זוג משתנים אקראיים

מטרה: לאפיין ניסוי אקראי בעל שתי תוצאות (בדידות או רציפות), והקשר בין התוצאות האלה.

משתנים אקראיים בדידים

PDF **התפלגות משותפת** (joint distribution)

$$p_{XY}[x_k, y_j] = p[X = x_k, Y = y_j]$$

משתנים, זוג, זוג ערכים, הסתברות לזוג ערכים

$$\{X = x_k\} \cap \{Y = y_j\}$$

דרך נוספת לרישום ההסתברות

תכונות:

הסתברות $\in[0,1]$

$$0 \leqslant p_{XY}[x_k, y_j] \leqslant 1 \ \forall i, k$$

$$\sum_{j,k} p_{XY}[x_k, y_j] = 1$$

דוגמה:

עבור תוצאת של זריקה של 2 קוביות, מתקבל הביטוי

$$p_{XY}[x_k, y_j] = \Pr[X = i] \Pr[Y = j] = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$$

$$F_{XY}(x, y) = p(X \leqslant x, Y \leqslant y)$$ CDF

$$\boxed{E[X] = \sum_i x_i p_X[x_i]} \rightarrow E[XY] = \sum_j \sum_k x_k y_j p_{XY}[x_k, y_j]$$

$$E[g(X, Y)] = \sum_j \sum_k g(x_k, y_j) p_{XY}[x_k, y_j]$$

תוחלת

$g(x,y)$ - בוני ערך, הסתברות לזוג

דוגמה:

תוצאות 5 אפשריות

x_k	y_j	$p_{XY}[x_k, y_j]$
1	1	1/5
1	2	1/5
1	3	1/5
2	1	1/5
2	2	1/5

$$\boxed{F_{XY}(x, y) = p(X \leqslant x, Y \leqslant y)}$$

* $F_{XY}(1.5, 2.5) = ?$

$$\{x_k \leqslant 1.5\} \cap \{y_j \leqslant 2.5\} = \underbrace{(1, 1), (1, 2)}$$

כל התוצאות שעומדות בתנאי

$$F_{XY}(1.5, 2.5) = p_{XY}[1, 1] + p_{XY}[1, 2] = \frac{2}{5}$$

*

$$E[X] = \sum_i x_i p_X[x_i]$$

$$= 1 \cdot p_X[1] + 2 \cdot p_X[2] = 1 \cdot \frac{3}{5} + 2 \cdot \frac{2}{5} = \frac{7}{5}$$

$$E\left[X^2\right] = \sum_i x_i^2 p_X[x_i]$$

$$= 1^2 \cdot p_X[1] + 2^2 \cdot p_X[2] = 1 \cdot \frac{3}{5} + 4 \cdot \frac{2}{5} = \frac{11}{5}$$

$$\mathrm{Var}[X] = E\left[X^2\right] - E^2[X] = \frac{11}{5} - \left(\frac{7}{5}\right)^2 = \frac{6}{25}$$

התפלגות שולית

$$p_{XY}[x_k, y_j] \rightarrow p_X[x_k], p_Y[y_j]$$

מלכה:

$$F_X(x) = F_{XY}(x, \infty)$$

$$F_Y(y) = F_{XY}(\infty, y)$$

$$p_X[x_k] = \sum_j p_{XY}[x_k, y_j]$$

$$p_Y[y_j] = \sum_k p_{XY}[x_k, y_j]$$

x_k	y_j	$p_{XY}[x_k, y_j]$
1	1	1/5
1	2	1/5
1	3	1/5

$p_X[1]$ * התפלגות שולית

$$p_X[1] = \sum_j p_{XY}[1, y_j] = p_{XY}[1, 1] + p_{XY}[1, 2] + p_{XY}[1, 3] = \frac{3}{5}$$

$$p_Y[3] = \sum_k p_{XY}[x_k, 3] = \frac{1}{5}$$

$$p_X[2] = p_{XY}[2,1] + p_{XY}[2,2] = \frac{2}{5}$$

*

$$E[XY] = \sum_j \sum_k x_k y_j p_{XY}[x_k, y_j] =$$

$$= \underbrace{1 \cdot 1 \cdot \frac{1}{5}}_{\text{שורה } 1} + \underbrace{1 \cdot 2 \cdot \frac{1}{5}}_{\text{שורה } 2} + 1 \cdot 3 \cdot \frac{1}{5} + 2 \cdot 1 \cdot \frac{1}{5} + 2 \cdot 2 \cdot \frac{1}{5} = \frac{12}{5}$$

אי תלות סטטיסטית

מטרה: לאפיין מצב של תוצאות ניסוי בלתי תלויות.

משתנים X, Y נקראים בלתי תלויים סטטיסטית **אם ורק אם** מתקיים

דוגמה: X, Y ב"ת $X, Y \sim Ber(p)$

$$\boxed{p_X[k] = \begin{cases} 1 - p & k = 0 \\ p & k = 1 \end{cases}}$$

משתנים X,Y נקראים בלתי תלויים סטטיסטית **אם ורק אם** מתקיים

$$p_{XY}[x_k,y_j]=p_X[x_k]p_Y[y_j]$$
$$F_{XY}(x,y)=F_X(x)F_Y(y)$$

התפלגות משותפת $\Rightarrow$ התפלגות שולית. תמיד

התפלגות שולית $\Leftrightarrow$ התפלגות משותפת רק במקרה של משתנים בלתי תלויים.

$$\boxed{p_X[k]=\begin{cases}1-p & k=0\\ p & k=1\end{cases}}$$

$$p_{XY}[x_k,y_j]=?$$

התפלגות משותפת

$$p_{XY}[0,0]=p_X[0]p_Y[0]=(1-p)^2$$
$$p_{XY}[0,1]=p_X[0]p_Y[1]=(1-p)p$$
$$p_{XY}[1,0]=p_X[1]p_Y[0]=p(1-p)$$
$$p_{XY}[1,1]=p_X[1]p_Y[1]=p^2$$

התפלגות שוליות

$$p_X[0]=p_{XY}[0,0]+p_{XY}[0,1]=(1-p)^2+(1-p)p=1-p$$
$$p_X[1]=p_{XY}[1,0]+p_{XY}[1,1]=(1-p)p+p^2=p$$

$$\boxed{p_X[x_k]=\sum_j p_{XY}[x_k,y_j]}$$

משתנים אקראיים רציפים

$$F_{XY}(x,y)=p(X\leqslant x,Y\leqslant y)$$ CDF

$$f_{XY}(x,y)=\frac{\partial^2F_{XY}(x,y)}{\partial x\partial y}$$ PDF

$$E[g(X,Y)]=\iint_{-\infty}^{\infty}g(x,y)f_{XY}(x,y)dxdy$$ תוחלת

התפלגות שוליות

$$f_X(x)=\int_{-\infty}^{\infty}f_{XY}(x,y)dy$$
$$f_Y(y)=\int_{-\infty}^{\infty}f_{XY}(x,y)dx$$
$$F_X(x)=F_{XY}(x,\infty)$$
$$F_Y(y)=F_{XY}(\infty,y)$$

תכונות:

$$f_{XY}(x,y)\geqslant 0$$
$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f_{XY}(x,y)dxdy=1=F_{XY}(\infty,\infty)$$

$$F_{XY}(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}f_{XY}(s,p)\,dp\,ds$$

$$E[XY]=\iint_{-\infty}^{\infty}xyf_{XY}(x,y)dxdy$$ מקרה פרטי חשוב

בלתי תלויים $\Rightarrow$

$$f_{XY}(x,y)=f_X(x)f_Y(y)$$
$$F_{XY}(x,y)=F_X(x)F_Y(y)$$

התפלגות אחידה - דוג' משתנים בלתי תלויים X,Y

$$f_{XY}(x,y)=\begin{cases}\frac{1}{b-a}\cdot\frac{1}{d-c} & a\leqslant x\leqslant b, c\leqslant y\leqslant d\\ 0 & \text{אחרת}\end{cases}$$

$X\sim U[a,b]$

$Y\sim U[c,d]$

חישוב התפלגות שולית

$$f_X(x)=\int_{-\infty}^{\infty}f_{XY}(x,y)dy$$
$$=\begin{cases}\frac{1}{b-a}\int_c^d\frac{1}{d-c}dy=\frac{1}{b-a} & a\leqslant x\leqslant b\\ 0 & \text{אחרת}\end{cases}$$

* $E[X]=?$

* $\Rightarrow \boxed{g(x,y)=x}$

$$\Rightarrow E[X]=\iint_{-\infty}^{\infty}g(x,y)f_{XY}(x,y)dxdy$$
$$=\iint x\cdot\frac{1}{b-a}\cdot\frac{1}{d-c}\,dx\,dy$$
$$=\frac{1}{d-c}\cdot\frac{1}{b-a}\int_c^d dy\int_a^b x\,dx=\frac{a+b}{2}$$

תכונות של תוחלת

$$E[aX+bY]=aE[X]+bE[Y]$$

תוחלת עבור משתנים בלתי תלויים

$$g(x,y)=g_1(x)g_2(y)$$ separatable

$$E[g_1(X)g_2(Y)]=E[g_1(X)]\,E[g_2(Y)]$$

$$E[XY]=E[X]\,E[Y]$$

$$E[XY]=E[X]E[Y]=\frac{a+b}{2}\cdot\frac{c+d}{2}$$

שונות משותפת

$E[XY] = E[X]\,E[Y]$

שונות משותפת

מטרה: אפיון קשר לינארי בין זוג משתנים

תרגיל הקדמה:

$\mathrm{Var}[X+Y]$

$Z = X+Y$

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

$$\boxed{E[aX+bY] = aE[X] + bE[Y]}$$

$a = b = 1$

$(a+b)^2 = a^2 + 2ab + b^2$

$$\mathrm{Var}[X+Y] = E\Big[(\overbrace{X+Y}^{Z} - \overbrace{E[X+Y]}^{Z})^2\Big]$$

$$= E\Big[\big(\underbrace{(X - E[X])}_{a} + \underbrace{(Y - E[Y])}_{b}\big)^2\Big]$$

$$= \underbrace{E\Big[\overbrace{(X - E[X])^2}^{a^2}\Big]}_{\mathrm{Var}[X]} + \underbrace{E\Big[\overbrace{(Y - E[Y])^2}^{b^2}\Big]}_{\mathrm{Var}[Y]}$$

$$\overbrace{+\,2\underbrace{E\big[(X - E[X])\,(Y - E[Y])\big]}_{\mathrm{Cov}[X,Y]}}^{2ab}$$

שונות משותפת (covariance) (הגדרה 3.10):

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

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

$\boxed{E[E[X]] = E[X]}$ – קבוע

$\boxed{E[XE[X]] = E[X]\,E[X] = E^2[X]}$

$E[b] = b$

הוכחה

$$\mathrm{Cov}[X,Y] = E\Big[\big(X - E[X]\big)\big(Y - E[Y]\big)\Big]$$
$$= E[XY] - E[XE[Y]] - E[Y\underbrace{E[X]}_{\text{קבוע}}] + E[X]\,E[Y]$$
$$= E[XY] - E[X]E[Y]$$

תכונות

$\mathrm{Cov}[X,X] = \mathrm{Var}[X]$

$\mathrm{Cov}[X,Y] = \mathrm{Cov}[Y,X]$ ש'מ' סדר X,Y

$\mathrm{Cov}[X,a] = 0$

$\mathrm{Cov}[aX,bY] = ab\mathrm{Cov}[X,Y]$

$\mathrm{Cov}[X,Y] = \mathrm{Cov}[X+a, Y+b]$

דוגמת הקדמה

$\mathrm{Var}[X \pm Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] \pm 2\mathrm{Cov}[X,Y]$

משתנים בלתי תלויים

$\mathrm{Cov}[X,Y] = 0$

$\mathrm{Var}[aX+bY] = a^2\mathrm{Var}[X] + b^2\mathrm{Var}[Y]$

מקדם קורלציה

Pearson product-moment correlation coefficient

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

* $E[XY] = 0$ אורתוגונליים

* שונות משותפת מנורמלת

מטרה: למדוד קשר לינארי בין X,Y

$$\rho_{XY} = \frac{\mathrm{Cov}[X,Y]}{\sqrt{\mathrm{Var}[X]\,\mathrm{Var}[Y]}}$$

$|\rho_{XY}| \leqslant 1$

* עבור $\rho_{XY} = \pm 1$ משתנים הינם בעלי תלות לינארית

* עבור X,Y חסרי קורלציה מתקיים

אין קשר לינארי

$\mathrm{Cov}[X,Y] = \rho_{XY} = 0$

דוגמה: חסרה דוגמה עם טבלה $\frac{1}{5}$

$$\mathrm{Cov}[X,Y] = E[XY] - E[X]\,E[Y] = \frac{12}{5} - \frac{7}{5}\cdot\frac{9}{5} = -\frac{3}{25}$$

$$\rho_{XY} = \frac{-\frac{3}{25}}{\sqrt{\frac{6}{25}\cdot\frac{14}{25}}} = -\frac{\sqrt{21}}{14} \approx -0.32732683535398857189914622 8$$

X,Y ב"ת

!!

חסרי קורלציה

כיוון אחד בלבד פרט למקרה מסוים בהמשך

$\boxed{\text{שונות תמיד חיובית} \quad \mathrm{Var}[X] \geq 0}$

דוגמה:

חשב $\mathrm{Var}[X]\,, \mathrm{Var}[Y]\,, \mathrm{Var}[X+Y]\,, \mathrm{Cov}[X,Y]\,, \rho_{XY}$ עבור

$X \sim U[0,1]$

קשר

חשב $\mathrm{Var}[X], \mathrm{Var}[Y], \mathrm{Var}[X+Y], \mathrm{Cov}[X,Y], \rho_{XY}$ עבור

$$X \sim U[0,1]$$
$$Y = -X \leftarrow$$

קשר לינארי בין X,Y

$$\boxed{\mathrm{Var}[aX] = a^2 \mathrm{Var}[x]}$$

$\boxed{a=-1} \Rightarrow \boxed{a^2=1}$

$Y = aX$

$$\mathrm{Var}[Y] = \mathrm{Var}[aX] = a^2 \mathrm{Var}[x] = \mathrm{Var}[x]$$

פתרון:

ע"פ נוסחה של התפלגות אחידה

$$\mathrm{Var}[X] = \mathrm{Var}[Y] = \frac{(b-a)^2}{12} = \frac{1}{12}$$

הצבה פשוטה

$$\mathrm{Var}[X+Y] = \mathrm{Var}[X-X] = 0$$

$$\mathrm{Cov}[X,Y] = \mathrm{Cov}[X,-X] \leftarrow$$ הצבה

$$= E[X(-X)] - E[X]\,E[-X] \leftarrow$$ חישוב ע"פ הגדרה

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

$$= -\mathrm{Var}[X]$$

$$\boxed{\sqrt{x^2} = |x| \qquad \frac{x}{|x|} = \mathrm{sign}(x)}$$

$$\rho_{XY} = -\frac{\mathrm{Var}[X]}{\sqrt{\mathrm{Var}[X]\,\mathrm{Var}[X]}} = -\frac{\mathrm{Var}[X]}{|\mathrm{Var}[X]|} = -1$$

דוגמה: מלרה בלתי תלויים $\Leftarrow$ חסרי קורלציה - החץ הוא רק בכיוון אחד

חשב $\mathrm{Cov}[X,Y]$ עבור $X \sim U[-1,1], Y = X^2$

הגדרה

$$\mathrm{Cov}[X,Y] = E[XY] - \underbrace{E[X]}_{0}\,E[Y]$$

$$\frac{1+(-1)}{2} = 0$$

$$E[XY] = E[X^3] = \int_{-1}^{1} \frac{1}{1-(-1)} x^3 dx = 0$$

הצבה

חישוב ע"פ הגדרה

$$\boxed{E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx}$$

$$g(x) = x^3$$

סיכום: X,Y תלויים וחסרי קורלציה.

התפלגות אחידה

$$\boxed{X \sim U[a,b] \qquad f_X(x) = \begin{cases} \frac{1}{b-a} & a \leqslant x \leqslant b \\ 0 & \text{אחרת} \end{cases}}$$

$a = -1, \ b = 1$