Scientific contributions in 2008

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 2008. Three main research topics are addressed: 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, 4].

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, 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 [5, 6, 7, 8, 9].

Besides, the use of high order moments techniques permits to use all the observations from the sample [10, 11, 12] similarly to the methods used for the production frontier estimation in econometrics.

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 [13] in the semi-supervised context.

I also developped dimension reduction methods for high dimensional regression problems [14, 15, 16].

Finally, I worked on the design of statistical models for optimizing power consumption of printers [17, 18].

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 [19, 20, 21]. 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.

References

[1] A. Gannoun, S. Girard, C. Guinot, and J. Saracco. Reference ranges based on nonparametric quantile regression. *Statistics in Medicine*, 21(20):3119–3135, 2002.

[2] 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.

[3] A. Gannoun, S. Girard, C. Guinot, and J. Saracco. Trois méthodes non paramétriques pour l'estimation de courbes de référence - application à l'analyse de propriétés biophysiques de la peau. *Revue de Statistique Appliquée*, L(1):65–89, 2002.

[4] 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.

[5] L. Gardes and S. Girard. A moving window approach for nonparametric estimation of the conditional tail index. *Journal of Multivariate Analysis*, 99:2368–2388, 2008.

[6] S. Girard and P. Jacob. A note on extreme values and kernel estimators of sample boundaries. *Statistics and Probability Letters*, 78:1634–1638, 2008.

[7] S. Girard and L. Menneteau. Smoothed extreme value estimators of non-uniform point processes boundaries with application to star-shaped supports estimation. *Communication in Statistics - Theory and Methods*, 37(6):881–897, 2008.

[8] L. Gardes, S. Girard, and A. Lekina. A moving window approach for nonparametric estimation of extreme level curves. In *18th conference of the Intenational Federation of Operational Research Societies (IFORS)*, Sandton, Afrique du Sud, 2008.

[9] S. Anquetin, B. Boudevillain, D. Ceresetti, J.D. Creutin, A. Godart, B. Hingray, G. Molinié, E. Leblois, C. Bernard-Michel, S. Girard, and L. Gardes. Rainfall features, forcing and estimation over the Cévennes-Vivarais region. In *2th HyMeX workshop*, Palaiseau, France, juin 2008.

[10] S. Girard and P. Jacob. Frontier estimation via kernel regression on high power-transformed data. *Journal of Multivariate Analysis*, 99:403–420, 2008.

[11] S. Girard and P. Jacob. Estimation de frontière par régression sur les puissances élevées des données. In *Colloque statistique non paramétrique et statistique des processus, en l'honneur du Professeur D. Bosq*, Paris, septembre 2008.

[12] S. Girard and P. Jacob. Frontier estimation via regression on high power-transformed data. In *Joint Meeting of the Statistical Society of Canada and the Société Française de Statistique*, Ottawa, Canada, mai 2008.

[13] C. Bouveyron and S. Girard. Robust supervised classification with Gaussian mixtures: learning from data with uncertain labels. In *Compstat, 18th symposium of the IASC*, Porto, Portugal, aout 2008.

[14] C. Bernard-Michel, L. Gardes, and S. Girard. A note on sliced inverse regression with regularizations. *Biometrics*, 64:982–986, 2008.

[15] C. Bernard-Michel, S. Douté, L. Gardes, and S. Girard. Inverting hyperspectral images with Gaussian regularized sliced inverse regression. In *16th European Symposium on Artificial Neural Networks*, pages 463–468, Bruges, Belgique, avril 2008.

[16] C. Bernard-Michel, L. Gardes, and S. Girard. Regularization methods for Sliced Inverse Regression. In *8th International Conference on Operations Research*, Havana, Cuba, février 2008.

[17] V. Ciriza, L. Donini, J.B. Durand, and S. Girard. A statistical model for optimizing power consumption of printers. In *Joint Meeting of the Statistical Society of Canada and the Société Française de Statistique*, Ottawa, Canada, mai 2008.

[18] V. Ciriza, L. Donini, J.B. Durand, and S. Girard. A statistical model for optimizing power consumption of printers. In *XIG R & T Conference, Xerox Corporation*, Webster, USA, mai 2008.

[19] J. Diebolt, L. Gardes, S. Girard, and A. Guillou. Bias-reduced estimators of the Weibull tail-coefficient. *Test*, 17:311–331, 2008.

[20] J. Diebolt, L. Gardes, S. Girard, and A. Guillou. Bias-reduced extreme quantiles estimators of Weibull-tail distributions. *Journal of Statistical Planning and Inference*, 138:1389–1401, 2008.

[21] L. Gardes and S. Girard. Estimation of the Weibull tail-coefficient with linear combination of upper order statistics. *Journal of Statistical Planning and Inference*, 138:1416–1427, 2008.