הסתברות מ' | הרצאה 7 (00094412)

שם: איל

February 29, 2024

נושאי השיעור: שונות, שונות משותפת

נושא ראשון - שונות

תזכורת:

- שונות של משתנה מקרי X היא $Var\left(X\right)=E\left[\left(X-E\left[X\right]\right)^{2}\right]$
- בהרצאה הקודמת אמרנו שזה מדד לגודל סטיות סביב הממוצע (כלומר כלי למדידת פיזור).

הערה: למה המדד שאנחנו בוחרים הוא דווקא שונות, למה הוא מוגדר להיות התוחלת של הסטייה בריבוע ולמה זה מדד שימושי?

- האפשרויות האחרון הן:

1. לנסות למדוד את התוחלת של הסטיות מהממוצע בלי להעלות בריבוע:

$$E\left[X-E\left[X\right]\right]=E\left[X\right]-E\left[X\right]=0$$

2. לנסות למדוד את התוחלת של הסטיות מהממוצע בערך מוחלט:

$$E\left[\left|X-E\left[X\right]\right|\right]$$

- זה מדד אפשרי, אבל קשה לעבוד עם פונקציית הערך המוחלט, לדוגמה כי היא לא גזירה ב-0.

3. $E\left[X\right]$ ו-$Var\left(X\right)$ יהיו:

$$Var\left(x\right),E\left[X\right]=\min,\arg\min\left\{ E\left[\left(X-m\right)^{2}\right]\mid m\in\mathbb{R}\right\}$$

– כאשר $\arg\min$ הוא הערך המינימלי של m שמביא את הביטוי למינימום.

- כלומר האפשרויות האחרות פחות נוחות מאשר שונות.

הערה: כדי "לפצות" על ההעלאה בריבוע הגדרנו את סטיית התקן להיות $SD=\sqrt{Var\left(X\right)}$

הערה: בדרך כלל לא מתקיים $E\left[g\left(X\right)\right]=g\left(E\left[X\right]\right)$.
בפרט, מתקיים

$$\sqrt{Var\left(X\right)}=\sqrt{E\left[\left(X-E\left[X\right]\right)^{2}\right]}\neq E\left[\sqrt{\left(X-E\left[X\right]\right)^{2}}\right]=E\left[\left|X-E\left[X\right]\right|\right]$$

הערה: נשים לב כי בדרך כלל מתקיים:

$$E\left[X^{2}\right]\neq\left(E\left[X\right]\right)^{2}$$

משפט 1. מהרצאה קודמת - תכונות של שונות.

- יהיו X,Y משתנים מקריים בדידים עם תוחלות ושונות מוגדרות.
- יהיו $a,b\in\mathbb{R}$ קבועים.
- מתקיים:

1. תמיד $Var\left(X\right)\geq0$
2. $Var\left(X\right)=0$ אמ"מ $X=E\left[X\right]$, כלומר $P\left(X=E\left[X\right]\right)=1$
3. $Var\left(a\cdot X+b\right)=a^{2}\cdot Var\left(X\right)$
4. אם X,Y בלתי תלויים אז $Var\left(X\pm Y\right)=Var\left(X\right)+Var\left(Y\right)$
5. $Var\left(X\right)=E\left[X^{2}\right]-\left(E\left[X\right]\right)^{2}$

הוכחה.

1. נראה כי $Var\left(X\right)\geq0$:

   - לפי הגדרה מתקיים:

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

– נשים לב כי $\left(X-E\left[X\right]\right)^{2}$ זהו משתנה מקרי אי שלילי, ולפי משפט מהרצאה קודמת מתקיים כי תוחלת של משתנה מקרי אי שלילי היא גם אי שלילית.

2. נראה כי $Var\left(X\right)=0$ אמ"מ $X=E\left[X\right]$:

- לפי משפט מהרצאה קודמת, תוחלת של משתנה מקרי אי שלילי שווה אפס אמ"מ המשתנה המקרי שווה אפס.
  - כלומר $\left(X-E\left[X\right]\right)^{2}=0$
    - ולכן $X=E\left[X\right]$.

3. נראה כי $Var\left(a\cdot X+b\right)=a^{2}\cdot Var\left(X\right)$:

- לפי הגדרה מתקיים:

$$Var\left(a\cdot X+b\right)=E\left[\left(a\cdot X+b-E\left[a\cdot X+b\right]\right)^{2}\right]$$

  - מליניארות התוחלת, ניתן להוציא את b החוצה:

$$=E\left[\left(a\cdot X+b-a\cdot E\left[X\right]+b\right)^{2}\right]$$

$$=E\left[\left(a\cdot X\cancel{+b}-a\cdot E\left[X\right]\cancel{+b}\right)^{2}\right]$$

  - שוב, מלינאריות התוחלת מתקיים:

$$=E\left[a^{2}\cdot\left(X-E\left[X\right]\right)^{2}\right]$$

$$=a^{2}\cdot E\left[\left(X-E\left[X\right]\right)^{2}\right]$$

  - נציב את הגדרת השונות $Var\left(X\right)=E\left[\left(X-E\left[X\right]\right)^{2}\right]$ ונקבל:

$$=a^{2}\cdot Var\left(X\right)$$

- נשים לב כי בפרט מתקיים $Var\left(-X\right)=Var\left(X\right)$

4. נוכיח מאוחר יותר.

5. נראה כי $Var\left(X\right)=E\left[X^{2}\right]-\left(E\left[X\right]\right)^{2}$:

- לפי הגדרה מתקיים:

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

- נפתח את ההעלאה בריבוע:

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

- מליניארות התוחלת מתקיים:

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

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

- נצרף את סעיף 1 ואת סעיף 5 כדי לקבל:

$$Var(X) = E\left[X^2\right] - (E[X])^2 \geq 0$$

$$\Rightarrow E\left[X^2\right] \geq (E[X])^2$$

■

דוגמה 2. שונות של משתנה מקרי בינומי והיפר גאומטרי.

- חשבו שונות של משתנה מקרי בינומי עם n, p והיפר גאומטרי N, G, n

פיתרון עבור המקרה הבינומי:

- נניח כי $X \sim Bin(n, p)$
- **צ"ל:** $Var(X)$.
- <u>דרך א' - לפי הגדרה:</u>
  - התוחלת של משנה בינומי היא $n \cdot P$ ולכן לפי הגדרת התוחלת מתקיים:

$$E\left[(X - E[X])^2\right] = E\left[(X - n \cdot p)^2\right] = \sum_{x=0}^{n} (x - n \cdot p)^2 \cdot P_X(x)$$

- <u>דרך ב':</u>

– לפי המשפט, מתקיים:

$$Var(X) = E\left[X^2\right] - \left(\overbrace{E[X]}^{=n\cdot p}\right)^2$$

$$Var(X) = E\left[X^2\right] - (n\cdot p)^2$$

- <u>דרך ג' - בעזרת שיטת האינדיקטורים:</u>

– נניח כי X סופר את מספר ההצלחות בסדרה של n ניסויי ברנולי בלתי תלויים עם הסתברות p.

– לכל $k = 1, \ldots, n$ נסמן A_k - הצלחה בניסוי ה-k.

– אזי X הוא סכימה של כל האינדיקטורים I_{A_k}:

$$X = \sum_{k=1}^{n} I_{A_k}$$

– ולכן השונות של X היא:

$$Var(X) = Var\left(\sum_{k=1}^{n} I_{A_k}\right)$$

– נשים לב כי I_{A_k} הם אינדיקטורים בלתי תלויים כי כל המאורעות A_k הם בלתי תלויים.

* לכן לפי סעיף 4 במשפט מתקיים:

$$Var(X) = \sum_{k=1}^{n} Var\left(I_{A_k}\right)$$

* מסעיף 5 במשפט מתקיים:

$$Var\left(I_{A_k}\right) = E\left[\left(I_{A_k}\right)^2\right] - \left(E\left[I_{A_k}\right]\right)^2$$

· מכיוון ש-I_{A_k} הוא אינדיקטור, מתקיים $I_{A_k} = \left(I_{A_k}\right)^2$ ולכן:

$$= E\left[I_{A_k}\right] - \left(E\left[I_{A_k}\right]\right)^2$$

* תוחלת של אינדיקטור I_{A_l} היא $P(A_k)$ ולכן מתקיים:

$$= E\left[I_{A_k}\right] - \left(E\left[I_{A_k}\right]\right)^2$$

$$= P\left(A_k\right) - \left(P\left(A_k\right)\right)^2$$

$$= p - p^2$$

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

$$= p \cdot q$$

– כלומר:

$$Var\left(X\right) = \sum_{k=1}^{n} p \cdot q = n \cdot p \cdot q$$

פיתרון עבור המקרה ההיפר גאומטרי:

- נסמן משתנה מקרי $Y \sim HG\left(N,G,n\right)$.
- נפתור בעזרת שיטת האינדיקטורים.
- נניח כי Y סופר את מספר הירוקים במדגם בגודל n ללא החזרה מתוך F ירוקים ו-$N-G$ כחולים.
  – נסדר את הכדורים $k = 1, \ldots, n$
  – נסמן מאורעות B_k - הכדור ה-k במדגם הוא ירוק.
- <u>הערה:</u> אם היינו שולפים עם החזרה אז היה מתקיים $Y \sim Bin\left(n, \frac{N}{G}\right)$
  – היינו מקבלים סיכוי $p = \frac{G}{N}$ לכל כדור ולכן היינו מקבלים $n \cdot p \cdot q = Var$
- כאשר מוציאים עם החזרה, השונות תהיה קטנה יותר.
- נסמן I_{B_k} להיות האינדיקטור שהמאורע B_k קרה.
- נשים לב כי מכיוון שהשליפות תלויות זו בזו, לא בהכרח מתקיים

$$Var\left(Y\right) = Var\left(\sum_{k=1}^{n} I_{B_k}\right) = \sum_{k=1}^{n} Var\left(I_{B_k}\right)$$

- ולכן נשתמש בסעיף 5 של המשפט הראשון:

$$Var\left(Y\right) = E\left[Y^2\right] - \left(\overbrace{E\left[Y\right]}^{=n\cdot\frac{G}{N}}\right)^2$$

– נציב $Y^2 = \left(\sum_{k=1}^{n} I_{B_k}\right)^2$

* נקבל:

$$E\left[Y^2\right] = E\left[\left(\sum_{k=1}^{n} I_{B_k}\right)^2\right]$$

$$= E\left[\sum_{t,k=1}^{n} I_{B_k} \cdot I_{B_t}\right]$$

$$= \sum_{t,k=1,\ \ t=k}^{n} E\left[I_{B_k} \cdot I_{B_t}\right] + \sum_{t,k=1,\ \ t\neq k}^{n} E\left[I_{B_k} \cdot I_{B_t}\right]$$

· יש לנו n איברים עבור המקיימים $t = k$.

· כל אחד מהם מקיים:

$$E\left[I_{B_k} \cdot I_{B_t}\right] = E\left[\left(I_{B_k}\right)^2\right] = E\left[I_{B_k}\right] = P\left(I_{B_k}\right) = \frac{G}{N}$$

· יש לנו $n \cdot (n-1)$ איברים המקיימים $t \neq k$.

· כל אחד מהם מקיים:

$$E\left[I_{B_k} \cdot I_{B_t}\right] = E\left[I_{B_k \cap B_t}\right]$$

$$= P\left(B_k \cap B_t\right) = \frac{G}{N} \cdot \frac{G-1}{N-1}$$

* נציב ונקבל:

$$= n \cdot \frac{G}{N} + n \cdot (n-1)\, \frac{G}{N} \cdot \frac{G-1}{N-1}$$

– כעת קיבלנו:

$$Var\left(Y\right) = \overbrace{\left(n \cdot \frac{G}{N} + n \cdot (n-1)\, \frac{G}{N} \cdot \frac{G-1}{N-1}\right)}^{=E\left[Y^2\right]} - \left(\overbrace{n \cdot \frac{G}{N}}^{=E[Y]}\right)^2$$

* (לא עשינו את החישוב האלגברי אבל) זה שווה ל:

$$Var\left(Y\right)=n\cdot\frac{G}{N}\cdot\left(1-\frac{G}{N}\right)\cdot\frac{N-n}{N-1}$$

- הערה: נשים לב שמתקיים $Var\left(Y\right)\leq Var\left(\text{"}Bin\left(n,\frac{G}{N}\right)\text{"}\right)$

נושא שני - שונות משותפת:

- נרצה להגדיר גודל שייצג את אופן הסטייה המשותף של זוג גדלים אקראיים מהממוצע שלהם.

הגדרה 3. שונות משותפת.

- בהינתן זוג משתנים מקריים X,Y עם תוחלות מוגדרות.
- ה**שונות המשותפת** שלהם (Co-variance) היא:

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

הערה: מה הגודל הזה מייצג?

- הוא מייצג ממוצע של הסטייה (מהממוצע) של גודל המיוצג על ידי X בניסוי, כפול הסטייה (מהממוצע) של גודל המיוצג על ידי Y בניסוי, על פני $n\to\infty$ חזרות של הניסוי.
- בפרט:
  - אם $0\leq Cov\left(X,Y\right)$ אז אינטואיטיבית, זה אומר שלרוב הסטיות של X,Y היא באותו כיוון מהממוצע (כלומר הסטיות של שניהם שליליות או חיוביות).
  - אם $Cov\left(X<y\right)\leq0$ אז הסטיות של X ו-Y (מהממוצע) נוטות להיות בכיוונים מנוגדים (לדוגמה X לרוב יחרוג מעל הממוצע ו-Y יהיה מתחת לממוצע).

הגדרה 4. משתנים מקרים מתואמים חיובית או שלילית

- יהיו X,Y משתנים מקריים עם שונות משותפת מוגדרת.
- אזי:

1. אם $0<Cov\left(X,Y\right)$, נאמר כי X,Y מתואמים חיובית.
2. אם $0>Cov\left(X,Y\right)$, נאמר כי X,Y מתואמים שלילית.
3. אם $0=Cov\left(X,Y\right)$, נאמר כי X,Y לא מתואמים. ניתן לומר "בלתי מתואמים".

משפט 5. תכונות של שונות משותפת.

- יהיו X,Y,Z משתנים מקריים עם תוחלת ושונות משותפת מוגדרת.
- יהיו $a,b,c,d\in\mathbb{R}$ קבועים.
- אזי:

1. השונות המשותפת מוגדרת (כלומר התוחלת בהגדרה של השונות המשותפת קיימת).
   - ומתקיים $\left|Cov\left(X,Y\right)\right|\leq SD\left(X\right)\cdot SD\left(Y\right)$
2. ניתן לחשב את השונות המשותפת על ידי הנוסחה: $Cov\left(X,Y\right)=E\left[X\cdot Y\right]-E\left[X\right]\cdot E\left[Y\right]$
3. מתקיים: $Cov\left(X,Y\right)=Cov\left(Y,X\right)$
4. מתקיים:

$$Cov\left(aX+b,c\cdot Y+d\right)=a\cdot c\cdot Cov\left(X,Y\right)$$

5. מתקיים:

$$Cov\left(X+Y,Z\right)=Cov\left(X,Z\right)+Cov\left(Y,Z\right)$$

6. מתקיים:

$$Cov\left(X,X\right)=Var\left(X\right)$$

7. מתקיים:

$$Var\left(X\pm Y\right)=Var\left(X\right)+Var\left(Y\right)\pm2\cdot Cov\left(X,Y\right)$$

8. אם X,Y בלתי תלויים אז $Cov\left(X,Y\right)=0$
   - הכיוון ההפוך לא בהכרח נכון.
     - כלומר אם שני משתנים מקריים הם בלתי מתואמים אז הם לא בהכרח בלתי תלויים.

דוגמה 6.

- בכד יש G כדורים ירוקים, B כדורים כחולים ו-R כדורים אדומים.
- בוחרים מדגם ללא החזרה של n כדורים.
- מצאו שונות משותפת בין מספר הכדורים הירוקים והכחולים במדגם.

פיתרון:

- נסמן:
  - X = מספר הירוקים במדגם.
  - Y = מספר הכחולים במדגם.
- **צ"ל:** $Cov\left(X,Y\right)$
  - נרצה לשים לב לסימן של $Cov\left(X,Y\right)$.
    - * נזכיר שהסימן של $Cov\left(X,Y\right)$ קובע האם X,Y מתואמים חיובית או שלילית.
- נשים לב שאם נוציא יותר כדורים ירוקים במדגם מהממוצע שלהם, כמות הירוקים תבוא על חשבון הכחולים.
  - לכן לכאורה נצפה ש-X,Y יהיו מתואמים שלילית.
- נשים לב כי X מפולג היפר גאומטרית, $X\sim HG\left(\overbrace{G+B+R}^{=N},G,n\right)$
  - מכאן נובע שהתוחלת של X היא:

$$E\left[X\right]=n\cdot\frac{G}{N}$$

- באותו אופן, Y מפולג היפר גאומטרית $Y\sim HG\left(N,B,n\right)$
  - והתוחלת של Y היא:

$$E\left[Y\right]=n\cdot\frac{B}{N}$$

- <u>דרך א' - שונות משותפת לפי הגדרה:</u>

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

$$=E\left[\left(x-n\cdot\frac{G}{N}\right)\left(Y-n\cdot\frac{B}{N}\right)\right]$$

  - זו תוחלת של פונקציה של משתנים מקריים:

$$=\sum_{x,y}\left(x-n\cdot\frac{G}{N}\right)\left(Y-n\cdot\frac{B}{N}\right)\cdot P_{X,Y}\left(x,y\right)$$

– נחשב את $P_{X,Y}(x,y)$, כלומר ההסתברות שיש לנו x כדורים ירוקים ו-y כדורים כחולים במדגם:

$$P_{X,Y}(x,y)=\begin{cases}\frac{\binom{G}{x}\binom{B}{y}\binom{R}{n-x-y}}{\binom{N}{n}} & 0\le x\le G\ ,\ 0\le y\le B,\ 0\le n-x-y\le R\\ 0 & otherwise\end{cases}$$

– נשים לב שזו דרך ארוכה.

• <u>דרך ב' - שימוש במשפט השני:</u>

$$Cov(X,Y)=E[X\cdot Y]-E[X]\cdot E[Y]$$

– בניסיון הקודם חישבנו את $E[X]$ ואת $E[Y]$.

– נשאר לחשב את $E[X\cdot Y]$. מכיוון שהם תלויים נצטרך לחשב לפי הגדרה:

$$E[X\cdot Y]=\sum_{x,y}x\cdot y\cdot P_{X,Y}(x,y)$$

* שוב קיבלנו ביטוי מסובך.

• <u>דרך ג' - נשתמש במשפט השני, סעיף 7:</u>

– לפי סעיף 7, מתקיים:

$$Var(X+Y)=Var(X)+Var(Y)+2\cdot Cov(X,Y)$$

* אנחנו מחפשים את $Cov(X,Y)$ ולכן נעביר אגפים כדי לקבל:

$$Cov(X,Y)=\frac{V(X+Y)-Var(X)-Var(Y)}{2}$$

* בדוגמה הראשונה של ההרצאה חישבנו את השונות של משתנה המתפלג היפר גאומטרית ולכן מתקיים:

$$Var(X)=n\cdot\frac{G}{N}\left(1-\frac{G}{N}\right)\left(\frac{N-n}{N-1}\right)$$

$$Var(Y)=n\cdot\frac{B}{N}\left(1-\frac{B}{N}\right)\left(\frac{N-n}{N-1}\right)$$

* נשים לב כי $X+Y\sim HG\left(N,B+G,n\right)$ כי $X+Y$ הוא מספר הכדורים הכחולים והירוקים במדגם בגודל n.
  · ולכן באופן דומה, השונות של $X+Y$ היא:

$$Var\left(X+Y\right)=n\cdot\frac{B+G}{N}\left(1-\frac{B+G}{N}\right)\left(\frac{N-n}{N-1}\right)$$

– נציב ונקבל:

$$Cov\left(X,Y\right)=\frac{V\left(X+Y\right)-Var\left(X\right)-Var\left(Y\right)}{2}$$

$$=\ldots=\left(-1\right)\cdot n\cdot\frac{G}{N}\cdot\frac{B}{N}\cdot\frac{N-n}{N-1}$$

• כלומר, קיבלנו ש-X,Y מתואמים שלילי כמו שציפינו.

הוכחה. הוכחות לתכונות השונות המשותפת:

1. לא נוכיח. נובע מאי-שוויון קושי שוורץ.
2. נראה כי $Cov\left(X,Y\right)=E\left[X\cdot Y\right]-E\left[X\right]\cdot E\left[Y\right]$:

• לפי הגדרה מתקיים:

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

$$=E\left[XY-E\left[X\right]Y-E\left[Y\right]\cdot X+E\left[X\right]\cdot E\left[Y\right]\right]$$

– ומלינאריות התוחלת מתקיים:

$$=E\left[XY\right]-E\left[X\right]\cdot E\left[Y\right]$$

7. נראה כי $Var\left(X\pm Y\right)=Var\left(X\right)+Var\left(Y\right)\pm2\cdot Cov\left(X,Y\right)$

• לפי הגדרה מתקיים:

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

– מליניארות התוחלת מתקיים:

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

* נפתח את הריבוע ומלינאריות התוחלת נקבל:

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

$$= Var\left(X\right) + Var\left(Y\right) \pm 2 \cdot Cov\left(X, Y\right)$$

8. נראה כי אם X, Y בלתי תלויים אז $Cov\left(X, Y\right) = 0$

• נניח כי X, Y בלתי תלויים

– לפי המשפט השני נקבל:

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

$$= E\left[X\right] \cdot E\left[Y\right] - E\left[X\right] \cdot E\left[Y\right]$$

$$= 0$$

• בפרט, אם X, Y בלתי תלויים אז מתקיים:

$$Var\left(X \pm Y\right) = Var\left(X\right) + Var\left(Y\right) \pm \overbrace{\cancel{2 \cdot Cov\left(X, Y\right)}}^{=0}$$

$$\Rightarrow Var\left(X \pm Y\right) = Var\left(X\right) + Var\left(Y\right)$$

■

הגדרה 7. מקדם מתאם.

- אם X,Y משתנים מקריים עם שונות ותוחלת מוגדרים אז:

$$Corr\left(X,Y\right)=\rho\left(X,Y\right)=\frac{Cov\left(X,Y\right)}{SD\left(X\right)\cdot SD\left(Y\right)}$$

<u>הערה:</u> המשמעות של מקדם המתאם היא ממוצע מכפלת הסטיות של X,Y מהתוחלת שלהם, מנורמלת לפי גודל הסטיות של כל אחד מהמשתנים המקריים בנפרד.

משפט 8. - ללא הוכחה.

- יהיו X,Y משתנים מקריים עם תוחלת ושונות.

1. הסימן של $\rho\left(X,Y\right)$ שווה לסימן של $Cov\left(X,Y\right)$.
2. $-1\leq\rho\left(X,Y\right)\leq1$
   - כלומר מקדם המתאם הוא תמיד בין 1 ל -1
3. מתקיים:
   (א) $\rho\left(X,Y\right)=1$ אם ורק אם $y=a\cdot X+b$ עבור $a>0$ ו-$b\in\mathbb{R}$
   (ב) $\rho\left(X,Y\right)=-1$ אם ורק אם $y=a\cdot x+b$ עבור $a<0$ ו-$b\in\mathbb{R}$
   - נשים לב ש $\rho\left(X,Y\right)=\pm1$ אם ורק אם מתקיים יחס ליניארי בין X ו-Y.

נושא שלישי - תוחלת מותנית.

הגדרה 9. תוחלת מותנית על מאורע.

- בהינתן משתנה מקרי בדיד X ומאורע A שהסתברותו אינה 0, המוגדרים על אותו מרחב הסתברות (Ω,P), נגדיר:

$$E\left(X\mid A\right)=\sum_{x}x\cdot P\left(\left\{X=x\right\}\mid A\right)$$

  - כל זאת בהנחה שהטור מתכנס בהחלט.

- הגודל הזה ייקרא התוחלת המותנית של X בהינתן A.

<u>הערה:</u>

- זאת בדיוק התוחלת של X ביחד למידת המותנית $P_{A}\left(t\right)=P\left(t\mid A\right)$
- ולכן כל הטענות והמשפטים שחלים על תוחלת רגילה תקפים גם פה.

הגדרה 10. תוחלת מותנית על ערכו של משתנה מקרי.

- בהינתן זוג משתנים מקריים בדידים X,Y המוגדרים על אותו (Ω,P):
  - **התוחלת המותנית של** X **בהינתן** $Y=y$ **עבור** $y\in\mathbb{R}$ מוגדרת על ידי:

$$E\left[X\mid Y=y\right]=\begin{cases}E\left[X\mid\left\{Y=y\right\}\right] & P_{Y}\left(y\right)>0\\0 & otherwise\end{cases}$$

* נשים לב כי $E\left[X\mid\left\{Y=y\right\}\right]$ מוגדר להיות $\sum_{x}x\cdot P\left(\left\{X=x\right\}\mid\left\{Y=y\right\}\right)$ בהגדרה הקודמת.

הערה:

- נבחין כי מתקיים:

$$E\left[X\mid Y=y\right]=\sum_{x}x\cdot P_{X|Y}\left(x,y\right)$$

- ונשים לב כי זו פונקציה של $y\in\mathbb{R}$.

דוגמה 11.

- נחזור לדוגמה עם הכדורים הירוקים, כחולים ואדומים.
- נרצה לחשב את התוחלת המותנית של X בהינתן Y, לכל y.
- כלומר **צ"ל:** $E\left[X\mid\left\{Y=y\right\}\right]$

פיתרון:

- כאמור, מתקיים:

$$E\left[X\mid\left\{Y=y\right\}\right]=\sum_{x}x\cdot P_{X|Y}\left(x,y\right)$$

- נחשב את $P_{X|Y}\left(x,y\right)$:
  - אפשר לחשב לפי הגדרה ולקבל:

$$P_{X|Y}\left(x,y\right)=\begin{cases}\frac{P_{X,Y}\left(x,y\right)}{P_{Y}\left(y\right)} & P_{Y}\left(y\right)>0\\0 & otherwise\end{cases}$$

  - ראינו בדוגמה הזו כי

$$P_{X,Y}\left(x,y\right)=\begin{cases}\frac{\binom{G}{x}\binom{B}{y}\binom{R}{n-x-y}}{\binom{N}{n}} & 0\leq x\leq G\ ,\ 0\leq y\leq B,\ 0\leq n-x-y\leq R\\0 & otherwise\end{cases}$$

* וגם Y מפולג היפר גאומטרי ולכן:

$$P_Y(y) = \begin{cases} \dfrac{\binom{B}{y}\binom{N-B}{n-y}}{\binom{N}{n}} & 0 \le y \le B,\ 0 \le n-y \le R \\ 0 & otherwise \end{cases}$$

– נחלק את שתי ההסתברויות ונקבל:

$$\frac{P_{X,Y}(x,y)}{P_Y(y)} = \frac{\dfrac{\binom{G}{x}\cancel{\binom{B}{y}}\binom{R}{n-x-y}}{\cancel{\binom{N}{n}}}}{\dfrac{\cancel{\binom{B}{y}}\binom{N-B}{n-y}}{\cancel{\binom{N}{n}}}}$$

$$\Rightarrow \frac{P_{X,Y}(x,y)}{P_Y(y)} = \frac{\binom{G}{x}\binom{R}{n-x-y}}{\binom{G+R}{n-y}}$$

* כאשר $0 \le x \le G$, $0 \le n-y-x \le R$ ו- $0 \le Y \le B$

– נשים לב כי עבור $0 \le y \le B$ קבוע, אגף ימין היא פונקציית ההסתברות של משתנה מקרי המפולג היפר גאומטרי עם הפרמטרים $G+R=N$, G, $n-y$.

– באופן מקוצר אפשר לכתוב:

$$X \mid Y = y \sim HG\left(\frac{G+R}{N-B}, G, n-y\right)$$

• נזכור שאנחנו מחפשים את התוחלת המותנית $E[X \mid Y=y]$:

$$E[X \mid \{Y=y\}] = \sum_x x \cdot P_{X|Y}(x,y)$$

- נשים לב שזו תוחלת של משתנה מקרי המפולג היפר גאומטרי $HG\left(N-B,G,n-y\right)$ ולכן מתקיים:

$$E\left[X\mid\{Y=y\}\right]=\begin{cases}(n-y)\cdot\frac{G}{N-B} & 0\le y\le B,\ \ 0\le n-y\le N-B\\ 0 & otherwise\end{cases}$$

בדומה להסתברות מותנית, גם עבור תוחלת מותנית יש נוסחה שקושרת את התוחלת המותנית לתוחלת הרגילה:

משפט 12. נוסחת התוחלת השלמה.

- גרסת מאורעות:
  - יהיו $\left(A_k\right)_{k=1}^{\infty}$ מאורעות המהווים חלוקה של Ω כך ש $P\left(A_k\right)>0$
  - יהא X משתנה מקרי עם תוחלת מוגדרת על אותו מרחב הסתברות (Ω,P).
  - אזי התוחלת של X היא:

$$E\left[X\right]=\sum_{k=1}^{\infty}E\left[X\mid A_k\right]\cdot P\left(A_k\right)$$

- גרסת משתנים מקריים:
  - יהיו X,Y משתנים מקריים בדידים על אותו מרחב מדגם כך של-X יש תוחלת מוגדרת.
  - אזי התוחלת של X היא:

$$E\left[X\right]=\sum_{y}E\left[X\mid Y=y\right]\cdot P_Y\left(y\right)$$

דוגמה 13.

- בהמשך לדוגמה הקודמת, חשבו את התוחלת $E\left[X\right]$ באמצעות התוחלת המותנית $E\left[X\mid Y=y\right]$.

פיתרון:

- בסעיף הקודם ראינו ש:

$$E\left[X\mid Y=y\right]=\begin{cases}(n-y)\cdot\frac{G}{N-B} & P_Y\left(y\right)>0\\ 0 & otherwise\end{cases}$$

- בנוסף, ראינו ש-Y מפולג היפר גאומטרי, $Y \sim HG\left(N,B,n\right)$ ואמרנו ש:

$$P_{Y}\left(y\right)=\begin{cases} \frac{\begin{pmatrix}B\\ y \end{pmatrix}\begin{pmatrix}N-B\\ n-y \end{pmatrix}}{\begin{pmatrix}N\\ n \end{pmatrix}} & 0\leq y\leq B,\ \ 0\leq n-y\leq R\\ 0 & otherwise \end{cases}$$

- לכן, לפי המשפט מתקיים:

$$E\left[X\right]=\sum_{y}E\left[X\mid Y=y\right]\cdot P_{Y}\left(y\right)$$

- במקום להציב ולחשב את הסכום באופן ישיר, נגדיר פונקציה g:

$$g\left(y\right)=E\left[X\mid Y=y\right]=\begin{cases} \left(n-y\right)\cdot\frac{G}{N-B} & P_{Y}\left(y\right)>0\\ 0 & otherwise \end{cases}$$

  - עכשיו, קיבלנו:

$$E\left[X\right]=\sum_{y}g\left(y\right)\cdot P_{Y}\left(y\right)$$

    * זו תוחלת של פונקציה של משתנה מקרי, כלומר:

$$E\left[X\right]=E\left[g\left(Y\right)\right]$$

$$=E\left[\begin{cases} \left(n-y\right)\cdot\frac{G}{N-B} & P_{Y}\left(y\right)>0\\ 0 & otherwise \end{cases}\right]$$

· אפשר לוותר את ה-$otherwise$ במקרה הזה כי אם $P_{Y}\left(y\right)\leq0$ אז נקבל שהמחובר שהוא נמצא בו בסכום יהיה 0.

· ולכן:

$$=E\left[\left(n-y\right)\cdot\frac{G}{N-B}\right]$$

$$=\sum_{y}\left(n-y\right)\cdot\frac{G}{N-B}\cdot P_{Y}\left(y\right)$$

$$= \sum_{y} E\left[X \mid Y=y\right] \cdot P_Y\left(y\right)$$

- ולפי ליניאריות של תוחלת מתקיים:

$$= \frac{G}{N-B}\left(n - E\left[Y\right]\right)$$

$$= \frac{G}{N-B}\left(n - n \cdot \frac{B}{N}\right)$$

$$= n \cdot \frac{G}{N}$$

את דרך הפיתרון הזו ניתן להכליל באופן הבא:

הגדרה 14. תוחלת מותנית על משתנה מקרי.

- יהיו זוג משתנים מקריים בדידים X,Y המוגדרים על אותו מרחב מדגם כך של-X יש תוחלת.
- אזי, **התוחלת המותנית של** X **בהינתן** Y (לא בהינתן $Y=y$) היא המשתנה המקרי $g\left(y\right)$, כאשר:

$$g\left(y\right) = E\left[X \mid Y=y\right]$$

- סימון:

$$E\left[X \mid Y\right] = g\left(y\right)$$

משפט 15. תוחלת שלמה באמצעות תוחלת על משתנה מקרי, נוסחת ההחלקה.

- אם X,Y זוג משתנים מקריים בדידים המוגדרים על אותו מרחב מדגם כך של-X יש תוחלת.
  - אז:

$$E\left[X\right] = E\left[E\left[X \mid Y\right]\right]$$

הוכחה. כמו בדוגמה הקודמת. ■

דוגמה 16.

- יהיו $X_1, X_2, \ldots$ משתנים מקריים בדידים עם אותה התפלגות (שווי התפלגות) ותוחלת מוגדרת.
- יהא N משתנה מקרי המקבל ערכים אי שליליים שלמים בלבד ועם תוחלת מוגדרת.
- הניחו כי כל המשתנים המקריים בלתי תלויים.
- נגדיר משתנה מקרי שהוא סכום של N משתנים מקריים:

$$S = \sum_{k=1}^{N} X_k$$

צ"ל: מצאו תוחלת של S.

פיתרון:

- התוחלת המותנית $E\left[S \mid N=n\right]$ שווה ל:

$$E\left[S \mid N=n\right] = E\left[\sum_{k=1}^{N} X_k \mid N=n\right]$$

  – נוכל להחלף את N ב-n בסכימה:

$$= E\left[\sum_{k=1}^{n} X_k \mid N=n\right]$$

  – הנחנו כי $X_1, X_2, \ldots$ בלתי תלויים ולכן מתקיים:

$$= E\left[\sum_{k=1}^{n} X_k\right]$$

    * מליניאריות התוחלת מתקיים:

$$= \sum_{k=1}^{n} E\left[X_k\right]$$

    * מכיוון שכל ה-X שווי התפלגות, נקבל תוחלת מותנית כפונקציה של n:

$$= n \cdot E\left[X_1\right]$$

- נגדיר $g(n) = E[S \mid N = n] = n \cdot E[X_1]$
  - אזי מתקיים:

$$E[S \mid N] = g(N)$$

$$= E[X_1] \cdot N$$

  - ולפי המשפט שראינו עכשיו, מתקיים:

$$E[S] = E[E[S \mid N]]$$

$$= E[E[X_1] \cdot N]$$

$$= E[X_1] \cdot E[N]$$

- לזה קוראים נוסחת וואלד.