Electronic Journal of Statistics > Vol. 3 (2009) open journal systems

Extensions of smoothing via taut strings

Lutz Dümbgen, *University of Berne*
Arne Kovac, *University of Bristol*

Abstract

Suppose that we observe independent random pairs (X_1,Y_1), (X_2,Y_2), ..., (X_n,Y_n). Our goal is to estimate regression functions such as the conditional mean or $beta$--quantile of Y given X, where $0 < beta < 1$. In order to achieve this we minimize criteria such as, for instance, $$ sum_{i=1}^n rho(f(X_i) - Y_i) + lambda cdot TV(f) $$ among all candidate functions f. Here rho is some convex function depending on the particular regression function we have in mind, $TV(f)$ stands for the total variation of f, and $lambda > 0$ is some tuning parameter. This framework is extended further to include binary or Poisson regression, and to include localized total variation penalties. The latter are needed to construct estimators adapting to inhomogeneous smoothness of f. For the general framework we develop noniterative algorithms for the solution of the minimization problems which are closely related to the taut string algorithm (cf. Davies and Kovac, 2001). Further we establish a connection between the present setting and monotone regression, extending previous work by Mammen and van de Geer (1997). The algorithmic considerations and numerical examples are complemented by two consistency results.

AMS 2000 subject classifications: Primary 62G08; secondary 62G35.

Keywords: conditional means, conditional quantiles, modality, penalization, uniform consistency, total variation, tube method.

Full Text: PDF

Dümbgen, Lutz, Kovac, Arne, Extensions of smoothing via taut strings, Electronic Journal of Statistics, 3, (2009), 41-75 (electronic). DOI: 10.1214/08-EJS216.

References

Antoniadis, A. and Fan, J. (2001). Regularization of Wavelet Approximations (with discussion). J. Amer. Statist. Assoc. 96, 939–967. MR1946364

Chu. C. K., Glad, I. K., Godtliebsen, F. and Marron, J. S. (1998). Edge-preserving smoothers for image processing (with discussion). J. Amer. Statist. Assoc. 93, 526–553. MR1631321

Davies, P.L. and Kovac, A. (2001). Local extremes, runs, strings and multiresolution (with discussion). Ann. Statist. 29, 1–65. MR1833958

Donoho, D. L. (1993) Nonlinear wavelet methods for recovery of signals, images, and densities from noisy and incomplete data. In: Different Perspectives on Wavelets (I. Daubechies, editor), pp. 173–205, American Mathematical Society. MR1268002

Donoho, D. L. and Johnstone, I. M. (1994). Ideal spatial adaption by wavelet shrinkage. Biometrika 81, 425–455. MR1311089

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. and Picard, D. (1995). Wavelet shrinkage: asymptopia? (with discussion). J. Roy. Statist. Soc. Ser. B 57, 371–394. MR1323344

Dümbgen, L. and Spokoiny, V.G. (2001). Multiscale testing of qualitative hypotheses. Ann. Statist. 29, 124–152. MR1833961

Hoeffding, W. (1956). On the distribution of the number of sucesses in independent trials. Ann. Math. Statist. 27, 713–721. MR0080391

Huang, J. Z. (2003). Local asymptotics for polynomial spline regression. Ann. Statist. 31, 1600–1635. MR2012827

Ihaka, R. and Gentleman, R. (1996). R: A language for data analysis and graphics. J. Comp. Graph. Statist. 5, 299–314.

Knuth, D. (1998). The Art of Computer Programming. Addison-Wesley, Reading, MA. MR0378456

Koenker, R. and Mizera, I. (2004). Penalized triograms: Total variation regularization for bivariate smoothing. J. Roy. Statist. Soc. Ser. B 66, 145–163. MR2035764

Koenker, R., Ng, P. and Portnoy, S. (1994). Quantile smoothing splines. Biometrika 81, 673–680. MR1326417

Rice, J. (1984). Bandwidth choice for nonparametric regression. Ann. Statist. 12, 1215–1230. MR0760684

van de Geer, S. (2001). Least squares estimation with complexity penalties. Mathematical Methods of Statistics 10, 355–374. MR1867165

van de Geer, S. and Mammen, E. (1997). Locally adaptive regression splines. Ann. Statist. 25, 387–413. MR1429931