Nonparametric Estimation of Conditional PDF

Yasuyuki Matsumura

Graduate School of Economics, Kyoto University

October 16, 2024

Mixed Data

- $Y \in \mathbb{R}^1$：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．

条件付き密度

- $g(y|x) = \dfrac{f(x,y)}{\mu(x)}$: 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

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

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

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

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,

$$
\begin{aligned}
K_\gamma(x, X_i) &= W_h(x^c, {X_i}^c)\, L(x^d, {X_i}^d, \lambda), \\
W_h(x^c, {X_i}^c) &= \prod_{s=1}^{q} \frac{1}{h_s} w\left(\frac{{x_s}^c - {X_{is}}^c}{h_s}\right), \\
L(x^d, {X_i}^d, \lambda) &= \prod_{s=1}^{r} \left(\frac{\lambda_s}{c_s - 1}\right)^{N_{is}(x)} (1 - \lambda_s)^{1 - N_{is}(x)}, \\
N_{is}(x) &= \mathbf{1}(X_{is} \neq {x_s}^d), \\
k_{h_0}(y, Y_i) &= \frac{1}{h_o} k\left(\frac{y - Y_i}{h_0}\right)
\end{aligned}
$$

によって定める.

条件付き密度の推定

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

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

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，

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

と表わされる．

最小二乗 CV

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：

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

最小二乗 CV

同様にして，I_{2n} も次のように変形する：

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

where $Z = (X, Y)$.

最小二乗 CV

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

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：

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

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．

最小二乗 CV

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

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

である [1].

[1] 証明は Hall et al. (2004) を参照せよ.

最小二乗 CV

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，$\mathsf{Var}(\hat{f}(x,y)-\hat{\mu}(x)g(y|x))$ 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．

最小二乗 CV

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

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

と表現できる．ただし，

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

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．
- $\mathcal{X}_g(a_o, a, b)$ 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 5.1[3]

$$n^{\frac{1}{q+5}} \hat{h}_S \xrightarrow{p} {a_s}^0 \text{ for } s = 0, 1, \cdots, q,$$

$$n^{\frac{2}{q+5}} \hat{\lambda}_S \xrightarrow{p} {a_s}^0 \text{ for } s = 1, \cdots, r,$$

$$n^{\frac{4}{q+5}} \inf CV_g \to \inf \mathcal{X}_g \text{ in probability.}$$

- 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．

最尤 CV

- 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．

最尤 CV

次に示す尤度関数 $\mathcal{L}$ の maximizer $h_0, \cdots, h_q, \lambda_1, \cdots, \lambda_r$ 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

- 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．

Irrelevant Variables

- 説明変数に irrlevant 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 5.2[4]

$$n^{\frac{1}{q_1+5}}\hat{h}_s \xrightarrow{p} {a_s}^0 \text{ for } s = 0, 1, \cdots, q_1,$$
$$\mathbb{P}(\hat{h}_s > C) \to 1 \text{ for } q_1 + 1 \leq s \leq q \text{ and for all } C > 0,$$
$$n^{\frac{2}{q_1+5}}\hat{\lambda}_S \xrightarrow{p} {a_s}^0 \text{ for } s = 1, \cdots, r_1,$$
$$\hat{\lambda}_S \xrightarrow{p} \frac{c_s - 1}{c_s} \text{ for } r_1 + 1 \leq s \leq r,$$
$$n^{\frac{4}{q_1+5}} \inf CV_\gamma(\hat{h}, \hat{\lambda}) \xrightarrow{p} \inf \mathcal{X}.$$

[4] 証明は Hall et al. 2004 を参照せよ．

Irrelevant Variables

Theorm 5.3 （漸近正規性）[5]

$$\sqrt{n\hat{h}_1 \cdots \hat{h}_{q1}} \left(\hat{g}(y|x) - g(y|x) - \sum_{s=0}^{q_1} B_{1s}(\bar{x}, y)\hat{h}_s^2 - \sum_{s=1}^{r_1} B_{2s}(\bar{x}, y)\hat{\lambda}_S \right)$$

$$\xrightarrow{d} N(0, \sigma_g^2(\bar{x}, y)),$$

[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．

$Y \in \mathbb{R}^{q+r}$（多次元）の場合

次回.

応用例

省略.

References

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