תרגיל מועד ב שאלה 8-2-4 בשבוע 1

שאלה: נסתכל על $D_n=\sqrt{n}\sup_{x\in\mathbb{R}}|\hat{F}-F|$, $\hat{F}_n(x)=\frac{1}{n}\sum_{j=1}^{n}1\{x_j\le x\}$

כאשר F בדידה עם תומך $\{x_1,\dots,x_k\}$

עם הסתברויות $\{p_1,\dots,p_k\}$

(א) בדוק בהתפלגות ש D_n תלויה ב x_j או ב p_j

בעיקרון D_n ב"ת של F כי לא רציפה, נסתכל יש

נחזור למקרה שלנו: ל $\hat{F}_n$ יש קפיצות ב x_j לכן

$$\hat{F}_n(x_j)\underset{def}{=}\hat{p}_j=\frac{1}{n}\sum_{m=1}^{n}1\{x_m=x_j\}\qquad x_j\in\{x_1,\dots,x_k\}$$

מכיוון שמדובר כאן במקרה בדיד נקבל

$$D_n=\sqrt{n}\max_{1\le j\le k}|\hat{p}_j-p_j|$$

והתפלגות תלויה ב p_j ולא ב x_j

ראינו כי $n\hat{p}\sim Multi(n,p)$ כאשר $p=\begin{pmatrix}p_1\\ \vdots\\ p_k\end{pmatrix}$

לכן עבור בעיקר לערך זה לא תלוי בקבוצה.

הערה: גם ההתפלגות האסימפטוטית הגבולית של D_n

ובכלל עדיין תלויות ב-p_j

דוגמא:

עבור $k=2$

$$D_n = \max\{|p_1-\hat{p}_1|, |p_2-\hat{p}_2|\}$$

נשים לב ש $p_2 = 1-p_1 \Rightarrow \hat{p}_2 = 1-\hat{p}_1$

$$\mathbb{P}_F\left(D_n = \sqrt{n}(1-p_1)\right) = p_1^n$$

(זה יקרה כאשר כל התצפיות נפלו)

א/ צריך לעבוד משהו עם D_n ולא

להסתכל על המקרה הכללי

ב) האם D_n מתכנס ואם כן לאן?

לפי CLT רב מימדי

$$\sqrt{n}[\hat{p}-p] \xrightarrow{d} N(\vec{0}, \Sigma) = Y_j$$

$$\Sigma = diag(p) - pp^T = \begin{pmatrix} p_1 & & 0 \\ & \ddots & \\ 0 & & p_k \end{pmatrix}$$

$$\hat{p} = \begin{pmatrix} \hat{p}_1 \\ \vdots \\ \hat{p}_k \end{pmatrix}$$

(כל רכיב כאן תלוי ב-n)

כיוון ש $x \to \max_j x_j$ רציפה

$$D_n \xrightarrow{d} \max_j (|Y_j|)$$

[לא ידועה התפלגות מפורשת כי F לא רציפה]

הוכחה:

$$[F(X_1), \ldots, F(X_n)] \overset{d}{=} (V_1, \ldots, V_n)$$

כאשר F רציפה | $V_1, \ldots, V_n \overset{i.i.d}{\sim} U(0,1)$

אין צורך להניח הפיכות של F^{-1} כי רק רציפה לכן צריך

להשתמש בהגדרה "מוכללת"

נסמן $U(x) = F(x)$

$$\mathbb{P}[F(X) < x] = E_F \mathbb{1}_{\{F(X)<x\}} = \int_{\mathbb{R}} \mathbb{1}_{\{F(X)<x\}} \, dF(X)$$

$$\frac{dy}{dx} = f_x \Leftarrow u = F(x)$$

$$= \int_{\mathbb{R}} \mathbb{1}_{\{F(x)<x\}} f_x \, dx = \int_0^1 \mathbb{1}_{\{u<x\}} du = \int_0^x 1 \, du = x$$

F כי רציפה לכן f_x קיים

הוכחה: התפלגות של D_n כאשר F רציפה

1 נראה כי

$$D_n \overset{claim}{=} \sqrt{n} \max\left\{ \frac{j}{n} - F(X_{(j)}) \vee \frac{j-1}{n} - F(X_{(j)}) \right\}$$

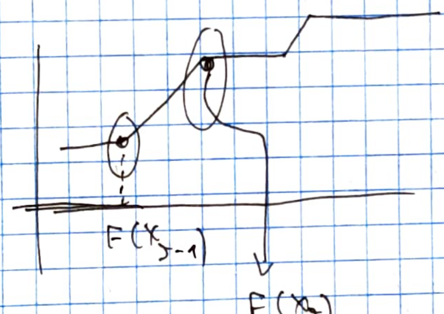
כאשר $X_{(j)}$ הוא סדר סטטיסטי $X_{(1)} \le \ldots \le X_{(n)}$

נסתכל על $\sqrt{n} D_n$

$$\frac{1}{\sqrt{n}} D_n = \sup_{x\in\mathbb{R}} \left| \frac{1}{n} \sum_{i=1}^{n} 1\!\!1_{\{X_i \le x\}} - F(x) \right|$$

$$= \sup_{u\in[0,1]} \left| \underbrace{\frac{1}{n} \sum_{i=1}^{n} 1\!\!1_{\{F(X_i) \le \underbrace{F(x)}_{u}\}} - \underset{F(x)}{\overset{=}{u}}}_{(*)} \right|$$

נשים לב שהפונקציה $(*)$ היא לינארית בקטעים $[F(X_{j-1}), F(X_j))$



$- U$

מספר

$= \sum 1\!\!1_{\{ - \}}$

לכן אסיק כי ה sup יכול להתקבל רק בנקודות קפיצה

מימין או משמאל ו F היא בנק' השבירה רציפה מימין

לכן הוא מושג ב $F(X_j)$ או בגבול $F(X_j-)$

ומכיוון שמספר הקפיצות הוא סופי מקבלים כי

יש j קפיצות כאלה

$$D_n = \frac{1}{n} \max_{1\le j\le n} \left[\left| \frac{1}{n} \sum_{i=1}^{n} 1\!\!1_{\{F(X_i) \le F(X_j)\}} - F(X_j) \right| \right.$$

$$\vee \left| \frac{1}{n} \left(\sum_{i=1}^{n} 1\!\!1_{\{F(X_i) < F(X_j)\}} \right) - F(X_j) \right|$$

ו $j-1$ קפיצות כאלה

סופרים