הרצאה 7

תזכורת:

ראינו שכאשר יש לנו שני משתנים מקריים המוגדרים מעל אותו מרחב מדגם, X,Y אפשר לדבר על $X|Y$ (וגם על $Y|X$)
כאל משתנה מקרי בפני עצמו - לכל ערך אפשרי k של Y, $X|Y=k$ הוא משתנה מקרי.

תוחלת שלמה

יהיו X,Y משתנים מקריים המוגדרים מעל אותו מרחב מדגם. מתקיים:

$$E(X)=E\big(E(X|Y)\big)$$

כלומר:

$$E(X)=\sum_k E(X|Y=k)\cdot P(Y=k)$$

השוויון הזה נקרא נוסחת התוחלת השלמה.

מה יכול להיות טוב בנוסחה הזו? לעיתים, קשה "לאכול" את המשתנה X בפני עצמו, ובפרט לחשב לו את התוחלת אבל את המשתנים $X|Y=k$ יחסית קל "לאכול". כלומר נוסחת התוחלת השלמה הופכת להיות שימושית כשאנו יודעים לתאר בקלות יחסית את המשתנה $X|Y=k$ אבל לא קל לתאר ישירות את X.

דוגמה:

מטילים מטבע הוגן n פעמים. בכל פעם שיוצא עץ אנחנו מכניסים כדור לבן לכד, ובכל פעם שיוצא פלי אנחנו מכניסים כדור שחור לכד.
לאחר מכן, אנחנו מוציאים מהכד m כדורים ללא החזרה ($m\le n$ מן הסתם).
נסמן ב- X את מספר הכדורים הלבנים שיצאו, חשבו את $E(X)$.

X דומה למשתנה היפרגיאומטרי - יש לנו בסך הכל n כדורים, אנחנו מוציאים m, והלבנים הם ה"מיוחדים".
מה הבעיה? חסר לנו ה- D, אנחנו לא יודעים כמה כדורים לבנים יש בכד.
במילים אחרות, המשתנה "X בהינתן שיש בכד D לבנים" הוא משתנה היפרגיאומטרי.

נסמן ב- Y מספר העצים שיצאו לנו ב- n הטלות מטבע, כך נקבל ש- $X|Y=k$ מתפלג היפרגיאומטרית, עם n כדורים בסך הכל, מתוכם יש k כדורים מיוחדים (לבנים) ומוציאים m כדורים ולכן $X|Y=k\sim HG(n,k,m)$
לפי נוסחת התוחלת השלמה:

$$E(X)=\sum_k E(X|Y=k)\cdot P(Y=k)$$

כפי שאמרנו $X|Y=k\sim HG(n,k,m)$ ולכן בשביל $E(X|Y=k)$ נוכל להשתמש בנוסחה הידועה לתוחלת של משתנה היפר-גיאומטרי.

$$E(X|Y=k)=\frac{mk}{n}$$

כמו כן, Y הוא משתנה בינומי - מטילים n פעמים וסופרים כמה פעמים יצא עץ. ההסתברות להצלחה (יצא עץ) היא $\frac{1}{2}$
כלומר $Y\sim Bin(n,\frac{1}{2})$ לכן אנחנו יודעים לחשב את $P(Y=k)$

$$P(Y=k)=\binom{n}{k}\cdot\left(\frac{1}{2}\right)^k\cdot\left(1-\frac{1}{2}\right)^{n-k}$$

לכן נקבל:

$$E(X)=\sum_k E(X|Y=k)\cdot P(Y=k)=\sum_k \frac{mk}{n}\cdot\binom{n}{k}\cdot\left(\frac{1}{2}\right)^k\cdot\left(1-\frac{1}{2}\right)^{n-k}$$

$$=\frac{m}{n}\sum_k k\cdot\underbrace{\binom{n}{k}\cdot\left(\frac{1}{2}\right)^k\cdot\left(1-\frac{1}{2}\right)^{n-k}}_{P(Y=k)}$$

$$=\frac{m}{n}\underbrace{\sum_k k\cdot P(Y=k)}_{E(Y)\text{ הגדרה של}}=\frac{m}{n}\cdot E(Y)$$

את $E(Y)$ ניתן לחשב לפי הנוסחה הידועה לתוחלת של משתנה בינומי $E(Y)=n\cdot\frac{1}{2}$ כלומר $E(Y)=\frac{n}{2}$ ולכן:

$$E(X)=\frac{m}{n}\cdot E(Y)=\frac{m}{n}\cdot\frac{n}{2}=\frac{m}{2}\Rightarrow E(X)=\frac{m}{2}$$

דוגמה:

בכד יש לנו 3 כדורים ירוקים, 4 כדורים כחולים ו- 3 כדורים אדומים. מוציאים כדור מהכד.
אם יצא כדור ירוק - מטילים קוביה הוגנת עד שמקבלים 1.
אם יצא כדור כחול - מטילים קוביה הוגנת עד שמקבלים 6.
אם יצא כדור אדום - מטילים קוביה הוגנת עד שמקבלים מספר אי זוגי.
נסמן ב- X את מספר הטלות הקוביה שהתבצעו. חשב את $E(X)$.

<u>פתרון:</u>

את X קשה לנו לתאר אבל את "X <u>בהינתן e</u> יצא כדור רצה או אחר" אותנו יודעים לתאר - זה משתנה גיאומטרי ("עד e")
אם כן, נסמן ב- Y את צבע הכדור שיצא. לפי הנתונים אנו נקבל:

$$X|Y=\text{ירוק}\sim G\left(\frac{1}{6}\right)$$

המשתנה $X|Y=\text{ירוק}$ הוא הטלת קוביה עד שמקבלים 1: זהו משתנה גיאומטרי, מנסים עד שמצליחים וההסתברות להצלחה היא $\frac{1}{6}$. באופן דומה:

$$X|Y=\text{כחול}\sim G\left(\frac{1}{6}\right)$$

$$X|Y=\text{אדום}\sim G\left(\frac{1}{2}\right)$$

אם כן, נשתמש בנוסחת התוחלת השלמה:

$$E(X)=\sum_k E(X|Y=k)\cdot P(Y=k)$$

בניגוד לדוגמה הקודמת, ל- Y עכשיו אין התפלגות ידועה, ואת ההסתברויות צריך לחשב באופן ידני. נקבל:

$$E(X)=\sum_k E(X|Y=k)\cdot P(Y=k)$$

$$E(X)=E(X|Y=\text{ירוק})\cdot P(Y=\text{ירוק})+E(X|Y=\text{כחול})\cdot P(Y=\text{כחול})+E(X|Y=\text{אדום})\cdot P(Y=\text{אדום})$$

התוחלת של משתנה גיאומטרי היא $\frac{1}{p}$ ולכן נקבל :

$$E(X)=\underbrace{\frac{1}{\frac{1}{6}}}_{E(X|Y=\text{ירוק})}\cdot\underbrace{\frac{3}{10}}_{P(Y=\text{ירוק})}+\underbrace{\frac{1}{\frac{1}{6}}}_{E(X|Y=\text{כחול})}\cdot\underbrace{\frac{4}{10}}_{P(Y=\text{כחול})}+\underbrace{\frac{1}{\frac{1}{2}}}_{E(X|Y=\text{אדום})}\cdot\underbrace{\frac{3}{10}}_{P(Y=\text{אדום})}=\frac{48}{10}=4.8$$

$$E(X)=4.8$$

אינדיקטורים

אינדיקטור זה משתנה ברנולי - משתנה שמקבל 1 בהסתברות p ו-0 בהסתברות של $1-p$.
לעיתים, בהינתן משתנה X, קשה לחשב את התוחלת או השונות שלו "ישירות", אבל אפשר להציג את X כסכום של אינדיקטורים $X_1+X_2+\dots$ ואז את התוחלת והשונות של הסכום להצליח לחשב ביתר קלות.

באופן כללי, אם X_i הוא אינדיקטור אז $E(X_i)=P(X_i=1)$.
נזכור תוחלת מוגדרת כך:

$$E(X_i)=\sum kP(X_i=k)$$

הערכים ש-X_i יכול לקבל הם רק 0 או 1 ולכן

$$E(X_i)=\sum kP(X_i=k)=0\cdot P(X_i=0)+1\cdot P(X_i=1)=P(X_i=1)$$

לכן, אם מצליחים לפרק את X כסכום של אינדיקטורים כלומר $X=X_1+X_2+\dots$ אז:

$$E(X)=E(X_1+X_2+\dots+X_n)=E(X_1)+\dots+E(X_n)=P(X_1=1)+\dots+P(X_n=1)$$

כלומר : $E(X)=\sum_{i=1}^{n}P(X_i=1)$ $\longrightarrow$ תוחלת של X ע"י פירוק לאינדיקטורים

דוגמה :

ראינו את בעיית הפקיד הרשלן - יש n מכתבים ו-n מעטפות שמיועדות להם, והפקיד מכניס את המכתבים רנדומלית למעטפות.
נסמן ב-X את מספר המכתבים שהפקיד מכניס למעטפה הנכונה. מה תוחלת של X?

פתרון :

נסמן ב-X_i את האינדיקטור הבא - X_i הוא 1 אם המכתב ה-i הוכנס למעטפה הנכונה, ו-X_i הוא 0 אחרת.
כך נקבל ש: $X=X_1+X_2+\dots+X_n$.
למשל נניח ש-$n=3$ ונניח שרק המכתב השני נכנס למעטפה הנכונה, כלומר המ"מ הוא $X=1$. במקרה כזה: $X_1=0, X_2=1, X_3=0$
המכתב הראשון לא נכנס למעטפה שלו ולכן $X_1=0$, המכתב השני כן נכנס למעטפה שלו ולכן $X_2=1$ והמכתב השלישי לא נכנס למעטפה שלו ולכן $X_3=0$. ולכן $X=X_1+X_2+X_3=1$.

כעת, נחשב את התוחלת:

$$E(X)=\sum_{i=1}^{n}P(X_i=1)$$

מה $P(X_i=1)$? כלומר מה ההסתברות שהמכתב שהוכנס אכן הוכנס למעטפה הנכונה? אם יש לנו n מעטפות
המכניסים אקראית אז ההסתברות היא $\frac{1}{n}$. נקבל :

$$E(X)=\sum_{i=1}^{n}P(X_i=1)=\sum_{i=1}^{n}\frac{1}{n}=\underbrace{\frac{1}{n}+\frac{1}{n}+\frac{1}{n}+\dots+\frac{1}{n}}_{\text{n פעמים}}=n\cdot\frac{1}{n}=1$$

$$E(X)=1$$

דוגמה:

נניח ש- X מתפלג היפרגאומטרית כלומר $X\sim HG(N,D,n)$. בהרצאה 5 ראינו ש- $E(X)=\frac{Dn}{N}$ (וגם בהרצאה הזו הצגנו את זה).

אנחנו מוציאים n פריטים מתוך ה- N ו- X מסמל את מספר המיוחדים שיצאו.

אנחנו יכולים לבטא את X כסכום של אינדיקטורים: $X_1+\dots+X_n$ במובן הבא: $X_i=1$ אם הפריט ה- i שהוצאנו הוא מיוחד ו- $X_i=0$ אם הפריט ה- i שהוצאנו הוא לא מיוחד.

למשל, בדוגמת המבחן בהרצאה 5, מתקיים $n=6$, $D=2$, $N=20$. אם למשל, בשאלה הראשונה והשאלה השישית יצאו לנו שאלה "מיוחדת" אז המ"מ הוא $X=2$ וניתן לפרק אותו לאינדיקטורים: $X_1=1, X_2=0, X_3=0, X_4=0, X_5=0, X_6=1$ (השאלה הראשונה מיוחדת, השאלות 2-5 לא מיוחדות והשאלה השישית מיוחדת) ואכן:

$$2=X=X_1+X_2+X_3+X_4+X_5+X_6=1+0+0+0+0+1=2$$

אם כן, נקבל:

$$E(X)=\sum_{i=1}^{n}P(X_i=1)$$

מכיוון שיש לנו D פריטים מיוחדים מתוך N פריטים בסך הכל. לכן אנו מקבלים $P(X_i=1)=\frac{D}{N}$ (להוציא משהו מיוחד) ונקבל:

$$E(X)=\sum_{i=1}^{n}P(X_i=1)=\sum_{i=1}^{n}\frac{D}{N}=\underbrace{\frac{D}{N}+\frac{D}{N}+\dots+\frac{D}{N}}_{\text{n פעמים}}=n\cdot\frac{D}{N}=\frac{Dn}{N}$$

שונות של X ע"י פירוק לאינדיקטורים:

כאשר, אם X_i אינדיקטור עם הסתברות p להיות 1 והסתברות $1-p$ להיות 0, השונות של האינדיקטור X_i היא:

$$V(X_i)=E(X_i^2)-(E(X_i))^2$$

כמו שראינו ש- $E(X_i)=P(X_i=1)$. כמו כן:

$$E(X_i^2)=\sum k^2p(X_i=k)=0^2\cdot p(X_i=0)+1^2\cdot p(X_i=1)=p(X_i=1)$$

$$V(X_i)=E(X_i^2)-(E(X_i))^2=P(X_i=1)-(p(X_i=1))^2=p(X_i=1)\left(1-p(X_i=1)\right)$$

למה שווה השונות של הסכום $V(X_1+X_2+\dots+X_n)$?

בהרצאה הקודמת ראינו ש- $V(X_i+X_j)=V(X_i)+V(X_j)+2cov(X_i,X_j)$. מכיוון ש- $cov(X_i,X_j)=cov(X_j,X_i)$ אפשר במקום $2cov(X_i,X_j)$ לרשום: $cov(X_i,X_j)+cov(X_j,X_i)$

$$V(X_i+X_j) = V(X_i)+V(X_j) + cov(X_i,X_j)+cov(X_j,X_i)$$

בנוסף לפי תכונות cov מתקיים $V(X_i) = cov(X_i,X_i)$, $V(X_j) = cov(X_j,X_j)$ לכן:

$$V(X_i+X_j) = cov(X_i,X_i) + cov(X_j,X_j) + cov(X_i,X_j)+cov(X_j,X_i)$$

נשים לב- אפשר לומר שהשונות של סכום היא סכום השונויות המשותפות של כל הזוגות האפשריים של שני משתנים.

$$V(X_1+X_2+\ldots+X_n) = \sum_{i=1}^{n}\sum_{j=1}^{n} cov(X_i,X_j)$$

נשים לב- הסכום מוכיל כל השונויות המשותפות של משתנה עם עצמו $cov(X_i,X_i)$ שאנחנו יודעים שהן שווה ל- $V(X_i)$.
כמו כן: $cov(X_i,X_j) = cov(X_j,X_i)$ ולכן בסוף אפשר לרשום שאנו מחשבים את הסכום הבא:

$$V(X_1+X_2+\ldots+X_n) = \sum_{i=1}^{n} V(X_i) + 2\sum_{i>j}^{n} cov(X_i,X_j)$$

כלומר, סכום השונויות ועוד פעמיים הסכום של כל השונויות של משתנים שונים זה מזה.
כמה שווה $V(X_i)$ כבר ראינו - $P(X_i=1)\left(1-P(X_i=1)\right)$. מה לגבי $cov(X_i,X_j)$? לפי הנוסחה:

$$cov(X_i,X_j) = E(X_iX_j) - E(X_i)E(X_j)$$

כעת:

$$E(X_iX_j) = \sum km\cdot P(X_i=k,\ X_j=m)$$

הערכים האפשריים של k,m הם $0,1$ ולכן המחובר היחידי בסכום הזה שלא מתאפס הוא זה שמתאים לערכים $k,m=1$.

$$E(X_iX_j) = \sum km\cdot P(X_i=k,\ X_j=m) = 1\cdot 1\cdot P(X_i=1,X_j=1) = P(X_i=1,X_j=1)$$

בנוסף לכך כבר אמרנו ש- $E(X_i)=P(X_i=1)$ וכן גם $E(X_j)=P(X_j=1)$ ולכן

$$cov(X_i,X_j) = E(X_iX_j) - E(X_i)E(X_j)$$

$$= P(X_i=1,X_j=1) - P(X_i=1)\cdot P(X_j=1)$$

וסה"כ במקום: $V(X_1+X_2+\ldots+X_n) = \sum_{i=1}^{n} V(X_i) + 2\sum_{i>j}^{n} cov(X_i,X_j)$ נרשום:

$$V(X_1+X_2+\ldots+X_n) = \sum_{i=1}^{n} P(X_i=1)\left(1-P(X_i=1)\right) + 2\sum_{i>j}^{n} P(X_i=1,X_j=1) - P(X_i=1)\cdot P(X_j=1)$$

דוגמה :

יש לנו n זוגות של בעל ואישה, כלומר $2n$ אנשים בסך הכל. אנחנו מסדרים את $2n$ האנשים בשורה.
נסמן ב- X את מספר הזוגות (מתוך ה- n) שבהם הבעל והאישה לא עומדים אחד ליד השני.
חשבו את $E(X), V(X)$.

פתרון :

אם כן נתאר X כסכום של אינדיקטורים - נסמן ב- X_i את האינדיקטור שמקיים $X_i=1$ - אם הבעל והאישה מהזוג ה- i לא עומדים זה ליד זה, ו- $X_i=0$ אם הבעל והאישה מהזוג ה- i עומדים זה ליד זה.

ראשית:

$$E(X)=\sum_{i=1}^{n}P(X_i=1)$$

כעת, נותר לנו לחשב את $P(X_i=1)$ כלומר את ההסתברות שהבעל והאישה לא יושבים זה לצד זה.
כאן כנראה שיותר קל לחשב את המשלים כלומר $P(X_i=0)$ ואז ממנו להסיק את $P(X_i=1)$.
נחשב את $P(X_i=0)$ - מה ההסתברות שהבעל והאישה מהזוג ה- i יושבים זה לצד זה.
יש לנו בסך הכל $(2n)!$ דרכים לסדר $2n$ אנשים בשורה. בכמה מתוכם הבעל והאישה מהזוג ה- i עומדים זה לצד זה?
יש לנו $2!$ דרכים לסדר אותם זה לצד זה (בעל - אישה / אישה - בעל) ואז אפשר לחשוב על שניהם כעל "בלוק אחד"
יש לנו עוד $2n-2$ אנשים וביחד עם ה"בלוק" נותר לנו לסדר $2n-1$ אנשים בשורה כלומר $(2n-1)!$ אפשרויות.
כלומר:

$$P(X_i=0)=\frac{2!\cdot(2n-1)!}{(2n)!}=\frac{2\cancel{(2n-1)\cdot(2n-2)\cdot\ldots\cdot1}}{2n\cdot\cancel{(2n-1)\cdot(2n-2)\cdot\ldots\cdot1}}=\frac{2}{2n}=\frac{1}{n}$$

$$P(X_i=1)=1-P(X_i=0)=1-\frac{1}{n}$$

$$P(X_i=1)=1-\frac{1}{n}$$

כעת נחשב את התוחלת :

$$E(X)=\sum_{i=1}^{n}P(X_i=1)=\underbrace{\left(1-\frac{1}{n}\right)\cdot\left(1-\frac{1}{n}\right)\cdot\ldots\cdot\left(1-\frac{1}{n}\right)}_{\text{פעמים } n}=n\left(1-\frac{1}{n}\right)=n-1$$

$$E(X)=n-1$$

שנית, נחשב את השונות:

$$V(X_1+X_2+\ldots+X_n)=\sum_{i=1}^{n}P(X_i=1)\left(1-P(X_i=1)\right)+2\sum_{i>j}P(X_i=1,X_j=1)-P(X_i=1)\cdot P(X_j=1)$$

כדי להקל על עצמנו נסמן $P(X_i=1)=P_i$, $P(X_j=1)=P_j$ ו- $P(X_i=1,X_j=1)=P_{ij}$ וכך נחסוך קצת בכתיבה:

$$V(X_1+X_2+\ldots+X_n)=\sum_{i=1}^{n}P_i(1-P_i)+2\sum_{i>j}(P_{ij}-P_iP_j)$$

מקרה אחד:

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

מקרה שני: אנחנו רוצים לחשב את $2\sum_{i>j}^{n}(P_{ij}-P_iP_j)$ כלומר אנחנו צריכים לחשב את ההסתברות $P(X_i=1,X_j=1)$
כלומר אנחנו רוצים שבשני הזוגות, ה- i וה- j הבעל והאישה לא עומדים זה ליד זה. כמו מקודם, יותר קל לחשב
כאן משלים - ההסתברות שבזוג ה- i הבעל והאישה עומדים זה ליד זה <u>או</u> שהבעל והאישה בזוג ה- j עומדים זה ליד זה.

$$P(X_i=1,X_j=1) = P(X_i=1\cap X_j=1) = 1-P\left(\overline{X_i=1\cap X_j=1}\right) = 1-P\left(\overline{X_i=1}\cup\overline{X_j=1}\right)$$

$$= 1-P(X_i=0\cup X_j=0)$$

מכאן אנו יודעים להמשיך באופן הבא:

$$1-P(X_i=0\cup X_j=0) = 1-\Big(P(X_i=0)+P(X_j=0)-P(X_i=0\cap X_j=0)\Big)$$

כפי שכבר ראינו $P(X_i=0)=P(X_j=0)=\frac{1}{n}$. באופן דומה, מה ההסתברות $P(X_i=0\cap X_j=0)$? אנו רוצים שבבעל והאישה מהזוג ה- i יעמדו זה ליד זה וגם הבעל והאישה מהזוג ה- j יעמדו זה לצד זה.
מתוך $(2n)!$ סידורים אפשריים יש $2!\cdot 2!(2n-2)!$ סידורים בהם זה קורה (בדומה לנימוק הקודם) - יש $2!$ דרכים לסדר בעל ואישה מזוג i, $2!$ דרכים לסדר בעל ואישה מזוג j. נחשוב על כל זוג כבלוק אחד ונדבר בסידור בשורה של שאר $2n-4$ אנשים עם עוד 2 בלוקים כאלה כלומר סידור בשורה של $2n-4+2$ כלומר סידור $2n-2$ אנשים בשורה כלומר $(2n-2)!$. לפי עקרון הכפלה: $2!\cdot 2!(2n-2)!$ ולכן:

$$P(X_i=0,X_j=0) = \frac{2!\cdot 2!(2n-2)!}{(2n)!} = \frac{2!\cdot 2!\cdot\cancel{(2n-2)\cdot(2n-3)\cdots 1}}{2n\cdot(2n-1)\cdot\cancel{(2n-2)(2n-3)\cdots 1}} = \frac{4}{2n(2n-1)} = \frac{2}{n(2n-1)}$$

$$P(X_i=1,X_j=1) = 1-\Big(\underbrace{P(X_i=0)}_{\frac{1}{n}}+\underbrace{P(X_j=0)}_{\frac{1}{n}}-\underbrace{P(X_i=0\cap X_j=0)}_{\frac{2}{n(2n-1)}}\Big)$$

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

מכאן:

$$2\sum_{i>j}^{n}(P_{ij}-P_iP_j) = 2\sum_{i>j}^{n}P(X_i=1,X_j=1)-P(X_i=1)\cdot P(X_j=1)$$

$$=2\sum_{i>j}^{n}1-\left(\frac{1}{n}+\frac{1}{n}-\frac{2}{n(2n-1)}\right)-\left(1-\frac{1}{n}\right)\left(1-\frac{1}{n}\right)$$

סיכום :

נלמד שיטה לחישוב תוחלת ושונות של משתנה מקרי, על ידי פירוקו לסכום של משתני אינדקטור. אינדיקטור הינו משתנה שפונקציית ההסתברות של נראית כך :

X	1	0
$P(X)$	P	$1-P$

נגיד ש- X_i הינו משתנה אינדיקטור כאשר : $i=1,2,\ldots,n$ ו- $X=\sum_{i=1}^{n} X_i$.

ניעזר בנוסחאות תוחלת ושונות סכום משתנים מקרים כדי לחשב את התוחלת והשונות של X .

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

$$V(X)=V\left(\sum_{i=1}^{n} X_i\right)=\sum_{i=1}^{n} V(X_i)+2\cdot\sum_{i<j} COV(X_i,X_j)$$

כאשר עבור משתנים אינדיקטורים מתקיים ש :

$$E(X_i)=P(X_i=1)$$
$$V(X_i)=P(X_i=1)\cdot P(X_i=0)$$
$$COV(X_i,X_j)=P(X_i=1,X_j=1)-P(X_i=1)P(X_j=1)$$

דוגמה

מסדרים סידור פרחים לאורך של שורה ובה 40 פרחים.
עבור כל מקום בהסתברות של 0.4 בוחרים פרג עם 4 עלי כותרת ובהסתברות של 0.6 בוחרים כלנית עם 5 עלי כותרת.
נסמן ב- X את <u>מספר השלשות הרצופות</u> בשורה שבהם סך מספר עלי הכותרת של כל שלישיית פרחים הוא זוגי.

(א) חשבו את $E(X)$

(ב) חשבו את $V(X)$

פתרון:

נגדיר אינדיקטור X_i . ערכו של X_i יהיה 1 אם מספר עלי הכותרת של הפרחים במקומות $i, i+1, i+2$ הוא זוגי.
וערכו של X_i יהיה 0 בכל מקרה אחר. כלומר

$$\text{אינדיקטור} \quad X_i = \begin{cases} 1 \longleftarrow & \text{מספר עלים של פרחים במקומות } i, i+1, i+2 \text{ הוא זוגי} \\ 0 \longleftarrow & \text{אחרת} \end{cases} \qquad 1 \le i \le 38$$

ניתן לפרק את X שהוא מ"מ שסופר את <u>מספר השלשות הרצופות</u> בשורה שבהם סך מספר עלי הכותרת של כל שלישיית פרחים הוא זוגי לאינדיקטורים.

$$X = \sum_{i=1}^{38} X_i$$

לדוגמה אם $X_1 = 1$ זה אומר שמספר עלי הכותרת של שלושת הפרחים במקומות 1,2,3 הוא זוגי.
אם $X_{38} = 0$ זה אומר שמספר עלי הכותרת של שלושת הפרחים במקומות 38,39,40 הוא אי זוגי.
כעת לאחר שפירקנו את X לאינדיקטורים נוכל לחשב את $E(X)$:

$$E(X) = \sum_{i=1}^{38} P(X_i = 1)$$

כעת נותר לנו לחשב את $P(X_i = 1)$. כלומר נותר לנו לחשב את ההסתברות שמספר עלי הכותרת בשלישיית פרחים שנבחר הוא זוגי.
עכשיו אנחנו נכנסים לקומבינטוריקה: לפרח מסוג פרג יש 4 עלים ולפרח מסוג כלנית יש 5 עלים. אז כדי לבחור שלישיית פרחים של מספר עלי כותרת זוגי יש כמה אופציות: או לבחור 3 פעמים פרג כלומר 0.4^3 או לבחור פעם אחת פרג כלומר $\binom{3}{1}\cdot 0.4$ ואז לבחור 2 כלניות כלומר $\binom{3}{1}\cdot 0.4\cdot 0.6^2$. לכן סך הכל:

$$P(X_i = 1) = 0.4^3 + \binom{3}{1}\cdot 0.4 \cdot 0.6^2 = 0.496$$

כעת נוכל לחשב את $E(X)$

$$E(X) = \sum_{i=1}^{38} P(X_i = 1) = 38 \cdot 0.496 = 18.848$$

$E(X) = 18.848$

ב) ראשית ניזכר בנוסחא לחישוב $V(X)$ באמצעות אינדיקטור:

$$V(X) = \sum_{i=1}^{n} V(X_i) + 2\sum_{i<j} cov(X_i, X_j)$$

במקרה שלנו $n=38$ כי יש 38 שלשות אפשריות ולכן

$$V(X) = \sum_{i=1}^{38} V(X_i) + 2\sum_{i<j} cov(X_i, X_j)$$

נחשב כל אחד מהחלקים ה- בנוסחא בנפרד.

שונות של אינדיקטור

$$V(X_i) = \underbrace{P(X_i=1)}_{\text{0.496 חישבנו בסעיף א}}\underbrace{(1-P(X_i=1))}_{1-0.496=0.504} = 0.496\cdot 0.504 = 0.249984$$

כעת נחשב את השונות המשותפת $cov(X_i, X_j)$

את התוחלת המשותפת נצטרך להפריד ל- 3 כמות של מקרים.

(1) $j = i+1$ (יש 37 מקרים)

סכום עלי הכותרת של רביעיית פרחים הוא זוגי (הבהרתי מספרים כך להבין שזאת רביעייה)

$$P(X_i=1, X_{i+1}=1) = \underbrace{0.4^4}_{\text{כולם פרג}} + \underbrace{0.4^2\cdot 0.6^2}_{\text{2 פרג, 2 כלנית}} + \underbrace{\binom{2}{1}\cdot 0.4\cdot 0.6^3}_{\text{1 פרג, 3 כלנית}} = 0.256$$

שונות משותפת של אינדיקטור

$$cov(X_i, X_{i+1}) = \underbrace{P(X_i=1, X_{i+1}=1)}_{0.256} - \underbrace{P(X_i=1)\cdot P(X_{i+1}=1)}_{(0.496\cdot 0.496)} = 0.009984$$

(2) $j = i+2$ (יש 36 מקרים)

סכום עלי הכותרת של חמישיית פרחים הוא זוגי.

$$P(X_i=1, X_{i+2}=1) = \underbrace{0.4^5}_{\text{5 פרג}} + \underbrace{0.4\cdot 0.6^4}_{\text{1 פרג, 4 כלנית}} + \underbrace{4\cdot 0.4^2\cdot 0.6^3}_{\text{2 פרג, 3 כלנית}} + \underbrace{2\cdot 0.4^3\cdot 0.6^2}_{\text{3 פרג, 2 כלנית}} = 0.2464$$

שונות משותפת של אינדיקטור

$$cov(X_i, X_{i+2}) = \underbrace{P(X_i=1, X_{i+2}=1)}_{0.2464} - \underbrace{P(X_i=1)\cdot P(X_{i+2}=1)}_{(0.496\cdot 0.496)} = 0.000384$$

(3) לכל $j > i+2$ המשתנים X_i, X_j בת"ל ולכן $cov(X_i, X_j) = 0$

סיכום:

$$V(X_i) = 0.249984 \qquad cov(X_i, X_j) = (37\cdot 0.009984 + 36\cdot 0.000384)$$

ולכן:

$$V(X) = \sum_{i=1}^{38} V(X_i) + 2\sum_{i<j} cov(X_i, X_j)$$

$$= 38\cdot 0.249984 + 2(37\cdot 0.009984 + 36\cdot 0.000384) = 10.265 \Rightarrow V(X) = 10.265$$

דוגמה

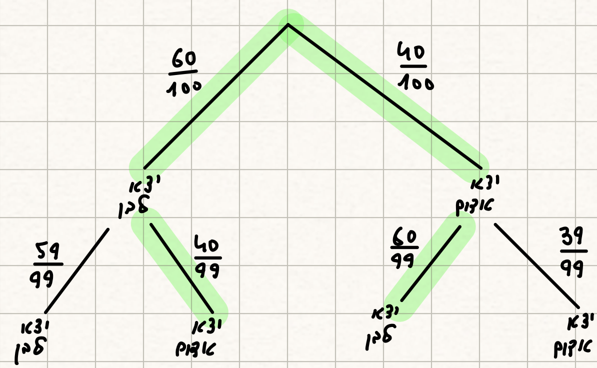
בחניון יש 100 מכוניות - 40 אדומות ו- 60 לבנות. בכל דקה נבחרת באקראי מכונית והיא מוצאת מהחניון.

(א) ידוע שאחרי 2 דקות נותרו בחניון 59 מכוניות לבנות. מה הסיכוי שהמכונית הראשונה שהוצאה הייתה אדומה?

(ב) יהי X מ"מ שסופר את מספר המכוניות האדומות שיצאו מהחניון ב- 50 הדקות הראשונות. חשב את $E(X)$.

פתרון:

(א) בסעיף זה מדברים על מה קורה אחרי 2 דקות, כלומר אחרי 2 הוצאות.



$$P\left(\text{מכונית ראשונה שיצאה היא אדומה} \;\middle|\; \text{ידוע שאחרי 2 הוצאות יש 59 מכוניות לבנות}\right) = \frac{\frac{40}{100}\cdot\frac{60}{99}}{\left(\frac{40}{100}\cdot\frac{60}{99}\right)+\left(\frac{60}{100}\cdot\frac{40}{99}\right)} = \frac{1}{2}$$

(ב) נתון X משתנה מקרי שסופר את מספר המכוניות האדומות שיצאו מהחניון אחרי 50 דקות. צריך לחשב $V(X)$

נגדיר X_i אינדיקטור.

X_i יהיה שווה ל- 1 כאשר יצאה מכונית אדומה מהחניון ו- X_i יהיה שווה ל- 0 כאשר יצאה מכונית לבנה מהחניון.

בחניון יש 40 מכוניות אדומות ו- 60 לבנות. בכל דקה מוציאים מכונית אחת (אדומה/לבנה). שואלים על 50 דקות.

$$\underset{0\le i\le 40}{X_i} = \begin{cases} 1 & \leftarrow \text{יצאה מכונית אדומה} \\ 0 & \leftarrow \text{יצאה מכונית לבנה} \end{cases}$$

כעת ניתן לפרק את המ"מ X שסופר את מספר המכוניות האדומות שיצאו מהחניון אחרי 50 דקות בעזרת אינדיקטורים

$$X = \sum_{i=0}^{50} X_i$$

כעת נותר לנו לחשב את $P(X_i=1)$ כלומר את ההסתברות שיצאה מכונית אדומה וזה $\frac{4}{10}$. לכן התוחלת היא:

$$E(X) = \sum_{i=0}^{50} P(X_i=1) = 50\cdot\frac{4}{10} = 20$$