שיעור 11 | מקדם מתאם, חסמים על הסתברויות | 7/1/18

- **שאלה על מקדם מתאם**

מבצעים 3 הטלות ב"ת של מטבע הוגן. יהי X – מספר תוצאות העץ ב-2 ההטלות הראשונות. יהי Y – מספר תוצאות העץ בשתי ההטלות האחרונות. מהו $r(X,Y)$?

- פתרון:

נצפה למקדם מתאם חיובי. יותר הצלחות בשתי הראשונות מרמז על הצלחה בשנייה ולכן מגדיל את התחזית בשתי האחרונות שהשנייה היא אחת מהן. אבל אי אפשר לבטא את Y כפונקציה לינארית של X ולכן אין מתאם 1 (כלומר אין קשר מוחלט). ולכן נצפה ש-$0 < r < 1$.

$$X = X_1 + X_2$$
$$Y = X_2 + X_3$$

כאשר X_i הם אינדיקטורים להטלות השונות.

$$r(X_1+X_2, X_2+X_3) = \frac{Cov(X_1+X_2, X_2+X_3)}{Var(\sqrt{V(X_1+X_2)V(X_2+X+3\,)}}$$

נחשב את המונה והמכנה בנפרד:

$$V(X_1+X_2) = V(X_2+X_3)$$
$$(X_1+X_2) \sim Bin\left(2, \frac{1}{2}\right) \Rightarrow Var(\sqrt{V(X_1+X_2)V(X_2+X+3\,)} = \frac{1}{2}$$
$$Cov(X_1+X_2, X_2+X_3) = Cov(X_1+X_2) + Cov(X_1+X_3) + Cov(X_2+X_2) + Cov(X_2, X_3)$$
$$= V(X_2) = \frac{1}{2}\cdot\frac{1}{2} = \frac{1}{4}$$

לכן:

$$r(X_1+X_2, X_2+X_3) = \frac{\frac{1}{4}}{\frac{1}{2}} = \frac{1}{2}$$

- **שאלה**

מבצעים 10 הטלות בלתי תלויות של קובייה תקינה. X_i – מספר הפעמים שקיבלנו i. חשבו $r(X_1, X_6)$.

- פתרון:

נצפה למקדם מתאם שלילי אבל לא מאוד שלילי.

<u>דרך א'</u>:

$$r(X_1, X_6) = \frac{Cov(X_1, X_6)}{\sqrt{V(X_1)V(X_6)}} = \frac{Cov(X_1, X_6)}{V(X_6)}$$
$$Cov(X_1+X_2+X_3+X_4+X_5+X_6, X_6) = Cov(10, X_6) = 0$$

*קווריאנס בין קבוע למשתנה שווה ל-0

$$Cov(X_1+X_2+X_3+X_4+X_5+X_6,X_6)=Cov(X_1,X_6)+Cov(X_2,X_6)+\cdots+Cov(X_6,X_6)$$
$$=5\cdot Cov(X_1,X_6)+Cov(X_6,X_6)=5\cdot Cov(X_1,X_6)+V(X_6)$$

אבל זה גם שווה ל-0 ולכן:

$$Cov(X_1,X_6)=-\frac{1}{5}V(X_6)$$

ולכן:

$$r(X_1,X_6)=-\frac{\frac{1}{5}V(X_6)}{V(X_6)}=-\frac{1}{5}$$

<u>דרך ב׳</u>:

$$r(X_1,X_6)=\frac{Cov(X_1,X_6)}{\sqrt{V(X_1)V(X_6)}}$$

$$X_1\sim Bin\left(10,\frac{1}{6}\right),\qquad X_6\sim Bin\left(10,\frac{1}{6}\right),\qquad X_1+X_6\sim Bin\left(10,\frac{2}{6}\right)$$

$$V(X_1)=V(X_6)=10\cdot\frac{1}{6}\cdot\frac{5}{6}$$

$$V(X_1+V_6)=10\cdot\frac{2}{6}\cdot\frac{4}{6}=\frac{80}{36}$$

$$V(X_1+X_6)=V(X_1)+V(X_6)+2Cov(X_1,X_6)$$

$$Cov(X_1,X_6)=\frac{V(X_1+X_6)-V(X_1)-V(X_6)}{2}=\frac{\frac{80}{36}-\frac{50}{36}-\frac{50}{6}}{2}=-\frac{10}{36}$$

ולכן:

$$r(X_1,X_6)=\frac{-\frac{10}{36}}{\frac{50}{36}}=-\frac{1}{5}$$

<u>דרך ג׳</u>:

יהי $X_{1,j}$ אינדיקטור לקבלת התוצאה 1 בהטלה ה-j.

יהי $X_{6,k}$ אינדיקטור לקבלת התוצאה 6 בהטלה ה-k.

מחפשים $Cov\left(\Sigma X_{1,j},\Sigma X_{6,k}\right)$.

$$Cov\left(\Sigma X_{1,j},\Sigma X_{6,k}\right)=\Sigma_j\Sigma_k Cov\left(X_{1,j},X_{6,k}\right)$$

יש תלות רק כאשר $j=k$. לכן הסכום שווה ל: $\Sigma_j Cov\left(X_{1,j},X_{6,j}\right)$.

למה שווה אינדיקטור Cov בודד?

$$Cov\left(X_{1,j},X_{6,j}\right)=E\left(X_{1,j}\cdot X_{6,j}\right)-E\left(X_{1,j}\right)\cdot E\left(X_{6,j}\right)=0-\frac{1}{6}\cdot\frac{1}{6}=-\frac{1}{36}$$

ומכאן:

$$\Sigma Cov\left(X_{1,j},X_{6,j}\right)=-\frac{10}{36}$$

$$\sqrt{V(X_1)V(X_6)} = V(X_1) = 10 \cdot \frac{1}{6} \cdot \frac{5}{6} = \frac{50}{36}$$

ולכן:

$$r(X_1, X_6) = \frac{-\frac{10}{36}}{\frac{50}{36}} = -\frac{1}{5}$$

חסמים על הסתברויות

אי שוויון מרקוב:

יהי X משתנה מקרי שמקיים $P(X \geq 0) = 1$ ובעל תוחלת סופית $E(X)$, אזי מתקיים עבור כל $a > 0$ ש:

$P(X \geq a) \leq \frac{E(X)}{a}$. זה לא נותן הסתברות מדויקת, אבל זה נותן חסם.

- **דוגמא:**

נניח ש: $P(X \geq 0) = 1$ ו-$E(X) = 2$. אז:

$$P(X \geq 8) \leq \frac{E(X)}{8} = \frac{1}{4}$$

הוכחת אי השוויון:

יהי X משתנה מקרי שמקיים $P(X \geq 0) = 1$ ובעל תוחלת סופית $E(X)$, צריך להוכיח כי מתקיים עבור כל $a > 0$ ש:

$P(X \geq a) \leq \frac{E(X)}{a}$.

$$E(X) = \Sigma_k P(X = k) \cdot k =$$

$$\Sigma_{k<a} P(X = k) \cdot k + \Sigma_{k \geq a} P(X = k) \cdot k \geq$$

$$\Sigma_{0<k<a} P(X = k) \cdot 0 + \Sigma_{k \geq a} P(X = k) \cdot a =$$

$$0 + a \cdot \Sigma_{k \geq a} P(X = k) =$$

$$a \cdot P(X \geq a)$$

קיבלנו ש: $E(X) \geq a \cdot P(X \geq a)$ כלומר: $P(X \geq a) \leq \frac{E(X)}{a}$

- **דוגמא**

יש בניין שבו 500 מדרגות. אני מטיל קובייה 100 פעמים. בכל פעם אני מתקדם לפי תוצאת הקובייה. מצאו חסם עליון שאגיע לפסגה לאחר לכל היותר 100 הטלות.

- פתרון:

X – סכום 100 ההטלות

$$E(X) = 100 \cdot 3.5 = 350$$

כדי להגיע לפסגה צריך שיתקיים $X \geq 500$.

$$P(X \geq 500) \leq \frac{E(X)}{500} = \frac{350}{500} = 0.7$$

אי שוויון צ'בישב

יהי X משתנה מקרי בעל תוחלת μ ושונות σ^2. אז עבור כל $a > 0$ מתקיים:

$$P(|X-\mu| \geq a) \leq \frac{\sigma^2}{a^2} = \frac{V(X)}{a^2}$$

הוכחת אי שוויון צ'בישב:

$$P(|X-\mu| \geq a) = P((|X-\mu|)^2 \geq a^2)$$

$(X-\mu)^2$ זה משתנה אי שלילי. לכן ניתן להפעיל את אי שוויון מרקוב:

$$P((X-\mu)^2 \geq a^2) \leq \frac{E((X-\mu)^2)}{a^2} = \frac{V(X)}{a^2}$$

❖ **דוגמא**

$X \sim Bin\left(100, \frac{1}{2}\right)$. מצאו חסם עליון על $P(X \geq 60)$.

➢ פתרון:

$$P(X \geq 60) \leq P(|X-50| \geq 10)$$

50 זו התוחלת של X.

$$P(|X-50| \geq 10) \leq \frac{V(X)}{10} = \frac{100 \cdot \frac{1}{2} \cdot \frac{1}{2}}{10^2} = \frac{1}{4}$$

האם ניתן למצוא חסם יותר טוב?

אפשר כאן להוסיף משהו מעבר:

$$P(|X-50| \geq 10) = P(X \leq 40) + P(X \geq 60)$$

לכן נוכל לקבל חסם: $P(X \geq 60) \leq \frac{1}{2} \cdot \frac{1}{4}$

*הערה:

אם למשל $X \sim Bin\left(100, \frac{1}{4}\right)$ אז אין סימטריה סביב 25. למשל סטייה של 30 ימינה אפשרית אבל סטייה של 30 שמאל לא אפשרית.

תרגיל מסכם לאי-שוויונות

❖ בעיית המזכירה הרשלנית:

מזכירה שולחת ל-n אנשים n מכתבים. X – מספר האנשים שקיבלנו מכתב נכון.

מצאו חסמים עליונים ל-$P(X \geq n-1)$, $P(X \geq 2)$, $P(X \geq 3)$.

➢ פתרון:

$$P(X \geq 2) \leq \frac{E(X)}{2} = \frac{1}{2} \text{ (לפי מרקוב)}$$

$$P(X \geq 2) \leq \frac{V(X)}{(2-1)^2} = \frac{1}{1} = 1 \text{ (לפי צבישב)}$$

צ'בישב לא נותן שום אינפורמציה.

$$P(X \geq 3) \leq \frac{E(X)}{3} = \frac{1}{3} \text{ (לפי מרקוב)}$$

$$P(X \geq 3) \leq \frac{V(X)}{(3-1)^2} = \frac{1}{4} \text{ (לפי צבישב)}$$

$$P(X \geq n-1) \leq \frac{E(X)}{n-1} = \frac{1}{n-1} \text{ (לפי מרקוב)}$$

$$P(X \geq n-1) \leq \frac{V(X)}{\left((n-1)-1\right)^2} = \frac{1}{(n-2)^2} \text{ (לפי צבישב)}$$

$$P((X-\mu)^2 \leq a) \leq \frac{V(X)}{a^2}$$

באופן מדויק:

$$P(X \geq n-1) = P(X = n-1) + P(X = n) = 0 + P(X = n) = \frac{1}{n!}$$

משתנים מקריים רציפים

צריך להתמודד עם מצב שכל נקודה מתקבלת בהסתברות 0 אבל בסופו של דבר איזושהי נקודה מתקבלת. לכל קטע יש הסתברות. אנו נוכל לחשב את ההסתברות לקבל קטע מסוים. עבור על משתנה מקרי לכל קטע יש הסתברות כלשהי של 0 או חיובית.

נשתמש בפונקציית ההסתברות המצטברת של משתנה X.

$F_X(x)$ היא ההסתברות שהמשתנה יקבל כרך קטן או שווה ל-x: $F_X(x) = P(X \leq x)$

אם ידוע לנו $F_X(x)$ לכל נקודה בישר אז ניתן לחשב את ההסתברות של כל קטע.

$$P(a < X \le b) = F_X(b) - F_X(a)$$

$$F_X(b) = P(X \le b),\ \ F_X(a) = P(X \le a)$$

ומתקיים: $P(X \le b) - P(X \le a) = P(a \le X \le b)$

אם עבור כל נקודה נדע מה ההסתברות המצטברת אז נדע לכל קטע מה ההסתברות לקבל אותו.

- **דוגמא:**

נמצא את פונקציית ההסתברות של אינדיקטור בעל הסתברות חצי:

$$F_X(x) = \begin{cases} 0 & X < 0 \\ \frac{1}{2} & 0 \le X < 1 \\ 1 & X \ge 1 \end{cases}$$

$$F_X(x) = P(X \le x), \qquad P(X \le 2) = 1$$

- **דוגמא אחרת של משתנה רציף:**

משתנה $X \sim U(0,1)$ – שימו לב שכאן הסוגריים עגולות, זה מסמן משתה אחיד רציף.

$$F_X(x) = \begin{cases} 0 & X < a \\ \frac{x-a}{b-a} & a \le X < b \\ 1 & X \ge b \end{cases}$$

זה משתנה שמקבל ערכים רק בין a ל-b. הוא מקבל ערכים בקטע מסוים בהסתברות ששווה לאורכו של הקטע הזה חלקי ארוך כל הקטע (a,b). זהו משתנה שפונקציית ההסתברות המצטברת שלו עולה באופן רציף.

לכל משתנה מתקיים $\lim_{X\to\infty^-} F_X(x) = 0$, $\lim_{X\to\infty^+} F_X(x) = 1$

הגדרה: אם $F_X(x)$ היא פונקציה רציפה אז נגיד שהמשתנה הוא משתנה רציף.

הגדרה: אם עבור כל $-\infty < x < \infty$ קיימת פונקציה $f_X(x)$ המקיימת $F_X(x) = \int_{-\infty}^{\infty} f_X(x)$ אז נגיד שהמשתנה X רציף בהחלט והפונקציה $f_X(X)$ היא פונקציית הצפיפות של המשתנה. באופן טבעי אם $F_X(x)$ גזירה אז הנגזרת שלה מועמדת להיות פונקציית הצפיפות.

למשל ל**משתנה אחיד** $\boldsymbol{X \sim U(a,b)}$ יש פונקציית הסתברות מצטברת:

$$F_X(x) = \begin{cases} 0 & X < a \\ \frac{x-a}{b-a} & a \le X < b \\ 1 & X \ge b \end{cases}$$

פונקציית צפיפות היא:

$$f_X(x)=\begin{cases}0 & X<a\\ \dfrac{1}{b-a} & a\le X<b\\ 0 & X\ge b\end{cases}$$

נכיר משפחה נוספת של התפלגויות:

משתנה אקספוננסיאלי/מעריכי $\boldsymbol{X\sim \exp(\lambda)}$ כאשר $\lambda>0$ הוא הפרמטר היחיד:

$$F_X(x)=\begin{cases}0 & X<0\\ 1-e^{-\lambda x} & 0\le X\end{cases}$$

פונקציית צפיפות היא:

$$f_X(x)=\begin{cases}0 & X<0\\ \lambda e^{-\lambda x} & 0\le X\end{cases}$$

❖ **שאלה**

נתון משתנה $X\sim\exp(2)$.

א. מהו $P(1\le X\le 2)$?

ב. מהו $P(-3\le X\le -2)$?

ג. מהו $P(-1\le X\le 3)$?

➢ פתרון:

א. $P(1\le X\le 2)=P(X\le 2)-P(X\le 1)=(1-e^{-2\cdot 2})-(1-e^{-2\cdot 1})=e^{-2}-e^{-4}$

ב. $P(-3\le X\le -2)=F_X(-2)-F_X(-3)=0-0=0$

ג. $P(-1\le X\le 3)=F_X(3)-F_X(-1)=F_X(3)=1-e^{-2\cdot 3}$

נגדיר משפחת התפלגויות נוספת:

משתנה מתפלג נורמאלית $\boldsymbol{X\sim N(\mu,\sigma^2)}$**:**

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

זו פונקציית צפיפות של משתנה שי לו שני פרמטרים μ ו-σ^2. זהו משתנה רציף כי הצגנו פונקציית צפיפות.

נכיר מקרה פרטי של משפחה זו: $X\sim N(0,1)$, **זהו משתנה נורמלי סטנדרטי**:

כאן הפרמטרים הם 0 ו-1 ופוקנציית הצפיפות היא:

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

יש לו פונקציית הסתברות מצטברת של: $F_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}} dt$

עבור משתנה נורמלי סטנדרטי מתקיים: $F_X(0) = \frac{1}{2}$. פונקציית ההסתברות המצטברת של משתנה זה נקרה $\phi(t)$. היא ניתנת בטבלה ידועה שמופיעה בהרבה מקומות.

בטבלה מופיעים ההסתברויות המצטברות לכל $X \geq 0$.

$$P(X \leq x) = \phi(x)$$

הודות לסימטריה מתקיים $P(X \leq -x) = P(X \geq x) = 1 - P(X \leq x) = 1 - \phi(x)$

- **דוגמא**

  $X \sim N(0,1)$. מהו $P(-2 < X < 3)$?

  - פתרון:

$$P(-2 < X < 3) = P(X < 3) - P(X < -2) = \phi(3) - [1 - \phi(2)] = \phi(3) + \phi(2) - 1$$