

# “Visualization of multivariate data: What is a valid pattern?”

## “Application to textual data”

**Ludovic LEBART**

(Telecom-ParisTech, Paris) (ludovic@lebart.org)

### ***Abstract:***

We show that several types of resampling techniques could be performed to assess the quality of the visualisations provided by multivariate descriptive methods. We limit ourselves here to those methods that involve a singular values decomposition (mainly: Principal Components Analysis, Simple and Multiple Correspondence analysis).

a) The simplest resampling scheme is certainly the *partial bootstrap*, which considers the replicates as supplementary variables, without new diagonalization of the replicated moment-product matrices.

b) Much more conservative, the *total bootstrap type 1* performs a new diagonalization for each replicate, with corrections limited to possible reflections of the axes.

c) *Total bootstrap type 2* adds to the preceding one a correction for the possible exchanges of axes.

d) *Total bootstrap type 3*, implies procrustean transformations of all the replicates striving to take into account both rotations and exchanges of axes. In the case of textual data, several levels of statistical units may coexist: the *specific bootstrap* implies a resampling at different levels (case of a hierarchy of statistical units: texts, responses to open-ended questions, sentences, context units, words).

Examples are presented and discussed for each technique.

### ***Keywords:***

Bootstrap, perturbation theory, validation, SVD (Singular Values Decomposition), CA (Correspondence Analysis), PCA (Principal Component Analysis), SOM (Self Organizing Maps).