Scientific outputs in 2005

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 report summarizes my scientific contributions in 2005. Three main research topics are addressed: Boundary or frontier estimation, High dimensional statistical learning and Extreme-value analysis.

1 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. In this context, I proposed a Bayesian procedure to estimate it [1]. 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 [2, 3, 4, 5]. 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 [6]. The proposed methods have been included in a software [7] freely available from my webpage.

2 Boundary or frontier estimation

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 [8]. Estimators based on projection as well as on kernel regression methods are applied on the extreme values set [9, 10].

Besides, the use of optimization techniques permits to select automatically the relevant points from all the observations of the sample [11, 12] similarly to the methods used in Support Vector Machines (SVM).

3 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 [13, 14, 15, 16].

I also developped dimension reduction methods for high dimensional regression problems [17].

Image analysis and computer vision are two important application domains for high dimensional data analysis and, more precisely, for dimension reduction methods. Indeed, a $M \times M$ grey-level image can be represented as a p-dimensional vector with $p = M^2$ or by a set of local descriptors. In both case, even with moderate image sizes, one obtains data living in very high-dimensional spaces. Principal Component Analysis (PCA) is usually an efficient tool for reducing the dimension of such data. However, even simple transformations between images can yield strong non-linearities in the p-dimensional space and thus strongly reduce the PCA efficiency. To overcome this problem, I have introduced Auto-Associative models allowing to build new nonlinear dimension reduction methods. The dataset is approximated by a differentiable manifold generalizing PCA's linear subspaces. The approximation algorithm is simple: it consists in incrementing the dimension of the manifold step by step [18].

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 [19]

References

[1] J. Diebolt, M. El-Aroui, M. Garrido, and S. Girard. Quasi-conjugate Bayes estimates for GPD parameters and application to heavy tails modelling. *Extremes*, 8:57–78, 2005.

[2] L. Gardes and S. Girard. Estimating extreme quantiles of Weibull tail-distributions. *Communication in Statistics - Theory and Methods*, 34:1065–1080, 2005.

[3] L. Gardes and S. Girard. Statistical inference for Weibull-tail distributions. In *Workshop on risk analysis and extreme values*, Paris, juin 2005.

[4] L. Gardes and S. Girard. Estimating extreme quantiles of Weibull-tail distributions. In *STATDEP, Statistics for dependent data*, Paris-Malakoff, janvier 2005.

[5] L. Gardes and S. Girard. Inférence statistique pour les lois à queue de type Weibull. In *37èmes Journées de Statistique organisées par la Société Française de Statistique*, Pau, juin 2005.

[6] L. Gardes and S. Girard. Asymptotic distribution of a Pickands type estimator of the extreme value index. *Comptes-Rendus de l'Académie des Sciences, Série I*, 341:53–58, 2005.

[7] J. Diebolt, M. Garrido, S. Girard, and J. Ecarnot. The Extremes software. In *Fourth International Symposium on Extreme Value Analysis*, Gothenburg, Suède, aout 2005.

[8] L. Gardes and S. Girard. Estimation d'une fonction quantile extrême. In *3èmes Journées de Statistique Fonctionnelle et Opératorielle*, pages 25–27, Toulouse, juin 2005.

[9] S. Girard and L. Menneteau. Central limit theorems for smoothed extreme value estimates of point processes boundaries. *Journal of Statistical Planning and Inference*, 135(2):433–460, 2005.

[10] S. Girard and P. Jacob. Asymptotic normality of the L_1-error for Geffroy's estimate of Poisson point process boundaries. *Publications de l'Institut de Statistique de l'Université de Paris*, XLIX:3–17, 2005.

[11] G. Bouchard, S. Girard, A. Iouditski, and A. Nazin. Some linear programming methods for frontier estimation. *Applied Stochastic Models in Business and Industry*, 21(2):175–185, 2005.

[12] S. Girard, A. Iouditski, and A. Nazin. L_1-optimal nonparametric frontier estimation via linear programming. *Automation and Remote Control*, 66(12):2000–2018, 2005.

[13] C. Bouveyron, S. Girard, and C. Schmid. Classification of high dimensional data: High dimensional discriminant analysis. In *Subspace, latent structure and feature selection techniques: statistical and optimisation perspectives workshop*, Bohinj, Slovénie, février 2005.

[14] C. Bouveyron, S. Girard, and C. Schmid. High dimensional discriminant analysis. In *International Symposium on Applied Stochastic Models and Data Analysis*, Brest, mai 2005.

[15] C. Bouveyron, S. Girard, and C. Schmid. Une nouvelle méthode de classification pour la reconnaissance de formes. In *Vingtième colloque GRETSI*, Louvain-la-Neuve, Belgique, septembre 2005.

[16] C. Bouveyron, S. Girard, and C. Schmid. Une méthode de classification des données de grande dimension. In *37èmes Journées de Statistique organisées par la Société Française de Statistique*, Pau, juin 2005.

[17] A. Gannoun, S. Girard, C. Guinot, and J. Saracco. Reference curves estimation via sliced inverse regression. In *International Symposium on Applied Stochastic Models and Data Analysis*, Brest, mai 2005.

[18] S. Girard and S. Iovleff. Auto-associative models and generalized principal component analysis. *Journal of Multivariate Analysis*, 93(1):21–39, 2005.

[19] C. Amblard and S. Girard. Estimation procedures for a semiparametric family of bivariate copulas. *Journal of Computational and Graphical Statistics*, 14(2):1–15, 2005.