THE MULTIVARIATE ANALYSIS RESEARCH GROUP

Carles M Cuadras
Departament d’Estadística
Facultat de Biologia
Universitat de Barcelona

The set of statistical methods known as Multivariate Analysis covers a wide group of theoretical and applied methods, including factor analysis, classification, multivariate analysis of variance and multidimensional scaling. Since 1980, and especially since 1989, a group of researchers in Barcelona has been studying geometric aspects of statistics, with applications in statistical inference, regression and multivariate analysis.

Distance-based regression: In this approach a response variable is predicted using several explanatory variables on quantitative and qualitative measurement scales. The key idea is to define a distance between observations and to project the response variable on the principal dimensions obtained via multidimensional scaling. This distance-based (DB) model, which may be named «principal coordinate regression», generalizes and improves the multiple regression model, as well as the non-linear regression model by using suitable distance functions (Cuadras, 1989; Cuadras and Arenas, 1990; Cuadras, Arenas and Fortiana, 1996). Related results can be found in Cuadras (1993) and Fortiana and Cuadras (1997).

Distance-based discriminant analysis: The classic problem of allocating an observation to two or more known groups, can also be solved by using the DB approach. Using only distance matrices, one for each group, we can define a proximity function which reduces to the classic discriminant function for particular distances, such as Euclidean or Mahalanobis. This method allows us to handle nominal variables and missing data, and approach the problem of typicality in classification (Cuadras, 1989; Cuadras, 1992; Cuadras, Fortiana and Oliva, 1997; Cuadras, Atkinson and Fortiana, 1997; Cuadras and Fortiana, 2000; Villarroya, Rios and Oller, 1995).

Related metric scaling: A natural extension of this DB approach is to define a joint distance as a function of two given distances on the same set, which satisfies some specific rules (e.g., additivity in the case of independence) and preserves the redundancy between both distances. This approach allows us to relate distances and to represent multivariate data under two different kinds of information (Cuadras and Fortiana, 1995, 1996; Cuadras, 1998; Arenas et al., 2000).

Principal components of a random variable: Just as we can obtain the principal components of a finite set of variables, we can define and obtain the principal directions of a Bernoulli process associated with a continuous random variable. This construction
is useful in goodness-of-fit assessment, in expanding a random variable in terms of its principal components, in constructing bivariate distributions with given marginals and in studying the asymptotic distribution of some statistics related to Rao’s quadratic entropy (Cuadras and Fortiana, 1995; Cuadras and Lahlou, 2000).

**Distributions with given marginals:** The construction of joint distributions given the marginals and some dependence parameters, is of great interest in probability and statistics. Families of distributions have been obtained when an interdependence matrix between marginals is given and when the regression curve is given. A continuous extension of correspondence analysis has also been derived (Cuadras and Auge, 1981; Ruiz-Rivas and Cuadras, 1988; Cuadras, 1992; Cuadras, 1996; Cuadras and Fortiana, 1997; Cuadras, Fortiana and Greenacre, 2000; Cuadras, 2002).

**Differential geometry in statistics:** If we interpret a statistical model as a Riemannian manifold, and define a distance between parameters through geodesics by using the Fisher information matrix as the metric tensor, we obtain the Rao distance, a natural extension of the Mahalanobis distance. This allows us to study the geometry of any regular statistical model and approach some problems in statistical inference, such as intrinsic estimation, invariance, testing of hypotheses and representing parametric models. (Oller and Cuadras, 1985; Oller, 1989; Calvo and Oller, 1990; Oller and Corcuera, 1995; Rios, Villarroya and Oller, 1992; Villarroya and Oller, 1993; Villarroya, Rios and Oller, 1995; Garcia and Oller, 2001).

**References**


