הסתברות - תרגול 7

29 בנובמבר 2018

אי־שוויון מרקוב

משפט (א"ש מרקוב)

יהי X משתנה מקרי אי־שלילי. אזי לכל $a > 0$ מתקיים

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

ההוכחה ממש פשוטה ומתבססת על ההערכה הבאה (הגסה) (בכיתה ראיתם הוכחה טיפה שונה)

$$aP(X \geq a) = \sum_{s \in Supp(X) \cap [a,\infty]} ap_X(s) \leq \sum_{s \in Supp(X) \cap [a,\infty]} sp_X(s) \leq E(X)$$

דוגמא 1

מטילים מטבע שנופל על ראש בהסתברות p 20 פעמים. חסמו את ההסתברות שהרצף "HH" התקבל פחות מפעמיים.

פיתרון

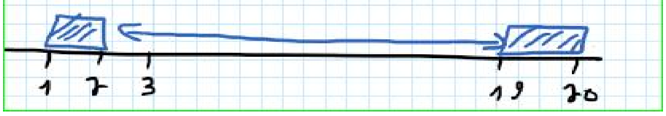
יהי $\Omega = \{H,T\}^{20}$ מרחב המדגם. נסמן ב־ X את מספר הפעמים שהרצף "HH" הופיע, ונגדיר $X_i = 1_{\{\omega_i = \omega_{i+1} = H\}}$. מתקיים $X_i \sim Ber\left(p^2\right)$ ולכן:

$$E\left(X\right) = E\left(\sum_{i=1}^{19} X_i\right) = \sum_{i=1}^{19} E\left(X_i\right) = 19p^2$$

מכיוון ש־ $X \leq 19$, נוכל להשתמש בא"ש מרקוב כדי להעריך:

$$P\left(X \leq 1\right) = P\left(19 - X \geq 18\right) \leq \frac{E\left(19 - X\right)}{18} = \frac{19\left(1 - p^2\right)}{18}$$

אני מסתכל על ציר הזמן
כל פרק זמן ; כל
מ 1 עד 19



$$E(19 - x) \underset{\text{לפי תוחלת}}{=} \underbrace{E(19)} - E(x) = \underbrace{19} - p^2$$

דוגמא 2

נסיק את א"ש בול מא"ש מרקוב.

תזכורת - א"ש בול: יהיו $\{A_n\}_{n=1}^{\infty}$ סדרת מאורעות במ"ה (Ω, P), אזי

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} P\left(A_n\right)$$

א"ש בול אמר לנו מה שידענו בקלות למקרה הסופי - שהסתברות של איחוד מאורעות קטנה או שווה לסכום הסתברות - למקרה האינסופי. אי־השיוויון נובע מכך שאולי יש חפיפות בין המאורעות. מי שלא זוכר, שיחזור לדיונים על נוסחת ההכלה וההדרה והגדרת הסיגמא־אדיטיביות.

נוכיח את א"ש בול בעזרת א"ש מרקוב.

הוכחה: יהיו $\{A_n\}_{n=1}^{\infty}$ סדרת מאורעות, ולכל n יהי $X_n = 1_{A_n}$ המ"מ המציין של A_n. לכל $m \in \mathbb{N}$ מתקיים

$$\begin{aligned} P\left(\bigcup_{n=1}^{m} A_n\right) &= P\left(\sum_{n=1}^{m} X_n \geq 1\right) \leq \frac{E\left[\sum_{n=1}^{m} X_n\right]}{1} \\ &= \sum_{n=1}^{m} E\left[X_n\right] = \sum_{n=1}^{m} P\left(A_n\right) \end{aligned}$$

כשהאי־שיוויון זה א"ש מרקוב ואז השתמשנו בלינאריות התוחלת ובנוסחת התוחלת של משתנה מציין. הסדרה $\left(\bigcup_{n=1}^{m} A_n\right)_{m=1}^{\infty}$ היא סדרה עולה, ולכן מרציפות ההסתברות

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{m\to\infty} P\left(\bigcup_{n=1}^{m} A_m\right) \leq \lim_{m\to\infty} \sum_{n=1}^{m} P\left(A_n\right) = \sum_{n=1}^{\infty} P\left(A_n\right)$$

אם אגף ימין הוא אינסוף, אז א"ש בול נכון תמיד, ואם הוא מתכנס אז קיבלנו את א"ש בול. ∎

הערה 0.1 שימו לב שהשתמשנו בלינאריות התוחלת במקרה הסופי - כי זה מה שהמשפט נותן לנו. באופן כללי אין לנו "לינאריות אינסופית" .

שונות

התוחלת נתנה לנו מעין ממוצע של התוצאות האפשריות, משוקלל ע"י הסבירות שלהן.

השונות נותנת לנו עוד מידע על המ"מ - כמה "רחוק" הוא מתפזר ממרכז הכובד שלו.

גם השונות והתוחלת ביחד לא מספיקים לאפיין מ"מ - אבל זה צעד בכיוון הנכון.

הגדרה 0.2 השונות של משתנה מקרי X מוגדרת

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

כשאת השוויון השני הוכחתם בכיתה.

דוגמאות

1. גלגול קוביה הוגנת הוא $X \sim U[6]$ וראינו ש־$E[X] = 3.5$. חישוב לפי הנוסחה נותן $VarU[N] = \frac{N^2-1}{12}$

$$Var(X) = \frac{36-1}{12} = \frac{35}{12}$$

שזה קצת קטן מ־3. זה מאוד מתאים למה שאנחנו חושבים על משתנה אחיד. התוחלת היא ממוצע אבל תוצאות רחוקות ממנה סבירות בדיוק כמו תוצאות קרובות אליה.

2. הטלת מטבע עם הסתברות q ל־H היא מ"מ ברנולי $X \sim Ber(q)$ עם תוחלת q ושונות $Var(X) = q(1-q)$. שימו לב שאם q מאוד גדול או מאוד קטן אז השונות תהיה מאוד קרובה ל־0, וזה מסתדר אם התפיסה שלנו את הניסוי ־ הסיכוי להיכשל (q קטן) או להצליח (q גדול) הוא מאוד גדול.

3. ספירה של H ברצף של N הטלות של המטבע מהדוגמא הקודמת זה מ"מ בינומי $X \sim Bin(N,q)$ בעל תוחלת $E[X] = Nq$ ושונות $Var(X) = Nq(1-q)$.

דיון קצר (אפשר להפנות לקריאה בלבד)

השונות בעצם מודדת מרחק מהתוחלת. ההגדרה כוללת ריבוע, כמו הנורמה האוקלידית, שנובעת ממשפט פיתגורס ־ עבור $u \in \mathbb{R}^n$

$$|u| = \sqrt{\sum_{i=1}^{n} u_i^2}$$

זה נראה במבט ראשון לא כ"כ טבעי, ושימוש בערך מוחלט בהגדרה, דהיינו $E[|X-E[X]|]$, נראה יותר טבעי. למרות זאת, ההגדרה עם הריבוע היא טובה יותר מבחינה מתמטית. קודם כל

$$P(|X-E[X]| > a) = P\left((X-E[X])^2 > a^2\right)$$

כך שההגדרה החדשה כן מודדת מרחק במובן היותר טבעי של המילה. שנית, פונקציית הערך המוחלט היא לא גזירה. זה אולי פחות בעייתי במקרה הבדיד, אך במקרה של מ"מ רציפים ־ שנגיע אליהם בהמשך ־ זה נהיה משמעותי. אי לכך השימוש בריבוע הוא נוח יותר, בייחוד לאור מה שצוין, שזה מודד את אותו דבר. כשרוצים לדבר ממש על המרחק אז לוקחים שורש של השונות, וזה מה שנקרא **סטיית התקן**.

א"ש צ'בישב

משפט 0.3 (א"ש צ'בישב) יהי X מ"מ בעל תוחלת ושונות סופיות. <u>אזי</u> לכל $a > 0$ מתקיים

$$P(|X-E[X]| > a) \le \frac{\mathrm{Var}[X]}{a^2}$$

א"ש זה נותן לנו להבין למה השונות אומרת כמה ההסתברות של המ"מ נופלת כשהתוצאות מתרחקות מהתוחלת.

דוגמא נניח ש $X \sim U[11]$. מה ההסתברות ש־X רחוק מהתוחלת שלו ביותר מ 4?
את התוחלת אנו יודעים לחשב:

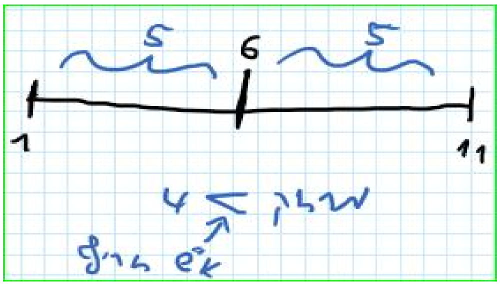
$$E[X] = \frac{11+1}{2} = 6$$

ו

$$\mathrm{Var}[X] = \frac{11^2-1}{12} = 10$$

ומכאן

$$P(|X - E[X]| > 4) \le \frac{\mathrm{Var}[X]}{16} = \frac{10}{16}.$$



מצד שני, אפשר לחשב את ההסתברות הזאת גם באופן ישיר:

$$P(|X - \mathbb{E}(X)| > 4) = P(X=1) + P(X=11) = \frac{2}{11}$$

רואים שהחסם שקיבלנו בעזרת צ'בישב איננו הדוק.

שונות משותפת

הגדרה 0.4 לכל שני משתנים מקריים X, Y בעלי תוחלת סופית השונות המשותפת שלהם מוגדרת ע"י

$$Cov(X,Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y]$$

שימו לב ש $Cov(X,X) = Var(X)$.
זהו מדד לפיזור מכפלת המ"מ סביב מכפלת התוחלות.
ברמה הטכנית, מטענה שתכף נזכיר, החשיבות המרכזית של השונות המשותפת היא בחישוב שונות של סכום מ"מ (לאו דוקא ב"ת).

ננסה להבין מה אומרת לנו השונות המשותפת
ההגדרה היא

$$\mathrm{Cov}[X,Y] = E[(X - E[X])(Y - E[Y])]$$

זה ארוך קצת, אבל נשים לב שהמ"מ שבסוגריים (כולל החיסור) הם בעצם מ"מ בעלי תוחלת 0, והתפלגות שאפשר לנסח ע"י המ"מ המקוריים. אז נכתוב מחדש תחת ההנחה שהתוחלות של X ו־Y הם 0.

$$\mathrm{Cov}[X,Y] = E[XY]$$

זה יותר פשוט. נזכיר שתוחלת היא ממוצע משוקלל. מתי $E[XY]$ מאוד גדול? כשההסתברות ש־X ו־Y גדולים ביחד היא גדולה, וההסתברות שכשאחד גדול השני קטן היא נמוכה. מתי $E[XY]$ מאוד קטן (הכוונה למספר שלילי שערכו המוחלט גדול)? בדיוק במצב ההפוך, כלומר שכש־X מאוד קטן אז Y מאוד גדול וההיפך - בהסתברות גבוהה. יש כאן אמירה

קצת יותר חזקה מתלות - אנחנו מקבלים מושג על איך - בגדול - המשתנים מתנהגים כשמסתכלים עליהם בו־זמנית.

מאידך גיסא - מתי $E[XY]=0$? יש שתי אפשרויות, בגדול - אחת שתמיד כש־$X\neq 0$ אז $Y=0$, וההיפך. אפשרות שניה היא שהמקרים שבהם X,Y גדולים ביחד מאזנים את המקרים בהם X קטן ו־Y גדול וההיפך. המקרה הראשון מתאר תלות חזקה מאוד. המקרה השני יכול להיות גם במקרה של תלות וגם במקרה של אי־תלות.

מכך אנו למדים שהשונות המשותפת מודדת תיאום בהתנהגות של המ"מ, כשתיאום יכול להיות ישר (שניהם מתנהגים מאוד דומה) או הפוך (כשאחד גדול השני קטן וההיפך).

כמה תכונות שהוכחו בכיתה

יהיו X,Y,Z מ"מ בעלי שונות סופית ויהי $a\in\mathbb{R}$.

1. $\mathrm{Var}[X+a]=\mathrm{Var}[X]$. זה הגיוני מאוד - הזזנו את כל טווח התוצאות ב־a, ולכן הממוצע זז ב־a, אבל הפיזור של התוצאות סביב הממוצע המוזז זהה לפיזור סביב הממוצע הקודם!
2. $\mathrm{Var}[aX]=a^2\mathrm{Var}[X]$. שוב - הגיוני. ניפחנו את התוצאות ב־a, וגם הפיזור התנפח בהתאם. הריבוע של ה־a בא מההגדרה של התוחלת באמצעות ריבוע במקום ערך מוחלט.
3. אם X,Y ב"ת <u>אזי</u> $\mathrm{Cov}[X,Y]=0$.
4. $\mathrm{Var}[X+Y]=\mathrm{Var}[X]+\mathrm{Var}[Y]+2\mathrm{Cov}[X,Y]$, ובפרט אם X,Y ב"ת, אז הראיתם שהם גם לא מתואמים - כלומר בעלי שונות משותפת 0 ומתקבל
$$\mathrm{Var}[X+Y]=\mathrm{Var}[X]+\mathrm{Var}[Y]$$
5. ובאופן כללי -

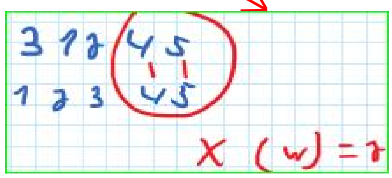
$$\mathrm{Var}\left[\sum_{n=1}^{N}X_n\right]=\sum_{n,k\leq N}\mathrm{Cov}[X_n,X_k]=\sum_{n=1}^{N}\mathrm{Var}[X_n]+2\sum_{n<k\leq N}\mathrm{Cov}[X_n,X_k]$$

בעיית הדוור המבולבל:

מסדרים את המספרים $\{1,...,n\}$ בשורה, בהסתברות אחידה על כל אפשרות סידור. יהא X המ"מ המקרי המתאר את כמות המספרים שנמצאים במקומם. לדוגמה, עבור $n=5$, בסידור

$$\omega=(3\,1\,2\,4\,5)$$

מתקיים $X(\omega)=2$. נרצה לחשב את השונות של המ"מ X. ניעזר בשיטת החלוקה למציינים:
יהא $X_i\sim ber(\frac{1}{n})$ המ"מ שנותן 1 אם המספר i נמצא במקום ה־i, ו 0 אחרת. מכיוון ש



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

המשתנים בין תלויים מכיון ש: $P(X_i \cap X_j) \neq P(X_i) \cdot P(X_j)$ נניח ש X_1 בודק אם 1 גר או לא

X_2 בודק האם 2 גר או לא אז $P(X_1 \cap X_2) = P($ 1,2 לא גרים $)$ $=$ (1 2) 3 4 5 צמודים יש לנו $3!$ אפשרויות

לסדר את 3,4,5. לעומת זאת $P(X_1) = P($ 1 לא גר $)$ $=$ (1) 2 3 4 5 הכבוד $4! =$ מאורגן דורש עבור 2 לכן

$\frac{1}{3!} \neq \frac{1}{4!} \cdot \frac{1}{4!}$

נקבל מלינאריות התוחלת ש־

$$\mathbb{E}(X) = \mathbb{E}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \mathbb{E}(X_i) = n \cdot \frac{1}{n} = 1.$$

החיים שלנו היו קלים לו המ"מ היו ב"ת ־ אז היינו יכולים להעזר בלינאריות ולסיים את החישוב בקלות. זה לא המצב (מדוע?). נבחין, אם כך, שהמשתנה המקרי $X_1 \cdot X_2 \sim ber(\frac{1}{n(n-1)})$, וכך נכון לכל $X_i X_j$ כך ש $i < j$. לכן

$$COV(X_1, X_2) = \mathbb{E}(X_1 X_2) - \mathbb{E}(X_1)\mathbb{E}(X_2) = \frac{1}{n(n-1)} - \frac{1}{n^2} = \frac{1}{n}(\frac{1}{n-1} - \frac{1}{n}) = \frac{1}{n^2(n-1)}$$

לבסוף, מהנוסחא לשונות של סכום מ"מ ושונות של מ"מ ברנולי מתקיים

$$V(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} V(X_i) + 2\sum_{i<j} COV(X_i, X_j) = n \cdot \frac{1}{n}(1 - \frac{1}{n}) + 2 \cdot \binom{n}{2} \frac{1}{n^2(n-1)} = 1$$

נשאל כעת שאלה נוספת:
מה ההסתברות שיותר מ $\frac{n}{2}$ אנשים קיבלו את המכתב שיועד להם?
בתרגול 8 (דוגמה 4) ראינו ש $\mathbb{E}(X) = V(X) = 1$. לכן, מאי שוויון מרקוב נקבל:

$$P(X \geq \frac{n}{2}) \leq \frac{2}{n}$$

ואילו מאי שוויון צ'בישב:

$$P(X \geq \frac{n}{2}) = P(X - 1 \geq \frac{n}{2} - 1) \leq P(|X - 1| \geq \frac{n}{2} - 1) \leq \frac{V(X)}{(\frac{n}{2} - 1)^2} = \frac{4}{(n-2)^2}$$

כעת נניח שיש $n + 1$ אנשים, כל אחד שולח n מכתבים ל־n האנשים מלבדו באופן אקראי. מה ההסתברות שכל האנשים טעו ביותר מ־$\frac{n}{2}$ מהשליחות?
לכל $1 \leq i \leq n + 1$ יהא X_i המ"מ המתאר את מספר המכתבים שהאיש ה־i שלח והגיעו ליעדם. מהאי שיוון האחרון נקבל:

$$\begin{aligned} P(X_1 < \frac{n}{2}, ..., X_{n+1} < \frac{n}{2}) &= 1 - P\left(\bigcup_{i=1}^{n+1} \{X_i \geq \frac{n}{2}\}\right) \\ &\geq 1 - \sum_{i=1}^{n+1} P(X_i \geq \frac{n}{2}) \\ &\geq 1 - \frac{4(n+1)}{(n-2)^2} \end{aligned}$$

כאשר הא"ש הראשון נובע מא"ש בול.