Scientific contributions in 2004

Stéphane Girard

Inria Grenoble Rhône-Alpes & LJK (team MISTIS).

655, avenue de l'Europe, Montbonnot. 38334 Saint-Ismier Cedex, France

Stephane.Girard@inria.fr

Abstract

This short report describes my scientific contributions in 2004. Three main research domains are explored: Boundary or frontier estimation, High dimensional statistical learning and Extreme-value analysis.

1 Boundary or frontier estimation

Boundary or frontier estimation, and more generally, level sets estimation, are recurrent functional estimation problems in statistics which are linked to outlier detection. In biology, one is interested in estimating reference curves, that it to say curves which bound 90% (for instance) of the population. Points outside this bound are considered as outliers compared to the reference population. Here, reference curves are computed through nonparametric regression quantile estimations [1, 2, 3].

In image analysis, the boundary estimation problem arises in image segmentation as well as in supervised learning. Two different and complementary approaches are developed. In the extreme quantiles approach the boundary bounding the set of points is viewed as the larger level set of the points distribution. Its estimation is thus an extreme quantile curve estimation problem. Estimators based on projection as well as on kernel regression methods are applied on the extreme values set [4, 5].

Besides, the use of optimization techniques permits to select automatically the relevant points from all the observations of the sample [6] similarly to the methods used in Support Vector Machines (SVM).

2 High dimensional statistical learning

I have proposed a parametrization of the Gaussian mixture model for classification purposes. It is assumed that the high-dimensional data live in subspaces with intrinsic dimensions smaller than the dimension of the original space and that the data of different classes live in different subspaces with different intrinsic dimensions. New high-dimensional data classifiers are introduced on the basis of this model [7].

Another aspect of multivariate data analysis is the modeling of dependence between variables. Copula provides a relevant tool to build multivariate probability laws, from fixed marginal distributions and required degree of dependence. From Sklar's Theorem, the dependence properties of a continuous multivariate distribution can be entirely summarized, independently of its margins, by a copula. We have introduced a new semiparametric family of bivariate copulas. The family is generated by a univariate function, determining the symmetry (radial symmetry, joint symmetry) and dependence property (quadrant dependence, total positivity, ...) of the copulas. An estimation procedure has also been introduced [8].

3 Extreme-value analysis

The decay of the survival function $P(X > x)$ is driven by a real parameter called the extreme-value index. When this parameter is positive, the survival function is said to be heavy-tailed. When this parameter is negative, the survival function vanishes above its right end-point. If this parameter is zero, then the survival function decreases to zero at an exponential rate. An important part of my work is dedicated to the study of such distributions [9]. For instance, in reliability, the distributions of interest are included in a semi-parametric family whose tails are decreasing exponentially fast. These so-called Weibull-tail distributions include Gaussian, gamma, exponential and Weibull distributions, among others. I also introduced an estimator of the extreme-value index valid in all cases [10, 11, 12]. The proposed methods have been included in a software [13] freely available from my webpage.

References

[1] A. Gannoun, S. Girard, C. Guinot, and J. Saracco. Sliced inverse regression in reference curves estimation. *Computational Statistics and Data Analysis*, 46(1):103–122, 2004.

[2] A. Gannoun, S. Girard, C. Guinot, and J. Saracco. Estimations non-paramétriques et semi-paramétriques de courbes de référence pour l'analyse de propriétés biophysiques de la peau. In *8ème congrès Agro-industrie et Méthodes Statistiques*, Rennes, mars 2004.

[3] A. Gannoun, S. Girard, C. Guinot, and J. Saracco. Implémentation en C d'estimateurs non-paramétriques de quantiles conditionnels. Application au tracé de courbes de référence. *La revue de Modulad*, 31:59–70, 2004.

[4] S. Girard and P. Jacob. Extreme values and kernel estimates of point processes boundaries. *ESAIM: Probability and Statistics*, 8:150–168, 2004.

[5] S. Girard. On the asymptotic normality of the L_1- error for Haar series estimates of Poisson point processes boundaries. *Statistics and Probability Letters*, 66:81–90, 2004.

[6] G. Bouchard, S. Girard, A. Iouditski, and A. Nazin. Nonparametric frontier estimation by linear programming. *Automation and Remote Control*, 65(1):58–64, 2004.

[7] C. Bouveyron, S. Girard, and C. Schmid. Dimension reduction and classification methods for object recognition in vision. In *5th French-Danish Workshop SSIAB*, pages 109–113, Saint-Pierre de Chartreuse, mai 2004.

[8] C. Amblard and S. Girard. Estimation procedures for a semiparametric family of bivariate copulas. In *Conference on Dependence Modelling: Statistical Theory and Applications in Finance and Insurance*, Québec, Canada, mai 2004.

[9] S. Girard. A Hill type estimate of the Weibull tail-coefficient. *Communication in Statistics - Theory and Methods*, 33(2):205–234, 2004.

[10] L. Gardes and S. Girard. A Pickands type estimator of the extreme value index. In *Third International Symposium on Extreme Value Analysis*, page 28, Aveiro, Portugal, juillet 2004.

[11] L. Gardes and S. Girard. A Pickands type estimator of the extreme value index. In *Workshop on Power laws in probability and statistics*, CIRM, Luminy, mars 2004.

[12] L. Gardes and S. Girard. Un estimateur de l'indice de valeur extrême de type Pickands. In *36èmes Journées de Statistique organisées par la Société Française de Statistique*, Montpellier, mai 2004.

[13] M. Garrido, S. Girard, and J. Ecarnot. The Extremes software. In *Third International Symposium on Extreme Value Analysis*, page 27, Aveiro, Portugal, juillet 2004.