תרגול מספר ב

מה הבעיות Bootstrap נועד לאמוד שונות/הטיה של אומדנים

יהיו $X_1, \dots, X_n \overset{iid}{\sim} F$ ו-$\hat{F}_n$ פונק' התפלגות אמפירית

$$t \in \mathbb{R} \qquad E_F\, \Psi(F, \hat{F}_n, t) = 0$$

נרצה למצוא את $\hat{t}$ שפותר את המשוואה הנ"ל ב-t

הצעה לפתרון: Empirical Bootstrap

$$E_{\hat{F}_n}(\hat{F}_n, F_n^*, t)$$

כאשר F_n^* פונק' התפלגות אמפירית של $\hat{F}_n$

$X_1^*, \dots, X_n^* \overset{iid}{\sim} \hat{F}_n$ דוגמים באופן אחיד מתוך $X_1, \dots, X_n$

ניתן להשתמש ב-$\hat{t}$ בשביל לחשב רווחי סמך

דוגמא 1.5 מהסיכומים של פבל:

מבוא אומדן Bootstrap לשונות עבור הפונקציונל

$$T(F) = \left[\int x\, dF\right]^r$$

פתרון: באופן כללי עבור $T(\hat{F}_n)$ נקבל

$$\underbrace{\mathrm{var}\left[T(\hat{F}_n)\right]}_{t} = E_F\left[T(\hat{F}_n)^2\right] - \left(E_F\, T(\hat{F}_n)\right)^2$$

$$\Downarrow$$

$$\hat{t} = E_{\hat{F}_n}\left[T(F_n^*)^2\right] - \left[E_{\hat{F}_n} T(F_n^*)\right]^2$$

נחשב ע"י המשוואה לעיל

המעבר לעבור בפילוג מ-X^* ⇐ X בגלל זה הניסויים

הסמויים

$$T(\hat{F}_n) = (\overline{X}_n)^2$$

$$\hat{\varepsilon} = \underbrace{E_{\hat{F}_n}(\overline{X}_n^*)^4}_{\mathrm{I}} - \underbrace{\left(E_{\hat{F}_n}(\overline{X}^{*2})\right)^2}_{\mathrm{II}}$$

II $\quad E_{\hat{F}_n}(\overline{X}_n^{*2}) = \mathrm{var}_{\hat{F}_n} \overline{X}_n^* + \left(E\, \overline{X}_n^*\right)^2$

$$= \frac{\mathrm{var}(X_1^*)}{n} + \left(E_{\hat{F}_n} X_1^*\right)^2$$

$$E_{\hat{F}_n}(X_1^*) = \sum_{i=1}^{n} \frac{1}{n} x_i = \overline{X}_n$$

$$\mathrm{var}_{\hat{F}_n}(X_1^*) = E_{\hat{F}_n}(X_1^* - \overline{X}_n)^2 = \sum_{i=1}^{n} \frac{1}{n}(x_i - \overline{X}_n) = \hat{\mu}_2$$

[illegible]

K [illegible] $\quad \hat{\mu}_k = \frac{1}{n}\sum_{i=1}^{n} (x_i - \overline{X}_n)^k = E_F (X_1^* - \overline{X}_n)^k$

$$= \frac{\hat{\mu}_2}{n} + (\overline{X}_n)^2$$

I $\quad E_{\hat{F}_n}\left[(\overline{X}_n^*)^4\right] = E_{\hat{F}_n}\left[(\overline{X}_n^* - \overline{X}_n + \overline{X}_n)^4\right]$

$$= E_{\hat{F}_n} \overbrace{(\overline{X}_n^* - \overline{X}_n)^4}^{B} + 4\overline{X}_n \overbrace{E_{\hat{F}_n}(\overline{X}_n^* - \overline{X}_n)^3}^{a} \quad \frac{\hat{\mu}_3}{n}$$

$\simeq$

$$+ 6\overline{X}_n^2 \underbrace{E_{\hat{F}_n}(\overline{X}_n^* - \overline{X}_n)^2}_{} \rightarrow \mathrm{var}_F(\overline{X}_n^*)$$

$$\cancel{-4\,\bar{X}_n^{3}\,E_{\hat{F}_n}\left(\bar{X}_n^{*}-\bar{X}_n\right)}+\bar{X}_n^{4}$$

נחשבים

$$A:\quad E_{\hat{F}_n}\left(\bar{X}_n^{*}-\bar{X}_n\right)^{3}=E_{\hat{F}_n}\left[\frac{1}{n}\sum_{i=1}^{n}\left(X_i^{*}-\bar{X}_n\right)\right]^{3}$$

$$=\frac{1}{n^{3}}\,E_{\hat{F}_n}\sum_{i=1}^{n}\left(X_i^{*}-\bar{X}_n\right)^{3}=\frac{1}{n^{3}}\,n\,E_{\hat{F}_n}\left(X_i-\bar{X}_n\right)^{3}$$

הכל מוכפל בגורמים

$$\left[E_{\hat{F}_n}\left(X_i^{*}-\bar{X}_n\right)=0\right]$$

$$=\frac{1}{n^{2}}\cdot\hat{\mu}_3$$

$$B:\quad E_{\hat{F}_n}\left(\bar{X}_n^{*}-\bar{X}_n\right)^{4}=E_{\hat{F}_n}\left[\frac{1}{n}\sum_{i=1}^{n}\left(X_i^{*}-\bar{X}_n\right)\right]^{4}$$

$$=\frac{1}{n^{4}}\cdot n\cdot\underbrace{E_{\hat{F}_n}\left(X_i^{*}-\bar{X}_n\right)^{4}}_{\hat{\mu}_4}+\frac{1}{n^{4}}\,3n(n-1)\Big(\underbrace{E_{\hat{F}_n}\left(X_i^{*}-\bar{X}_n\right)^{2}}_{\hat{\mu}_2}\Big)^{2}$$

$$=\frac{1}{n^{3}}\,\hat{\mu}_4+3\,\frac{1}{n^{3}}\,(n-1)\,\hat{\mu}_2$$

נציב הכל ונחשב

$$\hat{t}_n=\frac{1}{n^{2}}\left(\hat{\mu}_4-3\hat{\mu}_2^{2}\right)+\frac{1}{n^{2}}\left(2\hat{\mu}_2+4\hat{\mu}_3\,\bar{X}_n\right)+\underline{\frac{1}{n}\,4\bar{X}_n^{2}\,\hat{\mu}_2}$$

זה שווה ל שונות של $T(\hat{F}_n)$

[illegible]

$$Var_F\left[T(\hat{F}_n)\right]=Var_F\left(\bar{X}_n^{2}\right)=\overbrace{4\mu^{2}\sigma_F^{2}\cdot\frac{1}{n}}+O\left(n^{-\frac{3}{2}}\right)$$

* הפעם נדבר מעט יותר על שאלות Bootstrap

שאלה: עבור מדגם בגודל $n=3$ ובהנחה שהתצפיות שונות זו מזו

מצאו את התפלגות Bootstrap של הממוצע האמפירי

פתרון: צריך למצוא את $\mathbb{P}_{\hat{F}_n}(\overline{X}_n^* = x)$ עבור כל ערך x רלוונטי

נסמן את המדגם (X_1, X_2, X_3) ואת מדגם (X_1^*, X_2^*, X_3^*)

Bootstrap

X_i^* מקבל כל ערך מערכי המדגם (X_1, X_2, X_3) בהסתברות $\frac{1}{3}$

יש 3^3 אפשרויות ל-X^* כי זה מדגם עם החזרה

unordered entries of X^*	frequency	value x of $\overline{X}^*$	$\mathbb{P}_{\widehat{F_n}}(\overline{X}^* = x)$
(X_1, X_1, X_1)	1	$\frac{1}{3}X_1 + \frac{1}{3}X_1 + \frac{1}{3}X_1$	$\frac{1}{27}$
(X_2, X_2, X_2)	1	X_2	$\frac{1}{27}$
(X_3, X_3, X_3)	1	X_3	$\frac{1}{27}$
(X_1, X_1, X_2)	3	$\frac{2}{3}X_1 + \frac{1}{3}X_2$	$\frac{3}{27}$
(X_1, X_1, X_3)	3	$\frac{2}{3}X_1 + \frac{1}{3}X_3$	$\frac{3}{27}$
(X_2, X_2, X_1)	3	$\frac{2}{3}X_2 + \frac{1}{3}X_1$	$\frac{3}{27}$
(X_2, X_2, X_3)	3	$\frac{2}{3}X_2 + \frac{1}{3}X_3$	
(X_3, X_3, X_2)	3	$\frac{2}{}$	
(X_1, X_2, X_3)	6	$\frac{1}{3}X_1 + \frac{1}{3}X_2 + \frac{1}{3}X_3$	$\frac{6}{27} = \frac{2}{9}$

תרגיל:

בנו **רווח סמך** ברמת $1-\alpha$ בכלי ה Bootstrap עבור $\tau(F)=F(x_0)$

ו x_0 נקודה נתונה

פתרון: בנייה הנ״ל מהרצאות Bootstrap

$$\mathbb{P}_{\hat{F}_n}\left(\left|\hat{F}_n(x_0)-F_n^*(x_0)\right|\le\epsilon\right)=1-\alpha$$

נרצה למצוא את ϵ ואז רווח הסמך יהיה

$$\hat{F}_n(x_0)\pm\hat{\epsilon}_{n,\alpha}$$

$$\mathbb{P}_{\hat{F}_n}\left(\left|\hat{F}_n(x_0)-F_n^*(x_0)\right|\le\epsilon\right)=\mathbb{P}_{\hat{F}_n}\left[\hat{F}_n(x_0)-\epsilon\le F_n^*(x_0)\le\hat{F}_n(x_0)+\epsilon\right]$$

$$=\mathbb{P}_{\hat{F}_n}\left[\underbrace{n\hat{F}_n(x_0)-n\epsilon}_{I_1}\le\underbrace{\sum\mathbb{1}_{\{X_i^*\le x_0\}}}_{\sim Bin(n,\ \ )}\le\underbrace{n\hat{F}_n(x_0)+n\epsilon}_{I_2}\right]$$

$$\mathbb{E}_{\hat{F}_n}\mathbb{1}_{\{X_i^*\le x_0\}}==\hat{F}_n(x_0)$$

$$\left(\sum_{i=1}^{n}\frac{1}{n}\cdot\mathbb{1}_{\{X_i\le x_0\}}\right)$$

$$=\sum_{k\in[I_1,I_2]}\binom{n}{k}\left[\hat{F}_n(x_0)\right]^k\left[1-\hat{F}_n(x_0)\right]^{n-k}$$

$$k:\left|\frac{k}{n}-\hat{F}_n(x_0)\right|\le\epsilon$$

$$\left(\sum\mathbb{1}_{\{X_i^*\le x_0\}}=k\ \ \text{כ׳}\right)$$

במקרה הנ״ל אי אפשר לגלגל את ε

לכן נגדיל את ε בכל שלב כך שההסתברות תהיה לפחות $1-\delta$

ד) הגדרנו את האינטרוול עבור

$\delta = 0.2$

$n = 10$

$F_n^{-1}(x_0) = \frac{1}{2}$

פתרון:

נרצה למצוא את ה-ε המינימלי עבורו

$$\sum_{k : \left|\frac{k}{10} - \frac{1}{2}\right| \le \varepsilon} \binom{10}{k} \cdot \left(\frac{1}{2}\right)^k \cdot \left(\frac{1}{2}\right)^{10-k} = \left(\frac{1}{2}\right)^{10} \sum_{k : \left|\frac{k}{n} - \frac{1}{2}\right| \le \varepsilon} \binom{10}{k} \ge 1-\delta = 0.8$$

עבור $\varepsilon = 0$:

$$\varepsilon = 0 \Rightarrow \frac{n}{2} = 5 \Rightarrow \left(\frac{1}{2}\right)^{10} \cdot \binom{10}{5} = 0.246$$

לא מתאים

עבור $\varepsilon = \frac{1}{n}$:

$$\varepsilon = \frac{1}{n} \Rightarrow \left|\frac{k}{10} - \frac{1}{2}\right| \le 0.1 \Leftrightarrow 10 \cdot (0.5 - 0.1) \le k \le (0.1 + 0.5) \cdot 10$$

$$4 \le k \le 6$$

$$\Rightarrow \left(\frac{1}{2}\right)^{10} \left[\binom{10}{4} + \binom{10}{5} + \binom{10}{6}\right] = 0.65$$

לא מתאים

עבור $\varepsilon = \frac{2}{n}$: מקבלים הסתברות גבוהה 0.89

ולכן רווח הסמך שלנו יהיה

$$F_n^{-1}(x_0) \pm \frac{2}{n}$$