

Nonparametric Estimation of Conditional PDF

Yasuyuki Matsumura

Graduate School of Economics, Kyoto University

October 13, 2024



- 1 Conditional Density Estimation: Relevant Variables
- 2 Conditional Density Bandwidth Selection
- 3 Conditional Density Estimation: Irrelevant Variables
- 4 The Multivariate Dependent Variables Case
- 5 Applications
- 6 References

- $Y \in \mathbb{R}^1$: 独立変数 (被説明変数). まずは連続の場合を考える.
- $X \in \mathbb{R}^{q+r}$: 共変量. q 個の連続 r.v. と r 個の離散 r.v. からなる.

- $g(y|x) = \frac{f(x,y)}{\mu(x)}$: Y の X で条件付けた条件付き密度.
 - $f(\cdot)$: (X, Y) の結合密度.
 - $\mu(\cdot)$: X の周辺密度.

条件付き密度の推定

条件付き密度 $g(y|x)$ をノンパラメトリック推定量

$$\hat{g}(y|x) = \frac{\hat{f}(x, y)}{\hat{\mu}(x)},$$

$$\hat{f}(x, y) = \frac{1}{n} \sum_{i=1}^n \underline{K_{\gamma}(x, X_i)} \times \underline{k_{h_0}(y, Y_i)},$$

$$\hat{\mu}(x) = \frac{1}{n} \sum_{i=1}^n \underline{K_{\gamma}(x, X_i)}$$

によって推定する.

条件付き密度の推定

ただし、各変数に対するカーネルは、

$$K_\gamma(x, X_i) = W_h(x^c, X_i^c) L(x^d, X_i^d, \lambda),$$

$$W_h(x^c, X_i^c) = \prod_{s=1}^q \frac{1}{h_s} w\left(\frac{x_s^c - X_{is}^c}{h_s}\right),$$

$$L(x^d, X_i^d, \lambda) = \prod_{s=1}^r \left(\frac{\lambda_s}{c_s - 1}\right)^{N_{is}(x)} (1 - \lambda_s)^{1 - N_{is}(x)},$$

$$N_{is}(x) = \mathbf{1}(X_{is} \neq x_s^d),$$

$$k_{h_0}(y, Y_i) = \frac{1}{h_0} k\left(\frac{y - Y_i}{h_0}\right)$$

によって定める。

分析者自身で設定しなければならない Tuning Parameters は,

- $h = (h_1, \dots, h_q) \in \mathbb{R}^q,$
- $\lambda = (\lambda_1, \dots, \lambda_r) \in \mathbb{R}^r,$
- $h_0 \in \mathbb{R}^1$

の3種類のバンド幅である.

最小二乗 CV

CV 基準 (CV 関数) として, weighted-ISE を考える. 表記を簡単にするため,

$$\int dx = \sum_{x^d} \int dx^c$$

と書くことにすると, weighted-ISE は,

$$\begin{aligned} ISE &= \int [\hat{g}(y|x) - g(y|x)]^2 \mu(x) M(x^c) dx dy \\ &= \int \hat{g}(y|x)^2 \mu(x) M(x^c) dx dy \\ &\quad - 2 \int \hat{g}(y|x) g(y|x) \mu(x) M(x^c) dx dy \\ &\quad + \int g(y|x)^2 \mu(x) M(x^c) dx dy \\ &=: I_{1n} - 2I_{2n} + I_{3n} \end{aligned}$$

と表わされる.

I_{3n} は推定量 \hat{g} を含まないので, Tuning Parameters に関する最小化に関係がない. そこで, 改めて CV 基準を

$$\begin{aligned} I_{1n} - 2I_{2n} \\ = \int \hat{g}(y|x)^2 \mu(x) M(x^c) dx dy - 2 \int \hat{g}(y|x) g(y|x) \mu(x) M(x^c) dx dy \end{aligned}$$

とする.

最小二乗 CV

I_{1n}, I_{2n} はなんらかの推定対象ではあるものの、このままの形ではどのように推定すればよいかわかりにくい。そこで、 I_{1n} を次のように変形する：

$$\begin{aligned} I_{1n} &= \int \hat{g}(y|x)^2 \mu(x) M(x^c) dx dy \\ &= \int \frac{\hat{f}(x, y)^2}{\hat{\mu}(x)^2} \mu(x) M(x^c) dx dy \\ &= \int \left[\int \hat{f}(x, y)^2 dy \right] \frac{1}{\hat{\mu}(x)^2} M(x^c) \mu(x) dx \\ &= \int \hat{G}(x) \frac{1}{\hat{\mu}(x)^2} M(x^c) \mu(x) dx \quad \left(\leftarrow \hat{G}(x) = \int \hat{f}(x, y)^2 dy \right) \\ &= \mathbb{E} \left[\frac{\hat{G}(X)}{\hat{\mu}(X)^2} M(X^c) \right]. \end{aligned}$$

同様にして, I_{2n} も次のように変形する:

$$\begin{aligned} I_{2n} &= \int \hat{g}(y|x) \underline{g(y|x)\mu(x)} M(x^c) dx dy \\ &= \int \hat{g}(y|x) \underline{f(x,y)} M(x^c) dx dy \quad (\leftarrow g(y|x)\mu(x) = f(x,y)) \\ &= \mathbb{E}_Z[\hat{g}(Y|X)M(X^c)] \\ &= \mathbb{E}_Z \left[\frac{\hat{f}(X,Y)}{\hat{\mu}(X)} M(X^c) \right], \end{aligned}$$

where $Z = (X, Y)$.

$$I_{1n} = \mathbb{E} \left[\frac{\hat{G}(X)}{\hat{\mu}(X)^2} M(X^c) \right], \quad I_{2n} = \mathbb{E}_Z \left[\frac{\hat{f}(X, Y)}{\hat{\mu}(X)} M(X^c) \right]$$

を標本版で推定する．ただし， $\hat{G}(X), \hat{\mu}(X), \hat{f}(X, Y)$ は，Leave-One-Out 推定量で置き換える：

$$\begin{aligned} \hat{I}_{1n} &= \frac{1}{n} \sum_{i=1}^n \frac{\hat{G}_{-i}(X_i)}{\hat{\mu}_{-i}(X_i)^2} M(X_i^c), \\ \hat{I}_{2n} &= \frac{1}{n} \sum_{i=1}^n \frac{\hat{f}_{-i}(X_i, Y_i)}{\hat{\mu}_{-i}(X_i)} M(X_i^c) \end{aligned}$$

こうして，CV 基準の I_{1n}, I_{2n} を推定量で置き換えたものを

$$CV_g(h_0, h, \lambda) = \hat{I}_{1n} - 2\hat{I}_{2n}$$

とおく．

CV 基準 $CV_g(h_0, h, \lambda) = \hat{I}_{1n} - 2\hat{I}_{2n}$ の主要項は,

$$CV_{g0}(h_0, h, \lambda) = \text{IMSE} \\ = \int \left\{ \mathbb{E}[\hat{f}(x, y) - \hat{\mu}(x)g(y|x)]^2 \frac{M(x^c)}{\mu(x)^2} \right\} dx dy$$

である¹.

¹証明は Hall et al. (2004) を参照せよ.

まず, $\mathbb{E} \left\{ \hat{f}(x, y) - \hat{\mu}(x)g(y|x) \right\}$ を評価する.

- $\mathbb{E} \{ \hat{f}(x, y) \}$
- $\mathbb{E} \{ \hat{\mu}(x) \}$
- $\mathbb{E} \{ \hat{\mu}(x)g(y|x) \}$
- $\mathbb{E} \left\{ \hat{f}(x, y) - \hat{\mu}(x)g(y|x) \right\}$

次に, 同様の計算により, $\text{Var}(\hat{f}(x, y) - \hat{\mu}(x)g(y|x))$ を評価する. そして, バイアス項の二乗と分散のオーダーが「つりあう」ように, バンド幅 $\hat{h}_0, \hat{h}_s, \hat{\lambda}_s$ を選択する².

²これらの計算過程をスライドで表示するには気の遠くなるようなコーディングが必要なので, 板書する.

まず,

Letting $\rho_s = \lambda_s / \{(1 - \lambda_s)(c_s - 1)\}$, and letting $f_{00}(x^c, x^d, y)$ and $f_{ss}(x^c, x^d, y)$ denote the second derivative of $f(x^c, x^d, y)$ with respect to y and x_s^c respectively, we may write

$$\begin{aligned} E \{ \hat{f}(x, y) \} &= \sum_{v^d} \left\{ \prod_{s=1}^r (1 - \lambda_s) \rho_s^{1(v_s^d \neq x_s^d)} \right\} \left\{ \int \prod_{s=1}^q k(z_s) \right. \underbrace{M(v)}_{\text{red line}} \left. \right\} \\ &\quad \times f(x^c - hz, x^d, y - h_0 v) dz_1 \dots dz_q dv \\ &= f(x, y) + \sum_{s=1}^r \lambda_s \left\{ \frac{1}{c_s - 1} \sum_{v^d} 1_s(v^d, x^d) f(x^c, v^d, y) - f(x, y) \right\} \\ &\quad + \frac{1}{2} \kappa_2 h_0^2 f_{00}(x, y) + \frac{1}{2} \kappa_2 \sum_{s=1}^q h_s^2 f_{ss}(x, y) + o(\eta_2), \end{aligned} \tag{5.6}$$

次に,
て,
幅 \hat{h}_0

し
ド

²これらの計算過程をスライドで表示するには気の遠くなるようなコーディングが必要なので、板書する。

まず

$$\begin{aligned}
 E \{ \hat{\mu}(x) \} &= \sum_{v^d} P(X^d = v^d) \left\{ \prod_{s=1}^r (1 - \lambda_s) \rho_s^{1(v_s^d \neq x_s^d)} \right\} \\
 &\quad \times \left\{ \int \prod_{s=1}^q k(z_s) \right\} \cancel{M(v)} \cancel{\left(x^c - hz, x^d, y - h_0 v \right)} dz_1 \dots dz_q \cancel{dv} \\
 &= \mu(x) + \sum_{s=1}^r \lambda_s \left\{ \frac{1}{c_s - 1} \sum_{v^d} 1_s(v^d, x^d) \mu(x^c, v^d) - \mu(x) \right\} \\
 &\quad + \frac{1}{2} \kappa_2 \sum_{s=1}^q h_s^2 \mu_{ss}(x, \cancel{x}) + o(\eta_2).
 \end{aligned}
 \tag{5.7}$$

次に、
幅

ド

²これらの計算過程をスライドで表示するには気の遠くなるようなコーディングが必要なので、板書する。

先ほどの計算により, 適切な $\mathcal{X}_g(a_0, a, b)$ を用いて

$$CV_{g0}(h_0, h, \lambda) = n^{\frac{-4}{q+5}} \mathcal{X}_g(a_0, a, b) \quad \leftarrow \frac{-4}{q+5}$$

と表現できる. ただし,

$$h_s = a_s \times n^{\frac{-1}{q+5}} (s = 0, 1, \dots, q),$$
$$\lambda_s = b_s \times n^{\frac{-2}{q+5}} (s = 1, \dots, r)$$

である.

最小二乗 CV

- $CV_g(h_0, h, \lambda)$ の minimizer を $\hat{h}_0, \hat{h}_s^0, \hat{\lambda}_s^0$ とする.
- $CV_{g0}(h_0, h, \lambda)$ の minimizer を $h_0^0, h_s^0, \lambda_s^0$ とする.
- $\mathcal{X}_g(a_0, a, b)$ の minimizer を a_0^0, a_s^0, b_s^0 とする.
- $CV_g = CV_{g0} + (\text{s.o.})$ より, $\hat{h}_s = h_s^0 + (\text{s.o.}), \hat{\lambda}_s = \lambda_s^0 + (\text{s.o.})$.
- $CV_{g0}(h_0, h, \lambda) = n^{\frac{-q}{q+4}} \mathcal{X}_g(a_0, a, b)$.

Theorem 5.1³

$$\begin{aligned} n^{\frac{1}{q+5}} \hat{h}_S &\xrightarrow{p} a_s^0 \text{ for } s = 0, 1, \dots, q, \\ n^{\frac{2}{q+5}} \hat{\lambda}_S &\xrightarrow{p} a_s^0 \text{ for } s = 1, \dots, r, \\ n^{\frac{4}{q+5}} \inf CV_g &\rightarrow \inf \mathcal{X}_g \text{ in probability.} \end{aligned}$$

- これにより, 推定量 $\hat{g}(y|x)$ の漸近分布を求めることができる.

³証明は Hall et al. 2004 を参照せよ.

- 最小二乗 CV は，サンプルサイズが大きいときに計算が大変である．
- 実装を考慮すると，最尤 CV の方が計算コストが軽い．

次に示す尤度関数 \mathcal{L} の maximizer $h_0, \dots, h_q, \lambda_1, \dots, \lambda_r$ をバンド幅として採用する：

$$\mathcal{L} = \sum_{i=1}^n \ln \hat{g}_{-i}(Y_i|X_i),$$

where

$$\hat{g}_{-i}(Y_i|X_i) = \frac{\hat{f}_{-i}(X_i, Y_i)}{\hat{\mu}_{-i}(X_i)},$$

and $\hat{f}_{-i}(X_i, Y_i)$ and $\hat{\mu}_{-i}(X_i)$ are the leave-one-out kernel estimators of $f(X_i, Y_i)$ and $\mu(X_i)$, respectively.

- 最尤 CV は，裾の厚い分布に対して弱い．
- 過剰平滑化や一致性を失うことにつながり得る．
- 詳しくは，Hall 1987a, 1987b を参照せよ．

- 説明変数に irrelevant variables が含まれていても，漸近的には smoothed out されるので，結果に大きな変更はない。

Theorem 5.2⁴

$$n^{\frac{1}{q_1+5}} \hat{h}_s \xrightarrow{p} a_s^0 \text{ for } s = 0, 1, \dots, q_1,$$

$$\mathbb{P}(\hat{h}_s > C) \rightarrow 1 \text{ for } q_1 + 1 \leq s \leq q \text{ and for all } C > 0,$$

$$n^{\frac{2}{q_1+5}} \hat{\lambda}_S \xrightarrow{p} a_s^0 \text{ for } s = 1, \dots, r_1,$$

$$\hat{\lambda}_S \xrightarrow{p} \frac{c_s - 1}{c_s} \text{ for } r_1 + 1 \leq s \leq r,$$

$$n^{\frac{4}{q_1+5}} \inf CV_\gamma(\hat{h}, \hat{\lambda}) \xrightarrow{p} \inf \mathcal{X}.$$

⁴証明は Hall et al. 2004 を参照せよ。

Theorem 5.3 (漸近正規性)⁵

$$\sqrt{n\hat{h}_1 \cdots \hat{h}_{q_1}} \left(\hat{g}(y|x) - g(y|x) - \underbrace{\sum_{s=0}^{q_1} B_{1s}(\bar{x}, y) \hat{h}_s^2}_{\text{red underline}} - \underbrace{\sum_{s=1}^{r_1} B_{2s}(\bar{x}, y) \hat{\lambda}_S}_{\text{red underline}} \right) \\ \xrightarrow{d} N(0, \sigma_g^2(\bar{x}, y)),$$

⁵証明は Hall et al. 2004 を参照せよ.

Y が離散変数の場合

ここまでは、 $Y \in \mathbb{R}^1$ が連続変数である場合を考えてきた。 Y が離散変数であるとき、 Y に対するカーネルを

$$l(y, Y_i, \lambda_0) = \lambda_0^{N_i(y)} (1 - \lambda_0)^{1 - N_i(y)},$$

where

$$N_i(y) = \mathbf{1}(Y_i \neq y)$$

に変更する。この変更によって、定理 5.2 が主張するような適切なバンド幅のオーダーも変更する必要がある。

Y が離散変数の場合

So far we have assumed that Y is a continuous random variable. Now we turn our attention to the case for which Y is discrete. If Y assumes c_0 different values, then we replace the continuous kernel $k_{h_0}(y, Y_i) = h_0^{-1}k((y - Y_i)/h_0)$ by the categorical kernel $l(y, Y_i, \lambda_0) = \lambda_0^{N_i(y)}(1 - \lambda_0)^{1-N_i(y)}$, where $N_i(y) = 1(Y_i \neq y)$. Theorem 5.2 can be modified by replacing $q_1 + 5$ with $q_1 + 4$, and by replacing $n^{1/(q_1+5)}\hat{h}_0 \rightarrow a_0^0$ with $n^{2/(q_1+4)}\hat{\lambda}_0$, where b_0^0 is defined in a way similar to b_s^0 for $s = 1, \dots, r$. Theorem 5.3 also requires minor modification (see Exercise 5.1).

Y が離散変数の場合

Exercise 5.1. When Y is a discrete variable, say,

$$Y \in \mathcal{S}_0^d = \{0, 1, \dots, c_0 - 1\},$$

show that (5.17) should be modified to

$$\begin{aligned} & (n\hat{h}_1 \dots \hat{h}_{q_1})^{1/2} (\tilde{g}(y|x) - g(y|x) - \sum_{s=1}^{q_1} B_{1s}(\bar{x}, y) \hat{h}_s^2 - \sum_{s=1}^{r_1} B_{2s}(\bar{x}, y) \hat{\lambda}_s) \\ & \rightarrow N(0, \sigma_g^2(\bar{x}, y)), \end{aligned}$$

where

$$\begin{aligned} B_{1s}(\bar{x}, y) &= \frac{1}{2} \kappa_2 \sum_{s=1}^{q_1} \hat{h}_s^2 \left\{ \frac{\bar{f}_{ss}(y|\bar{x})}{\bar{\mu}(\bar{x})} - \frac{\bar{\mu}_{ss}(\bar{x})}{\bar{\mu}(\bar{x})} \bar{g}(y|\bar{x}) \right\}, \\ B_{2s}(\bar{x}, y) &= \frac{\hat{\lambda}_0}{c_0 - 1} \sum_{z^d \in \mathcal{S}_0^d} \mathbf{1}_0(z^d, y) \left\{ \bar{g}(z|\bar{x}^c, \bar{v}^d) - \frac{\bar{\mu}(\bar{x}^c, \bar{v}^d)}{\bar{\mu}(\bar{x})} \bar{g}(y|\bar{x}) \right\} \\ & \quad + \sum_{s=1}^{r_1} \frac{\hat{\lambda}_s}{c_s - 1} \sum_{\bar{v}^d} \mathbf{1}_s(\bar{v}^d, x_s^d) \left\{ \bar{g}(y|\bar{x}^c, \bar{v}^d) - \frac{\bar{\mu}(\bar{x}^c, \bar{v}^d)}{\bar{\mu}(\bar{x})} \bar{g}(y|\bar{x}) \right\}, \\ \sigma_a^2(\bar{x}, y) &= \kappa^{q_1} \bar{g}(y|\bar{x}) / \bar{\mu}(\bar{x}), \end{aligned}$$

where $\hat{\lambda}_0$ is the cross-validation selected smoothing parameter that is associated with y .

ここまで
数であると

where

に変更する
幅のオーダー

が離散変

刃なバンド

- Li, Q. and J. S. Racine, (2007). *Nonparametric Econometrics: Theory and Practice*, Princeton University Press.