Cattaneo, Jansson and Ma (2020, JASA)

Yasuyuki Matsumura (Kyoto University)

Last Updated: May 3, 2025

https://yasu0704xx.github.io

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2020, VOL. 115, NO. 531, 1449–1455: Theory and Methods
https://doi.org/10.1080/01621459.2019.1635480

Simple Local Polynomial Density Estimators

Matias D. Cattaneo[a], Michael Jansson[b], and Xinwei Ma[c]

[a]Department of Operations Research and Financial Engineering, Princeton University, Princeton, NJ; [b]Department of Economics, CREATES, University of California, Berkeley, CA; [c]Department of Economics, University of California, San Diego, CA

ABSTRACT

This article introduces an intuitive and easy-to-implement nonparametric density estimator based on local polynomial techniques. The estimator is fully boundary adaptive and automatic, but does not require pre-binning or any other transformation of the data. We study the main asymptotic properties of the estimator, and use these results to provide principled estimation, inference, and bandwidth selection methods. As a substantive application of our results, we develop a novel discontinuity in density testing procedure, an important problem in regression discontinuity designs and other program evaluation settings. An illustrative empirical application is given. Two companion Stata and R software packages are provided.

ARTICLE HISTORY
Received September 2017
Accepted May 2019

KEYWORDS
Density estimation; Local polynomial methods; Manipulation test; Regression discontinuity

CONTACT Matias D. Cattaneo cattaneo@princeton.edu Department of Operations Research and Financial Engineering, Sherrerd Hall, Charlton Street, Princeton University, Princeton, NJ 08544

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/r/JASA.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA.

This slide is available on

https://github.com/yasu0704xx/ArticleReview.

Contents

Introduction

Nonparametric Density Estimation

- Flexible (nonparametric) estimation of probability density function features prominently in empirical work in statistics, economics, and many other disciplines. Sometimes the density function is the main object of interest, while in other cases it is a useful ingredient in forming up two-step nonparametric or semiparametric procedures.
- Examples: manipulation testing, distributional treatment effect and counterfactual analysis, instrumental variables treatment effect specification and heterogeneity analysis, and common support/overlap testing.
- See Imbens and Rubin (2015) [14] and Abadie and Cattaneo (2018) [1] for reviews and further references.

Evaluation Points on the Boundary

- A common problem faced when implementing density estimators in empirical work is the presence of of evaluation points that lie on the boundary of the support of the variable of interest.
- Whenever the density estimator is constructed at or near boundary points, which may or may not be known by the researcher, the finite- and large-sample properties of the estimator are affected.

Evaluation Points on the Boundary

- Standard kernel density estimators are invalid at or near boundary points, while other methods may remain valid but usually require choosing additional tuning parameters, transforming the data, a priori knowledge of the boundary point location, or some other boundary-related specific information or modification.
- Furthermore, it is usually the case that one type of density estimator is used for evaluation points at or near the boundary, while a different type is used for interior points.

Boundary Adaptive Density Estimator

- We introduce a novel nonparametric estimator of a density function constructed using local polynomial techniques (Fan and Gijbels, 1996 [10]).
- The estimator is intuitive, easy to implement, does not require prebinning of the data, and enjoys all the desirable features assocated with local polynomial regression estimation.
- In particular, the estimator automatically adapts to the boundaries of the support the density without requiring specific data modification or additional tuning parameter choices, a feature that is unavailable for most other density estimators in the literature. See Karunamuni and Alberts (2005) [15] for a review on this topic.

Related Approaches

- The most closely related approaches currently available in the literature are the local polynomial density estimators of Cheng, Fan and Marron (1997) [8] and Zhang and Karunamuni (1998) [22].
- Their estimators require knowledge of the boundary location and prebinning of the data (or, more generally, pre-estimation of the density near the boundary), and hence introduce additional tuning parameters that need to be chosen.

Heuristic Idea

- The heuristic idea underlying our estimator, and differentiating it from other existing ones is simple to explain.
- Whereas other nonparametric density estimators are constructed by soothing out a histogram-type estimator of the density, our estimator is constructed by smoothing out the empirical distribution function using local polynomial techniques.
- Accordingly, our density estimator is constructed using preliminary tuning-parameter-free and $\sqrt{n}$-consistent distribution function estimator (where n denotes the sample size), implying in particular that the only tuning-parameter required by our approach is bandwidth associated with the local polynomial fit at each evaluation point.

Statistical Properties

- Asymptotic expansions of the leading bias and variance
- Asymptotic Gaussian distributional approximation and valid statistical inference
- Consistent standard error estimators
- Consistent data-driven bandwidth selection based on an asymptotic mean squared error (MSE) expansion
- Note that all these results apply to both interior and boundary points in a fully automatic and data-driven way, without requiring boundary specific transformations of the estimator or of the data, and without employing additional tuning parameters (beyond the main bandwidth present in any kernel-based nonparametric method).

Manipulation Testing

- McCrary (2008) [19] proposed the idea of manipulationtesting via discontinuity in density testing for regression discontinuity (RD) designs, and developed an implementation thereof using the density estimator of Cheng, Fan and Marron (1997) [8], which requires prebinning of the data and choosing two tuning parameters.
- On the other hand, the new proposed discontinuity in density test employing our density density estimator only requires the choice of one tuning parameter, and enjoys other features associated with local polynomial methods.
- Empirical Illustration: We also illustrate its performance with an empirical application employing the canonical Head Start data in the context of RD designs (Ludwig and Miller, 2007 [18]).

RD Designs

- For introductions to RD designs, and futher references, see Imbens and Lemieux (2008) [13], Lee and Lemieux (2010) [16], Cattaneo, Titiunik and Vazquez-Bare (2017) [7].
- For recent papers on modern RD methodology, see, for example, Arai and Ichimura (2018) [2], Ganong and Jager (2018) [11], Hyytinen, Merilainen, Saarimaa, Toivanen and Tukiainen (2018) [12], Dong, Lee and Gou (2019) [9], and references therein.

Software Packages

- `rddensity` (Cattaneo, Jansson and Ma, 2018 [5]): Specifically tailored to manipulation testing (i.e., two-sample doscontinuity in density testing).
- `lpdensity` (Cattaneo, Jansson and Ma, 2019 [6]): Providing generic density estimation over the support of the data.

Boundary Adaptive Density Estimation

Main Technical Results

Application to Manipulation Testing

Empirical Illustration

Conclusion

References

References i

Abadie, A., and Cattaneo, M. D. (2018). Econometric methods for program evaluation. *Annual Review of Economics*, **10**, 465-503.

Arai, Y., and Ichimura, H. (2018). Simultaneous selection of optimal bandwidths for the sharp regression discontinuity estimator. *Quantitative Economics*, **9**, 441-482.

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

- Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, **82**, 2295-2326.
- Cattaneo, M. D., Jansson, M., and Ma, X. (2018). Manipulation testing based on density discontinuity. *Stata Journal*, **18**, 234-261.
- Cattaneo, M. D., Jansson, M., and Ma, X. (2019). `lpdensity`: Local Polynomial Density Estimation and Inference. arXiv:1906.06529.

- Cattaneo, M. D., Titiunik, R., and Vazquez-Bare, G. (2017). Comparing inference approaches for RD designs: A reexamination of the effect of Head Start on child mortality. *Journal of Policy Analysis and Management*, **36**, 643-681.
- Cheng, M.-Y., Fan, J., and Marron, J. S. (1997). On automatic boundary corrections. *Annals of Statistics*, **25**, 1691-1708.
- Dong, Y., Lee, Y.-Y., and Gou, M. (2019). Regression discontinuity designs with a continuous treatment. SSRN Working Paper No. 3167541.
- Fan, J., and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. New York: Chapman & Hall/CRC.

- Ganong, P., and Jager, S. (2018). A permutation test for the regression kink design. *Journal of the American Statistical Association*, **113**, 494-504.
- Hyytinen, A., Merilainen, J., Saarimaa, T., Toivanen, O., and Tukiainen, J. (2018). When does regression discontinuity design work? Evidence from random election outcomes. *Quantitative Economics*, **9**, 1019-1051.
- Imbens, G., and Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, **142**, 615-635.

- Imbens, G. W., and Rubin, D. B. (2015). *Causal Inference in Statistics, Social, and Biomedical Sciences*. New York: Cambridge University Press.
- Karunamuni, R., and Alberts, T. (2005). On boundary correction in kernel density estimation. *Statistical Methodology*, **2**, 191-212.
- Lee, D. S., and Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, **48**, 281-355.
- Li, K.-C. (1987). Asymptotic optimality for C_p, C_L, cross-validation and generalized cross-validation: Discrete index set. *Annals of Statistics*, **15**, 958-975.

- Ludwig, J., and Miller, D. L. (2007). Does Head Start improve children ' s life chances? Evidence from a regression discontinuity design. *Quarterly Journal of Economics*, **122**, 159-208.
- McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, **142**, 698-714.
- Otsu, T., Xu, K.-L., and Matsushita, Y. (2014). Estimation and inference of discontinuity in density. *Journal of Business and Economic Statistics*, **31**, 507-524.
- Wand, M., and Jones, M. (1995). *Kernel Smoothing*. New York: Chapman & Hall/CRC.

Zhang, S., and Karunamuni, R. J. (1998). On kernel density estimation near endpoints. *Journal of Statistical Planning and Inference*, **70**, 301-316.