ノンパラメトリックカーネル密度推定

Li & Racine (2007, Sections 1.1 – 1.5)

松村 泰行（京都大学経済学研究科・博士後期課程１年）

最終更新: 2026 年 4 月 13 日

matsumura.yasuyuki.w85@kyoto-u.jp
https://yasu0704xx.github.io

参考文献

- Li & Racine (2007) Chapter 1
  - Chapter 1 のみ出版社のページから PDF をダウンロードできます。
  - Errata
  - Solutions Manual
  - R code
- Tsybakov (2009) Chapter 1
- 末石 (2015) 第 9 章
- 西山・人見 (2023) 第 1 章
- R パッケージ：`np`

- このスライドの最新版は ここ にあります。

目次（教科書：第 1.1 – 1.5 節）

ノンパラメトリック法とは何か？

カーネル密度推定量（１次元の場合）

漸近理論

バンド幅の選択

バイアスの処理

分布関数の推定

補遺：数学の準備

ノンパラメトリック法とは何か？

ノンパラメトリックな問題とは何か？

- **パラメトリック**な問題
  - 推定対象を有限次元のパラメータで表現できる場合。
  - 例えば，線形回帰モデル $Y = \beta_1 X_1 + \beta_2 X_2 + \varepsilon$ を推定したいとする。
  - 推定する母数は $(\beta_1, \beta_2) \in \mathbb{R}^2$ なので，これはパラメトリックな問題である。

- **ノンパラメトリック**な問題
  - 推定対象を有限次元のパラメータで表現できない場合。
  - 例えば，モデル $Y = g(X) + \varepsilon$ を推定したいとする。
  - $g(\cdot)$ の関数形についての情報が何もないとする。
  - $g(\cdot)$ は無限次元の母数なので，これはノンパラメトリックな問題である。

- セミパラメトリックな問題
  - 有限次元の母数と無限次元の母数の組合せを推定する場合。
  - 興味のある部分のみパラメトリックに表現し，興味のない部分（局外パラメータ）はノンパラメトリックに表現することがある。

  - 部分線形回帰モデル (Robinson, 1988)

$$Y = \beta_1 X_1 + \beta_2 X_2 + g(Z) + \varepsilon$$

  - インデックスモデル (Ichimura, 1993, etc.)

$$Y = g(\beta_1 X_1 + \beta_2 X_2) + \varepsilon$$

なぜノンパラメトリック法を用いるのか？

- 確率変数の分布や回帰関数の関数形などは，ノンパラメトリックな母数である。
    - 元来，そうである（はず）。
- ノンパラメトリックな母数に対して何か特定の分布や関数形を仮定することは，ややもすれば致命的な間違いの原因になり得る。
    - 誤差項は本当に正規分布にしたがっているのか？
    - 回帰関数は本当に（有限次元の母数に関して）線形なのか？

パラメトリック法とノンパラメトリック法のどちらを使うべきか？

- もし仮にパラメトリックな仮定が正しいのであれば，パラメトリックモデルを想定して母数を推定するのが**効率性** (efficiency) の観点からは望ましいだろう。
- しかし，**頑健性** (robustness) を考慮すると，パラメトリック法よりも弱い仮定で済むノンパラメトリック法が望ましい。

特定化の誤り

例 (Li and Racine, Example 1.1)

- 未知の確率変数 X の密度関数 $f(x)$ に興味があるとする。分析者は IID 標本 X_i $(i = 1, \cdots, n)$ を観測しているとする。
- モデル $X \sim \text{Normal}(\mu_0, \sigma_0^2)$ を仮定する。推定対象は $(\mu_0, \sigma_0^2) \in \mathbb{R}^2$ である。
- μ_0 の最尤推定量は標本平均 $\widehat{\mu} = \sum_{i=1}^n X_i/n$ であり，σ_0^2 の最尤推定量は標本分散 $\widehat{\sigma}^2 = \sum_{i=1}^n (X_i - \widehat{\mu})^2/n$ である。
- もし仮に正規分布の仮定が正しいならば，一般的な極値推定量の漸近理論から，$n \to \infty$ のとき，一致性

$$\widehat{\mu} \xrightarrow{p} \mu_0, \qquad \widehat{\sigma}^2 \xrightarrow{p} \sigma_0^2$$

が成り立つ。さらに，最尤推定量は漸近的に有効でもある。

- では，もし仮に $X \sim \mathrm{Normal}(\mu, \sigma^2)$ というパラメトリックモデルの仮定が間違っていたらどうなるだろうか？
- 先ほどの MLE で考えていた尤度関数を $L(\mu, \sigma^2)$ と書くことにする。
- 特定化が誤っているので，$\widehat{\mu}$ や $\widehat{\sigma}^2$ は擬似最尤推定量 (QMLE)

$$\underset{(\mu, \sigma^2)}{\arg\max}\, L(\mu, \sigma^2)$$

であり，擬似真値 (pseudo-true value) [1] に対する一致性しか得られない。

$$\widehat{\mu} \xrightarrow{p} \mu_{\mathrm{pseudo}}, \qquad \widehat{\sigma}^2 \xrightarrow{p} \sigma^2_{\mathrm{pseudo}}$$

- 一般に $\mu_0 \neq \mu_{\mathrm{pseudo}}$, $\sigma_0^2 \neq \sigma^2_{\mathrm{pseudo}}$ なので，QMLE は漸近的にもバイアスを持つ。

[1] $\mathrm{KLD}\,(f_0(x) || f(x;\theta))$ を最小にする θ のこと。詳しくは Hansen (2022, Chapters 8 and 23) 等を参照。

代表的なノンパラメトリック法

局所的推定と呼ばれる方法

- **カーネル推定（教科書** Chapters 1 – 6）
- k-nearest neighbor **推定（教科書** Chapter 14）

大域的推定と呼ばれる方法

- **シリーズ推定（教科書** Chapter 15）

カーネル密度推定量（１次元の場合）

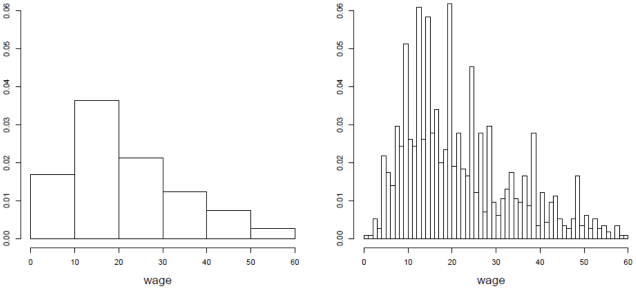
設定

- ある 1 次元の確率変数 $X \sim F(x)$ の IID 標本 $X_i \ (i = 1, \cdots, n)$ を観測している。
- $F(x)$ は連続（すなわち，滑らかな関数）であるとする。

- 分布関数 $F(x)$ に特定化の仮定を置かずに，密度関数 $f(x)$ を推定したい。
- つまり，密度関数 $f(x)$ をノンパラメトリックに推定したい。

ヒストグラム

- もっともシンプルな密度推定は，ヒストグラムを用いることである。
- 密度関数 $f(x)$ のサポートを J 個の幅 h の区間に分割したうえで，第 j 区間における推定量を $\widehat{f}_j(x) = n_j/nh$ （n_j は区間 j 内の観測値の個数）とすればよい。



(a) Bin Width = 10   (b) Bin Width = 1

Figure 17.1: Histogram Estimate of Wage Density for Asian Women

Hansen (2022, Chapter 17)

- ヒストグラムは直観的で簡単である。

- しかし，問題もある。
- 3 変量以上のとき，図に描くのが困難である。
- 各区間の端点近傍でのパフォーマンスには疑問が残る。
- 幅 h の良い決め方は非自明である。
- 連続でない（微分できない）ので，漸近理論を考えるのが難しい。

ナイーブ推定量

- 確率変数 X の密度関数 (PDF) は

$$f(x) = \frac{d}{dx}F(x) = \frac{d}{dx}\Pr(X \leq x) \tag{1}$$

によって定義される。ただし，$F(x)$ は累積分布関数 (CDF) である。

- $F(x)$ に対する最も自然な推定量は，経験分布関数 (EDF) である：

$$\widehat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}(X_i \leq x).$$

- 一般に，EDF は CDF の一致推定量であり，$\sqrt{n}$-consistency や漸近正規性を持つ。一様収束することとと，そのオーダーも知られている。

- (1) に $\widehat{F}_n(x)$ を代入したいが，$\widehat{F}_n(x)$ は指示関数を含むため微分できない。

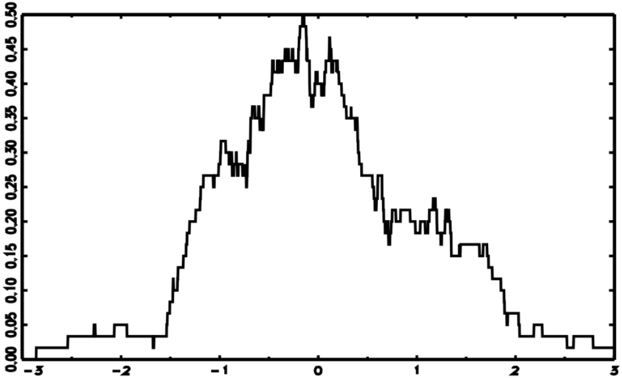
- かわりに，微分の定義

$$f(x) = \lim_{h \to 0} \frac{F(x+h) - F(x)}{h} = \lim_{h \to 0} \frac{F\left(x + \frac{h}{2}\right) - F\left(x - \frac{h}{2}\right)}{h}$$

にしたがって，次のような近似を考えてみる：

$$\begin{aligned}
\widehat{f}_{\mathrm{naive}}(x) &= \frac{\widehat{F}_n\left(x + \frac{h}{2}\right) - \widehat{F}_n\left(x - \frac{h}{2}\right)}{h} \\
&= \frac{1}{nh} \sum_{i=1}^{n} \left[\mathbf{1}\left(X_i \leq x + \frac{h}{2} \right) - \mathbf{1}\left(X_i \leq x - \frac{h}{2} \right) \right] \\
&= \frac{1}{nh} \sum_{i=1}^{n} \mathbf{1}\left(x - \frac{h}{2} < X_i \leq x + \frac{h}{2} \right) = \frac{1}{nh} \sum_{i=1}^{n} \mathbf{1}\left(-\frac{1}{2} < \frac{X_i - x}{h} \leq \frac{1}{2} \right).
\end{aligned}$$

- $\widehat{f}_{\text{naive}}(x)$ のことをナイーブ推定量[2] という。
- 標準正規分布の密度関数をナイーブ推定量 $(n = 100, h = 0.3)$ で推定した結果が，以下である（引用元：LN-1）。



[2]Rosenblatt 推定量 (Rosenblatt, 1956) ともいう。

- ナイーブ推定量も指示関数を含むため，やはり微分できない。

- $k(v) = \mathbf{1}\left(-\frac{1}{2} < v \leq \frac{1}{2}\right)$ を用いて $\widehat{f}_{\mathrm{naive}}(x)$ を書き換えると，

$$\widehat{f}_{\mathrm{naive}}(x) = \frac{1}{nh}\sum_{i=1}^{n} \mathbf{1}\left(-\frac{1}{2} < \frac{X_i - x}{h} \leq \frac{1}{2}\right) = \frac{1}{nh}\sum_{i=1}^{n} k\left(\frac{X_i - x}{h}\right)$$

となる。

- ナイーブ推定量を微分できないのは $k(v)$ が指示関数であることが原因なので，この $k(v)$ を滑らかな関数に取り換えてやるとうまくいくのではないだろうか？

カーネル密度推定量 (Kernel Density Estimator)

- $k(v)$ が以下を満たしているとする：

$$\int_{-\infty}^{\infty} k(v) dz = 1. \tag{2}$$

- このような $k(v)$ を用いた密度関数 $f(x)$ に対する推定量

$$\widehat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} k\left(\frac{X_i - x}{h} \right) \tag{3}$$

のことをカーネル密度推定量 (KDE) という。

- $k(v)$ をカーネル関数という。(2) のとおり，カーネル自身も密度関数である。
- $h > 0$ をバンド幅（平滑化パラメータ）という。

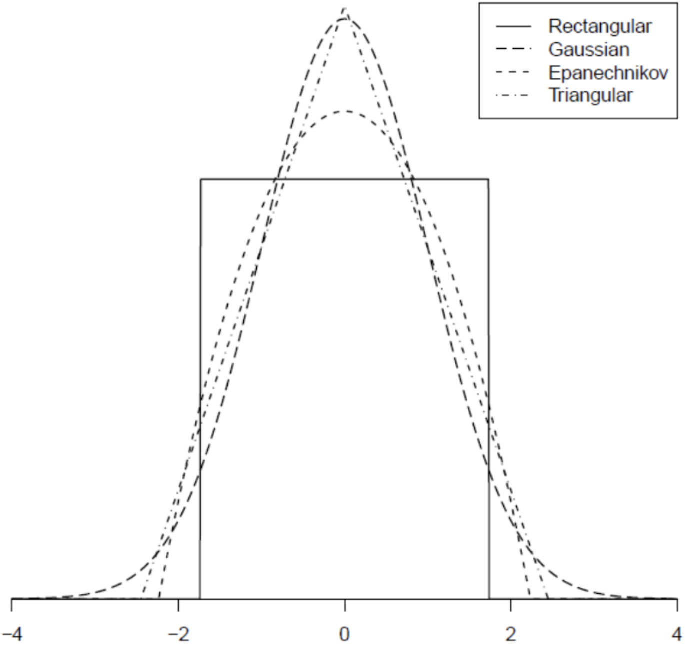
代表的なカーネル関数

- ナイーブ推定量も KDE の一種である。ナイーブ推定量のカーネル $k(v) = \mathbf{1}\left(-\frac{1}{2} < v \leq \frac{1}{2}\right)$ は一様カーネルとか rectangular カーネルと呼ばれる。

- 他によく用いられているのは，例えば，以下の 3 つのカーネルである。

- Gaussian: $k(v) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}v^2\right)$
- Epanechnikov: $k(v) = \frac{3}{4\sqrt{5}}\left(1 - \frac{v^2}{5}\right)\mathbf{1}\left(|v| \leq \sqrt{5}\right)$
- Triangular: $k(v) = \frac{1}{\sqrt{6}}\left(1 - \frac{|v|}{\sqrt{6}}\right)\mathbf{1}\left(|v| \leq \sqrt{6}\right)$

- どのカーネルを用いても，推定結果が大きく変わるといったことはない。
- 漸近的にはどのカーネルを用いても収束のオーダーには影響しない。



Rectangular
Gaussian
Epanechnikov
Triangular
−4
−2
0
2
4

バンド幅

- バンド幅を大きくとると，KDE はより滑らかになる。
  - 分散が抑えられるかわりに，バイアスが大きくなる。Oversmooth するという。
- バンド幅を小さくとると，KDE はより凸凹になる。
  - 分散は大きくなるが，バイアスを小さくできる。Undersmooth するという。

- このようなバンド幅を通じた bias-variance tradeoff は，**平均二乗誤差** (MSE) がバイアス（の二乗）と分散に分解されることを確認すると理解しやすい。

- 後ほど，Li and Racine, Theorem 1.1 において考察する。

シミュレーション

設定

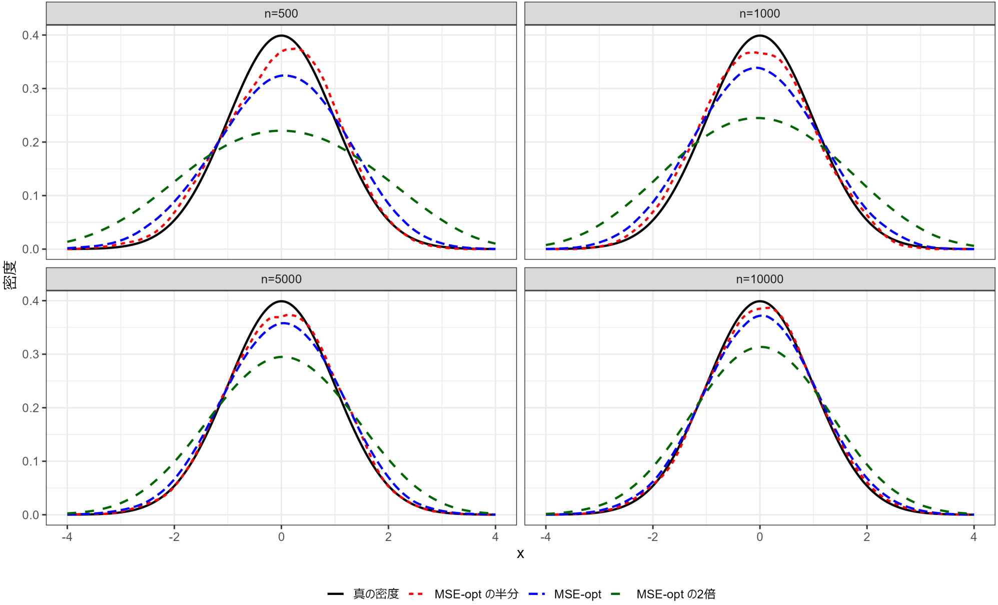
- データ： $X_1, \cdots, X_n \overset{\text{IID}}{\sim} \text{Normal}(0,1), \quad n \in \{500, 1000, 5000, 10000\}$
- バンド幅： MSE-optimal, MSE-optimal の半分, MSE-optimal の 2 倍
- カーネル： Epanechnikov

注意

- 今回は KDE のグラフを図示したいだけなので，繰り返し計算は 1 回だけ行った。
- 適切なバンド幅の選び方については後ほどきちんと考える。
- R のスクリプトはここからダウンロードできます。
- 必要に応じて `install.packages(c("np", "ggplot2"))` してください。

カーネル密度推定量

データ生成過程：標準正規分布，カーネル：Epanechnikov



漸近理論

漸近理論

平均二乗収束

- $n \to \infty$ のとき， $h \to 0$ かつ $nh \to \infty$ ならば，
- $\mathrm{MSE}\left(\widehat{f}(x)\right) \to 0$ となる。よって，$\widehat{f}(x) \xrightarrow{p} f(x)$ となる。

一様収束

- $n \to \infty$ のとき， $nh^2 \to \infty$ ならば，
- $\sup_x |\widehat{f}(x) - f(x)| \xrightarrow{p} 0$ となる。

一様収束の速度

- バンド幅 $h > 0$ について，$h = C\left(\frac{\log\log n}{n}\right)^{\frac{1}{6}}$ であるとき，
- $\sup_x |\widehat{f}(x) - f(x)| = O\left(\left(\frac{\log\log n}{n}\right)^{\frac{1}{3}}\right)$ a.s. である。

漸近正規性

- $n \to \infty$ のとき， $h \to 0$ かつ $nh \to \infty$ かつ $nh^5 = O(1)$ ならば，
- $\sqrt{nh}\left(\widehat{f}(x) - f(x) - \frac{h^2}{2} f^{(2)}(x) \int v^2 k(v) dv\right) \xrightarrow{d} \text{Normal}\left(0,\ f(x)\int k(v)^2 dv\right)$，
- もしくは，$\sqrt{nh}\left(\widehat{f}(x) - \mathbb{E}\left[\widehat{f}(x)\right]\right) \xrightarrow{d} \text{Normal}\left(0,\ f(x)\int k(v)^2 dv\right)$ となる。
- $nh^5 = o(1)$ ならば $\sqrt{nh}\left(\widehat{f}(x) - f(x)\right) \xrightarrow{d} \text{Normal}\left(0,\ f(x)\int k(v)^2 dv\right)$ となる。

平均二乗収束

Li and Racine: Theorem 1.1（平均二乗収束）

$\{X_i\}_{i=1}^n$ を $X \sim f(x)$ からの IID 標本とする。$f(x)$ は 3 階連続微分可能で，各 $f^{(s)}(x)$ はサポート上で一様に有界であるとする。x は f の連続点かつサポートの内点であるとする。カーネル関数 $k(v)$ は

$$\int k(v)dv = 1, \qquad k(v) = k(-v), \qquad 0 < \int v^2 k(v)dv = \kappa_2 < \infty. \tag{4}$$

を満たすとする。$n \to \infty$ のとき， $h \to 0$ かつ $nh \to \infty$ であるならば，

$$\mathrm{MSE}\left(\widehat{f}(x)\right) = \frac{h^4}{4}\left[\kappa_2 f^{(2)}(x)\right]^2 + \frac{\kappa f(x)}{nh} + o\left(h^4 + \frac{1}{nh}\right) = O\left(h^4 + \frac{1}{nh}\right)$$

となる。ただし，$\kappa = \int k^2(v)dv$ である。

証明の方針　平均二乗誤差をバイアスと分散に分解する。

- バイアスの評価： $\mathrm{bias}\left(\widehat{f}(x)\right) = \frac{h^2}{2}\kappa_2 f^{(2)}(x) + O(h^3)$
- 分散の評価： $\mathrm{var}\left(\widehat{f}(x)\right) = \frac{1}{nh}\{\kappa f(x) + O(h)\}$

証明　平均二乗誤差は，

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

のように，バイアス（の二乗）と分散の和に分解される。

まず，バイアスを評価しよう。バイアスの定義は

$$\mathrm{bias}\left(\widehat{f}(x)\right) \equiv \mathbb{E}\left[\widehat{f}(x)\right] - f(x) = \mathbb{E}\left[\frac{1}{nh}\sum_{i=1}^{n} k\left(\frac{X_i - x}{h}\right)\right] - f(x)$$

であるが，IID なので，

$$\mathrm{bias}\left(\widehat{f}(x)\right) = \frac{1}{h}\mathbb{E}\left[k\left(\frac{X_i - x}{h}\right)\right] - f(x)$$

と簡単にできる。期待値の定義より，

$$\mathrm{bias}\left(\widehat{f}(x)\right) = \frac{1}{h}\int k\left(\frac{z - x}{h}\right) f(z)dz - f(x)$$

である。変数変換 $v = \dfrac{z - x}{h}$ を行うと， $z = x + hv,\ dz = hdv$ に注意して，

$$\mathrm{bias}\left(\widehat{f}(x)\right) = \int k(v)f(x + hv)dv - f(x)$$

となる。

$f(x)$ は 3 階連続微分可能で，各 $f^{(s)}(x)$ は一様に有界 $(s = 0, 1, 2, 3)$ であると仮定していたので，$f(x + hv)$ に関して以下のテイラー展開が成り立つ。

$$\begin{aligned}\mathrm{bias}\left(\widehat{f}(x)\right) &= \int k(v)\left\{f(x) + f^{(1)}(x)hv + \frac{h^2v^2}{2}f^{(2)}(x) + \frac{h^3v^3}{6}f^{(3)}(\tilde{x})\right\}dv - f(x) \\ &= f(x)\int k(v)dv + f^{(1)}(x)h\int vk(v)dv + \frac{h^2}{2}f^{(2)}(x)\int v^2k(v)dv + O(h^3) \\ &\quad - f(x)\end{aligned}$$

カーネル $k(v)$ は (4) を満たすので（2 次カーネルなので），バイアスは

$$\mathrm{bias}\left(\widehat{f}(x)\right) = \frac{h^2}{2}f^{(2)}(x)\int v^2k(v)dv + o(h^2)$$

のように書くことができる。

次に，分散を評価しよう。

$$\mathrm{var}\left(\widehat{f}(x)\right) = \mathrm{var}\left[\frac{1}{nh}\sum_{i=1}^{n} k\left(\frac{X_i - x}{h}\right)\right] = \frac{1}{n^2h^2}\mathrm{var}\left[\sum_{i=1}^{n} k\left(\frac{X_i - x}{h}\right)\right]$$

IID なので，

$$\mathrm{var}\left(\widehat{f}(x)\right) = \frac{1}{nh^2}\mathrm{var}\left[k\left(\frac{X_i - x}{h}\right)\right]$$

である。分散は「二乗の期待値」と「期待値の二乗」の差なので，

$$\mathrm{var}\left(\widehat{f}(x)\right) = \frac{1}{nh^2}\left\{\mathbb{E}\left[k^2\left(\frac{X_i - x}{h}\right)\right] - \mathbb{E}\left[k\left(\frac{X_i - x}{h}\right)\right]^2\right\}$$

である。期待値の定義より，バイアスの計算と同様に積分の形でかける。

$$\mathrm{var}\left(\widehat{f}(x)\right) = \frac{1}{nh^2}\left\{\int k^2\left(\frac{z - x}{h}\right) f(z)dz - \left(\int k\left(\frac{z - x}{h}\right) f(z)dz\right)^2\right\}$$

変数変換 $v = \dfrac{z-x}{h}$ を行うと， $z = x + hv,\ dz = hdv$ に注意して，

$$
\begin{aligned}
\operatorname{var}\left(\widehat{f}(x)\right) &= \frac{1}{nh^2}\left\{\int k^2(v)f(x+hv)hdv - \left(\int k(v)f(x+hv)hdv\right)^2\right\} \\
&= \frac{1}{nh}\left\{\int k^2(v)f(x+hv)dv - h\left(\int k(v)f(x+hv)dv\right)^2\right\}
\end{aligned}
$$

となる。やはり，$f(x+hv)$ に関して以下のテイラー展開が成り立つ。

$$
\begin{aligned}
\operatorname{var}\left(\widehat{f}(x)\right) &= \frac{1}{nh}\left\{\int k^2(v)\left\{f(x) + f^{(1)}(\tilde{x})hv\right\}dv - h\left(f(x) + \operatorname{bias}\left(\widehat{f}(x)\right)\right)^2\right\} \\
&= \frac{1}{nh}\left\{f(x)\int k^2(v)dv + O(h)\right\} \\
&= \frac{1}{nh}\left\{\kappa f(x) + O(h)\right\}
\end{aligned}
$$

バイアスと分散のオーダーはそれぞれ，

$$\mathrm{bias}\left(\widehat{f}(x)\right) = \frac{h^2}{2} f^{(2)}(x) \int v^2 k(v) dv + o(h^2), \quad \mathrm{var}\left(\widehat{f}(x)\right) = \frac{1}{nh}\left\{\kappa f(x) + O(h)\right\}$$

であることが分かったので，平均二乗誤差は次のように評価される。

$$\mathrm{MSE}\left(\widehat{f}(x)\right) = \frac{h^4}{4}\left(\kappa_2 f^{(2)}(x)\right)^2 + \frac{\kappa f(x)}{nh} + o\left(h^4 + \frac{1}{nh}\right) = O\left(h^4 + \frac{1}{nh}\right)$$

□

補足 (Li and Racine, Theorem 1.1)

- Thm 1.1 よりただちに，$\widehat{f}(x) - f(x) = O_p\left(h^2 + \frac{1}{\sqrt{nh}}\right) = o_p(1)$ も成り立つ。
  - ただし，X のサポートの内点における理論保証であることに注意が必要である。
  - 境界付近では「片側のみ」にしか観測値が存在しないため，バイアスを持つ。
  - 境界バイアスに対しては LPD (Cattaneo, Jansson and Ma, 2020) が有用である。
- 追加的に $f(x)$ が一様連続であることを仮定すると，

$$\sup_x \mathrm{MSE}\left(\widehat{f}(x)\right) \to 0$$

  も成り立つ。
- 一様確率収束 $\sup_x \left|\widehat{f}(x) - f(x)\right| \xrightarrow{p} 0$ を主張するためにはさらに強い条件 $(nh^2 \to \infty)$ が必要である。

IMSE-optimal なバンド幅

- Theorem 1.1 の結果より，$f(x)$ の一致推定のみが目的である場合は，バンド幅は，$n \to \infty$ のときに $h \to 0$ と $nh \to \infty$ を満たすように，

$$h = c \times n^{-\frac{1}{\alpha}} \quad \exists c > 0, \quad \exists \alpha > 1$$

のオーダーで選べばよい。c や α は，通常，Integrated MSE (IMSE, MISE)

$$\mathrm{IMSE}\left(\widehat{f}\right) \equiv \int \mathbb{E}\left[\widehat{f}(x) - f(x)\right]^2 dx = \frac{h^4 {\kappa_2}^2}{4} \int \left[f^{(2)}(x)\right]^2 dx + \frac{\kappa}{nh} + o\left(h^4 + \frac{1}{nh}\right)$$

を最小化するようなバンド幅

$$h = c_0 \times n^{-\frac{1}{5}} \qquad c_0 = {\kappa_2}^{-\frac{2}{5}} \kappa^{\frac{1}{5}} \left(\int \left[f^{(2)}(x)\right]^2 dx\right)^{-\frac{1}{5}} (> 0)$$

を選択する。各点 x での pointwise な MSE ではなくて，X のサポート全体で MSE を最小化したいからである。

- c_0 は未知なので，何らかの方法で推定する必要がある（後ほど考える）。

一様収束とその速度

西山・人見：定理 1.13（一様収束）

$\{X_i\}_{i=1}^n$ を $X \sim f(x)$ からの IID 標本とする。$f(x)$ は 3 階連続微分可能で，各 $f^{(s)}(x)$ $(s = 0,1,2,3)$ はサポート上で一様に有界であるとする。x は f の連続点かつサポートの内点であるとする。カーネル関数は (4) と $|vk(v)| \to 0$ (as $|u| \to \infty$) を満たし，さらに，$k(v)$ の特性関数 $\psi(t) = \int \exp(itv)k(v)dv$ が絶対可積分であるとする。$n \to \infty$ のとき $nh^2 \to \infty$ であるならば，一様確率収束

$$\sup_x \left| \widehat{f}(x) - f(x) \right| \xrightarrow{p} 0$$

が成り立つ。

西山・人見：定理 1.14（一様収束の速度）

$\{X_i\}_{i=1}^n$ を $X \sim f(x)$ からの IID 標本とする。$f(x)$ は 3 階連続微分可能で，各 $f^{(s)}(x)$ $(s = 0, 1, 2, 3)$ はサポート上で一様に有界であるとする。x は f のサポートの内点であるとする。 カーネル関数は (4) と $|vk(v)| \to 0$ (as $|u| \to \infty$) を満たし，さらに，以下を満たしているものとする。

- $k(v)$ のサポートを $a = v_0 < v_1 < \cdots < v_M = b$ によって分割する。[a]
- $k(v)$ の全変動は有界 $\mu = \sup_{v_0, \cdots, v_M} \sum_m |k(v_{m+1}) - k(v_m)| < \infty$ である。[b]

（続く）

[a] $a = -\infty, b = \infty$ でもかまわない。
[b] $k(v)$ が微分可能なとき $\mu = \int |k'(v)|\, dv$ である。

（続き）

バンド幅 $h > 0$ について，$h = C\left(\frac{\log\log n}{n}\right)^{\frac{1}{6}}$ であるとき，

$$\sup_x \left|\widehat{f}(x) - f(x)\right| = O\left(\left(\frac{\log\log n}{n}\right)^{\frac{1}{3}}\right) \text{ a.s.}$$

が成り立つ。

- 教科書 Sections 10 and 12 では，多変量の場合の KDE の一様収束とその速度について考察している。証明はノートを参照してください。
- 平均二乗収束のときと同様，$\sup$ をとる範囲は f のサポートの内点の全体であり，バイアスの出る境界は除外して考えている。

漸近正規性の準備：三角配列 (Triangular Arrays)

- 極値推定量の漸近正規性を導出する際には，Lindeberg-Levy CLT を用いた。

Li and Racine: Lemma A.3 (Lindeberg-Levy CLT)

X_i $(i = 1, \cdots, n)$ が IID で，期待値 μ と分散 σ^2 が存在するとき，

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - \mu) \xrightarrow{d} \text{Normal}(0, \sigma^2)$$

が成り立つ。

- KDE の漸近正規性を考える際，Lindeberg-Levy CLT を用いることはできない。
- $n \to \infty$ のときに $h \to 0$ である必要があるため，

$$\frac{1}{h} k \left(\frac{X_i - x}{h} \right)$$

が n にも依存するからである（バンド幅が h_n であるということ）。

- このように $\{Z_{n,i} : 1 \leq i \leq n\}$ の形で表される確率変数の列を三角配列と呼ぶ。
- KDE の漸近正規性を導くためには，三角配列に対する中心極限定理である Liapunov CLT を利用する必要がある。
- Billingsley (1995) *Probability and Measure* の Chapter 27 等を参照されたい。

漸近正規性の準備：Liapunov CLT

Li and Racine: Lemma A.5 (Liapunov CLT)

確率変数列 $\{Z_{n,i} : 1 \leq i \leq n\}$ について，

- $Z_{n,1}, \cdots, Z_{n,n}$ が独立で，平均 $\mu_{n,i}$，分散 $\sigma^2_{n,i}$ であり，
- $\mathbb{E}\left[|Z_{n,i}|^{2+\delta}\right] < \infty$ なる $\delta > 0$ が存在し，
- $S_n = \sum_{i=1}^n Z_{n,i}$，$\sigma_n^2 = \mathrm{var}(S_n) = \sum_{i=1}^n \sigma^2_{n,i}$ とすると，$\sigma_n^2 = \sigma^2 + o(1)$（$\sigma^2$ は定数）であり，
- 後述する Liapunov 条件 (5) が満たされているとき，

中心極限定理が以下のように成り立つ。

$$\frac{1}{\sigma_n}(S_n - \mathbb{E}[S_n]) = \frac{1}{\sigma_n}\sum_{i=1}^n (Z_{n,i} - \mathbb{E}[Z_{n,i}]) \xrightarrow{d} \text{Normal}(0,1)$$

漸近正規性の準備：Liapunov 条件

Liapunov 条件

$$\lim_{n\to\infty}\sum_{i=1}^{n}\mathbb{E}\left[\left|Z_{n,i}-\mu_{n,i}\right|^{2+\delta}\right]=0 \quad \exists\,\delta>0 \tag{5}$$

漸近正規性

西山・人見：定理 1.15 & 定理 1.16（漸近正規性）

$\{X_i\}_{i=1}^n$ を $X \sim f(x)$ からの IID 標本とする。$f(x)$ はサポート上で一様に有界であるとする。カーネル関数は，$|vk(v)| \to 0$ (as $|u| \to \infty$) と $\sup |k(v)| < \infty$ を満たすものとする。x を f のサポートの内点であるとする。$n \to \infty$ のとき $nh \to \infty$ であるならば，漸近正規性

$$\sqrt{nh}\left(\widehat{f}(x) - \mathbb{E}\left[\widehat{f}(x)\right]\right) \xrightarrow{d} \text{Normal}\left(0, \kappa f(x)\right)$$

が成り立つ。さらに，追加的にバンド幅が $nh^5 \to 0$ を満たすとき，

$$\sqrt{nh}\left(\widehat{f}(x) - f(x)\right) \xrightarrow{d} \text{Normal}\left(0, \kappa f(x)\right)$$

が成り立つ。ただし，$\kappa = \int k^2(v)dv$ である。

補足（西山・人見：定理 1.15 & 定理 1.16）

- Li and Racine, Theorem 1.1 でやったように，分散のオーダーは

$$\text{var}\left(\widehat{f}(x)\right) = \frac{1}{nh}\left\{\kappa f(x) + O(h)\right\}$$

なので，収束のレートは $\sqrt{nh}$ である。

- 多変量の場合の収束のレートは $\sqrt{nh_1, \cdots h_q}$ である（教科書 Section 1.9）。

- $\mathbb{E}\left[\widehat{f}(x)\right] = f(x) + \text{bias}\left(\widehat{f}(x)\right) = \frac{h^2}{2} f^{(2)}(x) \int v^2 k(v) dv + o(h^2)$ であるので，$\mathbb{E}\left[\widehat{f}(x)\right]$ の周りでの分布を求めることになる。

- IMSE-optimal なバンド幅は $nh^5 = O(1)$ であるため，バイアスが残る。ナイーブな解決方法は，$nh^5 \to 0$ なるバンド幅を選ぶ過小平滑化である。[3]

[3] しかし，後述する理由によって，過小平滑化は非推奨である。

Liapunov 条件の十分条件

- Liapunov 条件 (5) を直接確かめるのは面倒かもしれない。実際，

$$\sum_{i=1}^{n} \mathbb{E}\left[|v_{n,i}|^{2+\delta}\right], \quad \text{where } v_{n,i} = \frac{k_{n,i} - \mathbb{E}\left[k_{n,1}\right]}{\sqrt{n \times \mathrm{var}\left(k_{n,i}\right)}}, \quad k_{n,i} = \frac{1}{h} k\left(\frac{X_i - x}{h}\right)$$

  の極限について調べることになる（西山・人見の p.21 を参照されたい）。

- Pagan and Ullah (1999, p.40) では，十分条件

$$\int k(v)^{2+\delta} dv < \infty, \quad \exists\, \delta > 0$$

  をチェックすることを提案している。この条件であれば，$k(v)$ が非負かつ有界であり，全範囲の積分は 1 になることより，

$$\int k(v)^{2+\delta} dv \leq \sup_{v} k(v)^{1+\delta} \int k(v) dv = \sup_{v} k(v)^{1+\delta} < \infty$$

  のようにして確認できる。

バンド幅の選択

IMSE-optimal なバンド幅

- IMSE-optimal なバンド幅は，

$$h_{\text{opt}} = c_0 \times n^{-\frac{1}{5}} \qquad c_0 = \kappa_2^{-\frac{2}{5}} \kappa^{\frac{1}{5}} \left(\int \left[f^{(2)}(x) \right]^2 dx \right)^{-\frac{1}{5}}$$

  であるが，c_0 は未知なので，何らかの方法で推定する必要がある。

1. Plug-in 法
2. Silverman (1986) のバンド幅
3. Rule-of-Thumb
4. 最小二乗クロスバリデーション (LS-CV)
5. 最尤クロスバリデーション (Likelihood-CV)

Plug-in 法

- Plug-in 法とは，h の適当な初期値を分析者が選択し c_0 のうち未知の部分

$$\int \left[f^{(2)}(x)\right]^2 dx$$

をカーネル法で推定してから，その推定値を代入して，KDE に用いる最適なバンド幅 h_{opt} を求める方法のことである。
- h の初期値の選び方は様々である。
- ここでは，Silverman (1986) の方法をとりあげる。

Silverman (1986) のバンド幅

- Silverman (1986) は h の初期値として

$$h_{\text{pilot}} = 1.06\sigma \times n^{-\frac{1}{5}} \tag{6}$$

を用いることを提案した。

- Silverman (1986) は，初期値の式 (6) にあらわれる 1.06σ という数値を，以下のような計算によって求めている。
- $X \sim \text{Normal}(\mu, \sigma^2)$ というパラメトリックな仮定をおくと，

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \Longrightarrow \int \left[f^{(2)}(x)\right]^2 dx = \frac{3}{8}\pi^{-\frac{1}{2}}\sigma^5$$

となる。

- Gaussian カーネルを用いるとき $\kappa_2{}^{-\frac{2}{5}}\kappa^{\frac{1}{5}} = (4\pi)^{-\frac{1}{10}}$ となる。

- すると，バンド幅の初期値として

$$(4\pi)^{-\frac{1}{10}} \times \left(\frac{3}{8}\pi^{-\frac{1}{2}}\sigma^5\right)^{-\frac{1}{5}} = (4\pi)^{-\frac{1}{10}} \times \left(\frac{3}{8}\pi^{-\frac{1}{2}}\right)^{-\frac{1}{5}} \times \sigma$$

  を使うことが考えられる。
- $(4\pi)^{-\frac{1}{10}} \times \left(\frac{3}{8}\pi^{-\frac{1}{2}}\right)^{-\frac{1}{5}}$ の部分を計算するとおよそ 1.06 になる。
- (6) にあらわれる σ は未知である。Silverman (1986) は，

$$\min\left\{\text{標準偏差}, \quad \frac{75\%\text{分位点} - 25\%\text{分位点}}{1.34}\right\}$$

  を用いることを提案している。
- Silverman (1986) は，正規分布に限らず，X が特定のパラメトリックな分布族に属するという仮定の下で $\int \left[f^{(2)}(x)\right]^2 dx$ を計算することを提案している。

Rule-of-Thumb

- Silverman (1986) が提案した初期値 (6) をそのまま h_{opt} として用いることもある。これは Silverman の Rule-of-Thumb と呼ばれている。
- 一般に，$\int \left[f^{(2)}(x)\right]^2 dx$ を推定するのに最適なバンド幅と，$f(x)$ を推定するのに最適なバンド幅は大きく異なるので，h_{pilot} がうまく機能するとは限らない。
- X の分布が正規分布に近いならば，(6) は最適なバンド幅に近いと考えられる。

- しかし，気になる点は残る。
- そもそも，分布形に仮定を置かないことが，ノンパラメトリック法を用いるモチベーションである。
- にもかかわらず，最適なバンド幅を選ぶために分布形を仮定してしまうというのは，いわばダブルスタンダードのようなもので，納得できないかもしれない。

Leave-one-out KDE

- クロスバリデーションによるバンド幅の選択を考えるために，まず，1 個抜きカーネル密度推定量 (Leave-one-out KDE) を定義しておく。

1 個抜きカーネル密度推定量 (Leave-one-out KDE)

i 番目の観測値を除く $(n-1)$ 個の観測値を用いた KDE を

$$\widehat{f}_{-i}(x) = \frac{1}{(n-1)h} \sum_{j=1,\, j \neq i}^{n} k\left(\frac{X_i - X_j}{h}\right)$$

で定義する。

最小二乗クロスバリデーション (Least Squares Cross-Validation)

- 最小二乗クロスバリデーションによって選ばれるバンド幅は

$$h_{\mathrm{CV}} = \arg\min_{h} \left[\int \left(\widehat{f}(x) \right)^2 dx - \frac{2}{n} \sum_{i=1}^{n} \widehat{f}_{-i}(X_i) \right]$$

である。h_{CV} は，IMSE の不偏推定量を最小にするような h を選んでいる。

- h_{CV} は以下のように導出される。
- まず，IMSE は，

$$\begin{aligned} \mathrm{IMSE}(h) &= \int \mathbb{E} \left[\left(\widehat{f}(x) - f(x) \right)^2 \right] dx \\ &= \mathbb{E} \left[\int \left(\widehat{f}(x) \right)^2 dx \right] - 2\mathbb{E} \left[\int \widehat{f}(x) f(x) dx \right] + \int (f(x))^2 dx \end{aligned} \tag{7}$$

のように分解できる。

- (7) の第 3 項は h に関する最小化には関係ないので無視して良い。
- 第 1 項 $\mathbb{E}\left[\int \left(\widehat{f}(x)\right)^2 dx\right]$ に対する不偏推定量として

$$\int \left(\widehat{f}(x)\right)^2 dx \tag{8}$$

を使うことが考えられる。なお，計算上は，(8) を数値積分するのではなく，

$$\begin{aligned}
\int \left(\widehat{f}(x)\right)^2 dx &= \int \left(\frac{1}{nh}\sum_{i=1}^{n} k\left(\frac{X_i - x}{h}\right)\right)^2 dx \\
&= \frac{1}{n^2h^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\int k\left(X_i - x\right) k\left(X_j - x\right) dx \\
&= \frac{1}{n^2h}\sum_{i=1}^{n}\sum_{j=1}^{n}\bar{k}\left(\frac{X_i - X_j}{h}\right), \quad \text{with} \quad \bar{k}(x) = \int k(v)k(x-v)dx
\end{aligned}$$

のように k の畳み込み (convolution) を使って解析的に求める方が単純である。

- (7) の第 2 項のうち $\mathbb{E}\left[\int \widehat{f}(x)f(x)dx\right]$ について，もし仮に $\widehat{f}(x)$ がランダムではなく確定的な関数であるとすると，

$$\int \widehat{f}(x)f(x)dx = \mathbb{E}\left[\widehat{f}(X_i)\right]$$

となるので，KDE の標本平均 $\frac{1}{n}\sum_{i=1}^{n}\widehat{f}(X_i)$ を用いた推定が可能になる。
- しかし，実際には $\widehat{f}(x)$ は X_i にも依存するランダムな関数なので，X_i を用いてこれを推定してもバイアスが生じる。
- そこで，先ほど定義しておいた leave-one-out KDE を用いた標本平均

$$\frac{1}{n}\sum_{i=1}^{n}\widehat{f}_{-i}(X_i) = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{1}{(n-1)h}\sum_{j\neq i}k\left(\frac{X_j - X_i}{h}\right)\right\}$$

を用いる。これが $\mathbb{E}\left[\int \widehat{f}(x)f(x)\right]$ に対する不偏推定量になっていることは，簡単な計算によって確認できる。

- 以上をまとめて，クロスバリデーション基準 (CV criterion) は

$$\mathrm{CV}_f(h) = \int \left(\widehat{f}(x)\right)^2 dx - \frac{2}{n}\sum_{i=1}^{n} \widehat{f}_{-i}(X_i)$$

によって与えられる。バンド幅 h に依存しない定数部分を除けば，$\mathrm{CV}_f(h)$ は $\mathrm{IMSE}(h)$ の不偏推定量になっている。

- IMSE を最小化するバンド幅を $h_0 = \arg\min_h \mathrm{IMSE}(h)$ とおくと，$n \to \infty$ のとき

$$\frac{h_{\mathrm{CV}} - h_0}{h_0} \xrightarrow{p} 0$$

が成り立つ。ただし，Hardle, Hall and Marron (1988) によると，収束レートは $O_p\left(n^{-\frac{1}{10}}\right)$ で遅い。

- 最小二乗クロスバリデーションに関するコメントは，教科書 pp.17–18 や過去のレクチャーノート（末石，2014）を参照されたい。

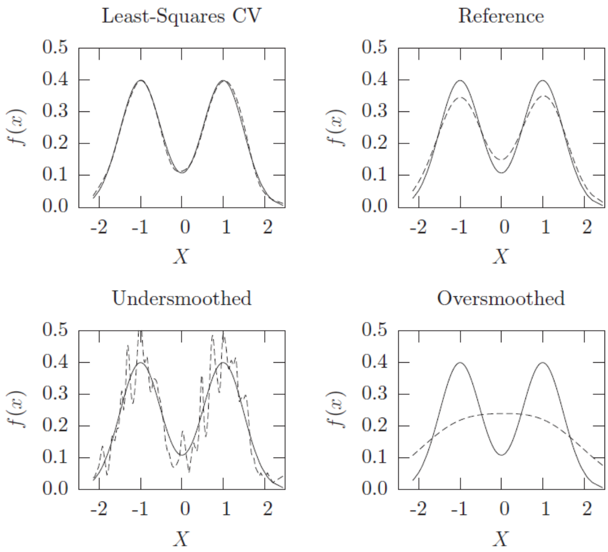
最尤クロスバリデーション (Likelihood Cross-Validation)

- 最尤クロスバリデーションでは，

$$\arg\max_h \sum_{i=1}^{n} \log \widehat{f}_{-i}(X_i)$$

を最大化するようにバンド幅を選択する。
- 基本的なアイデアは，良いバンド幅を用いて推定された KDE は真の密度関数と（KLD の意味で）近いはずだ，ということである。
- 真の分布が裾の厚い分布であるとき，最尤 CV で選ばれたバンド幅は一致性を失ってしまう。

バンド幅の比較 (Li and Racine, 2007, Figure 1.1)



IMSE-optimal バンド幅は $oh^5 = o(1)$ を満たさない

- 漸近正規性の主張は，$nh^5 = O(1)$ **のとき**

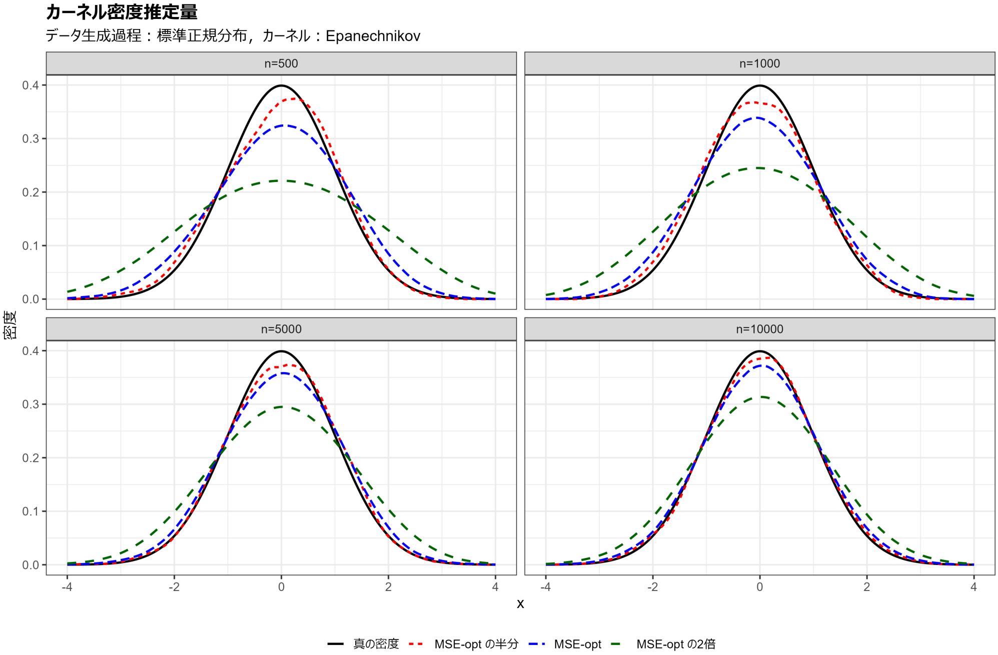
$$\sqrt{nh}\left(\widehat{f}(x) - f(x) - \frac{h^2}{2}f^{(2)}(x)\int v^2 k(v)dv\right) \xrightarrow{d} \text{Normal}\left(0,\ f(x)\int k(v)^2 dv\right)$$

でバイアスが残り，$nh^5 = o(1)$ **のとき**

$$\sqrt{nh}\left(\widehat{f}(x) - f(x)\right) \xrightarrow{d} \text{Normal}\left(0,\ f(x)\int k(v)^2 dv\right)$$

でバイアスが消えるというものであった。

- IMSE-optimal なバンド幅は $h = O(n^{-1/5})$ なので，$nh^5 = O(1)$ である。
- つまり，IMSE-optimal なバンド幅を選択すると，**中心化のために未知のバイアスを知ることが必要**となってしまう。これは応用上の大きな問題である。



カーネル密度推定量
データ生成過程：標準正規分布，カーネル：Epanechnikov
n=500
n=1000
n=5000
n=10000
0.4
0.3
0.2
0.1
0.0
密度
-4
-2
0
2
4
x
真の密度
MSE-opt の半分
MSE-opt
MSE-opt の2倍

バイアスの処理

バイアスの処理

- 2 次カーネルと IMSE-optimal なバンド幅を選択すると，未知の漸近バイアスが残ってしまうことを確認した。
- 過小平滑化 (undersmoothing) でバイアスを消すことはできるが，分散が大きくなってしまうし，分析者の主観に依存するので，あまり良い方法とは言えない。

バイアスの軽減方法

1. 高次カーネルの利用　（教科書 Section 1.11）
2. 局所多項式密度推定 (LPD; Cattaneo, Jansson and Ma, 2020)

バイアスを補正する方法

1. ロバストバイアス補正推論 (RBC Inference; Calonico, Cattaneo and Farrell, 2018)
2. 一般化ジャックナイフ法 (Schucany and Sommers, 1977)
3. 平滑化ブートストラップ

参考文献

- 西山 (2026) は，KDE のバイアス処理についてまとめた概説論文である。
- ブートストラップについては，西山・人見 (2023) の第 5 章が特に詳しい。

高次カーネル

- カーネル $k(v)$ が

$$\int k(v)dv = 1, \quad \int v^l k(v)dv = 0\,(l = 1, \cdots, j-1), \quad \int v^j k(v)dv = \kappa_j > 0$$

を満たすとき，$k(v)$ は j 次カーネルであるという。

- たとえば，(4) の条件を満たすカーネルは 2 次カーネルである。
- Li and Racine, Theorem 1.1 の証明を高次カーネルに拡張する。テイラー展開の次数を上げると，

$$\text{bias}\left(\widehat{f}(x)\right) = \frac{h^j}{j!}\kappa_j f^{(j)}(x) + O(h^{j+1})$$

のように，バイアスのオーダーが $O(h^j)$ に改善される。

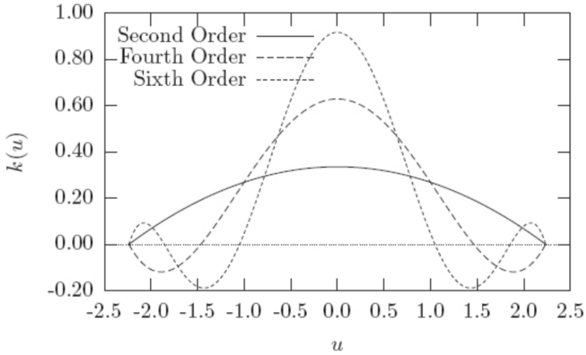


Figure 1.2: Epanechnikov kernels of varying order.

Li and Racine (2007, Figure 1.2)

- 悪い面もある。
- 高次カーネルはサポート上で負になる部分があるため，推定値が負になり得る。
- すると，密度関数の推定量としての解釈が難しくなってしまう。

局所多項式密度推定 (Local Polynomial Density Estimator)

- Cattaneo, Jansson and Ma (2020) は，局所多項式密度推定 (LPD)

$$\widehat{f}(x) = e_1^\top \widehat{\beta}(x), \qquad e_1 = (0, 1, 0, \cdots, 0) \in \mathbb{R}^{p+1}$$

$$\widehat{\beta}(x) = \mathop{\arg\min}_{b \in \mathbb{R}^{p+1}} \sum_{i=1}^{n} \left(\widehat{F}_n(X_i) - b_0 - b_1(X_i - x) - \cdots - b_p(X_i - x)^p \right) k\left(\frac{X_i - x}{h} \right)$$

  による境界バイアスへの対処を提案した (彼らは $p = 2$ を推奨している) 。
- LPD のアイデア自体は局所多項式回帰 (LPR) で考案されたものである。[4]
- ポイントは，ヒストグラムではなく，経験分布関数を平滑化する点である。

[4] LPR について，詳細は教科書の Section 2.5 を参照されたい。

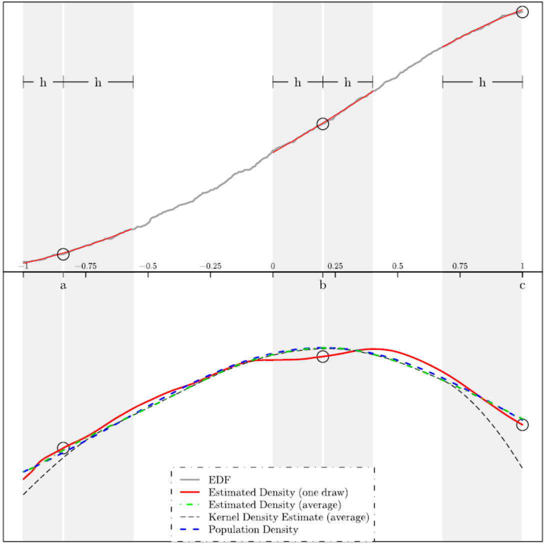


Figure 1. Graphical illustration of density estimator. Constructed using companion R (and Stata) package described in Cattaneo, Jansson, and Ma (2019) with simulated data.

Cattaneo, Jansson and Ma (2020, Figure 1)

- LPD は，x がサポートの内点でも境界でも，理論的に良い性質を持つ。
  - 例えば，$nh^2 \to \infty$, $nh^{2p+1} = O(1)$ のもとで漸近正規性が成り立つ。

$$\sqrt{nh}\left(\frac{\widehat{f}(x) - f(x) - h^p \mathcal{B}(x)}{\sqrt{\mathcal{V}(x)}}\right) \xrightarrow{d} \text{Normal}(0,1)$$

  - そのため，不連続回帰デザイン (RDD) における恣意的操作の検定 (manipulation test) でよく使われている (Cattaneo and Titiunik, 2022, Section 4.2) 。
  - R パッケージ：`lpdensity`, `rddensity`

- Imai and Okamoto (2026) は，LPD で境界の密度を推定する際は，三角カーネルではなく Gaussian カーネルか Laplace カーネルを用いることを推奨している。
  - 三角カーネルを使うと，MSE が大きくなる，信頼区間が長くなる，`rddensity` の検出力が低くなるといった問題が生じる。
  - さらに，三角カーネルのようにサポートがコンパクトなカーネルを使うと，有限標本での分散が無限大に発散することがある。
  - Imai and Okamoto (2026) の Section 2 に種々の数値実験の結果が列挙されている。

ロバストバイアス補正推論 (Robust Bias-Corrected Inference)

- Calonico, Cattaneo and Farrell (2018) は，ロバストバイアス補正 (RBC) によるバイアスの補正を提案した。
  - RBC 以前は，過小平滑化かナイーブなバイアス推定が主流であったが，どちらもあまり良い方法ではない。
  - 過小平滑化は，バイアスを無視できるが，分散が大きくなってしまうし，分析者の主観に依存してしまう。
  - ナイーブなバイアス推定は「バイアスの推定に関する不確実性」を無視しているため，最終的な KDE の分散を過小評価してしまう 。

- RBC が「頑健」なのは，バイアスの推定に関する不確実性に対してである。

- Calonico, Cattaneo and Farrell (2018) の提案する RBC は，漸近正規性をもつ。

$$\sqrt{nh}\left(\frac{\widehat{f}(x) - \widehat{\mathcal{B}}(x) - f(x)}{\widehat{\sigma}(x)}\right) \xrightarrow{d} \text{Normal}(0,1) \tag{9}$$

- $\widehat{\mathcal{B}}(x)$ はバイアスの推定量である。

$$\widehat{\mathcal{B}}(x) = \mu_L(k)\,\tilde{f}^{(L)}(x)\,h^L, \qquad \mu_L(k) = \frac{1}{L!}\int u^L k(u)du$$

ただし，$\tilde{f}^{(L)}(x)$ は，カーネル関数 k_{pilot} と バンド幅 h_{pilot} を用いた，密度関数の L 次微分 $f^{(L)}$ に対するパイロット推定量である。

$$\tilde{f}^{(L)}(x) = \frac{1}{n h_{\text{pilot}}{}^{L+1}} \sum_{i=1}^{n} k_{\text{pilot}}\left(\frac{X_i - x}{h_{\text{pilot}}}\right)$$

- $\widehat{\sigma}^2(x)$ は $\widehat{f}_{\text{RBC}}(x) = \widehat{f}(x) - \widehat{\mathcal{B}}(x)$ のバイアス補正を考慮した分散推定量である。

- (9) に基づく信頼区間は，高次の意味で，ナイーブなバイアス補正よりも望ましい被覆確率を達成する。

- バンド幅 h, h_{pilot} は，(9) に基づく信頼区間の被覆確率の誤差を小さくするように選ぶことが推奨されている。
  - $h = h_{\mathrm{pilot}}$ としてよい。
  - h と h_{pilot} の比率が漸近的な性質に影響を与えることが分かっている。

- このバンド幅は，MISE-optimal なバンド幅よりも小さいオーダーになるため，実質的には過小平滑化の一種ではないかという批判もある。[5]

[5]たしかに過小平滑化ではあるが，理論的根拠に基づく一定の基準を与えているという意味で，RBC を用いるのが望ましいように思う。

一般化ジャックナイフ法

準備中

平滑化ブートストラップ

準備中

分布関数の推定

KDE から分布関数を「復元」する

- 既に見たとおり，経験分布関数 (EDF) は分布関数 (CDF) の良い推定量である。
- しかし，EDF は滑らかではないので，応用上は不都合が生じるかもしれない。
- 滑らかな CDF の推定量が欲しい場合には，KDE を積分してはどうだろうか？

$$\widehat{F}_K(x) = \int_{-\infty}^{x} \widehat{f}(v)dv = \frac{1}{n}\sum_{i=1}^{n} G\left(\frac{x - X_i}{h}\right) \quad \text{with} \quad G(x) = \int_{-\infty}^{x} k(v)dv$$

- 1 次の漸近理論では，$\widehat{F}_K(x)$ は EDF と同じ漸近分布をもたらす。
- 一方で，$\widehat{F}_K(x)$ は高次の効率性の意味において EDF よりも優れている。
- これは，IMSE を考えるとわかりやすい。

MSE, IMSE

Li and Racine: Theorem 1.2

$F(x)$ が 2 階連続微分可能で，$d^2F(x)/dx^2$ がヘルダー連続であるとする。$k(v) = dG(v)/dv$ が有界・対称で，サポートはコンパクトであるとする。バンド幅が $0 \le h \le C\,n^{-\varepsilon}$, $\exists 0 < \varepsilon < 1/8$ であるとすると，$n \to \infty$ のとき，

$$\mathrm{MSE}\left[\widehat{F}(x)\right] = c_0(x)\frac{1}{n} - c_1(x)\frac{h}{n} + c_2(x)h^4 + o\left(h^4 + \frac{h}{n}\right),$$

$$\mathrm{IMSE}\left[\widehat{F}(x)\right] = C_0(x)\frac{1}{n} - C_1(x)\frac{h}{n} + C_2(x)h^4 + o\left(h^4 + \frac{h}{n}\right)$$

が成り立つ。

- 各 c, C の定義や証明は，ノートを参照してください。

バンド幅

- Theorem 1. より，IMSE-optimal なバンド幅のオーダーは $n^{-1/3}$ である。

$$h_0 = a_0 \times n^{-\frac{1}{3}}, \quad \text{with} \quad a_0 = \left(\frac{C_1}{4C_2}\right)^{\frac{1}{3}}$$

- Bowman, Hall and Prvan (1998) は CV を提案した。

$$\mathrm{CV}_F(h) = \frac{1}{n}\sum_{i=1}^{n}\int\left\{\mathbf{1}\left(X_i \le x\right) - \widehat{F}_{-i}(x)\right\}^2 dx,\ \widehat{F}_{-i}(x) = \frac{1}{n-1}\sum_{j\neq i} G\left(\frac{x - X_j}{h}\right)$$

- $h_{\mathrm{CV}} = \arg\min_h \mathrm{CV}_F(h)$ とすると，

$$\frac{h_{\mathrm{CV}}}{h_0} \xrightarrow{p} 1$$

が成り立つ。

高次の効率性

- Bowman, Hall and Prvan (1998) のバンド幅を用いると，一次の漸近理論では

$$\sqrt{n}\left(\widehat{F}_K(x) - F(x)\right) \xrightarrow{d} \text{Normal}\left(0,\ F(x)\left(1 - F(x)\right)\right)$$

であり，EDF $\widehat{F}_N(x)$ と同じ漸近分布をもたらす。

- IMSE-optimal なバンド幅 $h = O\left(n^{-\frac{1}{3}}\right)$ を用いるとき，$\widehat{F}_K(x)$ の IMSE は

$$\text{IMSE}\left(\widehat{F}_K(x)\right) = O\left(n^{-1} - n^{-\frac{4}{3}}\right)$$

のオーダーである。一方で，EDF $\widehat{F}_N(x)$ の IMSE は

$$\text{IMSE}\left(\widehat{F}_N(x)\right) = \frac{1}{n}\int \left(F(x)\left(1 - F(x)\right)\right) dx = O(n^{-1})$$

のオーダーである。

- よって，高次の意味では，EDF よりも $\widehat{F}_K(x)$ の方が効率的であると考えられる。

補遺：数学の準備

経験分布関数 (EDF) の性質

- 西山・人見（2023，第 1.1 節）が簡潔でわかりやすい。
- さらに詳細な議論は，van der Vaart and Wellner (2023, Chapter 2) などの経験過程に関するテキストを参照されたい。

漸近理論

- 漸近理論を証明するために，単調収束定理 (MCT)，Fatou の補題，優収束定理 (DCT)，カーネルのウェイト付きの積分，Borel-Cantelli の補題などを用いる。
- 教科書 Appendix A のほかには，西山・人見（2023，第 1 章）や清水（2021, 第 4 章，第 7 章）等が参考になる。