Smoothed Nonparametric Derivative Estimation Using Weighted Difference Quotients

Paper Author: *Yu Liu* and *Kris De Brabanter*

Presented By **Yikun Zhang**

Department of Statistics,
University of Washington

April 18, 2023

Department of STATISTICS

UNIVERSITY of WASHINGTON

Motivations

Given a differentiable function $m : \mathbb{R} \to \mathbb{R}$, its derivative is defined as:

$$m'(x) = \lim_{h \to 0} \frac{m(x+h) - m(x)}{h}.$$

Studying $m'(x)$ has significant impacts in Statistics:

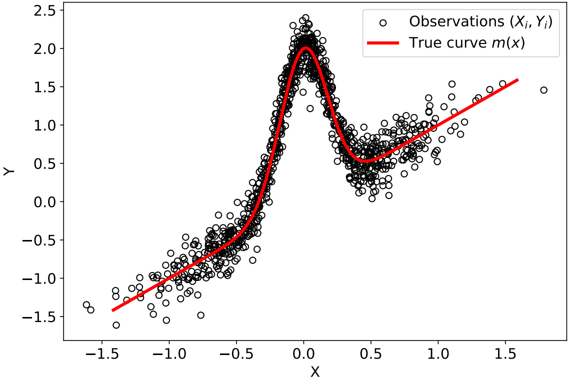
- Explore the structures in curves (Chaudhuri and Marron, 1999).
- Signify the changing trend in time series (Rondonotti et al., 2007).
- Correct the bias for density/regression estimators to conduct valid inference (Eubank and Speckman, 1993; Calonico et al., 2018; Cheng and Chen, 2019).
- ...

How to Estimate Derivatives in Practice?

Data setting:

$$Y_i = m(X_i) + e_i, \quad \text{with } X_i \in [a,b] \text{ for } i = 1, ..., n,$$

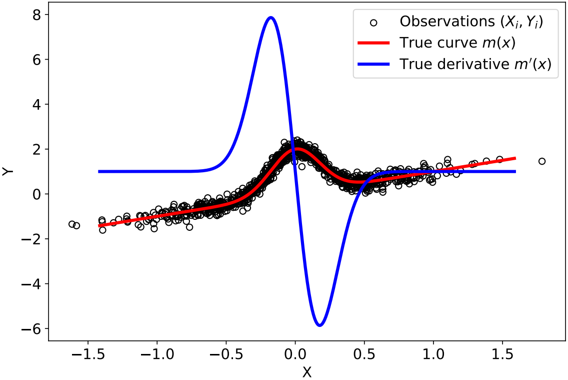
with e_i being independent of X_i and $\mathbb{E}(e_i) = 0, \mathrm{Var}(e_i) = \sigma_e^2 < \infty$.



Data setting:

$$Y_i = m(X_i) + e_i, \quad \text{with } X_i \in [a,b] \text{ for } i = 1, ..., n,$$

with e_i being independent of X_i and $\mathbb{E}(e_i) = 0, \mathrm{Var}(e_i) = \sigma_e^2 < \infty$.



Question: How do we estimate $m'(x)$ from the data $\mathcal{D} = \{(X_i, Y_i)\}_{i=1}^n$?

Parametric Methods: Assume $m(x)$ lying in some parametric family $\{f(x;\theta) : \theta \in \Theta\}$ and fit

$$\widehat{\theta} \in \arg\min_{\theta \in \Theta} \sum_{i=1}^{n} \left[Y_i - f(X_i;\theta)\right]^2 \quad \Longrightarrow \quad \widehat{m}'(x) = f'(x;\widehat{\theta}).$$

- *Drawback:* It is difficult to propose a correct family $\{f(x;\theta) : \theta \in \Theta\}$.

Nonparametric Methods: Make no model assumptions on $m(x)$ and estimate $m'(x)$ from the data $\mathcal{D}$.

- **Smoothing Splines (Zhou and Wolfe, 2000):** Estimate $m'(x)$ through the derivative of smoothing splines (*i.e.*, piecewise polynomial curves).

- **Local Polynomial Regression (Fan and Gijbels, 1996):** It solves the weighted least-square problem at each query point x: [1]

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^{p+1}} \sum_{i=1}^{n} \left[Y_i - \sum_{j=0}^{p} \beta_j (X_i - x)^j\right]^2 K\left(\frac{X_i - x}{h}\right)$$

  and estimate $m'(x)$ as $\widehat{m}'(x) = \widehat{\beta}_1$.

[1] K is a symmetric kernel and $h > 0$ is the bandwidth parameter.

Nonparametric Methods: Difference Quotients

We order the data $\mathcal{D} = \{(X_i, Y_i)\}_{i=1}^n$ according to the increasing order of $X_i, i = 1, ..., n$:

$$Y_i = m(X_{(i)}) + e_i, \quad i = 1, ..., n.$$

The first-order difference quotients are defined as (Müller et al., 1987; Härdle, 1990):

$$\widehat{q}^{(1)}(X_{(i)}) = \frac{Y_i - Y_{i-1}}{X_{(i)} - X_{(i-1)}}, \quad i = 1, ..., n.$$

Under some conditions, it estimates $m'(X_{(i)})$ in the asymptotic rate:

$$\widehat{q}^{(1)}(X_{(i)}) - m'(X_{(i)}) = O_P\left(n^2\right).$$

Drawback: The variance of $\widehat{q}^{(1)}(X_{(i)})$ scales *quadratically* with respect to the sample size n.

To reduce the variance, Iserles (2009); Charnigo et al. (2011) considered

$$\widehat{Y}_i^{(1)} \equiv \widehat{Y}_i^{(1)}(X_{(i)}) = \sum_{j=1}^{k} w_j \left(\frac{Y_{i+j} - Y_{i-j}}{X_{(i+j)} - X_{(i-j)}} \right)$$

for $k+1 \leq i \leq n-k$ and $k \leq \frac{(n-1)}{2}$.

- The weights $w_j, j = 1, ..., k$ with $\sum_{j=1}^{k} w_j = 1$ are chosen to minimize the variance $\mathrm{Var}\left(\widehat{Y}_i^{(1)} | X_{(1)}, ..., X_{(n)}\right)$.
- The asymptotic rate of convergence becomes

$$\widehat{Y}_i^{(1)} - m'(X_{(i)}) = O\left(\frac{k}{n}\right) + O_P\left(\frac{n^2}{k^3}\right).$$

Drawback: It only estimates m' at the design points $X_{(1)} \leq \cdots \leq X_{(n)}$.

Contributions of This Paper

De Brabanter et al. (2013) proposed using local polynomial regression to smooth out the noisy derivative estimates $\widehat{Y}_i^{(1)}, i = k+1, ..., n-k$.

- Noises $\widetilde{e}_i, i = k+1, ..., n-k$ are *correlated* for the noisy derivative estimates
$$\widehat{Y}_i^{(1)} = m'(X_{(i)}) + \widetilde{e}_i, \quad i = 1, ..., n.$$
- The kernel K needs to be bimodal and satisfies $K(0) = 0$.

Drawback: Their method only works for equispaced design, *i.e.*,

$$X_{(i)} = a + \frac{i \cdot (b-a)}{n-1}, \quad i = 1, ..., n.$$

In this paper (Liu and De Brabanter, 2020), the author will extend the above framework to random designs.

Thank you!

Reference I

S. Calonico, M. D. Cattaneo, and M. H. Farrell. On the effect of bias estimation on coverage accuracy in nonparametric inference. *Journal of the American Statistical Association*, 113(522):767–779, 2018.

R. Charnigo, B. Hall, and C. Srinivasan. A generalized c p criterion for derivative estimation. *Technometrics*, 53(3):238–253, 2011.

P. Chaudhuri and J. S. Marron. Sizer for exploration of structures in curves. *Journal of the American Statistical Association*, 94(447):807–823, 1999.

G. Cheng and Y.-C. Chen. Nonparametric inference via bootstrapping the debiased estimator. *Electronic Journal of Statistics*, 13(1):2194 – 2256, 2019.

K. De Brabanter, J. De Brabanter, I. Gijbels, and B. De Moor. Derivative estimation with local polynomial fitting. *Journal of Machine Learning Research*, 14(1):281–301, 2013.

R. L. Eubank and P. L. Speckman. Confidence bands in nonparametric regression. *Journal of the American Statistical Association*, 88(424):1287–1301, 1993.

J. Fan and I. Gijbels. *Local polynomial modelling and its applications*, volume 66. Chapman & Hall/CRC, 1996.

W. Härdle. *Applied nonparametric regression*. Number 19. Cambridge university press, 1990.

A. Iserles. *A first course in the numerical analysis of differential equations*. Number 44. Cambridge university press, 2009.

Y. Liu and K. De Brabanter. Smoothed nonparametric derivative estimation using weighted difference quotients. *Journal of Machine Learning Research*, 21(1):2438–2482, 2020.

H.-G. Müller, U. Stadtmüller, and T. Schmitt. Bandwidth choice and confidence intervals for derivatives of noisy data. *Biometrika*, 74(4):743–749, 1987.

V. Rondonotti, J. S. Marron, and C. Park. SiZer for time series: A new approach to the analysis of trends. *Electronic Journal of Statistics*, 1(none):268 – 289, 2007.

S. Zhou and D. A. Wolfe. On derivative estimation in spline regression. *Statistica Sinica*, pages 93–108, 2000.