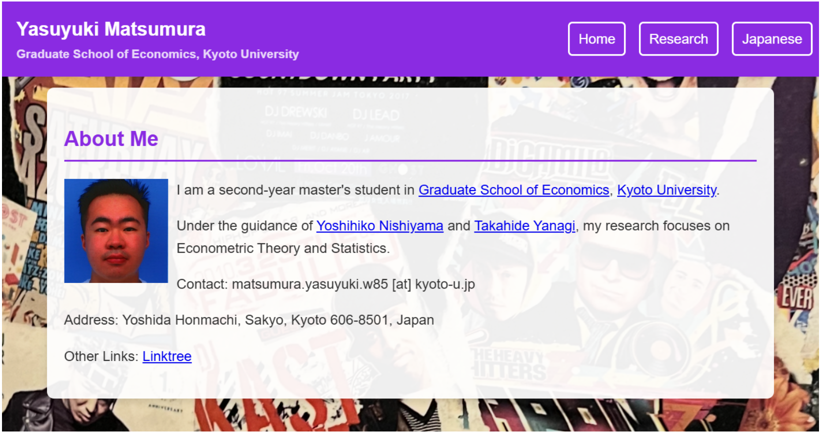
Nonparametric Density Estimation

Sections 17.1-17.8 of Hansen (2022)

Yasuyuki Matsumura (Kyoto University)

Last Updated: June 16, 2025

https://yasu0704xx.github.io



Available on https://github.com/yasu0704xx/Econometrics2025.

Introduction

- As a general rule, density functions can take any shape. They are inherently nonparametric and cannot be described by a finite set of parameters.
- That is, functional and/or distributional specifications relied on when estimating density functions may be incorrect.
- If we assume that such specifications are "true," we might obtain incorrect empirical conclusions.
- Thus, it would be desirable if we develop estimation procedures without requiring functional and/or distributional specifications.
- Nonparametric kernel methods achieve such a goal.

Setup

- Here we review Sections 17.1-17.8 of Hansen (2022) [9].
- We proceed with a discussion of how to estimate the probability density function $f(x)$ of a real-valued random variable X for which we have n IID observations $X_1, \cdots X_n$.
- We assume that $f(x)$ is continuous.
- The goal is to estimate $f(x)$ either at a single point x or a set of points in the interior of the support of X.

Literature

- Excellent textbooks on nonparametric density estimation include Silverman (1986) [22] and Scott (1992) [21].
- The following textbooks are often referred to:
    - van der Vaart (2000, Chapter 24) [23],
    - Pagan and Ullah (1999, Chapter 2) [18], and
    - Li and Racine (2007, Chapter 1) [15].
- 日本語の文献：
    - 西山・人見 (2023，第 1 章) [28]
    - 末石 (2015，第 9 章) [26]
    - 清水 (2023，第 5 章) [25]

Contents

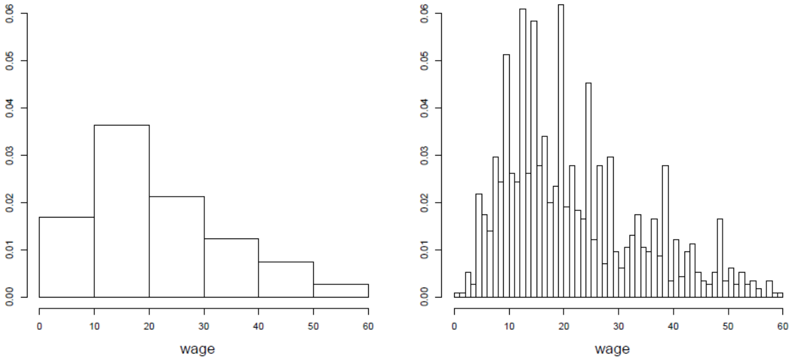
Idea behind Kernel Density Estimation

Histogram

- A simple and familiar estimator of $f(x)$ is a histogram.
- Divide the range of $f(x)$ into B bins of width w.
- Counting the number of observations n_j in each bin j, we obtain the histogram estimator of $f(x)$ for x in the j-th bin:

$$\hat{f}(x) = \frac{n_j}{nw}. \tag{1}$$

- The histogram is the plot of these heights, displayed as rectangles.



(a) Bin Width = 10

(b) Bin Width = 1

Figure 17.1: Histogram Estimate of Wage Density for Asian Women

Empirical Distribution Function

- Let us generalize the concept of histogram estimator.
- The empirical distribution function is defined as

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} 1(X_i \leq x).$$

- Let $F(x) = \int_{-\infty}^{x} f(x)dx$ denote the (cumulative) distribution function.
- By L.L.N. and C.L.T.,[1] we obtain

$$F_n(x) \xrightarrow{p} F(x),$$
$$\sqrt{n}(F_n(x) - F(x)) \xrightarrow{d} \text{Normal}(0, F(x)(1 - F(x))).$$

[1] We discuss these convergences in Chapter 18 of Hansen (2022) [9].

Naive Estimator

- Since $F_n(x)$ includes an indicator function, the empirical distribution function is not differentiable.
- Instead, let us consider approximate the "derivative" of $F_n(x)$.
- Note that, for $h \to 0$, it holds that

$$f(x) \approx \frac{F(x+h) - F(x-h)}{2h}.$$

- Replacing F with F_n, we obtain the naive estimator[2] of $f(x)$:

$$f_n(x) = \frac{F_n(x+h) - F_n(x-h)}{2h}.$$

- Under certain conditions, it can be shown that $f_n(x) \xrightarrow{p} f(x)$.

[2] The Rosenblatt estimator (Rosenblatt, 1956) [20]

- The naive estimator of $\phi(x)$ using IID observations $X_1, \cdots X_{100} \sim \text{Normal}(0, 1)$: [3]

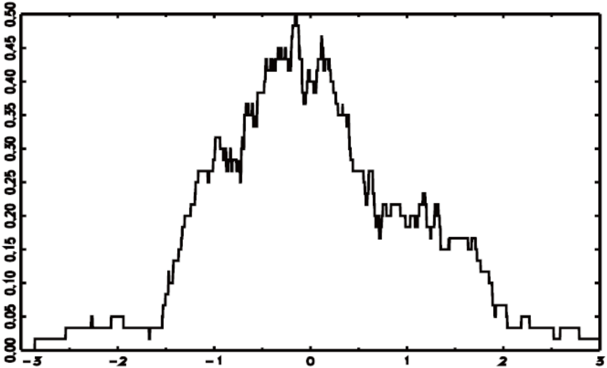


図 1: ナイーブ推定量 ($h = 0.3$)

[3] Cited from the lecture note 01 by N. Sueishi.

Idea behind Kernel Density Estimation

- The naive estimator can be rewritten as

$$
\begin{aligned}
f_n(x) &= \frac{1}{2nh} \sum_{i=1}^{n} 1(x - h \le X_i \le x + h) \\
&= \frac{1}{nh} \sum_{i=1}^{n} k_0\left(\frac{X_i - x}{h}\right),
\end{aligned}
$$

where $k_0(\cdot)$ is given by

$$
k_0(u) = \frac{1}{2} \cdot 1(-1 \le u \le 1).
$$

- Replacing $k_0(\cdot)$ with some smooth function, we can obtain a differentiable, smooth estimator of $f(x)$ ...?

Kernel Density Estimator

Kernel Density Estimator

- The kernel density estimator[4] of $f(x)$ is given by

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right). \tag{2}$$

- $K(u)$ is a weighting function known as a kernel function. The kernel $K(u)$ weights observations based on the distance between X_i and x.
- $h > 0$ is a scalar known as a bandwidth. The bandwidth h determines what is meant by "close."
- The kernel density estimator (2) critically depends on the bandwidth rather than the kernel function.

[4] The Parzen-Rosenblatt estimator (Parzen, 1962) [19]

Kernel Function

Definition 17.1

- A kernel function $K(u)$ satisfies

1. $0 \leq K(u) \leq \bar{K} < \infty,$ (3)
2. $K(u) = K(-u),$ (4)
3. $\int_{-\infty}^{\infty} K(u)du = 1,$ and (5)
4. $\int_{-\infty}^{\infty} |u|^r K(u)du < \infty$ for all positive integers r. (6)

- Essentially, a kernel function is a bounded PDF which is symmetric about zero.
- Assumption (6) is not essential for most results but is a convenient simplification and does not exclude any kernel functions used in standard empirical practice.

Definition 17.2

- A normalized kernel function $K(u)$ satisfies

$$\int_{-\infty}^{\infty} u^2 K(u) du = 1.$$

- The j-th moment of a kernel is defined as

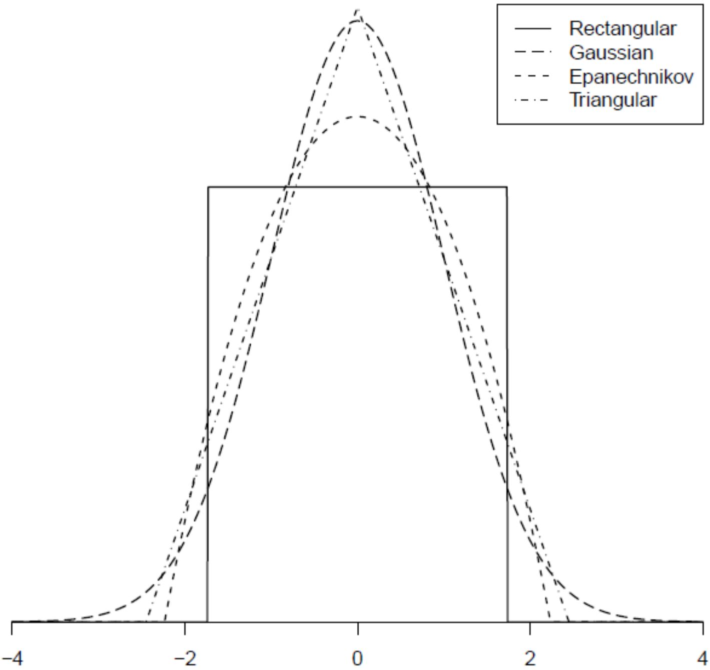
$$\kappa_j(K) = \int_{-\infty}^{\infty} u^j K(u) du.$$

- The order of a kernel ν is defined as the order of the first non-zero moment.

Examples of Second-Order Kernel

- Rectangular kernel: $K(u) = \frac{1}{2\sqrt{3}} 1(|u| \leq \sqrt{3})$
- Gaussian kernel: $K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right)$
- Epanechnikov kernel:[5] $K(u) = \frac{3}{4\sqrt{5}}\left(1 - \frac{u^2}{5}\right) 1(|u| \leq \sqrt{5})$
- Triangular kernel: $K(u) = \frac{1}{\sqrt{6}}\left(1 - \frac{|u|}{\sqrt{6}}\right) 1(|u| \leq \sqrt{6})$
- Quartic (Biweight) kernel: $K(u) = \frac{15}{16}(1 - u^2)^2 1(|u| \leq 1)$
- Triweight kernel: $K(u) = \frac{35}{32}(1 - u^2)^3 1(|u| \leq 1)$

[5] Epanechnikov (1969) [5]



Rectangular
Gaussian
Epanechnikov
Triangular
−4
−2
0
2
4

- Higher-order kernels can be used. See Hansen (2005) [8], and Section 1.11 of Li and Racine (2007) [15] for details.

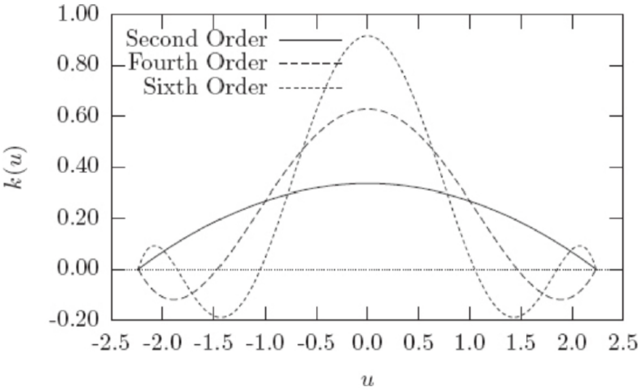


Figure 1.2: Epanechnikov kernels of varying order.

Bandwidth

Definition 17.3

- A bandwidth or tuning parameter $h > 0$ is a real number used to control the degree of smoothing of a nonparametric estimator.

- Larger values of h result in smoother estimators.
- Smaller values h result in less smooth estimators.

Properties of Kernel Density Estimator

- Invariance to rescaling the kernel function and bandwidth: The estimator (2) using $K(u)$ and h is equal for any $b > 0$ to the one using $K\left(\frac{u}{b}\right)$ and $\frac{h}{b}$.
- Invariance to data scaling: Suppose that $Y = cX$ for some $c > 0$, which means the (true) density of Y is

$$f_Y(y) = \frac{f_X(\frac{y}{c})}{c}.$$

Letting $\hat{f}_X(x)$ and $\hat{f}_Y(x)$ be the estimator (2) using $\{X_i\}_{i=1}^n$ and h and the one using $\{Y_i\}_{i=1}^n = \{cX_i\}_{i=1}^n$ and ch, respectively, Then, it holds that

$$\hat{f}_Y(y) = \frac{\hat{f}_X(\frac{y}{c})}{c}.$$

- The kernel density estimator (2) is non-negative, and integrates to 1: Letting $u = \dfrac{X_i - x}{h}$, we obtain

$$\begin{aligned}\int_{-\infty}^{\infty} \hat{f}(x)dx &= \frac{1}{nh}\sum_{i=1}^{n}\int_{-\infty}^{\infty} K\left(\frac{X_i - x}{h}\right) dx \\ &= \frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{\infty} K(u)du = 1.\end{aligned}$$

where the second equality holds because

$$dx = d(X_i + hu) = hdu$$

Bias, Variance, MSE

Bias, Variance, MSE

- The mean squared error (MSE) of a generic estimator $\hat{f}(x)$ can be decomposed as follows:

$$
\begin{aligned}
\text{MSE}\left(\hat{f}(x)\right) &\equiv \mathbb{E}\left[\{\hat{f}(x) - f(x)\}^2\right] \\
&= \mathbb{E}\left[\{\hat{f}(x) - \mathbb{E}[\hat{f}(x)] + \mathbb{E}[\hat{f}(x)] - f(x)\}^2\right] \\
&= \mathbb{E}\left[\{\hat{f}(x) - \mathbb{E}[\hat{f}(x)]\}^2\right] + \left[\mathbb{E}[\hat{f}(x)] - f(x)\right]^2 \\
&\equiv \text{var}\left(\hat{f}(x)\right) + \left[\text{bias}\left(\hat{f}(x)\right)\right]^2.
\end{aligned}
$$

Bias Evaluation

- Since $\{X_i\}_{i=1}^n$ is an IID sample, it holds that

$$\mathbb{E}[\hat{f}(x)] = \mathbb{E}\left[\frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right)\right]$$
$$= \mathbb{E}\left[\frac{1}{h}K\left(\frac{X - x}{h}\right)\right].$$

- By definition,

$$\mathbb{E}\left[\frac{1}{h}K\left(\frac{X - x}{h}\right)\right] = \int_{-\infty}^{\infty} \frac{1}{h}K\left(\frac{v - x}{h}\right) f(v)dv.$$

- Let $u = \dfrac{v-x}{h}$. Under certain conditions, we obtain

$$
\begin{aligned}
\int_{-\infty}^{\infty} \frac{1}{h} K\left(\frac{v-x}{h}\right) f(v)dv &= \int_{-\infty}^{\infty} K(u) f(x+hu) du \\
&= \int_{-\infty}^{\infty} K(u) \left\{ f(x) + f'(x)hu + \frac{1}{2} f''(x)h^2u^2 + o(h^2) \right\} du \\
&= f(x) \int_{-\infty}^{\infty} K(u)du + f'(x)h \int_{-\infty}^{\infty} uK(u)du \\
&\quad + \frac{h^2}{2} f''(x) \int_{-\infty}^{\infty} u^2 K(u)du + o(h^2) \\
&= f(x) + 0 + \frac{h^2}{2} f''(x)\kappa_2 + o(h^2)
\end{aligned}
$$

- Thus, the bias of the kernel density estimator (2) is described as

$$
\text{bias}\left(\hat{f}(x)\right) \equiv \mathbb{E}[\hat{f}(x)] - f(x) = \frac{h^2}{2} f''(x)\kappa_2 + o(h^2).
$$

Theorem 17.1

- Letting $\mathcal{N}$ denote the neighborhood of x, assume that $f(x)$ is continuous in $\mathcal{N}$. Then, as $h \to 0$,

$$\mathbb{E}[\hat{f}(x)] \to f(x).$$

- Assume additionally that $f''(x)$ is continuous in $\mathcal{N}$. Then, as $h \to 0$,

$$\text{bias}\left(\hat{f}(x)\right) \equiv \mathbb{E}[\hat{f}(x)] - f(x) = \frac{h^2}{2} f''(x)\kappa_2 + o(h^2).$$

Variance Evaluation

Theorem 17.2

- Assume that $f(x)$ is continuous in $\mathcal{N}$. Then, as $h \to 0$ and $nh \to \infty$,

$$\text{var}\left(\hat{f}(x)\right) = \frac{R_K f(x)}{nh} + o\left(\frac{1}{nh}\right)$$

where $R_K = \int K^2(u)du$ denotes the roughness of $K(u)$.

- The variance of kernel density estimator can be estimated by the sample analogue of $\mathbb{E}\left[\{\hat{f}(x) - \mathbb{E}[\hat{f}(x)]\}^2\right]$, or by $\dfrac{\kappa \hat{f}(x)}{nh}$.

MSE Evaluation

- Combining Theorems 17.1 and 17.2, we obtain the following result:

Theorem 1.1 of Li and Racine (2007) [15]

- Supppose that $f(x)$ is three-times differentiable.
- Assume that $K(\cdot)$ satisfies (3) and (4).
- As $n \to \infty$, $h \to 0$ and $nh \to \infty$,

$$\text{MSE}\left(\hat{f}(x)\right) = \frac{h^4}{4}\left[\kappa_2 f''(x)\right]^2 + \frac{\kappa f(x)}{nh} + o\left(h^4 + \frac{1}{nh}\right),$$

where $\kappa_2 = \int u^2 K(u)du$ and $\kappa = \int K^2(u)du$.

- This result implies that $\text{MSE}\left(\hat{f}(x)\right) \to 0$ and that $\hat{f}(x)$ is a consistent estimator of $f(x)$.

Take Away

- バイアス，分散，MSE それぞれの漸近的な評価のために必要な仮定については，Li and Racine (2007, Chapter 1) [15] や西山・人見 (2023，第 1 章) [28] が詳しい．
- 漸近的な評価を導出するために必要な定理や補題については，Li and Racine (2007, Appendix A) [15]，清水 (2021，第 4 章) [24]，西山・人見 (2023，第 1 章) [28] や瀬戸・細川 (2024，第 5 章) [27] が詳しい．

IMSE, AIMSE

IMSE, AIMSE

- The integrated mean squared error (IMSE) is a useful measure of precision of a kernel density estimator:

$$\text{IMSE} = \int_{-\infty}^{\infty} \text{MSE}\left(\hat{f}(x)\right) dx = \int_{-\infty}^{\infty} \mathbb{E}\left[\{\hat{f}(x) - f(x)\}^2\right] dx$$

- Suppose that $f''(x)$ is uniformly continuous. By similar arguments as we discuss MSE, it can be shown that as $n \to \infty$, $h \to 0$, and $nh \to \infty$,

$$\text{IMSE} = \frac{1}{4} R(f'') h^4 + \frac{\kappa}{nh} + o\left(h^4 + \frac{1}{nh}\right), \tag{7}$$

where $R(f'') = \int \{f''(x)\}^2 \, dx$ denotes the roughness of $f''(x)$.

- The leading term in (7) is called the asymptotic integrated mean squared error (AIMSE).

Optimal Bandwidth

- Bias-Variance Trade-Off: The first term of AIMSE is increasing in h, while the second term is decreasing in h.
- For a fixed second-order $K(\cdot)$, we can obtain AIMSE optimal bandwidth h_0 by solving the FOC: [6]

$$h_0 = \left(\frac{R_K}{R(f'')}\right)^{\frac{1}{5}} n^{-\frac{1}{5}}. \tag{8}$$

- In reality, AIMSE optimal h_0 depends on the second derivative of unknown $f(x)$. Thus, researchers need to select a bandwidth h by certain procedures. [7]

[6] Explanations on the estimation of $R(f'')$ can be found in Hall and Marron (1987) [7] and Jones and Sheather (1991) [12] among others.

[7] See Sections 17.9-17.11 and 17.15 of Hansen (2022) [9] for further discussions on bandwidth selection.

Optimal Kernel

Theorem 17.4

- AIMSE is minimized by the Epanechnikov kernel:[a]

$$K(u) = \frac{3}{4\sqrt{5}}\left(1 - \frac{u^2}{5}\right)1(|u| \leq \sqrt{5})$$

[a] Epanechnikov (1969) [5]

- See Section 17.8 of Hansen (2022) [9] for the proof.
- Imai and Okamoto (2024) [10] and Kanaya and Okamoto (2025) [13] suggest to use other kernel functions for certain optimality.

Appendix: Application to Pretesting for Manipulation

Manipulation Testing

- In (sharp) regression discontinuity designs, researchers would like to examine whether the identification condition

$$\text{For each } t = 0, 1,$$
$$\mathbb{E}[Y_i(t)|X_i = x] \text{ is continuous at } x = c$$

  is actually fulfilled.
- The fundamental problem: Note that only one out of $\mathbb{E}[Y_i(0)|X_i = x]$ and $\mathbb{E}[Y_i(1)|X_i = x]$ can be observed. That is, the above condition cannot be directly examined.
- Instead, researchers often examine certain necessary conditions (manipulation test).

Examples of Manipulation Testing

- Testing the continuity of the density of the assignment variable:[8]
  - McCrary (2008) [16]
  - Otsu, Xu and Matsushita (2013) [17]
  - Cattaneo, Jansson and Ma (2020) [4]
- Testing other necessary conditions:
  - Lee (2008) [14]
  - Canay and Kamat (2018) [3]
  - Fusejima, Ishihara and Sawada (2024) [6]
  - Arai, Hsu, Kitagawa, Mourifie and Wan (2022) [2]
- Let us briefly review the heuristic idea of the local polynomial density estimator proposed by Cattaneo, Jansson and Ma (2020) [4].

[8] Cited from Yanagi, 2024 (a lecture slide).

Cattaneo, Jansson and Ma (2020, JASA)

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

Taylor & Francis
Taylor & Francis Group

Check for updates

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.

www.tandfonline.com/doi/full/10.1080/01621459.2019.1635480

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) [11] 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.

Heuristic Idea

- Whereas other nonparametric density estimators are constructed by soothing out a histogram-type estimator of the density, the estimator proposed by Cattaneo, Jansson and Ma (2020) [4] is constructed by smoothing out the empirical distribution function using local polynomial techniques.
- Accordingly, their 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).

Take Away

- See Cattaneo, Jansson and Ma (2020) [4] and their supplementary materials for details.
- Software packages: `rddensity`, `lpdensity`

References

- Abadie, A., and Cattaneo, M. D. (2018). Econometric methods for program evaluation. *Annual Review of Economics*, **10**, 465-503.
- Arai, Y., Y. C. Hsu, T. Kitagawa, I. Mourifie and Y. Wan (2022). Testing identifying assumptions in fuzzy regression discontinuity designs. *Quantitative Economics 13*, 1-28.
- Canay, I. A. and V. Kamat (2018). Approximate permutation tests and induced order statistics in the regression discontinuity design. *Review of Economic Studies 85*, 1577-1608.

References ii

- Cattaneo, M. D., M. Jansson and X. Ma (2020). Simple local polynomial density estimators. *Journal of the American Statistical Association 115*(531), 1449-1455.
- Epanechnikov, V. A. (1969). "Non-parametric Estimation of a Multivariate Probability Density," *Theory of Probability and its Applications, 14,* 153-158.
- Fusejima, K., T. Ishihara and M. Sawada (2024). A unified diagnostic test for regression discontinuity designs. arXiv:2205.04345v4.
- Hall, P. and J. S. Marron (1987). "Estimation of Integrated Squared Density Derivatives," *Statistics & Probability Letters, 6,* 109-115.

- Hansen, B. E. (2005). "Exact Mean Integrated Squared Error of Higher Order Kernel Estimators," *Econometric Theory, 21,* 1031-1057.
- Hansen, B. E. (2022). *Probability and Statistics for Economists.* Princeton.
- Imai, S. and Y. Okamoto (2024). "Kernel Choice Matters for Boundary Inference Using Local Polynomial Density: With Application to Manipulation Testing," arXiv:2306.07619.
- Imbens, G. W., and Rubin, D. B. (2015). *Causal Inference in Statistics, Social, and Biomedical Sciences*. New York: Cambridge University Press.

- Jones, M. C. and S. J. Sheather (1991). "Using Non-Stochastic Terms to Advantage in Kernel-Based Estimation of Integrated Squared Density Derivatives," *Statistics & Probability Letters, 11*, 511-514.
- Kanaya, S. and Y. Okamoto (2025). "Optimal Kernels for Density and Derivative Estimation Revisited," Presented at Kansai Econometric Study Group, January 12, 2025, Kobe, Japan.
- Lee, D. S. (2008). Randomized experiments from non-random selection in US House elections. *Journal of Econometrics 142*(2), 675-697.

References v

- Li, Q. and J. S. Racine (2007). *Nonparametric Econometrics: Theory and Practice.* Princeton.
- McCrary (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics 142*(2), 698-714.
- Otsu, T., K. L. Xu, and Y. Matsushita (2013). Estimation and inference of discontinuity in density. *Journal of Business and Economic Statistics 31*(4), 507-524.
- Pagan, A. and A. Ullah (1999). *Nonparametric Econometrics.* Cambridge.

References vi

- Parzen, E. (1962). "On the Estimation of a Probability Density Function and Mode," *Annals of Mathematical Statistics, 33,* 1065-1076.
- Rosenblatt, M. (1956). "Remarks on Some Nonparametric Estimates of a Density Function," *Annals of Mathematical Statistics, 27,* 832-837.
- Scott, D. W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization.* Wiley.
- Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman and Hall.
- van der Vaart, A. W. (2000). *Asymptotic Statistics.* Cambridge.

- 清水泰隆 (2021)『統計学への確率論，その先へ』内田老鶴圃.
- 清水泰隆 (2023)『統計学への漸近論，その先は』内田老鶴圃.
- 末石直也 (2015) 『計量経済学：ミクロデータ分析へのいざない』日本評論社.
- 瀬戸道生，細川卓也 (2024) 『関数解析入門のためのフーリエ変換・ラプラス変換・積分方程式・ルベーグ積分』内田老鶴圃.
- 西山慶彦，人見光太郎 (2023) 『ノン・セミパラメトリック統計解析』共立出版.