Inequalities in Statistics and Probability
IMS Lecture Notes–Monograph Series Vol. 5 (1984), 68–77

CONVEXITY PROPERTIES OF ENTROPY FUNCTIONS AND ANALYSIS OF DIVERSITY

BY C. RADHAKRISHNA RAO
University of Pittsburgh

Some natural conditions which a diversity measure (variability) of a probability distribution should satisfy imply that it must have certain convexity properties, considered as a functional on the space of probability distributions. It is shown that some of the well known entropy functions, which are used as diversity measures do not have all the desirable properties and are, therefore, of limited use. A new measure called the quadratic entropy has been introduced, which seems to be well suited for studying diversity.

Methods for apportioning diversity (APDIV) at various levels of a hierarchically classified set of populations are described. The concept of analysis of diversity (ANODIV), as a generalization of ANOVA, applicable to observations of any type, is developed and its use in the analysis of cross classified data is demonstrated. The choice of a suitable measure of diversity for the above purpose is discussed.

1. Introduction. There is an extensive literature on the measurement and analysis of diversity. A unified approach to these problems is given in Rao (1982a), and a complete bibliography of papers on this subject is complied by Denis, Patil, Rossi and Taille (1979). The choice of a diversity (DIV) measure for the analysis of given data poses a serious problem. An attempt is made in this paper to lay down some natural conditions which a diversity measure should satisfy (Section 2) and discuss the methodology for data analysis through an appropriate diversity measure. Some of the situations where such an analysis is needed are as follows.

Geneticists are interested in comparing populations by the diversity exhibited in certain measurements (Karlin, Kennett and Bonne-Tamir (1979)), and in apportioning diversity (APDIV) in a substructured population as due to between and within groups (Lewontin (1972), Nei (1973), Chakraborty (1974), Rao (1982a) and Rao and Boudreau (1982)).

In analysis of variance (ANOVA) of quantitative data, we choose the *variance* as a measure of diversity and partition it into a number of additive components. 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 factors. A natural question arises as to whether other measures of diversity such as *mean absolute deviation* could be used for this purpose. Further, what is the natural extension of ANOVA to observations which are not quantitative in nature?

In this paper, the concept of ANOVA is extended to more general analysis of diversity (ANODIV) applicable to observations belonging to *any sample space* by an appropriate choice of a diversity measure satisfying some convexity properties.

The choice of the well known entropy functions due to Shannon (1948), Havrda and Charvát (1967) and Rényi (1961) as diversity measures have only limited use as they do

---

AMS 1980 subject classifications, 94A17, 62J10.

Key words and phrases: Analysis of diversity, Complete convexity, Diversity measure, Entropy, Jensen difference, Rao's Quadratic entropy.

not have strong convexity properties (Section 3). A new measure called *quadratic entropy* which is *completely convex* (Section 4) is introduced and shown to be a useful tool in APDIV (Section 5) and ANODIV (Sections 6 and 7).

2. Measures of Diversity. Consider a measurable space $(\mathcal{X},\mathcal{B})$ and a convex set $\mathcal{P}$ of probability measures defined on it. A function H mapping $\mathcal{P}$ into the real line R is said to be a measure of diversity if

$$C_0\colon H(P)\geqslant 0 \quad \text{for every } P\in\mathcal{P} \text{ and } H(P)=0 \text{ iff } P \text{ is degenerate.} \tag{2.1}$$

The condition C_0 is a natural one since a measure of diversity should preferably be nonnegative and take the value zero only when all the individuals of a population are identical.

Consider two measures P_1 and P_2 and a mixture $\lambda_1P_1+\lambda_2P_2$, $(\lambda_1\geqslant0,\ \lambda_2\geqslant0,\ \lambda_1+\lambda_2=1)$. It is again a natural requirement that the amount of diversity in a mixture of populations should not be smaller than the average of the diversities within the individual populations. We may formulate this requirement (diversity is possibly increased by mixing populations) as

$$C_1\colon H(\lambda_1P_1+\lambda_2P_2)\geqslant[\lambda_1H(P_1)+\lambda_2H(\mathrm{P}_2)] \tag{2.2}$$

with $>$ sign if $P_1\neq P_2$. The condition C_1 is equivalent to saying that H is a strictly concave function or $-H$ is a strictly convex function.

Let us denote $J^{(0)}=-H$ and define

$$J^{(1)}(P_1,P_2\colon\lambda_1,\lambda_2)=\lambda_1J^{(0)}(P_1)+\lambda_2J^{(0)}(P_2)-J^{(0)}(\lambda_1P_1+\lambda_2P_2) \tag{2.3}$$

as the first *Jensen difference* between P_1 and P_2. From the Condition C_1, the first Jensen difference $J^{(1)}$ is positive if the measures P_1 and P_2 are different, and hence may be considered as a measure of dissimilarity (distance) between P_1 and P_2. Now consider pairs of measures (P_{11},P_{12}), (P_{21},P_{22}) and the mixture $(\mu_1P_{11}+\mu_2P_{21},\ \mu_1P_{12}+\mu_2P_{22})$ all belonging to $\mathcal{P}^2=\mathcal{P}\times\mathcal{P}$. It is a natural requirement that the distance between the two mixed populations $\mu_1P_{11}+\mu_2P_{21}$ and $\mu_1P_{12}+\mu_2P_{22}$ should not be larger than the average of the distances between P_{11} and P_{12} and between P_{21} and P_{22}. This requirement (dissimilarity is possibly decreased by mixing) leads to

$$\begin{aligned}C_2\colon&\ \mu_1J^{(1)}(P_{11},P_{12}\colon\lambda_1,\lambda_2)+\mu_2J^{(1)}(P_{21},P_{22}\colon\lambda_1,\lambda_2)\\&-J^{(1)}(\mu_1P_{11}+\mu_2P_{21},\ \mu_1P_{12}+\mu_2P_{22}\colon\lambda_1,\lambda_2)\geqslant0\end{aligned} \tag{2.4}$$

where the left side of (2.4) is defined on $\mathcal{P}^4=\mathcal{P}^2\times\mathcal{P}^2$. We denote this expression by

$$J^{(2)}(\{P_{ij}\}\colon\{\lambda_i\mu_j\}) \tag{2.5}$$

observing that it can be alternatively written as

$$\begin{aligned}&\lambda_1J^{(1)}(P_{11},P_{21}\colon\mu_1,\mu_2)+\lambda_2J^{(1)}(P_{12},P_{22}\colon\mu_1,\mu_2)\\&-J^{(1)}(\lambda_1P_{11}+\lambda_2P_{12},\ \lambda_1P_{21}+\lambda_2P_{22}\colon\mu_1,\mu_2)\end{aligned} \tag{2.6}$$

exhibiting row and column symmetry. The condition C_2 means that the first Jensen difference $J^{(1)}$ is a convex function on $\mathcal{P}^2$.

Generalizing the above concepts, the need for which is demonstrated in Section 6, we lay down a series of conditions

$$C_i\colon J^{(i)}\geqslant0,\ i=1,2,\ \ldots \tag{2.7}$$

where $J^{(i)}$ is defined on $\mathcal{P}^{2^i}$ in a recursive way.

We use the following definitions:

Definition 2.1. A diversity measure H satisfying C_0 *is said to be completely convex* *if* $J^{(0)} = -H_1$, $J^{(1)}$, ... *defined on appropriate spaces are all convex, or the conditions* C_i *are satisfied for all* i.

Definition 2.2. *A diversity measure is said to be j-th order convex if the conditions* $C_0, C_1, \ldots, C_j$ *are satisfied.*

3. Entropy as a Diversity Measure. A number of diversity measures have been introduced through the concept of entropy and information and applied in different areas of research. When $\mathcal{P}$ is the simplex of all multinomial distributions in k cells

$$\mathcal{P} = \{p = (p_1, \ldots, p_k)', p_i \geqslant 0, \Sigma p_i = 1\}, \tag{3.1}$$

some of the well known measures of entropy are

$$H_S(p) = -\Sigma p_i \log p_i, \quad \text{(Shannon (1948))}, \tag{3.2}$$

$$H_\alpha(p) = \frac{1 - \Sigma p_i^\alpha}{\alpha - 1}, \alpha > 0, \alpha \neq 1, \quad \text{(Havrda and Charvát (1967))}, \tag{3.3}$$

$$H_R(p) = \frac{\log \Sigma p_i^\alpha}{1 - \alpha}, \alpha > 0, \alpha \neq 1, \text{(Rényi (1961))}. \tag{3.4}$$

All these measures are non-negative and take the value zero only when one of the p_i is unity and the rest are zero. They all attain the maximum when $p_i = 1/k$ for every i. Thus they satisfy the condition C_0.

It is easy to verify that H_S and H_α for any $\alpha > 0$ satisfy the concavity condition C_1, while H_R satisfies C_1 only for $0 < \alpha < 1$.

Burbea and Rao (1982a, 1982b) have shown: (i) H_S satisfies C_2 but not $C_3, C_4, \ldots$, and (ii) H_α satisfies C_2 for $1 \leqslant \alpha \leqslant 2$ when $k > 2$ and for $\alpha \in [1,2] \cup [3, 11/3]$ when $k = 2$ and does not satisfy $C_3, C_4, \ldots$ except when $\alpha = 2$. It is not known whether H_R satisfies C_2 or not.

It may be noted that the continuous analogues of (3.2)–(3.4) are not necessarily nonnegative functionals and their interpretation as diversity measures poses some difficulties.

4. Quadratic Entropy. Let us consider a measurable space $(\mathcal{X}, \mathcal{B})$ and a function d defined on $\mathcal{X}^2$ such that

$$\begin{aligned} d(X_1, X_2) = d(X_2, X_1) &> 0 \quad \text{if } X_1 \neq X_2 \\ &= 0 \quad \text{if } X_1 = X_2. \end{aligned} \tag{4.1}$$

Using such a nonnegative function (kernel), we may define the diversity of a probability measure P defined on $(\mathcal{X}, \mathcal{B})$ by

$$H(P) = \int d(X_1, X_2) P(dX_1) P(dX_2). \tag{4.2}$$

The motivation for such a definition in a biological context was given in Rao (1982a). The expression (4.2) is the average difference (as defined by 4.1) between two individuals (observations) drawn at random from the population specified by the probability measure P.

The functional (4.2) satisfies the condition C_0 for a diversity measure. The condition C_1 requires

$$J^{(1)}(P_1, P_2 : \lambda_1, \lambda_2) = 2\lambda_1\lambda_2[H(P_1, P_2) - \tfrac{1}{2}H(P_1) - \tfrac{1}{2}H(P_2)] \geqslant 0 \tag{4.3}$$

where $H(P_1, P_2)$ is the average difference between two randomly drawn individuals, one from P_1 and another from P_2. The concavity of H or nonnegativity of $J^{(1)}$ reflects the intuitive requirement that two individuals drawn from different populations are on the average more different than those coming from the same population. The expression

$$D(P_1, P_2) = H(P_1, P_2) - \tfrac{1}{2}[H(P_1) + H(P_2)] \tag{4.4}$$

which provides the excess variability, represents the amount of dissimilarity between the populations defined by P_1 and P_2.

Not all nonnegative kernels, $d(x_1,x_2)$, lead to concave diversity measures $H(P)$ as defined in (4.2). We shall investigate the conditions under which this happens. For this purpose we introduce the concept of a conditionally negative definite (C.N.D.) function studied by Schoenberg (1938), Wells and Williams (1975) and Parthasarathy and Schmidt (1972).

Definition. A nonnegative function $d(\cdot,\cdot)$ on $\mathcal{X}^2$ *is said to be conditionally negative definite (C.N.D.) if*

$$\Sigma_1^n\Sigma_1^n d(X_i,X_j)a_ia_j \le 0 \tag{4.5}$$

for any n and choices of real numbers $a_1, \ldots, a_n$ *such that* $\Sigma\, a_i = 0$. *If* (4.5) *is satisfied for all* $a_1, \ldots, a_n$, *then d is said to be a negative definite function.*

The following is a well known theorem of Schoenberg (1938).

THEOREM 4.1. *Let* $d(\cdot,\cdot)$ *be a nonnegative symmetric function on* $\mathcal{X}^2$. *Then* (i) *d is C.N.D. iff* $\rho = d^{1/2}$ *is a metric (i.e., satisfies the postulates of a distance function), and there is an isometry which embeds the metric space* $(\mathcal{X},\rho)$ *into a Hilbert space.* (ii) *If d is C.N.D., then* d^β *is also C.N.D. for* $0\le\beta\le 1$. (iii) *If* $\mathcal{X}$ *is a set of not more than four points, then for any metric d,* $(\mathcal{X}, d^\beta)$, $0\leqslant\beta\leqslant\frac{1}{2}$, *can be isometrically embedded into the Euclidean space of dimension* 3.

The following theorem is given in Parthasarathy and Schmidt (1972).

THEOREM 4.2. *Let* $d(\cdot,\cdot)$ *be a nonnegative symmetric function on* $\mathcal{X}^2$. *Then the following conditions are equivalent:* (i) *d is C.N.D.* (ii) *For any fixed* $X_0 \in \mathcal{X}$, *the kernel defined by*

$$d_0(X_1,X_2)=d(X_1,X_2)-d(X_1,X_0)-d(X_0,X_2)+d(X_0,X_0) \text{ for every } X_1,X_2$$

is negative definite. (iii) *For every* $t>0$, e^{-td} *is positive definite.*

The following theorem gives the conditions under which the diversity measure (4.2) is concave.

THEOREM 4.3. *The diversity measure H defined in (4.2) is concave if d is a C.N.D. function.*

The result follows by observing (see Rao (1982a, 1982b)) that

$$\begin{aligned} &J^{(1)}(P_1,P_2{:}\,\lambda_1,\lambda_2)\\ &= -\lambda_1\lambda_2\int d(X_1,X_2)[P_1(dX_1)-P_2(dX_1)][P_1(dX_2)-P_2(dX_2)]. \end{aligned}$$

Theorem 4.3 shows that when d is chosen as a conditionally negative definite function, the condition C_1 for the diversity measure H defined in (4.2) is satisfied. What further property should d satisfy in order that $C_2,C_3, \ldots$ hold? Fortunately, no further condition seems to be necessary as demonstrated in the following theorem (see Rao (1982a)).

THEOREM 4.4. *If d is C.N.D., then –H, where H is as defined in* (4.2), *is completely convex, i.e., H satisfies all the conditions* $C_1,C_2, \ldots$.

Consider two different pairs of probability measures $P_{11},P_{12}{:}P_{21},P_{22}$. Then $J^{(2)}$ as defined in (2.4) is seen to be, apart from a constant,

$$= -\int d(X_1,X_2)P_{(1-2)(1-2)}(dX_1)P_{(1-2)(1-2)}(dX_2)$$

where $P_{(1-2)(1-2)} = P_{11}+P_{22}-P_{12}-P_{21}$. Then $J^{(2)}\geqslant 0$, since the total measure of $P_{(1-2)(1-2)}$ is zero and d is C.N.D. Similarly, all higher order Jensen differences are convex, which proves the theorem.

A diversity measure H defined as in (4.2) and completely convex in the sense of Theorem 4.4 is called *quadratic entropy*.

The quadratic entropy seems to have better properties as a measure of diversity than the traditional entropy measures considered in Section 3. Now, we raise the question as to whether a completely convex diversity measure is a quadratic entropy. The following theorem due to Lau (1982) provides the answer in the affirmative, and thus gives a characterization of the quadratic entropy of Rao (1982a).

THEOREM 4.5. *Let $\mathcal{X}$ be a normed topological space and $\mathcal{P}$ be the space of probability measures on $(\mathcal{X},\mathcal{B})$ with weak topology. Let $H: P \rightarrow R_+$ be a continuous function such that* (i) *$H(P)=0$ if P is degenerate, and* (ii) *$-H$ is completely convex. Then there exists a unique C.N.D. function d on $\mathcal{X}^2$ such that*

$$H(P) = \int d(X_1,X_2)P(dX_1)P(dX_2) \tag{4.6}$$

i.e., H is a quadratic entropy.

Note 1. For example, if $\mathcal{X} = R^1$, then

$$d(X_1,X_2) = (X_1-X_2)^2 \text{ for every } X_1,X_2, \in R^1 \tag{4.7}$$

is a C.N.D. function, and the diversity measure associated with it

$$\int (X_1-X_2)^2 P(dX_1)P(dX_2) = 2\sigma_p^2 \tag{4.8}$$

is the variance functional of P. It is well known that the variance functional is completely convex.

Note 2. It follows from Theorem 4.1, result (ii), that if $\mathcal{X} = R^1$,

$$H(P) = \int |X_1-X_2|^\beta P(dX_1)P(dX_2) \tag{4.9}$$

is completely convex for $0 \leq \beta \leq 2$, and in particular the *city block distance* functional

$$\int |X_1-X_2| P(dX_1)P(dX_2) \tag{4.10}$$

is completely convex.

Note 3. If $\mathcal{X}$ is a space of not more than four points, then the quadratic entropy based on any function d which is a metric on $\mathcal{X}$ is completely convex. [Note that this may not be true when $\mathcal{X}$ has more than four points.]

Note 4. Let $\mathcal{X}$ be a space of n points and

$$d(X_1,X_2) = 1, \quad \text{if } X_1 \neq X_2 \quad \text{and} = 0, \ \text{if } X_1 = X_2. \tag{4.11}$$

Then the diversity measure (4.2) based on (4.11) is completely convex.

5. Apportionment of Diversity (APDIV). Biologists are interested in apportioning the total diversity in a population as due to differences between and within subpopulations. A concave diversity measure H is ideally suited for this purpose. If $P_1, \ldots, P_k$ are probability distributions in k subpopulations with prior probabilities $\lambda_1, \ldots, \lambda_k$, then we have the decomposition of the total diversity (T).

$$\begin{aligned} H(\Sigma\, \lambda_i P_i) &= \Sigma\, \lambda_i H(P_i) + J^{(1)}(\{P_i\}:\{\lambda_i\}) \\ T \ &= \qquad W \quad + \quad B \end{aligned} \tag{5.1}$$

where the components W and B are nonnegative. In (5.1), W is the average diversity within subpopulations and B may be interpreted as the diversity between the subpopulations. The ratio

$$G = \frac{B}{B+W} \tag{5.2}$$

called the index of diversity between subpopulations compared to the total has been used in genetic studies (see for instance Lewontin (1972), Nei (1973) and Chakraborty (1974)). Different diversity measures give different values of G, which raises the problem of choosing an appropriate measure in practical applications. For a discussion of this problem and some illustrative examples, the reader is referred to Rao (1982a) and Rao and Boudreau (1982). It is seen that if the object is APDIV (apportionment of diversity), we need only a concave diversity measure, i.e., one which satisfies only the condition C_1.

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 apriori probabilities are known, then the distributions of groups at any level of classification and the associated apriori probabilities can be computed. This would enable us to compute the average diversity within groups at any level of classification. Then we have the apportionment of the total diversity H_0 (for all populations mixed together) as is shown in Table 1.

TABLE 1. APDIV for a hierarchical classification

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

The only property required of a diversity measure for APDIV is concavity.

Note 1. It is interesting to note that if we use a quadratic entropy for APDIV, the decomposition (5.1) can be written as

$$H(\Sigma\,\lambda_i P_i) = \Sigma\,\lambda_i H(P_i) + \Sigma\Sigma\,\lambda_i\lambda_j D_{ij} \tag{5.3}$$

where

$$\begin{aligned} D_{ij} &= 2H\left(\frac{P_i+P_j}{2}\right) - H(P_i) - H(\mathrm{P_j}) \\ &= H(P_i,P_j) - \tfrac{1}{2}H(P_i) - \tfrac{1}{2}H(P_j) \end{aligned}$$

is the dissimilarity between the populations i and j. The second term on the right hand side of (5.3) is the average dissimilarity between populations. Such an interpretation is available only if a quadratic entropy is used.

Note 2. Let $P_1 = N(\mu_1,\sigma^2)$ and $P_2 = N(\mu_2,\sigma^2)$. If we use the variance functional (4.8), then $H(P_1) = 2\sigma^2 = H(P_2)$ and $H(P_1,P_2) = (\mu_1-\mu_2)^2+2\,\sigma^2$ so that the dissimilarity between P_1 and P_2 is

$$D_{12} = H(P_1,P_2) - \tfrac{1}{2}H(P_1) - \tfrac{1}{2}H(P_2) = (\mu_1-\mu_2)^2 = \delta^2. \tag{5.4}$$

On the other hand, if we use the city block functional (4.10)

$$H(P_1) = 2\sigma/\sqrt{\pi} = H(P_2)$$

$$H(P_1,P_2) = (2\sigma/\sqrt{\pi})e^{-\delta^2/4\sigma^2} + 2\delta\phi(-\delta/\sigma\sqrt{2}) - \delta$$

and the dissimilarity is

$$D_{12} = (2\sigma/\sqrt{\pi})e^{-\delta^2/4\sigma^2} - 1) + 2\delta\phi(-\delta/\sigma\sqrt{2}) - \delta \tag{5.5}$$

6. ANODIV: Generalization of ANOVA. Apportionment of diversity corresponds to analysis of variance (ANOVA) of one-way classified data, where the populations are identified by the levels of a single factor. Let us now consider two factors A_1 and A_2 and represent the probability distribution associated with the i-th level of A_1 and the j-th level of A_2 by P_{ij} with apriori probability $\lambda_i^{(1)}\lambda_j^{(2)}$, $i=1, \dots, p$; $j=1, \dots, q$ ($\Sigma\, \lambda_i^{(1)}=1$, $\Sigma\, \lambda_j^{(2)} = 1$). Define

$$P_{..} = \Sigma\Sigma\, \lambda_i^{(1)}\lambda_j^{(2)}P_{ij}$$

$$P_{i.} = \Sigma_j\lambda_j^{(2)}P_{ij},\ P_{.j} = \Sigma_i\lambda_i^{(1)}P_{ij}$$

where $P_{..}$ is the overall distribution and $P_{i.}$ and $P_{.j}$ are the marginal distributions for the levels of the individual factors A_1 and A_2 respectively. Consider the analysis of diversity (ANODIV), i.e., a decomposition of the overall diversity, $H(P_{..})$, as in Table 2, using any measure of diversity.

TABLE 2. ANODIV for two way data

row no.	*due to*	*diversity*
1	factor(A_1)	$J^{(1)}(\{P_{i.}\}:\{\lambda_i^{(1)}\})$
2	factor(A_1)	$J^{(1)}(\{P_{.j}\}:\{\lambda_j^{(2)}\})$
3	interaction (A_1A_2)	$J^{(2)}(\{P_{ij}\}:\{\lambda_i^{(1)}\lambda_j^{(2)}\})$
4	between populations	$J^{(1)}(\{P_{ij}\}:\{\lambda_i^{(1)}\lambda_j^{(2)}\})$
5	within populations	$\Sigma\Sigma\, \lambda_i^{(1)}\lambda_j^{(2)}H(P_{ij})$
6	Total	$H(P_{..})$

Rows (4) and (5) provide the analysis of diversity as between and within populations defined by the pq cells of the two way classification.

Rows (1) and (2) measure the diversities in the marginal distributions or the main effects of the factors A_1 and A_2. The residual diversity in row (3) represents the interaction between the factors A_1 and A_2. Thus the rows (1)-(3) and (5) provide an analysis of (6), the total diversity, as assignable to different causes. For a two way ANODIV, the diversity measure (H) need only satisfy the conditions C_1 and C_2 to ensure that $J^{(1)}$ and $J^{(2)}$ representing the main effects and interation are nonnegative.

If we have three way data, we can obtain a similar decomposition. The main effects are computed from one factor marginal distributions, two factor interactions from two factor marginal distributions, while the three factor interaction, the third order Jensen difference, is obtained by the formula

$$\begin{aligned} & J^{(3)}(\{P_{ijk}\} : \{\lambda_i^{(1)}\lambda_j^{(2)}\lambda_k^{(3)}\}) \\ & = H(P_{...}) - \Sigma\Sigma\Sigma\, \lambda_i^{(1)}\lambda_j^{(2)}\lambda_k^{(3)}H(P_{ijk}) \\ & -(A_1)-(A_2)-(A_3)-(A_1A_2)-(A_2A_3)-(A_3A_1) \end{aligned} \tag{6.1}$$

as in the case of analysis of variance with balanced data.

In order that all main effects and interactions are nonnegative in the ANODIV of three way classified data it is necessary to choose a diversity measure for which $J^{(1)}$, $J^{(2)}$ and $J^{(3)}$ are nonnegative, i.e., which satisfies the conditions C_1, C_2 and C_3. The results are easily generalized to analysis of m-way data for which we need a diversity measure whose Jensen differences up to order m are nonnegative. Note that $J^{(m)}$ represents the m factor interaction.

The generalization of ANOVA to ANODIV suggested above is quite general and can be applied on *any type of data* with observations *in any space,* by choosing a diversity measure with the ***appropriate order*** of convexity. Thus, if we have two way classified qualitive data, Shannon's entropy could be used, but not for higher order classified data.

7. Sampling Problems. In Section 6, we have discussed ANODIV in terms of population distributions, which provided various diversity components. In practice we have only observations from different populations, in which case we have problems of estimating the diversity components and testing 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 data as in Table 3.

TABLE 3. Populations and observations

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

We estimate the probability distribution function F_i for the i-th population by the empirical distribution function $\hat{F}_i$ based on the observations $x_{i1}, \dots, x_{in_i}$. Let us choose $\lambda_i = n_i/n.$ (where $n. = \Sigma n_i$) as the apriori probability of the i-th population. Further let us consider the variance functional

$$H(F) = \int (X_1 - X_2)^2 dF(X_1) dF(X_2) \tag{7.1}$$

as the diversity measure. Substituting $\hat{F}_i$ for F_i *and* $\lambda_i = n_i/n.$ in the basic decomposition formula (5.1), we have

$$H(\Sigma(n_i/n.)\hat{F}_i) = \Sigma(n_i/n.)H(\hat{F}_i) + J^{(1)}(\{\hat{F}_i\}:\{\lambda_i\}). \tag{7.2}$$

Computing the various expressions in (7.2), using (7.1) for the H function, we obtain

$$n.^{-1}\Sigma_i\Sigma_j(x_{ij} - \bar{x}_{..})^2 = \Sigma_i(n_i/n.)n_i^{-1}\Sigma_j(x_{ij} - \bar{x}_{i.})^2 + n.^{-1}\Sigma_i n_i(\bar{x}_{i.} - \bar{x}_{..})^2 \tag{7.3}$$

which is the usual ANOVA as within and between populations. The decomposition (7.3) is used in testing the hypothesis that the populations are the same, and in estimating the magnitude of differences between populations when the null hypothesis is rejected.

Instead of the variance functional, we can also use the city block distance functional

$$H(F) = \int |X_1 - X_2| dF(X_1) dF(X_2). \tag{7.4}$$

This leads to the decomposition

$$n_{.}^{-2}\Sigma_{ij}\Sigma_{rs}|x_{ij}-x_{rs}| = \Sigma_i(n_i/n.)n_i^{-2}\Sigma_r\Sigma_s|x_{ir}-x_{is}| + J^{(1)}\mathrm{L}(\{\hat{F}_i\}:\{\lambda_i\})$$

which could provide valid tests of significance for non-normal populations.

The ANODIV for one way classified categorical data (two way contingency table) using the second order entropy $(1 - \Sigma\, p_i^2)$, also called the Gini-Simpson index, is already illustrated in a paper by Light and Margolin (1971). We shall extend the analysis to two way classified categorical data using a more general quadratic entropy for a multinomial distribution with $p_1, \ldots, p_k$ as cell probabilities

$$H(p) = \Sigma\Sigma\, p_i p_j d_{ij} \tag{7.5}$$

where d_{ij} are chosen such that the $(k-1) \times (k-1)$ matrix

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

is nonnegative definite to ensure complete convexity of the diversity measure (7.5). For further details and a characterization of (7.5), reference may be made to Rao (1982c).

Let us represent the observed numbers for k different categories in the (i,j)-th cell by n_{ijr}, $r = 1, \ldots, k$, and the estimated probabilities by $p_{ijr} = n_{ijr}/n_{ij.}$ where $n_{ij.} = n_{ij1}+\ldots+n_{ijk}$. If the cell numbers satisfy the conditons

$$n_{ij.} = n_{...}\lambda_i^{(1)}\lambda_j^{(2)},\ \lambda_i^{(1)} = n_{i..}/n_{...},\ \lambda_j^{(2)} = n_{.j.}/n_{...} \tag{7.6}$$

then we can obtain the ANODIV using the diversity measure (7.5) as shown in Table 4.

TABLE 4. ANODIV: Two way categorical data

due to	*diversity*
factor (A_1)	$\Sigma\Sigma\, d_{rs}p_{..r}p_{..s} - \Sigma\, \lambda_i^{(1)}\Sigma\Sigma\, d_{rs}p_{i.r}p_{i.s}$
factor (A_2)	$\Sigma\Sigma\, d_{rs}p_{..r}p_{..s} - \Sigma\, \lambda_j^{(2)}\Sigma\Sigma\, d_{rs}p_{.jr}p_{.js}$
interaction $(\mathrm{A}_1 A_2)$	* (by subtraction)
between populations	* (by subtraction)
within populations	$\Sigma\Sigma\, \lambda_i^{(1)}\lambda_j^{(2)}\Sigma\Sigma\, d_{rs}p_{ijr}p_{ijs}$
total	$\Sigma\Sigma\, d_{rs}p_{..r}p_{..s}$

If $n_{ij.}$ do not satisfy the conditions (7.6), we can still carry out the ANODIV by choosing appropriate values of $\lambda_i^{(1)}$ and $\lambda_j^{(2)}$ (see Rao (1982c) for example).

The sampling distributions of the various expressions in Table 4 are likely to be complicated even in large samples. Their use in tests of significance and estimation of diversity components is under investigation.

REFERENCES

BURBEA, J. and RAO, C. R. (1982a). On the convexity of some divergence measures based on entropy functions. *IEEE Trans. Inf. Theory 28* 489–495.

BURBEA, J. and RAO, C. R. (1982b). On the convexity of higher order Jensen differences based on entropy functions. *IEEE Trans. Inf. Theory 28* 961–963.

CHAKRABORTY, R. (1974). A note on Nei's measure of gene diversity in a substructured population. *Humangenetik 21* 85–88.

DENNIS, B., PATIL, G. P., ROSSI, O. 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.

HAVRDA, J. and CHARVÁT, F. (1967). Quantification method in classification processes: concept of structural α-entropy. *Kybernetika 3* 30–35.

KARLIN, S., KENNETT, 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. *Ann. J. Hum. Genet. 31* 341–365.

LAU, K.-S. (1983). Characterization of the quadratic entropies of Rao. *Sankhyā Ser.* (in press).

LEWOTIN, 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. Statist. Assoc. 66* 534–544.

NEI, M. (1973). Analysis of gene diversity in subdivided populations. *Proc. Nat. Acad. Sci. 70* 3321–3323.

PARTHASARATHY, K. R. and SCHMIDT, K. (1972). *Positive Definite Kernels, Continuous Tensor Products and Central Limit Theorems of Probability Theory*. Springer-Verlag, Lecture Notes in Mathematics, 272.

RAO, C. RADHAKRISHNA (1982a). Diversity and dissimilarity coefficients: a unified approach. *Theoret. Pop. Biol. 21* 24–43.

RAO, C. RADHAKRISHNA (1982b). Diversity: its measurement, decomposition, apportionment and analysis. *Sankhyā Ser. A. 44* 1–22.

RAO, C. RADHAKRISHNA (1982c). Gini-Simpson index of diversity: a characterization, generalization and applications. *Utilitas Mathematica 218* 273–282.

RAO, C. RADHAKRISHNA and BOUDREAU, R. (1982). Diversity and cluster analyses of blood group data on some human populations. *Tech. Report 82–35,* Department of Mathematics and Statistics, University of Pittsburgh, Pittsburgh, PA.

RÉNYI, A. (1961). On measures of entropy and information. *Proc. Fourth Berkeley Symposium 1* 547–561.

SCHOENBERG, I. J. (1938). Metric spaces and positive definite functions. *Trans. Amer. Math. Soc. 44* 522–536.

SHANNON, C. E. (1948). A mathematical theory of communications. *Bell System Tech J. 27* 379–423, 623–656.

WELLS, J. and WILLIAMS, L. (1975). *Embeddings and Extensions in Analysis*. Springer-Verlag, Lecture Notes in Mathematics.