A Comparative Study of Nonparametric Derivative Estimators

National University of Ireland, Galway
Ollscoil na hÉireann, Gaillimh

University of Durham

John Newell[1], Jochen Einbeck[2],

[1]NUI Galway, [2]University of Durham

Motivation

Nonparametric estimation of derivatives is important in a variety of disciplines. Specifically, when considering a regression problem of type

$$y_i = m(x_i) + \varepsilon_i, \quad i = 1, \ldots, n,$$

one is often not interested in in $m(\cdot)$ itself, but rather in $m'(\cdot)$. An important special case is when x represents time, in which the 1st derivative of m has the interpretation of a speed, and the 2nd derivative of an acceleration, which is of interest in the analysis of growth curves.

The importance of estimating derivatives goes far beyond the end in itself. Often one relies on asymptotic approximations in order to obtain bias and variance estimates, confidence intervals, optimal bandwidths, etc., and these expressions usually involve derivatives of $m(\cdot)$, which are normally unknown and have to be estimated.

Nonparametric Derivative Estimation

There are two main approaches to nonparametric derivative estimation.

Local polynomials of degree *p*.

The estimator of the j^{th} derivative $m^{(j)}(x)$ (where $0 < j \le p$) at point x is given by $\hat{m}^j(x) = j!\hat{\beta}_j(x)$ according to Taylor's Theorem, where $\hat{\beta}_j(x)$ is obtained by minimising

$$\sum_{i=1}^{n} K\left(\frac{x_i - x}{h}\right)\left\{y_i - \sum_{j=0}^{p}\beta_j(x)(x_i - x)^j\right\}^2$$

in terms of the vector $(\beta_0(x), \ldots, \beta_p(x))$ where K is a kernel function and h the bandwidth controlling the degree of smoothing.

Spline smoothing.

If $\hat{m}(x)$ is a spline estimate of $m(x)$ one considers $\frac{d^j}{dx^j}\hat{m}(x)$ as an estimator of $m^{(j)}(x)$. Different authors have pursued this idea, using splines with (Heckman & Ramsay, 2000) or without penalisation.

Papers originating from the local polynomial smoothing community gave the impression that the entire issue of nonparametric derivative estimation is solved, and as a result the research activity about this topic stalled to some extent. This is unfortunate, as most problems are treated rather cursorily in the literature and many open questions remain.

Ramsay (1998) for example noted that \`typically one sees derivatives go wild at the extremes, and the higher the derivative, the wilder the behavior', and that further problems arise when it comes to smoothing parameter (bandwidth) selection, where CV and GCV can be \`poor guides'.

Comparison of Available Routines

There are several R packages available for derivative estimation and a summary of their features is as follows:

Package (Version)	function	j(max)	Smooth.Par
locfit (1.5-3)	locfit	2	GCV
KernSmooth (2.22-19)	locpoly	No limit	—
lokern (1.0-4)	glkerns	4*	plug-in**
lpridge (1.0-3)	lpridge	9	—
pspline (1.0-10)	smooth.Pspline	4**	CV/GCV
sfsmisc (0.95-9)	D1D2	2	GCV
SemiPar (1.0-2)	spm	7***	RE(ML)

*if bandwidth selected automatically, then `j(max)` =2. **a variant `lokerns` featuring a variable bandwidth is also implemented. ***no formal requirement, but from our experience it breaks down computationally for higher orders.

For illustration, we consider data generated by contaminating the function $m(x)=x+2\exp(-16x^2)$, $x \in [-2,2]$, with very small Gaussian noise (σ= 0.1). A moderate outlier at the left boundary with coordinates (-1.97, -1.75) and a further outlier at (0.95,0) were added by hand, giving a total sample size n=60 (Figure 1). A small simulation study was carried out to compare a selection of the routines in terms of their ability to recover the first two derivatives from the contaminated function (Figure 2).

Local Polynomial Methods

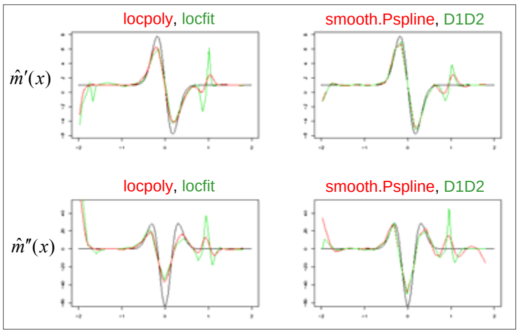
The packages considered were `locfit` (**locfit**) and `locpoly` (**KernSmooth**) with usual default setting $p=j+1$ (Fan & Gijbels, 1996). The bandwidths were chosen using the result of locfit's `gcvplot` for the 2nd derivative, but undersmoothed for the first. Both functions produce a considerable bias which cannot be cured by modifying the bandwidth as otherwise the outlier and boundary effects get even worse.

There is a systematic problem with this kind of estimator: note that the asymptotic bias of the derivative estimate based on a quadratic fit with bandwidth h is given by

$$Bias(\hat{m}'(x) \mid x_1, \ldots, x_n) = c \cdot m'''(x)h^2 + op(h^2)$$

($c>0$ being a constant depending on kernel moments (Fan & Gijbels, 1996)). This implies that, where $m'(\cdot)$ is concave, the bias is negative, and where $m'(\cdot)$ is convex, the bias is positive, i.e. a downward smoothing bias similar as observed by Stoker (1993) for density derivative estimation. This bias (left panel of Figure 1) tends to increase with the derivative order j; one reason is that the necessary bandwidth h (appearing in the bias generally as a factor h^{p+1-j}) increases with j. The smoothing bias diminishes when setting $p=j+2$ as suggested by Ruppert (1997), at the expense of increased outlier and boundary effects (not shown).

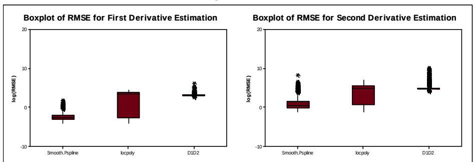
Figure 1.
Behavior of derivative estimators for a random simulation



Smoothing Splines

For comparison we considered two penalized smoothing packages only: `smooth.Pspline` (**`pspline`**) and `D1D2` (**`sfsmisc`**). The latter is restricted to cubic splines, whereas we use for the former a quintic and septic spline for the 1st and 2nd derivative, respectively (Ramsay, 1998). The smoothing parameter is selected for `smooth.Pspline` using the built-in GCV routine, and for `D1D2` such that the fits pass equally well through the central part. Both fits are much less biased than the local polynomial estimators, and more stable at the left boundary (Figure 1).

Figure 2.



P-Spline (Penalized B- Spline) Software Note

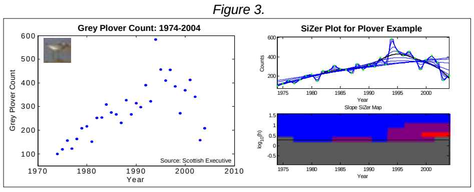
R function names are quite confusing:

- smooth.Pspline *does not* actually use B-Splines, but penalized smoothing splines.
- smooth.spline *does* use B-Splines if the number of knots specified is smaller than *n*, and it also features penalisation..
- There seems to be no simple immediate tool in R to fit P-Splines (in the above sense) directly.
- However, the function smooth.construct (in package mgcv) offers a way to construct P Spline smoothers for the use inside a generalized additive model.

Examples

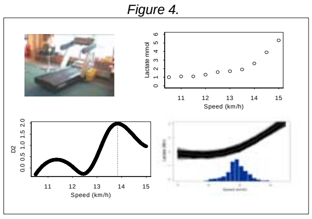
The change in bird populations is considered a good environmental indicator. The Grey Plover for example is a used as an indicator of water quality. A scatterplot and a SiZer (Significance Zero Crossings of Derivatives) plot (Chaudhuri and Marron, 1999) suggest that a possible significant decrease in Plover number from 2000 onwards (Figure 3).

Figure 3.



When analysing blood lactate data of elite athletes, a useful marker has been identified (Newell et al., 2005) which is the workload corresponding to the maximum of the 2nd derivative of the underlying lactate curve (Figure 4).

Figure 4.



Outlook

There is a general lack of *robust* derivative estimators. Smoothing parameter selection tools are based on optimising the estimate of the *regression function* and not of the derivative, which can lead to serious undersmoothing (Jarrow et al., 2004).

References

Chaudhuri, P., and Marron, J.S. (1999) SiZer for Exploration of Structures in Curves, *JASA*, Vol. 94.

Fan, J., and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. London: Chapman & Hall.

Heckman, N.E., and Ramsay, J. (2000). Penalized regression with model-based penalties. *The Canadian Journal of Statistics*, 28, 241-258.

Jarrow, R., Ruppert, D., and Yu, Y. (2004). Estimating the term structure of corporate debt with a semiparametric penalized spline model, *JASA*, 99, 57-66.

Newell, J., Einbeck, J., Madden, N., and McMillan, K. (2005). Model free endurance markers based on the second derivative of blood lactate curves. In: Francis et al. (Eds), *Proceedings of the 20th IWSM*, 357-364, Sydney.

Ramsay, J. (1998). Derivative estimation. StatLib— S-News Thu, 12 March 1998, www.math.yorku.ca/Who/Faculty/Monette/S-news/0556.html

Ruppert, D. (1997). Empirical-bias bandwidths for local polynomial nonparametric regression and density estimation. *JASA*, 92, 1049–1062.

Stoker, T.M. (1993). Smoothing bias in density derivative estimation. *JASA*, 88, 855–863.