IEMS 402 Statistical Learning - 2025 Winter
HW4

Tuorui Peng[1]

Exercise 1 Estimating the Derivatives via Kernel Smoothing

- Bias term:

$$
\begin{aligned}
\left|d_n(x)-p'(x)\right| =& \left|\int_{-1}^{1}\frac{1}{h^2}K\left(\frac{X_i-x}{h}\right)p(x)\,\mathrm{d}x-p'(x)\right|\\
=&\frac{1}{h}\left|\int K(\|v\|)\big(p(x+hv)-vp'(x)\big)\,\mathrm{d}v\right|\\
=&\frac{1}{h}\left|\int K(\|v\|)\big(p(x+hv)-vp'_{x,\beta}(x+hv)+vp'_{x,\beta}(x+hv)-vp'(x)\big)\,\mathrm{d}v\right|\\
\leq&\frac{1}{h}\int K(\|v\|)\left|p(x+hv)-vp'_{x,\beta}(x+hv)\right|\,\mathrm{d}v+\frac{1}{h}\int K(\|v\|)\left|vp'_{x,\beta}(x+hv)-vp'(x)\right|\,\mathrm{d}v\\
\lesssim& Lh^{\beta-1}\int K(\|v\|)\left|v\right|\,\mathrm{d}v
\end{aligned}
$$

  in which the last inequality is because $p(\,\cdot\,)-p_{x,\beta}(\,\cdot\,)$ is still a β-Hölder function.

$$
\begin{aligned}
var(d_n(x)) \leq& \frac{1}{nh^4}\int K^2(\frac{X_i-x}{h})p(x)\,\mathrm{d}x\\
=&\frac{1}{nh^3}\int K^2(v)p(x+hv)\,\mathrm{d}v\\
\lesssim&\frac{1}{nh^3}\sup_{x\in[-1,1]}p(x)\int K^2(v)\,\mathrm{d}v
\end{aligned}
$$

  where the last inequality is because only the first order term of $p(\,\cdot\,)$ gives non-zero kernel integration.

Put together, we have

$$
\mathrm{MSE}\lesssim h^{2(\beta-1)}+\frac{1}{nh^3}
$$

Optimal h_n is at $h_n\asymp n^{-\frac{1}{2\beta+1}}$, and the optimal MSE is $n^{-\frac{2(\beta-1)}{2\beta+1}}$.

Exercise 2 An average treatment effect estimator

2.(a)

In completely randomized experiment, we can see that

$$
\begin{aligned}
\mathbb{E}\left[Y_i(a)1\{A_i=a\}\right] =& \frac{1}{2}\mathbb{E}\left[Y_i(a)1\{A_i=a\}|A_i=0\right]+\frac{1}{2}\mathbb{E}\left[Y_i(a)1\{A_i=a\}|A_i=1\right]\\
=&\begin{cases}\frac{1}{2}\mathbb{E}\left[Y_i(0)|A_i=0\right], & \text{if } a=0\\ \frac{1}{2}\mathbb{E}\left[Y_i(1)|A_i=1\right], & \text{if } a=1\end{cases}\\
=&\frac{1}{2}\mathbb{E}\left[Y_i(a)\right]=\frac{1}{2}\mathbb{E}\left[Y(a)\right]
\end{aligned}
$$

[1] TuoruiPeng2028@u.northwestern.edu

and we have ATE being

$$\tau = \mathbb{E}\left[Y(1)\right] - \mathbb{E}\left[Y(0)\right] = 2\mathbb{E}\left[Y(1)1\{A=1\}\right] - 2\mathbb{E}\left[Y(0)1\{A=0\}\right].$$

2.(b)

Note that we can write $\hat{\tau}_n$ as

$$\begin{aligned}
\hat{\tau}_n =& \frac{1}{n}\big[2Y_i(1)1\{A_i=1\} - 2Y_i(0)1\{A_i=0\}\big] \\
\mathbb{E}\left[2Y_i(1)1\{A_i=1\} - 2Y_i(0)1\{A_i=0\}\right] =& \tau \\
var(2Y_i(1)1\{A_i=1\} - 2Y_i(0)1\{A_i=0\}) =& var(\mathbb{E}\left[2Y_i(1)1\{A_i=1\} - 2Y_i(0)1\{A_i=0\}|A\right]) \\
&+ \mathbb{E}\left[var(2Y_i(1)1\{A_i=1\} - 2Y_i(0)1\{A_i=0\}|A)\right] \\
=& (\mathbb{E}\left[Y(1)\right] + \mathbb{E}\left[Y(0)\right])^2 + 2\big(var(Y(1)) + var(Y(0))\big)
\end{aligned}$$

By CLT

$$\sqrt{n}(\hat{\tau}_n - \tau) \xrightarrow{\text{d}} N(0, (\mu_1+\mu_0)^2 + 2(\sigma_1^2+\sigma_0^2)), \quad \mu_a = \mathbb{E}\left[Y(a)\right], \sigma_a^2 = var(Y(a))$$

2.(c)

Note that for completely randomized experiment, we have $|S_1| = n - |S_0| \sim \text{Binomial}(n, \dfrac{1}{2})$. Thus we get (take $a=1$ example)

$$\begin{cases} \sqrt{n}(1 - 2\left|S_1\right|/n) \xrightarrow{\text{d}} N(0,1) \\ 2\left|S_1\right|/n \xrightarrow{\text{p}} 1 \end{cases} \overset{\text{Slutsky}}{\Rightarrow} \sqrt{n}(\frac{n}{2\left|S_1\right|} - 1) \xrightarrow{\text{d}} N(0,1)$$

and similarly for $a=0$ we have $\sqrt{n}(\dfrac{n}{2\left|S_0\right|} - 1) \xrightarrow{\text{d}} N(0,1)$.

2.(d)

We have

$$\begin{aligned}
\sqrt{n}(\hat{\tau}_n^{\text{norm}} - \tau) =& \frac{\sqrt{n}}{\sqrt{|S_1|}}\sqrt{|S_1|}\Big(\frac{1}{|S_1|}\sum_i (Y_i(1) - \mathbb{E}\left[Y(1)\right])1\{A_i=1\}\Big) \\
&+ \frac{\sqrt{n}}{\sqrt{|S_0|}}\sqrt{|S_0|}\Big(\frac{1}{|S_0|}\sum_i (Y_i(0) - \mathbb{E}\left[Y(0)\right])1\{A_i=0\}\Big)
\end{aligned}$$

For now we treat A as given, and we have

$$\sqrt{n}(\hat{\tau}_n^{\text{norm}} - \tau)\big|A \xrightarrow{\text{d}} \frac{\sqrt{n}}{\sqrt{|S_1|}}N(0,\sigma_1^2) + \frac{\sqrt{n}}{\sqrt{|S_0|}}N(0,\sigma_0^2)$$

On the other hand we notice that we already have

$$\frac{\sqrt{n}}{\sqrt{|S_1|}} \xrightarrow{\text{p}} \sqrt{2}, \quad \frac{\sqrt{n}}{\sqrt{|S_0|}} \xrightarrow{\text{p}} \sqrt{2}, \quad cov\left(\frac{n}{|S_1|}, \frac{n}{|S_0|}\right) = -1$$

Thus by Slutsky's theorem, we have

$$\sqrt{n}(\hat{\tau}_n^{\text{norm}} - \tau) \xrightarrow{\text{d}} N(0, \sigma_{\text{norm}}^2)$$

where

$$\sigma_{\text{norm}}^2 = 2\sigma_1^2 + 2\sigma_0^2$$

2.(e)

It seems that based on our results up to now, the conlusion would be that:

$$\sqrt{n}(\hat{\tau}_n - \tau) \xrightarrow{\text{d}} N(0, \sigma^2), \quad \sigma^2 = (\tau_1 + \tau_0)^2 + 2(\sigma_1^2 + \sigma_0^2)$$
$$\sqrt{n}(\hat{\tau}_n^{\text{norm}} - \tau) \xrightarrow{\text{d}} N(0, \sigma_{\text{norm}}^2), \quad \sigma_{\text{norm}}^2 = 2(\sigma_1^2 + \sigma_0^2)$$

so we have $\sigma^2 > \sigma_{\text{norm}}^2$ as long as $\tau_1 + \tau_0 > 0$.

Exercise 3 A weighted average treatment effect estimator

3.(a)

With the covariate X involved, we have

$$\begin{aligned}\tau =&\mathbb{E}\left[Y(1)\right] - \mathbb{E}\left[Y(0)\right]\\ =&\mathbb{E}\left[\mathbb{E}\left[Y(1)\big(1\{A=1\}+1\{A=0\}\big)\big|X=x\right]\right] - \mathbb{E}\left[\mathbb{E}\left[Y(0)\big(1\{A=1\}+1\{A=0\}\big)\big|X=x\right]\right]\end{aligned}$$

now note that since $(Y(1), Y(0)) \perp A|X$, we have

$$\begin{aligned}\mathbb{E}\left[Y(A)1\{A=1\}|X=x\right] =&Y(1)e(x)\\ \mathbb{E}\left[Y(A)1\{A=0\}|X=x\right] =&Y(0)(1-e(x))\end{aligned}$$

substitute this back to the above equation, we have

$$\begin{aligned}\tau =&\mathbb{E}\left[\frac{Y(A)1\{A=1\}}{e(x)}\big(1\{A=1\}+1\{A=0\}\big) - \frac{Y(A)1\{A=0\}}{1-e(x)}\big(1\{A=1\}+1\{A=0\}\big)\Big|X=x\right]\\ =&\mathbb{E}\left[\frac{Y(A)1\{A=1\}}{e(x)} - \frac{Y(A)1\{A=0\}}{1-e(x)}\Big|X=x\right]\\ =&\mathbb{E}\left[\frac{Y(A)1\{A=1\}}{e(X)}\right] - \mathbb{E}\left[\frac{Y(A)1\{A=0\}}{1-e(X)}\right]\end{aligned}$$

3.(b)

3.(c)

Exercise 4 Logistic regression