משתנה מקרי דו מימדי

נתונים לנו שני משתנים מקריים המוגדרים על אותו מרחב מדגם, ואנו רוצים לעסוק בהסתברויות משותפות שלהם.
למשל, אנחנו מטילים מטבע הוגן 3 פעמים. נסמן ב- X את מספר הפעמים שיצא עץ בשתי ההטלות הראשונות ונסמן ב- Y את מספר העצים שיצא בשתי ההטלות האחרונות.
אנחנו יודעים לתאר את X ו- Y טוב מאוד, כל אחד בנפרד. כעת, נתחיל לתאר אותם ביחד - באותה טבלה.
אצלנו בדוגמה, הערכים האפשריים של X הם $0,1,2$ והערכים האפשריים של Y הם $0,1,2$.
בטבלה, נשים את הערכים האפשריים של X בעמודות ואת של Y בשורות. נקבל, במקרה שלנו:

$Y \backslash X$	0	1	2	$P(Y)$
0	$\frac{1}{8}$	$\frac{1}{8}$	0	$\frac{1}{4}$
1	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$
2	0	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$
$P(X)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	1

את ההסתברויות של $X=k$ ושל $Y=m$ אנחנו יודעים לחשב כי אנחנו יודעים לתאר כל מ"מ בקיף בנפרד.
למשל $P(Y=0)=\frac{1}{4}$ זאת ההסתברות שבהטלת 2 מטבעות מספר העצים יהיה 0.
הקוביות "הפנימיות" מתארות את ההסתברויות של חיתוכים - $P(X=k \cap Y=m)$. למשל, הקוביה שמתאימה לעמודה 1 ושורה 2 נרשום את ההסתברות של המאורע $X=1 \cap Y=2$. אם כן, נחשב את ההסתברויות האלו:

$X=0 \cap Y=0$ - לא יצאו עצים בכלל, וההסתברות לכך היא $\frac{1}{8} = \frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}$.

$X=2 \cap Y=0$ - נשים לב שבהטלה השנייה היא גם אחת משתי ההטלות הראשונות וגם אחת משתי ההטלות האחרונות.
לכן המאורע הזה לא מתרחש - $X=2$ פירושו 2 עצים בשתי ההטלות הראשונות ולכן בפרט בהטלה השנייה
בהכרח יצא עץ. בנוסף ש- $Y=0$ פירושו פלי בשתי ההטלות האחרונות ובפרט פלי בהטלה השנייה. לא יכול
לצאת עץ ופלי באותה הטלה (ההטלה השנייה) ולכן ההסתברות למאורע זה היא 0.

$X=1 \cap Y=0$ - אפשר להסיק שזה $\frac{1}{8}$ מההסתברויות שכבר מילאנו באותה שורה, מבלי לחשב ישירות. אם בכל
זאת רוצים ישירות אז המאורע המתאים הוא עץ, פלי, פלי וזה $\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2} = \frac{1}{8}$

$X=0 \cap Y=2$ - לא יכול להתרחש, בדומה למה שאמרנו על $X=2 \cap Y=0$, כלומר ההסתברות היא 0.
$X=0 \cap Y=1$ - באמצעות העמודה של $X=0$ אפשר להסיק שההסתברות היא $\frac{1}{8}$.
נשים לב - אם נחשב את ההסתברות של $X=1 \cap Y=1$, נוכל למצוא את שאר ההסתברויות ע"י השלמה.
אם: $X=1 \cap Y=1$ פירושו שיצא עץ אחד בשתי ההטלות הראשונות ויצא עץ אחד בשתי ההטלות האחרונות.
המקרים שבהם זה קורה הם: פלי, עץ, פלי + עלי, פלי, עץ = $\frac{1}{4} = \frac{1}{8} + \frac{1}{8}$

2 אחרונות 2 ראשונות 2 אחרונות 2 ראשונות

את השאר אפשר להשלים. באופן כללי אם יודעים את ההסתברויות של X,Y בנפרד צריך לחשב יחסית מעט הסתברויות משותפות ואת השאר אפשר להשלים.

סימון ומינוח:

במקום $X=k \cap Y=m$ נהוג לסמן עם פסיק. למשל, בדוגמה שלנו $P(X=2,Y=0)=0$

הטבלה הזו נקראת טבלת התפלגות משותפת של Y, X.

להתפלגויות של X בנפרד ושל Y בנפרד קוראים התפלגויות שוליות.

נשים לב: ההסתברויות השוליות הן הסכומים של השורות/העמודות המתאימות בטבלה. למשל, בדוגמה שלנו - מהי ההסתברות $P(X=1)$? המאורע $X=1$ הוא בעצם איחוד של שלושה מאורעות זרים:

$X=1$ כאשר $Y=0$, $X=1$ כאשר $Y=1$, $X=1$ כאשר $Y=2$

רצים על כל הערכים האפשריים של Y. כלומר:

$$P(X=k)=\sum_m P(X=k,Y=m)$$

בדוגמה שלנו:

$$P(X=1)=\sum_m P(X=1,Y=m)=P(X=1,Y=0)+P(X=1,Y=1)+P(X=1,Y=2)$$

באופן דומה ההסתברות ש- $P(Y=m)$ היא:

$$P(Y=k)=\sum_k P(X=k,Y=m)$$

דוגמה:

מסובבים שלושה סביבונים. שניים מהם הוגנים - ההסתברות לכל אות היא $\frac{1}{4}$ ובסביבון השלישי ההסתברות לקבלת האות ג' היא $\frac{1}{3}$. בוחרים סביבון באופן אקראי ומסובבים אותו פעמיים. כעת נגדיר שני משתנים מקריים:

X - הערך של X הוא 1 אם בחרנו סביבון הוגן ו-0 אחרת.

Y - מספר הפעמים שקיבלנו את האות ג'. ל-Y יש שלושה ערכים אפשריים $0, 1, 2$

נבנה טבלה:

$Y \backslash X$	0	1	$P(Y)$
0	$\frac{4}{27}$	$\frac{3}{8}$	$\frac{113}{216}$
1	$\frac{4}{27}$	$\frac{1}{4}$	$\frac{43}{108}$
2	$\frac{1}{27}$	$\frac{1}{24}$	$\frac{17}{216}$
$P(X)$	$\frac{1}{3}$	$\frac{2}{3}$	1

כעת, מכיוון שיש 2 סביבונים הוגנים מתוך 3, נוכל לרשום את ההסתברויות השוליות של X.
שנית את ההסתברויות של Y דווקא נוח יותר לחשב באמצעות ההסתברויות המשותפות עם X.

$X=0 \cap Y=0$ – $X=0$ פירושו שהסביבון שאנו מסובבים אינו הוגן, $Y=0$ פירושו שלא יצא ג' בכלל. בסביבון שאינו הוגן, ההסתברות שלא "יצא" ג' היא $\frac{2}{3}$, ולכן ההסתברות ש- $Y=0$ במקרה זה היא $\underbrace{\frac{1}{3}}_{\text{בחירת סביבון לא הוגן}} \cdot \underbrace{\frac{2}{3}}_{\text{תוצאה לא ג'}} \cdot \underbrace{\frac{2}{3}}_{\text{תוצאה לא ג'}}$

$X=0 \cap Y=1$ – שוב סביבון שאינו הוגן ויצא ג' פעם אחת בדיוק מתוך שתיים, יש שני מקרים - יצא ג' בסיבוב ראשון ולא בסיבוב שני או יצא ג' בסיבוב שני ולא בסיבוב הראשון:

$$\frac{1}{3}\left(\frac{1}{3}\cdot\frac{2}{3}+\frac{2}{3}\cdot\frac{1}{3}\right)=\frac{4}{27}$$

$X=0 \cap Y=2$ – באמצעות השלמה של לאחדה.

$X=1 \cap Y=0$ – בחרנו סביבון הוגן בהסתברות $\frac{2}{3}$ ולא יצא ג' בכלל. ההסתברות שלא יצא ג' בסביבון הוגן היא $\frac{3}{4}$. ולכן ב- 2 סיבובים ההסתברות היא:

$$\underbrace{\frac{2}{3}}_{\text{בחירת הוגן}} \cdot \underbrace{\frac{3}{4}}_{\text{לא ג' בהטלה 1}} \cdot \underbrace{\frac{3}{4}}_{\text{לא ג' בהטלה 2}} = \frac{3}{8}$$

$X=1 \cap Y=2$ – בחרנו סביבון הוגן בהסתברות $\frac{2}{3}$ ויצא ג' פעמיים. ההסתברות שיצא ג' בסביבון הוגן היא $\frac{1}{4}$, ולכן ההסתברות היא

$$\underbrace{\frac{2}{3}}_{\text{בחירת הוגן}} \cdot \underbrace{\frac{1}{4}}_{\text{יצא ג'}} \cdot \underbrace{\frac{1}{4}}_{\text{יצא ג'}} = \frac{1}{24}$$

$X=1 \cap Y=1$ – באמצעות השלמה.

ועכשיו אפשר לרשום את ההסתברויות השוליות של Y.

אי תלות :

בהינתן שני משתנים מקריים על אותו מרחב מדגם אנו אומרים ש- X,Y <u>בלתי תלויים</u> אם לכל ערך k של X ולכל ערך m של Y מתקיים :

$$P(X=k,Y=m)=P(X=k)\cdot P(Y=m)$$

במקרה זה המאורעות $X=k$ ו- $Y=m$ הם בלתי תלויים לכל k ולכל m.

ב- 2 הדוגמאות שלנו, המשתנים - אינטואיטיבית - תלויים. ניתן דוגמה שמוכיחה זאת בכל שאלה.
בשאלה של המטבעות:

$$0=P(X=2,Y=0)\neq P(X=2)\,P(Y=0)=\frac{1}{4}\cdot\frac{1}{4}$$

בדוגמה השנייה למשל :

$$\frac{4}{27}=P(X=0,Y=0)\neq P(X=0)\cdot P(Y=0)=\frac{1}{3}\cdot\frac{113}{216}$$

תזכורת:

משתנה מקרי דו מימדי - שני משתנים מקריים X,Y שמוגדרים על אותו מרחב מדגם. במקום לסמן $P(X=k \cap Y=m)$ אנחנו מסמנים $P(X=k, Y=m)$ ומתארים את שני המשתנים יחדיו באותה הטבלה.

עשינו בין היתר את הדוגמה הבאה - מטילים מטבע 3 פעמים, נסמן ב-X את מספר העצים שיצאו ב-2 ההטלות הראשונות וב-Y את מספר העצים שיצאו בשתי ההטלות האחרונות.

קיבלנו את הטבלה הבאה:

$Y \backslash X$	0	1	2	$P(Y)$
0	$\frac{1}{8}$	$\frac{1}{8}$	0	$\frac{1}{4}$
1	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$
2	0	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$
$P(X)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	1

הערכים האפשריים של X,Y הם $0,1,2$ (כמובן, הערכים האפשריים של X יכולים להיות שונים מאלו של Y)

הגדרנו אי תלות בין משתנים מקריים - X,Y הם בלתי תלויים אם לכל k,m מתקיים: $P(X=k, Y=m) = P(X=k)\cdot P(Y=m)$

אז מה היתרון בלתאר את שני המשתנים "בבת אחת"? לעיתים, כאשר שני המשתנים רשומים בבת אחת, יותר נוח לטפל במשתנה שמוגדר כביטוי של שניהם.

למשל בדוגמה שלנו, נסמן: $W=XY$, $Z=X+Y$. אנחנו רוצים לחשב את התוחלת של Z,W ונראה שלעיתים זה יחסית נוח לעשות זאת כשהטבלה איתנו.

ראשית, $E(Z)=E(X+Y)$ ולפי תכונות של תוחלת $E(X+Y)=E(X)+E(Y)$ ורק אפשר לחשב כל אחת מהתוחלות בנפרד ולחבר:

$$E(X)=\sum kP(X=k) = 0\cdot\frac{1}{4}+1\cdot\frac{1}{2}+2\cdot\frac{1}{4} = 1$$

$$E(Y)=\sum mP(Y=m) = 0\cdot\frac{1}{4}+1\cdot\frac{1}{2}+2\cdot\frac{1}{4} = 1$$

$$E(Z)=E(X+Y)=E(X)+E(Y)=1+1=2$$

נשים לב, לא היינו צריכים למצוא במפורש מה הם הערכים האפשריים של Z, מה הן ההסתברויות של כל ערך של Z, הידע על X ועל Y הספיק כדי לחשב את התוחלת של Z.

שנית, הזכרנו בעבר (אך כי לא הוכחנו) שאם X,Y בלתי תלויים אז: $V(X+Y)=V(X)+V(Y)$ אבל אצלנו X,Y הם תלויים ולכן זה לא יעזור. עם הידע שיש לנו כרגע, אנו צריכים לרשום במפורש את הטבלה של Z.

מה הם הערכים האפשריים של Z? $Z=X+Y$ והערכים האפשריים של X,Y הם $0,1,2$ ולכן הערכים האפשריים של Z הם $0,1,2,3,4$. מה ההסתברות לכל אחד?

$Z=0$ פירושו שגם X וגם Y שווים ל-0 וההסתברות לכך היא $\frac{1}{8}$ (לפי הטבלה)

$Z=1$ פירושו $X=0, Y=1$ או $X=1, Y=0$ וההסתברות לכך היא $\frac{1}{8}+\frac{1}{8}=\frac{1}{4}$ (לפי הטבלה)

$Z=2$ פירושו $X=1, Y=1$ או $X=2, Y=0$ או $X=0, Y=2$ וההסתברות לכך היא $\frac{1}{4}+0+0=\frac{1}{4}$ (לפי הטבלה)

$Z=3$ פירושו $X=2, Y=1$ או $X=1, Y=2$ וההסתברות לכך היא $\frac{1}{8}+\frac{1}{8}=\frac{1}{4}$ (לפי הטבלה)

$Z=4$ פירושו $X=2, Y=2$ וההסתברות לכך היא $\frac{1}{8}$ (לפי הטבלה)

לכן:

Z	0	1	2	3	4	total
$P(Z)$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{8}$	1

מכאן:

$$E(Z^2) = \sum k^2 P(Z=k) = 0^2\cdot\frac{1}{8} + 1^2\cdot\frac{1}{4} + 2^2\cdot\frac{1}{4} + 3^2\cdot\frac{1}{4} + 4^2\cdot\frac{1}{8} = \frac{11}{2}$$

$$V(Z) = E(Z^2) - (E(Z))^2 = \frac{11}{2} - (2)^2 = \frac{11}{2} - 4 = \frac{3}{2}$$

נעבור ל- W, נחשב את התוחלת שלו. אנחנו יודעים שלא בהכרח מתקיים $E(XY) = E(X)E(Y)$ ולכן לכאורה אי אפשר לחשב את $E(XY)$ מבלי למצוא את ההתפלגות של XY מפורשות.

מה הם הערכים האפשריים של $W = XY$? הערכים האפשריים של X,Y הם $0,1,2$ ולכן הערכים האפשריים של $W = XY$ הם $0,1,2,4$. מה ההסתברות לכל אחד?

$W=0$ פירושו $XY=0$ כלומר $X=0$ או $Y=0$ ולכן ההסתברות היא:

$$P(W=0) = P(X=0 \cup Y=0) = P(X=0) + P(Y=0) - P(X=0, Y=0) = \frac{1}{4} + \frac{1}{4} - \frac{1}{8} = \frac{3}{8}$$

$W=1$ פירושו $XY=1$ כלומר $X=1$ וגם $Y=1$ וההסתברות לכך היא $\frac{1}{4}$ (לפי הטבלה). $P(W=1) = \frac{1}{4}$

$W=2$ פירושו $XY=2$ כלומר $X=1$ וגם $Y=2$ או $X=2$ וגם $Y=1$ ולכן ההסתברות לכך היא: $P(W=2) = \frac{1}{8} + \frac{1}{8} = \frac{1}{4}$

$W=4$ פירושו $XY=4$ כלומר $X=2$ וגם $Y=2$ ולכן ההסתברות לכך לפי הטבלה היא $\frac{1}{8}$. $P(W=4) = \frac{1}{8}$.

לכן:

W	0	1	2	4	total
$P(W)$	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{8}$	1

$$E(W) = \sum k\cdot P(W=k) = 0\cdot\frac{3}{8} + 1\cdot\frac{1}{4} + 2\cdot\frac{1}{4} + 4\cdot\frac{1}{8} = \frac{10}{8} = \frac{5}{4}$$

שאלה:

האם יש דרך לחשב תוחלת של ביטוי של X,Y מתוך הטבלה המשותפת של X,Y מבלי למצוא את הטבלה של הביטוי?

והאם יש דרך לחשב את השונות של ביטוי של X,Y מתוך הטבלה המשותפת של X,Y מבלי למצוא את הטבלה של הביטוי?

לגבי התוחלת, התשובה היא המשפט הבא.

משפט: תהי f פונקציה של שני משתנים, התוחלת $f(X,Y)$ מחושבת באופן הבא:

$$E(f(X,Y)) = \sum f(k,m)\cdot P(X=k, Y=m)$$

כשהסכום רץ על כל הערכים האפשריים של X,Y.

למשל, בדוגמה האחרונה שלנו, הפונקציה $f(X,Y) = XY$ כך ש- $f(k,m) = km$. נקבל ש-

$$E(XY) = \sum km \cdot P(X=k, Y=m)$$

אצלנו, הערכים האפשריים של X ושל Y הם $0,1,2$ כך שנקבל 9 מחוברים:

$$E(XY) = \sum km \cdot P(X=k, Y=m)$$

$$= 0 \cdot 0 \cdot P(X=0, Y=0) + 0 \cdot 1 \cdot P(X=0, Y=1) + 0 \cdot 2 \cdot P(X=0, Y=2)$$

$$+ 1 \cdot 0 \cdot P(X=1, Y=0) + 1 \cdot 1 \cdot P(X=1, Y=1) + 1 \cdot 2 \cdot P(X=1, Y=2)$$

$$+ 2 \cdot 0 \cdot P(X=2, Y=0) + 2 \cdot 1 \cdot P(X=2, Y=1) + 2 \cdot 2 \cdot P(X=2, Y=2)$$

$$\Rightarrow E(XY) = 0+0+0+0+1 \cdot \frac{1}{4} + 2 \cdot \frac{1}{8} + 0 + 2 \cdot \frac{1}{8} + 4 \cdot \frac{1}{8} = \frac{1}{4}+\frac{1}{4}+\frac{1}{4}+\frac{2}{4} = \frac{5}{4}$$

מה לגבי השונות? נתמקד ב- $V(X+Y)$. כדי לטפל ב- $V(X+Y)$ אנחנו צריכים למושג של שונות משותפת.

שונות משותפת

יהיו X,Y משתנים מקריים. השונות המשותפת של X,Y מסומנת $cov(X,Y)$, ומוגדרת באופן הבא:

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

טענה

באופן מתקיים: $cov(X,Y) = E(XY) - E(X)E(Y)$

<u>הוכחה</u>

נפתח את הסוגריים:

$$cov(X,Y) = E\Big((X-E(X))(Y-E(Y))\Big) = E\Big(XY - XE(Y) - YE(X) + E(X)E(Y)\Big)$$

תוחלת של סכום היא סכום התוחלות

$$\Rightarrow E\Big(XY - XE(Y) - YE(X) + E(X)E(Y)\Big) = E(XY) - E(XE(Y)) - E(YE(X)) + E(E(X)E(Y))$$

כעת, $E(X), E(Y)$ הם תוחלות משמע הם מספרים קבועים ולכן ניתן להוציא אותם החוצה מהתוחלות.

$$E(XY) - E(XE(Y)) - E(YE(X)) + E(E(X)E(Y)) = E(XY) - E(Y) \cdot E(X) - E(X)E(Y) + E(X)E(Y)$$

כעת:

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

למשל, בדוגמה האחרונה, ראינו ש- $E(X)=1,\ E(Y)=1$ וגם $E(XY)=\frac{5}{4}$ ומכאן השונות המשותפת היא:

$$\mathrm{Cov}(X,Y) = E(XY)-E(X)E(Y) = \frac{5}{4}-1\cdot 1 = \frac{1}{4}$$

הערה: מה המשמעות של השונות המשותפת? בקווים כלליים, כאשר השונות המשותפת חיובית, פירוש הדבר שבין המשתנים יש קשר "חיובי" - כאשר אחד מהם גדל, גם השני גדל. כאשר השונות המשותפת שלילית, פירוש הדבר שבין המשתנים יש קשר שלילי, כאשר אחד מהם גדל אז השני קטן.

מה קורה כאשר $\mathrm{Cov}(X,Y)=0$?

במקרה כזה המשתנים נקראים בלתי מתואמים.

משפט: אם X,Y בלתי תלויים אז הם בלתי מתואמים (ההיפך איננו נכון, כלומר יכול להיות שמשתנים הם בלתי מתואמים וכן תלויים)

למשל, נתבונן במשתנים X,Y עם הטבלה הבאה:

Y \ X	2	4	6	8	P(Y)
1	$\frac{1}{4}$	0	0	$\frac{1}{4}$	$\frac{1}{2}$
2	0	$\frac{1}{4}$	$\frac{1}{4}$	0	$\frac{1}{2}$
P(X)	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	1

מצד אחד, מכיוון שיש בתוך הטבלה הסתברויות ששוות ל-0, המשתנים בהכרח תלויים.
מצד שני, נחשב את השונות המשותפת, ראשית:

$$E(X) = 2\cdot\frac{1}{4}+4\cdot\frac{1}{4}+6\cdot\frac{1}{4}+8\cdot\frac{1}{4} = \frac{20}{4} = 5$$

$$E(Y) = 1\cdot\frac{1}{2}+2\cdot\frac{1}{2} = \frac{3}{2}$$

שנית:

$$E(XY) = \sum km\cdot P(X=k,Y=m)$$

$$= 2\cdot1\cdot P\underbrace{(X=2,Y=1)}_{1/4} + 2\cdot2\cdot P\underbrace{(X=2,Y=2)}_{0} + 4\cdot1\cdot P\underbrace{(X=4,Y=1)}_{0} + 4\cdot2\cdot P\underbrace{(X=4,Y=2)}_{1/4}$$

$$+ 6\cdot1\cdot P\underbrace{(X=6,Y=1)}_{0} + 6\cdot2\cdot P\underbrace{(X=6,Y=2)}_{\frac{1}{4}} + 8\cdot1\cdot P\underbrace{(X=8,Y=1)}_{\frac{1}{4}} + 8\cdot2\cdot P\underbrace{(X=8,Y=2)}_{0}$$

$$\Rightarrow E(XY) = \frac{2}{4}+\frac{8}{4}+\frac{12}{4}+\frac{8}{4} = \frac{30}{4} = \frac{15}{2}$$

כעת נחשב את השונות המשותפת:

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

$$\mathrm{Cos}(X,Y) = \frac{15}{2}-5\cdot\frac{3}{2} = \frac{15}{2}-\frac{15}{2} = 0$$

תכונות השונות המשותפת

① $cov(X,Y) = cov(Y,X)$

הוכחה

$$cov(X,Y) = E(XY) - E(X)E(Y) \underset{\text{תכונת תוחלת}}{=} E(YX) - E(Y)E(X) = cov(Y,X)$$

② $cov(X,X) = V(X)$

הוכחה:

$$cov(X,X) = E(XX) - E(X)E(X) = E(X^2) - \big(E(X)\big)^2 = V(X)$$

③ $cov(X,Y+Z) = cov(X,Y) + cov(X,Z)$

הוכחה:

$$cov(X,Y+Z) = E\big(X(Y+Z)\big) - E(X)E(Y+Z) = E(XY+XZ) - E(X)\big(E(Y)+E(Z)\big) = E(XY)+E(XZ) - E(X)E(Y) - E(X)E(Z)$$

$$= \underbrace{E(XY) - E(X)E(Y)}_{cov(X,Y)} + \underbrace{E(XZ) - E(X)E(Z)}_{cov(X,Z)} = cov(X,Y) + cov(X,Z)$$

④ $cov(aX+b, cY+d) = ac \cdot cov(X,Y)$

הוכחה:

$$cov(aX+b, cY+d) = E\big((aX+b)(cY+d)\big) - E(aX+b)E(cY+d)$$

$$= E\big(acXY + bcY + daX + bd\big) - \big(aE(X)+b\big)\big(cE(Y)+d\big)$$

$$= E(acXY) + E(bcY) + E(daX) + E(bd) - \big(acE(X)E(Y) + bcE(Y) + adE(X) + bd\big)$$

$$= acE(XY) + bcE(Y) + adE(X) + bd - acE(X)E(Y) - bcE(Y) - adE(X) - bd$$

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

$$= ac\big(E(XY) - E(X)E(Y)\big) = ac \cdot cov(XY)$$

⑤ הקשר לשונות של $X+Y$ - מתקיים השוויון הבא:

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

הוכחה:

$$V(X+Y) = E\big((X+Y)^2\big) - \big(E(X+Y)\big)^2 = E(X^2 + 2XY + Y^2) - \big(E(X)+E(Y)\big)^2$$

$$= E(X^2) + 2E(XY) + E(Y^2) - \Big(\big(E(X)\big)^2 + 2E(X)E(Y) + \big(E(Y)\big)^2\Big)$$

$$= E(X^2) - \big(E(X)\big)^2 + E(Y^2) - \big(E(Y)\big)^2 + 2E(XY) - 2E(X)E(Y)$$

$$= \underbrace{E(X^2)-\left(E(X)\right)^2}_{V(X)} + \underbrace{E(Y^2)-\left(E(Y)\right)^2}_{V(Y)} + \underbrace{2\left(E(XY)-E(X)E(Y)\right)}_{2cov(X,Y)} = V(X)+V(Y)+2cov(X,Y)$$

מקדם מתאם

יהיו X,Y שני משתנים מקריים. מקדם המתאם של X,Y מסומן $\rho(X,Y)$ והוא מוגדר באופן הבא:

$$\rho(X,Y) = \frac{cov(X,Y)}{\sqrt{V(X)V(Y)}}$$

כפי שאמרנו, הסימן של השונות המשותפת - למקדם המתאם יש את אותו סימן - הסימן הזה מצביע על סוג הקשר בין המשתנים X,Y, כאשר הסימן חיובי פירושו שכאשר X גדל גם Y גדל וכאשר הסימן שלילי פירושו שכאשר X גדל Y קטן.

החלוקה בביטוי $\sqrt{V(X)V(Y)}$ מאפשרת לנו גם לומר על כמה הקשר "חזק" - כלומר עד כמה הקשר בין X ל- Y דומה לקשר לינארי $Y=aX+b$.

למשל, בדוגמה של הטלת מטבע 3 פעמים ראינו ש- $cov(X,Y)=\frac{1}{4}$ כמו כן X,Y הם בעצם משתנים בינומיים - מטילים מטבע פעמיים וסופרים כמה עצים יוצאו כך שבעצם $X\sim Bin(2,\frac{1}{2})$, $Y\sim Bin(2,\frac{1}{2})$ ולכן: $V(X)=V(Y)=np(1-p)=2\cdot\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{4}$

כעת נחשב את מקדם המתאם:

$$\rho(X,Y) = \frac{cov(X,Y)}{\sqrt{V(X)V(Y)}} = \frac{\frac{1}{4}}{\sqrt{\frac{1}{2}\cdot\frac{1}{2}}} = \frac{1}{2}$$

איך אנחנו מכמתים את החוזק של קשר בהתאם למקדם המתאם?

אם $0.7 \le \rho \le 1$ אנו אומרים שהקשר חזק.

אם $0.4 \le \rho \le 0.7$ אנו אומרים שהקשר בינוני.

אם $0 \le \rho \le 0.4$ אנו אומרים שהקשר חלש.

תכונות של מקדם המתאם:

1. $-1 \le \rho \le 1$
2. $\rho=1$ אמ"מ הקשר בין X ל- Y הוא לינארי חיובי, כלומר $Y=aX+b$ עבור $a>0$.
   $\rho=-1$ אמ"מ הקשר בין X ל- Y הוא לינארי שלילי, כלומר $Y=-aX+b$ עבור $a<0$.
3. ככל ש- $|\rho|$ קרוב יותר ל- 1, עוצמת הקשר חזקה יותר.
4. $\rho=0$ אמ"מ $cov(X,Y)=0$ - שבר מתאפס אמ"מ המונה מתאפס.
5. למה שווה $\rho(aX+b,cY+d)$?
   אם ל- a,c יש את אותו סימן, אז: $\rho(aX+b,cY+d)=\rho(X,Y)$
   אם ל- a,c יש סימנים מנוגדים אז: $\rho(aX+b,cY+d)=-\rho(X,Y)$

דוגמה :

בקערה יש n אגסים, m בננות ו- k תפוחים, לפחות 10 מכל סוג.
אנחנו מוציאים 10 פירות מהקערה ללא החזרה.
נסמן ב- X את מספר האגסים שהוצאנו, ב- Y את מספר הבננות שהוצאנו וב- Z את מספר התפוחים שהוצאנו.

(א) חשבו את השונות $V(X+Y+Z)$

נשים לב : $X+Y+Z=10$ כי זה סך כל הפירות שהוצאנו ונתון שהוצאנו בסך הכל 10 פירות. לכן:

$$V(X+Y+Z)=V(10)=0$$

שונות של מספר קבוע היא 0

(ב) עבור $n,m,k=10$ חשבו את $P(X=5,Y=2)$.

אנחנו רוצים לבחור 5 אגסים, 2 בננות ולכן בהכרח 3 תפוחים. יש 10 פירות מכל סוג, ובחירת הפירות נעשית ללא חזרה וללא חשיבות לסדר (פירות מאותו סוג זהים זה לזה) ולכן: $\binom{10}{5}\binom{10}{2}\binom{10}{3}$
סך כל האפשרויות להוציא 10 פירות מתוך ה- 30 שבקערה הוא $\binom{30}{10}$ ולכן ההסתברות המבוקשת היא:

$$P(X=5,Y=2)=\frac{\binom{10}{5}\binom{10}{2}\binom{10}{3}}{\binom{30}{10}}=0.045$$

(ג) חשבו את $\rho(X,Y)$, בהנחה שיש 10 פירות מכל סוג.
לפי הנוסחה:

$$\rho(X,Y)=\frac{cov(X,Y)}{\sqrt{V(X)V(Y)}}$$

גם ל- X וגם ל- Y יש 11 ערכים אפשריים $0,1,\dots,10$. כלומר בטבלה משותפת תהיינה 121 קוביות - ולכן לחשב ישירות מהטבלה את $E(XY)$ בשביל $cov(XY)$ זה לא נעים בכלל. צריך לחפש משהו טריק, כמו בסעיף א'.
ראשית, גם X וגם Y מתפלגים היפרגיאומטרית באופן הבא: $X,Y\sim HG(30,10,10)$ - יש $N=30$ פירות בסך הכל, מתוכם $D=10$ מוחפים (אגסים במקרה של X ובננות במקרה של Y) ואנו מוציאים $n=10$ פירות.
לכן בשביל השונויות של X,Y נוכל להשתמש בשונות הידועה של ההתפלגות ההיפרגיאומטרית

$$V(X)=n\cdot\frac{\frac{D}{N}\cdot\left(1-\frac{N}{D}\right)(N-n)}{N-1}=\frac{10\cdot\frac{10}{3}\cdot\left(1-\frac{10}{30}\right)(30-10)}{30-1}=\frac{4000}{261}$$

$$V(Y)=n\cdot\frac{\frac{D}{N}\cdot\left(1-\frac{N}{D}\right)(N-n)}{N-1}=\frac{10\cdot\frac{10}{3}\cdot\left(1-\frac{10}{30}\right)(30-10)}{30-1}=\frac{4000}{261}$$

מה לגבי $cov(X,Y)$? אנחנו יודעים ש:

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

ולכן:

$$cov(X+Y)=\frac{V(X+Y)-V(X)-V(Y)}{2}\underset{V(X)=V(Y)}{=}\frac{V(x+Y)-2V(X)}{2}$$

כמו כן, על סמך מה שכבר הצגנו $X+Y = 10-Z$ ולכן:

$$V(X+Y) = V(10-Z) = V(-1\cdot Z + 10) \underset{\text{תכונות שונות}}{=} -1^2 V(Z) = V(Z)$$

ושוב גם ל- Z ההתפלגות היפר גיאומטרית כמו ל- X ולכן $V(X)=V(Y)=V(Z)$.

$$\frac{V(X+Y)-2V(X)}{2} = \frac{V(Z)-2V(X)}{2} = \frac{V(X)-2V(X)}{2} = \frac{-V(X)}{2}$$

לסיום:

$$cov(X+Y) = \frac{-V(X)}{2} \qquad V(X) = V(Y) = V(Z)$$

$$\Downarrow$$

$$\rho(X,Y) = \frac{cov(X,Y)}{\sqrt{V(X)V(Y)}} = \frac{\frac{-V(X)}{2}}{\sqrt{V(X)^2}} = \frac{\frac{-V(X)}{2}}{V(X)} = -\frac{1}{2}$$