \text{ iff } P \text{ is degenerate.}$$

$$C_2: H \text{ is a concave function on } \mathcal{P}.$$

We shall refer to $H(P)$ as the diversity within a population π characterized by the probability measure P .

The condition C_1 is a natural one since a measure of diversity should be nonnegative and take the value zero when all the individuals of a population are identical, i.e., when the associated probability measure is concentrated at a particular point of \mathcal{X} . The condition C_2 is motivated by the consideration that the diversity in a mixture of populations should not be smaller than the average of the diversities within individual populations.

2.2 General theory. We start first by choosing a nonnegative symmetric function $d(X_1, X_2)$ which is a measure of difference between two individuals with $X = X_1$ and $X = X_2$, without any reference to the probability distributions of X_1 and X_2 . The choice of $d(X_1, X_2)$ naturally depends on the nature of the practical problem under investigation. We define the DIV (diversity) of π_i as

$$H(P_i) = H_i = \int d(X_1, X_2) P_i(dX_1) P_i(dX_2) \quad \dots (2.2.1)$$

i.e., as the average difference between two randomly chosen individuals from π_i . Suppose that one individual is drawn from π_i and another from π_j . Then the average difference is

$$H_{ij} = \int d(X_1, X_2) P_i(dX_1) P_j(dX_2). \quad \dots (2.2.2)$$

We expect H_{ij} to be larger than the average of H_i and H_j , in which case the DIS (dissimilarity) between π_i and π_j may be defined by what may be termed as the Jensen difference

$$\begin{aligned} D_{ij} &= H_{ij} - \frac{1}{2}(H_i + H_j) \\ &= 2 \left[H \left(\frac{P_i + P_j}{2} \right) - \frac{1}{2}H(P_i) - \frac{1}{2}H(P_j) \right]. \quad \dots (2.2.3) \end{aligned}$$

The expression (2.2.3) will be non-negative for any i and j iff $d(X_1, X_2)$ is chosen such that the function H defined on \mathcal{P} as in (2.2.1) is concave. This can be easily verified by considering $P_\lambda \in \mathcal{P}$ where

$$P_\lambda = \lambda P_i + (1-\lambda)P_j, \quad 0 < \lambda < 1$$

and computing

$$\begin{aligned} H_\lambda &= \int d(X_1, X_2) P_\lambda(dX_1) P_\lambda(dX_2) \\ &= \lambda^2 H_i + (1-\lambda)^2 H_j + 2\lambda(1-\lambda) H_{ij}. \end{aligned}$$

Then

$$\begin{aligned} H_0 - (\lambda H_1 + (1-\lambda) H_2) \\ = 2\lambda(1-\lambda)(H_{ij} - \frac{1}{2}H_{i1} - \frac{1}{2}H_{1j}) = 2\lambda(1-\lambda) D_{ij}. \end{aligned}$$

The concavity of H ensures that $D_{ij} \geq 0$ and vice-versa.

2.3 *Some examples.* (1) Let $X \in R^m$, a real vector space of m dimension furnished with an inner product $(x, y) = x' Ay$, where A is a positive definite matrix. Define

$$d(X_1, X_2) = (X_1 - X_2, X_1 - X_2).$$

Let $X \sim (\mu_i, \Sigma_i)$ in π_i (i.e., X is distributed with mean vector μ_i and dispersion matrix Σ_i). Then

$$\begin{aligned} H_i &= 2 \operatorname{tr} A \Sigma_i \\ H_{ij} &= \operatorname{tr} A \Sigma_i + \operatorname{tr} A \Sigma_j + \delta'_{ij} A \delta_{ij} \end{aligned}$$

where tr stands for the trace of a matrix and $\delta'_{ij} = \mu_i - \mu_j$. Applying the formula (2.2.3)

$$D_{ij} = \delta'_{ij} A \delta_{ij}. \quad \dots (2.3.1)$$

If $\Sigma_i = \Sigma$ for all i and $A = \Sigma^{-1}$, (2.3.1) becomes the Mahalanobis D^2 between π_i and π_j .

(2) Let $X = (x_1, \dots, x_m)$ where x_i can take only a finite number of values. For instance, x_i may stand for the type of gene allele at a given locus i on a chromosome. In such a case an appropriate measure of difference between two vectors X_1 and X_2 is

$$d(X_1, X_2) = m - \sum \delta_r \quad \dots (2.3.2)$$

where $\delta_r = 1$ if the r -th components of X_1 and X_2 agree and zero otherwise. Let x_r take k_r different values with probabilities

$$p_{ir_j}, \quad j = 1, \dots, k_r$$

in population π_i . Define

$$j_{ir}^{(i)} = E(\delta_r) = \sum_{s=1}^{k_r} p_{ir_s}^2$$

when X_1, X_2 are independently drawn from π_i and

$$j_{ij}^{(i)} = E(\delta_r) = \sum_{s=1}^{k_r} p_{ir_s} p_{jr_s}$$

when X_1 is drawn from π_i and X_2 from π_j . Then

$$H_i = \sum_{r=1}^m (1-j_{ir}^2) = m(1-J_{ii})$$

$$H_{ij} = \sum_{r=1}^m (1-j_{ijr}^2) = m(1-J_{ij}) \quad \dots (2.3.3)$$

$$\begin{aligned} D_{ij} &= H_{ij} - \frac{1}{2}(H_i + H_j) \\ &= m[\frac{1}{2}(J_{ii} + J_{jj}) - J_{ij}] \\ &= \frac{1}{2} \sum_{r=1}^m \sum_{s=1}^m (p_{ir} - p_{jr})^2. \quad \dots (2.3.4) \end{aligned}$$

The expression (2.3.4) without the factor m has been called by Nei (1978) as "a minimum estimate of the net codon difference per locus" and used by him and his colleagues (see the list of references in Nei, 1978) as a measure of genetic distance in phylogenetic studies.

Note 1: When $m = 1$, we have a single multinomial and the expression for H_i in (2.3.3) reduces to the Gini-Simpson index

$$1 - \sum_{i=1}^k p_i^2 \quad \dots (2.3.5)$$

where p_1, \dots, p_k are the cell probabilities. [This measure was introduced by Gini, 1912 and by Simpson, 1949 in biological work]. The properties of (2.3.5) have been studied by various authors (Bhargava and Doyle, 1974; Bhargava and Uppuluri, 1975; Agresti and Agresti, 1978).

Note 2: It is seen that H_i as defined in (2.3.3) depends only on the marginal distributions of x_i , $i = 1, \dots, m$ and is additive with respect to the characters examined. These properties arise from the way the difference function (2.3.2) is defined. The DIS (2.3.4) is specially useful in evolutionary studies as suggested by Nei (1978).

Note 3: We may consider the joint distribution of (x_1, \dots, x_m) as a combined multinomial with $k = k_1 \times \dots \times k_m$ classes and apply the formula (2.2.1) to measure diversity. In such a case the difference between two individuals takes the value 1 when all the components x_i agree and the value zero if at least one is different. This leads to an expression different from (2.3.4), as the basic function for assessing differences between individuals

is not the same. When x_1, \dots, x_m are independently distributed, an explicit expression for the DIV based on the combined multinomial reduces to

$$H = 1 - [1 - H(1)] \dots [1 - H(m)]$$

where $H(r)$ is the DIV based on x_r , the r -th character only. It may be noted that the expression for DIV given in (2.3.3) is $H = \Sigma H(r)$ whether x_i are independently distributed or not.

Note 4: If we consider the component x_r in the vector X of *Note 2* as the genotype of a diploid organism as determined by a pair of alleles at locus r and define δ_r in (2.3.2) as 1 if two individual organisms have the same value for x_r and zero otherwise, then we get the DIV and DIS based on the measure of genotypic identity devised by Hedrick (1971) and applied by Mitton (1977) in genetic studies. The expressions are the same as in (2.3.3) and (2.3.4) with p_{ir} interpreted as genotypic frequencies instead of gene frequencies.

Note 5: If we consider x_r as a genotype as in *Note 4*, but define the difference as $2m - \Sigma \delta_r$ where δ_r is the number of genes common to the individuals at the r -th pair of loci (which may be two, one or zero) then we obtain the DIV and DIS based on a measure introduced by Lattor (1973, 1980).

(3) The Gini-Simpson index $H(p)$ defined in (2.3.5) can be interpreted as the expected distance between two individuals drawn at random from a multinomial distribution when the distance is defined to be zero if the individuals belong to the same category and unity otherwise (see Rao, 1981a). But in practice it may be more meaningful, in some situations as in sociological problems, to assign different distance values to different pairs of categories, say $d_{ij} > 0$ between the categories i and j . Introducing d_{ii} for the distance of the category i with itself, which is usually zero but may be taken as different from zero in some problems, the expected distance is

$$H_{\Delta}(p) = \Sigma \Sigma d_{ij} p_i p_j = p' \Delta p \quad \dots (2.3.6)$$

where Δ is the matrix (d_{ij}) . For the Gini-Simpson index, the diagonal entries are zero and the rest are unities. For $H_{\Delta}(\cdot)$ to qualify as a diversity measure it should satisfy some properties.

- (i) It should be zero (minimum) when all the individuals belong to the same category.
- (ii) It should be a concave function on the set $S = \{p : p_i \geq 0, \Sigma p_i = 1\}$, so that the diversity in a mixture of distributions is not smaller than the average diversity of the individual distributions constituting the mixture.

These conditions place a restriction on the choice of the elements d_{ij} such as

$$d_{11} = \dots = d_{kk} \quad \dots \quad (2.3.7)$$

and the $(k-1) \times (k-1)$ matrix

$$(d_{ik} + d_{jk} - d_{ij} - d_{kk}), \quad i, j = 1, \dots, k-1, \quad \dots \quad (2.3.8)$$

is nonnegative definite.

We call $H_{\Delta}(p)$ with Δ chosen subject to the conditions (2.3.7) and (2.3.8) as the generalized entropy function of degree 2.

The choice of d_{ij} is not a statistical problem and will depend on an individual's assessment of differences between qualitative categories with reference to a given problem. However, one can use methods of multi-dimensional scaling in estimating d_{ij} by using *supplementary information* such as inequality relationships between d_{ij} and d_{rs} for different combinations (i, j) and (r, s) . The use of such estimates may be more appropriate than taking $d_{ii} = 0$ and $d_{ij} = 1$ in the analysis of qualitative variation using methods such as those discussed in Section 5 of this paper.

2.4 Similarity (SIM) measures. Instead of a difference measure between two individuals, it may be natural to consider a similarity function $s(X_1, X_2)$ and define S_i, S_j and S_{ij} by taking expectations analogous to H_i, H_j and H_{ij} . Then the DIV of π_i may be defined by a suitable decreasing function of S_i , such as $1 - S_i$ or $-\log S_i$, specially when the range of S_i is $(0, 1)$. The DIS obtained by choosing $H_i = 1 - S_i$ is

$$D_{ij} = \frac{1}{2}(S_i + S_j) - S_{ij} \quad \dots \quad (2.4.1)$$

and that by choosing $H_i = -\log S_i$ is

$$\begin{aligned} D_{ij} &= \frac{1}{2}(\log S_i + \log S_j) - \log S_{ij} \\ &= -\log \frac{S_{ij}}{\sqrt{S_i S_j}}. \end{aligned} \quad \dots \quad (2.4.2)$$

For instance, in the second example of Section 2.3, a natural definition of $s(X_1, X_2) = (\Sigma k_r)/m$, which lies in the range $(0, 1)$. Then

$$S_i = J_{ii}, \quad S_j = J_{jj}, \quad S_{ij} = J_{ij} \quad \dots \quad (2.4.3)$$

where J_{ij} are as defined in (2.3.3). Using (2.4.1) and (2.4.2), we have the alternative forms

$$D_{ij} = \frac{1}{2}(J_{ii} + J_{jj}) - J_{ij}, \quad \dots \quad (2.4.4)$$

$$D_{ij} = -\log \frac{J_{ij}}{\sqrt{J_{ii} J_{jj}}}. \quad \dots \quad (2.4.5)$$

The expression (2.4.4) is the same as the "minimum genetic distance" (2.3.4) of Nei (1978), and (2.4.5) is what he calls the "standard genetic distance".

Again, in the example (2), we may define the similarity function as $(\delta_1 \dots \delta_m)^{1/m}$ instead of $(\delta_1 + \dots + \delta_m)/m$. The new function has the value unity when the gene alleles coincide at all the loci and zero otherwise. In such a case, when the characters are independent,

$$\begin{aligned} S_i &= j_i^{(1)} \dots j_i^{(m)} = (J'_i)^m \\ S_{ij} &= j_{ij}^{(1)} \dots j_{ij}^{(m)} = (J'_{ij})^m \end{aligned} \quad \dots (2.4.6)$$

where j'_{ij} are as defined in example (2) of Section 2.3. Taking logarithms of (2.4.6), the corresponding DIS is

$$D_{ij} = -\log \frac{J'_{ij}}{\sqrt{J'_i J'_j}} \quad \dots (2.4.7)$$

which Nei calls the "maximum genetic distance".

2.5 Measures of entropy. A wide variety of DIV measures have been introduced through the concept of entropy and information. The general approach in these cases is basically different from that of Section 2.2 where a function $d(X_1, X_2)$ measuring the difference between individuals X_1 and X_2 is chosen first and probability distributions of X_1 and X_2 are used only to find the average of $d(X_1, X_2)$. In practice, $d(X_1, X_2)$ would be chosen to reflect some intrinsic difference between individuals relevant to a particular investigation. On the other hand, a measure of entropy is directly conceived of as a function defined on the space of distribution functions, satisfying some postulates. Some of the postulates are that it is nonnegative, attains the maximum for the uniform distribution and has the minimum when the distribution is degenerate. Thus a measure of entropy is an index of similarity of a distribution function with the uniform distribution, and hence a measure of DIV.

We shall consider the space of all multinomial distributions for simplicity of presentation of results, observing that the formulae for the continuous case can be obtained by replacing the summation by the integral sign. We represent the probabilities in the k cells of a general multinomial by p_1, \dots, p_k and for a particular population π_i by p_{i1}, \dots, p_{ik} . Mathai and Rathio (1975) mention three general forms for entropy :

$$H = (1-\alpha)^{-1} \log(\sum p_r^{\alpha+\beta-1} / \sum p_r^\beta) \quad \dots (2.5.1)$$

$$H = [(\sum p_r^{\alpha+\beta-1} / \sum p_r^\beta) - 1] \div (2^{1-\alpha} - 1) \quad \dots (2.5.2)$$

$$H = -\sum p_r^\beta \log p_r / \sum p_r^\beta \quad \dots (2.5.3)$$

where all the summations are taken from 1 to k . When $\beta_r = 1$ for all r , (2.5.1), (2.5.2) and (2.5.3) are respectively the expressions introduced by Renyi (1961), Havrda and Charvát (1967) and Shannon (1948).

With $\beta_r = 1$ for all r , all the functions (2.5.1) – (2.5.3) are nonnegative, attain the maximum when p_i are equal (maximum diversity) and are zero when $p_i = 1, p_j = 0, j \neq i$ (minimum diversity). Mathai and Rathie (1975) discuss the various additional mathematical postulates which lead to these functions. Patil and Taille (1979) and Pielou (1975) provide interpretations of some of these functions in the context of ecological studies.

It is easy to show that when $\beta_r = 1$ for all r , the diversity measures (2.5.2) and (2.5.3) are concave but (2.5.1) is not so in general. This appears to be a drawback of Renyi's entropy as a measure of diversity.

3. DECOMPOSITION OF DIVERSITY (DEDIV)

3.1 *General formula.* In some problems it may be of some interest to decompose the total diversity within a population into a number of components each of which is attributable to a given or a conceptual factor. Such a decomposition can be done in a natural way if H is defined as in Section 2.2, by considering a decomposition of the basic difference function

$$d(X_1, X_2) = d_1(X_2, X_2) + \dots + d_c(X_1, X_2).$$

Taking expectations, we obtain the decomposition of diversity in a population $P_i \in P$ as

$$H(P_i) = H_i = H_i^{(1)} + \dots + H_i^{(c)} \quad \dots (3.1.1)$$

where $H_i^{(r)} = H^{(r)}(P_i) = E(d_r(X_1, X_2) | P_i)$. We obtain the corresponding decomposition of DIS between populations P_i and P_j in the form

$$D_{ij} = D_{ij}^{(1)} + \dots + D_{ij}^{(c)} \quad \dots (3.1.2)$$

where

$$2^{-1}D_{ij}^{(r)} = H^{(r)} \left(\frac{P_i + P_j}{2} \right) - \frac{1}{2}H^{(r)}(P_i) - \frac{1}{2}H^{(r)}(P_j).$$

3.2 *Some examples.* (1) Let x be an m -vector random variable with $E(x) = \mu_i$ and $E(x - \mu_i)(x - \mu_i)' = \Sigma$ in the population π_i . Let $\theta_1 \geq \dots \geq \theta_m$ be the eigen values and L_1, \dots, L_m be the corresponding eigen vectors of Σ . If we choose

$$d(X_1, X_2) = (X_1 - X_2)'(X_1 - X_2) \quad \dots (3.2.1)$$

i.e., the square of the Euclidean distance in R^m , then

$$d(X_1, X_2) = [L'_1(X_1 - \bar{X}_1)]^2 + \dots + [L'_m(X_1 - \bar{X}_1)]^2 \quad \dots \quad (3.2.2)$$

which gives the decomposition of H_t for π_t as

$$H_t = 2 \operatorname{tr} \Sigma = 2\theta_1 + \dots + 2\theta_m. \quad \dots \quad (3.2.3)$$

The expression (3.2.3) is the familiar hierarchical decomposition of the total variability in terms of principal components. The corresponding decomposition of DIS between π_t and π_j is

$$D_{ij} = \delta'_{ij} \delta_{ij} = (L'_1 \delta_{ij})^2 + \dots + (L'_m \delta_{ij})^2$$

where $\delta_{ij} = \mu_i - \mu_j$, the difference in the mean vectors of π_t and π_j .

(2) Let $(X_1 - \bar{X}_1) = (d_1, \dots, d_m)$ with d_i representing the i -th component of $X_1 - \bar{X}_1$. Then the expression (3.2.1) is Σd_i^2 . Penrose (1954) considered the decomposition

$$\Sigma d_i^2 = m\bar{d}^2 + \Sigma(d_i - \bar{d})^2 \quad \dots \quad (3.2.4)$$

where $m\bar{d} = \Sigma d_i$. If the individual measurements are standardized, i.e., $E(d_i^2) = 2$ for all i , then the diversity in a population π_t is

$$H_t = E(\Sigma d_i^2) = E(m\bar{d}^2) + E[\Sigma(d_i - \bar{d})^2]$$

$$2m = 2\{(1 + m^{-1} \bar{\rho})\} + 2\{(m-1)(1 - \bar{\rho})\} \quad \dots \quad (3.2.5)$$

where $\bar{\rho}$ is the average value of all the correlations. The first term of (3.2.5) represents the diversity due to "size" and the second due to "shape" as defined by Penrose.

Spielman (1973) considers a decomposition of the type

$$\Sigma d_i^2 = [\sqrt{\Sigma \bar{X}_{1i}^2} - \sqrt{\Sigma \bar{X}_{2i}^2}]^2 + 2[\sqrt{\Sigma \bar{X}_{1i}^2} \sqrt{\Sigma \bar{X}_{2i}^2} - \Sigma X_{1i} X_{2i}] \quad \dots \quad (3.2.6)$$

where X_{1i} and X_{2i} are the components of X_1 and X_2 respectively. Taking expectations of both sides of (3.2.6), we obtain a decomposition of the diversity due to "size" and "shape" as defined by Spielman. The expectations of the expressions in (3.2.6) depend on the actual distributions of the variables. If X_1 and X_2 have the multivariate normal distribution, then the expectations can be computed from the noncentral chi-square distribution.

Instead of the principal components as in (3.2.2), we may choose any set of orthonormal vectors M_1, \dots, M_m and consider the decomposition

$$(X_1 - X_2)'(X_1 - X_2) = [M'_1(X_1 - X_2)]^2 + \dots + [M'_m(X_1 - X_2)]^2 \quad \dots \quad (3.2.7)$$

and obtain the corresponding decomposition of diversity by taking expectations. We can choose the vectors M_1, \dots, M_m to represent conceptual factors such as shape and size as shown in Rao (1962, 1971b). For instance, a choice of M_1 such as

$$M_1 = (\sigma' \Sigma^{-1} \sigma)^{-1} \Sigma^{-1} \sigma,$$

where Σ is the variance-covariance matrix of X and σ is the vector of standard deviations of the components of X (i.e., the square roots of the diagonal elements of Σ), represents the size factor. Then taking expectations on both sides of (3.2.7)

$$\begin{aligned} E[(X_1 - X_2)'(X_1 - X_2)] \\ &= E[M_1'(X_1 - X_2)]^2 + E\{[M_2'(X_1 - X_2)]^2\} + \dots + \{M_m'(X_1 - X_2)\}^2 \\ 2 \operatorname{tr} \Sigma &= 2 \frac{\sigma' \Sigma^{-1} \sigma}{\sigma' \Sigma^{-1} \sigma} + H_{1h} = H_{1s} + H_{1h} \quad \dots \quad (3.2.8) \end{aligned}$$

we have a decomposition of the total diversity as due to size (H_{1s}) and shape (H_{1h}).

(3) Decomposition of diversity in a population where the individuals are classified by categories of a number of attributes has been considered by ecologists through measures of entropy (see Patil and Taille, 1979 and Pielou, 1975). For instance, a bird's nest may be classified by the species to which the bird belongs and the level of height at which it builds its nest. If p_{ij} denotes the probability that a nest belongs to a bird of species i and is at level j of height, then the diversity with respect to both the attributes, species (S) and height (L), is

$$H(SL) = H(\{p_{ij}\}, i = 1, 2, \dots; j = 1, 2, \dots)$$

choosing some measure of entropy defined over all multinomial distributions. The diversity due to species alone is

$$H(S) = H(\{p_{i.}\}, i = 1, 2, \dots)$$

where $p_{i.} = \sum_j p_{ij}$, $i = 1, 2, \dots$. We may formally consider the decomposition

$$\begin{aligned} H(SL) &= H(S) + [H(SL) - H(S)] \\ &= H(S) + H_S(L) \end{aligned}$$

and define $H_S(L)$ as the diversity due to L given S .

A natural way of defining conditional diversity is

$$H(L|S) = \sum w_i H \left\{ \left\{ \frac{p_{ij}}{p_i} \right\}, j = 1, 2, \dots \right\}$$

where w_i is a suitable weight attached to category i of attribute S . If H is Shannon's entropy [formula (2.5.3) with $\beta_r = 1$], then

$$H_S(L) = H(L|S) \text{ with } w_i = p_i.$$

If H is the entropy of Havrda and Charvát [formula (2.5.2) with $\beta_r = 1$] then

$$H_S(L) = H(L|S) \text{ with } w_i = p_i^{\alpha}.$$

There is no such simple relation in the case of Renyi's entropy [formula (2.5.1) with $\beta_r = 1$]. In this case

$$H_S(L) = (1-\alpha)^{-1} \log \left[\sum_i \frac{p_i^{\alpha}}{\sum_i p_i^{\alpha}} \sum_j \frac{p_{ij}^{\alpha}}{p_i^{\alpha}} \right]$$

whereas

$$H(L|S) = (1-\alpha)^{-1} \sum w_i \log \sum_j \frac{p_{ij}^{\alpha}}{p_i^{\alpha}}.$$

From the point of view of decomposition of diversity in the form considered by ecologists, the use of Shannon's entropy seems to be a natural one.

4. APPORTIONMENT OF DIVERSITY (APDIV)

4.1 *General formula.* If P_1, \dots, P_k are the distributions of X in populations π_1, \dots, π_k with a priori probabilities $\lambda_1, \dots, \lambda_k$, then the distribution in the mixture π_0 is $P_0 = \lambda_1 P_1 + \dots + \lambda_k P_k$. If H is a DIV measure defined on \mathcal{P} and is concave, then we can write

$$\begin{aligned} H(P_0) &= \sum \lambda_i H(P_i) + J \{(\lambda_i); \{P_i\}\} \\ &= W + B \end{aligned} \quad \dots (4.1.1)$$

as the sum of two nonnegative components. The first term W is the average diversity within the individual populations and the second term B , which is called the Jensen difference or mutual information using the terminology of *Information Theory*, is the diversity between the populations. The ratio

$$G = \frac{B}{W+B} \quad \dots (4.1.2)$$

is called the index of diversity between populations in genetic studies (see Lewontin, 1972; Nei, 1973 and Rao, 1980).

More generally, let us consider a number of populations grouped in a hierarchical classification such as populations within regions and regions within species and so on. If the distributions within populations and their a priori probabilities are known, then the distributions at any level of classification such as regions and species and the associated a priori probabilities can be computed. This would enable us to compute the average diversities H_P , H_R and H_S within populations, within regions and within species respectively and H_0 the diversity in the mixture of all the populations put together. Then we have the apportionment of diversity as in Table 1.

TABLE 1. APDIV IN A HIERARCHICAL CLASSIFICATION

due to	diversity	ratio
within populations	H_P	
between populations (within regions)	$H_R - H_P$	$(H_R - H_P)/H_R$
between regions (within species)	$H_S - H_R$	$(H_S - H_R)/H_S$
between species	$H_0 - H_S$	$(H_0 - H_S)/H_0$
total	H_0	

4.2 *Examples.* (1) It is interesting to note that if H is defined as in (2.2.1) using a basic difference function $d(X_1, X_2)$ then the formula (4.1.1) can be written as

$$H(P_0) = \sum \lambda_i H(P_i) + \sum \Sigma \lambda_i \lambda_j D_{ij}$$

where

$$D_{ij} = 2H\left(\frac{P_i + P_j}{2}\right) - H(P_i) - H(P_j).$$

is the DIS between π_i and π_j .

Let us consider k populations as in example (1) of Section 2.3 where in π_i , the m -vector variable $X \sim (\mu_i, \Sigma)$ and choose $d(X_1, X_2)$ as the Mahalanobis D^2 [formula (2.3.1) with $A = \Sigma^{-1}$]. Further let π_0 be a mixture of π_1, \dots, π_k with a priori probabilities $\lambda_1, \dots, \lambda_k$. Then, the decomposition (4.1.1) becomes

$$\begin{aligned} H_0 &= H(w) + D(b) \\ &= 2m + \Sigma \Sigma \lambda_i \lambda_j \delta'_{ij} \Sigma^{-1} \delta_{ij} \\ &= 2m(1 + V) \end{aligned} \quad \dots (4.2.1)$$

where $\delta_{ij} = \mu_i - \mu_j$. Thus the diversity within populations is $2m$ and the ratio $C(b)$ of (4.1.2) is V which is the weighted combination of Mahalanobis D^2 's for all pairs of populations. The author has suggested (see Mahalanobis, Majumdar and Rao, 1949) the use of an estimate of V in the selection of variables to maximize dissimilarity between populations.

Let us consider example (2) of Section 2.3 and denote by π_0 , the mixture of π_1, \dots, π_k with apriori probabilities $\lambda_1, \dots, \lambda_k$. In this case (4.1.1) becomes

$$H_0 = m[\Sigma \lambda_i(1 - J_{ii}) + \Sigma \Sigma \lambda_i \lambda_j (\frac{1}{2}J_{ii} + \frac{1}{2}J_{jj} - J_{ij})] \quad \dots (4.2.2)$$

which is the decomposition obtained by Nei (1973) and Chakraborty (1974). The ratio $G(b)$ defined in (4.1.2) is

$$G(b) = \frac{\Sigma \Sigma \lambda_i \lambda_j (\frac{1}{2}J_{ii} + \frac{1}{2}J_{jj} - J_{ij})}{1 - \Sigma \Sigma \lambda_i \lambda_j J_{ij}} \quad \dots (4.2.3)$$

The ratio (4.2.3) obtained by considering only the two populations π_i and π_j with equal prior probabilities

$$\theta_{ij} = \frac{J_{ii} + J_{jj} - 2J_{ij}}{4 - J_{ii} - J_{jj} - 2J_{ij}} \quad \dots (4.2.4)$$

is the hybridity coefficient of Morton (1973) who used it as a DIS between π_i and π_j in phylogenetic studies.

(2) Let p_{i1}, \dots, p_{ir} be the probabilities of r attributes in population π_i with apriori probability λ_i , $i = 1, \dots, k$. If we use Shannon's entropy, then (4.1.1) becomes

$$\begin{aligned} & - \sum_{j=1}^r (\lambda_1 p_{1j} + \dots + \lambda_k p_{kj}) \log (\lambda_1 p_{1j} + \dots + \lambda_k p_{kj}) \\ & = - \sum_{i=1}^k \lambda_i \sum_{j=1}^r p_{ij} \log p_{ij} \\ & \quad + \sum_{i=1}^k \lambda_i \sum_{j=1}^r p_{ij} \log \frac{p_{ij}}{\sum_i \lambda_i p_{ij}}. \end{aligned}$$

If we use the entropy of Havrda and Charvát, then (4.1.1) becomes (apart from the multiplier $1 - 2^{1-a}$)

$$\begin{aligned} & 1 - \sum_{j=1}^r (\lambda_1 p_{1j} + \dots + \lambda_k p_{kj})^a \\ & = \sum_{i=1}^k \lambda_i \left(1 - \sum_{j=1}^r p_{ij}^a \right) \\ & \quad + \sum_{j=1}^r [\lambda_1 p_{1j}^a + \dots + \lambda_k p_{kj}^a - (\lambda_1 p_{1j} + \dots + \lambda_k p_{kj})^a]. \end{aligned}$$

Renyi's entropy cannot be used in APDIV since the concavity condition does not hold (see Rao, 1980).

5. ANALYSIS OF DIVERSITY (ANODIV)

5.1. *General theory.* Apportionment of diversity considered in Section 4 corresponds to analysis of variance of one-way classified data, where the populations are identified by the levels of a single factor, with P_i as the probability measure of the population π_i . Let us now consider two factors A_1 and A_2 and represent the population specified by the i -th level of A_1 and the j -th level of A_2 by π_{ij} and its associated probability measure by P_{ij} , $i = 1, \dots, p$ and $j = 1, \dots, g$. Further let $\lambda_i^{(1)}, \dots, \lambda_p^{(1)}$ be the prior probabilities for the levels of A_1 and $\lambda_j^{(2)}, \dots, \lambda_g^{(2)}$ be independent prior probabilities for the levels of A_2 so that the prior probability associated with P_{ij} is $\lambda_i^{(1)} \lambda_j^{(2)}$. Define

$$\begin{aligned} P_{..} &= \sum_i \lambda_i^{(1)} \lambda_j^{(2)} P_{ij} \\ P_{i.} &= \sum_j \lambda_j^{(2)} P_{ij}, \quad P_{.j} = \sum_i \lambda_i^{(1)} P_{ij} \quad \dots \quad (5.1.1) \end{aligned}$$

where $P_{..}$ is the overall distribution and $P_{i.}$ and $P_{.j}$ are the marginal distributions for the levels of the factors A_1 and A_2 respectively. Consider the decomposition of $H(P_{..})$ as indicated in Table 2.

TABLE 2. ANODIV OF TWO-WAY DATA

due to	diversity
factor (A_1)	$J((\lambda_i^{(1)}); \{P_{i.}\})$
factor (A_2)	$J((\lambda_j^{(2)}); \{P_{.j}\})$
interaction (A_1, A_2)	$I((\lambda_i^{(1)}), (\lambda_j^{(2)}); \{P_{ij}\})$
within populations	$\sum \lambda_i^{(1)} \lambda_j^{(2)} H(P_{ij})$
total	$H(P_{..})$

The J functions in Table 2 are as defined in (4.1.1) with respect to the appropriate marginal distributions. The I function, obtained by subtracting the rest of the values from the total $H(P_{..})$, is in the nature of diversity due to *interaction* of the two factors A_1 and A_2 . We need I to be nonnegative for a proper interpretation of the interaction component. While the concavity of H ensures that the J function is nonnegative, the nonnegativity of I may require some further conditions on the choice of the diversity function H .

As in the case of analysis of variance we can consider three-way or higher order classified data and define higher order interactions in a recursive way. Thus, the three factor interaction (A_1, A_2, A_3) can be computed by the formula

$$\begin{aligned} I(\{\lambda_i^{(1)}\}, \{\lambda_j^{(2)}\}, \{\lambda_k^{(3)}\}; \{P_{ijk}\}) \\ = II(P_{...}) - \sum \sum \sum \lambda_i^{(1)} \lambda_j^{(2)} \lambda_k^{(3)} II(P_{ijk}) \\ - (A_1) - (A_2) - (A_3) - (A_1, A_2) - (A_1, A_3) - (A_2, A_3) \end{aligned} \quad \dots (5.1.2)$$

where $(A_1), \dots, (A_2, A_3), \dots$ are obtained as in Table 1 by considering the appropriate two dimensional marginal distributions. Similarly higher order interactions could be computed.

We shall investigate the choice of diversity functions which make all interactions nonnegative when the apriori probabilities associated with the different factors are independent.

5.2. *Choice of the diversity measure.* We consider the diversity coefficient

$$II(P) = \int d(X_1, X_2) P(dX_1) P(dX_2) \quad \dots (5.2.1)$$

and show that if $II(P)$ is a concave function on P , then all interactions as defined in Table 2 and formula (5.1.2) are nonnegative.

For instance, the interaction term in Table 2 is

$$\begin{aligned} I(\{\lambda_i^{(1)}\}, \{\lambda_j^{(2)}\}; \{P_{ij}\}) \\ = \sum \lambda_i^{(1)} II(P_{i..}) + \sum \lambda_j^{(2)} II(P_{.j}) - \sum \sum \lambda_i^{(1)} \lambda_j^{(2)} II(P_{ij}) - II(P_{..}) \\ = \sum \lambda_i^{(1)} J(\{\lambda_j^{(2)}\}; \{P_{ij}\}) - J(\{\lambda_i^{(1)}\}; P_{.j}). \end{aligned} \quad \dots (5.2.2)$$

It is a straight forward computation to show that (5.2.2) is equal to

$$\begin{aligned} -\sum \sum \sum \lambda_i^{(1)} \lambda_j^{(2)} \lambda_k^{(2)} \\ \times \int d(X_1, X_2) P_{(i-j)(r-s)}(dX_1) P_{(i-j)(r-s)}(dX_2) \end{aligned} \quad \dots (5.2.3)$$

where $P_{(i-j)(r-s)} = P_{ir} + P_{js} - P_{is} - P_{jr}$. Note that the term under the integral sign in (5.2.3) is proportional to

$$\begin{aligned} II\left(\frac{P_{ir} + P_{js} + P_{is} + P_{jr}}{4}\right) - \frac{1}{2} \\ II\left(\frac{P_{ir} + P_{js}}{2}\right) - \frac{1}{2} \\ II\left(\frac{P_{jr} + P_{is}}{2}\right) \end{aligned}$$

which is non-negative if $II(P)$ defined in (5.2.1) is concave over P , which proves the desired result.

Similarly it can be shown that all higher order interactions are non-negative.

5.3 *Sampling theory.* Up to now, we have discussed ANODIV in a mixture of populations in terms of various components assuming that the individual population distributions are known. In practice we have only observations from different populations, in which case we can only estimate the various components of diversity and, if necessary, test hypotheses concerning them. We shall briefly describe how the appropriate methodology could be developed for this purpose.

To indicate how ANODIV provides a unified approach to the analysis of different types of data, let us consider the familiar analysis of variance of one-way classified quantitative data as in Table 3.

TABLE 3. POPULATIONS AND OBSERVATIONS

1	2	...	k
x_{11}	x_{21}	...	x_{k1}
\vdots	\vdots		\vdots
x_{1n_1}	x_{2n_2}	...	x_{kn_k}

We estimate the probability distribution function F_t for the i -th population by the empirical distribution function \hat{F}_t based on the observations x_{i1}, \dots, x_{in_i} . Let us choose $n_1/n, \dots, n_k/n$, ($\sum n_i = n$), as the prior probabilities ($\lambda_1, \dots, \lambda_k$) used in the theoretical development of Sections 4.1 and 5.1. Further let us consider the diversity measure

$$H(F) = \int (X_1 - X_2)^2 F(dX_1) F(dX_2). \quad \dots (5.3.1)$$

Substituting the estimates \hat{F}_t for F_t and $\lambda_t = (n_t/n)$ in the basic decomposition formula (4.1.1)

$$H(\sum \lambda_t F_t) = \sum \lambda_t H(F_t) + J(\{\lambda_t\}; \{F_t\}) \quad \dots (5.3.2)$$

we have

$$H\left(\sum \frac{n_t}{n} \hat{F}_t\right) = \sum \frac{n_t}{n} H(\hat{F}_t) + J\left(\left\{\frac{n_t}{n}\right\}, \{\hat{F}_t\}\right). \quad \dots (5.3.3)$$

Computing the various expressions in (5.3.2), using (5.3.1) for the H function, we obtain (in the standard notation)

$$\frac{1}{n} \sum \sum (x_{ij} - \bar{x}_{..})^2 = \sum \frac{n_t}{n} \frac{1}{n_t} \sum_f (x_{if} - \bar{x}_{t.})^2 + \frac{1}{n} \sum_t n_t (\bar{x}_{t.} - \bar{x}_{..})^2 \quad \dots (5.3.4)$$

which is the usual analysis of sum of squares as *within* and *between* populations. The decomposition (5.3.4) is relevant for tests of significance. But

if the problem is one of estimating the variance between populations defined in terms of given prior probabilities $\lambda_1, \dots, \lambda_k$, then the formula (5.3.2) substituting \hat{F}_i for F_i only may have to be used.

We can, in an analogous manner, derive the analysis of variance for the general m -way classified data when there are equal number of observations in cells or when the number in each cell can be expressed as a product of numbers specific to the levels of each factor.

The ANODIV for one-way classified categorical data (two way contingency table) using the Gini-Simpson index is already illustrated in the paper by Light and Margolin (1971). As mentioned earlier, it may be more appropriate in some practical problems to use a more general diversity index of the form $p'\Delta p$ as defined in (2.3.6) for a multinomial distribution $p' = (p_1, \dots, p_k)$ choosing an appropriate distance matrix Δ . Consider the following contingency table giving the responses of people to four alternatives in an opinion survey.

TABLE 4. FREQUENCIES OF DIFFERENT OPINIONS
IN k OBSERVED POPULATIONS

opinions	populations				total
	1	2	...	k	
A_1	n_{11}	n_{12}	...	n_{1k}	$n_{1.}$
A_2	n_{21}	n_{22}	...	n_{2k}	$n_{2.}$
A_3	n_{31}	n_{32}	...	n_{3k}	$n_{3.}$
A_4	n_{41}	n_{42}	...	n_{4k}	$n_{4.}$
total	$n_{.1}$	$n_{.2}$...	$n_{.k}$	$n_{..}$

In data of this kind it is more appropriate to assign different distances to different pairs of alternatives. We shall illustrate the ANODIV using the diversity measure $p'\Delta p$ (see Rao, 1981b for a numerical example). The estimated multinomial distribution for the j -th population is

$$p_{ij} = n_{ij}/n_{.j}, \quad i = 1, \dots, 4$$

and that for the mixture of all populations choosing $\lambda_i = n_{.i}/n_{..}$ is

$$p_i = n_{.i}/n_{..}, \quad i = 1, \dots, 4.$$

Using the basic decomposition formula (4.1.1) with the estimated distributions and choosing $\lambda_i = n_{.i}/n_{..}$ we have the following analysis of the total diversity

$$\sum \sum d_{ij} p_i \cdot p_j = \sum \frac{n_{.i}}{n_{..}} \sum \sum d_{r\ell} p_{r1} p_{\ell 1} + B. \quad \dots (5.3.5)$$

The first term on the right hand side of (5.3.5) is the diversity "within populations" and the second term B , due to "between populations", has the explicit representation

$$B = \Sigma \Sigma \frac{n_{.i}}{n..} \frac{n_{.j}}{n..} \Sigma \Sigma d_{rs} (p_{ri} - p_{rj}) (p_{ri} - p_{rj}). \quad \dots (5.3.6)$$

Of course, we could have carried out the analysis using any appropriately chosen set of prior probabilities instead of the ratios determined by the sample sizes for the various populations. The large sample distributions of the expressions in (5.3.5) can be obtained on standard lines (see Agresti and Agresti, 1978; Bhargava and Uppuluri, 1977; Light and Margolin, 1971 and Chapter 6 of Rao, 1973).

We can extend the ANODIV to two-way classified categorical data (three way contingency table) using the formulae developed in Section 5.1 provided the number of observations in each cell is the same or can be expressed as the product of two numbers specific to the levels of the two factors associated with the cell and the total sample size. Let us represent the observed numbers for k different categories in the (i, j) -th cell by n_{ijr} , $r = 1, \dots, k$ and the estimated probabilities by $p_{ijr} = (n_{ijr}/n_{ij})$. If the cell numbers n_{ij} satisfy the conditions stated above, then we can obtain the ANODIV as shown in Table 4 where $p_{i..}$, $p_{.jr}$, $p_{..r}$ are defined in the usual way from the appropriate totals and $\lambda_i^{(1)} = (n_{i..}/n..)$, $\lambda_j^{(2)} = (n_{.j}/n..)$.

TABLE 5. ANODIV : TWO-WAY DATA

due to	diversity
factor (A_1)	$\Sigma \Sigma d_{rs} p_{..r} - \Sigma \lambda_i^{(1)} \Sigma \Sigma d_{rs} p_{i..} p_{i..}$
factor (A_2)	$\Sigma \Sigma d_{rs} p_{..r} - \Sigma \lambda_j^{(2)} \Sigma \Sigma d_{rs} p_{.jr} p_{.jr}$
interaction (A_1, A_2)	* (by subtraction)
within populations	$\Sigma \Sigma \lambda_i^{(1)} \lambda_j^{(2)} \Sigma \Sigma d_{rs} p_{ijr} p_{ijr}$
total	$\Sigma \Sigma d_{rs} p_{..r} p_{..r}$

As observed earlier, when n_{ij} do not satisfy the conditions stated above, we can still carry out the ANODIV by choosing appropriate values of $\lambda_i^{(1)}$ and $\lambda_j^{(2)}$ (see Rao, 1981 for an example). The formula for the interaction component when the prior probability for the (i, j) -th population $\lambda_{ij} \neq \lambda_i^{(1)} \lambda_j^{(2)}$ is somewhat complicated.

The sampling distribution of the various expressions in Table 5 when n_{ij} are large, which are currently under investigation, will be reported elsewhere. The extension of the ANODIV to a general m -way classified data can be done in the usual way.

5.4 *Entropy measures for ANODIV.* The following entropy measures

$$\begin{aligned} H(p) &= -\sum p_i \log p_i, \text{ Shannon (1948),} \\ &= (\alpha-1)^{-1}(1-\sum p_i^\alpha), \text{ Havrda and Charvát (1967)} \end{aligned}$$

are concave functions and take the value zero when all p_i are zero except one. Thus they satisfy the conditions C_1 and C_2 of a diversity measure. Any one of them can be used to apportion the diversity in a mixture of populations as between and within populations. For instance Lewontin (1972) used the Shannon entropy and Nei (1973) used the Havrda and Charvát entropy with $\alpha = 2$ (which is the same as the Gini-Simpson index) on the analysis of gene diversity as between and within populations. Rao (1980) indicated the possibility of using other entropy functions for this purpose.

Can these functions be used for ANODIV in an m -way classified data as in the case of the diversity measures constructed in Section 2.2? This question has been completely answered in two papers by Burbea and Rao (1980, 1981) and Rao (1981b). They have shown that Shannon's entropy may be used for the analysis of two-way classified data, and there is a possibility of higher order interactions being negative when the number of factors is more than 2. Similarly, the Havrda and Charvát entropy can be used for two-way classified data provided $\alpha \in [1, 2]$ when $k > 2$ and $\alpha \in [1, 2] \cup \left[3, \frac{11}{3}\right]$ when $k = 2$. But for data classified by 3 or more factors, higher order interactions may take negative values except when $\alpha = 2$, which corresponds to the Gini-Simpson index.

REFERENCES

- AGRESTI, A. and AGRESTI, B. F. (1978): Statistical analysis of qualitative variation. *Social Methodology* (K. F. Schuessler, Ed.), 204-237.
- ANDERSON, R. J. and LANDIS, J. R. (1980): CATANOVA for multidimensional contingency tables: Nominal scale response. *Comm. Statist.*, A9, 1191-1200.
- BHARGAVA, T. N. and DOYLE, P. H. (1974): A geometric study of diversity. *J. Theoretical Biology*, 43, 241-251.
- BHARGAVA, T. N. and UPFULURI, V. R. R. (1975): On diversity in human ecology. *Matron*, 34, 1-13.

- BISHOP, Y. M. M., FEINBERG, S. E. and HOLLAND, P. W. (1976): *Discrete Multivariate Analysis: Theory and Practice*. The MIT Press.
- DURBEA, J. and RAO, C. R. (1982): On the convexity of some divergence measures based on entropy functions. *IEEE Trans. Information Theory*, 28 (in press)
- (1981): On the convexity of higher order Jensen differences based on entropy functions. *Tech. Rept. 81-20*, University of Pittsburgh.
- CAVALLI-SFORZA, L. L. (1969): Human diversity. *Proc. XII Internat. Congr. Genetics*, Tokyo, 3, 405-16.
- CHAKRABORTY, R. (1974): A note on Nei's measure of gene diversity in a substructural population. *Humangenetik*, 21, 85-88.
- DENNIS, B., PATIL, G. P., ROSSI, O., STERMAN, S. and TAILLE, C. (1979): A bibliography of literature on ecological diversity and related methodology. In *Ecological Diversity in Theory and Practice*, 1 CPH, 319-354.
- GENI, C. (1912): Variabilità e mutabilità. Studi Economico-Giuridici della facoltà di Giurisprudenza dell'Università di Cagliari III, Parte II.
- HAMBERMAN, S. J. (1974): *The Analysis of Frequency Data*, Chicago University Press.
- HAYRADA, J. and CHARVAT, F. (1967): Quantification method in classification processes: concept of structural α -entropy. *Kybernetika*, 3, 30-35.
- HEDRICK, P. W. (1971): A new approach to measuring genetic similarity. *Evolution*, 25, 276-280.
- KARLIN, S., KENETT, R. and BONNE-TAMIR, B. (1979): Analysis of biochemical genetic data on Jewish populations: II. Results and interpretations of heterogeneity indices and distance measures with respect to standardis. *Am. J. Hum. Genet.*, 31, 341-365.
- LATTER, B. D. H. (1973): Measures of genetic distance between individuals and populations. Pages 27-39 in N. E. Morton, (ed.) *Genetic Structure of Populations*. Univ. of Hawaii Press.
- (1980): Genetic differences within and between populations of the major human subgroups. *Am. Nat.*, 116, 220-237.
- LEWONTIN, R. C. (1972): The apportionment of human diversity. *Evolutionary Biology*, 6, 381-398.
- LIGHT, R. J. and MARGOLIN, B. H. (1971): An analysis of variance for categorical data. *J. Amer. Stat. Assoc.*, 66, 534-544.
- MAHALANOBIS, P. C., MAJUMDAR, D. N. and RAO, C. R. (1949): Anthropometric survey of the United Provinces, 1941: a statistical study. *Sankhyā*, 9, 90-324.
- MAJUMDAR, D. N. and RAO, C. R. (1958): Bengal anthropometric survey, 1945: a statistical study. *Sankhyā*, 19, 203-408.
- MATRAI, A. and RATHIE, P. N. (1974): *Basic Concepts in Information Theory and Statistics*, Wiley (Halested Press).
- MATSUBATA, K. (1957): Decision rule based on the distance for the classification problem. *Ann. Inst. Statist. Math.*, 8, 67-77.
- MITTON, J. (1977): Genetic differentiation of races of man as judged by single locus and multi-locus analysis. *Am. Nat.*, 111, 203-212.
- MORTON, N. E. and LAPOVEL, J. M. (1973): Topology of kinship in Micronesia. *Amer. J. Hum. Genet.*, 25, 422-432.

- MORTON, N. E. (1973): Kinship and population structure. *In Genetic Structure of Populations*, pp. 66-71 (N. E. Morton, ed.).
- NEI, M. (1973): Analysis of gene diversity in subdivided populations. *Proc. Nat. Acad. Sci.*, 70, 3321-3323.
- (1978): The theory of genetic distance and evolution of human races. *Jap. J. Human Genet.*, 23, 341-369.
- PATIL, G. P. and TALLE, C. (1970): An overview of diversity. *In Ecological Diversity in Theory and Practice*. 1, CPH, 3-28.
- PENROSE, L. S. (1954): Distance, size and shape. *Ann. Eugen.*, 18, 337-343.
- PIELOU, E. C. (1975): *Ecological Diversity*, John Wiley, New York.
- (1970): Interpretation of paleoecological similarity matrices. *Paleobiology*, 5, 435-443.
- RAO, C. R. (1948): The utilization of multiple measurements in problems of biological classification. *J. Roy. Statist. Soc. B* 10, 150-193.
- (1962): Use of discriminant and allied functions in multivariate analysis. *Sankhyā A24*, 149-154.
- (1964): The use and interpretation of principal components in applied research. *Sankhyā A28*, 329-358.
- (1965, 1973): *Linear Statistical Inference and its Applications*, John Wiley, New York.
- (1971a): *Advanced Statistical Methods in Biometric Research*, Haffner (first edition 1952). John Wiley, New York.
- (1971b): Taxonomy in anthropology. *In Mathematics in the Archaeological and Historical Sciences*. 19-20. Edin. Univ. Press.
- (1977): Cluster analysis applied to a study of race mixture in human populations. *Proc. Michigan Univ. Symp.*, 175-197.
- (1980): Diversity and dissimilarity coefficients: a unified approach. *Tech. Rept. No. 80-10*. University of Pittsburgh.
- (1981a): Gini-Simpson index of diversity: a characterization, generalization and applications. *Tech. Rept. No. 81-22*, Univ. of Pittsburgh.
- (1981b): Analysis of diversity: a unified approach. *Tech. Rept. No. 81-28*. Univ. of Pittsburgh.
- RENYI, A. (1961): On measures of entropy and information. *Proc. Fourth Berkeley Symp.*, 1, 547-561.
- SANOHVI, L. D. (1953): Comparison of genetic and morphological methods for a study of biological differences. *Amer. J. Phy. Anthropol.*, 11, 385-404.
- SANOHVI, L. D. and BALAKRISHNAM, V. (1972): Comparison of different measures of genetic distance between human populations. *In The Assessment of Population Affinities in Man*. (J. S. Weiner and J. Huizinga, Eds.), pp. 25-36.
- SEN, A. (1973): *On Economic Inequality*, Clarendon Press, Oxford.
- SHANNON, C. E. (1948): A mathematical theory of communications. *Bell System Tech. J.*, 27, 379-423, 623-656.
- SIMPSON, E. H. (1949): Measurement of diversity. *Nature*, 163, 688.
- SPIELMAN, R. S. (1973): Do the natives all look alike? Size and shape components of anthropometric differences among Yanomama Indian villages. *American Naturalist*, 107, 694-708.
- SOKAL, R. R. and SNEATH, P. H. A. (1963): *Principles of Numerical Taxonomy*, W. H. Freeman.

Paper received: December, 1980.

Paper revised: December, 1981.