4.1.15

הסתברות - שיעור אחת עשרה

שונות של משתנה אינדיקטורי

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

$$P(X=1)=p$$

כבר הראינו ש $E(X)=P(X=1)\cdot 1 \; (+P(X=0)\cdot 0)=p$

חישוב הנוסחה לפי הגדרה:

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

$$=(1-p)\cdot p^2+p(1-p)^2=p(1-p)\left[p+(1-p)\right]=p(1-p)$$

למשל אם $p=\frac{1}{2}$ אז $V(X)=\frac{1}{2}\left(1-\frac{1}{2}\right)=\frac{1}{4}$.

חישוב השונות לפי אינדיקטור לפי הנוסחה:

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

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

שונות של סכום משתנים מקריים

X,Y שני משתנים מקריים בעלי תוחלות μ_x, μ_y

$\mu_Y=E(Y),\ \mu_X=E(X)$

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

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

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

$$=V(X)+V(Y)+2E(X-\mu_X)(Y-\mu_Y)$$

נגדיר: $cov(X,Y)=E\left((X-\mu_X)(Y-\mu_Y)\right)$ (covarians)

cov נקרא שונות משותפת.

הראינו שמתקיים:

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

נוסחה לחישוב $cov(X,Y)$

כלומר: $cov(X,Y)=E(X\cdot Y)-E(X)\cdot E(Y)$

או $cov(X,Y)=E(X\cdot Y)-\mu_x\cdot\mu_y$

הוכחה לפי הגדרה

$$cov(X,Y)\overset{\uparrow}{=}E((X-\mu_x)(Y-\mu_y))=E(XY-\mu_y X-\mu_x Y+\mu_x\cdot\mu_y)$$

$$=E(X\cdot Y)-E(\mu_y\cdot X)-E(\mu_x\cdot Y)+\mu_x\cdot\mu_y=$$

$$=E(X\cdot Y)-\mu_y\cdot\mu_x-\mu_x\cdot\mu_y+\mu_x\cdot\mu_y=E(X\cdot Y)-\mu_x\cdot\mu_y$$

תרגיל לחישוב cov

נתונים X,Y עם התפלגות משותפת

מצאו $cov(X,Y)$

X \ Y	2	4	P_X
1	0.1	0.2	0.3
3	0.3	0.4	0.7
P_Y	0.4	0.6	

$cov(X,Y)=E(X\cdot Y)-E(X)-E(Y)$

$E(X)=0.3\cdot 1+0.7\cdot 3=2.4$

$E(Y)=0.4\cdot 2+0.6\cdot 4=3.2$

מכפילים "מסקלות" כפול ההסתברות.

$E(X\cdot Y)=0.1\cdot 1\cdot 2+0.2\cdot 4\cdot 1+0.3\cdot 3\cdot 2+0.4\cdot 4\cdot 3$

$=0.2+0.8+1.8+4.8=7.6$

$cov(X,Y)=E(XY)-E(X)-E(Y)=7.6-2.4-3.2=$

$V(X+Y)=V(X)+V(Y)+2\,cov(X,Y)$

באופן דומה כמו שקיבלנו לגבי שני משתנים X,Y

מתקיים לגבי סכום n משתנים

$$V\left(\sum_{i=1}^{n}X_i\right)=\sum_{i=1}^{n}V(X_i)\underset{\text{חיבור}}{+}\sum_i\sum_j cov(X_i,X_j)$$

שאלה ג:

נתונים 3 משתנים X, Y, Z

$V(X)=8 \qquad V(Y)=5 \qquad V(Z)=9$

$cov(X,Y)=1 \qquad cov(X,Z)=2 \qquad cov(Y,Z)=-1$

מה יהיה $V(X+Y+Z)$?

$$V(X+Y+Z)=V(X)+V(Y)+V(Z)+2\cdot cov(X,Y)+2cov(X,Z)$$
$$+2cov(Y,Z)=8+5+9+2\cdot 1+2\cdot 2+2\cdot(-1)$$

תכונות של שונות משותפת cov

א. $cov(X,Y)=cov(Y,X)$

ב. $cov(X,X)=V(X)$

ג. $cov(aX+b, cY+d)=ac\cdot cov(X,Y)$

ד. $cov(\sum_i X_i, \sum_j Y_j)=\sum_i\sum_j cov(X_i,Y_j)$

ה. אם X,Y ב"ת אז $cov(X,Y)=0$ (לא בכיוון ההפוך).

הוכחות:

א.
$$cov(X,Y)=E(XY)-E(X)\cdot E(Y)=E(YX)-E(X)\cdot E(Y$$
$$=cov(YX)$$

ב.
$$cov(X,X)=E(XX)-E(X)\cdot E(X)=E(X^2)-E^2(X)=V(X)$$

ג.
$$cov(aX+b, cY+d)=E((aX+b)(cY+d))-E(aX+b)\cdot E(cY+d)$$
$$=E(acXY)+E(adX)+E(bcY)+bd-[(aE(X)+b)(cE(Y)+d)]$$
$$=acE(XY)+ad\cdot \cancel{E(X)}+bc\cancel{E(Y)}+\cancel{bd}-[acE(X)\cdot E(Y)+ad\cancel{E(X)}$$
$$+bc\cancel{E(Y)}+\cancel{bd}]=acE(X\cdot Y)-acE(X)\cdot E(Y)=$$
$$=ac(E(XY)-E(X)E(Y))=ac\cdot cov(X,Y)$$

ה. אם X ו Y הם ב"ת אז $cov(X,Y)=0$

הוכחה:

$$cov(X,Y)=E(X\cdot Y)-E(X)\cdot E(Y)$$

$$E(X\cdot Y)=\sum_{x}\sum_{y}P(X=x,Y=y)\cdot xy$$ ב"ת

$$=\sum_{x}\sum_{y}\left(P(X=x)\cdot P(Y=y)\right)\cdot x\cdot y=\sum_{x}\sum_{y}P(X=x)\cdot x\cdot P(Y=y)\cdot y$$

$$=\sum_{x}\left(P(X=x)\cdot x\sum_{y}P(Y=y)\cdot y\right)=E(X)\cdot E(Y)$$

לכן אם X,Y ב"ת (בלתי תלויים). $E(X\cdot Y)=E(X)\cdot E(Y)$

אזי $cov(X,Y)=E(X\cdot Y)-E(X)\cdot E(Y)$

$=0.$

שונות של משתנה $X\sim Bin(n,p)$

$X=\sum_{i=1}^{n}X_i$ כאשר X_i הם אינדיקטורים בעלי הסתברות p.

$$V(X_i)=p(1-p)=p\cdot q$$

אזי $V(X)=n\cdot p\cdot q$

שונות של משתנה $X\sim HG(n,a,b)$

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

X הוא $\sum_{j=1}^{n}X_j$ כאשר עבור כל j, X_j הוא אינדיקטור עם סיכוי $\frac{a}{a+b}$ אזי $V(X_j)=\frac{a}{a+b}\cdot\frac{b}{a+b}$

לכן $\sum V(X_j)=n\cdot\frac{a}{a+b}\cdot\frac{b}{a+b}$ אבל השונות כוללת גם סכום

ה cov: עבור כל $cov(X_j,X_i)=cov(X_1,X_2)$

אזי $V(X)=n\cdot\frac{a}{a+b}\cdot\frac{b}{a+b}+2\binom{n}{2}cov(X_1,X_2)$

$$cov(X_1,X_2)=E(X_1,X_2)-E(X_1)\cdot E(X_2)$$

$$E(X_1)=E(X_2)=\frac{a}{a+b}$$

X_1 ו X_2 יכולים לקבל רק ערכים 0 או 1 ולכן

המכפלה שלהם יכולה לקבל רק ערכים 0 או 1.

לכן $E(X_1\cdot X_2)=P(X_1=1,X_2=1)$

$$E(X_1\cdot X_2)=P(X_1=1,X_2=1)\cdot 1+P(X_1X_2=0)\cdot 0$$

$$P(X_1=1,X_2=1)=\frac{a}{a+b}\cdot\frac{a-1}{a+b-1}$$

$$E(X_1\cdot X_2)=\frac{a}{a+b}\cdot\frac{a-1}{a+b-1}$$

$$cov(X_1,X_2)=E(X_1\cdot X_2)-E(X_1)\cdot E(X_2)=\frac{a}{a+b}\cdot\frac{a-1}{a+b-1}-\left(\frac{a}{a+b}\right)^2$$

$$V(x)=V\left(\sum_{i=1}^{n}X_i\right)=n\cdot V(X_1)+2\binom{n}{2}\cdot cov(X_1,X_2)=$$

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

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

שונות של $X\sim NB(n,p)$

תזכורת: מספר שלבי ניסוי ראשון $Y\sim G(p)$ אז $V(Y)=\frac{q}{p^2}$

X מייצג שלבי הניסוי סכום של n ~~בלתי~~ משתנים גאומטריים

ב"ת $G(p)$. לכן $V(x)=n\cdot\frac{q}{p^2}$.

שונות מספר האנשים שמוצאים לזוגם אחרי ההגרלה הרשלנית.

X - מספר האנשים שמוצאים לזוגם.

$X=\sum_{i=1}^{n} X_i$ כאשר X_i הוא אינדיקטור לכך שהאדם ה-i יצא לזוגתו.

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

X_i הוא אינדיקטור עם סיכוי $\frac{1}{n}$

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

עבור כל i,j: $\operatorname{cov}(X_i,X_j)=\operatorname{cov}(X_1,X_2)$

לכן $V(X)=n\cdot V(X_1)+2\binom{n}{2}\cdot\operatorname{cov}(X_1,X_2)$

$$V(X_1)=\frac{1}{n}\left(1-\frac{1}{n}\right)$$

$$\operatorname{cov}(X_1,X_2)=E(X_1,X_2)-\underbrace{E(X_1)}_{\frac{1}{n}}\cdot\underbrace{E(X_2)}_{\frac{1}{n}}$$

$$E(X_1\cdot X_2)=P(X_1=1,X_2=1)=\frac{1}{n}\cdot\frac{1}{n-1}$$

בדרך אחרת:

$$E(X_1\cdot X_2)=\frac{(n-2)!}{n!}=\frac{1}{n}\cdot\frac{1}{n-1}$$

לכן $\operatorname{cov}(X_1,X_2)=\frac{1}{n(n-1)}-\frac{1}{n^2}=\frac{1}{n}\left[\frac{1}{n-1}-\frac{1}{n}\right]$

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

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

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

גם השונות היא 1.

$$V(X)=1$$

אני מבולבל. מספר הזוגי וגם מספר מספר הזוגי.

מה שונית מספר הצדדים שלנו זהה אותו כ״ת.

$$V(x)+V(Y) = 2V(x) = 2\cdot\frac{1}{2}\cdot\frac{1}{2} = \frac{1}{2}$$

האם זה בהכרח נכון או רק מקרי?

סעיף א:

אם כן מקבל בכל מקרה תוצאה הפוכה ממני יש מה

השוויונות?

$$\underset{\text{תמיד}}{X+Y=1}$$

כי בכל מקרה אחד מהיינו מקבל 1 (בכיוון אחד מהיינו)

$$V(X+Y)=0$$

$$V(X+Y)= V(X)+V(Y)+2\cdot cov(X,Y)$$

$$= V(X)+V(Y)+2\cdot cov(X,1-X)=$$

$$= V(X)+V(Y)+2cov(X,-X)=$$ ← הוצאנו את המינוס בגלל הלינאריות.

$V(X)=V(Y)$

$$=V(X)+V(X)-2cov(X,X)$$

$$=V(X)+V(X)-2\cdot V(X)=0$$

אם נניח שכן מתקיים ממני את התוצאה מהי

$$V(X+Y)= V(X+X)=V(2X)=2^2\cdot V(X)$$

במקרה כזה $cov(X,Y)=cov(X,X)=V(X)$.

לכן גם $V(X+Y)= V(X)+V(Y)+2cov(X,Y)=$

$$=V(X)+V(X)+2cov(X,X)=$$

$$=V(X)+V(X)+2\cdot V(X)=4V(X)$$

יהיו X ו- Y שני משתנים מקריים ב"ת שווי התפלגות מהו $E(X+Y)$?

תשובה $E(X+Y)=2E(X)$

מהו $V(X+Y)$?

$$V(X+Y)=V(X)+V(Y)+2\cdot\underbrace{cov(X,Y)}_{\text{ב"ת}}=2V(X)$$

מהו $E(X+X)$? תשובה $2E(X)$

מהו $V(X+X)$?

נראה ביתר פירוט:

$$V(X+X)=V(2X)=2^2V(X)=4V(X)$$

בדרך אחרת:

$$V(X+X)=V(X)+V(X)+2cov(X,X)=V(X)+V(X)+2V(X)$$

$$=4\cdot V(X)$$

באופן כללי אם X_i סדרת משתנים ב"ת אז: (ושווי התפלגות).

$$V\left(\sum_{i=1}^{n}X_i\right)=\sum V(X_i)=n\cdot V(X_1)$$

$$V(n\cdot X_1)=n^2\,V(X_1)$$

אם יש n משתנים ב"ת שווי התפלגות ~~עם~~ שונות סופית σ^2 לכל אחד, אז כאשר $n\to\infty$, שונות הממוצע שלהם שואפת לאפס. כך אם נסכום הרבה משתנים כאלה אז השונות הממוצעת ישאף לאפס.

סכום המשתנים האלו $\sum_{i=1}^{n}X_i$ הממוצע שלהן הוא $\frac{\sum_{i=1}^{n}X_i}{n}$ נראה שכאשר $n\to\infty$ אז שונות הממוצע שואפת לאפס.

$$V\left(\frac{\sum_{i=1}^{n}X_i}{n}\right)=\frac{1}{n^2}\cdot V\left(\sum_{i=1}^{n}X_i\right)$$

$$=\frac{1}{n^2}\cdot\sum V(X_i)=\frac{1}{n^2}\cdot n\cdot V(X_1)=\frac{V(X_1)}{n}$$

ואכן כאשר $n\to\infty$ זה הביטוי שואף לאפס.

הערה:

אם משתנים ~~בלתי תלויים~~ אזי $cov=0$

וקיימים גם בלתי תלויים.

אלה משתנים בעלי שונות משותפת של אפס.

דוגמא:

מטילים מטבע סדרה של 10 הטלות ב"ת של מטבע הוגן. מהי תוחלת ומהי שונות מספר הרצפים HH?

יהי X מספר הרצפים HH.

* $X=\sum X_i$ כאשר X_i הוא אינדיקטור לכך שבמקום i מתחיל רצף.

$1\le i\le 9$ רצף יתחיל אם במקום מסוים i לא יהיה רצף של לפחות שניים.

$\frac{1}{2}$ $\frac{1}{2}$ *ב'

2 + * תלוי $\quad E(X_i)=\frac{1}{4} \qquad E(X)=\frac{9}{4}$

$$V(X_i)=\frac{1}{4}\cdot\frac{3}{4}=(q\cdot p)=\frac{3}{16}$$

$$V(X)=\sum V(X_i)+2\sum cov(X_i,X_j)$$

$$\sum V(X_i)=9\cdot\frac{3}{16}$$

אם הפרש בין i ל-j הוא לפחות 2 אזי $cov(X_j,X_i)=0$

ואילו ~~א~~ בין X_i ל-X_{i+1} יש cov שונה מאפס.

$$cov(X_i,X_{i+1})=cov(X_1,X_2)$$

$E(X_1\cdot X_2)$

$\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{8}$

$$cov(X_1,X_2)=\underbrace{E(X_1\cdot X_2)}_{\frac{1}{8}}-\underbrace{E(X_1)}_{\frac{1}{4}}\cdot\underbrace{E(X_2)}_{\frac{1}{4}}=\frac{1}{8}-\frac{1}{16}=\frac{1}{16}$$

$$Var(X)=9\cdot\frac{3}{16}+8\cdot2\cdot\frac{1}{16}=9\cdot\frac{3}{16}+8\cdot2\cdot\frac{1}{16}$$

כעת מבצעים 10 הטלות ב"ת ומחפשים את מספר
הרצפים של hth.

תוחלת מספר הרצפים היא $8(\frac{1}{2})^3=1$

כי X_i היא אינד' הצלחה הסת' $(\frac{1}{2})^3$ $(1\le i\le 8)$.

מהו $cov(X_j,X_i)$?

אם המרחק בין הזוגות הוא לפחות 3 אז $cov(X_i,X_j)=0$.
כי הם ב"ת.

$$E(X_i)=\frac{1}{8}$$

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

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

אך יכול להיות ~~של~~ שמתחיל עם מקום i ושם מתחיל
$i+1$.

$$cov(X_i,X_{i+2})=(\tfrac{1}{2})^5-\frac{1}{8}\cdot\frac{1}{8}>0=\frac{1}{64}$$

יש 7 רצפים של $i,i+1$ ויש 6 רצפים של $i,i+2$.

לכן

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