

# Nonparametric Density Estimation

Sections 17.1-17.8 of Hansen (2022)

---

Yasuyuki Matsumura (Kyoto University)

Last Updated: April 22, 2025

<https://yasu0704xx.github.io>

## About Me



I am a second-year master's student in [Graduate School of Economics, Kyoto University](#).

Under the guidance of [Yoshihiko Nishiyama](#) and [Takahide Yanagi](#), my research focuses on Econometric Theory and Statistics.

Contact: [matsumura.yasuyuki.w85 \[at\] kyoto-u.jp](mailto:matsumura.yasuyuki.w85[at]kyoto-u.jp)

Address: Yoshida Honmachi, Sakyo, Kyoto 606-8501, Japan

Other Links: [Linktree](#)

This slide is 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.

- Here we review Sections 17.1-17.8 of Hansen (2022) [3].
- 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  $X_1, \dots, X_n$ .
- 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  $X$ .

- Excellent textbooks on nonparametric density estimation include Silverman (1986) [11] and Scott (1992) [10].
- The following textbooks are often referred:
  - Silverman (1986) [11],
  - Scott (1992) [10],
  - van der Vaart (2000, Chapter 24) [12],
  - Pagan and Ullah (1999, Chapter 2) [7], and
  - Li and Racine (2007, Chapter 1) [6].
- 日本語の文献：
  - 西山・人見 (2023, 第1章) [16]
  - 末石 (2015, 第9章) [15]
  - 清水 (2023, 第5章) [14]

Idea behind Kernel Density Estimation

Kernel Density Estimator

Bias, Variance, MSE

IMSE, AIMSE

Refernces

## Idea behind Kernel Density Estimation

---

# Histogram

- A simple and familiar estimator of  $f(x)$  is a histogram.
- Devide 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:

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

- 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.,<sup>1</sup> we obtain

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

---

<sup>1</sup>We discuss these convergences in Chapter 18 of Hansen (2022) [3].

# 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  $h \rightarrow 0$ , it holds that

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

- Replacing  $f$  and  $F$  with  $f_n$  and  $F_n$ , respectively, we obtain the naive estimator<sup>2</sup> of  $f(x)$ :

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

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

---

<sup>2</sup>The Rosenblatt estimator (Rosenblatt, 1956) [9]

- The naive estimator of  $\phi(x)$  using IID observations  $X_1, \dots, X_{100} \sim \text{Normal}(0, 1)$ :<sup>3</sup>

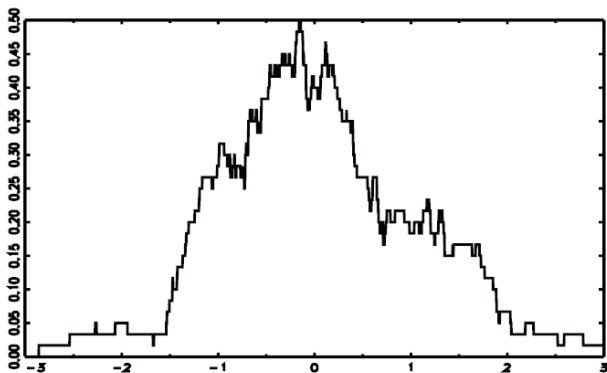


図 1: ナイーブ推定量 ( $h = 0.3$ )

---

<sup>3</sup>Cited from [the lecture note 01](#) by N. Sueishi.

# Idea behind Kernel Density Estimation

- The naive estimator can be rewritten as

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

where  $k_0(\cdot)$  is given by

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

- 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<sup>4</sup> of  $f(x)$  is given by

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

- $K(u)$  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.

---

<sup>4</sup>The Parzen-Rosenblatt estimator (Parzen, 1962) [8]

## 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

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

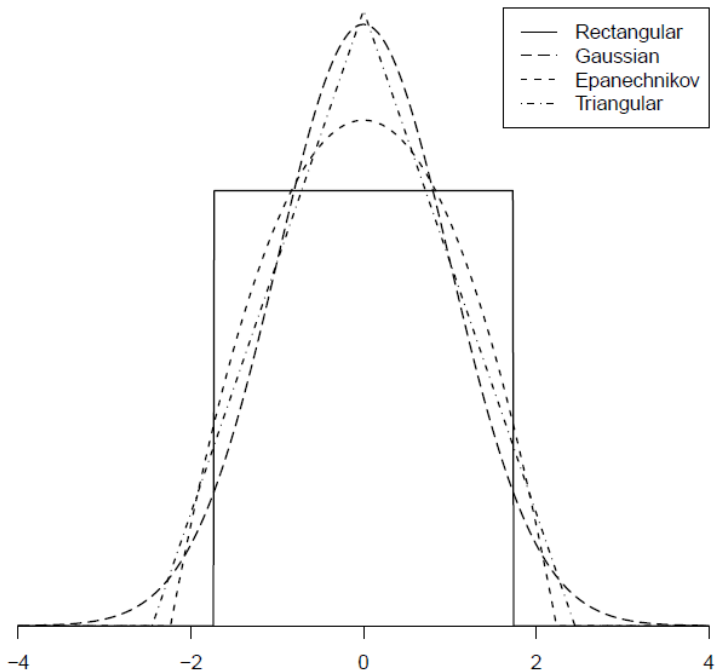
- The order of a kernel  $\nu$  is defined as the order of the first non-zero moment.

## 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:<sup>5</sup>  $K(u) = \frac{3}{4\sqrt{5}} \left(1 - \frac{u^2}{5}\right) 1(|u| \leq \sqrt{5})$
- 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:  $K(u) = \frac{35}{32}(1 - u^2)^3 1(|u| \leq 1)$

---

<sup>5</sup>Epanechnikov (1969) [1]



# Higher-Order Kernel

- Higher-order kernels can be used. See Section 1.11 of Li and Racine (2007) [6] for details.

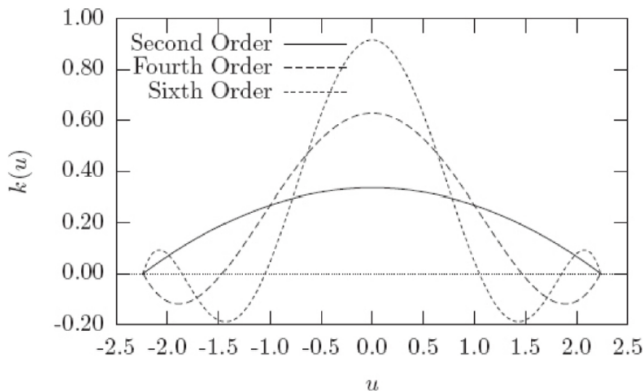


Figure 1.2: Epanechnikov kernels of varying order.

## 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

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

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 hold that

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

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

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

## Bias, Variance, MSE

---



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

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

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

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

- 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  $u = \frac{v-x}{h}$ . 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  $h \rightarrow 0$ ,

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

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

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

## Theorem 17.2

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

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

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  $\frac{\kappa \hat{f}(x)}{nh}$ .

# MSE Evaluation

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

## Theorem 1.1 of Li and Racine (2007) [6]

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

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

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

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

- 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) [6] や西山・人見 (2023, 第1章) [16] 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) [6], 清水 (2021, 第4章) [13] や西山・人見 (2023, 第1章) [16] が詳しい.

## IMSE, AIMSE

---



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

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

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

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

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: <sup>6</sup>

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

- 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. <sup>7</sup>

---

<sup>6</sup>Explanations on the estimation of  $R(f'')$  can be found in Hall and Marron (1987) [2] and Jones and Sheather (1991) [4] among others.

<sup>7</sup>See Sections 17.9-17.11 and 17.15 of Hansen (2022) [3] for further discussions on bandwidth selection.

## Theorem 17.4

- Letting  $h$  be the AIMSE optimal bandwidth  $h_0$  given in (8), AIMSE is minimized by the Epanechnikov kernel:<sup>a</sup>

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





---





<sup>a</sup>Epanechnikov (1969) [1]







- See Section 17.8 of Hansen (2022) [3] for the proof.
- Kanaya and Okamoto (2025) [5] suggest to use another kernel function for certain optimality.

## Refernces

---

-  Epanechnikov, V. A. (1969). "Non-parametric Estimation of a Multivariate Probability Density," *Theory of Probability and its Applications*, 14, 153-158.
-  Hall, P. and J. S. Marron (1987). "Estimation of Integrated Squared Density Derivatives," *Statistics & Probability Letters*, 6, 109-115.
-  Hansen, B. E. (2022). *Probability and Statistics for Economists*. Princeton.
-  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.
-  Li, Q. and J. S. Racine (2007). *Nonparametric Econometrics: Theory and Practice*. Princeton.
-  Pagan, A. and A. Ullah (1999). *Nonparametric Econometrics*. Cambridge.
-  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, その先へ』 内田老鶴圃.

-  末石直也 (2015) 『計量経済学：マイクロデータ分析へのいざない』 日本評論社.
-  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.