Contributions to functional estimation

Stéphane Girard

INRIA Rhône-Alpes & LJK (team MISTIS).

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

Stephane.Girard@inria.fr

June 9, 2017

Abstract: This report summarizes my contributions to functional estimation. Most of my work addresses the problem of frontier or boundary estimation. Three methods are developped: Estimators based on the extreme values of the sample, estimators based on linear programming techniques and estimators based on high order moments. Some applications to reference curves estimation and production frontier estimation are provided.

Contributions

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, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22]. In this framework, we obtain the asymptotic distribution of the error between the estimators and the true frontier [23, 24, 25].

In the linear programming approach, the boundary of a set of points is defined as a closed curve bounding all the points and with smallest associated surface. It is thus natural to reformulate the boundary estimation problem as a linear programming problem. This method belongs to the Support Vector Machines (SVM) techniques [26, 27, 28, 29].

Besides, the use of high order moments techniques permits to use all the observations from the sample [30, 31, 32, 33] similarly to the methods used for the production frontier estimation [34, 35] in econometrics.

Finally, I developped dimension reduction methods for high dimensional regression problems [36, 37, 38, 39, 40, 41, 42, 43, 44, 2].

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] S. Girard and S. Louhichi. On the strong consistency of the kernel estimator of extreme conditional quantiles. In E. Ould-Said et al., editor, *Functional Statistics and Applications*, pages 59–77. Springer, 2015.

[6] L. Gardes and S. Girard. On the estimation of the functional Weibull tail-coefficient. *Journal of Multivariate Analysis*, 146:29–45, 2016.

[7] L. Gardes and S. Girard. Nonparametric estimation of the conditional tail copula. *Journal of Multivariate Analysis*, 137:1–16, 2015.

[8] J. El Methni, L. Gardes, and S. Girard. Nonparametric estimation of extreme risks from conditional heavy-tailed distributions. *Scandinavian Journal of Statistics*, 41:988–1012, 2014.

[9] A. Daouia, L. Gardes, and S. Girard. On kernel smoothing for extremal quantile regression. *Bernoulli*, 19:2557–2589, 2013.

[10] L. Gardes and S. Girard. Functional kernel estimators of large conditional quantiles. *Electronic Journal of Statistics*, 6:1715–1744, 2012.

[11] J. Carreau and S. Girard. Spatial extreme quantile estimation using a weighted log-likelihood approach. *Journal de la Société Française de Statistique*, 152(3):66–83, 2011.

[12] L. Gardes and S. Girard. Functional kernel estimators of conditional extreme quantiles. In F. Ferraty, editor, *Recent advances in functional data analysis and related topics*, pages 135–140. Springer, Physica-Verlag, 2011.

[13] A. Daouia, L. Gardes, S. Girard, and A. Lekina. Kernel estimators of extreme level curves. *Test*, 20(14):311–333, 2011.

[14] L. Gardes and S. Girard. Conditional extremes from heavy-tailed distributions: An application to the estimation of extreme rainfall return levels. *Extremes*, 13(2):177–204, 2010.

[15] L. Gardes, S. Girard, and A. Lekina. Functional nonparametric estimation of conditional extreme quantiles. *Journal of Multivariate Analysis*, 101:419–433, 2010.

[16] 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.

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

[18] 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.

[19] 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.

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

[21] S. Girard and P. Jacob. Projection estimates of point processes boundaries. *Journal of Statistical Planning and Inference*, 116(1):1–15, 2003.

[22] S. Girard and P. Jacob. Extreme values and Haar series estimates of point process boundaries. *Scandinavian Journal of Statistics*, 30(2):369–384, 2003.

[23] J. Geffroy, S. Girard, and P. Jacob. Asymptotic normality of the L_1-error of a boundary estimator. *Nonparametric Statistics*, 18(1):21–31, 2006.

[24] 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.

[25] 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.

[26] 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.

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

[28] 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.

[29] A. Nazin and S. Girard. L_1-optimal linear programming estimator for periodic frontier functions with Holder continuous derivative. *Automation and Remote Control*, 75(12):2152–2169, 2014.

[30] S. Girard, A. Guillou, and G. Stupfler. Uniform strong consistency of a frontier estimator using kernel regression on high order moments. *ESAIM: Probability and Statistics*, 18:642–666, 2014.

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

[32] S. Girard and P. Jacob. Frontier estimation with local polynomials and high power-transformed data. *Journal of Multivariate Analysis*, 100:1691–1705, 2009.

[33] S. Girard, A. Guillou, and G. Stupfler. Frontier estimation with kernel regression on high order moments. *Journal of Multivariate Analysis*, 116:172–189, 2013.

[34] A. Daouia, S. Girard, and A. Guillou. A γ-moment approach to monotonic boundaries estimation: with applications in econometric and nuclear fields. *Journal of Econometrics*, 178:727–740, 2014.

[35] A. Daouia, L. Gardes, and S. Girard. Nadaraya's estimates for large quantiles and free disposal support curves. In I. Van Keilegom and P. Wilson, editors, *Exploring research frontiers in contemporary statistics and econometrics*, pages 1–22. Springer, 2012.

[36] A. Chiancone, F. Forbes, and S. Girard. Student sliced inverse regression. *Computational Statistics and Data Analysis*, 2017. to appear.

[37] A. Chiancone, S. Girard, and J. Chanussot. Collaborative sliced inverse regression. *Communication in Statistics - Theory and Methods*, 46(12):6035–6053, 2017.

[38] M. Fauvel, S. Girard, S. Douté, and L. Gardes. Machine learning methods for the inversion of hyperspectral images. In A. Reimer, editor, *Horizons in World Physics*, volume 290, pages 51–77. Nova Science, New-York, 2017.

[39] S. Girard and J. Saracco. An introduction to dimension reduction in nonparametric kernel regression. In D. Fraix-Burnet and D. Valls-Gabaud, editors, *Regression methods for astrophysics*, volume 66, pages 167–196. EDP Sciences, 2014.

[40] R. Coudret, S. Girard, and J. Saracco. A new sliced inverse regression method for multivariate response. *Computational Statistics and Data Analysis*, 77:285–299, 2014.

[41] M. Chavent, S. Girard, V. Kuentz-Simonet, B. Liquet, T. M. N. Nguyen, and J. Saracco. A sliced inverse regression approach for data stream. *Computational Statistics*, 29:1129–1152, 2014.

[42] C. Bernard-Michel, S. Douté, M. Fauvel, L. Gardes, and S. Girard. Retrieval of Mars surface physical properties from Omega hyperspectral images using regularized sliced inverse regression. *Journal of Geophysical Research - Planets*, 114, 2009. E06005.

[43] C. Bernard-Michel, L. Gardes, and S. Girard. Gaussian regularized sliced inverse regression. *Statistics and Computing*, 19:85–98, 2009.

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