DIVERSITY AND DISSIMILARITY COEFFICIENTS: A UNIFIED APPROACH (U)
DIVERSITY AND DISSIMILARITY COEFFICIENTS:
A UNIFIED APPROACH

C. Radhakrishna Rao*

July 1980

Technical Report No. 80-10

Institute for Statistics and Applications
Department of Mathematics and Statistics
University of Pittsburgh
Pittsburgh, PA. 15260

*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.
DIVERSITY AND DISSIMILARITY COEFFICIENTS:
A UNIFIED APPROACH

C. Radhakrishna Rao*

Summary: 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.
A criterion is developed for choosing a measure of diversity
in a given class of measures. The Gini-Simpson index of
diversity is derived as the solution to a functional
equation.

AMS Classification 62H30

Key Words: Entropy, Information, Diversity, Dissimilarity,
Similarity, Mahalanobis $D^2$, Size and Shape
Factors, Discrimination, Geodesic Distance

*The work of this author is sponsored by the Air Force Office
of Scientific Research under Contract F49620-79-0161. Reproduc-
duction in whole or in part is permitted for any purpose of
the United States Government.
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, Majumdar and Rao, 1949; Majumdar and Rao, 1958; Rao, 1971a, b, 1977), in genetics (Cavalli-Sfroza, 1969; Karlin et al, 1979; Morton and Lalouel, 1973; Nei, 1978; Sanghvi, 1953; Sanghvi and Balakrishnan, 1972), in economics (Gini, 1912; Sen, 1973) in sociology (Agresti and Agresti, 1972) and in biology (Sokhal 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).

Most of these measures are based on heuristic considerations; some are derived from mathematically well postulated axioms, while others are constructed using possible 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 measurable space. The probability distribution function of \( X \) in \( \pi_i \) is denoted by \( P_i \) and the convex set generated by \( \{ P_i \} \) is denoted by \( P \). A diversity coefficient
(DIVC) is a mapping from $P$ into the real line, which reflects differences between individuals (X's) within a population. We denote the DIVC of $\tau_i$ by $H_i$ (the symbol $H$ is used to indicate heterogeneity). A dissimilarity coefficient (DISC) or a similarity coefficient (SIMC) is a mapping from $P \times P$ into the real line, which reflects the differences or similarities between populations. We denote a DISC between $\tau_i$ and $\tau_j$ by $D_{ij}$ and a SIMC by $S_{ij}$.

2. COEFFICIENTS 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 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 DIVC (diversity coefficient) of $\tau_i$ as

$$H_i = \int d(X_1,X_2) \, P_1 (dX_1) \, P_1 (dX_2)$$

(2.1.1)

i.e., as the average difference between two randomly drawn individuals from $\tau_i$. Suppose that one individual is drawn from $\tau_i$ and another from $\tau_j$. 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 \( \pi_i \) and \( \pi_j \) may be defined by what can be termed as the Jensen difference.

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

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 \( P \) as in (2.1.1) is concave. This can be easily verified by considering \( P_0 \in P \) where

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

and computing

\[ H_0 = \int d(X_1, X_2) P_0 (dX_1) P_0 (dX_2) \]

\[ = \lambda^2 H_i + (1-\lambda)^2 H_j + 2\lambda (1-\lambda)H_{ij}. \quad (2.1.5) \]

Then

\[ H_0 - (\lambda H_i + (1-\lambda)H_j) \]

\[ = 2\lambda(1-\lambda)(H_{ij} - \frac{1}{2} H_i - \frac{1}{2} H_j) = 2\lambda(1-\lambda)D_{ij}. \quad (2.1.6) \]

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

### 2.2 Some Examples

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

\[
d(X_1, X_2) = (X_1 - X_2, X_1 - X_2).
\] (2.2.1)

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

\[
H_i = 2 \text{tr} \ A \Sigma_i
\]

\[
H_{ij} = \text{tr} \ A \Sigma_i + \text{tr} \ A \Sigma_j + \delta_{ij} \ A \delta_{ij}
\] (2.2.2)

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

\[
D_{ij} = \delta_{ij} A \delta_{ij}.
\] (2.2.3)

If \( \Sigma_i = \Sigma \) for all \( i \) and \( A = \Sigma^{-1} \), (2.2.3) becomes the Mahalanobis \( D^2 \) between \( \pi_i \) and \( \pi_j \).

(2) Let \( X = (x_1, \ldots, x_m) \) 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

\[
d(X_1, X_2) = m - \Sigma \delta_r
\] (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
in population \( \pi_1 \). Define

\[
j_{ii}^{(r)} = E(\delta_{ri}) = \frac{k_r}{\sum_{s=1}^{k_r} p_{irs}}
\]

(2.2.5)

when \( X_1, X_2 \) are independently drawn from \( \pi_i \) and

\[
j_{ij}^{(r)} = E(\delta_{rij}) = \frac{k_r}{\sum_{s=1}^{k_r} p_{irs} p_{jrs}}
\]

(2.2.6)

when \( X_1 \) is drawn from \( \pi_i \) and \( X_2 \) from \( \pi_j \). Then

\[
H_i = \sum_{r=1}^{m} (1 - j_{ii}^{(r)}) = m(1 - j_{ii})
\]

\[
H_{ij} = \sum_{r=1}^{m} (1 - j_{ij}^{(r)}) = m(1 - j_{ij})
\]

(2.2.7)

\[
D_{ij} = H_{ij} - \frac{1}{2} (H_i + H_j)
\]

\[
= m \left( \frac{1}{2} (j_{ii} + j_{jj}) - j_{ij} \right) = \frac{1}{2} \sum_{r=1}^{m} \sum_{s=1}^{k_r} (p_{irs} - p_{jrs})^2.
\]

(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

\[
1 - \sum_{i=1}^{k} p_i^2
\]  

(2.2.9)

where \( p_1, \ldots, p_k \) 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 \( x_i, i = 1, \ldots, m \), 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 \((x_1, \ldots, x_m)\) as a combined multinomial with \( k = k_1 \times \ldots \times k_m \) 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 \( x_1, \ldots, x_m \) are independently distributed, an explicit expression for the DIVC based on the combined multinomial reduces to
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 \( H = \sum H(r) \) whether \( x_i \) are independently distributed or not.

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 \( P_1, \ldots, P_k \) are the distributions of \( X \) in \( \pi_1, \ldots, \pi_k \) and \( \lambda_1, \ldots, \lambda_k \) are the apriori probabilities, then the distribution in the mixture \( \pi_o \) is \( \lambda_1 P_1 + \ldots + \lambda_k P_k \). It is easily seen that

\[
H_o = \Sigma \lambda_i H_i + \Sigma \lambda_i \lambda_j D_{ij}
\]

\[
= H(w) + D(b) \quad (2.3.1)
\]

where \( D_{ij} = H_{ij} - (H_i + H_j)/2 \) is the DISC between \( \pi_i \) and \( \pi_j \). \( H(w) \), the DIV within populations, is the weighted average of the DIV's within populations and \( D(b) \), the DISC between populations, is the weighted average of the DISC's between all pairs of populations. The ratio

\[
G(b) = \frac{D(b)}{H_o} \quad (2.3.2)
\]

is an index of diversity between populations.
Different choices of the difference function \( d(X_1, X_2) \) may give different values to the ratio \( G(b) \). 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 \( \pi_i \), 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 \( \pi_o \) be a mixture of \( \pi_1, \ldots, \pi_k \) with apriori probabilities \( \lambda_1, \ldots, \lambda_k \). Then using the expressions (2.2.2), the decomposition (2.3.1) becomes

\[
H_O = H(w) + D(b) \\
= 2m + \Sigma \lambda_i \lambda_j \delta_{ij} \Sigma^{-1} \delta_{ij} \\
= 2m(1 + V) \tag{2.3.3}
\]

where \( \delta_{ij} = \mu_i - \mu_j \). Thus the diversity within populations is \( 2m \) and the ratio \( G(b) \) of (2.3.2) 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 2.2 and denote by \( \pi_o \), the mixture of \( \pi_1, \ldots, \pi_k \) with apriori probabilities \( \lambda_1, \ldots, \lambda_k \). In this case (2.3.1) becomes, with \( J_{ij} \) as defined in (2.2.7),

\[
H_O = m[\Sigma \lambda_i (1 - J_{ii}) + \Sigma \Sigma \lambda_i \lambda_j (\frac{1}{2} J_{ii} + \frac{1}{2} J_{jj} - J_{ij})] \tag{2.3.4}
\]
which is the decomposition obtained by Nei (1973) and Chakravarthy (1974). The ratio \( G(b) \) defined in (2.3.2) is

\[
G(b) = \frac{\sum \lambda_i \lambda_j \left( \frac{1}{2} J_{ij} + \frac{1}{2} J_{jj} - J_{ij} \right)}{1 - \sum \lambda_i \lambda_j J_{ij}}.
\]

The ratio (2.3.5) obtained by considering only the two populations \( \pi_i \) and \( \pi_j \) with equal prior probabilities

\[
\theta_{ij} = \frac{J_{ii} + J_{jj} - 2 J_{ij}}{4 - J_{ii} - J_{jj} - 2 J_{ij}}
\]

is the hybridity coefficient of Morton (1973) who used it as a DISC between \( \pi_i \) and \( \pi_j \) 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

\[
d(X_1,X_2) = d_1(X_1,X_2) + \ldots + d_c(X_1,X_2)
\]

(2.4.1)

provides us with a corresponding decomposition of the DIVC for \( \pi_i \)

\[
H_i = H^{(1)}_i + \ldots + H^{(c)}_i
\]

(2.4.2)

where \( H_i^{(s)} = E[d_s(X_1,X_2) | P_i] \), and of the DISC between \( \pi_i \) and \( \pi_j \).
\[ D_{ij} = D_1^{(s)} + \ldots + D_m^{(c)} \quad (2.4.3) \]

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 \( \pi_i \) and denote the eigen values of \( \Sigma \) by \( \theta_1 \geq \ldots \geq \theta_m \) and the corresponding eigen vectors by \( L_1, \ldots, L_m \).

If we choose

\[ d(X_1, X_2) = (X_1 - X_2)'(X_1 - X_2) \]

i.e., the simple Euclidean distance in \( \mathbb{R}^m \), then

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

gives the decomposition of DVC for \( \pi_i \)

\[ H_i = 2 \text{ tr } \Sigma = 2 \theta_1 + \ldots + 2 \theta_m \quad (2.4.5) \]

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 \( \pi_i \) and \( \pi_j \) is

\[ D_{ij} = \delta_{ij}^2 + \ldots + \delta_{ij}^2 \quad (2.4.6) \]

where \( \delta_{ij} = \mu_i - \mu_j \), the difference in the mean vectors for \( \pi_i \) and \( \pi_j \). 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

\[ D_{ij} = \delta_{ij}^2 \Sigma^{-1} \delta_{ij} = \frac{1}{\theta_1} (L_1 \delta_{ij})^2 + \ldots + \frac{1}{\theta_m} (L_m \delta_{ij})^2. \quad (2.4.6) \]

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 \( M_1, \ldots, M_m \) such that \( M_i^T \Sigma M_j = 0 \) for \( i \neq j \) and \( M_i^T \Sigma M_i = 1 \), to obtain a decomposition

\[ D_{ij} = \delta_{ij}^2 \Sigma^{-1} \delta_{ij} = (M_1^T \delta_{ij})^2 + \ldots + (M_m^T \delta_{ij})^2 \]

\[ = D^{(1)}_{ij} + \ldots + D^{(m)}_{ij}. \quad (2.4.7) \]

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

\[ M_1 = (\sigma L^{-1} \sigma)^{1/2} L^{-1} \sigma \quad (2.4.8) \]

in (2.4.7), where \( \sigma \) is the vector of standard deviations of the individual characters (i.e., square roots of diagonal elements of \( \Sigma \)), then
\[(\mathbf{M}_1\delta_{ij})^2 = D_{si}^2 \]  

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

\[D_{ij} = D^2 = D_{si}^2 + D_{sh}^2 \]  

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 (SIMC's)

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 \(\pi_i\) 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 \(H_i = 1 - S_i\) is

\[D_{ij} = \frac{1}{2}(S_i + S_j) - S_{ij} \]  

(2.5.1)
and that by choosing $H_i = - \log S_i$ is

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

$$= - \log \frac{S_{ij}}{\sqrt{S_i S_j}} . \quad (2.5.2)$$

For instance, in the second example of Section 2.2, a natural definition of $s(X_1, X_2) = (\sum \delta_r)/m$, 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,
\[ S_i = j^{(1)}_{ii} \ldots j^{(m)}_{ii} = (J'_{ii})^m \]
\[ S_{ij} = j^{(1)}_{ij} \ldots j^{(m)}_{ij} = (J'_{ij})^m \quad (2.5.6) \]

where \( j^{(r)}_{ij} \) 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".

2.6 A Functional Equation

Consider a multinomial distribution in \( k \) classes with probabilities \( p = (p_1, \ldots, p_k) \), and let \( H(p) \) be a DIVC. The maximum DIV obtains when \( p = (k^{-1}, \ldots, k^{-1}) = e \), say (for evenness), so that we may have the condition:

\[ C_1: \max_{p} H(p) = H(e). \quad (2.6.1) \]

Using \( H(p) \) as a DIVC, we can construct a DISC between the multinomials defined by \( p \) and \( e \) by using (2.1.3),

\[ D_{pe} = H \left( \frac{p+e}{2} \right) - \frac{1}{2} H(p) - \frac{1}{2} H(e). \quad (2.6.2) \]

The larger the value of \( H(p) \), the closer \( p \) is to \( e \), which suggests an alternative way of defining the DIS between the populations defined by \( p \) and \( e \) as a quantity proportional to
\[ \max_p H(p) - H(p) = H(e) - H(p). \]  
(2.6.3)

Equating (2.6.2) to a constant multiple of (2.6.3) we obtain the functional equation

\[ H\left(\frac{p+e}{2}\right) - \frac{1}{2} [H(p)+H(e)] = c [H(e)-H(p)] \]

or

\[ H\left(\frac{p+e}{2}\right) = \left(\frac{1}{2} + c\right) H(e) + \left(\frac{1}{2} - c\right) H(p). \]  
(2.6.4)

where \( c \) is a constant. There may be many solutions to (2.6.4) subject to the condition \( C_1 \). We shall impose some regularity conditions on \( H(p) \) in order to restrict the solutions to a smaller class:

- \( C_2: \) \( H(p) \) is symmetric in \( p_1, \ldots, p_k \)
- \( C_3: \) \( H(p) \) admits first and second order partial derivations with respect to \( p_1, \ldots, p_{k-1} \) and the \((k-1) \times (k-1)\) matrix

\[ H''(p) = \begin{pmatrix} \frac{\partial^2}{\partial p_1 \partial p_j} H(p) \end{pmatrix} \]

is continuous and not null at \( p = e \).

Of course \( H'(p) = 0 \) at \( p = e \) in view of the condition \( C_1 \) and the condition \( C_3 \) ensures that the diversity measure is locally sensitive when \( p \) deviates from \( e \).

We shall show that under the conditions \( C_1, C_2 \) and \( C_3 \), the function \( H(p) \) satisfying the equation (2.6.4) is of the form

\[ H(p) = a \left(1 - \Sigma p_1^2\right) + b \]  
(2.6.5)
where \( a > 0 \) and \( b \) are constants, i.e., \( H(p) \) is essentially the Gini-Simpson index.

(i) Using the condition \( C_3 \), we obtain on taking the first and second derivatives of both sides of (2.6.5) with respect to \( p_1, \ldots, p_{k-1} \),

\[
\frac{1}{2} H'(\frac{p+e}{2}) = (\frac{1}{2} - c) H'(p) \quad (2.6.6)
\]

\[
\frac{1}{4} H''(\frac{p+e}{2}) = (\frac{1}{2} - c) H''(p) \quad (2.6.7)
\]

where \( H' \) is a \( k-1 \) vector and \( H'' \) is a \((k-1) \times (k-1)\) matrix. Putting \( p = e \) in (2.6.7)

\[
\frac{1}{4} H''(e) = (\frac{1}{2} - c) H''(e) \quad (2.6.8)
\]

which implies that \( c = 1/4 \), using the condition \( H''(e) \neq 0 \)

(ii) The equation (2.6.7) becomes

\[
H''(\frac{p+e}{2}) = H''(p). \quad (2.6.9)
\]

Repeated use of (2.6.9) gives

\[
H''(p) = H'' [2^{-n}(p-e)+e] + H''(e). \quad (2.6.10)
\]

The equation (2.6.10) implies that \( H(p) \) is quadratic in \( p_1, \ldots, p_{k-1} \), which may be written, using the condition of symmetry,
\[ H(p) = \lambda_1 \sum p_1^2 + \lambda_2 \sum p_i p_j + \lambda_3 \sum p_i + \lambda_4 \]
\[ = \mu_1 \sum p_1^2 + \lambda_2 (1-p_k)^2 + \lambda_3 (1-p_k) + \lambda_4 \]  
(2.6.11)

where all the summations are taken from 1 to \( k-1 \). The condition \( C_2 \) demands symmetry with respect to \( p_1, \ldots, p_k \), in which case (2.6.11) assumes the form

\[ H(p) = \mu_1 \sum \frac{k}{l} p_i^2 + \mu_2. \]  
(2.6.12)

Using the condition \( C_1 \), we find that \( \mu_1 < 0 \) in which case \( H(p) \) is of the form

\[ a(1- \sum p_i^2) + b \]  
(2.6.13)

where \( a > 0 \), which is required to be proved.

3. ENTROPY AND INFORMATION

3.1 Measures of Entropy

A wide variety of DIVC's 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 en-
tropy 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 \( p_1, \ldots, p_k \) and for a particular population \( \pi_i \) by \( p_{i1}, \ldots, p_{ik} \). Mathai and Rathie (1975) consider three general forms for entropy:

\[
H = (1 - \alpha)^{-1} \log \left( \sum p_r^{a+\beta} / \sum p_r^\beta \right) 
\]  
(3.1.1)

\[
H = \left[ \left( \sum p_r^{a+\beta} / \sum p_r^\beta \right) - 1 \right] \times (2^{1-\alpha} - 1) 
\]  
(3.1.2)

\[
H = - \sum p_r^\beta \log p_r / \sum p_r^\beta 
\]  
(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.

The functions (3.1.1) - (3.1.3) are all concave and the method of Section 2.1 can be used to construct a DISC between \( \pi_i \) and \( \pi_j \). For instance, choosing (3.1.3) with \( \beta_r = 1 \) as a DIVC, and a mixture \( \pi_o \) of populations \( \pi_i \) and \( \pi_j \) with apriori probabilities \( \lambda_1 \) and \( \lambda_2 \), we have

\[
H_i = - \sum_{r=1}^{k} p_{ir} \log p_{ir}
\]

\[
H_0 = - \sum_{r=1}^{k} (\lambda_1 p_{ir} + \lambda_2 p_{jr}) \log (\lambda_1 p_{ir} + \lambda_2 p_{jr}) \tag{3.1.4}
\]

\[
D_{ij} = H_0 - \lambda_1 H_i - \lambda_2 H_j
\]

\[
= \lambda_1 \sum_{r=1}^{k} p_{ir} \log \frac{p_{ir}}{\lambda_1 p_{ir} + \lambda_2 p_{jr}} + \lambda_2 \sum_{r=1}^{k} p_{jr} \log \frac{p_{jr}}{\lambda_1 p_{ir} + \lambda_2 p_{jr}} \tag{3.1.5}
\]

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

Similarly, the DISC between \( \pi_i \) and \( \pi_j \) obtained by choosing (3.1.2) with \( \beta_r = 1 \) is
\[ D_{ij} = \left[ \sum (\lambda_1 p_{ir} + \lambda_2 p_{jr}) - \lambda_1 \sum p_{ir}^\alpha - \lambda_2 \sum p_{jr}^\alpha \right] / (2^{1-\alpha} - 1) \tag{3.1.6} \]

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

\[ 2 \lambda_1 \lambda_2 \sum (p_{ir} - p_{jr})^2. \tag{3.1.7} \]

The DISC obtained by choosing (3.1.1) with \( \beta_r = 1 \) is

\[ D_{ij} = \log \frac{\sum (\lambda_1 p_{ir} + \lambda_2 p_{jr})^\alpha}{(\sum p_{ir})^{\lambda_1}(\sum p_{jr})^{\lambda_2}}. \tag{3.1.8} \]

The formulae (3.1.5)-(3.1.8) involve explicitly the prior probabilities \( \lambda_1, \lambda_2 \). 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 \( \pi_o \) of populations \( \pi_1, \ldots, \pi_m \) with prior probabilities \( \lambda_1, \ldots, \lambda_m \) we can obtain a decomposition of \( \text{DIV} \) in \( \pi_o \), based on any choice of the \( H \) functions (3.1.1)-(3.1.3),

\[ H_o = \sum \lambda_r H_r + (H_o - \sum \lambda_r H_r) \]

\[ = H(w) + D(b) \tag{3.2.1} \]

as \( \text{DIV} \) within and \( \text{DIS} \) between populations. It may be noted that \( D(b) \) cannot in general be obtained as a weighted combin-
ation of DISC's between all pairs of populations as in (2.3.1) for the choice of DIVC's 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 DIVC's 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 Chakravarthy (1974) used (3.1.2) with \( \alpha = 2 \) and \( \beta_r = 1 \). This raises the question as to what is the optimum choice of a DIVC in a given class \( \{H\} \) to study the apportionment of DIV as between and within populations. A natural choice appears to be one which maximizes the ratio \( G(b) = D(b)/H_0 \) or minimizes the ratio \( H(w)/H_0 \). Such a choice will depend on populations under study and the prior probabilities.

To examine the extent to which the optimum choice depends on the population distributions, 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 1 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 \( \alpha \), and the fifth entry within brackets gives \( \alpha^* \), the optimum value of \( \alpha \). The blanks for certain combinations indicate that the values are the same as for the combination with the complimentary values of \((p_1, q_1)\), the binomial proportions of the two populations. It is seen that the optimum value \( \alpha^* \) depends on the values of \( p_1, q_1 \), although it is stable for a wide range of values. If \( p_1 \) and \( q_1 \) are both small or both large \( \alpha^* \) is small and tends to zero as \( p_1 \) and \( q_1 \) approach zero or unity.

For values of \( p_1, q_1 \) near the boundary determined by the points \((.005,.7), (.01,.6), (.05,.5), (.1,.4), (.2,.3)\), \( \alpha^* \) is close to unity which corresponds to the Shannon DIVC. For other ranges of \( (p_1,q_1) \), \( \alpha^* \) is nearly 2.5, although \( \alpha = 2 \), which corresponds to the Gini-Simpson index is a close competitor.

The values of the ratio \( G(b) \) for the heptoglobin diversity in 25 Caucasian populations considered by Lewontin (1972) for different values of \( \alpha \) are as follows:

\[
\begin{array}{ccc}
\alpha: & 1.0 & 2.0 & 2.5 \\
G(b): & .0209 & .0249 & .0251 \\
\end{array}
\]

The frequency of the heptoglobin allele in these cases varied between 21% and 45% except in one case it was 12%. The optimum \( \alpha \) in such cases is about 2.5.
### TABLE 1

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

<table>
<thead>
<tr>
<th>$p_1$</th>
<th>$q_1$</th>
<th>.005</th>
<th>.05</th>
<th>.10</th>
<th>.30</th>
<th>.50</th>
<th>.70</th>
<th>.90</th>
<th>.95</th>
</tr>
</thead>
<tbody>
<tr>
<td>.995</td>
<td>.955*</td>
<td>.834</td>
<td>.741</td>
<td>.505</td>
<td>.359</td>
<td>.248</td>
<td>.134</td>
<td>.087</td>
<td></td>
</tr>
<tr>
<td>.980</td>
<td>.985</td>
<td>.808</td>
<td>.529</td>
<td>.324</td>
<td>.168</td>
<td>.045</td>
<td>.019</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.981</td>
<td>.987</td>
<td>.812</td>
<td>.530</td>
<td>.321</td>
<td>.161</td>
<td>.040</td>
<td>.016</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.981</td>
<td>.987</td>
<td>.812</td>
<td>.530</td>
<td>.363</td>
<td>.276</td>
<td>.173</td>
<td>.124</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2.43)</td>
<td>(2.44)</td>
<td>(2.45)</td>
<td>(2.45)</td>
<td>(0.80)</td>
<td>(0.60)</td>
<td>(0.54)</td>
<td>(0.54)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.95</td>
<td></td>
<td>.714</td>
<td>.622</td>
<td>.388</td>
<td>.242</td>
<td>.127</td>
<td>.017</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.810</td>
<td>.724</td>
<td>.451</td>
<td>.254</td>
<td>.108</td>
<td>.009</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.815</td>
<td>.730</td>
<td>.454</td>
<td>.253</td>
<td>.105</td>
<td>.008</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.815</td>
<td>.730</td>
<td>.454</td>
<td>.255</td>
<td>.127</td>
<td>.019</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2.44)</td>
<td>(2.45)</td>
<td>(2.45)</td>
<td>(1.69)</td>
<td>(0.97)</td>
<td>(0.73)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.90</td>
<td></td>
<td>.531</td>
<td>.305</td>
<td>.167</td>
<td>.065</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.640</td>
<td>.375</td>
<td>.191</td>
<td>.063</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.647</td>
<td>.380</td>
<td>.191</td>
<td>.061</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.647</td>
<td>.380</td>
<td>.191</td>
<td>.066</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2.45)</td>
<td>(2.46)</td>
<td>(2.37)</td>
<td>(1.25)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.70</td>
<td></td>
<td>.119</td>
<td>.031</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.160</td>
<td>.042</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.163</td>
<td>.043</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.163</td>
<td>.043</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2.47)</td>
<td>(2.47)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* The first four vertical entries correspond to $a = 1, 2, 2.5$ and $a_*$ respectively. The last entry within brackets is $a_*$, the optimal value.
Table 2 gives the values of $H_o$ and $G(b)$ for 9 blood group and 7 protein loci in the case of Makiritare Indians from 7 different villages. These were computed using the data kindly supplied by Chakravarthy (1974), assuming equal population sizes for the villages. It is seen from Table 2 that for the blood group loci, where $p$ values are in the interval (30%, 70%), the optimum $a$ is 2.5; and for the biochemical markers, where $p$ values are in the interval (5%, 20%), the optimum $a$ is 1, although the differences in $G$ values are not large. The value of $\alpha = 2.5$ comes out better on the criterion suggested for the choice of a DIVC. However, the value of $\alpha = 2.0$ is a close competitor and has other desirable properties (see Burbea and Rao, 1980).

4. DISCRIMINATION INDEX

A general method of constructing DISC's 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 $\pi_i$ and $p_j(\cdot)$ in $\pi_j$. The best decision rule based on an observed value $x$ of $X$, for discriminating between $\pi_i$ and $\pi_j$ with prior probabilities in the ratio 1:1 is to assign $x$ to
TABLE 2
Gene DIV of Makiritare Indians in Seven Villages
and Index of DIS Between Villages

<table>
<thead>
<tr>
<th>Locus</th>
<th>Ave p</th>
<th>$\alpha = 1$</th>
<th>$\alpha = 2$</th>
<th>$\alpha = 2.5$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$H_0$</td>
<td>G(b)</td>
<td>$H_0$</td>
<td>G(b)</td>
</tr>
<tr>
<td>Serological</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diego</td>
<td>.196</td>
<td>.7139</td>
<td>.1743</td>
<td>.6303</td>
</tr>
<tr>
<td>Kidd</td>
<td>.336</td>
<td>.9209</td>
<td>.0250</td>
<td>.8924</td>
</tr>
<tr>
<td>Rh(C)</td>
<td>.418</td>
<td>.9805</td>
<td>.0401</td>
<td>.9731</td>
</tr>
<tr>
<td>P</td>
<td>.434</td>
<td>.9874</td>
<td>.0172</td>
<td>.9826</td>
</tr>
<tr>
<td>Lewis</td>
<td>.466</td>
<td>.9967</td>
<td>.0791</td>
<td>.9954</td>
</tr>
<tr>
<td>Ss</td>
<td>.470</td>
<td>.9974</td>
<td>.0575</td>
<td>.9964</td>
</tr>
<tr>
<td>Rh(E)</td>
<td>.563</td>
<td>.9885</td>
<td>.0058</td>
<td>.9841</td>
</tr>
<tr>
<td>MN</td>
<td>.714</td>
<td>.8635</td>
<td>.0263</td>
<td>.8168</td>
</tr>
<tr>
<td>Duffy</td>
<td>.736</td>
<td>.8327</td>
<td>.0122</td>
<td>.7772</td>
</tr>
<tr>
<td>Average</td>
<td>.9202</td>
<td>.0415</td>
<td>.8943</td>
<td>.0448</td>
</tr>
<tr>
<td>Biochemical</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ap</td>
<td>.0557</td>
<td>.3101</td>
<td>.0647</td>
<td>.2104</td>
</tr>
<tr>
<td>Hp</td>
<td>.424</td>
<td>.9833</td>
<td>.0650</td>
<td>.9769</td>
</tr>
<tr>
<td>Gc</td>
<td>.820</td>
<td>.6801</td>
<td>.0431</td>
<td>.5904</td>
</tr>
<tr>
<td>$PGM_1$</td>
<td>.848</td>
<td>.6148</td>
<td>.0592</td>
<td>.5156</td>
</tr>
<tr>
<td>Lp</td>
<td>.876</td>
<td>.5407</td>
<td>.0084</td>
<td>.4345</td>
</tr>
<tr>
<td>Alb</td>
<td>.9857</td>
<td>.1081</td>
<td>.1293</td>
<td>.0564</td>
</tr>
<tr>
<td>6PGD</td>
<td>.991</td>
<td>.0741</td>
<td>.2503</td>
<td>.0357</td>
</tr>
<tr>
<td>Average</td>
<td>.4730</td>
<td>.0561</td>
<td>.4028</td>
<td>.0521</td>
</tr>
</tbody>
</table>
population $\pi_i$ if $p_i(x) > p_j(x)$

population $\pi_j$ if $p_i(x) < p_j(x)$  \hspace{1cm} (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

$$C_{ij} = \frac{1}{2} \int_{R_1} p_i(x) \, dx + \frac{1}{2} \int_{R_2} p_j(x) \, dx$$  \hspace{1cm} (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 $\pi_i$ and $\pi_j$ as

$$D_{ij} = C_{ij} - \frac{1}{2}$$  \hspace{1cm} (4.1.3)

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

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

$$[\int |p_i(x) - p_j(x)|^t \, dx]^{1/t}, \ t \geq 1.$$  \hspace{1cm} (4.1.5)
In the development of decision theory, Wald (1950) introduced the distance function between \( \pi_i \) and \( \pi_j \)

\[
D_{i,j} = \max_R \left| \int_R [p_i(x) - p_j(x)] \, dx \right|
\]  

(4.1.6)

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

\[
D_{i,j} = 1 - \int \min[p_i(x), p_j(x)] \, dx
\]  

(4.1.7)

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

(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 \( \theta \) is \( M = [m_{ij}(\theta)] \) where

\[
m_{ij}(\theta) = \int \frac{1}{p} \frac{dp}{d\theta_i} \frac{dp}{d\theta_j} \, dx.
\]  

(4.2.1)

We endow the space \( \Theta \) with the quadratic differential metric
and define the distance between two points $\theta_1$ and $\theta_2$ 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 $\theta_1$ to state $\theta_2$. In a recent paper Atkinson and Mitchell (1980) 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

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

$$I_0 = \int \left[ p_i(x) - p_j(x) \right] \log \frac{p_i(x)}{p_j(x)} \, dx \quad (4.3.1)$$

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 $m = 1$,

$$I_1 = \int \left| p_i(x) - p_j(x) \right| \, dx \quad (4.3.3)$$
which is Kolmogorov's variational distance (overlap distance of Rao, 1948). When \( m = 2 \)

\[
I_2 = \int \left[ \sqrt{p_1(x)} - \sqrt{p_j(x)} \right]^2 dx
\]

\[
= 2 \left( 1 - \int \sqrt{p_1(x)} \sqrt{p_j(x)} dx \right)
\]  

(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

\[
\cos^{-1} \int \sqrt{p_1(x) p_j(x)} dx.
\]  

(4.3.5)

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

\[- \log_e \int \sqrt{p_1(x) p_j(x)} dx.\]  

(4.3.6)

The measure (4.3.5) was proposed by Bhattacharya (1946) as a DISC between populations \( \pi_i \) and \( \pi_j \) 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 is seen that there are various approaches for measuring DIV and DIS and some of the controversies on the choice of these measures in practical investigations (see Li, 1978; Nei, 1978; Morton, 1973; and Smith, 1977) may be resolved through the concepts developed in the present paper. Some further work in this direction, which is in progress, will be reported elsewhere.
REFERENCES


Diversity and Dissimilarity Coefficients: A Unified Approach

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. A criterion is developed for choosing a measure of diversity in a given class of measures. The Gini-Simpson index of diversity is derived as the solution to a functional equation.