HW 5

סטטיסטיקה 2

גיל קפלן - [redacted]

ראקל אופנשטיין - [redacted]

תרגיל בית 5

שאלה 1

יהיו $X_1,\dots,X_n \sim F$. ותהי $\hat{F}_n(x)$ פונקציית ההתפלגות האמפירית. הוכיחו כי עבור x נתון מתקיים:

א. $\mathbb{E}\left[\hat{F}_n(x)\right] = F(x)$

ב. $Var\left(\hat{F}_n(x)\right) = \frac{F(x)(1-F(x))}{n}$

ג. $\hat{F}_n(x)$ הוא אומד עקיב ל-$F(x)$

ד. $\hat{F}_n(x)$ מתפלג אסימפטוטית נורמלית

הדרכה: שימו לב שעבור x נתון, המשתנה $I(X \le x)$ הוא משתנה ברנולי.

שאלה 2

נתון מדגם של זוגות משתנים מקריים $(X_1,Y_1),\dots,(X_n,Y_n)$. המתאם בין שני משתנים מקריים X ו-Y מוגדר להיות:

$$\rho = \frac{cov(X,Y)}{\sigma_x \sigma_y}$$

הראו איך ניתן לאמוד את המתאם באמצעות אומדי plug-in.

שאלה 3

יהיו $X_1,\dots,X_n \sim F$ ותהי $\hat{F}_n(x)$ פונקציית ההתפלגות האמפירית.

א. עבור x,y נתונים ושונים מצאו את $Cov\left(\hat{F}_n(x),\hat{F}_n(y)\right)$

ב. יהיו $a,b \in \mathbb{R}$ כך ש-$a<b$ ונגדיר $\theta = T(F) = P(a < x \le b)$. הראו כי זהו פונקציונל לינארי, הציעו אומד plug-in ל-θ ומצאו רווח סמך ברמת סמך מקורבת $1-\alpha$ עבור θ.

שאלה 4

נגדיר את השברון ה-α של פונקציית התפלגות אמפירית באופן הבא:

$$\hat{F}_n^{-1}(\alpha) = \inf\left\{x :\ \hat{F}_n(x) \ge \alpha\right\}$$

יהי $\hat{\theta}_\alpha^*$ השברון ה-α של ערכי הבוטסטראפ שהתקבלו. הראו שרווח הסמך הפיווטלי ניתן לכתיבה כ-

$$\left[2\hat{\theta}_n - \hat{\theta}^*_{1-\frac{\alpha}{2}}, 2\hat{\theta}_n - \hat{\theta}^*_{\frac{\alpha}{2}}\right]$$

מה אתם יכולים להסיק על אורך רווח הסמך הפיווטלי לעומת אורך רווח סמך מבוסס אחוזונים?

שאלה 5 – רשות (בונוס 10 נקודות)

יהיו $X_1,\dots,X_n \sim F$ ו-$\theta = T(F)$ פונקציונל כלשהו, כך ש- $\hat{\theta}_n \equiv T_n = T(X_1,\dots,X_n)$.

הוכיחו/ הפריכו:

$$E_{\hat{F}_n}[T_n] = T_n$$

במילים אחרות, האם תוחלת האומד המחושבת על פי ההתפלגות האמפירית שווה בהכרח לאומד המקורי?

שאלה 1

יהיו $X_1,\dots,X_n \sim F$. ותהי $\hat{F}_n(x)$ פונקציית ההתפלגות האמפירית. הוכיחו כי עבור x נתון מתקיים:

א. $\mathbb{E}\left[\hat{F}_n(x)\right] = F(x)$

$$E\left(\hat{F}_n(x)\right) \overset{\text{הגדרה}}{=} E\left(\frac{1}{n}\sum_{i=1}^{n} I(X_i \le x)\right) \overset{\text{Linearity}}{=} \frac{1}{n}\sum_{i=1}^{n} E\left(I(X_i \le x)\right)$$

$$\underset{\substack{\text{נתון כי } X_i \\ \text{ש"ה לכל } i}}{=} \cancel{\frac{n}{n}} E\left(I(X_1 \le x)\right) \underset{\substack{\text{תוחלת של אינדיקטור} \\ \text{שווה להסתברות של המאורע}}}{=} P(X_1 \le x) \underset{\text{הגדרה}}{=} F(x)$$

ההוכחה פה קוציונ' ■

ב. $Var\left(\hat{F}_n(x)\right) = \frac{F(x)\left(1-F(x)\right)}{n}$

$$Var\left(\hat{F}_n(x)\right) = E\left(\hat{F}_n^2(x)\right) - E^2\left(\hat{F}_n(x)\right) \underset{\substack{\text{ההצבה לפי הגדרה} \\ \text{והצבה וחילוק פי } n}}{=} E\left(\hat{F}_n^2(x)\right) - \frac{F^2(x)\cdot n}{n} \overset{*}{=}$$

$$\overset{*}{=} \frac{F(x)\left(1+(n-1)F(x)\right)}{n} - \frac{F(x)\left(nF(x)\right)}{n} = \frac{F(x)\left(1+(n-1-n)F(x)\right)}{n}$$

$$= \frac{F(x)\left(1-F(x)\right)}{n} \quad ■$$

$$(*)\quad E\left(\hat{F}_n^2(x)\right) = E\left(\left(\frac{1}{n}\sum_{i=1}^{n} I(X_i \le x)\right)^2\right) = \frac{1}{n^2}E\left(\sum_{i=1}^{n} I(X_i \le x)\sum_{j=1}^{n} I(X_j \le x)\right)$$

$$= \frac{1}{n^2}E\left(\sum_{i=1}^{n} I^2(X_i \le x) + \sum_{\substack{i \ne j \\ i,j}} I(X_i \le x, X_j \le x)\right) = \frac{1}{n^2}\left(n E\left(I(X_i \le x)\right)\right) + \frac{1}{n^2}(n-1)\cdot n\, E\left(\overbrace{I(X_1 \le x, X_2 \le x)}^{\text{ב"ת}}\right)$$

האינדיקטור לכן
$$I^2(X_i \le x) = I(X_i \le x) = \begin{cases} 1 & X_i \le x \\ 0 & \text{else} \end{cases}$$

$$E\left(I(X_1 \le x, X_2 \le x)\right) = E\left(I(X_1 \le x)\right)\cdot E\left(I(X_2 \le x)\right)$$

$$\underset{\text{תוחלת של הסתברות}}{=} \frac{F(x)}{n} + \frac{n-1}{n}F^2(x) = \frac{F(x)\left(1+(n-1)F(x)\right)}{n}$$

$$E\left(I(X \le x)\right) = P(X \le x) = F(x)$$

ג. $\hat{F}_n(x)$ הוא אומד עקיב ל-$F(x)$

הוכחנו בסעיף ב' כי ב. $Var\left(\hat{F}_n(x)\right)=\frac{F(x)\left(1-F(x)\right)}{n}$

$$\Rightarrow \frac{\overbrace{F(X)(1-F(X))}^{\text{קבוע כתלות ב-}\hat{x}}}{n} \xrightarrow{n\to\infty} 0$$

לכן עקיב ל-$F(x)$

ד. $\hat{F}_n(x)$ מתפלג אסימפטוטית נורמלית

מתקיים כי $I(X \le x)$ הוא משתנה ברנולי. לכן $\sum_{i=1}^{n} I(X\le x)$

כאשר $n\to\infty$ מתפלג אסימפטוטית נורמלית מכיוון שהוכחנו בכיתה כי ברנולי היא אסימפטוטית נורמלית.

שאלה 2

נתון מדגם של זוגות משתנים מקריים $(X_1, Y_1), \ldots, (X_n, Y_n)$. המתאם בין שני משתנים מקריים X ו-Y מוגדר להיות:

$$\rho = \frac{cov(X, Y)}{\sigma_x \sigma_y}$$

הראו איך ניתן לאמוד את המתאם באמצעות אומדי plug-in.

$X_1, \ldots, X_n \sim X$
$Y_1, \ldots, Y_n \sim Y$, $\hat{F}_n(x, y) = \sum_{i=1} \hat{p}_i$

$$\rho = T(F_n(X, Y)) = \frac{E(XY) - E(X)E(Y)}{E(X)E(Y)} = \frac{\int\int_{\substack{x \in dom(X) \\ y \in dom(Y)}} xy \, dF(x) \cdot dF(y)}{E(X)E(Y)} - 1$$

$$\Rightarrow \quad \hat{\rho} = \frac{\cancel{\frac{1}{n}} \hat{\sum_{i=1}} \hat{\sum_{j=1}} x_i Y_j}{\cancel{\frac{1}{n}} \left(\hat{\sum_{i=1}} x_i \hat{\sum_{j=1}} Y_i \right)} - 1 = \frac{\hat{\sum_{i=1}} \hat{\sum_{j=1}} x_i y_j}{\hat{\sum_{i=1}} \hat{\sum_{j=1}} x_i Y_j} - 1 = 1 - 1 = 0$$

שאלה 3

יהיו $X_1,\dots,X_n\sim F$ ותהי $\hat{F}_n(x)$ פונקציית ההתפלגות האמפירית.

א. עבור x,y נתונים ושונים מצאו את $Cov\left(\hat{F}_n(x),\hat{F}_n(y)\right)$

ב. יהיו $a,b\in\mathbb{R}$ כך ש-$a<b$ ונגדיר $\theta=T(F)=P(a<x\le b)$. הראו כי זהו פונקציונל לינארי, הציעו אומד plug-in ל-θ ומצאו רווח סמך ברמת סמך מקורבת $1-\alpha$ עבור θ.

(א

$$Cov\left(\hat F_n(x),\hat F_n(y)\right)=E\left(\hat F_n(x)\hat F_n(y)\right)-E\left(\hat F_n(x)\right)\cdot E\left(\hat F_n(y)\right)$$

$$=\underbrace{E\left(\frac1n\sum_{i=1}^{n}I(X_i\le x)\frac1n\sum_{j=1}^{n}I(X_i\le y)\right)}_{*}-E\left(\frac1n\sum_{i=1}^{n}I(X_i\le X)\right)\cdot E\left(\frac1n\sum_{i=1}^{n}I(X_i\le X)\right)$$

הוכחנו בשאלה 2א.

$$\overset{*}{=}\frac1n F_x(\min(x,y))+\frac{n-1}{n}F_x(x)F_x(y)-F_x(x)\cdot F_x(y)=\frac1n\left(F_x(\min(x,y))-F_x(x)F_x(y)\right)$$

כיוון

$$(*)\quad E\left(\frac1n\sum_{i=1}^{n}I(X_i\le x)\frac1n\sum_{i=1}^{n}I(X_i\le y)\right)=\frac1{n^2}\left(E\left(\sum_{i=1}^{n}I(X\le\min(x,y))\right)+E\left(\underbrace{\sum_{i\ne j}^{n}I(X_i\le x,X_j\le y)}_{\text{ב"ת כי } i\ne j}\right)\right)$$

$$=\frac{n}{n^2}F(\min(x,y))+\frac{n(n-1)}{n^2}F(x)F(y)=\frac1n F(\min(x,y))+\frac{n-1}{n}F(x)F(y)$$

(ב

$$P(a<X\le b)=\int_a^b F(x)\,dx=\int_a^b 1\,dF(x)$$

לכן אם נגדיר $r(x)=I(a<x_i\le b)$, נקבל

$$T(F)=\int_{-\infty}^{\infty}r(x)\,dF(x)$$

לכן הפונקציונל לינארי.

אומד plug-in: הפונקציונל לינארי ולכן מהתירגול:

$$\hat{\theta} = T(\hat{F}_n(x)) = \frac{1}{n}\sum_{i=1}^{n} r(X_i) = \frac{1}{n}\sum_{i=1}^{n} I(a < X_i \le b)$$

רווח סמך: מתקיים כי:

$$Var(T(\hat{F}_n(x))) = Var\left(\frac{1}{n}\sum_{i=1}^{n} I(a<X_i\le b)\right) =$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} Var(I(a<X_i\le b))$$

$$\underset{\uparrow}{=} \frac{1}{n} P(a<X\le b)(1-P(a<X\le b))$$

$I(a<X_i\le b) \sim Ber(P(a<X_i\le b))$

לכן האומד לשונות הוא:

$$\hat{var}(\hat{\theta}) = Var(T(\hat{F}_n(x))) =$$

$$= \frac{1}{n}[\hat{F}_n(b) - \hat{F}_n(a)][1 + \hat{F}_n(a) - \hat{F}_n(b)]$$

$$\Rightarrow \hat{se} = \sqrt{\frac{1}{n}[\hat{F}_n(b) - \hat{F}_n(a)][1 + \hat{F}_n(a) - \hat{F}_n(b)]}$$

מהקירוב הנורמלי, נקבל רווח סמך ברמת סמך $1-\alpha$ הבא:

$$T(\hat{F}_n(x)) \pm z_{\frac{\alpha}{2}}\sqrt{\frac{1}{n}[\hat{F}_n(b) - \hat{F}_n(a)][1 + \hat{F}_n(a) - \hat{F}_n(b)]}$$

שאלה 4

נגדיר את השברון ה-α של פונקציית התפלגות אמפירית באופן הבא:

$$\hat{F}_n^{-1}(\alpha) = \inf\{x : \hat{F}_n(x) \geq \alpha\}$$

יהי $\hat{\theta}_\alpha^*$ השברון ה-α של ערכי הבוטסטראפ שהתקבלו. הראו שרווח הסמך הפיווטלי ניתן לכתיבה כ-

$$\left[2\hat{\theta}_n - \hat{\theta}^*_{1-\frac{\alpha}{2}}, 2\hat{\theta}_n - \hat{\theta}^*_{\frac{\alpha}{2}}\right]$$

מה אתם יכולים להסיק על אורך רווח הסמך הפיווטלי לעומת אורך רווח סמך מבוסס אחוזונים?

$$\hat{g}_n = \frac{1}{B}\sum_{i=1}^{n} I\{\sqrt{n}(\hat{\theta}_n^{*b} - \hat{\theta}) \leq x\}$$

$$\frac{\alpha}{2} = P(\hat{\theta}_n^{*b} \leq \hat{\theta}^*_{\frac{\alpha}{2}}) = P(\hat{\theta}_n^{*b} - \hat{\theta}_n \leq \hat{\theta}^*_{\frac{\alpha}{2}} - \hat{\theta}_n) =$$

$$= P(\sqrt{n}(\hat{\theta}_n^{*b} - \hat{\theta}_n) \leq \sqrt{n}(\theta_{\frac{\alpha}{2}} - \hat{\theta}_n))$$

$$\Rightarrow \sqrt{n}(\hat{\theta}^*_{\frac{\alpha}{2}} - \hat{\theta}_n) = \hat{g}_{\frac{\alpha}{2}}$$

נציב בר"ס הנלמד בתרגול:

$$\left[\hat{\theta}_n - \frac{\hat{g}_{1-\frac{\alpha}{2}}}{\sqrt{n}}, \hat{\theta}_n - \frac{\hat{g}_{\frac{\alpha}{2}}}{\sqrt{n}}\right]$$

ונקבל:

$$\left[2\hat{\theta}_n - \hat{\theta}^*_{1-\frac{\alpha}{2}}, 2\hat{\theta}_n - \hat{\theta}^*_{\frac{\alpha}{2}}\right]$$

אורך ר"ס פיווטלי היא כמעט פי 2 יותר גבוה מהר"ס המבוסס אחוזונים.

שאלה 5 – רשות (בונוס 10 נקודות)

יהיו $X_1, \dots, X_n \sim F$ ו-$\theta = T(F)$ פונקציונל כלשהו, כך ש- $\hat{\theta}_n \equiv T_n = T(X_1, \dots, X_n)$.

הוכיחו/ הפריכו:

$$E_{\hat{F}_n}[T_n] = T_n$$

במילים אחרות, האם תוחלת האומד המחושבת על פי ההתפלגות האמפירית שווה בהכרח לאומד המקורי?

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

$$\mathbb{E}_{\hat{F}_n}(T_n) = \int_{-\infty}^{\infty} T_n \, d\hat{F}_n = \frac{1}{n} \sum_{i=1}^{n} T(X_1, \dots, X_n) = \frac{n}{n} T_n = T_n$$

לכן הטענה נכונה.