מועד ב' 2021

שאלה 1:

נתונים $X_1,\dots,X_n \overset{iid}{\sim} F$ מעל $\mathbb{R}$.

$$T(F) = F\left(\int x\, dF(x)\right)$$

(1) תנו פירוש הסתברותי ל-$T(F)$ ומצאו את אומד ההצבה.

פתרון:

נשים לב: $\int x\, dF(x) = \mathbb{E}_F X_1 = \mu_F$

$$T(F) = F(\mu_F) = \mathbb{P}_F(X_1 \le \mu_F)$$

כלומר, מדובר בהסתברות שהמ"מ קטן מהתוחלת שלו.

כדי למצוא את אומד ההצבה, נחליף את F ב-$\hat{F}_n$.

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{X_i \le x\}$$

$$T(\hat{F}_n) = \hat{F}_n(\underbrace{\mu_{\hat{F}_n}}_{\bar{X}_n}) = \hat{F}_n(\bar{X}_n) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{X_i \le \bar{X}_n\}$$

//

(2) הוכחנו כי $\mathbb{E}_F T(\hat{F}_n) = \mathbb{E}_F F(\bar{X}_{n-1})$

ונשתמש בשוויון זה כדי להראות כי ההטיה של $T(\hat{F}_n)$ נעלמת כאשר $n \to \infty$, עבור F רציפה.

פתרון:

מה עשו

$$\mathbb{E}_F T(\hat{F}_n) = \mathbb{E}_F \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{X_i \le \bar{X}_n\} \overset{\downarrow}{=} \mathbb{P}_F(X_1 \le \underbrace{\bar{X}_n}_{\frac{1}{n}X_1 + \frac{1}{n}X_2 + \cdots})$$

$$= \mathbb{P}_F\Big(\underbrace{\big(1-\tfrac{1}{n}\big)}_{\frac{n-1}{n}} X_1 \le \cancel{\tfrac{1}{n}} \sum_{i=2}^{n} X_i\Big) = \mathbb{P}_F\Big(X_1 \le \underbrace{\frac{1}{n-1}\sum_{i=2}^{n} X_i}_{\bar{X}_{n-1}}\Big)$$

$\mathbb{P}(X\in A) = \mathbb{E}\,\mathbb{P}(X\in A \mid Y)$

$\mathbb{P}(X\in A)$

$= \mathbb{E}\,\mathbb{1}\{X\in A\}$

$= \mathbb{E}\,\mathbb{E}(\mathbb{1}_{X\in A} \mid Y)$

$= \mathbb{E}\,\mathbb{P}(X\in A \mid Y)$

$$= \mathbb{E}_F \mathbb{P}_F(X_1 \le \bar{X}_{n-1} \mid \bar{X}_{n-1}) = \mathbb{E}_F F(\bar{X}_{n-1})$$

$$\boxed{\mathbb{P}_F(X_1 \le a) = F(a)}$$

כדי להראות שההטיה נעלמת, (כאשר $n\to\infty$), נרצה להראות

$$\mathbb{E}_F T(\hat{F}_n) \xrightarrow[n\to\infty]{} T(F)$$

$$\| $$

$$\mathbb{E}_F F(\bar{X}_{n-1})$$

לפי LLN, $\bar{X}_{n-1} \xrightarrow[n\to\infty]{a.s.} \underbrace{\mathbb{E}_F X_1}_{\mu_F}$, כיוון ש-$F$ רציפה, אז

$$F(\bar{X}_{n-1}) \xrightarrow[n\to\infty]{a.s.} F(\mu_F) = T(F)$$

אולי תחת תנאים נוספים,

משפט ההתכנסות הנשלטת: אם $X_n \xrightarrow[n\to\infty]{} X$; $\mathbb{P}(|X_n|\le Y) = 1$, $\mathbb{E}|Y| < \infty$

$$\lim_{n\to\infty} \mathbb{E} X_n = \mathbb{E} \lim_{n\to\infty} X_n = \mathbb{E}X$$

נרצה להפעיל תוחלת על איבר שמאל ועדיין לקבל את אותו גבול.

לשם נצטרך להשתמש במשפט ההתכנסות הנשלטת (או החסומה) כדי להחליף בין גבול ותוחלת. כיוון ש- F חסומה בין 0 ל-1,

$$\lim_{n\to\infty} \mathbb{E}_F T(\hat F_n) = \lim_{n\to\infty} \mathbb{E}_F F(\bar X_{n-1}) = \mathbb{E}_F \lim_{n\to\infty} F(\bar X_{n-1}) = T(F)$$

משפט ההתכנסות הנשלטת.

//

(3) נמצא את פונ' הרגישות (Influence function) של $T(F)$ על קבוצת ההתפלגויות הרציפות עם תומך חסום.

פתרון:

נזכור כי צריך לחשב את נגזרת Hadamard של F בכיוון $G-F$:

$$\dot T_F(G-F) := \lim_{t\downarrow 0} \frac{1}{t}\Big(T\big(F+t(G_t-F)\big) - T(F)\Big) \quad ; \quad \|G_t - G\|_\infty \xrightarrow[t\to 0]{} 0$$

אנו נחשב את הנגזרת בנקודה $G = \Delta_x$; $\Delta_x(u) = 1_{\{x\le u\}}$

אם כן, יהי G_t מסלול כך ש- $\|G_t - G\|_\infty \to 0$ ויהי

$$F_t = F + t(G_t - F)$$

$$T(F_t) = F_t\left(\int x\, dF_t(x)\right) = \underbrace{F\left(\int x\, dF_t(x)\right)}_{*} + t\left(G_t\left(\int x\, dF_t(x)\right) - F\left(\int x\, dF_t(x)\right)\right)$$

$$* \quad F\left(\int x\, dF_t(x)\right) = F\left(\int x\, dF(x) + t\int x\, d\left(G_t(x) - F(x)\right)\right)$$

(טיילור) $$= F\left(\int x\, dF(x)\right) + t\, f\left(\int x\, dF(x)\right)\int x\, d\left(G_t(x) - F(x)\right) + o(t), \quad t\to 0$$

צפיפות.

$$\frac{1}{t}\left(T(F_t) - T(F)\right) = \frac{1}{t}\Big(\underbrace{\cancel{F\left(\int x\, dF(x)\right)}}_{T(F)} + t\, f\left(\int x\, dF(x)\right)\int x\, d\left(G_t(x) - F(x)\right) + o(t)$$

$$+ t\left(G_t\left(\int x\, dF_t(x)\right) - F\left(\int x\, dF_t(x)\right)\right) - \cancel{T(F)}\Big)$$

$$= f(\mu_F)\int x\, d\underbrace{\left(G_t(x) - F(x)\right)}_{\to G - F} + \underbrace{o(1)}_{\to 0} + \underbrace{G_t}_{\to G}\left(\int x\, d\underbrace{F_t(x)}_{\to F}\right) - F\left(\int x\, d\underbrace{F_t(x)}_{\to F}\right)$$

$$\xrightarrow[t\downarrow 0]{} f(\mu_F)\int x\, d\left(G(x) - F(x)\right) + G(\mu_F) - F(\mu_F)$$

//

נגדיר $G = \Delta_x$ $\qquad \mu_G = x$

$$\int \psi(u)\, d\delta_x(u) = \psi(x)$$

ניקח $\psi(u) = u$. ואז מה קיבלת תוחלת ביחס ל- $d\delta_x(u)$.

ולכן

$$L_F(x) = f(\mu_F)(x - \mu_F) + \mathbb{1}\{x \le \mu_F\} - F(\mu_F)$$

//

שאלה ג:

נסתכל על בעיית הרגרסיה

$$Y_j = f(X_j) + \varepsilon_j \quad ; \quad j=1,\dots,n$$

$$\varepsilon_j \overset{iid}{\sim} N(0,1) \quad ; \quad X_1,\dots,X_n \overset{iid}{\sim} U([0,1])$$

אומד גרעין ל- f נתון ע"י:

$$\hat{f}(x) = \frac{1}{nh}\sum_{j=1}^{n} Y_j K\left(\frac{x-X_j}{h}\right) \quad ; \quad h>0 \; ; \; supp(K)\subseteq[-1,1]$$

(1) הראו כי ניתן לבחור את K כך שההטיה מקיימת את החסם

$$\sup_{f\in\Sigma(\beta,L)} \left|\mathbb{E}_f \hat{f}(x_0) - f(x_0)\right| \leq C_1 h^{\beta}$$

בכל נקודה פנים $x_0\in(0,1)$, כאשר $\Sigma(\beta,L)$ מחלקת Hölder ו- C_1

קבוע שלא תלוי ב- n,h.

פתרון:

$$\mathbb{E}_f \hat{f}(x_0) - f(x_0) = \frac{1}{nh}\sum_{j=1}^{n} \mathbb{E}_f\left(f(X_j)+\varepsilon_j\right) K\left(\frac{x_0-X_j}{h}\right) - f(x_0)$$

$\mathbb{E}\varepsilon_j = 0$, i.i.d.

$$= \frac{1}{h}\mathbb{E}_f\left(f(X_1)K\left(\frac{x_0-X_1}{h}\right)\right) - f(x_0)$$

$X_1 \sim U([0,1])$

$$= \frac{1}{h}\int_0^1 f(u) K\left(\frac{x_0-u}{h}\right) du - f(x_0)$$

$t = \frac{x_0-u}{h}$

$u=0 \Rightarrow t=\frac{x_0}{h}$

$u=1 \Rightarrow t=\frac{x_0-1}{h}$

$$= \int_{\frac{x_0-1}{h}}^{x_0/h} f(x_0-th)K(t)dt - f(x_0)$$

עבור h קטן $\cong \int_{-1}^{1} f(x_0-th)K(t)dt - f(x_0)$

וזה נכון כיוון ש-
$x_0\in(0,1)$

$$= \int_{-1}^{1} \left(f(x_0-th)-f(x_0)\right)K(t)\,dt$$

$$\Rightarrow \left|\mathbb{E}_f \hat{f}(x_0)-f(x_0)\right| = \left|\int_{-1}^{1} \frac{1}{l!}\left(f^{(l)}(x_0-\xi th)-f^{(l)}(x_0)\right)(-ht)^l K(t)\,dt\right|$$

וגם ש- K הוא
גרעין מסדר l
$l=\lfloor\beta\rfloor$
$\int_{-1}^{1} u^j K(u)\,du=0$
$j=1,\dots,l-1$

$$\le \int_{-1}^{1} \frac{1}{l!} L\,|ht|^{\beta-l}\cdot|ht|^l\cdot|K(t)|\,dt$$

$f\in\Sigma(\beta,L)$

$$= \underbrace{\frac{L}{l!}\int_{-1}^{1}|t|^{\beta}|K(t)|\,dt}_{C_1}\cdot h^{\beta}$$

אז סה"כ קיבלנו

$$\sup_{f\in\Sigma(\beta,L)}\left|\mathbb{E}_f \hat{f}(x_0)-f(x_0)\right|\le C_1\cdot h^{\beta}$$

כנדרש.

(3) יהי K גרעין סימטרי שמקיים את התנאים הקודמים.
הראו כי לכל $f\in\Sigma(\beta,L)$ אם $f(0)\neq 0$, אזי הגרעין לא קונסיסטנטי בנקודה
$x_0=0$, וזה לא תלוי בבחירת h.

נראה: הסתכלו על רכיב ההטיה.

פתרון

$$\mathbb{E}_f \hat{f}(0)-f(0)$$

$$\mathbb{E}_f \hat{f}(0) = \frac{1}{h}\int_0^1 f(s)\,K\!\left(\frac{0-s}{h}\right)ds = \frac{1}{h}\int_0^1 f(s)\,K(s/h)\,ds$$

$$= \int_0^{1/h} f(th)\,K(t)\,dt \xrightarrow[h\to 0]{} f(0)\int_0^1 K(t)\,dt = \frac{1}{2}f(0)$$

החלפת משתנים.

$$1 = \int_{-1}^{1} K(t)\,dt = 2\int_{0}^{1} K(t)\,dt$$

כלומר, קיבלנו שבקצה ההטיה לא נעלמת כאשר $h \to 0$ ולכן אין קונסיסטנטיות בנקודה זו.

//.