שיעור 9 | שונות | 24/12/17

תוחלת של משתנה $X \sim NB(n,p)$

X- הוא סכום של n משתנים $G(p)$. לכל אחד מהם יש תוחלת $\frac{1}{p}$ לכן $E(X)=\frac{n}{p}$.

מדדים לפיזור של התפלגות

1. רעיון ראשון של מדד לפיזור: ממוצע הסטיות מהתוחלת: $\Sigma P(X=k)\cdot\left(k-E(X)\right)$. הרעיון הזה נופל, כי ממוצע הסטיות מהתוחלת הוא תמיד שווה ל-0. מתקיים:

$$\Sigma P(X=k)\left(k-E(X)\right)=\Sigma P(X=k)\cdot k-\Sigma P(X=k)\cdot E(X)=E(X)-E(X)=0$$

2. רעיון אחר שהוא רעיון טוב: ממוצע הערך המוחלט של הסטיות: כאן הסטיות לא מתקזזות. אבל הדבר שאיתו עובדים זה ממוצע ריבועי הסטיות: $\boldsymbol{E\left(\left(X-E(X)\right)^2\right)}$.

<u>הגדרה</u>: $Var(X)=E\left(\left(X-E(X)\right)^2\right)$, זוהי השונות של משתנה X.

❖ **דוגמא**

מהי השונות של משתנה $X\sim U[1,6]$ (משתנה קובייה)?

➢ פתרון:

לפי הגדרת השונות:

$$Var(X)=\Sigma P(X=k)\left(X-E(X)\right)^2=$$

$$\frac{1}{6}(1-3.5)^2+\frac{1}{6}(2-3.5)^2+\frac{1}{6}(3-3.5)^2+\frac{1}{6}(4-3.5)^2+\frac{1}{6}(5-3.5)^2+\frac{1}{6}(6-3.5)^2=\frac{35}{12}$$

נוסחה מרכזית לחישוב שונות: $V(X)=Var(X)=E(X^2)-E^2(X)$

נראה שהשונות שווה לביטוי הזה:

$$V(X)=E\left(\left(X-E(X)\right)^2\right)=E\left(X^2-2E(X)X+E^2(X)\right)=E(X^2)-E(2E(X)X)+E^2(X)$$

$$=E(X^2)-2E(X)E(X)+E^2(X)=E(X^2)-2E^2(X)+E^2(X)=E(X^2)-E^2(X)$$

❖ **דוגמא**

חישוב שונות של משתנה קובייה לפי הנוסחא:

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

$$E^2(X)=\left(E(X)\right)^2=3.5^2$$

$$E(X^2)=\Sigma P(X=x)x^2=\frac{1}{6}\cdot 1^2+\frac{1}{6}\cdot 2^2+\frac{1}{6}\cdot 3^2+\frac{1}{6}\cdot 4^2+\frac{1}{6}\cdot 5^2+\frac{1}{6}\cdot 6^2=\frac{1+4+9+25+36}{6}$$

למעשה עבור על משתנה $X\sim U[a,b]$, מתקיים: $V(X)=\frac{(b-a+1)^2-1}{12}$

אז באמת למשתנה קובייה יש שונות: $\frac{35}{12}=\frac{(6-1+1)^2-1}{12}$.

שונות של אינדיקטור בעל הסתברות p

$$P(X=1)=p, \qquad P(X=0)=q=1-p$$

1. בדרך ראשונה:

$$E(X)=p\cdot 1+q\cdot 0=p$$

לפי הגדרת השונות:

$$V(X)=E\left(\left(X-E(X)\right)^2\right)=q(0-p)^2+p(1-p)^2=q\cdot p^2+p\cdot q^2=pq(p+q)=pq\cdot 1=pq$$

2. בדרך שנייה:

לפי הנוסחה:

$$V(X)=E(X^2)-E^2(X)=(p\cdot 1^2+q\cdot 0^2)-p^2=p-p^2=p(1-p)=pq$$

לינאריות השונות:

$$V(aX+b)=V(aX)=E((aX)^2)-E^2(aX)=$$

$$E(a^2X^2)-E(aX)E(aX)=a^2E(X^2)-aE(X)\cdot aE(X)-a^2E^2(X)=a^2\left(E(X^2)-E^2(X)\right)=a^2\cdot V(X)$$

שונות של סכום משתנים

$$V(X+Y)=E\left(\left(X+Y-E(X+Y)\right)^2\right)=$$

$$E\left(\left(X+Y-E(X)-E(Y)\right)^2\right)=E\left(\left(X-E(X)+\left(Y-E(Y)\right)\right)^2\right)=$$

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

$$E\left(\left(X-E(X)\right)^2\right)+E\left(\left(Y-E(Y)\right)^2\right)+2\cdot E\left(\left(X-E(X)\right)\left(Y-E(Y)\right)\right)=$$

$$V(X)+V(Y)+2\cdot E\left(\left(X-E(X)\right)\left(Y-E(Y)\right)\right)$$

אנו רואים ששונו הסכום שווה לסכום השוניות ועוד יצור חדש.

הגדרה:

$$Cov(X,Y) = E\left(\left(X - E(X)\right)\left(Y - E(Y)\right)\right)$$

ולכן:

$$V(X+Y) = V(X) + V(Y) + 2Cov(X,Y)$$

ל-$Cov(X,Y)$ קוראים גם **השונות המשותפת** של X ו-Y. אם $Cov(X,Y) = 0$ אז אומרים ש-X,Y בלתי מתואמים.

נוסחה לחישוב $Cov(X,Y)$

$$Cov(X,Y) = E\left(\left(X - E(X)\right)\left(Y - E(Y)\right)\right) =$$

$$E\left(XY - X \cdot E(Y) - E(X) \cdot Y + E(X)E(Y)\right) =$$

$$E(XY) - E\left(X \cdot E(Y)\right) - E(E(X) \cdot Y) + E(X)E(Y) =$$

$$E(XY) - E(Y)E(X) - E(X)E(Y) + E(X)E(Y) =$$

$$E(XY) - E(X)E(Y)$$

טענה: $Cov(X,Y) = Cov(Y,X)$

הסבר: $Cov(X,Y) = E(XY) - E(X)E(Y) = E(YX) - E(Y)E(YX) = Cov(Y,X)$

טענה: אם X,Y הם בלתי תלויים אז $Cov(X,Y) = 0$.

הסבר: $Cov(X,Y) = E(XY) - E(X)E(Y)$. נסתכל על $E(XY)$:

$$E(XY) = \Sigma_X \Sigma_Y P(X=x, Y=y) \cdot x \cdot y$$

אם הם בלתי תלויים אז ניתן לכתוב: $\Sigma_X \Sigma_Y P(X=x, Y=y) \cdot x \cdot y = \Sigma_X \Sigma_Y P(X=x) \cdot x \cdot P(Y=y) \cdot y$

וזה שווה: $\Sigma_X P(X=x)x \cdot \Sigma_Y P(Y=y)y = E(X) \cdot E(Y)$.

ואז: $Cov(X,Y) = E(X) \cdot E(Y) - E(X)E(Y) = 0$. וקבלנו שאם הם בלתי תלויים $Cov(X,Y) = 0$.

נראה שבכיוון ההפוך זה לא נכון:

נראה ש-$Cov(X,Y) = 0$ לא גורר בהכרח אי תלות. נראה זאת כל ידי דוגמא.

- **דוגמא:**

$$P(X=0,Y=0)=P(X=-1,Y=1)=P(X=1,Y=1)=\frac{1}{3}$$

X,Y הם תלויים כי $(Y=1|X=0)=0\neq P(Y=1)$.

או קצת אחרת: $P(X=0,Y=1)=0\neq P(X=0)P(Y=0)$.

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

X סימטרי סביב 0 ולכן $E(X)=0$. ולכן: $Cov(X,Y)=E(XY)=\frac{1}{3}\cdot 0\cdot 0+\frac{1}{3}\cdot(-1)1+\frac{1}{3}1\cdot 1$.

<u>**טענה**</u>: $Cov(X,X)=Var(X)$

הסבר: $Cov(X,X)=E(XX)-E(X)E(X)=E(X^2)-E^2(X)=Var(X)$.

תכונות נוספות:

1. $Cov\left(\Sigma_i X_i,\Sigma_j Y_j\right)=\Sigma_{i,j}Cov\left(X_i,Y_j\right)$
2. $Cov(aX+b,cY+d)=a\cdot c\cdot Cov(X,Y)$
3. שונות של סכום משתנים:

$$Var(\Sigma_{i=1}^{n}X_i)=\Sigma_i V(X_i)+2\Sigma_{i<j}Cov\left(X_i,Y_j\right)$$

למשל:

$$V(X+Y+Z)=V(X)+V(Y)+V(Z)+2Cov(X,Y)+2Cov(X,Z)+2Cov(Y,Z)$$

שונות של משתנה $\boldsymbol{X\sim Bin(n,p)}$

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

X_i – אינדיקטורים בלי הסתברות p.

$$V(X)=V(\Sigma_{i=1}^{n}X_i)=\Sigma_{i=1}^{n}V(X_i)+2\Sigma_{i<j}Cov\left(X_i,X_j\right)=\Sigma_{i=1}^{n}pq+0=npq$$

שונות של משתנה $\boldsymbol{X\sim HG(n;a,b)}$

זהו סכום של אינדיקטורים. יש n אינדיקטורים וכל אינדיקטור הוא להוצאת כדור לבן בהוצאה מסויימת.

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

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

$$V(X_i) = \frac{a}{a+b} \cdot \frac{b}{a+b}$$

יש $2\binom{n}{2}$ זוגות של אינדיקטורים שצריך לחשב את ה-Cov ביניהם. יש $\binom{n}{2}$ זוגות של משתנים וכל זוג סופרים את ה-Cov ביניהם.

$$Cov(X_i, X_j) = E(X_i \cdot X_j) - E(X_i)E(X_j)$$

מתקיים:

$$E(X_i, X_j) = P(X_i = 1, X_j = 1) = \frac{a}{a+b} \cdot \frac{a-1}{a+b-1}$$

גישה אחרת:

$$E(X_i, X_j) = P(X_i = 1, X_j = 1) = \frac{\binom{a}{2}}{\binom{a+b}{2}}$$

קיבלנו: $Cov(X_i, X_j) = \frac{a}{a+b} \cdot \frac{a-1}{a+b-1} - \left(\frac{a}{a+b}\right)^2$. ולכן:

$$V(X) = n \cdot \frac{a}{a+b} \cdot \frac{b}{a+b} + 2\binom{n}{2} \cdot \left[\frac{a}{a+b} \cdot \frac{a-1}{a+b-1} - \left(\frac{a}{a+b}\right)^2\right] =$$

$$n \cdot \frac{a}{a+b} \cdot \frac{b}{a+b} + \left[1 - \frac{n-1}{a+b-1}\right]$$

בעיית המזכירה הרשלנית

מזכירה שולחת באקראי n מכתבים ל-n אנשים.

X – מספר המכתבים שיגיעו ליעדם הנכון.

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

X_i – אינדיקטור לכך שהמכתב ה-i הגיע ליעדו.

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

$$V(X_i) = \frac{1}{n}\left(1 - \frac{1}{n}\right) = \frac{1}{n} \cdot \frac{n-1}{n}$$

$$Cov(X_i, X_j) = E(X_i \cdot X_j) - E(X_i) \cdot E(X_j) = E(X_i \cdot X_j) - \frac{1}{n} \cdot \frac{1}{n}$$

$$E(X_i)\cdot E(X_j)=P(X_i=1,X_j=1)=\frac{1}{n}\cdot\frac{1}{n-1}=\frac{(n-2)!}{n!}$$

$$Cov(X_i,X_j)=E(X_i\cdot X_j)-E(X_i)\cdot E(X_j)=\frac{1}{n}\cdot\frac{1}{n-1}-\frac{1}{n^2}=\frac{1}{n}\left(\frac{1}{n-1}-\frac{1}{n}\right)=\frac{1}{n^2(n-1)}$$

$$V(X)=n\cdot\frac{1}{n}\cdot\frac{n-1}{n}+2\binom{n}{2}\frac{1}{n^2(n-1)}=\frac{n-1}{n}+\frac{1}{n}=1$$

❖ **שאלה:**

מטילים 10 קוביות תקינות הטלות בלתי תלוית. יהי X – מספר הרצפים של 6,6 שיתקבלו. מהו $V(X)$?

➢ פתרון:

X_i – אינדיקטור לרצף במקומות ה-$i,i+1$.

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

$$V(X_i)=\frac{1}{36}\cdot\left(1-\frac{1}{36}\right)=\frac{1}{36}\cdot\frac{35}{36}$$

יש 9 אינדיקטורים כאלה (לא יכול להיות רצף במקום ה-10) ולכן:

$$\Sigma V(X_i)=9\cdot\frac{1}{36}\cdot\frac{35}{36}$$

בין X_i ו-X_j שאינם שכנים אין תלות ולכן השונות המשותפת שביניהם היא 0. Cov שונה מ-0 הוא רק כאשר יש להם הטלה משותפת.

$$Cov(X_i,X_{i+1})=E(X_i,X_{i+1})-E(X_i)\cdot E(X_{i+1})=$$

$$P(X_i=1,X_{i+1}=1)-\frac{1}{36}\cdot\frac{1}{36}=\left(\frac{1}{6}\right)^3-\frac{1}{36}\cdot\frac{1}{36}$$

יש 8 שלשות ולכן יש 8 זוגות של X_i,X_{i+1}. לכן:

$$V(X)=9\cdot\frac{1}{36}\cdot\frac{35}{36}+2\cdot 8\cdot\left[\left(\frac{1}{6}\right)^3-\frac{1}{36}\cdot\frac{1}{36}\right]$$

❖ **שאלה**:

מבצעים 10 הטלות ב"ת של מטבע תקין. מהי שונות מספר הרצפים hth?

➢ פתרון:

X – מספר הרצפים

X_i – אינדיקטור לרצף החל מהמקום ה-i. יש 8 אינדיקטורים.

$$V(X_i)=\frac{1}{8}\cdot\frac{7}{8},\qquad \Sigma V(X_i)=8\cdot\frac{1}{8}\cdot\frac{7}{8}=\frac{7}{8}$$

יש תלות רק בין שלשות שיש להם חיתוך לא ריק.

נתסכל על $Cov(X_i,X_{i+1})$:

$$Cov(X_i, X_{i+1}) = E(X_i, X_{i+1}) - E(X_i) \cdot E(X_{i+1}) = E(X_i, X_{i+1}) - \frac{1}{8} \cdot \frac{1}{8} = 0 - \frac{1}{64}$$

כי לא יתכן שהחל מהמקום ה-i יהיה לי: hth ובמקום ה-$i+1$ יהיה לנו גם hth.

יש זוגות X_i, X_{i+1} כמספר הרביעיות, שזה 7.

כעת נסתכל על $Cov(X_i, X_{i+2})$:

$$Cov(X_i, X_{i+2}) = E(X_i, X_{i+2}) - E(X_i) \cdot E(X_{i+2}) = \left(\frac{1}{2}\right)^5 - \frac{1}{8} \cdot \frac{1}{8} = \frac{1}{64}$$

כמה זוגות X_i, X_{i+2} יש לנו? כמספר החמישיות, שזה 6.

לסיכום:

$$V(X) = 8 \cdot \frac{1}{8} \cdot \frac{7}{8} + 2 \cdot 7\left(-\frac{1}{64}\right) + 2 \cdot 6 \cdot \frac{1}{64}$$