MATHEMATICAL FORMULATION OF MULTIVARIATE EUCLIDEAN MODELS FOR DISCRIMINATION METHODS

KENNETH MULLEN
DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF GULEPH

DANIEL M. ENNIS
PHILIP MORRIS RESEARCH CENTER

Multivariate models for the triangular and duo-trio methods are described in this paper. In both cases, the mathematical formulation of Euclidean models for these methods is derived and evaluated for the bivariate case using numerical quadrature. Theoretical results are compared with those obtained using Monte Carlo simulation which is validated by comparison with previously published theoretical results for univariate models of these methods. This work is discussed in light of its importance to the development of a new theory for multidimensional scaling in which the traditional assumption can be eliminated that proximity measures and perceptual distances are monotonically related.

Key words: triangular, duo-trio, multidimensional scaling, proximity measures.

Introduction

Since the pioneering work of Shepard (1962, 1963) and Kruskal (1964a, 1964b), the theory of multidimensional scaling has rested squarely on the assumption that proximity measures and perceptual distances are monotonically related. Direct similarity scaling, confusion matrices, and techniques for grouping stimuli have led to different approaches to providing proximity measures. Under the assumption that proximities and distances are related monotonically, or even linearly, spatial representations of stimuli in n-dimensional space have been estimated based on different assumptions about the distance metric (Euclidean, city block and other Minkowski γ-metrics.)

In attempting to understand the relationship between certain multivariate parameters and the subject's response to tasks which involve grouping stimuli on the basis of their similarity, it has been shown by the present authors that the monotonicity assumption will be invalid under certain multivariate conditions (Ennis & Mullen, 1986a). Other proximity measures, such as the probability of confusing one stimulus with another, should depend on the same multivariate parameters in a similar fashion. Multivariate parameter values for which proximity measures are monotonic functions of perceptual distances, then, are a particular subset of a larger set of possible multivariate specifications. A knowledge of the relationship between proximity measures and perceptual distances as determined by different multivariate scenarios will lead to a new, and more
general, theory for multidimensional scaling where the monotonicity assumption will not be needed.

Most of the work leading to these ideas was based on Monte Carlo simulation of multivariate models based on Thurstonian concepts. The goal of this paper is to present the framework for developing an analytical approach to computing the proximity measure, the probability of a correct response, from multivariate parameters using the Euclidean metric for sensory discrimination methods. An analytical solution is found for two of the discrimination methods, the triangular and duo-trio methods, in the bivariate case and the basis for the n-dimensional case is derived, but not evaluated.

Aside from their potential use in mapping stimuli with multivariate sensory attributes, sensory discrimination methods are used to study the discriminability of pairs of univariate and multivariate stimuli. In fact, it was because multivariate theory for these commonly used methods is nonexistent that the present authors initiated research in this area. The triangular and duo-trio methods are widely used in sensory food research (Amerine, Pangborn, & Roessler, 1965). In the triangular method, the subject is instructed to select out of three stimuli (two randomly drawn from one stimulus set and one from another stimulus set) the one which is perceptually different from the other two. In the duo-trio method, the subject is instructed to select out of two stimuli (each drawn randomly from separate stimulus sets) the one which is identical to or different from a third stimulus which was drawn from one of the stimulus sets (the standard). The standard may be presented before or after the two unknown stimuli. Frijters (1979) discussed much of the earlier work on univariate models for the triangular and duo-trio methods (Ura, 1960; David & Trivedi, 1962) and developed the psychometric basis on which the triangular method could be introduced as a signal detection theory method for stimuli with univariate momentary sensory values.

Evaluation of the multivariate Euclidean model for the triangular method using Monte Carlo stimulation (Ennis & Mullen, 1985, 1986a, 1986b) showed that the relationship between the probability of a correct response, $P_c$, and the sensory distance, $\delta$, between the stimulus sets, depended on: (i) the number of sensory dimensions of the momentary sensory values, (ii) the correlation structure of the sensory dimensions for each stimulus object, (iii) the relative orientations of the stimulus objects to each other in a multidimensional space and (iv) the relative magnitudes of the variances of the momentary sensory values on the different dimensions.

The Triangular Method

The Multivariate Model: Assumptions

The multivariate model can be applied under experimental conditions identical to those given by Frijters (1979) without the constraint that the stimuli possibly vary with respect to one particular sensory attribute only. The assumptions underlying the model are:

1. The stimuli $S_{x_1}$, $S_{x_2}$ and $S_y$ give rise to corresponding sensory values of the respective magnitudes $x_1$, $x_2$ and $y$, where $x_1' = (x_{11}, x_{12}, \ldots, x_{1n})$, $x_2' = (x_{21}, x_{22}, \ldots, x_{2n})$ and $y' = (y_1, y_2, \ldots, y_n)$ where $x_i'$ indicates the transpose of the vector, $x_i$, and where $n$ is the number of sensory dimensions. The momentary sensory values are mutually independently distributed, with $x_1$ and $x_2$ having density functions $f(x_1)$ and $f(x_2)$, and $y$ having density function $f(y)$.

2. The probability densities $f(x)$ and $f(y)$ are multivariate normal distributions with means $\mu_x$ and $\mu_y$ (where $\mu_x' = (\mu_{x1}, \mu_{x2}, \ldots, \mu_{xn})$ and $\mu_y' = (\mu_{y1}, \mu_{y2}, \ldots, \mu_{ym})$) and variance-