סטטיסטיקה

שיעור שעבר

- תכונות אומדן:
  - הטיה.
  - שונות.
  - MSE
- אמידת שונות נורמלית.

היום:

- המשך תכונות:
  עקיבות.
- גישות לאמידה:
  נראות מקסימלית. (נזכיר: בייזיאני).

עקיבות

נזכר: יש לנו $\underline{X}=(X_1,\dots,X_n)\sim F_\theta$ מדגם מקרי. נניח יש לנו אומד מוגדר

ל-θ, נסמן $T(\theta)$ בצורה יותר מפורשת, נכתוב $T_n(\underline{X}_n)$, למשל:

$$T_n(\underline{X}_n)=\bar{X}_n=\frac{\sum_{i=1}^{n}X_i}{n}$$

חוסר הטיה: $E_\theta T_n(\underline{X}_n)=\theta$

עקיבות: $T_n(\underline{X}_n)\xrightarrow[n\to\infty]{P\text{ (בהסתברות)}}\theta$.

האומד הולך ומתקרב לערך האמיתי.

במילים: "בסופו של דבר" ($n\to\infty$) האומד שלנו יהיה "נכון" ($T\to\theta$), אם יהיו לנו מספיק נתונים.

הגדרה פורמלית של התכנסות בהסתברות:

$$P(|T_n(\underline{X}_n)-\theta|>\varepsilon)\xrightarrow[n\to\infty]{}0\quad\forall\varepsilon$$

$$\Updownarrow$$

$$T_n(\underline{X}_n)\xrightarrow[n\to\infty]{P}\theta$$

דוגמא

$X_i\sim N(\mu,\sigma^2)$, $T_n(X_n)=\bar{X}$ אומד ל-μ.

האם עקיב?

למשל, מחוק המספרים הגדולים: $\bar{X}_n\xrightarrow[n\to\infty]{P}EX=\mu$

באופן כללי, $\bar{X}$ תמיד עקיב לתוחלת של X (EX), ותמיד חסר הטיה ל-EX.

לפי אותו סימון, תקף גם לגבי $\hat{p}$ במקרה הבינומי.

דוגמא פחות טריוויאלית

נניח $X_i\sim U(0,\theta)$ וניקח $T_n(X_n)=\max\limits_{i\in\{1,\dots,n\}}X_i$

① האם $T_n(X_n)$ חסר הטיה?

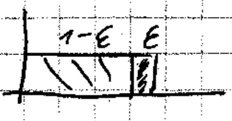
נשים לב כי $P(T_n(\underline{X}_n)\geq\theta)=0$ תמיד.

$P(T_n(X_n)<\theta(1-\varepsilon))=$ עבור $\varepsilon>0$ כלשהו.

קטן מ-θ

$$= P_r(X_1 \le (1-\varepsilon)\cdot\theta)\cdot P_r(X_2 \le (1-\varepsilon)\cdot\theta)\cdots P_r(X_n \le (1-\varepsilon)\theta) =$$
$$= (1-\varepsilon)^n$$

$$F_\theta((1-\varepsilon)\cdot\theta) = 1-\varepsilon$$



$$P(T_n \ge \theta) = 0 \ (\text{אפס})$$

$$P(T_n \le (1-\varepsilon)\cdot\theta) = (1-\varepsilon)^n$$

זהו חישוב כרגיל: $E(T_n(X_n)) = \int_0^\theta f_\theta(T_n(X_n))(x)\, x\, dx$

פונקצית ההתפלגות המצטברת של $T_n(X_n)$:

$$P_n(T_n(X_n) \le x) = \left(\frac{x}{\theta}\right)^n$$

צפיפות:

$$\frac{\partial\left(\frac{x}{n}\right)^n}{\partial x} = \frac{n\cdot x^{n-1}}{\theta^n} \longleftarrow \text{צפיפ}$$

$$E(T_n(X_n)) = \int_0^\theta \frac{n\cdot x^{n-1}}{\theta^n}\, x\, dx = \frac{n}{\theta^n}\cdot\int_0^\theta x^n\, dx = \frac{x^{n+1}}{n+1}\Big|_{x=0}^{\theta}\cdot\frac{n}{\theta^n} =$$
$$= \frac{n}{n+1}\cdot\frac{\theta^{n+1}}{\theta^n} = \frac{n}{n+1}\cdot\theta$$

← אומד מוטה, הטיה $\frac{1}{n+1}\cdot\theta$ $\left(\theta - \frac{n}{n+1}\theta =\right)$.

$$P(|\theta - T_n| > \varepsilon) = P(T_n < (1-\varepsilon)\cdot\theta) = (1-\varepsilon)^n \xrightarrow[n\to\infty]{} 0$$

← הוכחנו עקיבות המקסימום $T_n(X_n)$ עבור θ בהתפלגות אחידה.

<u>שאלות</u>

א. האם עקיבות גוררת ח"ה? לא בהכרח, ראינו דוגמא.

ב. האם חוסר הטיה גורר עקיבות?

דוגמא, בהתפלגות הנורמלית: $T_n(X_n) = X_1$

התוחלת של X_1 שווה ל- $\mu = E(X)$. $[E(X_1) = E(X) = \mu]$

← מסביר - השונות תהיה σ^2 כשגודל המדגם שואף ל-∞

<u>שיטות לבניית אומדים</u>

עד עכשיו לקחנו אומדים ש"המצאנו" כגון: $\bar{X}$, $\hat{p}$, אומד לפלס: $\hat{p}_{lap} = \frac{\sum x_i + 1}{n+2}$,

$\hat{p}_{conse} = 0.4$ ואז חקרנו את תכונותיהם.

השאלה היא, איך "מייצרים" אומדים בצורה מובנית.

<u>שלוש השיטות המרכזיות</u>

- נראות מקסימלית. (היום).
- שיטה בייזיאנית (סוף הקורס)
- שיטת המומנטים

פונקציית הנראות

נניח כרגיל: $X_1,\dots,X_n \sim P_\theta$ מדגם מקרי.

כמו כן נסמן את ההסתברות (במקרה הבדיד) / הצפיפות (במקרה הרציף) של המדגם:

הבדיד: $P_\theta(\underline{X}) = P_\theta(x_1)\,P_\theta(x_2)\cdots P_\theta(x_n)$

הרציף: $f_\theta(\underline{x}) = f_\theta(x_1)\cdot f_\theta(x_2)\cdots f_\theta(x_n)$

הרעיון המרכזי: נסתכל על $f_\theta(\underline{X})$ או $P_\theta(\underline{X})$ כפונקציה של θ במקום של $\underline{X}$, ונקרא לזה נראות.

$$\underline{\underline{L(\theta;\underline{x})}} = \begin{cases} P_\theta(\underline{X}) & \text{במקרה הבדיד} \\ f_\theta(\underline{X}) & \text{במקרה הרציף} \end{cases}$$

פונקציית נראות: likelihood

נשתמש הרבה בלוג הנראות:

$$\ell(\theta;\underline{x}) = \log L(\theta;\underline{x})$$

דוגמאות:

$\underline{X_i \sim Bernulli(p)}$ $\Leftarrow$

$$\underset{(P_p)}{P_\theta(x)} = \begin{cases} p & x=1 \\ 1-p & x=0 \end{cases}$$

$$L(p) = \prod_{i=1}^{n} p^{I(x_i=1)}\cdot(1-p)^{I(x_i=0)} = p^{\sum x_i}\cdot(1-p)^{n-\sum x_i}$$

כאשר I היא פונקציית אינדיקטור.

$$\ell(p) = \sum x_i \cdot \log p + (n-\sum x_i)\cdot\log(1-p)$$

$\underline{X_i \sim N(\mu,\sigma^2)}$ נניח σ^2 ידוע:

$$L(\mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}}\cdot e^{-\frac{(x_i-\mu)^2}{2\cdot\sigma^2}} =$$

מכפילים n פעמים ושורש נותן $\frac{n}{2}$.

$$= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left\{-\frac{1}{2\cdot\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$$

$$\ell(\mu) = -\frac{n}{2}\cdot\log(2\pi\sigma^2) - \frac{1}{2\cdot\sigma^2}\cdot\sum_{i=1}^{n}(x_i-\mu)^2$$

אומד נראות מקסימלי

$$\hat{\theta}_{MLE} = \underset{\theta\in\Theta}{\arg\max}\, L(\theta)$$

במילים: אומד הנראות המקסימלי הוא הערך של θ שמביא למקסימום את פונקציית (לוג) הנראות.

דוגמאות [illegible] MLE

ברנולי:

$$\ell(p;x) = \sum x_i\cdot\log p + (n-\sum x_i)\log(1-p)$$

מציאת המקסימום:

$$\frac{\partial \ell}{\partial p} = \frac{\sum x_i}{p} - \frac{n-\sum x_i}{1-p} = \frac{(1-p)\cdot\sum x_i - p(n-\sum x_i)}{p\cdot(1-p)} =$$

$$= \frac{\sum x_i - p\cdot\sum x_i - p\cdot n + p\cdot\sum x_i}{p\cdot(1-p)} = \frac{\sum x_i - p(x)}{p(1-p)}$$

נחשב נגזרת II כדי לוודא ש- $\frac{\partial l}{\partial p}=0$ הוא מקסימום:

$$\frac{\partial^2 l}{\partial p^2} = \frac{-np\cdot(1-p) - (1-2p)\cdot(\sum x_i - np)}{(p\cdot(1-p))^2} \overset{?}{<} 0$$

$$= \frac{-\cancel{np} + \cancel{np^2} - \sum x_i + 2p\cdot\sum x_i + \cancel{np} - 2\cdot\cancel{p^2 n}}{>0} = \frac{-\sum x_i + 2p\cdot\sum x_i - p^2 n}{>0}$$

ניתן להוכיח שהמונה <0.

$\Rightarrow$ הנקודה בה $f=0$ היא מקס' גלובלי.

לכן באומדן הנראות נגזרת ב- $\hat{p} = \frac{\sum x_i}{n} = \hat{P}_{MLE}$ $\Leftarrow$ אומד הנראות המקס' במקרה הבינומי.

נורמלי (השונות ידועה)

$$l(\mu;x) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\frac{\partial l}{\partial \mu} = \frac{1}{2\sigma^2}\cdot 2\sum_{i=1}^{n}(x_i-\mu) = \frac{1}{\sigma^2}\cdot(\sum x_i - n\cdot\mu)$$

$$\frac{\partial^2 l}{\partial \mu^2} = \frac{-n}{2\sigma^2} < 0$$

$$\Rightarrow \hat{\mu}_{MLE} = \frac{\sum x_i}{n} = \bar{x}$$

אחיד

$X_i \sim Unif(0,\theta)$ מחפשים אומד ל- θ.

$$f_\theta(x) = \begin{cases} \frac{1}{\theta} & 0\le x\le\theta \\ 0 & \text{אחרת} \end{cases}$$

או במילים אחרות: $\min(x_i)\ge 0$, $\max(x_i)\le\theta$

$$\Rightarrow L(\theta) = \begin{cases} (\frac{1}{\theta})^n & 0\le x_1,\dots,x_n\le\theta \\ 0 & \text{אחרת} \end{cases}$$

$\Leftarrow$ אומד ל- θ הוא הערך ה"תקין" המינימלי.

$$\hat{\theta}_{MLE} = \max_{i\in\{1,\dots,n\}}\{x_i\}$$

אומד לשונות במודל הנורמלי

$X\sim N(\mu,\sigma^2)$, σ^2 לא ידוע. הגדר הצגנו: $\hat{\sigma}^2 = \frac{1}{n-1}\sum(x_i-\bar{x})^2$

(אומד חסר הטיה באומד ל- σ^2).

כעת, אם נחפש לשאר $\hat{\mu}_{MLE}$, $\hat{\sigma}^2_{MLE}$ הם שונים. נקבל $\hat{\mu}_{MLE}=\bar{X}$

$$E(\hat{\mu}_{MLE}) = \frac{n-1}{n}\cdot\sigma^2 \quad , \quad \hat{\sigma}^2_{MLE} = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

תכונות אומד (הם תחת תנאי רגולריות קיים).

א. עקביות, אם $X\sim F_\theta$ אזי $\hat{\theta}_{MLE}\xrightarrow[n\to\infty]{P}\theta$

ב. מכוון תכונות אופטימליות אסימפטוטית.

במקרה [illegible] מתקיים: $\hat{\theta} \sim N(\theta, \frac{1}{[illegible]})$

כאשר $I(\theta)=E_\theta\left(\frac{\partial \log f(\theta)}{\partial\theta}\right)^2$

נכונה התכונה השניה

ג. אינווריאנטיות פונקציונלית

נניח שאנחנו יודעים את האמד $\hat{\theta}_{MLE}$ עבור פרמטר θ אז האמד של $g(\theta)$

(פונקציה חד חד ערכית) הוא: $(\widehat{g(\theta)})_{MLE} = g(\hat{\theta}_{MLE})$

למשל אם נניח במודל הנורמלי θ הוא μ אזי $\hat{\mu}_{MLE}=\bar{x}$

$(\hat{\mu}^2)_{MLE}=\bar{x}^2$

רווחי סמך

יש לנו התפלגות F_θ, נניח לפרמטר θ שנרצה לאמוד.

בהינתן אומד $T(X)$ ר"ס ברמת סמך $1-\alpha$, הוא מוגדר ע"י ערכים $\varepsilon_1,\varepsilon_2>0$

כך ש: $Pr_\theta(\theta\in[T-\varepsilon_1, T+\varepsilon_2])\ge 1-\alpha$

$\iff Pr_\theta(T(X)\in[\theta-\varepsilon_2,\theta+\varepsilon_1])\ge 1-\alpha$

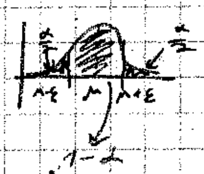
בד"כ ניקח $\alpha=0.05$ ונגיד שיש לנו רווח סמך ברמת סמך 95% $(1-\alpha)$

דוגמא

ר"ס (רווח סמך) לפרמטר הנורמלי, כאשר σ^2 ידוע, בהתחשב באומד $\bar{X}$ $T(X)=\bar{X}$

אנחנו יודעים: $\bar{X}\sim N(\mu,\frac{\sigma^2}{n})$

$P(\bar{X}\notin[\mu-\varepsilon,\mu+\varepsilon])=\alpha=$



אם נניח שבחרנו $\varepsilon_1=\varepsilon_2=\varepsilon$

וניקח $\alpha=0.05$.

אנחנו רוצים שההסתברות: $Pr(\bar{x}\in[\mu-\varepsilon,\mu+\varepsilon])=0.95$

$$p(\bar{X}\in[\mu-\varepsilon,\mu+\varepsilon])=Pr(\mu-\varepsilon\le\bar{X}\le\mu+\varepsilon)=Pr(\bar{X}\le\mu+\varepsilon)-Pr(\bar{X}\le\mu-\varepsilon)$$

$$=Pr\Big(\underbrace{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}_{N(0,1)}\le\frac{\varepsilon}{\sigma/\sqrt{n}}\Big)-Pr\Big(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\le\frac{-\varepsilon}{\sigma/\sqrt{n}}\Big)$$

$$=\Phi\Big(\frac{\varepsilon}{\sigma/\sqrt{n}}\Big)-\Phi\Big(-\frac{\varepsilon}{\sigma/\sqrt{n}}\Big)$$

מה אנחנו יודעים על $\Phi(-\frac{\varepsilon}{\sigma/\sqrt{n}})$?

$$\Phi\Big(\frac{-\varepsilon}{\sigma/\sqrt{n}}\Big)=1-\Phi\Big(\frac{\varepsilon}{\sigma/\sqrt{n}}\Big)$$

$$=2\Phi\Big(\frac{\varepsilon}{\sigma/\sqrt{n}}\Big)-1=0.95$$

אנחנו רוצים

$$\Rightarrow\Phi\Big(\frac{\varepsilon}{\sigma/\sqrt{n}}\Big)=0.975$$

$$\Rightarrow\frac{\varepsilon}{\sigma/\sqrt{n}}=Z_{0.975}=1.96\Rightarrow\varepsilon=\frac{1.96\cdot\sigma}{\sqrt{n}}$$

$\Leftarrow$ רווח סמך ב-95% עבור μ

$$\Big[\bar{x}-\frac{1.96\cdot\sigma}{\sqrt{n}},\ \bar{x}+\frac{1.96\cdot\sigma}{\sqrt{n}}\Big]$$

דוגמא: גבהי סטודנטים, נניח מתפלג נורמלית. נניח: $\sigma^2=16$, $n=25$, $\bar{X}=175$ ס"מ

רווח סמך לגובה ממוצע ברמת סמך 95%:

$$\mu \in \left[175-\frac{1.96\cdot 4}{5},\ 175+1.96\cdot\frac{4}{5}\right]=\underbrace{[173.4,\ 176.6]}_{I}$$

הגדלתי פי 100 את המדגם ⇐ הרווח יקטן פי 10. כי זה קטן פי

$\sqrt{}$ גודל המדגם.

בגלל שמספר של ממוצע הוא לא תלוי בנפרד, תלוי ב-$\sqrt{}$ (השורש) של המדגם.

<u>שאלה:</u> מה משמעות רווח הסמך?

א. האם הנ"ל אומר ש- $\mu \in I$.

ב. האם הנ"ל אומר שיש לנו בטחון של 95% ש- $\mu \in I$

ג. האם הנ"ל אומר שאם נמדוד מחדש בקבוצה של 25, הסיכוי $\bar{X} \in I$ 95%.

א לא נכון. אנחנו לא יודעים שרווח הסמך מכיל את הפרמטר.

ב. התשובה של רווח סמך.

ג. אם יודעים את המדגם הבא, הסיכוי לפול ברווח סמך הוא 95% הסיכוי יהיה קטן יותר.

$\bar{X}_1 \sim N(\mu,\frac{\sigma^2}{n})$, $\bar{X}_2 \sim N(\mu,\frac{\sigma^2}{n})$ ובת"ל

$$\bar{X}_1-\bar{X}_2 \underset{\text{בת"ל}}{\sim} N\left(0,\frac{2\cdot\sigma^2}{n}\right)$$

$$P(\bar{X}_2\in \bar{X}_1\text{ של סמך ר"ס}) = P\left(\bar{X}_2\in\left[\bar{X}_1-1.96\cdot\frac{\sigma}{\sqrt{n}},\ \bar{X}_1+1.96\cdot\frac{\sigma}{\sqrt{n}}\right]\right)=$$

$$=P_r\left(\frac{(\bar{X}_2-\bar{X}_1)\cdot\sqrt{n}}{\sqrt{2}\cdot\sigma}\in\left[-\frac{1.96}{\sqrt{2}},\ \frac{1.96}{\sqrt{2}}\right]\right)=2\cdot\underbrace{\Phi\left(\frac{1.96}{\sqrt{2}}\right)}_{0.916}-1\approx 0.83$$

$\frac{1.96}{\sqrt{2}}$ = 1.38

<u>רווח סמך לתוחלת נורמלית כש שונות לא ידועה</u>

אנחנו צריכים את השונות (או סטיית התקן) כדי לחשב ר"ס

פתרון מתבקש: נעריך את השונות ונציב בתוך הנוסחא, ז"א במקום σ^2

ניקח את: $\hat{\sigma}^2=\frac{1}{n-1}\cdot\sum(x_i-\bar{x})^2$

$\boxed{\hat{\sigma}\ \text{נקרא סטיית תקן}}$

האם נכון להגיד:

$$\frac{\bar{X}-\mu}{\hat{\sigma}/\sqrt{n}} \overset{??}{\sim} N(0,1)$$

הטענה שעדיין מתפלג נורמלי אינה נכונה. כי החלפנו סטיית תקן $\boxed{\text{בסטיית תקן}}$

<u>אבל</u> ל- $\hat{\sigma}$ יש התפלגות (מסובכת) אך ידועה.

$\frac{\bar{X}-\mu}{\hat{\sigma}/\sqrt{n}}$ לכן ניתן לחשב את ההתפלגות <u>הנכונה</u> של

ולהמציא לה שם: התפלגות t של סטודנט.

"סטודנט": william Gos[illegible] [illegible]

בגלל הצבא ומקום העבודה, הוא פרסם את המאמר בפסבדונים: "a student".

2. ההגדרה המפורשת: אם $X_i \sim N(\mu,\sigma^2)$ אזי $\frac{\bar{X}-\mu}{\hat{\sigma}/\sqrt{n}} \sim t_{n-1}$

באופן יותר מפורש: $\frac{(n-1)\cdot\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-1}$

ואנחנו יודעים: $\sim t_{n-1}$; במקרה שלנו: $\frac{\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}}{\left(\frac{\hat{\sigma}/\sqrt{n}}{\sigma/\sqrt{n}}\right)} \sim \frac{N(0,1)}{\sqrt{\frac{\chi^2_{n-1}}{n-1}}}$; $\frac{\mu-\bar{x}}{\sigma/\sqrt{n}}$ מספר נורמלי סטנדרטי

מה שחשוב:

⇐ רווח סמך כש-σ ידוע $(1-\alpha)$

$$\mu \in \left(\bar{x} - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}},\ \bar{x} + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right]$$

רווח סמך כש-σ לא ידוע:

$$\mu \in \left[\bar{x} - t_{n-1,1-\frac{\alpha}{2}}\frac{\hat{\sigma}}{\sqrt{n}},\ \bar{x} + t_{n-1,1-\frac{\alpha}{2}}\frac{\hat{\sigma}}{\sqrt{n}}\right]$$

<u>דוגמא: גבהי סטודנטים</u>

לפי Z, אם $\sigma^2 = 16$ (ידוע) קיבלנו $[173.4, 176.5]$ (ברמת סמך של $z_{0.975}=1.96$)

נדרשים אותם נתונים אך $t_{24,0.975} = 2.064$

⇐ נניח שחישבנו $\hat{\sigma}^2 = 16$, אזי רווח הסמך של סמך t: $[173.3, 176.7]$

שיעור הבא:

לא שיעור בחנוכה.