Multiresolution and model choice

Arne Kovac

We consider various settings of the nonparametric regression problem where for given data $y_1, \ldots, y_n$ at time points $t_1, \ldots, t_n$ we require an approximation f that is simple and close to the data. Most approaches develop first an algorithm that takes the data and some additional parameters like bandwidth and kernel function for kernel estimators. In a second step another method is developed for choosing the additional parameters, very often based on minimizing error criteria on test beds like cross-validation. Typically these methods do not produce simple approximations for complex data sets.

In this talk we study approaches that work the other way round and define first a criterion for approximation, giving rise to a set of functions each being an adequate model for the data. In a second step we aim to find a particular simple function among them and try to minimize measures such as the number of local extreme values.

The multiresolution criterion has turned out to be useful for defining approximation. Given noisy data $y_1, \ldots, y_n$ we require a function f to satisfy

$$\left| \sum_{i \in I} (y_i - f_i) \right| < w_I \cdot \sigma \tag{1}$$

with $w_I = \sqrt{|I| \cdot 2 \log(n)}$ for all intervals I of some family $\mathcal{I}$ of subintervals of $\{1, \ldots, n\}$ (Davies and Kovac, 2001; Davies, Kovac and Meise, 2007). This criterion is very strict in the sense that approximations from most popular smoothing methods like smoothing splines with cross validation, adaptive weights smoothing or kernel estimators using local plug-in bandwidths do not usually satisfy this criterion for complex data sets. Wavelet thresholding equipped with the universal $\tau = \sqrt{2 \log(n)}$ threshold (Donoho et al, 1995) have residuals that satisfy similar multiresolution conditions, but usually still hurt some of the multiresolution conditions in (1).

By replacing $y_i - f_i$ with terms such as $\mathrm{sign}(y_i - f_i)$ (Kovac, 2002) or, more generally, $R_i'(\hat{f}_i)$ with data-dependent functions R_i (Dümbgen and Kovac, 2005) the multiresolution criterion can be adapted to situations with outliers, quantile regression or Poisson regression. An extension to inverse problems is also straightforward: Assume that K is some linear operator and that we want to use Kf instead of f to approximate the data. Then we require a function to satisfy

$$\left| \sum_{i \in I} (y_i - (Kf)_i) \right| < w_I \cdot \sigma \quad \text{for all } I \in \mathcal{I}.$$

The multiresolution criterion can also be used in the context of estimating parameters of an ordinary differential equation. Here we model the data y as noisy observations from an ODE

$$\dot{x}(t) = f(x, u, t|\theta)$$

and want to estimate θ. Again it makes sense to only allow values for θ such that the residuals of x satisfy the multiresolution criterion.

Extending the multiresolution criterion to two or more dimensions is not straightforward, one possibility is to use a decomposition of the residuals using wedgelets (Polzehl and Spokoiny, 2003).

There are several possible ways for maximizing simplicity among all adequate functions. One way consists in minimising total variation (Davies, Kovac and Meise, 2007):

$$\sum_{i=1}^{n-1} |f_{i+1} - f_i| = \min \quad \text{s.t. } f \text{ satisfies (1).}$$

This leads to a linear program which can be computationally relatively expensive for some data sets. The computational complexity of problems like

$$\sum_{i=1}^{n} R_i(f_i) + \sum_{i=1}^{n-1} \lambda_i |f_{i+1} - f_i|$$

is considerably smaller and is for common choices of R_i not larger than $O(n \log(n))$ using a generalization of the taut string algorithm (Dümbgen and Kovac, 2005). The local penalty parameters can be chosen by the local squeezing technique (Davies and Kovac, 2001) to make sure that the solution satisfies the multiresolution criterion. Finally by using quick update steps it is possible to calculate the solution for the first n data from the solution for the first $n-1$ data without recalculating most of the solution. This allows an extension to online processing (Kovac and Wei, 2007).

References

[1] P. L. Davies and A. Kovac, *Local extremes, runs, strings and multiresolution (with discussion)*, Annals of Statistics 29 (2001), 1–65.

[2] P. L. Davies, A. Kovac and M. Meise, *Confidence Regions, Regularization and Non-Parametric Regression.*, Technical report (2007).

[3] L. Dümbgen and A. Kovac, *Extensions of Smoothing via Taut Strings*, Technical report (2005).

[4] D. L. Donoho, I. .M. Johnstone, G. Kerkyacharian and D. Picard, *Wavelet shrinkage: asymptopia?*, Journal of the Royal Statistical Society, Ser. B **57** (1995), 371–394.

[5] A. Kovac, *Robust nonparametric regression and modality.* in: Developments in Robust Statistics, R. Dutter, P. Filzmoser, U. Gather, P. Rousseeuw (eds.), Physica, Heidelberg, p218-227.

[6] A. Kovac and Y. Wei, *A taut string method for online data*, Technical report (2007).

[7] J. Polzehl and V. Spokoiny, *Image denoising: Pointwise adaptive approach*, Annals of Statistics 31 (2003), 30–57.