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.