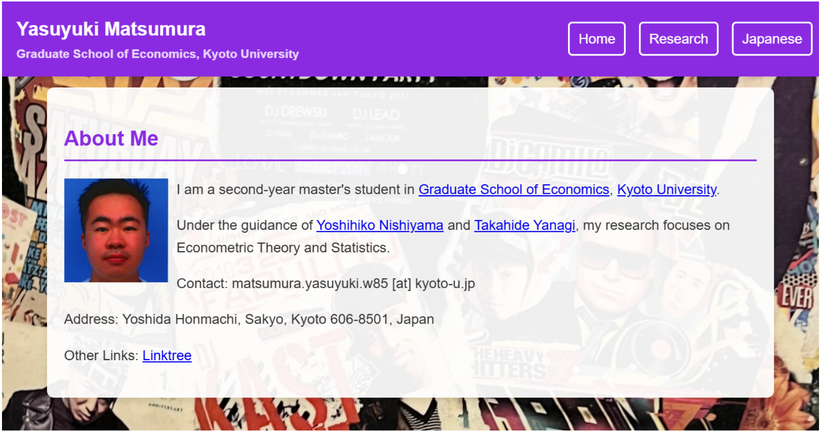
Nonparametric Density Estimation

Sections 17.1-17.8 of Hansen (2022)

Yasuyuki Matsumura (Kyoto University)

Last Updated: April 21, 2025

https://yasu0704xx.github.io



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 disirable 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) [1].
- 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.

References

- Excellent textbooks on nonparametric density estimation include Silverman (1986) [7] and Scott (1992) [6].
- The following textbooks are often refered:
  - Silverman (1986) [7],
  - Scott (1992) [6],
  - van der Vaart (1998, Chapter 24) [8],
  - Pagan and Ullah (1999, Chapter 2) [3], and
  - Li and Racine (2007, Chapter 1) [2].
- 日本語の文献：
  - 西山・人見 (2023，第 1 章) [11]
  - 末石 (2015，第 9 章) [10]
  - 清水 (2023，第 5 章) [9]

Contents

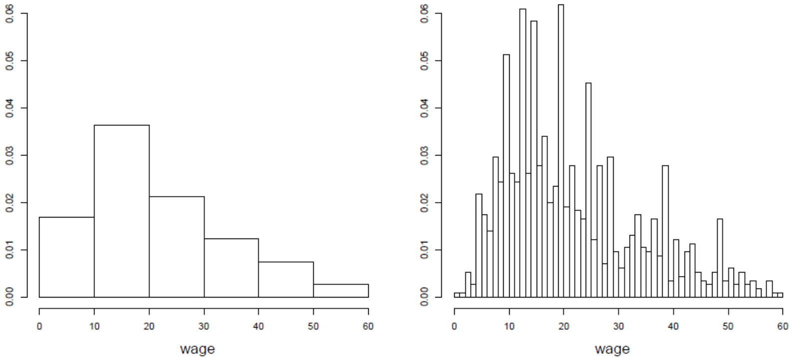
Idea behind Kernel Density Estimation

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:

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

- 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) [1].

Naive Estimator

- Since $F_n(x)$ is discrete, 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 and F with f_n and F_n, respectively, we obtain the naive estimator (the Rosenblatt 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] Rosenblatt (1956) [5]

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

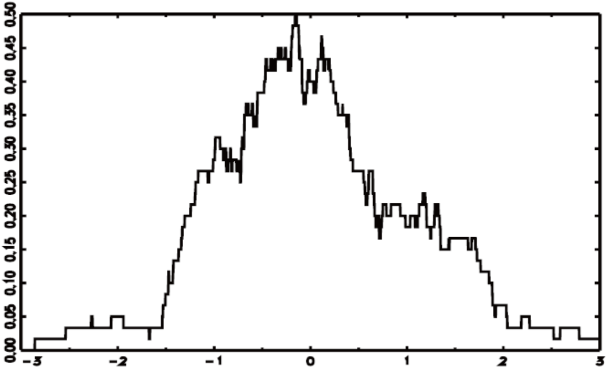


図 1: ナイーブ推定量 ($h = 0.3$)

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

Problems arising with Naive Estimator

- Although $f_n(x)$ is an asymptotically "good" estimator, it is still displayed as rectangles, and then looks different from $f(x)$.
- Also, $f_n(x)$ is not differentiable.

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) [4]

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

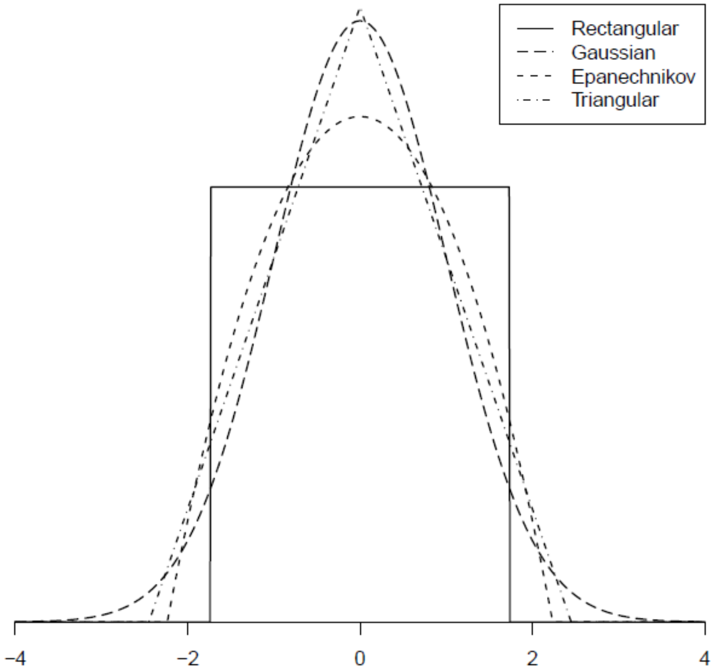
$$\int_{-\infty}^{\infty} u^2 K(u) du = 1.$$

- 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.

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


Rectangular
Gaussian
Epanechnikov
Triangular
−4
−2
0
2
4

Bandwidth

Bias, Variance

Integrated Mean Squared Error (IMSE)

Optimal Kernel

Refernces

References i

- Hansen, B. E. (2022). *Probability and Statistics for Economists.* Princeton.
- 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.

References ii

- 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. (1998). *Asymptotic Statistics.* Cambridge.
- 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.
- 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.