Lec 5,6. U-statistics

· 定义: $U = \frac{1}{\binom{n}{r}} \sum_{i_1<\cdots<i_r} h(X_{i_1},\cdots,X_{i_r})$, h: kernel, permutational symmetric, $Eh=\theta$

r: rank/order

① unbiased : $EU=\theta$.

② $Var(U) = \frac{1}{\binom{n}{r}^2} \sum_I \sum_J Cov(h(X_I), h(X_J))$

$= \frac{1}{\binom{n}{r}^2} \sum_{c=|I\cap J|=0}^{r} Cov(h(X_I), h(X_J))$

$= \frac{1}{\binom{n}{r}^2} \sum_{c=0}^{r} \binom{n}{r}\binom{r}{c}\binom{n-r}{r-c} \underline{Cov(h(X_1^c, X_{c+1}^r), h(X_1^c, X_{r+1}^{2r-c}))}$

$= Var(E[h(X_1^c, X_{c+1}^r)|X_1^c]) =: \zeta_c$.

$= \frac{1}{\binom{n}{r}} \sum_{c=1}^{r} \binom{r}{c}\binom{n-r}{r-c} \zeta_c$ → (accurate)

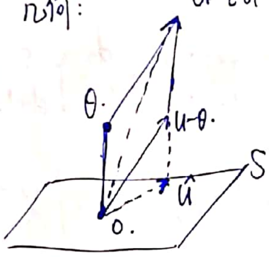
$\approx \sum_{c=1}^{r} ⊛ n^{-c} \zeta_c$ → (inaccurate)

$nVar(U) \longrightarrow r^2\zeta_1$, $\zeta_1 \neq 0$

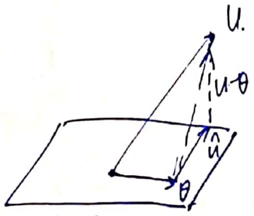
③ $\sqrt{n}(U-\theta) \longrightarrow N(0, r^2\zeta_1)$, $\zeta_1\neq 0$, $Eh^2<\infty$.

Sketch pf: i) 构造 $\hat{U}$ 和 $U-\theta$ 很像 (有种 Taylor 一阶近似感觉)

几何: $U: EU=\theta$: 在 θ 四处游荡



$S: \{f = f_1(X_1)+\cdots+f_n(X_n)\}$ $\boxed{\text{Hajék proj}}$



注意 $X_i \perp X_j$.

$\hat{U} = \frac{r}{n}\sum_{i=1}^{n} h_1(X_i)$, $h_1(x) = Eh(x, X_2,\cdots,X_r) - \theta$. $= E(U|X_1)$

ii) 对 $\hat{U}$ 用 CLT, $\sqrt{n}\,\hat{U} \rightarrow N(0, r^2\zeta_1)$

iii) 证明 $\sqrt{n}(\hat{U})$ 和 $\sqrt{n}(U-\theta)$ 同分布. "几乎" #

More generally,

• Q: ①. Project S on larger space? Hoeffding. decomposition.

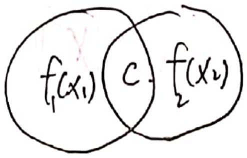
② Two sample X_i, Y_j ?

③ $\zeta_1 = 0$?

④ V-stat.? $A = \{a_1, \cdots, a_r\}$

① Hoeffding: let $H_A = \{f = f(X_{a_1}, \cdots, X_{a_r})\}$ 且 $H_A \perp H_B$, $\forall B \subsetneq A$, then.

$$P_A T = \sum_{B \subset A} (-1)^{|A|-|B|} E(T \mid X_i, i \in B)$$



② 2-sample U-stat. $X_{1 \sim m} \overset{iid}{\sim} F$, $Y_{1 \sim n} \overset{iid}{\sim} G$

$$U_{mn} = \frac{1}{\binom{m}{r}\binom{n}{s}} \sum_{\substack{\alpha_1 < \cdots < \alpha_r \\ \beta_1 < \cdots < \beta_s}} h(X_{\alpha_1}, \cdots, X_{\alpha_r}, Y_{\beta_1}, \cdots, Y_{\beta_s}), \quad Eh = \theta.$$

prop. $EU_{mn} = \theta$.

$$\operatorname{Var} U_{mn} = \frac{1}{\binom{m}{r}^2 \binom{n}{s}^2} \sum_{c,d=0}^{r,s} \binom{m}{r}\binom{r}{c}\binom{m-r}{r-c}\binom{n}{s}\binom{s}{d}\binom{n-s}{s-d} \zeta_{c,d}.$$

$$\zeta_{c,d} = \operatorname{Cov}\big(h(X_1^c, X_{c+1}^r, Y_1^d, Y_{d+1}^s), h(X_1^c, \tilde{X}_{c+1}^r, Y_1^d, \tilde{Y}_{d+1}^s)\big)$$

$$= \operatorname{Var}\big(E[h(X_1^c, X_{c+1}^r) \mid X_1^c]\big)$$

若 $\frac{m}{m+n} \to \lambda$, 则 $(m+n) U_{mn} \to N\left(0, \frac{r^2 \zeta_{1,0}}{\lambda} + \frac{s^2 \zeta_{0,1}}{1-\lambda}\right)$.

③. Degeneracy of order k: $\zeta_1 = \cdots = \zeta_k = 0$, $\zeta_{k+1} \neq 0\ (>0)$. !

Thm (order=1) (1) $r=2$, $n(U_n - \theta) \to \sum \lambda_j (Z_j^2 - 1)$, $Z_j \sim N(0,1)$ iid., λ_j 为 eigenvalues of $h(x_1, x_2) - \theta$.

(2) $r \geq 2$, $n(U_n - \theta) \to \binom{r}{2} \sum \lambda_j (Z_j^2 - 1)$, λ_j 为 $h_2(X_1, X_2) - \theta$ eigenvalue.

④ V-statistics: $V = \frac{1}{n^r} \sum_{i_1, \cdots, i_r} h(X_{i_1}, \cdots, X_{i_r})$. then U, V 渐近相同分布.

Lec 7. density: histo- and naive estimator, kde

- IMSE/Risk: $R(\hat f_n, f) \triangleq \int E(\hat f_n - f)^2 dx$

Integrated

$$= \int Var(\hat f_n)\,dx + \int (E\hat f_n - f)^2\,dx$$

- CV = Leave-one-out (Recall: Jackknife).

- Histogram: $[kh,(k+1)h)$ 中样本频率/h. 样本 $X_1,\dots,X_n$

$$\hat f_n(x) = \frac{\#\{X_i\in[kh,(k+1)h)\}/n}{h}, \quad \forall x\in[kh,(k+1)h).$$

Thm (informal). $R(\hat f_n,f) = \underbrace{\frac{h^2}{12}\|f'\|^2 + \frac{1}{nh}}_{\text{AIMSE (or typo: AMISE)}} + o(h^2) + O(\frac1n) \Rightarrow h_{opt} = (*)\, n^{-\frac13}$

对应 $R = (*)\, n^{-\frac23}$

Sketch pf: 作 2次 Taylor

$$f(u) = f(x) + f'(x)(u-x) + \frac12 f''(\xi)(u-x)^2, \quad u,x\in[kh,(k+1)h).$$

- Naive density estimator:

$$f(x) = \lim_{h\to0}\frac{1}{2h}(F(x+h)-F(x-h)) = \lim_{h\to0}\frac{1}{2h}P(x-h<X\le x+h)$$

$$\hat f_h(x) = \frac{1}{2h}\cdot\frac{\sum_{i=1}^n I(x-h<X_i\le x+h)}{n} \xrightarrow{P} f(x)$$

$$= \frac1n\sum_{i=1}^n \frac1h \cdot \cancel{2} K\left(\frac{X_i-x}{h}\right), \quad K(x) = \frac12 I(-1<x\le1)$$

- More general, KDE

$$\hat f_h(x) = \frac{1}{nh}\sum K\left(\frac{X_i-x}{h}\right), \quad K_{21}, K_{02} < \infty, \quad K_{ij} \triangleq \int u^i K^j du$$

Prop. $E(\hat f_h(x) - f(x)) = \frac12 h^2 f''(x) K_{21} + o(h^2)$

$Var(\hat f_h(x)) = \frac{1}{nh} f(x) K_{02} + o(\frac{1}{nh})$

AMSE $= \lim_{n\to\infty}$ MSE $= \frac14 h^4 f''^2(x) K_{21}^2 + \frac{1}{nh} f(x) K_{02}$

⇒ local h_{opt}

Sketch pf: 作 2次 Taylor

$h = cn^{-\frac15}$ 时, $n^{\frac25}(\hat f_h - f) \to N(\frac{c^2}{2} f''(x)K_{21}, \frac1c f(x)K_{02})$

AIMSE $= \frac14 h^4 \|f''\|^2 K_{21}^2 + \frac{1}{nh} K_{02}$

⇒ global $h_{opt} = C\cdot n^{-\frac15}$

Lec 8 KDE +

- KDE estimate $f^{(r)}(x)$:

$$\hat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right).$$

$$\Rightarrow \hat{f}_h^{(r)}(x) = \frac{1}{nh^{r+1}}\sum_{i=1}^{n} K^{(r)}\left(\frac{x-X_i}{h}\right)$$

Prop. $E\left(\hat{f}_h^{(r)}(x) - f^{(r)}(x)\right) = \frac{1}{2} f^{(r+2)}(x) k_{21} h^2 + o(h^2)$

$$Var\left(\hat{f}_h^{(r)}(x)\right) = \frac{f(x)}{nh^{2r+1}}\int (K^{(r)})^2 du + o\left(\frac{1}{nh^{2r+1}}\right)$$

$h_{opt} = C \cdot n^{-\frac{1}{2r+5}}$, $AMISE(h_{opt}) = C \cdot n^{-\frac{4}{2r+5}}$

$F(x)$.

$$\hat{F}_h(x) = \frac{1}{nh}\sum_{i=1}^{n}\int_{-\infty}^{x} K\left(\frac{y-X_i}{h}\right)dy$$

$$= \frac{1}{n}\sum G\left(\frac{x-X_i}{h}\right).$$

where $G(x) = \int_{-\infty}^{x} K(t)\,dt$

$E\left(\hat{F}_h(x) - F(x)\right) = \frac{1}{2}F''(x) k_{21} h^2 + o(h^2)$

$Var\left(\hat{F}_h(x)\right) = \frac{1}{n}F(x)(1-F(x)) - \frac{2hf(x)}{n}\int u G(u) K(u) du + o\left(\frac{h}{n}\right)$

$h_{opt} = C \cdot n^{-\frac{1}{3}}$

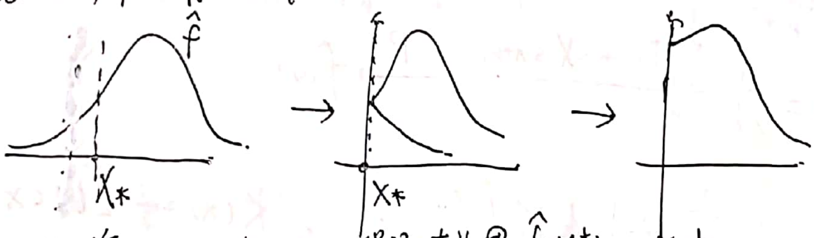
but $AMISE(h_{opt}) = C \cdot n^{-1}$!

para : $O(n^{-1})$.

- Extension:

① $h = h(x)$. (adaptive choice of bw)

② Boundary Correction: (i) 作反射 Reflection



(ii) 作 renormalization, 保证在边界 $\hat{f}(x^*) \to f(x^*)$

(iii) Generalized Jackknifing: combine kernel K and another ~~ker~~ kernel L to have better prop.