## Publicacions més rellevants de la línia de recerca:
# Estadística basada en distàncies i aplicacions a la bioinformàtica

**Referència:** Cuadras. C. M. and Cuadras, D. A parametric approach to correspondence analysis. *Linear Algebra and its Applications*, **417 (1),** (2006), 64-74.

**Abstract:** Correspondence analysis on two categorical variables is usually formulated on a two-way contingency table of frequencies. Assuming the variables to be ordinal and re-expressing the canonical correlation analysis, we show that correspondence analysis may also be approached in terms of the cumulative frequencies. A continuous extension is made. It is shown that the diagonal expansion of a bivariate density may be formulated using integral operators on the cumulative distribution function.

---

**Referència:** Sanchez, A., Salicru, M. and Ocaña, J. Statistical methods for the analysis of high-throughput data based on functional profiles derived from the Gene Ontology. *Journal of Statistical Planning and Inference*, **137 (12)**, (2007), 3975-3989.

**Abstract:** The increasing availability of high-throughput data, that is, massive quantities of molecular biology data arising from different types of experiments such as gene expression or protein microarrays, leads to the necessity of methods for summarizing the available information. As annotation quality improves it is becoming common to rely on biological annotation databases, such as the Gene Ontology (GO), to build functional profiles which characterize a set of genes or proteins using the distribution of their annotations in the database. In this work we describe a statistical model for such profiles, provide methods to compare profiles and develop inferential procedures to assess this comparison. An R-package implementing the methods will be available at publication time.

---

Garcia, G. and Oller, J. M. On the smoothness of projections in statistical models. *Communications in Statistics-Theory and Methods*, **35 (8)**, (2006), 1375-1385.

**Abstract:** In this paper we review different meanings of the word intrinsic in statistical estimation, focusing ourattention on the use of this word in the analysis of the properties of an estimator. We

review the intrinsic versions of the bias and the mean square error and results analogous to the Cram´er-Rao inequality and Rao-Blackwell theorem. Different results related to the Bernoulli and normal distributions are also considered.