Nonparametric Estimation of Conditional PDF

Yasuyuki Matsumura

Graduate School of Economics, Kyoto University

October 28, 2024

Introduction

- We consider the estimation of the conditional PDF $g(Y|X)$ when both Y and X are general multivariate vectors.

Settings

- $Z = (X, Y) = (Z^c, Z^d)$
- Z^d consists of r discrete variables.
  - $X^d \in \mathbb{R}^{r_x}$, $Y^d \in \mathbb{R}^{r_y}$.
- Z^c consists of q continuous variables.
  - $X^c \in \mathbb{R}^{q_x}$, $Y^c \in \mathbb{R}^{q_y}$.
- $r = r_x + r_y$, $q = q_x + q_y$.

Kernel Methods

- Let Z_{is}^d denote the s-th component of Z_i^d.
- Define a univariate kernel function

$$l(Z_{is}^d, Z_{js}^d, \lambda_s) = \begin{cases} 1 - \lambda_s & \text{if } Z_{is}^d = Z_{js}^d, \\ \dfrac{\lambda_s}{c_s - 1} & \text{if } Z_{is}^d \neq Z_{js}^d. \end{cases}$$

- The product kernel is given by

$$L_{\lambda, Z_i^d, Z_j^d} = \prod_{s=1}^{r} l(Z_{is}^d, Z_{js}^d, \lambda_s).$$

Kernel Methods

- Let Z_{is}^c denote the s-th component of Z_i^c.
- Define a univariate kernel function $w(\cdot)$ in the same way as Section 5.1.
- Then, the product kernel is given by

$$W_{h,Z_i^c,Z_j^c} = \prod_{s=1}^{q} \frac{1}{h_s} w\left(\frac{Z_{is}^c - Z_{js}^c}{h_s}\right).$$

- Define $K_{\gamma,Z_i,z} = L_{\lambda,Z_i^d,z} \times W_{h,Z_i^c,z^c}$.
- Apply the same notation $L(\cdot)$ and $W(\cdot)$ to Y and X.

Kernel Methods

- We estimate

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

by the following nonparametric estimator:

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

where

$$\hat{f}(x,y) = \hat{f}(z) = \frac{1}{n}\sum_{i=1}^{n} K_{\gamma,Z_i,z},$$

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

Bandwidth Selection: Least Squares CV

- Consider the case where the independent variables are all relevent.
  - Irrelevant independent variables will be asymptotically smoothed out.
- As in Section 5.2, we choose the tuning parameters h and λ that minimize a sample analogue of $I_{1n} - 2I_{2n}$:

$$CV(h, \lambda) = \frac{1}{n} \sum_{i=1}^{n} \frac{\hat{G}_{-i}(X_i)}{\hat{\mu}_{-i}(X_i)^2} - \frac{2}{n} \sum_{i=1}^{n} \frac{\hat{f}_{-i}(X_i)}{\hat{\mu}_{-i}(X_i)}.$$

Bandwidth Selection: Least Squares CV

- Take $\hat{h}$ and $\hat{\lambda}$ as the CV choices.
- Let ${h_s}^0$ and ${\lambda_s}^0$ denote the minimizers of the leading term of $CV(h, \lambda)$.

Racine et al. [2004]

Where ${h_s}^0 = c_{1s} n^{\frac{-1}{q+4}}$ and ${\lambda_s}^0 = c_{2s} n^{\frac{-2}{q+4}}$, it holds that

$$\frac{\hat{h}_s}{{h_s}^0} \xrightarrow{p} 1, \quad \frac{\hat{\lambda}_s}{{\lambda_s}^0} \xrightarrow{p} 1.$$

Bandwidth Selection: Least Squares CV

Theorem 5.4

Under several assumptions given in Racine et al. (2004),

$$\frac{\hat{h}_s - {h_s}^0}{{h_s}^0} = O_p\left(n^{\frac{-\alpha}{q+4}}\right), \quad \alpha = \min\left\{2, \frac{q}{2}\right\},$$

$$\frac{\hat{\lambda}_s - {\lambda_s}^0}{{\lambda_s}^0} = O_p\left(n^{-\beta}\right), \quad \beta = \min\left\{\frac{1}{2}, \frac{4}{q+4}\right\}.$$

Bandwidth Selection: Least Squares CV

Theorem 5.5 (Asymptotic Normality)

$$\sqrt{nh_1 \cdots h_q}\left(\hat{g}(y|x) - g(y|x) - \sum_{s=1}^{q} \hat{h}_s^2 B_{1s}(z) - \sum_{s=1}^{r} \hat{\lambda}_s B_{2s}(z)\right)$$
$$\xrightarrow{d} N(0, \Omega(z)),$$

where $B_{1s}(z), B_{2s}(z)$, $\Omega(z)$ are defined by

$$B_{1s}(z) = \begin{cases} \frac{1}{2}\kappa_2 \frac{f_{ss}(x,y)}{\mu(x)} \, (s = 1, \cdots, q_y), \\ \frac{1}{2}\kappa_2 \frac{f_{ss}(x,y) - \mu_{ss}(x) g(y|x)}{\mu(x)} \, (s = q_y + 1, \cdots, q), \end{cases}$$

$$B_{2s}(z) = \begin{cases} \frac{1}{c_s - 1} \sum_{v^d \in \mathcal{D}} \mathbf{1}_s(v^d, z^d) f(z^c, v^d) \, (s = 1, \cdots, r_y), \\ \frac{1}{c_s - 1} \sum_{u^d \in \mathcal{D}} \mathbf{1}_s(u^d, x^d) \frac{f(z^c, y^d, u^d) - g(y|x)\mu(x^c, u^d)}{\mu(x)} \, (s = r_y + 1, \cdots \end{cases}$$

$$\Omega(z) = \frac{\kappa^q g(y|x)}{\mu(x)}.$$

Stochastic Equicontinuity

- Take an empirical process $v_n(t)$.
- We say $v_n(t)$ is stochastic equicontinuous[1] at $t = t_0$ if

$$\text{For } {}^{\forall}\epsilon > 0, {}^{\forall}\eta > 0, {}^{\exists}\delta > 0;$$
$$\limsup_{n\to\infty} \mathbb{P}\left(\sup_{t\in T,\, \rho(t,t_0)<\delta} |v_n(t) - v_n(t_0)| > \eta \right) < \epsilon.$$

[1]確率的同程度連続.

Proof of Theorem 5.5

- The smoothing parameters $\hat{h}_s, \hat{\lambda}_s$ which we take as the minimizers of $CV(h, \lambda)$ are stochastic.
- On the other hand, the parameters h_s, λ_s that are defined as $h_s = c_{1s} \times n^{\frac{-1}{q+4}}, \lambda_s = c_{2s} \times n^{\frac{-2}{q+4}}$ (which satisfy $\frac{\hat{h}_s}{h_s} \xrightarrow{p} 1, \frac{\hat{\lambda}_s}{\lambda_s} \xrightarrow{p} 1$) are nonstochastic.
- By using stochastic equicontinuity arguments, we know that the asymptotic distribution of $\hat{g}(y|x)$ is the same WHETHER we use stochastic $\hat{h}_s, \hat{\lambda}_s$ OR nonstochastic h_s, λ_s.
- Therefore, we consider only the nonstochastic smoothing parameter case.

Proof of Theorem 5.5

- To derive the asymptotic normality of $\hat{g}(y|x)$, we write

$$\hat{g}(y|x) - g(y|x) = \frac{(\hat{g}(y|x) - g(y|x))\hat{\mu}(x)}{\hat{\mu}(x)} \equiv \frac{\hat{m}(y,x)}{\hat{\mu}(x)},$$

where $\hat{m}(y,x) = (\hat{g}(y|x) - g(y|x))\hat{\mu}(x) = \hat{f}(y,x) - g(y|x)\hat{\mu}(x)$.

Proof of Theorem 5.5

- We consider the computation of
  - $\mathbb{E}(\hat{m}(y,x))$,
  - $\text{var}(\hat{m}(y,x))$,

  so that we will obtain the asymptotic normality of $\hat{m}(y,x)$:

$$\sqrt{nh_1 \cdots h_q}\{\hat{m}(y,x) - (\text{bias terms})\} \xrightarrow{d} N(0, \kappa^q f(y,x)).$$

- Note that we have already established

$$\hat{\mu}(x) - \mu(x) = Op\left(\sum_{s=q_y+1}^{q} h_s^2 + \frac{1}{\sqrt{nh_{q_y+1} \cdots h_q}}\right).$$

- Then, we derive the asymptotic normality of $\hat{g}(y|x)$:

$$\sqrt{nh_1 \cdots h_q}(\hat{g}(y|x) - g(y|x) - (\text{bias temrs})) \xrightarrow{d} N(0, \Omega(z)).$$

References

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