סטטיסטיקה

הערה

בנושא הסטטיסטיקה, נדון בתחומים:

1. אמידה – אמידה נקודתית ורווחי סמך.
2. בדיקת השערות.

סטטיסטיקה

מודל ופרמטר $\Leftarrow$ מדגם $\Leftarrow$ אומד לפרמטר.

משתנה מקרי.

למשתנה יש התפלגות.

1. מודל: $X \sim \mathcal{F}_\vartheta$.

   ϑ - פרמטר לא ידוע.

בעיית האמידה

נתונים ערכים של משתנה מקרי $X \sim \mathcal{F}_\vartheta$, כאשר $\mathcal{F}$ משפחה של התפלגויות ו – ϑ פרמטר לא ידוע.

צריך להציע אומדן של ϑ.

■

מתוך נתוני המדגם ניתן לבנות "אומד":

$$T = t(X_1, \cdots X_n)$$

דוגמה

$$X \sim Poi(\lambda)$$

נתונים הערכים $X_1 = 5\,, X_2 = 8$.

נציע אומד לפרמטר λ.

הגדרה

אומד T לפרמטר ϑ נקרא אומד חסר הטיה אם $E(T) = \vartheta$ לכל ϑ.

דוגמה

$$X_1, \cdots, X_n \sim N(\mu, \sigma^2)$$

נבחר $T = X_1$. אכן, $E(T) = \mu$.

נבחר $T = \overline{X_n}$. אכן, $E(T) = \mu$.

$\square$

דוגמה

$$X_1, \cdots, X_n \sim b(p)$$

- ניתן להשתמש בערך יחיד.

  מתקיים:

$$E(X_1) = p$$

  לכן: X_1 הוא אומד הטיה.

- מקובל להשתמש בממוצע.

  מתקיים:

$$E(\bar{X}) = E\left(\frac{X_1 + \cdots + X_n}{n}\right) = p$$

  לכן: $\bar{X}$ הוא אומד הטיה.

אמידה נקודתית של פרמטרים של התפלגות נורמלית

משתנים מקריים $X_1, \cdots, X_n \sim N(\mu, \sigma^2)$ בלתי תלויים.

הממוצע $\overline{X_n}$ הוא אומד חסר הטיה לתוחלת μ.

נמצא אומד חסר הטיה לשונות σ^2.

ממוצע המדגם:

$$\overline{X_n} = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i$$

שונות המדגם:

$$s_n^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} (X_i - \overline{X_n})^2$$

אומד חסר הטיה לשונות של האוכלוסייה (נוכיח כעת):

$$s_{n-1}^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (X_i - \overline{X_n})^2$$

חישוב

$$E\left(\sum_{i=1}^{n} (X_i - \overline{X_n})^2\right) = \sum_{i=1}^{n} E((X_i - \overline{X_n})^2)$$

$$E\left(\sum_{i=1}^{n} (X_i - \overline{X_n})^2\right) = \sum_{i=1}^{n} E\left(X_i^2 - 2X_i \cdot \overline{X_n} + \overline{X_n}^2\right)$$

$$E\left(\sum_{i=1}^{n} (X_i - \overline{X_n})^2\right) = \sum_{i=1}^{n} E(X_i^2) - 2E(X_i \cdot \overline{X_n}) + E\left(\overline{X_n}^2\right)$$

$$E\left(\sum_{i=1}^{n} (X_i - \overline{X_n})^2\right) = \sum_{i=1}^{n} E(X_i^2) - 2 \cdot \sum_{i=1}^{n} E(X_i \cdot \overline{X_n}) + \sum_{i=1}^{n} E\left(\overline{X_n}^2\right)$$

- $\underline{E(X_i^2)}$:

$$V(X_i) = E(X_i^2) - E(X_i)^2$$
$$\Downarrow$$
$$E(X_i^2) = V(X_i) + E(X_i)^2$$
$$\Downarrow$$
$$E(X_i^2) = \sigma^2 + \mu^2$$

- $\underline{E\left(\overline{X_n}^2\right)}$:

$$V(\overline{X_n}) = E\left(\overline{X_n}^2\right) - E(\overline{X_n})^2$$
$$\Downarrow$$
$$E\left(\overline{X_n}^2\right) = V(\overline{X_n}) + E(\overline{X_n})^2$$
$$\Downarrow$$
$$E(X_i^2) = \frac{\sigma^2}{n} + \mu^2$$

- $E(X_i \cdot \overline{X_n})$:

$$E(\dot{X_\iota} \cdot \overline{X_n}) = E\left(X_i \cdot \frac{1}{n} \cdot \sum_{j=1}^{n} X_j\right)$$
$$\Downarrow$$

$$E(X_i \cdot \overline{X_n}) = \frac{1}{n} \cdot \sum_{j=1}^{n} E\left(X_i \cdot X_j\right)$$

$$\Downarrow$$

$$E(X_i \cdot \overline{X_n}) = \frac{1}{n} \cdot \left(E(X_i^2) + \sum_{i \neq j=1}^{n} E(X_i) \cdot E\left(X_j\right) \right)$$

$$\Downarrow$$

$$E(X_i \cdot \overline{X_n}) = \frac{\sigma^2 + \mu^2}{n} + \frac{(n-1) \cdot \mu^2}{n}$$

לכן :

$$E\left(\sum_{i=1}^{n} (X_i - \overline{X_n})^2 \right) = \sum_{i=1}^{n} (\sigma^2 + \mu^2) - 2 \cdot \sum_{i=1}^{n} \frac{\sigma^2 + \mu^2}{n} + \frac{(n-1) \cdot \mu^2}{n} + \sum_{i=1}^{n} \left(\frac{\sigma^2}{n} + \mu^2 \right)$$

$$\Downarrow$$

$$E\left(\sum_{i=1}^{n} (X_i - \overline{X_n})^2 \right) = n \cdot (\sigma^2 + \mu^2) - 2 \cdot (\sigma^2 + \mu^2 + (n-1) \cdot \mu^2) + n \cdot \left(\frac{\sigma^2}{n} + \mu^2 \right)$$

$$\Downarrow$$

$$E\left(\sum_{i=1}^{n} (X_i - \overline{X_n})^2 \right) = (n-1) \cdot \sigma^2$$

$$\Downarrow$$

$$E(s_{n-1}^2) = \sigma^2$$

$$\square$$

לכן :

$$s_{n-1}^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (X_i - \overline{X_n})^2$$

הוא אומד חסר הטיה לשונות σ^2.

$$\blacksquare$$

הערה

אין אומד חסר הטיה לסטיית התקן σ.

הערה

$$X_1, \cdots, X_n \sim \mathcal{F}_\vartheta$$

יהיו $T_1 = t(X_1, \cdots, X_n)$ ו - $T_2 = t'(X_1, \cdots, X_n)$ אומדים חסרי הטיה לפרמטר ϑ.

$$E((T - \vartheta)^2) = V_\vartheta(T)$$

הגדרה

אומד חסר הטיה T_1 עדיף על אומד חסר הטיה T_2 אם לכל ϑ מתקיים :

$$V(T_1) \leq V(T_2)$$

דוגמה

$$X_1, \cdots, X_n \sim N(\mu, \sigma^2)$$

X_1 ו - $\overline{X_n}$ אומדים חסרי הטיה לתוחלת μ.

מתקיים :

$$V(X_1) = \sigma^2 > \frac{\sigma^2}{n} = V(\overline{X_n})$$

לכן, $\overline{X_n}$ עדיף על T_1.

תרגיל

$$X_1, \cdots, X_n \sim N(\mu, \sigma^2)$$

נקבע מספרים $\alpha_1, \cdots, \alpha_n$.

נגדיר :

$$T = \sum_{i=1}^{n} \alpha_i \cdot X_i$$

מתקיים :

T אומד חסר הטיה

$$\Updownarrow$$

$$\sum_{i=1}^{n} \alpha_i = 1$$

<u>הוכח</u>: הממוצע עדיף על כל אומד חסר הטיה אחר.

פתרון

נחשב ונקבל ש:

$$V(T) = \sigma^2 \cdot \sum_{i=1}^{n} \alpha_i^2$$

נחשב ונקבל ש: $V(T)$ מינימלי עבור $\alpha_i = 1/n$ לכל i.

הגדרה

אומד T נקרא אומד חסר הטיה בעל שונות מינימלית במידה שווה $(UMVUE)$ אם לכל ϑ ולכל אומד חסר הטיה T' מתקיים:

$$V_{\vartheta}(T) \leq V_{\vartheta}(T')$$

דוגמה

הממוצע עבור התוחלת של התפלגות נורמלית הוא $UMVUE$.

דוגמה

$X_1, \cdots, X_n \sim U[0, \vartheta]$ משתנים מקריים בלתי תלויים.

נמצא אומד חסר הטיה עבור הפרמטר ϑ.

$$E(X_i) = \frac{\vartheta}{2}$$

לכן, $2X_i$ אומד חסר הטיה עבור ϑ.

$$E(\overline{X_n}) = \frac{\vartheta}{2}$$

לכן, $2\overline{X_n}$ אומד חסר הטיה עבור ϑ.

מתקיים:

$$V(2\overline{X_n}) = \frac{1}{3n} \cdot \vartheta^2$$

תזכורת

סטטיסטי הסדר: $Y_1 \leq \cdots \leq Y_n$.

$$E(Y_n) = \frac{n}{n+1} \cdot \vartheta$$

$$\Updownarrow$$

$\frac{n+1}{n} \cdot Y_n$ אומד חסר הטיה עבור ϑ

באופן כללי:

$$E(Y_k) = \frac{k}{n+1} \cdot \vartheta$$

$$\Updownarrow$$

$\frac{n+1}{k} \cdot Y_k$ אומד חסר הטיה עבור ϑ

מתקיים:

$$V\left(\frac{n+1}{k} \cdot Y_k\right) = \frac{n+1-k}{k \cdot (n+2)} \cdot \vartheta^2$$

בפרט:

$$V\left(\frac{n+1}{n} \cdot Y_n\right) = \frac{1}{n \cdot (n+2)} \cdot \vartheta^2$$

לכן, $\frac{n+1}{n} \cdot Y_n$ עדיף על $2\overline{X_n}$.

■

שיטה כללית למציאת אומדים (לאו דווקא חסרי הטיה)

שיטת הנראות המקסימלית

(נניח שההתפלגות רציפה).

אם $X \sim \mathcal{F}_\vartheta$, למעשה ידועה פונקציית הצפיפות של X, פרט לפרמטר ϑ.

$$f_X(t) = f(t;\vartheta)$$

הצפיפות המשותפת של משתנים מקריים בלתי תלויים $X_1, \cdots, X_n$:

$$f_{X_1,\cdots,X_n}(t;\vartheta) = \prod_{i=1}^{n} f_X(t_i;\vartheta)$$

מתבוננים בצפיפות של התוצאה אפוסטריורי (אחרי הניסוי):

$$L(X_1,\cdots,X_n;\vartheta) := \prod_{i=1}^{n} f_X(X_i;\vartheta)$$

רוצים למקסם את $L(X_1,\cdots,X_n;\vartheta)$.

במקום זה, נמקסם את: $\log L(X_1,\cdots,X_n;\vartheta)$.

$$\log L(X_1,\cdots,X_n;\vartheta) = \sum_{i=1}^{n} \log f_X(X_i;\vartheta)$$

ניתן לגזור לפי ϑ ולהשוות ל - 0.

$$\vartheta = \vartheta(X_1,\cdots,X_n)$$

ϑ זה נקרא אומד הנקראות המקסימלית.

דוגמה

$$X \sim N(\mu,1)$$

$$\Downarrow$$

$$f_X(t) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{(t-\mu)^2}{2}}$$

הצפיפות המשותפת של $X_1, \cdots, X_n$:

$$L(X_1, \cdots, X_n ; \mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{(X_i-\mu)^2}{2}}$$

נמקסם את $L(X_1, \cdots, X_n ; \mu)$.

במקום זה, נמקסם את $\log L(X_1, \cdots, X_n ; \mu)$.

$$\log L(X_1, \cdots, X_n ; \mu) = \sum_{i=1}^{n} \log\left(\frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{(X_i-\mu)^2}{2}}\right)$$

$$\Downarrow$$

$$\log L(X_1, \cdots, X_n ; \mu) = \sum_{i=1}^{n} \log\left(\frac{1}{\sqrt{2\pi}}\right) + \sum_{i=1}^{n} \log\left(e^{-\frac{(X_i-\mu)^2}{2}}\right)$$

$$\Downarrow$$

$$\log L(X_1, \cdots, X_n ; \mu) = n \cdot \log\left(\frac{1}{\sqrt{2\pi}}\right) - \sum_{i=1}^{n} \frac{(X_i-\mu)^2}{2}$$

נגזור לפי μ ונשווה ל – 0:

$$(\log L(X_1, \cdots, X_n ; \mu))' = 0$$

$$\Downarrow$$

$$\left(n \cdot \log\left(\frac{1}{\sqrt{2\pi}}\right) - \sum_{i=1}^{n} \frac{(X_i-\mu)^2}{2}\right)' = 0$$

$$\Downarrow$$

$$\sum_{i=1}^{n} \frac{-2 \cdot (X_i-\mu)}{2} = 0$$

$$\Downarrow$$

$$\sum_{i=1}^{n} X_i - \mu = 0$$

$$\Downarrow$$

$$\hat{\mu} = \widehat{X_n}$$

■

דוגמה

$$X \sim U[0, \vartheta]$$

$$\Downarrow$$

$$f_X(t\,; \vartheta) = \frac{1}{\vartheta} \cdot I_{[0,\vartheta]}(t)$$

עבור:

$$I_A(t) \coloneqq \begin{cases} 1, & t \in A \\ 0, & t \notin A \end{cases}$$

הצפיפות המשותפת של $X_1, \cdots, X_n$:

$$L(X_1, \cdots, X_n\,; \vartheta) = \prod_{i=1}^{n} \frac{1}{\vartheta} \cdot I_{[0,\vartheta]}(X_i)$$

$$\Downarrow$$

$$L(X_1, \cdots, X_n\,; \vartheta) = \frac{1}{\vartheta^n} \cdot \prod_{i=1}^{n} I_{[0,\vartheta]}(X_i)$$

$$\Downarrow$$

$$L(X_1, \cdots, X_n\,; \vartheta) = \frac{1}{\vartheta^n} \cdot I_{[0,\vartheta]}(\min X_i) \cdot I_{[0,\vartheta]}(\max X_i)$$

<u>הערה</u>: אם $\min X_i\,, \max X_i \in [0, \vartheta]$, אז: $X_i \in [0, \vartheta] \;:\; \forall 1 \le i \le n$.

נרצה למקסם את $L(X_1, \cdots, X_n\,; \vartheta)$.

אנו מחפשים את ה $-\vartheta$ הקטן ביותר עבורו הפונקציה אינה 0.

ה $-\vartheta$ הנ"ל היא ה - $\max X_i$ (הקטנה ביותר עבורה הביטוי הנ"ל אינו מתאפס).

לכן:

$$\hat{\vartheta} = \max X_i$$