(00094412) הסתברות מ' | תרגול 6

שם: איל

February 22, 2024

נושא התרגול: נוסחת הזנב, אינדיקטורים, שונות

נושא ראשון - נוסחת הזנב ואינדיקטורים

- לפעמים יהיה קשה לחשב את פונקציית ההסתברות של משתנה מקרי מסוים, ואנחנו נצטרך להסתפק בתכונות של המשתנה המקרי, לדוגמה תוחלת.
  - היום נכיר שתי דרכים נוספות לחשב את התוחלת, בלי לחשב באופן ישיר את פונקציית ההסתברות:
    1. נוסחת הזנב
    2. שיטת האינדיקטורים
- **נוסחת הזנב:**
  - עבור משתנה מקרי X המקבל ערכים שלמים אי שליליים $\sum_{x=1}^{\infty} P\left(X \geq x\right)$
- **שיטת האינדיקטורים:**
  - נסמן אינדיקטור של מאורע A להיות $I_A = \begin{cases} 1 & A_{happened} \\ 0 & A_{didn't\ \ happen} \end{cases}$
    * נשים לב כי אינדיקטור הוא משתנה ברנולי, כלומר $I_A \sim Ber\left(p\left(A\right)\right)$
      · מתקיים כי תוחלת של אינדיקטור שווה לפונקציית ההסתברות של המאורע.
  - כעת נוכל לחשב את התוחלת של האינדיקטור:

$$E\left[I_A\right] = 1 \cdot P\left(I_A = 1\right) + 0 \cdot P\left(I_A = 0\right) = P\left(A\right)$$

    * ואז נוכל לקבל:

$$E\left[X\right] = E\left[\sum_{i=1}^{n} I_{A_i}\right]$$

- ומלינאריות התוחלת מתקיים:

$$= E[X] = \sum_{i=1}^{n} E[I_{A_i}] = \sum_{i=1}^{n} P(A_i)$$

תרגיל 1.

מטילים קובייה הוגנת 10 פעמים באופן בלתי תלוי.

חשב בעזרת נוסחת הזנב את התוחלת של המ"מ הבאים:

א. סכום שתי התוצאות הגבוהות ביותר ב-3 ההטלות הראשונות.

ב. תוצאה מקסימלית ב-5 הטלות ראשונות.

תזכורת: בתרגול 5 פתרנו את הסעיפים הנ"ל ללא שימוש בנוסחת הזנב

פיתרון 1.א.

- נסמן ב-M את התוצאה המינימלית.
- נשתמש בנוסחת הזנב:

$$E[M] = \sum_{m=1}^{\infty} P(M \geq m)$$

$$P(M \geq m) = P(X_1 \geq m_1, X_2 \geq m_2, X_3 \geq m_3)$$

- נתון כי המשתנים בלתי תלויים ולכן מתקיים:

$$= P(X_1 \geq m_1) P(X_2 \geq m_2) P(X_3 \geq m_3)$$

$$= \left(\frac{7-m}{6}\right)^3$$

- נבחן את התוחלת ונראה שכאשר $m > 6$ אז ההסתברות $P(M \geq m) = 0$ ולכן:

$$E[M] = \sum_{m=1}^{\infty} P(M \geq m) = \sum_{m=1}^{6} P(M \geq m)$$

$$= \sum_{m=1}^{6} \left(\frac{7-m}{6}\right)^3 = 2.042$$

פיתרון 1.ב.

- נסמן ב-Y את המשתנה המקרי
- נמצא את התוחלת של Y לפי נוסחת הזנב:

$$E[Y]=\sum_{y=1}^{\infty}P(Y\geq y)$$

– מתקיים:

$$P(Y\geq y)=1-P(Y<y)$$

* נבחן את $P(Y<y)$

$$P(Y<y)=P(Y\leq y-1)$$

$$=P(X_1\leq y-1)^5$$

$$=\left(\frac{y-1}{6}\right)^5$$

* נציב ונקבל:

$$\begin{aligned}E[Y]&=\sum_{y=1}^{\infty}P(Y\geq y)=\sum_{y=1}^{6}\left(1-\left(\frac{y-1}{6}\right)^5\right)\\&=\sum_{y=1}^{6}\left(1-\left(\frac{y-1}{6}\right)^5\right)\end{aligned}$$

תרגיל 2. שאלה קשה.
בסדרת ביטים אקראית באורך n, כל ביט הוא 1 בהסתברות p, בלי תלות באחרים.
נסמן ב- X את מספר הרצפים (כלומר, מספר הקטעים בעלי אותו ערך).
חשבו את:
א. $E[X]$
ב. $E\left[X^2\right]$ עבודה עצמית - (סעיף לא קל, אבל הרבה פעמים ידרשו מאיתנו לחשב את התוחלת של X^2. צריך להעלות בריבוע את $X=1+\sum_{i=1}^{n-1}U_i$ ולחשב תוחלת לכל המחוברים)

2. א. פיתרון:

- נבדוק את שלושת הדרכים שיש לנו לחשב תוחלת:

- לפי הגדרה נצטרך למצוא את $P_X(x)$ וזה לא נשמע קל.
- גם לפי נוסחת הזנב (שהיא לרוב קשורה למינימום או מקסימום) לא נצליח בקלות.
- לכן ננסה לפי שיטת האינדיקטורים.

- נגדיר אינדיקטורים כך ש-X יהיה שווה לסכום האינדיקטורים U_i:
  - U_i = אינדיקטור של המאורע בו ביט i שונה ביט $i+1$.
    - כעת מתקיים:

$$X = 1 + \sum_{i=1}^{n-1} U_i$$

- נחשב את התוחלת של X:

$$E[X] = E\left[1 + \sum_{i=1}^{n-1} U_i\right]$$

$$= E[1] + E\left[\sum_{i=1}^{n-1} U_i\right]$$

$$\Rightarrow E[X] = 1 + \sum_{i=1}^{n-1} E[U_i]$$

  - התוחלת של כל U_i שווה להסתברות שביט i שונה מביט $i+1$. כלומר התוחלת של האינדיקטור שווה להסתברות שהמאורע קרה.
    - כלומר, מתקיים:

$$E[U_i] = \overbrace{p}^{first\ \ bit\ \ is\ \ 1} \cdot (1-p) + (1-p) \cdot p$$

$$= 2 \cdot p \cdot (1-p)$$

  - קיבלנו:

$$E[X] = 1 + \sum_{i=1}^{n-1} 2 \cdot p \cdot (1-p)$$

הסבר על שאלת העמקה - 2. ב.

- **צ"ל:** $E\left[X^2\right]$
- כאמור, השתמשנו בסימון U_i = האינדיקטור של המאורע "ביט i שונה מביט $i+1$"
- קיבלנו $X=1+\sum_{i=1}^{n-1}U_i$ וגם $E\left[X\right]=1+\left(n-1\right)\cdot2p\cdot\left(1-p\right)$
- נפתח את X^2:

$$E\left[X^2\right]=E\left[\left(1+\sum_{i=1}^{n-1}U_i\right)^2\right]$$

– נשים לב שצריך להעלות סכום בריבוע, שזה אומר שנקבל סכום כפול.

$$=E\left[1+2\cdot\sum_{i=1}^{n-1}U_i+\sum_{i=1}^{n-1}\sum_{j=1}^{n-1}U_iU_j\right]$$

$$=E\left[1\right]+\overbrace{2\cdot E\left[\sum_{i=1}^{n-1}U_i\right]}^{=2\cdot(n-1)\cdot2\cdot p\cdot(1-p)}+E\left[\sum_{i=1}^{n-1}\sum_{j=1}^{n-1}U_iU_j\right]$$

– כעת נחשב את $E\left[\sum_{i=1}^{n-1}\sum_{j=1}^{n-1}U_iU_j\right]$ על ידי פיצול הסכום לשני מקרים:

$$E\left[\sum_{i=1}^{n-1}\sum_{j=1}^{n-1}U_iU_j\right]=\overbrace{2}^{symetry\ \ of\ \ j>i}\cdot\sum_{i<j}E\left[U_iU_j\right]+\sum_{i=j}^{n-1}E\left[U_iU_i\right]$$

* לאינדיקטור בריבוע יש את אותה פונקציית הסתברות כמו האינדיקטור עצמו, ולכן נקבל:

$$\sum_{i=j}^{n-1}E\left[U_iU_i\right]=$$

$$=\sum_{i=i}^{n-1}\left(2p\cdot\left(1-p\right)\right)$$

$$=\left(n-1\right)\left(2p\cdot\left(1-p\right)\right)$$

– נחשב את $\sum_{i<j}E\left[U_iU_j\right]$:

* נשים לב כי אנו צריכים להפריד בין מקרים של ביטים עוקבים וביטים לא עוקבים:

- אם i ו-j לא עוקבים אז U_i ו-U_j בלתי תלויים, ולכן לפי כלל הכפל:

$$\sum_{i<j} E\left[U_iU_j\right] = E\left(U_i\right) \cdot E\left(U_j\right) = \left(2p \cdot (1-p)\right)^2$$

- אם i ו- j עוקבים, אז יש שני תתי סדרות כאלה בלבד ואנו מחפשים את ההסתברות של אחת מהם $P\left(\{010\} \cup \{101\}\right)$:

$$P\left(\{010\} \cup \{101\}\right) = p \cdot (1-p) \cdot p + (1-p) \cdot p \cdot (1-p)$$

$$= p \cdot (1-p)$$

* כעת עלינו להבין כמה איברים יש שמקיימים ש-i, j עוקבים:
  - כמות הביטים i, j השונים זה מזה היא $\binom{n-1}{2}$
  - מתוכם יש $n-2$ מקרים בהם i עוקב ל-j.
  - התשובה יוצאת קצת מכוערת ולכן לא כתבנו אותו בתרגול.

- העיקרון בשאלה הזו הוא הפיצול לסכומים שונים.

נושא שני - שונות וסטיית תקן

- נניח שיש לנו שתי חנויות:
  - חנות א' שהממוצע לפריט בה הוא 100 שקל אבל המחירים בה נעים בין 50-150 שקל.
  - חנות ב' שבה הממוצע לפריט הוא 110 שקל אבל המחירים נעים בין 100 ובין 120 שקל.
- נוכל לחשב את ה**שונות** או **סטיית תקן** על מנת להעריך באיזו חנות לקנות.
- תכונות של שונות:
  - היא אי שלילית
  - היא שווה אפס אם ורק אם X קבוע (כלומר אם המשתנה המקרי מקבל ערך כלשהו בהסתברות 1)
  - היא **לא** פעולה לינארית, מתקיים

$$Var\left(a \cdot X + b\right) = a^2 Var\left(X\right)$$

$$SD\left(a \cdot X + b\right) = |a|\, SD\left(X\right)$$

  - בניגוד לתוחלת, סכום השונות שווה לשונות של סכום המשתנים המקריים **רק אם** המשתנים המקריים בלתי תלויים.

תרגיל 3.

לפנינו 20 מעטפות.באחת מהן 100 והשאר ריקות.

באפשרותנו לבחור מבין 2 המשחקים הבאים:

משחק א': נפתח את המעטפות ללא החזרה, כאשר עבור כל בדיקה אנו משלמים 8, עד שנגיע למעטפה המיוחלת ונזכה ב-100

משחק ב': דומה למשחק א', רק שלאחר כל בדיקה מחזירים את המעטפה. לכן, משלמים עבור בדיקה רק 3.

בשאלה זו נדון איזה משחק יותר כדאי.

א. באיזה משחק תוחלת הרווח הסופי גבוהה יותר?

ב. באיזה משחק השונות של הרווח הסופי גבוהה יותר?

פיתרון 3. א.

- נסמן:
  - Y_1 = הרווח הסופי של משחק א'
  - Y_2 = הרווח הסופי של משחק ב'.
- נמצא את $E\left[Y_1\right]$ ואת $E\left[Y_2\right]$ ונבדוק מי גדול יותר.
- נשים לב שהתוחלת תלויה בכמה מעטפות הוצאנו.
  - לכן נסמן:
    - X_1 = מספר המעטפות שנפתחו במשחק א'
      - מתקיים:

$$Y_1 = 100 - 8 \cdot X_1$$

    - X_2 = מספר המעטפות שנפתחו במשחק ב'
      - מתקיים:

$$Y_2 = 100 - 3 \cdot X_2$$

- נשים לב כי $X_2 \sim Geo\left(\frac{1}{20}\right)$, כי אנחנו מחזירים בכל שליפה את המעטפה ולכן השליפות בלתי תלויות.
- נראה כי $X_1 \sim U\left(\{1,2,\ldots,20\}\right)$, כלומר מתפלג באופן אחיד על $\{1,2,\ldots,20\}$:
  - מתקיים

$$P\left(X_1 = 1\right) = \frac{1}{20}$$

$$P\left(X_2 = 2\right) = \frac{19}{20} \cdot \frac{1}{19} = \frac{1}{20}$$

$$P(X_3=3)=\frac{19}{20}\cdot\frac{18}{19}\cdot\frac{1}{18}=\frac{1}{20}$$

* ניתן להמשיך כך עד 20 (או להראות באינדוקציה)
* כלומר המיקום של המעטפה עם ה-100$ מתפלג באופן אחיד.

- נחשב את התוחלת של X_1, שמתפלג באופן אחיד (התוחלת שלו היא אמצע הקטע בין 1 ל-20):

$$E[X_1]=\frac{1+20}{2}=\frac{21}{2}$$

– ולכן התוחלת של Y_1 היא:

$$E[Y_1]=100-8\cdot E[X_1]=16$$

- כעת נחשב את התוחלת של X_2 שמתפלג באופן גאומטרי:

$$E[X_2]=\frac{1}{p}=20$$

– ולכן התוחלת של Y_2 היא:

$$E[Y_2]=100-3\cdot E[X_2]=40$$

- כלומר מתקיים $E[Y_2]>E[Y_1]$

פיתרון סעיף ב':

- הזזה בקבוע לא משנה את ה-Var.
- על מנת לחשב את $Var(Y_1)$, נחשב קודם את $Var(X_1)$

$$Var(X)=E\left[X_1^2\right]-(E[X_1])^2$$

– ומתקיים:

$$E\left[X_1^2\right]=\sum_{x=1}^{20}x^2\cdot\overbrace{P(X_1=x)}^{=\frac{1}{20}}$$

$$= \sum_{x=1}^{20} x^2 \cdot \frac{1}{20}$$

$$= \frac{1}{20} \cdot \sum_{x=1}^{20} x^2$$

* ולכן $Var\left(X_1\right) = 33.25$

- ומתכונות התוחלת מתקיים:

$$Var\left(Y_1\right) = 64 \cdot Var\left(X_1\right) = 2128$$

– נשים לב כי $Var\left(Y_1\right)$ הוא ביחידות של דולר בריבוע.

- נחשב את את $Var\left(Y_2\right)$:

$$Var\left(Y_2\right) = 9 \cdot Var\left(X_2\right) = 3420$$

- בסעיף הזה נקבל ש$Var\left(Y_2\right) > Var\left(Y_1\right)$.

– כלומר התוצאות במשחק השני רחוקות יותר מהתוחלת שלהן מאשר במשחק הראשון.

* אם נתייחס רק לVar אז לפי הסעיף הזה המשחק הראשון עדיף.

תרגיל 4.

בקופסה כדור אחד שחור ושניים לבנים.

מוציאים כדור באקראי ומחזירים אותו עם כדור נוסף מאותו צבע.

חוזרים על פעולה זו פעמיים.

מצאו התפלגות וחשבו תוחלת ושונות של X - מספר הכדורים הלבנים שבקופסה בסוף התהליך.

פיתרון:

- הערכים ש-X יכול לקבל הם $2, 3, 4$.

– לכן נמצא את התוחלת של X לפי הגדרה.

* נסמן B_1, B_2 - יצא כדור שחור.

* נסמן W_1, W_2 = יצא כדור לבן.

* נחשב את המחוברים בהגדרת התוחלת:

· עבור $x = 2$:

$$2 \cdot P\left(X = 2\right) = P\left(B_1 \cap B_2\right) = P\left(B_1\right) \cdot P\left(B_2 \mid B_1\right)$$

$$= \frac{1}{3} \cdot \frac{2}{4} = 2 \cdot \frac{1}{6}$$

- עבור $x = 3$:

$$3 \cdot P(X = 3) = P(W_1 \cap B_2) + P(B_1 \cap W_2) = \ldots = 3 \cdot \frac{1}{3}$$

- עבור $x = 4$:

$$4 \cdot P(X = 4) = P(W_1 \cap W_2) = 4 \cdot \frac{1}{2}$$

* כלומר התוחלת היא:

$$E[X] = \sum_{x=2}^{4} x \cdot P(X = x) = 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{3} + 4 \cdot \frac{1}{2} = 3\frac{1}{3}$$

- על מנת למצוא את השונות, נחשב לפי ההגדרה השנייה:

$$VarX = E\left[X^2\right] - (E[X])^2$$

  – ונקבל:

$$E\left[X^2\right] = \sum_{x=2}^{4} x^2 \cdot P(X = x) = 2^2 \cdot \frac{1}{6} + 3^2 \cdot \frac{1}{3} + 4^2 \cdot \frac{1}{2} = 11\frac{2}{3}$$

  – נציב ונקבל:

$$Var(X) = 11\frac{2}{3} - \left(3\frac{1}{3}\right)^2$$

תרגיל 5.

יש לנו n מכתבים שמכניסים ל-n מעטפות באופן אקראי.

יהא X מ"מ המתאר את מספר המכתבים שהגיעו ליעד.

חשב תוחלת ושונות של X.

פיתרון:

- לא נרצה ללכת לפי ההגדרה כי X לא משתנה מקרי מוכר.
- ולמרות שבתרגול הראשון חישבנו את $P\left(X\geq1\right)$, לא נרצה להשתמש בנוסחת הזנב כי נצטרך לחשב $P\left(X\geq x\right)$ לכל x.
- לכן נבחר בשיטת האינדיקטורים:
  - נסמן אינדיקטור $I_{j}=\begin{cases}1 & the\ j_{th}\ letter\ arrived\\0 & otherwise\end{cases}$
    * מתקיים:

$$X=\sum_{i=1}^{n}I_{i}$$

- כעת ניתן לחשב את התוחלת:

$$E\left[X\right]=E\left[\sum_{i=1}^{n}I_{i}\right]$$

$$=\sum_{i=1}^{n}P\left(i_{th}\ letter\ arrived\right)$$

  - ראינו בתרגול הראשון שמתקיים:

$$P\left(i_{th}\ letter\ arrived\right)=\frac{\left(n-1\right)!}{n!}=\frac{1}{n}$$

    * כלומר כל הפרמוטציות עבורן המכתב i הגיע ליעד, חלקי כל הפרמוטציות האפשריות.
  - ולכן מתקיים:

$$E\left[X\right]=\sum_{i=1}^{n}P\left(i_{th}\ letter\ arrived\right)$$

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

- נחשב שונות לפי הגדרה:

$$Var\left(X\right)=E\left[X^{2}\right]-\left(E\left[X\right]\right)^{2}$$

– נחשב את $E\left[X^2\right]$:

$$E\left[X^2\right]=E\left[\left(\sum_{i=1}^{n}I_i\right)^2\right]$$

$$=E\left[\sum_{i=1}^{n}\sum_{j=1}^{n}I_jI_i\right]$$

* **<u>טריק לסכומים כפולים:</u>** נרצה לפצל את הסכומים הכפולים הללו ונעשה זאת על ידי חלוקה למקרים בהם $i=j$ ולמקרים בהם $i\neq j$:

$$=E\left[\sum_{i=1}^{n}\sum_{j=1,\ \ j\neq i}^{n}I_jI_i+\sum_{i=1}^{n}\sum_{j=1}^{n}I_jI_j\right]$$

· מלינאריות התוחלת מתקיים:

$$=E\left[\sum_{i=1}^{n}\sum_{j=1,\ \ j\neq i}^{n}I_jI_i\right]+E\left[\sum_{i=1}^{n}\sum_{j=1}^{n}I_j\right]$$

$$=\sum_{i=1}^{n}\sum_{j=1,\ \ j\neq i}^{n}E\left[I_jI_i\right]+\sum_{i=1}^{n}\sum_{j=1}^{n}E\left[I_j\right]$$

· מכיוון ש-I_j^2 מפולג ברנולי עם אותו פרמטר $P\left(j_{th}\ \ letter\ \ arrived\right)$ כמו של I_j, מתקיים:

$$E\left(I_j^2\right)=P\left(j_{th}\ \ letter\ \ arrived\right)=\frac{1}{n}$$

· על מנת לחשב את $E\left[I_jI_i\right]$, נשים לב כי $E\left[I_jI_i\right]=P\left(A_j\cap A_i\right)$, כלומר ההסתברות שגם המכתב ה-i וגם המכתב ה-j הגיעו ליעד.

· כעת נחשב את $E\left[I_jI_i\right]$ כאשר $i\neq j$:

$$E\left[I_jI_i\right]=P\left(i_{th}\ \ letter\ \ arrived\ \ \cap\ \ j_{th}\ \ letter\ \ arrived\ \right)$$

$$=\frac{(n-2)!}{n!}$$

$$=\frac{1}{n\cdot(n-1)}$$

* קיבלנו:

$$E\left[X^2\right]=\sum_{i=1}^{n}\sum_{j=1,\ \ j\neq i}^{n}\frac{1}{n\cdot(n-1)}+\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{n}$$

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

$$=1+1$$

$$=2$$

- כעת מתקיים:

$$Var\left(X\right)=\overbrace{E\left[X^2\right]}^{=2}-\overbrace{\left(E\left[X\right]\right)^2}^{=1^2}=1$$