

Diversity and Dissimilarity Coefficients: A Unified Approach

C. RADHAKRISHNA RAO*

*Institute for Statistics and Applications,
Department of Mathematics and Statistics,
University of Pittsburgh, Pittsburgh, Pennsylvania 15260*

Received September 1, 1980

Three general methods for obtaining measures of diversity within a population and dissimilarity between populations are discussed. One is based on an intrinsic notion of dissimilarity between individuals and others make use of the concepts of entropy and discrimination. The use of a diversity measure in apportionment of diversity between and within populations is discussed.

1. INTRODUCTION

There is an extensive literature on measures of diversity within populations and dissimilarity or similarity between populations. They have been used in a wide variety of studies in anthropology (Rao, 1948; Mahalanobis *et al.*, 1949; Majumdar and Rao, 1958; Rao, 1971a,b, 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) and in biology (Sokal and Sneath, 1963; Pielou, 1975; Patil and Taille, 1979). An extensive bibliography of papers on measures of diversity and their applications is compiled by Dennis *et al.* (1979).

Most of these measures are based on heuristic considerations; some are derived from mathematically well-postulated axioms, while others are constructed assuming some models for genetic and environmental mechanisms causing differences between individuals and populations. The object of this paper is to review some of these measures and to provide some unified approaches for deriving them.

We consider a set of populations $\{\pi_i\}$, where the individuals of each population are characterized by a set of measurements $X \in (\Omega, \mathcal{B})$, a

* The work of this author is sponsored by the Air Force Office of Scientific Research under Contract F49620-79-0161. Reproduction in whole or in part is permitted for any purpose of the United States Government.

measurable space. The probability distribution function of X in π_i is denoted by P_i and the convex set generated by a mixture of $\{P_i\}$ by $\mathcal{P}$. A diversity coefficient (DIVC) is a mapping from $\mathcal{P}$ into the real line, which reflects differences between individuals (X 's) within a population. We denote the DIVC of π_i by H_i (the symbol H is used to indicate heterogeneity). A dissimilarity coefficient (DISC) or a similarity coefficient (SIMC) is a mapping from $\mathcal{P} \times \mathcal{P}$ into the real line, which reflects the differences or similarities between populations. We denote a DISC between π_i and π_j by D_{ij} and a SIMC by S_{ij} .

2. DIVCS AND DISCS BASED ON INTRINSIC DIFFERENCES BETWEEN INDIVIDUALS

2.1. General Theory

We start first by choosing a non-negative 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 underlying probability distributions of X . The choice of $d(X_1, X_2)$ naturally depends on the nature of the practical problem under investigation. We define the DIVC (diversity coefficient) of π_i as

$$H_i = \int d(X_1, X_2) P_i(dX_1) P_i(dX_2), \quad (2.1.1)$$

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

$$H_{ij} = \int d(X_1, X_2) P_i(dX_1) P_j(dX_2). \quad (2.1.2)$$

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

$$D_{ij} = H_{ij} - \frac{1}{2}(H_i + H_j). \quad (2.1.3)$$

The expression (2.1.3) 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 (2.1.1) is concave. This can be easily verified by considering $P_0 \in \mathcal{P}$, where

$$P_0 = \lambda P_i + (1 - \lambda) P_j, \quad 0 < \lambda < 1, \quad (2.1.4)$$

and computing

$$\begin{aligned} H_0^\lambda &= \int d(X_1, X_2) P_0(dX_1) P_0(dX_2) \\ &= \lambda^2 H_i + (1 - \lambda)^2 H_j + 2\lambda(1 - \lambda) H_{ij}. \end{aligned} \quad (2.1.5)$$

Then

$$\begin{aligned} H_0^\lambda - (\lambda H_i + (1 - \lambda) H_j) \\ = 2\lambda(1 - \lambda)(H_{ij} - \tfrac{1}{2}H_i - \tfrac{1}{2}H_j) = 2\lambda(1 - \lambda) D_{ij}. \end{aligned} \quad (2.1.6)$$

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

Note 1. In the definition (2.1.1) of diversity no condition is imposed on the function $d(X_1, X_2)$ except that it should be nonnegative and should have an intrinsic biological meaning. The logical requirement that the Jensen difference (2.1.3) should be nonnegative restricts the choice of $d(X_1, X_2)$ to functions which induce a concave functional on $\mathcal{P}$.

Note 2. We may define the expression (2.1.6) involving prior probabilities λ and $(1 - \lambda)$ for π_i and π_j as a more general Jensen difference between π_i and π_j . Fortunately, this difference is only a constant multiple of (2.1.3).

Note 3. The Jensen difference (2.1.3) is a function defined on $\mathcal{P} \times \mathcal{P}$. Suppose that we are measuring dissimilarity between say two linguistic groups L_1, L_2 residing in two different cities C_1 and C_2 . Let D_{12}^1 and D_{12}^2 be the DISCs between L_1 and L_2 in C_1 and C_2 , respectively, and D_{12} between L_1 and L_2 when the populations in the two cities are mixed. In such a case, we expect D_{12} to be smaller than $(D_{12}^1 + D_{12}^2)/2$. This logical requirement demands that the Jensen difference (2.1.3) should be a convex function on $\mathcal{P} \times \mathcal{P}$, which imposes a further restriction on the choice of $d(X_1, X_2)$. Some studies based on the convexity of (2.1.3) have been recently made by Burbea and Rao (1980) and Rao (1981).

2.2. Some Examples

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

$$d(X_1, X_2) = \langle X_1 - X_2, X_1 - X_2 \rangle. \quad (2.2.1)$$

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 and not necessarily m -variate normal). Then

$$\begin{aligned} H_i &= 2 \operatorname{tr} A \Sigma_i, \\ H_{ij} &= \operatorname{tr} A \Sigma_i + \operatorname{tr} A \Sigma_j + \delta'_{ij} A \delta_{ij}, \end{aligned} \quad (2.2.2)$$

where tr stands for the trace of a matrix and $\delta_{ij} = \mu_i - \mu_j$. Applying the formula (2.1.3)

$$D_{ij} = \delta'_{ij} A \delta_{ij}. \quad (2.2.3)$$

If $\Sigma_i = \Sigma$ for all i and $A = \Sigma^{-1}$, (2.2.3) 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

$$d(X_1, X_2) = m - \Sigma \delta_r, \quad (2.2.4)$$

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

$$p_{ir1}, \dots, p_{irk_r}$$

in population π_i . Define

$$j_{ii}^{(r)} = E(\delta_r) = \sum_{s=1}^{k_r} p_{irs}^2 \quad (2.2.5)$$

when X_1, X_2 are independently drawn from π_i and

$$j_{ij}^{(r)} = E(\delta_r) = \sum_{s=1}^{k_r} p_{irs} p_{jrs} \quad (2.2.6)$$

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

$$H_i = \sum_{r=1}^m (1 - j_{ii}^{(r)}) = m(1 - J_{ii}), \quad (2.2.7)$$

$$H_{ij} = \sum_{r=1}^m (1 - j_{ij}^{(r)}) = m(1 - J_{ij}),$$

$$\begin{aligned} D_{ij} &= H_{ij} - \frac{1}{2}(H_i + H_j) \\ &= m[\frac{1}{2}(J_{ii} + J_{jj}) - J_{ij}] = \frac{1}{2} \sum_{r=1}^m \sum_{s=1}^{k_r} (p_{irs} - p_{jrs})^2. \end{aligned} \quad (2.2.8)$$

The expression (2.2.8) 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 (2.2.8) reduces to the Gini-Simpson index

$$1 - \sum_{i=1}^k p_i^2, \quad (2.2.9)$$

where $p_1, \dots, p_k$ are the cell probabilities. [This measure was introduced by Gini (1912) and used by Simpson (1949) in biological work.] The properties of (2.2.9) 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 (2.2.7) 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 (2.2.4) is defined. The DISC (2.2.8) 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 (2.1.1) 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 (2.2.8), as the basic function for assessing the differences between individuals is not the same. When $x_1, \dots, x_m$ are independently distributed, an explicit expression for the DIVC based on the combined multinomial reduces to

$$H = 1 - [1 - H(1)] \dots [1 - H(m)], \quad (2.2.10)$$

where $H(r)$ is the DIVC based on x_r , the r th character only. It may be noted that the expression for DIVC given in (2.2.7) 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 (2.2.4) as 1 if two individual organisms have the same value for x_r and zero otherwise, then we get the DIVC and DISC 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 (2.2.5)–(2.2.8) with p_{irs} interpreted as genotype 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 DIVC and DISC based on a measure introduced by Latter (1973, 1980).

2.3. Apportionment of DIV

With the DIVC as defined by (2.1.1) and using the concavity property, the DIV in a mixture of populations can be apportioned in a natural way as between and within populations. If $P_1, \dots, P_k$ are the distributions of X in $\pi_1, \dots, \pi_k$ and $\lambda_1, \dots, \lambda_k$ are the a priori probabilities, then the distribution in the mixture π_0 is $\lambda_1 P_1 + \dots + \lambda_k P_k$. It is easily seen that

$$\begin{aligned} H_0 &= \Sigma \lambda_i H_i + \Sigma \Sigma \lambda_i \lambda_j D_{ij} \\ &= H(w) + D(b), \end{aligned} \quad (2.3.1)$$

where $D_{ij} = H_{ij} - (H_i + H_j)/2$ is the DISC between π_i and π_j , $H(w)$, the DIV within populations, is the weighted average of the DIVs within populations and $D(b)$, the DIS between populations, is the weighted average of the DISCs between all pairs of populations. The ratios

$$F(b) = \frac{D(b)}{H(w)} \quad \text{and} \quad G(b) = \frac{D(b)}{H_0} \quad (2.3.2)$$

have been used as measures of diversity between populations, of which the second expression $G(b)$ is in the form of an index lying in the range $(0, 1)$. Different choices of the difference function $d(X_1, X_2)$ may give different values to the ratios F and G . In Section 3, we shall discuss this problem in a more general context.

Let us consider k populations as in example (1) of Section 2.2, 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 (2.2.3) with $A = \Sigma^{-1}$). Further let π_0 be a mixture of $\pi_1, \dots, \pi_k$ with a priori probabilities $\lambda_1, \dots, \lambda_k$. Then using the expressions (2.2.2), the decomposition (2.3.1) becomes

$$\begin{aligned} H_0 &= H(w) + D(b) \\ &= 2m + \Sigma \Sigma \lambda_i \lambda_j \delta'_{ij} \Sigma^{-1} \delta_{ij} \\ &= 2m(1 + V), \end{aligned} \quad (2.3.3)$$

where $\delta_{ij} = \mu_i - \mu_j$. Thus the diversity within populations is $2m$ and the ratio $F(b)$ of (2.3.2) is V , which is the weighted combination of Mahalanobis D^2 's for all pairs of populations (see Appendix by Rao in Mahalanobis *et al.* (1949) for the use of an estimate of V in the selection of variables to maximize dissimilarity between populations). Note that normality of X is not required.

Let us consider example (2) of Section 2.2 and denote by π_0 , the mixture

of $\pi_1, \dots, \pi_k$ with a priori probabilities $\lambda_1, \dots, \lambda_k$. In this case (2.3.1) becomes, with J_{ij} as defined in (2.2.7),

$$H_0 = m[\Sigma \lambda_i(1 - J_{ii}) + \Sigma \Sigma \lambda_i \lambda_j (\frac{1}{2}J_{ii} + \frac{1}{2}J_{jj} - J_{ij})], \quad (2.3.4)$$

which is the decomposition obtained by Nei (1973) and Chakraborty (1974). The ratio $G(b)$ defined in (2.3.2) is

$$G(b) = \frac{\Sigma \Sigma \lambda_i \lambda_j (\frac{1}{2}J_{ii} + \frac{1}{2}J_{jj} - J_{ij})}{1 - \Sigma \Sigma \lambda_i \lambda_j J_{ij}}. \quad (2.3.5)$$

The ratio (2.3.5) obtained by considering only the two populations π_i and π_j with equal prior probabilities

$$\theta_{ij} = \frac{J_{ii} + J_{jj} - 2J_{ij}}{4 - J_{ii} - J_{jj} - 2J_{ij}} \quad (2.3.6)$$

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

2.4. Decomposition of DIVC and DISC

In the method outlined in Section 2.1, the basic expression which determines the DIVC and DISC is the difference function $d(X_1, X_2)$. Any decomposition of $d(X_1, X_2)$ such as

$$d(X_1, X_2) = d_1(X_1, X_2) + \dots + d_c(X_1, X_2) \quad (2.4.1)$$

provides us with a corresponding decomposition of the DIVC for π_i

$$H_i = H_i^{(1)} + \dots + H_i^{(c)}, \quad (2.4.2)$$

where $H_i^{(s)} = E[d_s(X_1, X_2) | P_i]$, and of the DISC between π_i and π_j

$$D_{ij} = D_{ij}^{(1)} + \dots + D_{ij}^{(c)}, \quad (2.4.3)$$

where $D_{ij}^{(s)}$ is obtained from $H_i^{(s)}$, $H_j^{(s)}$ and $H_{ij}^{(s)}$ using the formula (2.1.3).

Let $X \sim (\mu_i, \Sigma)$ in π_i and denote the eigenvalues of Σ by $\theta_1 \geq \dots \geq \theta_m$ and the corresponding normalized orthogonal eigenvectors by $L_1, \dots, L_m$. If we choose

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

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

$$d(X_1, X_2) = [L_1'(X_1 - X_2)]^2 + \dots + [L_m'(X_1 - X_2)]^2 \quad (2.4.4)$$

gives the decomposition of DIVC for π_i

$$H_i = 2 \operatorname{tr} \Sigma = 2\theta_1 + \cdots + 2\theta_m, \quad (2.4.5)$$

which is the familiar decomposition of total variability with respect to m characters in terms of principal components (Rao, 1964). The corresponding decomposition of DISC between π_i and π_j is

$$D_{ij} = \delta'_{ij} \delta_{ij} = (L'_1 \delta_{ij})^2 + \cdots + (L'_m \delta_{ij})^2, \quad (2.4.6)$$

where $\delta_{ij} = \mu_i - \mu_j$, the difference in the mean vectors for π_i and π_j . However, if we choose

$$d(X_1, X_2) = (X_1 - X_2)' \Sigma^{-1} (X_1 - X_2),$$

i.e., the Mahalanobis distance between two individuals then we have a different decomposition

$$D_{ij} = \delta'_{ij} \Sigma^{-1} \delta_{ij} = \frac{1}{\theta_1} (L'_1 \delta_{ij})^2 + \cdots + \frac{1}{\theta_m} (L'_m \delta_{ij})^2. \quad (2.4.6)$$

Note that the eigen vectors provide a transformation of the original measurements into uncorrelated variables, in which case the Mahalanobis distance can be written as the sum of Mahalanobis distances due to different uncorrelated variables. We can choose any arbitrary set of vectors $M_1, \dots, M_m$ such that $M'_i \Sigma M_j = 0$ for $i \neq j$ and $M'_i \Sigma M_i = 1$, to obtain a decomposition

$$\begin{aligned} D_{ij} &= \delta'_{ij} \Sigma^{-1} \delta_{ij} = (M'_1 \delta_{ij})^2 + \cdots + (M'_m \delta_{ij})^2 \\ &= D_{ij}^{(1)} + \cdots + D_{ij}^{(m)}. \end{aligned} \quad (2.4.7)$$

By combining some of the D_{ij} 's on the right-hand side of (2.4.7), we obtain decompositions of D_{ij} with a smaller number of components.

If we choose

$$M_1 = (\sigma' \Sigma^{-1} \sigma)^{-1/2} \Sigma^{-1} \sigma \quad (2.4.8)$$

in (2.4.7), where σ is the vector of standard deviations of the individual characters (i.e., square roots of diagonal elements of Σ), then

$$(M'_1 \delta_{ij})^2 = D_{si}^2 \quad (2.4.9)$$

represents the component of Mahalanobis D^2 between π_i and π_j due to the size factor as defined by Rao (1962, 1971b). Then

$$D_{ij} = D^2 = D_{si}^2 + D_{sh}^2, \quad (2.4.10)$$

where D_{sh}^2 , the residual after subtracting the D^2 due to size, represents the distance due to shape factors between the two populations.

Penrose (1954) obtained a similar decomposition of Karl Pearson's CRL (coefficient of racial likeness) in terms of size and shape. The Penrose indices do not take into account the correlations that may exist between characters. For further details regarding the use of size and particular shape factors reference may be made to Rao (1962, 1971b).

2.5. Similarity Coefficients (SIMCs)

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 DIVC 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 DISC obtained by choosing $H_i = 1 - S_i$, provided it satisfies the concavity condition, is

$$D_{ij} = \frac{1}{2}(S_i + S_j) - S_{ij} \quad (2.5.1)$$

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

$$\begin{aligned} D_{ij} &= \frac{1}{2}(\log S_i + \log S_j) - \log S_{ij} \\ &= -\log \frac{S_{ij}}{\sqrt{S_i S_j}}. \end{aligned} \quad (2.5.2)$$

For instance, in the second example of Section 2.2, a natural definition of $s(X_1, X_2) = (\Sigma \delta_r)/m$, which lies in the range $(0, 1)$. Then

$$S_i = J_{ii}, \quad S_j = J_{jj}, \quad S_{ij} = J_{ij}, \quad (2.5.3)$$

where J_{ij} are as defined in (2.2.7), and using (2.5.1) and (2.5.2) we have the alternative forms

$$D_{ij} = \frac{1}{2}(J_{ii} + J_{jj}) - J_{ij}, \quad (2.5.4)$$

$$D_{ij} = -\log \frac{J_{ij}}{\sqrt{J_{ii} J_{jj}}}. \quad (2.5.5)$$

The expression (2.5.4) is the same as the "minimum genetic distance" (2.2.8) of Nei (1978), and (2.5.5) is what he calls the "standard genetic distance."

Again, in the example (2), we may define the similarity function as $(\delta_1 \cdots \delta_m)^{1/m}$ instead of $(\delta_1 + \cdots + \delta_m)/m$. 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,

$$\begin{aligned} S_i &= j_{ii}^{(1)} \cdots j_{ii}^{(m)} = (J'_{ii})^m, \\ S_{ij} &= j_{ij}^{(1)} \cdots j_{ij}^{(m)} = (J'_{ij})^m, \end{aligned} \quad (2.5.6)$$

where $j_{ij}^{(r)}$ are as defined in (2.2.5) and (2.2.6). Taking logarithms of (2.5.6), the corresponding DISC is

$$D_{ij} = -\log \frac{J'_{ij}}{\sqrt{J'_{ii}J'_{jj}}}, \quad (2.5.7)$$

which Nei calls the “maximum genetic distance.”

Since $S_i + S_j \geq 2S_{ij}$, we may define an index of similarity $2S_{ij}/(S_i + S_j)$ as proposed by Hedrick (1971). This lies in the range (0, 1) and the corresponding DISC may be defined as

$$D_{ij} = -\log \frac{2S_{ij}}{S_i + S_j} \quad (2.5.8)$$

which is analogous to (2.5.7).

3. ENTROPY AND INFORMATION

3.1. Measures of Entropy

A wide variety of DIVCs have been introduced through the concept of entropy and information. The general approach in these cases is basically different from that of Section 2.1, 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 dissimilarity 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 non-negative, 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 Rathie (1975) consider three general forms for entropy:

$$H = (1 - \alpha)^{-1} \log(\Sigma p_r^{\alpha + \beta_r - 1} / \Sigma p_r^{\beta_r}), \quad (3.1.1)$$

$$H = [(\Sigma p_r^{\alpha + \beta_r - 1} / \Sigma p_r^{\beta_r}) - 1] \div (2^{1-\alpha} - 1), \quad (3.1.2)$$

$$H = -\Sigma p_r^{\beta_r} \log p_r / \Sigma p_r^{\beta_r}, \quad (3.1.3)$$

where all the summations are taken from 1 to k . When $\beta_r = 1$ for all r we have the familiar expressions introduced by Renyi (1961), Havrda and Charvat (1967) and Shannon (1948).

All the functions (3.1.1)–(3.1.3) are non-negative, 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.

With $\beta_r = 1$, (3.1.2), (3.1.3) and (3.1.1) for $\alpha < 1$ are concave and the method of Section 2.1 can be used to construct a DISC between π_i and π_j . For instance, choosing (3.1.3) with $\beta_r = 1$ as a DIVC, and a mixture π_0 of populations π_i and π_j with a priori probabilities λ_1 and λ_2 , we have

$$H_i = - \sum_{r=1}^k p_{ir} \log p_{ir}, \quad (3.1.4)$$

$$H_0 = - \sum_{r=1}^k (\lambda_1 p_{ir} + \lambda_2 p_{jr}) \log(\lambda_1 p_{ir} + \lambda_2 p_{jr}),$$

$$\begin{aligned} D_{ij} &= H_0 - \lambda_1 H_i - \lambda_2 H_j \\ &= \lambda_1 \Sigma p_{ir} \log \frac{p_{ir}}{\lambda_1 p_{ir} + \lambda_2 p_{jr}} + \lambda_2 \Sigma p_{jr} \log \frac{p_{jr}}{\lambda_1 p_{ir} + \lambda_2 p_{jr}}, \end{aligned} \quad (3.1.5)$$

which is the information radius defined by Sibson and Jardine (1971) from other considerations.

Similarly, the DISC between π_i and π_j obtained by choosing (3.1.2) with $\beta_r = 1$ is

$$D_{ij} = [\Sigma(\lambda_1 p_{ir} + \lambda_2 p_{jr})^\alpha - \lambda_1 \Sigma p_{ir}^\alpha - \lambda_2 \Sigma p_{jr}^\alpha] \div (2^{1-\alpha} - 1), \quad (3.1.6)$$

which, when $\alpha = 2$, reduced to the Euclidean distance, apart from a constant multiplier,

$$2\lambda_1 \lambda_2 \Sigma (p_{ir} - p_{jr})^2. \quad (3.1.7)$$

The DISC obtained by choosing (3.1.1) with $\beta_r = 1$ and $\alpha < 1$ is

$$D_{ij} = (1 - \alpha)^{-1} \log \frac{\Sigma(\lambda_1 p_{ir} + \lambda_2 p_{jr})^\alpha}{(\Sigma p_{ir}^\alpha)^{\lambda_1} (\Sigma p_{jr}^\alpha)^{\lambda_2}}. \quad (3.1.8)$$

which may be negative for $\alpha > 1$ since Renyi's entropy is not in general concave for $\alpha > 1$. The formulae (3.1.5)–(3.1.8) involve explicitly the prior probabilities λ_1, λ_2 . In many practical applications, it is appropriate to choose $\lambda_1 = \lambda_2 = 1/2$ to define a DISC between two populations.

3.2. Apportionment of Diversity

By considering a mixture π_0 of populations $\pi_1, \dots, \pi_m$ with prior probabilities $\lambda_1, \dots, \lambda_m$ we can obtain a decomposition of DIV in π_0 , based on any choice of the H functions (3.1.2), (3.1.3) and (3.1.1) with $\alpha < 1$,

$$\begin{aligned} H_0 &= \Sigma \lambda_r H_r + (H_0 - \Sigma \lambda_r H_r), \\ &= H(w) + D(b), \end{aligned} \quad (3.2.1)$$

as DIV within and DIS between populations. It may be noted that $D(b)$ cannot in general be obtained as a weighted combination of DISCs between all pairs of populations as in (2.3.1) for the choice of DIVCs derived by the method of Section 2.1. (It is, however, true when H is chosen as in (3.1.2) with $\beta_r = 1$ and $\alpha = 2$, in which case it also belongs to the class of DIVCs derived in Section 2.1). The ratio $G(b) = D(b)/H_0$ has been used by geneticists as an index of diversity between populations compared to within. However, as observed in Section 2.3, its value depends on the H function chosen. In their studies on diversity with respect to blood groups and biochemical markers, Lewontin (1972) used the H function (3.1.3) with $\beta_r = 1$, and Nei (1973) and Chakraborty (1974) used (3.1.2) with $\alpha = 2$ and $\beta_r = 1$. This raises the question as to what is a suitable choice of DIVC among the functions (3.1.1)–(3.1.3) to study the apportionment of diversity as between and within populations.

To examine this question, the following computations were made in the simple case of two binomial populations with equal prior probabilities. The class of H functions considered is a subclass of (3.1.2) and (3.1.3),

$$H^{(\alpha)} = (p_1^\alpha + p_2^\alpha - 1)(2^{1-\alpha} - 1),$$

where for $\alpha = 1$, the function is defined by the limiting value

$$H^{(1)} = -p_1 \log p_1 - p_2 \log p_2.$$

Table I gives the values of $D(b)/H_0$ for different combinations of the proportions for the two binomials. For each combination, the first entry corresponds to the value of $G(b)$ for $\alpha = 1$, the second for $\alpha = 2$, the third for $\alpha = 2.5$, the fourth for the optimum α , and the fifth entry within brackets gives α_* , the optimum value of α . The blanks for certain combinations indicate that the values are the same as for the combination with the complementary values of (p_1, q_1) , the binomial proportions of the two populations.

TABLE I

Values of $G(b)$ between Two Binomial Distributions Defined by p_1 and q_1

$p_1 \backslash q_1$	0.005	0.05	0.10	0.30	0.50	0.70	0.90	0.95
0.995	0.955*	0.834	0.741	0.505	0.359	0.248	0.134	0.087
	0.980	0.895	0.808	0.529	0.324	0.168	0.045	0.019
	0.981	0.897	0.812	0.530	0.321	0.161	0.040	0.016
	0.981	0.897	0.812	0.530	0.363	0.276	0.173	0.124
	(2.43)	(2.44)	(2.45)	(2.45)	(0.80)	(0.60)	(0.54)	(0.54)
0.95		0.714	0.622	0.388	0.242	0.127	0.017	
		0.810	0.724	0.451	0.254	0.108	0.009	
		0.815	0.730	0.454	0.253	0.105	0.008	
		0.815	0.730	0.454	0.255	0.127	0.019	
		(2.44)	(2.45)	(2.45)	(1.69)	(0.97)	(0.73)	
0.90			0.531	0.305	0.167	0.065		
			0.640	0.375	0.191	0.063		
			0.647	0.380	0.191	0.061		
			0.647	0.380	0.191	0.066		
			(2.45)	(2.46)	(2.37)	(1.25)		
0.70				0.119	0.031			
				0.160	0.042			
				0.163	0.043			
				0.163	0.043			
				(2.47)	(2.47)			

* The first four vertical entries correspond to $\alpha = 1, 2, 2.5$ and α_* , respectively. The last entry within brackets is α_* , the optimal value.

It is seen from Table I that the quantity $G(b)$ is fairly stable for different choices of α except for low values of p_1, q_1 in which case the DIVC based on Shannon entropy seems to exaggerate the diversity between populations. This is also seen from the computations presented in Table II on the geographical diversity between Makiritare Indians from seven different villages with respect to nine blood group and seven protein loci. The computations are based on the data kindly supplied by Chakraborty (see Chakraborty, 1974 for his analysis of the data) using the formula (3.2.1) with equal a priori probabilities for the villages. The values of $G(b)$ are fairly consistent for different values of α except when the average gene frequency is low as in the case of 6PGD (see Table II) for which the value of $G(b)$ is very high when $\alpha = 1$. The diversity based on Shannon entropy is extremely sensitive at such low values and provides an exaggerated value for diversity. This phenomenon was also observed by Latter (1980), where he compared the

TABLE II
Gene DIV of Makiritare Indians in Seven Villages
and Index of DIS between Villages

Locus	Ave <i>p</i>	<i>α</i> = 1		<i>α</i> = 2		<i>α</i> = 2.5	
		<i>H</i> ₀	<i>G</i> (<i>b</i>)	<i>H</i> ₀	<i>G</i> (<i>b</i>)	<i>H</i> ₀	<i>G</i> (<i>b</i>)
Serological							
Diego	0.196	0.7139	0.1743	0.6303	0.1711	0.6240	0.1693
Kidd	0.336	0.9209	0.0250	0.8924	0.0320	0.8899	0.0325
Rh(C)	0.418	0.9805	0.0401	0.9731	0.0542	0.9724	0.0554
P	0.434	0.9874	0.0172	0.9826	0.0232	0.9821	0.0237
Lewis	0.466	0.9967	0.0791	0.9954	0.1044	0.9952	0.1191
Ss	0.470	0.9974	0.0575	0.9964	0.0770	0.9963	0.0786
Rh(E)	0.563	0.9885	0.0058	0.9841	0.0079	0.9837	0.0081
MN	0.714	0.8635	0.0263	0.8168	0.0291	0.8128	0.0292
Duffy	0.736	0.8327	0.0122	0.7772	0.0142	0.7726	0.0166
Average		0.9202	0.0415	0.8943	0.0448	0.8921	0.0486
Biochemical							
AP	0.0557	0.3101	0.0647	0.2104	0.0238	0.2054	0.0213
Hp	0.424	0.9833	0.0650	0.9769	0.0866	0.9763	0.0884
Gc	0.820	0.6801	0.0431	0.5904	0.0432	0.5837	0.0427
PGM _i	0.848	0.6148	0.0592	0.5156	0.0504	0.5086	0.0488
Lp	0.876	0.5407	0.0084	0.4345	0.0052	0.4275	0.0047
Alb	0.9857	0.1081	0.1293	0.0564	0.1719	0.0547	0.1444
6PGD	0.991	0.0741	0.2503	0.0357	0.0678	0.0346	0.0934
Average		0.4730	0.0561	0.4028	0.0521	0.3987	0.0522

performance of diversity coefficients based on the Shannon entropy, Hedrick's measure of genotype identity and Latter's measure of gene identity. Based on these studies and Latter's observation that "Shannon's measure of information... has many convenient properties from a mathematical point of view, but is extremely difficult to interpret genetically," it appears that the DIVC based on Shannon's entropy is at a disadvantage compared to the others.

4. DISCRIMINATION INDEX

A general method of constructing DISCs is through the concept of discrimination between populations, i.e., the probability with which a given

individual can be identified as a member of one of two populations to which he possibly belongs.

4.1. *Overlap Distance* (Rao, 1948, 1977; Wald, 1950)

Let X be a set of measurements which has the probability density $p_i(\cdot)$ in π_i and $p_j(\cdot)$ in π_j . The best decision rule based on an observed value x of X , for discriminating between π_i and π_j with prior probabilities in the ratio 1 : 1 is to assign x to

$$\begin{aligned} \text{population } \pi_i & \quad \text{if } p_i(x) > p_j(x), \\ \text{population } \pi_j & \quad \text{if } p_i(x) < p_j(x), \end{aligned} \quad (4.1.1)$$

and to decide by tossing an unbiased coin when $p_i(x) = p_j(x)$. The probability of correct classifications for the optimum decision rule is

$$C_{ij} = \frac{1}{2} \int_{R_1} p_i(x) dx + \frac{1}{2} \int_{R_2} p_j(x) dx, \quad (4.1.2)$$

where R_1 is the region $p_i(x) \geq p_j(x)$ and R_2 , the region $p_j(x) > p_i(x)$. The minimum value of (4.2) is $1/2$ which is attained when $p_i(\cdot) = p_j(\cdot)$, and the maximum is unity when the supports of $p_i(\cdot)$ and $p_j(\cdot)$ are disjoint. The more dissimilar the populations are, the greater would be the probability of correct classifications. Then we may define the DISC between π_i and π_j as

$$D_{ij} = C_{ij} - \frac{1}{2}, \quad (4.1.3)$$

which is in the range $(0, \frac{1}{2})$. It is seen that

$$C_{ij} - \frac{1}{2} = \frac{1}{4} \int |p_i(x) - p_j(x)| dx, \quad (4.1.4)$$

which is a multiple of Kolmogorov's variational distance or city block distance, which is a special case of the Minkowski distance

$$\left[\int |p_i(x) - p_j(x)|^t dx \right]^{1/t}, \quad t \geq 1. \quad (4.1.5)$$

In the development of decision theory, Wald (1950) introduced the distance function between π_i and π_j ,

$$D_{ij} = \max_R \left| \int_R [p_i(x) - p_j(x)] dx \right|, \quad (4.1.6)$$

where R represents any arbitrary region. The expression (4.1.6) is identifiable as

$$D_{ij} = 1 - \int \min[p_i(x), p_j(x)] dx \quad (4.1.7)$$

$$= \int_{R_1} [p_i(x) - p_j(x)] dx, \quad (4.1.8)$$

where R_1 is the region $p_i(x) \geq p_j(x)$ as in (4.1.2). The expression (4.1.8) is the difference between the proportions of correct and wrong classifications by using the optimum decision rule (4.1.1). The expression (4.1.7) may be interpreted as the proportion of mismatched individuals in the two populations.

4.2. Quadratic Differential Metric (Rao, 1948)

Let us consider a family of probability densities $p(x, \theta)$, $\theta \in \Theta$, a k -vector parameter space. The Fisher information matrix at θ is $M = [m_{ij}(\theta)]$, where

$$m_{ij}(\theta) = \int \frac{1}{p} \frac{dp}{d\theta_i} \frac{dp}{d\theta_j} dx. \quad (4.2.1)$$

We endow the space Θ with the quadratic differential metric

$$\Sigma \Sigma m_{ij}(\theta) \delta\theta_i \delta\theta_j, \quad (4.2.2)$$

and define the distance between two points θ_1 and θ_2 as the geodesic distance determined by (4.2.2). The expression (4.2.2) is a measure of difference between two probability distributions close to each other and the distance defined by it may be useful in evolutionary studies where gradual changes take place in a population in moving from state θ_1 to state θ_2 . In a recent paper Atkinson and Mitchell (1981) have derived the expressions for geodesic distances based on (4.2.2) for well-known families of distributions.

4.3. Invariants of Jeffreys

Jeffreys (1948) defined what are called invariants between two distributions

$$I_m = \int [|p_i(x)|^{1/m} - |p_j(x)|^{1/m}]^m dx, \quad m > 0, \quad (4.3.1)$$

$$I_0 = \int [p_i(x) - p_j(x)] \log \frac{p_i(x)}{p_j(x)} dx,$$

where the second expression is the sum of Kullback–Leibler information numbers

$$I_{ij} = \int p_i(x) \log \frac{p_i(x)}{p_j(x)} dx, \quad I_{ji} = \int p_j(x) \log \frac{p_j(x)}{p_i(x)} dx. \quad (4.3.2)$$

When $m = 1$,

$$I_1 = \int |p_i(x) - p_j(x)| dx \quad (4.3.3)$$

which is Kolmogorov's variational distance (overlap distance of Rao, 1948). When $m = 2$

$$\begin{aligned} I_2 &= \int [\sqrt{p_i(x)} - \sqrt{p_j(x)}]^2 dx \\ &= 2 \left(1 - \int \sqrt{p_i(x) p_j(x)} dx \right), \end{aligned} \quad (4.3.4)$$

which is extensively used by Matusita (1957) in inference problems. The expression (4.3.4) is a function of the Hellinger distance

$$\cos^{-1} \int \sqrt{p_i(x) p_j(x)} dx. \quad (4.3.5)$$

Rao and Varadarajan (1963) have defined the Hellinger DISC to be

$$-\log_e \int \sqrt{p_i(x) p_j(x)} dx. \quad (4.3.6)$$

The measure (4.3.5) was proposed by Bhattacharya (1946) as a DISC between populations π_i and π_j and has been used in some genetic studies. The alternative expression (4.3.6) has an advantage over (4.3.5) in the sense that it is additive with respect to characteristics independently distributed in the populations.

It may be noted that (4.3.3), which is Jeffrey's invariant of order 1, is twice (4.1.8), which is Wald's distance. Again (4.1.8) is twice (4.1.3) the overlap distance of Rao.

5. SOME REMARKS

The paper discusses three unified methods of deriving measures of diversity and dissimilarity. Of these, the new method based on an intrinsic difference between individuals discussed in Section 2 offers a wide scope. The choice of the basic difference function has to be made on biological considerations depending on the problem under investigation. The discrimination approach of Section 4 also admits practical interpretation and the measures based on this approach are specially useful in the study of infraspecific variation among taxonomical units (Jardine and Sibson, 1971, p. 33). The measures based on the concept of entropy considered in Section 3 have attractive mathematical properties but are difficult to interpret genetically (Latter, 1980).

For the apportionment of genetic diversity between and within populations, a number of diversity measures have been proposed by different authors (see Latter, 1980, Lewontin, 1972 and Nei, 1973). Among those studied, the differences in the estimates of $G(b)$, the index of between population diversity, are of a small order except in the case of Shannon entropy (first used by Lewontin) which seems to give an exaggerated value for diversity between populations when the gene frequencies are at very low levels. Further work is needed to bring out finer distinctions among the other measures.

There has been considerable debate on the choice of a dissimilarity measure in comparing populations. The problem cannot be answered without a reasonably precise statement of the objectives as observed by Smith (1977). If only a descriptive classification of the populations is intended through dendograms or other clustering methods, a fairly large class of measures may lead to the same conclusions. In such cases, it may be useful to try more than one possible measure to test the consistency of final conclusions (see Karlin *et al.*, 1979). In some situations a choice can be made among a number of measures for particular applications, by first testing them on a group of populations whose interrelationships have been otherwise well established. That measure which gives results most consistent with available knowledge may be chosen. Such a procedure has been proposed by Pielou (1979) and applied in choosing one out of nine possible similarity measures in a study on the interpretation of paleoecological similarity matrices. In genetic studies, it is of some importance that the distance function should reflect in an easily interpretable way the basic genetic differences. In such a case the results may have some evolutionary significance (see Morton, 1973 and Nei, 1978). Further theoretical and empirical investigations are necessary to obtain more insight into these problems.

REFERENCES

- AGRESTI, A. AND AGRESTI, B. F. 1978. Statistical analysis of qualitative variation, in "Social Methodology" (K. F. Schussler, Ed.), pp. 204-237.
- ATKINSON, C. AND MITCHELL, A. F. S. 1981. Rao's distance measure, *Sankhya*, in press.
- BHARGAVA, T. N. AND DOYLE, P. H. 1974. A geometric study of diversity, *J. Theoret. Biol.* **43**, 241-251.
- BHARGAVA, T. N. AND UPPULURI, V. R. R. 1975. On diversity in human ecology, *Metron* **34**, 1-13.
- BHATTACHARYA, A. 1946. A measure of divergence between two multinomial populations, *Sankhya* **7**, 401.
- BURBEA, J. AND RAO, C. R. 1980. "On the Convexity of Jensen Difference Arising from Entropy of Order," Tech. Report, University of Pittsburgh.
- CAVALLI-SFORZA, L. L. AND EDWARDS, A. W. F. 1967. Phylogenetic analysis: Models and estimation procedures, *Amer. J. Hum. Genet.* **19**, 233-257.
- CAVALLI-SFORZA, L. L. 1969. Human diversity, in "Proceedings, XII Internat. Congr. Genetics, Tokyo," Vol. 3, pp. 405-416.
- 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., STEHMAN, S., AND TAILLE, C. 1979. A bibliography of literature on ecological diversity and related methodology, in "Ecological Diversity in Theory and Practice," Vol. 1 CPH, pp. 319-354.
- GINI, C. 1912. "Variabilità e Mutabilità," Studi Economicoaguridici della facoltà di Giurisprudenza dell'Università di Cagliari III, Parte II.
- HAVRADA, 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.
- JARDINE, N. AND GIBSON, R. 1971. "Mathematical Taxonomy," Wiley, New York.
- JEFFREYS, H. 1948. "Theory of Probability," 2nd edition, Oxford Univ. Press (Clarendon), London.
- 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 standards, *Amer. J. Hum. Genet.* **31**, 341-365.
- LATTER, B. D. H. 1973. Measures of genetic distance between individuals and populations, in "Genetic Structure of Populations" (N. E. Morton, Ed.) pp. 27-39, Univ. of Hawaii Press, Honolulu.
- LATTER, B. D. H. 1980. Genetic differences within and between populations of the major human subgroups, *Amer. Natur.* **116**, 220-237.
- LEWONTIN, R. C. 1972. The apportionment of human diversity, *Evolutionary Biol.* **6**, 381-398.
- MAHALANOBIS, P. C., MAJUMDAR, D. N., AND RAO, C. R. 1949. Anthropometric survey of the United Provinces, 1941: A statistical study, *Sankhya* **9**, 90-324.
- MAJUMDAR, D. N. AND RAO, C. R. 1958. Bengal anthropometric survey, 1945: A statistical study. *Sankhya* **19**, 203-408.
- MATHAI A. AND RATHIE, P. N. 1974. "Basic Concepts in Information Theory and Statistics," Wiley (Halsted Press), New York.
- MATUSITA, 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 multilocus analysis, *Amer. Natur.* **111**, 203-212.

- MORTON, N. E. AND LALOVEL, J. M. 1973. Typology of kinship in Micronesia, *Amer. J. Hum. Genet.* **25**, 422-432.
- MORTON, N. E. (1973). Kinship and population structure, in "Genetic Structure of Populations" (N. E. Morton, Ed.), pp. 66-71, Univ. of Hawaii Press, Honolulu.
- NEI, M. 1973. Analysis of gene diversity in subdivided populations, *Proc. Nat. Acad. Sci.* **70**, 3321-3323.
- NEI, M. 1978. The theory of genetic distance and evolution of human races, *Japan. J. Human Genet.* **23**, 341-369.
- PATIL, G. P. AND TAILLE, C. 1979. An overview of diversity, in "Ecological Diversity in Theory and Practice," Vol. 1 CPH, pp. 3-28.
- PENROSE, L. S. 1954. Distance, size and shape, *Ann. Eugen.* **18**, 337-343.
- PIELOU, E. C. 1975. "Ecological Diversity," Wiley, New York.
- PIELOU, E. C. 1979. Interpretation of paleoecological similarity matrices, *Paleobiology* **5**, 435-443.
- RAO, C. R. 1945. Information and accuracy attainable in the estimation of parameters, *Bull. Calcutta Math. Soc.* **37**, 81-91.
- RAO, C. R. 1948. The utilization of multiple measurements in problems of biological classification, *J. Roy. Statist. Soc. B* **10**, 159-193.
- RAO, C. R. 1954. On the use and interpretation of distance functions in statistics, *Bull. Internat. Statist. Inst.* **34**, 90.
- RAO, C. R. 1962. Use of discriminant and allied functions in multivariate analysis, *Sankhya A* **24**, 149-154.
- RAO, C. R. 1964. The use and interpretation of principal components in applied research, *Sankhya A* **26**, 329-358.
- RAO, C. R. 1965, 1973. "Linear Statistical Inference and its Applications," Wiley, New York.
- RAO, C. R. 1971a. "Advanced Statistical Methods in Biometric Research," Haffner (first edition 1952), Wiley, New York.
- RAO, C. R. 1971b. Taxonomy in anthropology, in "Mathematics in the Archaeological and Historical Sciences," pp. 19-20, Edinburgh Univ. Press, Edinburgh.
- RAO, C. R. 1977. Cluster analysis applied to a study of race mixture in human populations, in "Proceedings, Michigan Univ. Symp.," pp. 175-197.
- RAO, C. R. 1981. "Analysis of Diversity: A Unified Approach," Univ. of Pittsburgh. Tech. Rept. 81-20.
- RENYI, A. 1961. On measures of entropy and information, in "Proceedings, Fourth Berkeley Symp.," Vol. 1, pp. 547-561.
- SANGHVI, L. D. 1953. Comparison of genetic and morphological methods for a study of biological differences, *Amer. J. Phys. Anthropol.* **11**, 385-404.
- SANGHVI, L. D. AND BALAKRISHNAN, 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, Oxford Univ. Press, London.
- SEN, A. 1973. "On Economic Inequality," Oxford Univ. Press (Clarendon), London.
- 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.
- SMITH, C. A. B. 1977. A note on genetic distance, *Ann. Hum. Genet. (London)* **40**, 463-479.
- SOKAL, R. R. AND SNEATH, P. H. A. 1963. "Principles of Numerical Taxonomy," Freeman, San Francisco.
- WALD, A. 1950. "Statistical Decision Functions," Wiley, New York.