HW8. Solution

Friday, November 17, 2017 11:40 AM

(1)] $X_1, \dots, X_n$ are i.i.d., absolutely continuous
$X_1 \sim F, f$
F: CDF, f: PDF

] $\hat{F}_n$ is the emirical distribution based on the sample $\{X_i\}_{i=1}^n$

Consider a moving average estimator

$$\hat{f}_n(x) = \frac{\hat{F}_n(x+a_n) - \hat{F}_n(x-a_n)}{2a_n} := \frac{\hat{p}_n}{2a_n}$$

(a) Show that $\mathbb{E}(\hat{f}_n(x)) \to f(x)$, $a_n \to 0$

$2na_n \hat{f}_n(x) = \#$ of X_i in $(x-a_n, x+a_n] \sim Bin(n, p_n)$,
$p_n = F_n(x+a_n) - F_n(x-a_n)$

$$\Rightarrow \mathbb{E}(\hat{f}_n(x)) = \frac{np_n}{2na_n} = \frac{F(x+a_n) - F(x-a_n)}{2a_n} \xrightarrow[a_n \to 0]{} f(x)$$

by def. of derivative $f(x) = dF(x)/dx$

(b) Show that $Var(\hat{f}_n(x)) \to 0$, $a_n \to 0$ and $na_n \to \infty$

$$Var(\hat{f}_n(x)) = \frac{np_n(1-p_n)}{(2na_n)^2} = \underbrace{\frac{np_n}{2na_n}}_{\to f} \cdot \underbrace{\frac{1-p_n}{2na_n}}_{\to 0} \to 0$$

(c) Argue that $\hat{f}_n$ is a consistent estimator

(c) Argue that $\hat{f}_n$ is a consistent estimator for f under the conditions of (b)

$$\left.\begin{array}{l}\mathbb{E}\hat{f}_n \to f\\ \mathrm{Var}(\hat{f}_n)\to 0\end{array}\right\} \Rightarrow \hat{f}_n \xrightarrow[P]{} f \quad \text{, point-wise}$$

Indeed

$$P(|\hat{f}_n - \mathbb{E}\hat{f}_n| > \varepsilon) \underset{\substack{\uparrow\\ \text{Chebyshev}}}{\leq} \frac{\mathrm{Var}(\hat{f}_n)}{\varepsilon^2} \to 0$$

$$\Rightarrow \hat{f}_n - \mathbb{E}\hat{f}_n \xrightarrow[P]{} 0$$

$$\text{b/c } \mathbb{E}\hat{f}_n \to f \Rightarrow \hat{f}_n \xrightarrow[P]{} f$$

Indeed $P(|\hat{f}_n - f| > \varepsilon) \leq \underbrace{P\left(|\hat{f}_n - \mathbb{E}\hat{f}_n| > \frac{\varepsilon}{2}\right)}_{\to 0} + \underbrace{\begin{cases} 1, & |f-\mathbb{E}\hat{f}_n| > \frac{\varepsilon}{2}\\ 0, & \text{otherwise}\end{cases}}_{\to 0}$

$$\varepsilon < |\hat{f}_n - f| \leq |\hat{f}_n - \mathbb{E}\hat{f}_n| + |f - \mathbb{E}\hat{f}_n|$$

b/c $\mathbb{E}\hat{f}_n \to f$ pointwise $\Rightarrow \forall x \;\underset{\varepsilon>0}{\forall}\; \exists N: |f - \mathbb{E}\hat{f}_n| \leq \frac{\varepsilon}{2}, \forall n > N$

(d) Show that

$$\sqrt{2na_n}\cdot\left[\hat{f}_n(x) - f(x)\right] \rightsquigarrow \mathcal{N}(0, f(x))$$

$$2na_n\hat{f}_n = \sum_{i=1}^n Y_{ni}, \quad Y_{ni} := \begin{cases} 1, & X_i \in (x-a_n, x+a_n]\\ 0, & \text{otherwise}\end{cases}$$

$$Y_{ni} \sim \text{Binary}(p_n)$$

Check Lyapunov condition

$$\sum_{i=1}^n \mathbb{E}|Y_{ni} - \mathbb{E}Y_{ni}|^3 \leq \sum_{i=1}^n (1+p_n)\,\mathbb{E}|Y_{ni} - p_n|^2 \leq$$

$$\mathbb{E}Y_{ni} = p_n, \quad |Y_{ni} - \mathbb{E}Y_{ni}| \leq \underbrace{|Y_{ni}|}_{=0 \text{ or } 1} + |\mathbb{E}Y_{ni}| \leq 1 + p_n$$

$$\mathrm{Var}(Y_{ni}) = p_n(1-p_n) := \sigma_{ni}^2$$

$$\leq (1+p_n)\cdot n\cdot\sigma_{ni}^2, \qquad n\cdot\sigma_{ni}^2 = \sigma_n^2$$

$$\frac{1}{\ldots}\sum^n \mathbb{E}|Y_{ni} - p_n|^3 \leq (1+p_n)\,n\cdot\sigma_{ni}^2$$

$$= (1+p_n) \cdot n \cdot \sigma_{ni}^2 , \quad n \cdot \sigma_{ni}^2 = \sigma_n^2$$

$$\frac{1}{\sigma_n^3} \sum_{i=1}^{n} \mathbb{E}|Y_{ni} - p_n|^3 \le \frac{(1+p_n)\, n \cdot \sigma_{ni}^2}{n \cdot \sqrt{n}\, \sigma_{ni}^3} = \frac{(1+p_n)}{\sqrt{n}\sqrt{p_n(1-p_n)}} \xrightarrow[n\to\infty]{} 0$$

By Lyapunov CLT

$$\frac{2na_n(\hat{f}_n - \mathbb{E}\hat{f}_n)}{\sqrt{np_n(1-p_n)}} \rightsquigarrow \mathcal{N}(0,1)$$

By Slutsky

$$\sqrt{2na_n}\,(\hat{f}_n - \mathbb{E}\hat{f}_n) = (\quad) \cdot \sqrt{\frac{np_n(1-p_n)}{2na_n}} \rightsquigarrow$$

$$\rightsquigarrow \mathcal{N}(0,1)\sqrt{f}$$

$$\frac{np_n}{2a_n \cdot n} \to f \quad (\text{see } (a)), \quad \frac{p_n}{2na_n} \to 0$$

from (a) $\quad \mathbb{E}\hat{f}_n - f = \frac{p_n}{2a_n} - f = f'(x) \cdot a_n + a_n \cdot o(1)$

$$p_n = f(x) \cdot 2a_n + \frac{1}{2} f'(x) \cdot (2a_n)^2 + o(a_n^2)$$

when f is differentiable

$$\sqrt{2na_n} \cdot (\mathbb{E}\hat{f}_n - f) = \sqrt{2na_n^3}\,(f'(x) + o(1))$$

$$= \sqrt{2na_n^3} \cdot O(1)$$

When $n \to \infty$, $na_n \to \infty$, $a_n \to 0$, $na_n^3 \to 0$ $\Rightarrow \sqrt{2na_n}(\mathbb{E}\hat{f}_n - f) \to 0$

For example $a_n = \frac{a}{n^{1-\varepsilon}}$, $0 < \varepsilon < \frac{2}{3}$

$$\Rightarrow na_n^3 = a^3 \cdot n^{3\varepsilon - 2} \to 0 \Rightarrow \varepsilon < \frac{2}{3}$$

$$na_n = an^{\varepsilon} \to \infty$$

$$a_n \to 0$$

$\Rightarrow \sqrt{2na_n}(\hat{f}_n - f) \rightsquigarrow \mathcal{N}(0, f)$