<u>הסתברות וסטטיסטיקה – חוברת מתרגל</u>

<u>אוניברסיטת ת"א</u>

נכתב ונערך ע"י דודו לגזיאל

<u>הבהרה: הרשימות הללו הינן רשימות מתרגל</u> ולכן ייתכנו בהן טעויות ואי דיוקים. האחריות על השימוש בהן היא על המשתמש בלבד! בהצלחה! ☺.

הקדמה: החוברת נכתבה כרשימות מתרגל בהתבסס על חוברת התרגול של **אופיר הררי**, ובסיוע תרגילים שנאגרו לאורך שנים על ידי – **אסי כהן**, **רועי טפר**, **יובל הלר** ורבים וטובים אחרים. בחוברת זו ישנם תרגילים רבים שנכתבו על פי הספר המעולה 'הסתברות – קורס ראשון' של **שלדון רוס** ; 'מבוא הסתברות' של **אמיר בק** ו-'מבוא להסתברות וסטטיסטיקה – הסתברות' של **תלמה לויתן ואלונה רביב.**

תוכן עניינים:

פרק 1

<u>נושאים:</u>

- הקדמה – דברי פתיחה, הצגה כללית של הקורס, נהלים וכד'.
- תזכורת – תורת הקבוצות.
- אקסיומות פונקצית הסתברות.

<u>מושגים בסיסיים בתורת הקבוצות</u>

הגדרה – קבוצה היא אוסף של עצמים שונים, ללא חשיבות לסדר הופעתם וללא חשיבות להופעות חוזרות של אותם עצמים.

לדוגמא:

$$\phi = \{\ \ \}$$

$$\mathbb{N} = \{1, 2, 3, \ldots\}$$

$$\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$$

$$A = \{dog, cat, table, tau\}$$

הערה חשובה! מאחר ואנחנו דנים בהסתברות אז נשתמש בשפת הסתברות.
בהסתברות ישנה קבוצה כללית שקובעת את כל האיברים הקיימים בעולם (בו אנו דנים) המכונה 'מרחב המדגם' ונסמן לרוב על ידי Ω. כל קבוצה מעתה תקרא 'מאורע'.

<u>פעולות על מאורעות</u>

- האיחוד של A ו-B, המסומן על ידי $A \cup B$, הוא אוסף כל האיברים שנמצאים או ב-A או ב-B $A \cup B = \{x \mid x \in A \ or\ x \in B\}$.
- החיתוך של A ו-B, המסומן על ידי $A \cap B$, הוא אוסף כל האיברים שנמצאים גם ב-A וגם ב-B $A \cap B = \{x \mid x \in A \ and\ x \in B\}$.
- 'המשלים של A' הוא מאורע הכולל את כל האיברים מ- Ω שאינם ב- A
$\bar{A} = A^c = \{x \in \Omega \mid x \notin A\}$.
- המאורע 'A פחות B' הוא המאורע המורכב מכל האיברים השייכים ל-A ולא ל-B.
$A \setminus B = \{x \in \Omega \mid x \in A, x \notin B\}$.

<u>הגדרות – חזרה</u>:

- "מאורע A מוכל במאורע B" $\left(A\subseteq B\right)$, אם לכל $x\in A$, מתקיים $x\in B$.
- "מאורע A שווה למאורע B" $\left(A=B\right)$, אם $B\subseteq A$ ו- $A\subseteq B$.
- "מאורעות A ו-B זרים זה לזה" אם $A\cap B=\phi$.
- מאורעות B,A ו-C נקראים "זרים בזוגות, אם כל זוג מאורעות הם זרים (ניתן להכליל בקלות עבור n מאורעות).

<u>תרגיל 1</u>: מכלל הסטודנטים להנדסה נבחר סטודנט יחיד. נגדיר את המאורעות הבאים:

A – הסטודנט קורא ידיעות אחרונות.

B – הסטודנט קורא ישראל היום.

C – הסטודנט קורא הארץ.

השתמשו בפעולות האיחוד, חיתוך ומשלים בכדי לתאר את המאורעות:

1. הסטודנט אינו קורא ישראל היום.
2. הסטודנט קורא לפחות עיתון אחד.
3. הסטודנט קורא בדיוק 2 עיתונים.
4. הסטודנט קורא רק ידיעות אחרונות.

<u>פתרון תרגיל 1</u>:

1. המאורע הוא - $\overline{B}$.
2. המאורע הוא - $A\cup B\cup C$.
3. המאורע הוא - $\left(A\cap B\cap\overline{C}\right)\cup\left(B\cap C\cap\overline{A}\right)\cup\left(C\cap A\cap\overline{B}\right)$.
4. המאורע הוא - $A\cap\overline{B}\cap\overline{C}$.

<u>חוקי דה-מורגן</u>

- $\overline{A\cap B}=\overline{A}\cup\overline{B}$ או $\overline{\bigcap_{i\in I}A_i}=\bigcup_{i\in I}\overline{A_i}$.
- $\overline{A\cup B}=\overline{A}\cap\overline{B}$ או $\overline{\bigcup_{i\in I}A_i}=\bigcap_{i\in I}\overline{A_i}$.

אקסיומות פונקצית הסתברות

ישנן 3 אקסיומות לפונקצית ההסתברות $P(\cdot)$ המגודרת מקבוצות כל המאורעות לציר הממשי.

1. לכל מאורע $A \subseteq \Omega$ מתקיים - $0 \le P(A) \le 1$.
2. אם $A_1, A_2, A_3, \ldots, A_n \subseteq \Omega$ מאורעות זרים בזוגות אז - $P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} A_i$.
3. $P(\Omega) = 1$.

מכאן ניתן לגזור כל מיני תכונות נוספות כמו:

- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. נוסחת ההכלה והדחה לשני מאורעות.
- $P(\overline{A}) = 1 - P(A)$. הסתברות המשלים של A.
- לכל n מאורעות מתקיים : $P\left(\bigcup_{i=1}^{n} A_i\right) \le \sum_{i=1}^{n} P(A_i)$.

ועוד...

<u>תרגיל 2</u>: בן ואלון יוצאים לחגוג את תחילת השנה האקדמית. הם גרים בסמיכות לשני פאבים – A ו-B. ההסתברות שיילכו לפאב A היא 0.5 וההסתברות שיילכו לפאב B היא 0.4. כמו כן ההסתברות שיילכו לאחד מן המקומות לפחות היא 0.8.

א. תארו את מרחב המדגם.

ב. מה ההסתברות שיילכו גם לפאב A וגם לפאב B?

ג. מה ההסתברות שיילכו לפאב אחד בלבד?

<u>פתרון תרגיל 2</u>:

א. מרחב המדגם הוא - $\Omega = \{A \cap B, A \cap \overline{B}, \overline{A} \cap B, \overline{A} \cap \overline{B}\}$ כאשר כל איבר ב- Ω קובע לאיזה פאב/ים הלכו בן ואלון.

ב. נוכל לפתור בעזרת נוסחת הכלה והדחה או בעזרת טבלה. אנו מעוניינים למצוא את ההסתברות ל- $\{A \cap B\}$.

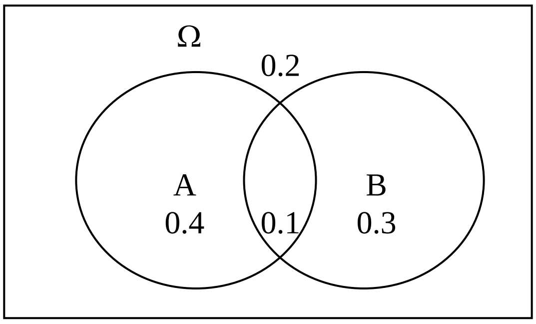
$$P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0.4 + 0.5 - 0.8 = 0.1$$

או לחילופין בעזרת טבלה:

חיתוך בין מאורעות	A	$\overline{A}$
B	0.1	0.3
$\overline{B}$	0.4	0.2

הסבר – נתון כי ההסתברות של צמד התאים העליונים יחד עם התא השמאלי התחתון היא 0.8. בנוסף נתון שהסתברות צמד התאים השמאליים היא 0.5 ולכן נוכל לקבוע שהסתברות התא הימני העליון היא 0.3. מכאן נסיק את כל היתר ובפרט את התשובה $P(A\cap B)=0.1$.

נרשום את התוצאות בדיאגרמת ון:



ג. אנו מעוניינים למצוא את ההסתברות של המאורע - $(A\cap\overline{B})\cup(B\cap\overline{A})$
ומהטבלה / דיאגרמה אנו רואים כי ההסתברות לכך היא 0.7.

<u>תרגיל 3</u>: בסקר שנערך בפקולטה להנדסה בשנה א' התקבלו התוצאות הבאות –

- 30% מהסטודנטים לוקחים את הקורס 'מבוא להסתברות'.
- 40% לוקחים את הקורס 'אינפי 1'.
- 35% לוקחים את הקורס 'נדידת ציפורים'.
- חצי מהסטודנטים לוקחים את 'מבוא להסתברות' או 'אינפי 1'.
- 55% לא לוקחים את הקורס 'מבוא להסתברות' ולא לוקחים 'נדידת ציפורים'.
- 20% לוקחים את 'אינפי 1' יחד עם 'נדידת ציפורים'.
- 15% לוקחים את כל הקורסים יחדיו.

מה ההסתברות שסטודנטית ספציפית לא לומדת אפילו קורס אחד מהנ"ל?

פתרון תרגיל 3:

נסמן את המאורעות באופן הבא:

M – הסטודנטית לוקחת את קורס מבוא להסתברות.

N – הסטודנטית לוקחת את קורס אינפי 1.

I – הסטודנטית לוקחת את קורס נדידת ציפורים.

אנו צריכים למצוא את ההסתברות למאורע $\overline{M} \cap \overline{N} \cap \overline{I}$.

$$P\left(\overline{M} \cap \overline{N} \cap \overline{I}\right) = P\left(\overline{M \cup N \cup I}\right) = 1 - P\left(M \cup N \cup I\right)$$

ולכן נמצא את - $P(M \cup N \cup I)$. נעזר בנוסחת הכלה והדחה ל-3 מאורעות.

$$\begin{aligned} P(M \cup N \cup I) &= P(M) + P(N) + P(I) \\ &\quad - P(M \cap N) - P(M \cap I) - P(N \cap I) + P(M \cap N \cap I) \\ &= 0.3 + 0.4 + 0.35 \\ &\quad - P(M \cap N) - P(M \cap I) - P(N \cap I) + 0.15 \\ &= 1.2 - P(M \cap N) - P(M \cap I) - P(N \cap I) \end{aligned}$$

נמצא את ההסתברויות החסרות,

$$\begin{aligned} P(M \cap N) &= P(M) + P(N) - P(M \cup N) \\ &= 0.3 + 0.35 - \left[1 - P\left(\overline{M \cup N}\right)\right] \\ &= 0.65 - \left[1 - P\left(\overline{M} \cap \overline{N}\right)\right] = 0.65 - [1 - 0.55] = 0.2 \end{aligned}$$

$$P(I \cap N) = 0.2$$

$$P(M \cap I) = P(M) + P(I) - P(M \cup I) = 0.3 + 0.4 - 0.5 = 0.2$$

ולכן,

$$P(M \cup N \cup I) = 1.2 - 0.2 - 0.2 - 0.2 = 0.6$$

ומכאן ש-

$$P\left(\overline{M} \cap \overline{N} \cap \overline{I}\right) = 1 - 0.6 = 0.4.$$

תרגיל 4: סטודנט עובר בדרכו לאוניברסיטה 3 רמזורים. הסיכוי שלא יהיה אף רמזור אדום בדרך הוא 0.4, רמזור אדום אחד – 0.1, 2 רמזורים אדומים – 0.2. מרחב המדגם מורכב ממספר הרמזורים בהם נתקל הסטודנט. מה ההסתברות של המאורעות הבאים:

A - לפחות רמזור אדום אחד.

B - לפחות רמזור ירוק אחד.

C - מספר אי זוגי של רמזורים אדומים.

D - לכל היותר רמזור ירוק אחד.

פתרון תרגיל 4:

$$P(A)=1-P(no\ red\ lights)=1-0.4=0.6$$

$$P(B)=P(0\ red\ lights)+P(1\ red\ light)+P(2\ red\ lights)=0.4+0.1+0.2=0.7$$

נצטרך למצוא את ההסתברות שיהיו 3 רמזורים אדומים –

$$\begin{aligned}P(3\ red\ lights)&=1-P(2\ red\ lights)-P(1\ red\ light)-P(no\ red\ lights)\\&=1-0.4-0.2-0.1=0.3\end{aligned}$$

ולכן –

$$P(C)=P(1\ red\ light)+P(3\ red\ lights)=0.1+0.3=0.4$$

$$P(D)=P(1\ green\ light)+P(no\ green\ lights)=0.2+0.3=0.5$$

פרק 2

נושאים:

- מרחבי הסתברות סימטריים.
- קומבינטוריקה.

מרחבי הסתברות סימטריים

נזכר שכבר הגדרנו את מרחב המדגם Ω בתור העולם בו אנו דנים. הצירוף (Ω, P) הוא מרחב הסתברות. כאשר Ω היא קבוצה סופית אזי מרחב ההסתברות נקרא סופי. אם בנוסף $P(\omega)$ זהה לכל $\omega \in \Omega$ נאמר שזה מרחב הסתברות סימטרי.

למה מרחב הסתברות סימטרי זה דבר מאוד נוח?

כאשר מדובר במרחבי הסתברות סימטריים, כל שעלינו לעשות הוא לחשב את גודל המאורע (קרי, לספור את מספר המצבים) ולחלק במספר המצבים האפשריים הכללי.

במצב כזה הנוסחה להסתברות היא מאוד פשוטה –

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\# A}{\# \Omega} = \frac{number\ of\ elements\ in\ A}{number\ of\ elements\ in\ \Omega}$$

קומבינטוריקה

הרקע התיאורטי בקומבינטוריקה הוא מאוד קצר ופשוט. למעשה, כל הרקע התיאורטי מתבסס על 2-3 נוסחאות כלליות ומהן נגזר כל היתר.

נוסחאות:

- מספר (התמורות) הדרכים לסדר בשורה k עצמים שונים הוא $k!$.
- מספר הדרכים לבחור עם החזרה וחשיבות לסדר k עצמים מתוך n עצמים שונים הוא n^k.
- מספר הדרכים לבחור ללא החזרה וללא חשיבות לסדר k עצמים מתוך n עצמים שונים הוא
$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$
- מספר הדרכים לסדר n אובייקטים שונים במעגל היא - $n-1!$.

<u>דוגמא</u>: את מספר האפשרויות לבחור 3 סטודנטים שונים מתוך קבוצת תרגול המונה 50 סטודנטים ניתן לחשב במספר דרכים.

דרך א' – מספר האפשרויות לבחור סטודנט ראשון (50) כפול מספר האפשרויות לבחור סטודנט שני (49) כפול מספר האפשרויות לבחור סטודנט שלישי (48) ונחלק ב-$3!$ כי סדר הבחירה לא משנה.

דרך ב' – פשוט לרשום $\binom{50}{3}=\frac{50!}{3!(47)!}=\frac{48\cdot49\cdot50}{3!}$.

<u>תרגיל 1</u>: 8 אנשים שביניהם יש את אלון, אלונה, בן ובתיה מסתדרים בשורה באקראי. מה ההסתברות ש:

א. אלון ואלונה נמצאים זה לצד זה?

ב. לפחות אחד מבין אלון ואלונה נמצא בקצה?

ג. אלון ואלונה נמצאים זה לצד זה אך בן ובתיה אינם נמצאים זה לצד זה?

<u>פתרון תרגיל 1</u>:

א. נסמן – A הוא המאורע בו אלון ואלונה נמצאים זה לצד זה. נחשב את $|A|$. נשים את אלון ואלונה יחדיו ונתייחס להם כ-'בלוק'. עתה יש לסדר 7 אובייקטים בשורה, אובייקט אחד הוא 'בלוק' של שני אנשים והיתר הם 6 אנשים נוספים. סה"כ האפשרויות לסידור הן - $7!$. נכפיל את $7!$ ב-$2!$ כי זה מספר הדרכים השונות לסדר את אלון ואלונה בינם לבין עצמם ולכן:

$$P(A)=\frac{|A|}{|\Omega|}=\frac{7!\cdot2!}{8!}=\frac{1}{4}$$

ב. בסעיף זה נשתמש דווקא במשלים (יותר קל באופן הזה). נסמן B כמאורע שאלונה או אלון נמצאים בקצה.

$$P(B)=1-P(\overline{B})=1-\frac{6\cdot5\cdot6!}{8!}=1-\frac{3\cdot5}{7\cdot4}=\frac{13}{28}$$

כאשר חישבנו את ההסתברות למשלים של B על ידי – מספר האפשרויות להציב את אלונה תחילה (6 אפשרויות), כפול מספר האפשרויות להציב את אלון (5 אפשרויות) כפול מספר האפשריות להציב את יתר האנשים ($6!$ אפשרויות).

ג. נסמן – C כמאורע שאלון ואלונה נמצאים יחדיו וגם בן ובתיה נמצאים יחדיו ואת המאורע D כמאורע שבו אלון ואלונה נמצאים יחדיו אך בן ובתיה אינם נמצאים יחדיו. נשים לב כי - $D=A\setminus C$ ובנוסף אנו רואים כי $C\subseteq A$ ולכן

$$P(D)=P(A\setminus C)=P(A)-P(C)=$$
$$=\frac{1}{4}-\frac{6!2!2!}{8!}=\frac{5}{28}$$

כאשר את $P(C)$ חישבנו על ידי קיבוע אלון ואלונה יחדיו ובן ובתיה יחדיו, סידור כל האובייקטים בשורה ($6!$ אפשרויות) כפול כל הסידורים הפנימיים האפשריים בין הבנים והבנות בזוגות.

<u>תרגיל 2</u>: 20 אנשים שהם 10 נשים ובעליהן מתיישבים באקראי סביב שולחן עגול. מה ההסתברות ש-

א. כל הזוגות יושבים יחדיו?
ב. רק 9 זוגות יושבים יחד?
ג. אף זוג לא יושב יחד?

<u>פתרון תרגיל 2</u>:

א. גודל מרחב שלנו הוא מספר הסידורים של אנשים במעגל ובמקרה זה ישנם 20 איש ולכן $|\Omega|=19!$. נסמן את המאורע המבוקש ב-A. מאחר ואנו מעוניינים שהזוגות יישבו יחדיו אזי צריך לקחת כל זוג כבלוק ולסדר את הבלוקים בפני עצמם ולאחר מכן לסדר כל בלוק באופן עצמאי (סידור פנימי). בגלל שישנם 10 זוגות, יש לנו 10 בלוקים וכמות הסידורים של 10 בלוקים במעגל היא $9!$. כעת נותר לנו לסדר כל בלוק בפני עצמו (הגברת יכולה לשבת מימין או משמאל לבן זוגה) ולכן נצטרך להכפיל ב- 2^{10}. לכן נקבל:

$$P(A)=\frac{9!\cdot 2^{10}}{19!}$$

ב. נסמן את המאורע ב-B ונפתור בעזרת חישוב ישיר. נתחיל בכך שאנחנו צריכים לבחור את הזוגות שיושבים יחדיו ואת הזוג שלא יושב יחד (10 אפשרויות). לאחר מכן צריכים לסדר את כל הזוגות שיושבים יחדיו במעגל. מאחר ומדובר ב-9 זוגות אזי מספר הסידורים שלהם במעגל הוא $8!$ ובנוסף צריכים להכפיל באפשרויות סידור של כל זוג בינו לבין עצמו (בת הזוג יושבת מימין או משמאל) ולכן להכפיל זאת ב- 2^9. לבסוף אנחנו צריכים למקם את הזוג הנוסף שלא יושב יחדיו. נוכל תחילה למקם את האישה בין הזוגות שכבר הצבנו (9 אפשרויות) ולאחר מכן את הגבר ב-8 המקומות הנוספים. נקבל כי

$$P(B)=\frac{10\cdot 8!\cdot 2^9\cdot 9\cdot 8}{19!}=\frac{10!\cdot 2^{12}}{19!}$$

ג. נזכר בנוסחת הכלה והדחה הכללית ביותר –

$$\begin{aligned}P(E_1\cup E_2\cup\ldots\cup E_n)=\sum_{i=1}^{n}P(E_i)-\sum_{i_1<i_2}P\left(E_{i_1}\cap E_{i_2}\right)+\ldots+\\+(-1)^{r+1}\sum_{i_1<i_2<\ldots<i_r}P\left(E_{i_1}\cap E_{i_2}\cap\ldots\cap E_{i_r}\right)+\\+\ldots+(-1)^{n+1}P(E_1\cap E_2\cap\ldots\cap E_n)\end{aligned}$$

או לחילופין,

$$P(E_1\cup E_2\cup\ldots\cup E_n)=\sum_{k=1}^{n}(-1)^{k+1}\sum_{i_1<i_2<\ldots<i_k}P\left(E_{i_1}\cap E_{i_2}\cap\ldots\cap E_{i_k}\right)$$

כאשר הסכום $\sum_{i_1<i_2<\ldots<i_k}P\left(E_{i_1}\cap E_{i_2}\cap\ldots\cap E_{i_k}\right)$ הוא על כל תת הקבוצות מכל r האפשריות שניתן לבחור מ-n המאורעות.

נשים לב כעת כי אם נגדיר את המאורע E_i כמאורע בו הזוג ה-I לא יושב יחדיו אזי אנחנו צריכים למצוא את

$$P\left(\bigcap_{i=1}^{n}E_i\right)=1-P\left(\overline{\bigcap_{i=1}^{n}E_i}\right)=1-P\left(\bigcup_{i=1}^{n}\overline{E_i}\right)$$

והמאורע $\bigcup_{i=1}^{n}\overline{E_i}$ הוא המאורע בו ישנו לפחות זוג אחד היושב יחדיו. כל שנותר כעת זה למצוא את $\sum_{i_1<i_2<\ldots<i_k}P\left(\overline{E_{i_1}}\cap\overline{E_{i_2}}\cap\ldots\cap\overline{E_{i_k}}\right)$ ולהציב בנוסחה למעלה. מאורע זה הוא המאורע בו ישנם לפחות k זוגות היושבים יחד, על כן הסכום שווה ל-

$$\sum_{i_1<i_2<\ldots<i_k}P\left(\overline{E_{i_1}}\cap\overline{E_{i_2}}\cap\ldots\cap\overline{E_{i_k}}\right)=\sum_{i_1<i_2<\ldots<i_k}\frac{2^k(19-k)!}{19!}=\binom{10}{k}\frac{2^k(19-k)!}{19!}$$

וזאת ניתן להציב בסכום לעיל ולקבל תשובה סופית.

<u>תרגיל 3</u>: מחלקים A כדורים ל-n תאים באופן מקרי. מה ההסתברות שכל התאים המלאים כאשר:

א. A = n.

ב. $A = n+1$.

ג. $A = n+2$.

פתרון תרגיל 3:

א. נשים לב כי בכל הסעיפים מרחב המדגם שלנו הוא n^A. בכדי שכל התאים יהיו מלאים צריך שבכל תא יהיה בדיוק כדור אחד ולכן אנחנו למעשה צריכים למצוא את מספר הדרכים להציב n כדורים ב-n תאים. זה שקול למספר התמורות על n איברים ולכן ההסתברות היא - $\frac{n!}{n^A}$.

ב. במצב זה צריך שיהיה תא עם 2 כדורים וכל היתר עם כדור אחד. נבחר תחילה את צמד הכדורים שהולכים לאותו התא, ישנן $\binom{n+1}{2}$ אפשרויות כאלה. את צמד הכדורים נחשיב כבלוק ולכן יש לנו n בלוקים שונים לסדר ב-n תאים. מכאן שמספר המצבים הכולל הוא - $\binom{n+1}{2}n!$ וההסתברות היא $\frac{\binom{n+1}{2}n!}{n^{n+1}}$.

ג. ישנם שני מצבים אפשריים – מצד אחד בו יש תא עם 3 כדורים והיתר עם 1 ומצב בו יש 2 תאים עם 2 כדורים והיתר עם בודד.

עבור המצב הראשון ישנן (כמו בסעיף ב') $\binom{n+2}{3}n!$ אפשרויות. במצב השני לעומת זאת יש $\binom{n+2}{2}\binom{n}{2}n!\frac{1}{2}$ אפשרויות (בוחרים פעמיים צמד כדורים ואז מסדרים את כל הבלוקים ; החלוקה ב-2 היא בגלל שסדר בחירת הצמדים חסר חשיבות). מכאן שההסתברות למאורע זה היא - $\frac{\binom{n+2}{3}n!+\binom{n+2}{2}\binom{n}{2}n!\frac{1}{2}}{n^{n+2}}$.

תרגיל 4: בכיתה יש 14 סטודנטים ו-16 סטודנטיות.

א. מה מספר הדרכים לבחור ועד בין 3 תלמידים כך ש-
- הועד לא ימנה רק בנים?
- ראש הועד תהיה בת?

- ייבחרו בנפרד ראש ועד, ממלא מקום וגזבר?

ב. בכמה אופנים נוכל לבחור ועד כיתה בן 5 תלמידים, ועדת קישוט בת 3 תלמידים ושני נציגים למועצת תלמידים?

<u>פתרון תרגיל 4</u>:

א.

- סך כל האפשרויות לבחור ועד בן 3 תלמידים הוא $\binom{30}{3}$. סך כל האפשרויות לבחור ועד בן 3 תלמידים המורכב מבנים בלבד הוא $\binom{14}{3}$. לכן מספר הועדים שלא מונים בנים בלבד הוא $\binom{30}{3}-\binom{14}{3}$.

  דרך נוספת היא לספור את מספר הועדים המונים רק בנות $\binom{16}{3}$, 2 בנות $\binom{14}{1}\binom{16}{2}$ או בת אחת $\binom{14}{2}\binom{16}{1}$ ולסכום. חישוב ייראה שהתוצאה זהה.

- תחילה נבחרת יושבת ראש לוועד (16 אפשרויות) ולאחר מכן נבחר צמד חברים נוספים $\binom{29}{2}$.

- במקרה הזה ישנה חשיבות לסדר הבחירה משום שאנחנו מבצעים בחירות לתפקידים ספציפיים. נוכל לבחור מראש את חברי הועד ואז למנות אותם לתפקידים או לחילופין לבצע בחירה ישירה לכל תפקיד.

  בדרך הראשונה מספר האפשרויות הוא - $\binom{30}{3}\cdot 3!$ שזה מספר האפשרויות לבחור ועד באופן כללי ולאחר מכן מספר האפשרויות לחלק 3 תפקידים ל-3 אנשים.

  בדרך השנייה נבחר תחילה יושב ראש (30 אפשרויות), לאחר מכן גזבר (29 אפשרויות) ולבסוף ממלא מקום (28 אפשרויות) ונקבל - $30\cdot 29\cdot 28$.

ב. נבחר תחילה ועד כיתה עם 5 חברים, $\binom{30}{5}$ אפשרויות. לאחר מכן ועדת קישוט ונציגים למועצת תלמידים, $\binom{25}{3}$ ו- $\binom{22}{2}$ בהתאמה. סה"כ - $\binom{22}{2}\binom{25}{3}\binom{30}{5}$ אפשרויות.

תרגיל 5: (בעיית ימי ההולדת) בכיתה n סטודנטים. מהי ההסתברות שיש לפחות זוג אחד של סטודנטים עם אותו יום הולדת בשנה?

פתרון תרגיל 5: תחילה נשים לב שגודל מרחב המדגם שלנו הוא 365^n (משום שלכל סטודנט יש 365 ימים אפשריים ליום הולדת וישנם n סטודנטים).

נסמן את המאורע – "לפחות זוג אחד של סטודנטים נולד באותו היום" ב-A ונשים לב ש- $\overline{A}$ הוא המאורע שכל הסטודנטים נולדו בימים שונים.

מספר האפשרויות לכך הוא $\binom{365}{n}\cdot n!$ שכן צריך לבחור n ימי הולדת שונים (בחירה ללא החזרה) וכמובן ישנה חשיבות לסדר ולכן נכפיל בכל הסידורים האפשריים.

לכן נקבל ש-

$$P(A)=1-P\left(\overline{A}\right)=1-\frac{\binom{365}{n}\cdot n!}{365^n}=1-\frac{\frac{365!}{(365-n)!}}{365^n}=1-\frac{365!}{(365-n)!365^n}$$

והתוצאה המפתיעה היא שעבור n=23 ההסתברות גדולה מ-50% ועבור 60 סטודנטים ההסתברות היא 0.994.

מה האינטואיציה לתוצאה הזאת? ובכן, ההסתברות שלכל זוג אנשים יהיה יום הולדת באותו היום היא $\frac{1}{365}=\frac{365}{365^2}$ ועבור כמות לא גדולה של אנשים ניתן לייצר הרבה זוגות...

תרגיל 6: במשחק פוקר, 5 קלפים מהווים "יד".

א. היד נקראת 'סטרייט' אם יש בה ערכים עוקבים שאינם מאותה צורה. 'סטרייט' יכול להיות לדוגמא: אס-2-3-4-1, או 10-9-8-7-נסיך. מה ההסתברות לקבל 'סטרייט' מיד בחלוקת הקלפים? (בהנחה שלכל יד יש אותה הסתברות להתקבל).

ב. היד נקראת full house אם שלושה קלפים שווים בערכם והשניים האחרים גם שווים בערכם (שלשה וזוג). מה ההסתברות לקבל פול האוס בחלוקת הקלפים?

פתרון תרגיל 6:

א. נשים לב תחילה שגודל מרחב המדגם שלנו הוא $\binom{52}{5}$. נצטרך כעת לספור את כמות הסטרייטים האפשריים שניתן לקבל. סטרייט אפשרי יכול להתחיל מ-אס, 2,... ולכל היותר מ-10 (הסטרייט הגבוה ביותר הוא 10 – נסיך – מלכה – מלך – אס) ולכן ישנם 10 רצפים אפשריים של ערכים ליד הזו. בנוסף, לכל קלף (מתוך החמישה שביד) יכולות להיות 4 צורות (תלתן, לב אדום וכד') ולכן סה"כ ישנם $10\cdot 4^5$

אפשרויות אבל (!) צריך לחסר מערך זה את האפשרות שכל הקלפים הם מאותה צורה (ויש לכך $10\cdot 4$ אפשרויות, 10 רצפים אפשריים כפול 4, מכיוון שיש 4 צורות). לכן ההסתברות היא –

$$\frac{10\cdot 4^5-10\cdot 4}{\binom{52}{5}}=\frac{10\left(4^5-4\right)}{\binom{52}{5}}\approx 0.0039 .$$

ב. בשביל לקבל full house נשים לב שנצטרך לבחור תחילה את הערך של הקלפים בשלשה ושל הקלפים בזוג ולכך ישנם $13\cdot 12$ אפשרויות. לאחר מכן נצטרך לבחור שלשה וזוג מאותם ערכים ולכך יש $\binom{4}{3}\binom{4}{2}$. לכן ההסתברות שנקבל היא –

$$\frac{13\cdot 12\cdot\binom{4}{3}\binom{4}{2}}{\binom{52}{5}}\approx 0.0014 .$$

<u>תרגיל 7</u>: כמה פתרונות במספרים טבעיים יש למשוואה - $x_1+x_2+\ldots+x_m=n$?

<u>פתרון תרגיל 7</u>: נמיר את הבעיה בבעיה שקולה. מה מספר הדרכים לחלק n כדורים זהים ל- m תאים שונים? נחשוב על מספר הכדורים בכל תא i כעל הערך של x_i (וישנם m כאלה המכילים סה"כ n כדורים) ונדמיין שהתאים מחולקים על ידי מקלות.

● ● | ● ● ● | ● | | ● | ● ● ●

בכדי לפצל את n הכדורים ל-m תאים דרושים לנו m-1 מקלות לפיכך מספר האפשרויות לחלק את הכדורים בין התאים היא לבחור n או m-1 מקומות מתוך n+m-1 מקומות אפשריים.

$$\frac{(n+m-1)!}{n!(m-1)!}=\binom{n+m-1}{n} .$$

<u>תרגיל 8</u>: בקו 271 לאוניברסיטת ת"א יש 6 סטודנטים והוא נעצר ב-9 תחנות בסמוך לאוניברסיטה. כל סטודנט בוחר תחנה אחת באופן מקרי ויורד בה. מהי ההסתברות ש-

א. בכל תחנה ירד סטודנט אחד לכל היותר?

ב. מה ההסתברות שבבית התפוצות יירדו 3 סטודנטים בדיוק?

<u>פתרון תרגיל 8</u>:

א. נתחיל בחישוב מרחב המדגם. במקרה זה נובע ש- $|\Omega| = 9\cdot9\cdot9\cdot9\cdot9\cdot9 = 9^6$

כי לכל סטודנט ישנן 9 אפשרויות וישנם 6 סטודנטים. גודל המאורע שלנו הוא -

$\binom{9}{6}\cdot 6!$ כי צריך לבחור 6 תחנות מתוך 9 אפשרויות ללא החזרה וההכפלה ב- $6!$

נובעת מכך שהסטודנטים שונים זה מזה ולכן יש חשיבות לסדר. ההסתברות היא:

$$\frac{\binom{9}{6}\cdot 6!}{9^6}$$

ב. נגדיר את המאורע המבוקש כ-A. נחשב את גודל A באופן הבא: נבחר תחילה 3

סטודנטים מתוך ה-6 ($\binom{6}{3}$ אפשרויות) ונחייב אותם לרדת בבית התפוצות. בנוסף

יתר הסטודנטים יכולים לרדת בכל תחנה אחרת שזה 8^3 אפשרויות, סה"כ:

$$P(A) = \frac{\binom{6}{3}\cdot 8^3}{9^6}$$

<u>תרגיל 9</u>: בקופסא עם 100 נורות יש 6 פגומות. מוציאים מדגם מקרי ללא החזרה של 5 נורות. מה הסיכוי ש –

א. יהיו לנו בדיוק 2 נורות תקינות?

ב. יהיו לנו מספיק נורות להחליף צמד נורות שנשרפו במטבח ובסלון?

<u>פתרון תרגיל 9</u>:

א. גודל מרחב המדגם שלנו הוא - $\binom{100}{5}$ (בחירת 5 נורות מתוך 100). קבוצת הנורות

הפגומות, שתקרא 'קבוצת המיוחדים', מונה 6 נורות. נזכור שמדובר במרחב מדגם סימטרי. אנחנו צריכים למצוא את המאורע בו יהיו לנו בדיוק 2 נורות תקינות מתוך ה-5. נגדיר את המאורע המבוקש בתור A_2.

$$P\left(A_2\right)=\frac{\binom{6}{3}\binom{94}{2}}{\binom{100}{5}}$$

ב. אנחנו חייבים שיהיו לנו במדגם לפחות 2 נורות תקינות אך יכול להיות גם שיהיו לנו יותר מ-2 נורות תקינות (למעשה יכולות להיות עד 5 נורות תקינות במדגם) ולכן ההסתברות שיהיו לפחות 2 נורות תקינות (מאורע A) היא:

$$P\left(A\right)=\frac{\sum_{i=2}^{5}\binom{6}{5-i}\binom{94}{i}}{\binom{100}{5}}=\frac{\binom{6}{3}\binom{94}{2}+\binom{6}{2}\binom{94}{3}+\binom{6}{1}\binom{94}{4}+\binom{6}{0}\binom{94}{5}}{\binom{100}{5}}$$

שזה בעצם בחירת נורות תקינות (2 ומעלה) מתוך ה-94 תקינות שיש בקופסא ובנוסף בחירת יתר הנורות מתוך 6 מיוחדות (פגומות).

קיבלנו במצב זה התפלגות הנקראת היפרגיאומטרית.

התפלגות היפרגיאומטרית קובעת מה ההסתברות לדגום מספר מסוים של אובייקטים מיוחדים, כאשר יש אוכלוסיה של n אובייקטים בכלל ובתוכה m אובייקטים מיוחדים ואנו דוגמים מכלל האוכלוסייה k אובייקטים שונים ללא החזרה.

פרק 3

נושאים:

- הסתברות מותנית.
- חוק בייס.
- נוסחת ההסתברות השלמה.
- אי-תלות.

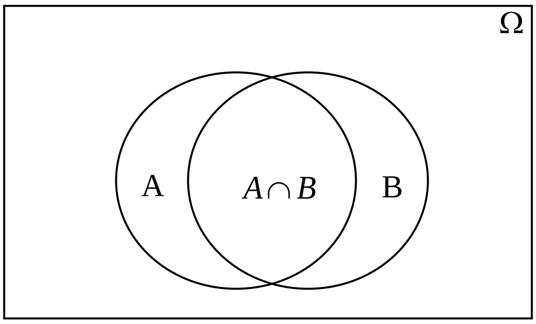
הסתברות מותנית

הרעיון מאחורי הסתברות מותנה היא לחשב את ההסתברות למאורע ספציפי כאשר ידוע לנו שמאורע אחר אכן התרחש.

הגדרה: ההסתברות למאורע A בהינתן שמאורע B התרחש (הסתברות חיובית) היא –

$$P(A|B):=\frac{P(A\cap B)}{P(B)}$$

נדגים זאת בעזרת דיאגרמת ון. יהיו מאורעות A ו-B כלשהם.



נתון כי B התרחש ולכן אנחנו יודעים שאנו נמצאים כעת במצב שמוכל בתוך המאורע B בדיאגרמה. ההסתברות של המאורע A בהינתן B זה בדיוק ההסתברות של החלק $A\cap B$ חלקי ההסתברות של B.

חוק בייס

במידה וגם למאורע A יש הסתברות חיובית אזי,

$$P(A|B)=\frac{P(B|A)P(A)}{P(B)}$$

<u>תרגיל 1</u>: כד מכיל 10 כדורים לבנים, 5 צהובים ו-10 שחורים. מוציאים באקראי כדור מן הכד ומתברר שהוא אינו שחור.

א. מה ההסתברות שהכדור צהוב?

ב. מה ההסתברות שהכדור לבן?

<u>פתרון תרגיל 1</u>:

א. נסמן ב-Y את המאורע שהכדור שהוצא הוא צהוב ונסמן ב-B את המאורע שהוא שחור.

$$P\left(Y\,|\,\overline{B}\right)=\frac{P\left(Y\cap\overline{B}\right)}{P\left(\overline{B}\right)}=\frac{P\left(Y\right)}{P\left(\overline{B}\right)}=\frac{\frac{5}{25}}{\frac{15}{25}}=\frac{1}{3}$$

המעבר $P\left(Y\cap\overline{B}\right)=P\left(Y\right)$ נובע מכך שהמאורע שהכדור יהיה צהוב מוכל במאורע שהכדור אינו שחור.

ב. בגלל שמצאנו בסעיף א' כי $P\left(Y\,|\,\overline{B}\right)=\frac{1}{3}$ ובגלל ש-

$P\left(B\,|\,\overline{B}\right)=\frac{P\left(B\cap\overline{B}\right)}{P\left(\overline{B}\right)}=\frac{P\left(\phi\right)}{P\left(\overline{B}\right)}=0$ נובע ש-

$$P\left(W\,|\,\overline{B}\right)=1-P\left(B\,|\,\overline{B}\right)-P\left(Y\,|\,\overline{B}\right)=1-0-\frac{1}{3}=\frac{2}{3}$$

כאשר W הוא המאורע שהכדור שהוצא הוא לבן.

<u>עקרון הכפל</u>

יהיו מאורעות $A_1,\ldots,A_n$ כלשהם, אזי

$$P\left(\bigcap_{i=1}^{n}A_i\right)=P\left(A_1\right)\cdot P\left(A_2\,|\,A_1\right)\cdot\ldots\cdot P\left(A_n\,|\,A_1\cap A_2\cap\ldots\cap A_{n-1}\right)$$

<u>תרגיל 2</u>: בכד יש 5 כדורים אדומים ו-4 לבנים. מוציאים מהכד בזה אחר זה באופן מקרי וללא החזרה 3 כדורים. מה ההסתברות שכל הכדורים לבנים?

<u>פתרון תרגיל 2</u>: נסמן A_i כמאורע בו הכדור ה-i הוא לבן, ולכן

$$P\left(A_1\cap A_2\cap A_3\right)=P\left(A_1\right)\cdot P\left(A_2\,|\,A_1\right)\cdot P\left(A_3\,|\,A_1\cap A_2\right)=\frac{4}{9}\cdot\frac{3}{8}\cdot\frac{2}{7}=\frac{1}{21}$$

נוסחת ההסתברות השלמה

נניח שיש לנו סדרת מאורעות זרים בזוגות המחלקים את מרחב המדגם, $A_1,\dots,A_n$

($\bigcup_{i=1}^{n} A_i = \Omega\,,\ A_i \cap A_j = \phi \quad \forall i \neq j$). נוכל להיעזר במאורעות אלה על מנת לחשב את ההסתברות של מאורע אחר B בעזרת הסתברות מותנה.

$$P(B)=\sum_{i=1}^{n} P(B\,|\,A_i)P(A_i)$$

<u>תרגיל 3</u>: סיגנאל בינארי מועבר במערכת רועשת. אם הסיגנל הוא '0' אז ההסתברות שיועבר הערך ה-'1' בטעות היא q_0 , אם הסיגנל הוא '1' אזי ההסתברות שיועבר הערך '0' בטעות היא q_1.

א. נניח שהמקור משדר '0' בהסתברות p ו-'1' בהסתברות p-1. מה ההסתברות שיתקבל הסיגנל הנכון?

ב. מה ההסתברות לשדר 1011 ללא טעויות?

ג. נניח שכל סיגנל משוגר 3 פעמים והפענוח נקבע לפי הרוב שיתקבל. מה הסיכוי ש-'0' שמשודר יפוענח כראוי?

ד. נניח שהמקור משדר 3 סיגנלים כמו בסעיף ג' ושההסתברויות של סעיף א' עדיין תקפות. אם נתון שהתקבל שדר 101, מה ההסתברות שהשדר המקורי היה '0'?

<u>פתרון תרגיל 3</u>:

א. נגדיר את המאורע W בתור "האות שהתקבל הוא שגוי" ואת המאורע R כ-"האות שהתקבל תקין". בעזרת נוסחת ההסתברות השלמה נקבל,

$$P(R)=P(R\,|\,0)P(0)+P(R\,|\,1)P(1)=(1-q_0)\,p+(1-q_1)(1-p)$$

ב. בכדי לשדר 1011 ללא טעויות צריך שכל השדרים של '1' יעברו טוב וכן כל השדרים של '0'. ההסתברות לכך היא - $(1-q_0)(1-q_1)^3$.

ג. בכדי ש-'0' יועבר נכון צריך שלפחות פעמים יתקבל האות '0' ולכן ישנם 2 מצבים אפשריים: התקבל פעמיים הסיגנל '0' והתקבל 3 פעמים הסיגנל '0'. ההסתברות לכך שיתקבל פעמיים הסיגנל '0' היא $\binom{3}{2}(1-q_0)^2 q_0$ וההסתברות לכך

שיתקבל 3 פעמים היא - $\left(1-q_0\right)^3$ והתשובה הסופית הינה סכום ההסתברויות.

ד. נשתמש בחוק בייס על מנת למצוא את ההסתברות –

$$P\left(0|101\right)=\frac{P\left(101|0\right)P\left(0\right)}{P\left(101\right)}$$

נמצא את $P\left(101\right)$ בעזרת נוסחת ההסתברות השלמה.

$$P\left(101\right)=P\left(101|0\right)P\left(0\right)+P\left(101|1\right)P\left(1\right)=$$
$$=\left(1-q_0\right)q_0^2p+\left(1-q_1\right)^2q_1\left(1-p\right)$$

ולכן,

$$P\left(0|101\right)=\frac{P\left(101|0\right)P\left(0\right)}{P\left(101\right)}=\frac{\left(1-q_0\right)q_0^2p}{\left(1-q_0\right)q_0^2p+\left(1-q_1\right)^2q_1\left(1-p\right)}.$$

<u>תרגיל 4</u>: בקזינו 2 מכונות מזל. ההסתברות לזכות במכונה א' היא 0.4 ובמכונה ב' היא 0.3. אחד בוחר באופן מקרי מכונה אחת ומשחק. אם הוא זוכה הוא משחק בה משחק נוסף ואם הוא מפסיד אז הוא עובר למכונה הבאה.

מה ההסתברות לכך ש-:

א. האדם הפסיד בשני המשחקים?

ב. ניצח בשני המשחקים?

ג. ניצח באחד והפסיד באחד?

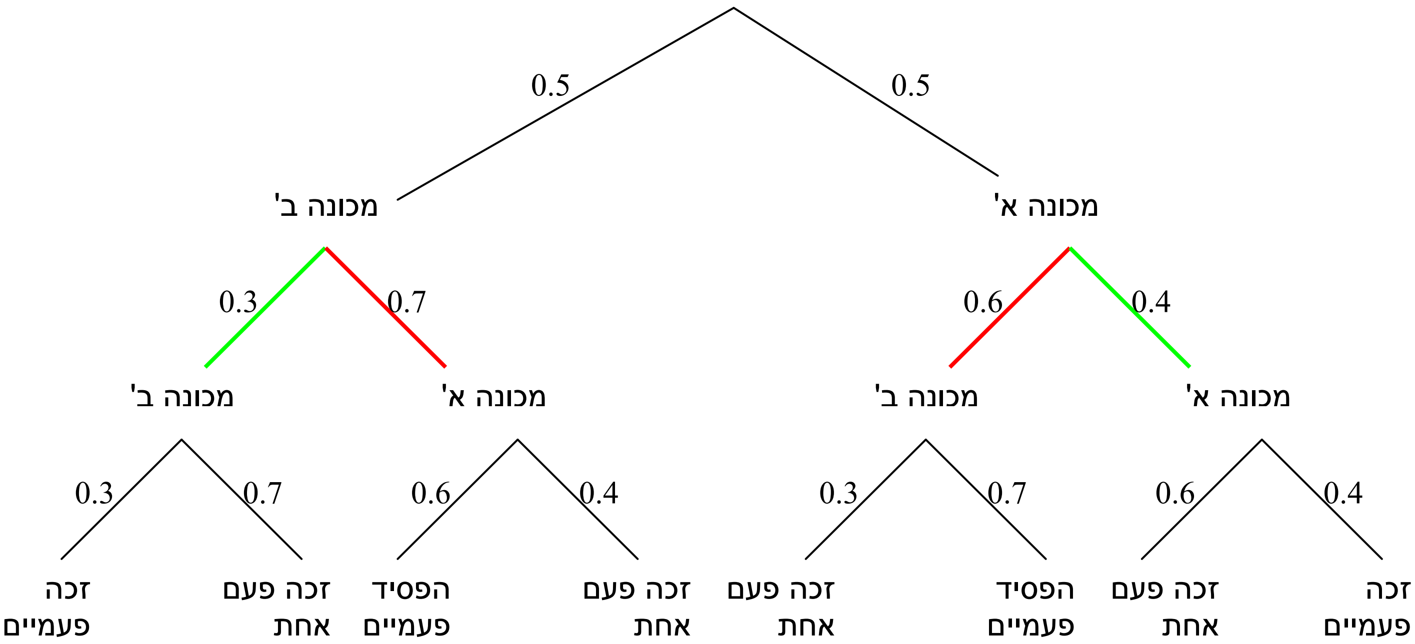
<u>פתרון תרגיל 4</u>:

א. נבנה תחילה את עץ ההסתברויות לשאלה (בעמוד הבא). כעת כל שנותר זה לעקוב אחרי הענפים הרלוונטיים לכל סעיף ולענות בהתאם. לפי העץ, ההסתברות שיפסיד פעמיים היא - $0.5\cdot 0.7\cdot 0.6+0.5\cdot 0.6\cdot 0.7=0.42$

ב. ההסתברות שיזכה פעמיים היא - $0.5\cdot 0.3\cdot 0.3+0.5\cdot 0.4\cdot 0.4=0.125$.

ג. ההסתברות שיזכה פעם אחת היא (בעזרת סעיפים א' ו-ב') -

$1-0.125-0.42=0.455$.

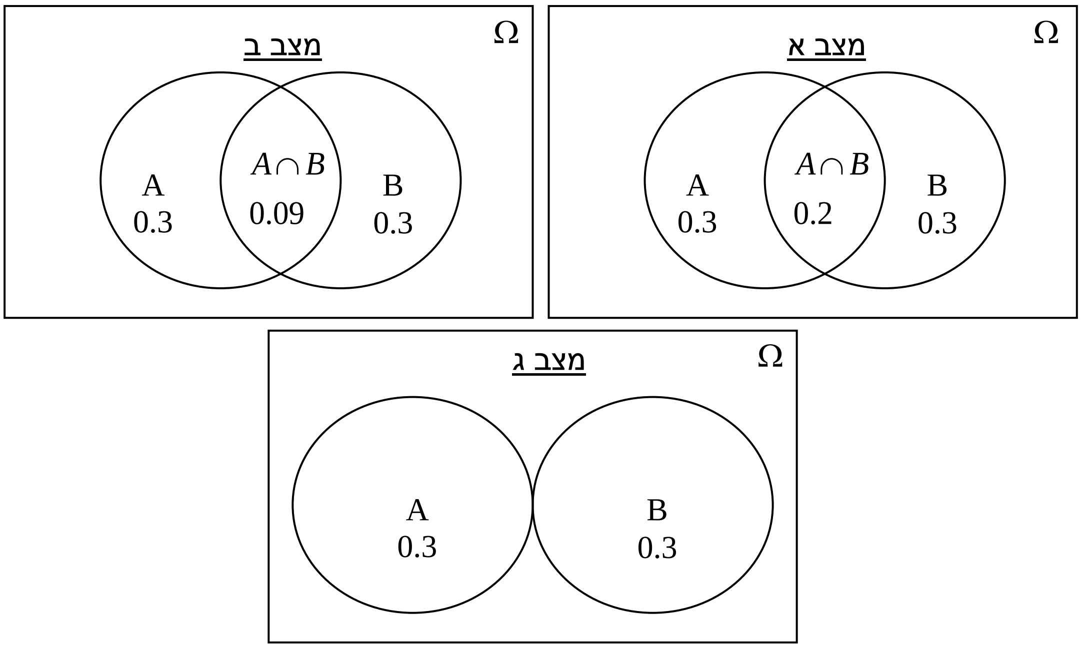


אי תלות בין מאורעות

יהיו זוג מאורעות A ו-B. המאורעות יכולים להיות תלויים או בלתי תלויים.

<u>הגדרה</u>: מאורעות A ו-B הם בלתי תלויים אם $P(A\cap B)=P(A)P(B)$.

<u>דוגמא</u>: באילו מצבים המאורעות A ו-B בלתי תלויים?



במצב א' המאורעות לא מקיימים את הגדרת האי תלות משום ש- $0.2\neq 0.3\cdot 0.3=0.09$

לעומת זאת במצב ב' הם כן מקיימים. מה לגבי מצב ג'?

במצב ג' הרבה יותר קל לשפוט בגלל שכאשר מאורע אחד מתרחש אזי השני בוודאות לא מתרחש (מאורעות זרים) ולכן הם תלויים.

<u>תרגיל 5</u>: מבצעים סדרה אינסופית של ניסויים בלתי-תלויים. לכל ניסוי יש הסתברות p להצלחה. מה הסתברות ש-:

א. מתקבלת לפחות הצלחה אחת ב-n הניסויים הראשונים.

ב. מתקבלות לפחות k הצלחות ב-n הניסויים הראשונים.

ג. בכל הניסויים מתקבלות הצלחות.

<u>פתרון תרגיל 5</u>:

א. הרבה יותר נוח לחשוב על המאורע המשלים למאורע "מתקבלת לפחות הצלחה אחת..." ומאורע זה הוא – "אין כלל הצלחות ב-n הניסויים הראשונים", נסמן מאורע זה ב-B. אם נסמן ב- A_i את המאורע שהייתה הצלחה בניסוי ה-I אזי:

$$P(B)=P\left(\bigcap_{i=1}^{n}\overline{A_i}\right)=P\left(\overline{A_1}\right)\cdot P\left(\overline{A_2}\right)\cdot\ldots\cdot P\left(\overline{A_n}\right)=(1-p)^n$$

ולכן ההסתברות לכן שהייתה לפחות הצלחה אחת היא - $1-(1-p)^n$.

ב. בכדי שיהיו k הצלחות צריכים להיות בנוסף גם n-k כשלונות. אנחנו צריכים לחשוב על כל הקומבינציות לקבל k הצלחות ב-n ניסויים (למעשה צריך למקם k איברים בתוך n מקומות או לחילופין לבחור k מקומות מתוך n) ולאחר מכן להכפיל בהסתברויות שיהיו אכן אותן ההצלחות ואותן הכישלונות. נניח כי המאורע המבוקש הוא A, אזי

$$P(A)=\sum_{i=k}^{n}\binom{n}{i}p^i(1-p)^{n-i}$$

ג. נמצא תחילה את ההסתברות שכל n הניסויים הראשונים מסתיימים בהצלחה.

$$P\left(\bigcap_{i=1}^{n}A_i\right)=P(A_1)\cdot P(A_2)\cdot\ldots\cdot P(A_n)=p^n$$

כעת נשים לב כי –

$$P\left(\bigcap_{i=1}^{\infty}A_i\right)=P\left(\lim_{n\to\infty}\bigcap_{i=1}^{n}A_i\right)=\lim_{n\to\infty}P\left(\bigcap_{i=1}^{n}A_i\right)=\lim_{n\to\infty}p^n=\begin{cases}0 & p<1\\ 1 & p=1\end{cases}$$

<u>תרגיל 6</u>: בספרייה 10 ספרי הסתברות, מהם 5 הכוללים פתרונות. סטודנט המעוניין לשאול ספר מקבל את אחד הספרים באקראי (עם או בלי פתרונות). נאמר שהגעתם לספריה במטרה לשאול ספר בהסתברות והספרן מיידע אתכם שאחד הספרים כבר הושאל. אתם מקבלים ספר באקראי ומחזירים אותו כעבור 3 ימים ואז מגיע אחד מחבריכם ושואל גם הוא את אחד הספרים. נסמן: A – "הספר ששאלתם כולל פתרונות", B – "הספר ששאל חברכם כולל פתרונות" ו-C – "הספר שכבר הושאל כלל פתרונות".
האם המאורעות A ו-B תלויים?

<u>פתרון תרגיל 6</u>: נתחיל בכך שנמצא את ההסתברות לכך ששאלנו ספר עם פתרונות, קרי

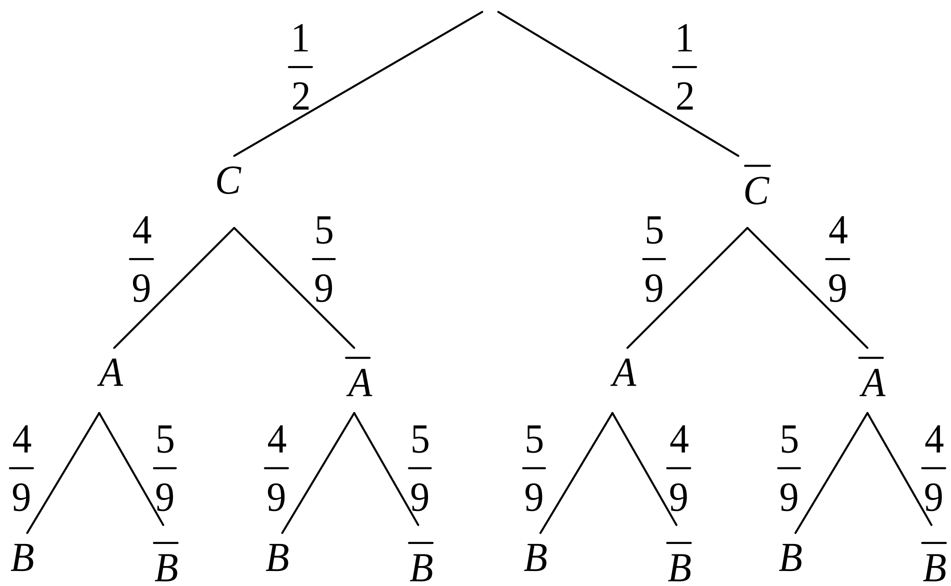
$$P(A)=P(B)=?$$

נשים לב שהסתברות של מאורע A שווה להסתברות של B כי ברגע שהחזרנו את הספר וחברנו הגיע לשאול ספר, אז, למעשה, ביצענו (אנחנו וחברנו) את אותו הניסוי באותם התנאים ולכן ההסתברויות זהות. נשתמש בנוסחת ההסתברות השלמה ונקבל:

$$P(A)=P(A|C)P(C)+P\left(A|\overline{C}\right)P\left(\overline{C}\right)=$$
$$=\frac{4}{9}\cdot\frac{5}{10}+\frac{5}{9}\cdot\frac{5}{10}=\frac{20}{90}+\frac{25}{90}=\frac{45}{90}=\frac{1}{2}$$

כלומר קיבלנו כי $P(A)=P(B)=\frac{1}{2}$. נותר כעת למצוא את ההסתברות של $A\cap B$.

נוכל לתאר את כל השתלשלות העניינים על ידי עץ הסתברויות:



כעת נעקוב אחרי הענפים שמתארים את המאורע $A\cap B$. נחשב את ההסתברויות הנובעות מכל ענף ונסכום יחדיו בכדי לקבל את הסתברות המאורע המבוקש:

$$P(A\cap B)=\frac{1}{2}\cdot\left(\frac{4}{9}\right)^2+\frac{1}{2}\cdot\left(\frac{5}{9}\right)^2=\frac{41}{162}\approx 0.253\neq 0.25=P(A)P(B)$$

קיבלנו למעשה שהמאורעות תלויים!!! איך זה ייתכן?!?!

ובכן, אם השאלנו ספר עם פתרונות הרי שהדבר מעיד על הסתברות גבוהה יותר לכך שכרגע יש רוב לספרים עם פתרונות בספרייה ולכן בהתאם ההסתברות שחברינו ייקח ספר עם פתרון גדלה בהתאם.

<u>תרגיל 7</u>: מטילים זוג קוביות עד אשר מתקבל לראשונה סכום מספרים שהוא גבוה מ-10.

1. מה ההסתברות שמספר הניסיונות יהיה קטן מ-10?
2. מה הסיכוי שהסכום 12 יתקבל לפני הסכום 11?

<u>פתרון תרגיל 7</u>:

1. נגדיר את A להיות המאורע המבוקש ותחילה נחשב את ההסתברות לקבל מספר גבוה מ-10. הזוגות האפשריים למצב זה הם $\{(5,6),(6,5),(6,6)\}$, ז"א סה"כ 3 מצבים מתוך 36 אפשריים ולכן הסיכוי לקבל מספר שווה ל-10 ומטה הוא $\frac{33}{36}=\frac{11}{12}$. נמצא את ההסתברות למאורע הנדרש בעזרת המשלים. המאורע המשלים אומר שיהיו לפחות 10 ניסיונות, ז"א נצטרך לקבל לפחות 9 "כשלונות" רצופים בניסוי בכדי להגיע להטלה ה-10. על כן:

$$P(A)=1-P(\overline{A})=1-\left(\frac{11}{12}\right)^9$$

2. נסמן ב-B את המאורע המבוקש. נשים לב שיש לנו מצב קלאסי בו אנחנו לא יודעים מתי יתקבל הסכום '12' ולכן אנחנו נתנה על המאורע הזה (הסתברות שלמה). נגדיר את A_k להיות המאורע לפיו הסכום 12 התקבל לראשונה בהטלה ה-k.

$$P(B)=\sum_{k=1}^{\infty}P(B\,|\,A_k)P(A_k)=$$

$$=\sum_{k=1}^{\infty}P(11\,after\,12\,|\,12\,on\,k\,round)\frac{1}{36}=$$

$$=\frac{1}{36}\sum_{k=1}^{\infty}P(10\,or\,less\;\;first\,k-1\,rounds)=$$

$$=\frac{1}{36}\sum_{k=1}^{\infty}\left(\frac{11}{12}\right)^{k-1}=\frac{1}{36}\sum_{k=0}^{\infty}\left(\frac{11}{12}\right)^{k}=$$

$$=\frac{1}{36}\frac{1}{1-\frac{11}{12}}=\frac{1}{3}$$

פרק 4

נושאים:

- משתנה מקרי בדיד.
- התפלגויות של מ"מ בדידים – ברנולי, בינומי, גיאומטרי ועוד.

משתנים מקריים

מהו משתנה מקרי? משתנה מקרי זו פונקציה מ- Ω ל- $\mathbb{R}$. בגלל שמשתנה מקרי זו פונקציה מ- Ω ל- $\mathbb{R}$, אנחנו יכולים לשאול מה ההסתברות שהיא תקבל ערך מסוים.

לדוגמא:

- אם X הוא תוצאת הטלת קובייה בודדת. אזי הערכים שהוא מקבל הם $\{1,2,3,4,5,6\}$ וכל ערך מתקבל בהסתברות $\frac{1}{6}$. לכן נוכל לומר שההסתברות ש-X גדול מ-4, קרי X הוא 5 או 6, היא שליש.
- אם X הוא סכום הטלת צמד קוביות, אזי הערכים ש-X מקבל הם - $\{2,3,\ldots,12\}$ וכל ערך מתקבל בהסתברות כלשהי (לא סימטרי).

פונקצית ההסתברות של משתנה מקרי בדיד X היא הפונקציה -

$$P_X(x)=P(X=x)=P\left(\{\omega\in\Omega \mid X(\omega)=x\}\right)$$

פונקצית ההתפלגות המצטברת

פונקצית התפלגות מצטברת של מ"מ X היא הפונקציה

$F_X(x)=P(X\le x)=P\left(\{\omega\in\Omega \mid X(\omega)\le x\}\right)$. זוהי פונקציה בעלת התכונות הבאות:

1. פונקציה מונוטונית לא יורדת.
2. רציפה מימין.
3. $\lim\limits_{x\to-\infty} F_X(x)=0, \quad \lim\limits_{x\to\infty} F_X(x)=1$.

תרגיל 1: מטילים 3 קוביות ויהי X התוצאה הגבוהה ביותר שהתקבלה.

א. חשבו את פונקצית ההסתברות של X, $P(X=k)$ עבור $k=1,\ldots,6$.

ב. חשבו ושרטטו את גרף פונקצית ההתפלגות המצטברת $F_X(t)=P(X\le t)$.

הערה: שימו לב כי באותה מידה יכולנו לתאר את הבעיה בתור הטלת קובייה הוגנת 3 פעמים ברצף.

פתרון תרגיל 1:

א. אנחנו נדרשים למצוא את $P\left(X=k\right)$ עבור $k=1,\dots,6$.

נתחיל בכך שמרחב המדגם של 3 תוצאות אפשריות מונה $6^3=216$ תוצאות אפשריות שוות הסתברות.

נגדיר את המאורעות הבאים:

- A_k - "התוצאה הגבוהה מבין השלוש קטנה ו/או שווה ל-k (לכל היותר k)".
- B_k - "התוצאה הגבוהה מבין השלוש היא בדיוק k".

כעת נשים לב כי:

$$B_1=A_1$$
$$B_k=A_k \setminus A_{k-1} \ \forall k>1$$

דבר זה נובע מכך שאם לוקחים את כל המאורעות בהם התוצאה היא שווה או קטנה מ-k ומורידים את כל המאורעות בהם התוצאה קטנה או שווה ל-k-1 אז נשארים עם כל המאורעות בהם התוצאה היא בדיוק k.

על כן, כל שנותר לנו הוא למצוא את ההסתברות של A_k ואז

$P\left(B_k\right)=P\left(A_k \setminus A_{k-1}\right)=P\left(A_k\right)-P\left(A_{k-1}\right) \ \forall k>1$ וזאת כי $A_k \supseteq A_{k-1}$.

המאורע A_k מורכב מכל השלשות שקיבלנו בכל הקוביות תוצאה בין 1 ל-k, ז"א שלכל הטלה ישנן k ערכים אפשריים ולכן,

$$P\left(A_k\right)=\frac{k^3}{216} \quad \Rightarrow \quad P\left(B_k\right)=\frac{k^3-\left(k-1\right)^3}{216}=\frac{3k^2-k+1}{216}$$

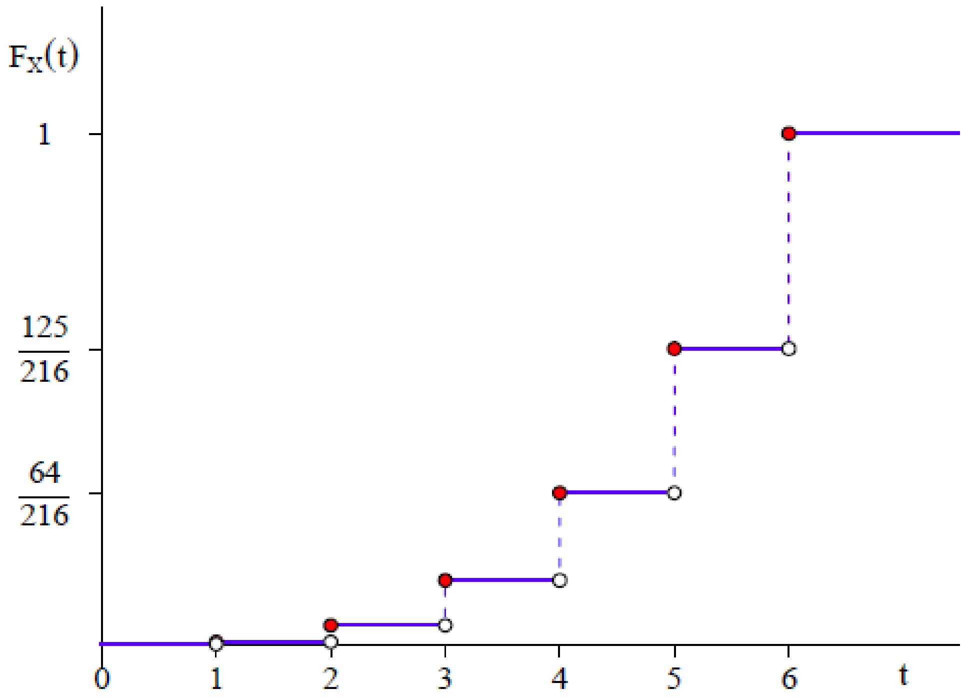
נרשום את ערכי $P\left(X=k\right)$ בטבלה:

K	1	2	3	4	5	6
$P\left(X=k\right)$	$1/216$	$7/216$	$19/216$	$37/216$	$61/216$	$91/216$

ב. נרשום את $F_X\left(t\right)=P\left(X\le t\right)$ מפורשות:

$$F_X\left(t\right)=\begin{cases} 0 & t<1 \\ 1/216 & 1\le t<2 \\ 8/216 & 2\le t<3 \\ 27/216 & 3\le t<4 \\ 64/216 & 4\le t<5 \\ 125/216 & 5\le t<6 \\ 1 & t\ge 6 \end{cases}$$

נשים לב כי הפונקציה היא רציפה מימין ומקיימת את התכונות שציינו קודם לכן.



<u>תרגיל 2</u>: הסיכוי להיכשל במועד א' ו-ב' במבוא להסתברות הוא 0.3 ו-0.5 בהתאמה. רק אלו שנכשלים במועד א' יכולים לגשת למועד ב' ואלו הנכשלים במועד ב' יכולים לבקש מועד מיוחד. בקורס 100 סטודנטים. מה התפלגות מספר הבחינות שייאלצו המרצה והמתרגל המסכנים לחבר? מהי פונקצית ההתפלגות המצטברת? (שאלה זו היפותטית לגמרי ולא בא לומר דבר על סיכויי ההצלחה בקורס... ☺).

<u>פתרון תרגיל 2</u>: נסמן ב-X את מספר המועדים שייאלצו המרצה והמתרגל לחבר.

מכיוון שחייבים לחבר לפחות מבחן 1 (מועד א') אזי:

$$P(X \le 0) = 0 .$$

אם נסמן ב-A את המאורע שכולם עבור במועד א', אז

$$P(X=1) = P(A) = 0.7^{100} .$$

נחשב כעת את $P(X=2)$. נסמן ב- A_k את המאורע "במועד א' נכשלו k סטודנטים ובמועד ב' כל ה-k עברו".

$$P(X=2)=\sum_{k=1}^{100}P(A_k)=\sum_{k=1}^{100}\binom{100}{k}0.3^k0.7^{100-k}0.5^k=$$
$$=\sum_{k=1}^{100}\binom{100}{k}0.15^k0.7^{100-k}=\sum_{k=0}^{100}\binom{100}{k}0.15^k0.7^{100-k}-\binom{100}{0}0.15^0 0.7^{100}=$$
$$=(0.7+0.15)^{100}-0.7^{100}=0.85^{100}-0.7^{100}$$

כאשר המעבר מהשורה השנייה לשלישית בוצעה בעזרת הבינום של ניוטון –

$$(x+y)^n=\sum_{k=0}^{n}\binom{n}{k}x^k y^{n-k}$$

בכדי למצוא את $P(X=3)$ נוכל להשתמש ב-2 דרכים:

- בעזרת מה שחישבנו קודם לכן נקבל

$$P(X=3)=1-P(X\neq 3)=1-P(X=1)-P(X=2)=$$
$$=1-\left(0.85^{100}-0.7^{100}\right)-0.7^{100}=1-0.85^{100}.$$

- בנוסף אם נגדיר את המאורע B כמאורע "אף אחד לא נכשל בשני המועדים" ונחשב את ההסתברות להיכשל בשני המועדים $0.3\cdot 0.5=0.15$ (), אזי

$$P(B)=(1-0.15)^{100}\quad\Rightarrow$$
$$P(X=3)=1-P(B)=1-(1-0.15)^{100}=1-0.85^{100}$$

נרשום כעת את פונקצית ההסתברות המצטברת בטבלה:

$$F_X(t)=\begin{cases}0 & t<1\\ 0.7^{100} & 1\le t<2\\ 0.85^{100} & 2\le t<3\\ 1 & t\ge 3\end{cases}$$

<u>התפלגויות בדידות מיוחדות – ברנולי, בינומי, פאוסוני, גיאומטרי והיפרגאומטרי</u>

<u>ניסוי ברנולי</u>- זהו ניסוי שתוצאותיו האפשריות הן 'הצלחה' בהסתברות p ו-'כישלון' בהסתברות 1-p. על בסיס 'ניסוי ברנולי' מגדירים 'משתנה מקרי ברנולי'. זהו משתנה מקרי המקבל ערכים 0,1 כאשר 0 מסמן 'כישלון' ו-1 'הצלחה'.

<u>מ"מ בינומי</u>- זהו משתנה מקרי הסופר את מספר ההצלחות ב-n ניסויי ברנולי בלתי תלויים כאשר ההסתברות לכל הצלחה היא p. הסימון של מ"מ X כזה היא - $X\sim Bin(n,p)$.

דוגמא אינטואיטיבית לניסוי ברנולי היא הטלת מטבע כאשר על צד אחד של המטבע יש את הספרה '1' ועל הצד השני את הספרה '0'. סכום התוצאות שנקבל אחרי n הטלות יהיה מ"מ בינומי.

פונקצית ההסתברות של מ"מ בינומי היא –

$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}, \quad k=1,2,\dots,n$$

<u>מ"מ גיאומטרי</u>- זהו מ"מ הסופר את מספר הניסויים לבצע ניסויי ברנולי בלתי תלויים עד שהתקבלה ההצלחה הראשונה. נסמן אותו באופן הבא: $X\sim G(p)$ ונגיד ש-X מתפלג גיאומטרית. פונקצית ההסתברות שלו תהיה –

$$P(X=k)=(1-p)^{k-1}p, \quad k=1,2,\dots$$

<u>מ"מ פאוסוני</u>- מ"מ X מתפלג פאוסוני עם פרמטר λ (יסומן כ- $X\sim Pois(\lambda)$) אם

$$P(X=k)=e^{-\lambda}\frac{\lambda^k}{k!}, \quad k=0,1,2,\dots$$

התפלגות פאוסונית מופיעה רבות כאשר דנים במספר אירועים בפרק זמן נתון.

לדוגמא:

- מספר התפרקויות של חלקיקי אלפא מחומר רדיואקטיבי בשנייה.
- מספר טעויות הדפוס בעמוד של עיתון.
- מספר האנשים שת"א בהגיעו לגיל 100.
- מספר השיחות המתקבלות במרכזייה בדקה.

ועוד רבים נוספים. הסיבה לכך היא שהתפלגות פואסונית היא במקרים רבים גבול של התפלגות בינומית כאשר ישנם המון ניסויים וההסתברות להצלחה בכל ניסוי קטנה מאוד (למעשה מדובר בגבול כאשר $\frac{\lambda}{n}=p$ ו-n שואף לאינסוף).

ישנן מספר תכונות חשובות להתפלגות זאת:

- אם מספר אירועים מתפלג פואסוני עם פרמטר λ ביחידת זמן / שטח אחת אזי מספר האירועים ב-n יחידות זמן / שטח יתפלג גם הוא פואסונית עם פרמטר $n\cdot\lambda$.
- בשני קטעי זמן / שטח זרים, מספר האירועים בלתי תלוי זה בזה.
- לא ייתכנו שני אירועים בו-זמנית.

<u>התפלגות פואסונית כקירוב לבינומית</u>

ניתן להגיע להתפלגות הפואסונית על ידי כך שנקרב התפלגות בינומית אם ניקח n שואף לאינסוף ונשמור על המכפלה $\lambda=n\cdot p$ קבועה, כפי שניתן לראות בחישוב הבא:

נניח ש- X הוא משתנה מקרי בינומי $X \sim Bin(n,p)$, ובנוסף נניח ש- n מאוד גדול (לפחות כמה מאות כדי שהקירוב יעבוד), והמכפלה $\lambda = n \cdot p$ קבועה ולא גדולה (ניקח שרירותית את התנאי שהיא קטנה מ-5).

$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}=\frac{n!}{k!(n-k)!}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k}$$
$$=\frac{\lambda^k}{k!}\left[\frac{n!}{(n-k)!n^k}\right]\frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^k}$$

כעת ניתן לקחת את הגבול עבור n שואף לאינסוף ולכן הביטוים שרשומים ראשונים לא יישתנו, מה שבתוך הסוגריים המרובעים מתכנס ל-1 (כאשר k קבוע ו-n שואף לאינסוף), והמונה בשבר האחרון נותן לנו בדיוק את האקספוננט שצריך (המכנה בשבר הזה מתכנס ל-1 מאחר ו-k קבוע ו-n גדל)

$$P(X=k)=\lim_{n\to\infty}\frac{\lambda^k}{k!}\left[\frac{n!}{(n-k)!n^k}\right]\frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^k}=\frac{\lambda^k}{k!}\cdot 1\cdot\frac{e^{-\lambda}}{(1)^k}=e^{-\lambda}\frac{\lambda^k}{k!}$$

<u>מ"מ היפרגאומטרי</u>- נניח שיש אוכלוסיה בגודל N ובה S פרטים מיוחדים. דוגמים ממנה ללא החזרה תת-קבוצה בגודל D. אם X הוא משתנה מקרי שסופר את מספר האיברים המיוחדים במדגם אזי הוא מ"מ היפרגאומטרי עם פרמטרים S,N ו-D. נסמן זאת בתור - $X \sim HG(N,S,D)$ ו-

$$P(X=k)=\frac{\binom{S}{k}\binom{N-S}{D-k}}{\binom{N}{D}} \quad k=1,2,\ldots,\min\{S,D\}$$

<u>תרגיל 3 (מ"מ בינומי)</u>: שחקן מהמר על אחד מן המספרים 1 עד 6 ואז מטילים 3 קוביות. אם המספר עליו הימר השחקן מתקבל בקוביות i פעמים, כאשר $i=1,2,3$, השחקן זוכה ב-i שקלים ואם המספר אינו מתקבל באף קובייה, השחקן מפסיד שקל אחד. מה התפלגות הפרס המתקבל במשחק?

פתרון תרגיל 3: נסמן ב-X את מספר השקלים שבהם זוכה השחקן במשחק. בהנחה שהקוביות תקינות ושהתוצאות המתקבלות בהן בלתי-תלויות, מספר הפעמים שיתקבל המספר שעליו הימר השחקן הוא משתנה מקרי שהוא "כמעט" בינומי עם הפרמטרים $\left(3, \frac{1}{6}\right)$ (כמעט בינומי מאחר והוא יכול לקבל ערכים שליליים). נמצא את התפלגות X:

$$P(X=-1)=\binom{3}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^3=\frac{125}{216}$$
$$P(X=1)=\binom{3}{1}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^2=\frac{75}{216}$$
$$P(X=2)=\binom{3}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^1=\frac{15}{216}$$
$$P(X=3)=\binom{3}{3}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^0=\frac{1}{216}$$

כל ערך אחר X מקבל רק בהסתברות 0.

תרגיל 4 (מ"מ בינומי וגיאומטרי): במבחן בהסתברות ישנן סה"כ 5 שאלות "כן/לא". אם עונים על לפחות 4 שאלות נכון, מקבלים ציון 'עובר'.

א. אלון שלא הספיק ללמוד למבחן מחליט לנחש את הפתרון לכל שאלה ע"י הטלת מטבע הוגן (הסתברות חצי ל-'כן' והסתברות חצי ל-'לא'). מה ההסתברות שעבר את המבחן?

ב. בן, המתרגל הרחמן, מאשר לסטודנטים לגשת לכמה מועדים שיירצו (שאלות שונות בין המועדים). יהי X מספר המועדים שאלון עשה. מה ההסתברות שמספר המועדים שאלון עשה הוא 3? מה ההסתברות שמספר המועדים לפחות 3?

פתרון תרגיל 4:

א. נסמן ב-Y את מספר השאלות שאלון צדק בהן. נשים לב כי $Y \sim Bin\left(5, \frac{1}{2}\right)$ משום שמ"מ בינומי $Bin(n,p)$ סופר את מספר ההצלחות ב-n ניסויים כאשר הניסויים בלתי תלויים וההסתברות להצלחה בכל ניסוי היא p. על כן שואלים אותנו בשאלה,

$$P(Y\ge 4)=P(Y=4)+P(Y=5)=\binom{5}{4}\cdot\left(\frac{1}{2}\right)^4\cdot\left(\frac{1}{2}\right)^1+\binom{5}{5}\cdot\left(\frac{1}{2}\right)^5\cdot\left(\frac{1}{2}\right)^0$$
$$=\binom{5}{4}\cdot\left(\frac{1}{2}\right)^5+\left(\frac{1}{2}\right)^5=\frac{6}{32}=\frac{3}{16}=0.1875$$

ב. X הוא משתנה מקרי גיאומטרי, למה? ההגדרה למ"מ גיאומטרי זה משתנה מקרי שסופר את מספר הניסויים עד להצלחה ראשונה כאשר הניסויים בלתי תלויים

והצלחה בכל ניסוי מתקבל בהסתברות p. כמו כן, את ההסתברות להצלחה במבחן כלשהו מצאנו בסעיף א', על כן, $X \sim G\left(\frac{3}{16}\right)=G(0.1875)$.

נשתמש בנוסחה למ"מ גיאומטרי ונקבל –

$$P(X=3)=(1-0.1875)^2\cdot 0.1875=0.1237$$
$$P(X\geq 3)=1-P(X\leq 2)=1-P(X=1)-P(X=2)=$$
$$=1-(1-0.1875)^0\cdot 0.1875-(1-0.1875)^1\cdot 0.1875=$$
$$=0.1875^2-2\cdot 0.1875+1=0.6601$$

<u>תרגיל 5 (מ"מ גיאומטרי)</u> : מטילים זוג מטבעות שוב ושב עד שאחד מהם מראה "עץ" והשני "פלי". המטבע הראשון מראה "עץ" בהסתברות p והשני בהסתברות q. כל ההטלות הן בלתי תלויות.

א. זהו את התפלגות מספר ההטלות.

ב. מה ההסתברות שבהטלה האחרונה המטבע הראשון הוא זה שמראה "עץ"?

<u>פתרון תרגיל 5</u>:

א. נשים לב כי

$$P(\{(H,T),(T,H)\})=p\cdot(1-q)+q\cdot(1-p)$$

ז"א שההסתברות ל-"הצלחה" בכל סיבוב היא בלתי תלויה בסיבוב שלפניו. בכל סיבוב יש לנו הסתברות להצליח והסתברות זאת היא פונקציה של p ו-q בלבד. נזכור שמספר הניסויים בסדרת ניסויים ברנולי בלתי תלויים מתפלג גיאומטרי לפי הגדרה. לכן אם נסמן ב-X את מספר הניסויים שנבצע (עד ההצלחה) אז נקבל ש-

$$X\sim G\big(p(1-q)+q(1-p)\big)$$

ב. שאלה זו דנה בהסתברות מותנה עם מ"מ גיאומטרי. נסמן ב-A את המאורע שהמטבע הראשון יוצא 'עץ' וב-S את המאורע שהמבוקש (מטבע אחד 'פלי' והשני 'עץ').

$$P(A|S)=\frac{P(A\cap S)}{P(S)}=\frac{p(1-q)}{p(1-q)+q(1-p)}$$

<u>תרגיל 6 (מ"מ בינומי והיפרגאומטרי)</u>: בכנסת ישנן 24 חברי כנסת מתוך ה-120 שהן נשים. בוחרים באקראי יושבי ראש ל-12 הועדות. מה ההסתברות ש-3 מתוכם יהיו נשים אם

א. כל ח"כ יכול לכהן בכמה ועדות שיירצה?

ב. כל ח"כ יכול לכהן בוועדה אחת בלבד?

<u>פתרון תרגיל 6</u>: נסמן ב-X את מס' חברות הכנסת בתפקיד ראש וועדה.

א. בגלל שכל ח"כ יכול לכהן בכמה ועדות שיירצה אזי יש לנו למעשה 12 ניסויים בלתי תלויים כאשר הסיכוי להצלחה בכל ניסוי הוא $p=\frac{24}{120}=\frac{1}{5}$, על כן $X \sim Bin\left(12,\frac{1}{5}\right)$ ו-

$$P(X=3)=\binom{12}{3}\cdot\left(\frac{1}{5}\right)^3\cdot\left(\frac{9}{5}\right)^9=0.2362$$

ב. הפעם הבחירה היא ללא החזרה ולכן יש לנו מצב חדש. הדרך הקלה לחשוב על הבעיה כעת היא באופן הבא: נניח כי קבוצת חברות הכנסת, 24 במספר, זו קבוצת ה-"מיוחדים" שלנו וכעת אנחנו לוקחים מדגם של 12 חברי כנסת מתוך קבוצה של 120 ו-X הוא בדיוק מספר המיוחדים במדגם הזה. על כן מדובר במשתנה מקרי היפרגאומטרי, $X \sim HG(120,24,12)$. מכאן נקבל

$$P(X=3)=\frac{\binom{24}{3}\binom{96}{9}}{\binom{120}{12}}=0.2489$$

באופן לא מפתיע קיבלנו תוצאות די דומות. למה זה לא מפתיע? כי האוכלוסייה יחסית גדולה ולכן ההבדל בין בחירה עם החזרה לבחירה ללא החזרה הוא זניח (למעשה באוכלוסיה גדולה ההסתברויות הן בקירוב בלתי תלויות). זו דוגמא טובה לכך שהתפלגות בינומית היא קירוב סביר להתפלגות היפרגאומטרית כאשר האוכלוסייה גדולה במיוחד.

<u>תרגיל 7 (מ"מ אחיד וגיאומטרי)</u>: בתוך כד יש 10 כדורים הממוספרים ב-1 עד 10. מוצאים מהכד כדורים בזה אחר זה עד שמוציאים את הכדור המסומן ב-5. יהיה X מספר ההוצאות. מצא את התפלגות X בהנחה ש:

א. מוציאים את הכדורים מהכד עם החזרה.

ב. מוציאים את הכדורים מהכד ללא החזרה.

<u>פתרון תרגיל 7</u>:

א. נשים לב כי כאשר מוציאים את הכדורים מן ההכד עם החזרה, אזי בפועל בכל הוצאה מבצעים את אותו הניסוי כאשר ההסתברות בכל שלב להוציא את הכדור שמספרו 5 היא $1/10$. על כן, אם X הוא מספר הפעמים שמוצאים כדור עד להוצאת הכדור עם המספר 5, אזי X הוא מ"מ גיאומטרי עם הסתברות $1/10$, $X \sim G\left(\frac{1}{10}\right)$.

מכאן אנו מקבלים ש- $k=1,2,\ldots$ $P_X(k)=\frac{1}{10}\cdot\left[\frac{9}{10}\right]^{k-1}$.

ב. כאשר מוציאים את הכדורים מהכד ללא החזרה, ניתן לעשות אנלוגיה מבעיה זו לבעיה אחרת, מהי?

מספר הדרכים להוציא את הכדורים מהכד שקול למעשה למספר הדרכים לסדר את הכדורים בשורה. זה נכון מאחר ובכל פעם שנוציא את כדור מהכד נוכל למקם אותו מאחורי הכדורים שהוצאו קודם לכן ולכן יש לנו העתקה חד-חד-ערכית ועל ממספר ההוצאות למספר הסידורים בשורה.

איך זה עזר לנו? ובכן, הרבה יותר קל לשאול את השאלה, מה ההסתברות כעת שהכדור 5 יהיה במקום ה-k כאשר מסדרים את הכדורים בשורה. משיקולי סימטריה, נובע שזה $1/10$ ולכן $k=1,\ldots,10$ $P_X(k)=\frac{1}{10}$ וזהו משתנה מקרי אחיד.

<u>תרגיל 8 (מ"מ פאוסוני ובינומי)</u> : בן ואלון מחפשים דירה בת"א. הם עשו לעצמם רשימה של מאפיינים הכרחיים לדירה (מיזוג, סמיכות לתחבורה ציבורית, ליד פאב שכונתי וכד'). בת"א ישנן $2\cdot10^6$ דירות וכל דירה עומדת ברשימת התנאים הנ"ל בהסתברות $\frac{1}{10^6}$ באופן בלתי תלוי. יהי X מספר הדירות העומדות בתנאי הרשימה.

א. כיצד X מתפלג?

ב. בהינתן שאלון ובן מצאו דירה, מה הסיכוי שהיא הדירה היחידה שעומדת בדרישות?

ג. בהינתן שאלון ובן מצאו 2 דירות, מה הסיכוי שהן היחידות הרלוונטיות?

<u>פתרון תרגיל 8</u>:

א. X הוא למעשה מ"מ בינומי. למה? כי יש לנו קבוצה של $2\cdot10^6$ דירות, הסיכוי שכל דירה מתאימה הוא $\frac{1}{10^6}$ באופן בלתי תלוי, אז זה שקול למצב שיש $2\cdot10^6$ ניסויים וכל ניסוי מצליח בהסתברות $\frac{1}{10^6}$ ו-X סופר את מספר הניסויים / הדירות שהצליחו / מתאימות. זו בדיוק ההגדרה למ"מ בינומי. על כן - $X\sim Bin\left(2\cdot10^6,10^{-6}\right)$.

הערה: נשים לב כי משתנה בינומי עם הערכים הללו הוא בדיוק המצב הקלאסי של קירוב פאוסוני עם פרמטר $\lambda=np=2\cdot10^6\cdot10^{-6}=2$ כי מדובר במספר ניסויים מאוד גדול, בהסתברות מאוד קטנה וההסתברות שואפת לאפס ביחס ישר למספר הניסיונות.

ב. אנו צריכים למצוא למעשה את ההסתברות הבאה $P(X=1|X\geq1)=?$. לשם כך נשתמש בקירוב הפואסוני

$$P\left(X=1|X\geq 1\right)=\frac{P\left(X=1\cap X\geq 1\right)}{P\left(X\geq 1\right)}=\frac{P\left(X=1\right)}{P\left(X\geq 1\right)}=$$
$$=\frac{e^{-2}\frac{2^1}{1!}}{1-e^{-2}\frac{2^0}{0!}}=\frac{2}{e^2-1}=0.313$$

ג. מבצעים חישוב דומה עבור X=2.

$$P\left(X=2|X\geq 2\right)=\frac{P\left(X=2\cap X\geq 2\right)}{P\left(X\geq 2\right)}=\frac{P\left(X=2\right)}{P\left(X\geq 2\right)}=$$
$$=\frac{e^{-2}\frac{2^2}{2!}}{1-e^{-2}\frac{2^0}{0!}-e^{-2}\frac{2^1}{1!}}=\frac{2}{e^2-3}=0.455$$

<u>תרגיל 9 (מ"מ פאוסוני)</u> : לאחר התרגול יוצאים כל הסטודנטים לפאב השכונתי הקרוב. לטובת האירוע, המציא אלון הסטודנט החרוץ משחק שתייה חדש. מהרגע שייכנסו כל הסטודנטים לפאב עליהם לדאוג לכך שבן הברמן יגיש צ'ייסרים לפי תהליך פואסון עם קצב של 2.5 צ'ייסרים בחצי דקה.

א. כיצד מתפלג מספר הצ'ייסרים שיגיש בן בחצי דקה הראשונה? בדקה הראשונה?
ב. מה ההסתברות שבשתי הדקות הראשונות יגישו לפחות 10 צ'ייסרים?
ג. בהינתן שב-3 דקות הראשונות קיבל כל סטודנט בקבוצה (סה"כ 30 סטודנטים) צ'ייסר על חשבון הבית. מה התפלגות מספר הצ'ייסרים בדקה הראשונה?

<u>פתרון תרגיל 9</u>:

א. נגדיר את X_k להיות מספר הצ'ייסרים שהוגשו עד הדקה ה-K. עבור $X_{0.5}$ - מספר הצ'ייסרים בחצי הדקה הראשונה נתון שהוא מתפלג פואסוני עם קצב של 2.5 צ'ייסרים $X_{0.5}\sim P\left(2.5\right)$. עבור X_1, מספר הצ'ייסרים בדקה הראשונה, אנחנו יודעים שישנה התכונה שהתפלגות היא גם פואסונית עם פרמטר $X_1\sim P\left(5\right)$.

ב. מספר הצ'ייסרים ב-2 דקות הראשונות מתפלג $X_2\sim P\left(10\right)$. נחשב מפורשות את ההסתברות בעזרת התפלגות פואסון:

$$P\left(X_2\geq 10\right)=1-P\left(X_2<10\right)=1-\sum_{k=0}^{9}P\left(X_2=k\right)=1-\sum_{k=0}^{9}e^{-10}\frac{10^k}{k!}$$

ג. ידוע כי מספר הצ'ייסרים בדקה הראשונה ומהדקה הראשונה עד השלישית הם בלתי תלויים (תכונת ההפרשים הבלתי תלויים) ולכן X_1 ו- X_3-X_1 הם משתנים מקריים בלתי תלויים. נחשב את הנדרש בעזרת הסתברות מותנה:

$$P\left(X_1=k \mid X_3=30\right)=\frac{P\left(X_1=k, X_3=30\right)}{P\left(X_3=30\right)}=\frac{P\left(X_1=k, X_3-X_1=30-k\right)}{P\left(X_3=30\right)}=$$

$$=\frac{P\left(X_1=k\right)P\left(X_3-X_1=30-k\right)}{P\left(X_3=30\right)}=\frac{e^{-5}\cdot\frac{5^k}{k!}\cdot e^{-10}\cdot\frac{10^{30-k}}{(30-k)!}}{e^{-15}\cdot\frac{15^{30}}{30!}}=$$

$$=\binom{30}{k}\left(\frac{5}{15}\right)^k\left(\frac{10}{15}\right)^{30-k}$$

פרק 5

נושאים:

- תוחלת (משתנה מקרי בדיד).
- שונות (משתנה מקרי בדיד).

תוחלת

יהי משתנה מקרי בדיד X המקבל את הערכים $x_1, x_2, \ldots$. התוחלת של X (expectation) מסומנת על ידי $E(X)$ ומוגדרת על ידי:

$$E(X) = \sum_i x_i \cdot P(X = x_i)$$

למעשה התוחלת היא ממוצע משוקלל של הערכים שהמשתנה המקרי יכול לקבל כפול ההסתברויות לקבל את אותם ערכים.
תכונה שמאוד (!) חשוב לזכור לגבי תוחלת היא שהתוחלת היא ליניארית, ז"א $E(X+Y) = E(X) + E(Y)$ עבור X ו-Y משתנים מקריים כלשהם ו- $E(aX+b) = aE(X) + b$ כאשר a,b קבועים כלשהם.

את הכלל הנ"ל כמובן ניתן להכליל לכל מספר של משתנים מקריים. לדוגמא: יהיו $X_1, \ldots, X_n$ משתנים מקריים וכן $a_1, \ldots, a_n \in \mathbb{R}$ מספרים ממשיים, אזי

$$E\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i E(X_i)$$

ומכך ניתן להסיק את התכונה הבאה:

$$E(g(X)) = \sum_x g(x) \cdot P(X = x)$$

עבור g פונקציה "טובה" כלשהי (נדון בהמשך באיזו פונקציות מדובר, בד"כ רציפות).

מה התוחלת של משתנים מקריים בדידים שראינו כבר? נעבור על כל משתנה מקרי שראינו ונחשב את התוחלת שלו מההגדרה.
<u>תוחלת של מ"מ אחיד</u>: יהי $X \sim U[a,b]$ ולכן X מקבל ערכים מ-a עד b כאשר ההסתברות לקבל כל ערך היא שווה. התוחלת של X היא

$$E(X)=\sum_{k=a}^{b}k\cdot P(X=k)=\sum_{k=a}^{b}k\cdot\frac{1}{b-a}=\frac{1}{b-a}\sum_{k=a}^{b}k$$
$$=\frac{1}{b-a}\frac{(b+a)\cdot(b-a)}{2}=\frac{(b+a)}{2}$$

<u>תוחלת של מ"מ ברנולי</u>: יהי $p\in[0,1]$, $X\sim Bernoulli(p)$ ולכן X מקבל את הערך 1 בהסתברות p ו-0 בהסתברות 1-p.

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

<u>תוחלת של מ"מ בינומי</u>: יהי $p\in[0,1]$, $n\in\mathbb{N}$, $X\sim Bin(n,p)$.

$$E(X)=\sum_{k=0}^{n}k\cdot P(X=k)=\sum_{k=0}^{n}k\cdot\binom{n}{k}p^k(1-p)^{n-k}=\sum_{k=1}^{n}k\cdot\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}$$
$$=np\sum_{k=1}^{n}\frac{n-1!}{k-1!(n-k)!}p^{k-1}(1-p)^{n-k}=np\sum_{k=1}^{n}\binom{n-1}{k-1}p^{k-1}(1-p)^{n-k}$$
$$=np\sum_{k=0}^{n}\binom{n-1}{k}p^k(1-p)^{n-k-1}=np(p+1-p)^{n-1}=np$$

<u>תוחלת של מ"מ גיאומטרי</u>: יהי $p\in[0,1]$, $X\sim G(p)$, אזי נסמן q=1-p ונקבל

$$E(X)=\sum_{k=1}^{\infty}k\cdot P(X=k)=\sum_{k=1}^{\infty}k\cdot pq^{k-1}$$
$$=p\sum_{k=1}^{\infty}kq^{k-1}=p\sum_{k=1}^{\infty}\frac{d}{dq}\left[q^k\right]=p\frac{d}{dq}\left[\sum_{k=1}^{\infty}q^k\right]$$
$$=p\frac{d}{dq}\left[\sum_{k=1}^{\infty}q^k\right]=p\frac{d}{dq}\left[\frac{1}{1-q}\right]=\frac{p}{(1-q)^2}=\frac{p}{p^2}=\frac{1}{p}$$
.

<u>תוחלת של מ"מ היפרגיאומטרי</u>: יהי $X\sim HG(N,S,D)$, אזי $E(X)=D\frac{S}{N}$.

נוותר על החישוב הפעם ☺...

שונות

השונות של משתנה מקרי X (במידה וקיימת) המסומנת על ידי $V(X)$ (variance) מוגדרת באופן הבא:

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

נשים לב לתכונה הבאה:

$$\begin{aligned}V(X)&=E\left(\left[X-E(X)\right]^2\right)=E\left(X^2-2XE(X)+E(X)^2\right)\\&=E\left(X^2\right)-E\left(2XE(X)\right)+E\left(E(X)^2\right)\\&=E\left(X^2\right)-2E(X)E(X)+E(X)^2=E\left(X^2\right)-2E(X)^2+E(X)^2\\&=E\left(X^2\right)-E(X)^2\end{aligned}$$

סטיית התקן של מ"מ X מוגדרת כ- $\sigma(X)=\sqrt{V(X)}$.

תכונות חשובות של שונות – תשימו לב כי בעזרת הצבה ופישוט אלגברי מקבלים כי עבור מ"מ X וקבועים $a,b\in\mathbb{R}$ מתקיים $V(aX+b)=a^2V(X)$.

תרגיל 1: לקבוצת כדורגל מתוכננים שני משחקים. יש לה סיכוי של 0.4 לא להפסיד את המשחק הראשון, ו-0.7 לא להפסיד את השני באופן בלתי תלוי בתוצאות המשחק הראשון. במידה והקבוצה לא מפסידה במשחק כלשהו, יש לה סיכוי שווה לתיקו ולניצחון. הקבוצה מקבלת 3 נקודות לניצחון, נקודה לתיקו ו-0 נקודות להפסד. מצאו את התפלגות, תוחלת ושונות מספר הנקודות של הקבוצה.

פתרון תרגיל 1: נגדיר את X להיות מספר הנקודות בהן זוכה הקבוצה לאחר שני המשחקים.הערכים ש-X יכול לקבל הם: 0,1,2,3,4,6. נסמן ניצחון ב-W, שוויון ב-T והפסד ב-L. נתון כי ההסתברות ל-L במשחק הראשון היא 0.6, ההסתברות ל-T ו-W היא 0.2. במשחק השני ההסתברות ל-L היא 0.3 וההסתברות ל-T או W היא 0.35 כ"א. נמצא את ההסתברות לקבלת כל ערך:

$$P(X=0)=P(\{(L,L)\})=0.6\cdot 0.3=0.18$$

$$P(X=1)=P(\{(L,T),(T,L)\})=0.2\cdot 0.3+0.6\cdot 0.35=0.27$$

$$P(X=2)=P(\{(T,T)\})=0.35\cdot 0.2=0.07$$

$$P(X=3)=P(\{(L,W),(W,L)\})=0.6\cdot 0.35+0.2\cdot 0.3=0.27$$

$$P(X=4)=P(\{(T,W),(W,T)\})=0.2\cdot 0.35+0.35\cdot 0.2=0.14$$

$$P(X=6)=P(\{(W,W)\})=0.2\cdot 0.35=0.07$$

נבדוק האם קיבל קיבלנו תוצאות הגיוניות:

$$P(X=0)+P(X=1)+P(X=2)+P(X=3)+(X=4)+P(X=6)$$
$$=0.18+0.27+0.07+0.27+0.14+0.07=1$$

שזה כבר טוב... ☺

לכן קיבלנו את ההתפלגות של מספר הנקודות של הקבוצה.

כעת נחשב את התוחלת.

$$E(X)=\sum_{\substack{k=0\\k\neq 5}}^{6}k\cdot P(X=k)=0\cdot 0.18+1\cdot 0.27+2\cdot 0.07$$
$$+3\cdot 0.27+4\cdot 0.14+6\cdot 0.07=2.2$$

באותו האופן נמצא את $E(X^2)$ ובעזרתו נחשב את השונות של X.

$$E(X^2)=\sum_{\substack{k=0\\k\neq 5}}^{6}k^2\cdot P(X=k)=0\cdot 0.18+1\cdot 0.27$$
$$+4\cdot 0.07+9\cdot 0.27+16\cdot 0.14+36\cdot 0.07=7.74$$

ולכן,

$$V(X)=E(X^2)-E(X)^2=7.74-4.84=2.9$$

תרגיל 2: בכד ישנם N כדורים ממוספרים מ-1 עד N. מוציאים n כדורים ללא החזרה. יהי X המספר הגבוה ביותר מבין הכדורים שהוצאו.

א. רשמו את התפלגות X.

ב. חשבו את תוחלת X.

<u>פתרון תרגיל 2</u>:

נשים לב כי הערכים היחידים ש-X מקבל הם $n,n+1,\dots,N$ וזה בגלל שברגע ששולפים n כדורים אז לכל הפחות הערך המקסימאלי יהיה n.

סך כל האפשרויות להוציא n כדורים מתוך N הוא כאמור $\binom{N}{n}$. אם בנוסף נדרוש שהמספר המרבי לא יעלה על k, הרי שעלינו להצטמצם לבחירה של n כדורים מתוך k הכדורים הנושאים את המספרים הנמוכים, ולכך יש $\binom{k}{n}$ אפשרויות. על כן,

$$P(X\le k)=\frac{\binom{k}{n}}{\binom{N}{n}}$$

באופן דומה נקבל,

$$P(X\le k-1)=\frac{\binom{k-1}{n}}{\binom{N}{n}}$$

לכן מכאן נובע ש-

$$P(X=k)=P(X\le k)-P(X\le k-1)=\frac{\binom{k}{n}-\binom{k-1}{n}}{\binom{N}{n}}$$

נפשט את הביטוי במונה באופן הבא:

$$\binom{k}{n}-\binom{k-1}{n}=\frac{k!}{n!(k-n)!}-\frac{(k-1)!}{n!(k-1-n)!}=\frac{(k-1)!}{n!(k-1-n)!}\left[\frac{k}{(k-n)}-1\right]$$
$$=\frac{(k-1)!}{n!(k-1-n)!}\left[\frac{n}{(k-n)}\right]=\frac{(k-1)!}{(n-1)!(k-n)!}=\binom{k-1}{n-1}$$

ולכן מקבלים כי,

$$P(X=k)=P(X\le k)-P(X\le k-1)=\frac{\binom{k-1}{n-1}}{\binom{N}{n}}\quad k=n,\dots,N$$

שימו לב כי מהידיעה שסכום ההסתברויות שווה ל-1 קיבלנו את הזהות הבאה:

$$P(X=k)=1 \quad \Rightarrow \quad \frac{1}{\binom{N}{n}}\sum_{k=n}^{N}\binom{k-1}{n-1}=1 \quad \Rightarrow \quad \sum_{k=n}^{N}\binom{k-1}{n-1}=\binom{N}{n}$$

$$\underset{k-1=l}{\Rightarrow} \quad \sum_{l=n-1}^{N-1}\binom{l}{n-1}=\binom{N}{n} \quad \Rightarrow \quad \sum_{l=n}^{N}\binom{l}{n}=\binom{N+1}{n+1}$$

נפתור כעת את סעיף ב'.

$$E(X)=\sum_{k=n}^{N}k\cdot P(X=k)=\frac{1}{\binom{N}{n}}\sum_{k=n}^{N}k\frac{(k-1)!}{(k-n)!(n-1)!}$$

$$=\frac{n}{\binom{N}{n}}\sum_{k=n}^{N}\frac{k!}{(k-n)!n!}=\frac{n}{\binom{N}{n}}\sum_{k=n}^{N}\binom{k}{n}=\frac{n}{\binom{N}{n}}\binom{N+1}{n+1}=\frac{n(N+1)}{(n+1)}$$

<u>תרגיל 3</u>: מסובבים סביבון 5 פעמים. יהיה X מספר הפעמים שיצא נס. כיצד מתפלג X ?
חשבו את ההתפלגות של $Y=2X+1$. חשבו את $E(X),E(Y),V(X)$.

<u>פתרון תרגיל 3</u>: קל לראות כי $X\sim Bin\left(5,\frac{1}{4}\right)$, לכן נותר לנו למצוא את ההתפלגות של Y. הערכים ש-Y יכול לקבל הם הערכים האי זוגיים שבין 1 ל- 11.
עבור $k=1,3,5,7,9,11$ נקבל:

$$P(Y=k)=P(2X+1=k)=P\left(X=\frac{k-1}{2}\right)=\binom{5}{\frac{k-1}{2}}\left(\frac{1}{4}\right)^{\frac{k-1}{2}}\left(\frac{3}{4}\right)^{5-\frac{k-1}{2}}$$

נמצא את התוחלת כעת של כל אחד מהמשתנים המקריים.

$$E(X)=5\cdot\frac{1}{4}=1.25 \quad \Rightarrow \quad E(Y)=E(2X+1)=2E(X)+1=3.5$$

$$V(X)=5\cdot\frac{1}{4}\cdot\frac{3}{4}=0.9375 \quad \Rightarrow \quad V(Y)=V(2X+1)=4E(X)=3.75$$

<u>תרגיל 4</u>: יהיו שני משתנים מקריים X,Y כך ש-

$X \sim Bin(n,p), Y \sim HG(N,D,S)$. מצאו את $E(X),E(Y)$.

<u>פתרון תרגיל 4</u>: נשתמש בטכניקה חדשה לטובת פתרון התרגיל והיא- פירוק לאינדיקאטורים. משתנה מקרי בינומי X ניתן לפרק לאינדיקאטורים כך שכל אינדיקאטור מציין את תוצאת כל ניסוי ומתפלג לפי:

$$X=\sum_{i=1}^{n} X_i \quad , \quad X_i=\begin{cases}1 & p\\ 0 & 1-p\end{cases}$$

בצורה כזאת, אם הייתה הצלחה בניסוי כלשהו, אזי המשתנה המקרי המציין שלו מקבל את הערך 1 ובמקביל נוסף ערך '1' לסכום $X=\sum_{i=1}^{n} X_i$. כעת נוכל להשתמש בליניאריות של תוחלת ולקבל:

$$E(X)=E\left(\sum_{i=1}^{n} X_i\right)=\sum_{i=1}^{n} E(X_i)=\sum_{i=1}^{n}\left[1\cdot p+0\cdot(1-p)\right]=\sum_{i=1}^{n} p=np$$

נבצע את אותו הפירוק עבור משתנה מקרי היפרגאומטרי Y.

$$Y=\sum_{i=1}^{D} Y_i$$

פירוק של משתנה מקרי היפרגאומטרי מתבצע לפי גודל המדגם. במקום להסתכל על Y שסופר את מספר המיוחדים במדגם בגודל D אנחנו יכולים להסתכל על סכום של D משתנים מקריים כאשר כל אחד מהם מתייחס לאובייקט במדגם ומקבל את הערך 1 אם הוא מיוחד ואת הערך 0 אם הוא לא מיוחד (נשים לב שמשתנים מקריים אלו תלויים, אבל בשל הליניאריות של התוחלת זה לא יפריע לנו).

$$E(Y)=E\left(\sum_{i=1}^{D} Y_i\right)=\sum_{i=1}^{D} E(Y_i)=\sum_{i=1}^{D}\left[1\cdot\frac{S}{N}+0\cdot\left(1-\frac{S}{N}\right)\right]=\sum_{i=1}^{D}\frac{S}{N}=D\frac{S}{N}$$

אומנם המשתנים המקריים המציינים הללו הם תלויים, אך כאשר בוחנים מה ההסתברות שכל אחד מהם יהיה מיוחד אזי (משיקולי סימטריה) לכולם יש את אותה הסתברות להיות מיוחדים ולהיות לא מיוחדים ולכן כל התוחלת הללו שווות למה שמצוין לעיל.

פרק 6

נושאים:

- משתנים מקריים רציפים.
- תהליך פואסון.
- התפלגויות אחידה ומעריכית (רציפות).
- תוחלת ושונות של מ"מ צריף.
- התפלגות נורמלית.

משתנה מקרי רציף

עד כה נתקלנו רק במשתנים מקריים בדידים שמקבל ערכים בדידים (לעיתים מספר סופי של ערכים ולעיתים אינסופי). ישנה עוד מחלקה של משתנים מקריים וזו מחלקת המשתנים המקריים הרציפים. משתנים מקריים אלו מקבלים רצף של ערכים על הציר הממשי (יכול להיות גם רק על חלק מהציר הממשי).

למשתנים מקריים רציפים ישנה התפלגות ופונקצית התפלגות מצטברת כפי שראינו עד כה, אולם במקרה של משתנה מקרי רציף – פונקצית ההתפלגות המצטברת שלו היא רציפה ולא רק רציפה מימין.

בנוסף, ניתן להגדיר למשתנה מקרי רציף X את פונקצית הצפיפות שלו שמסומנת על ידי $f_X(\cdot)$ והיא למעשה מבטאת את צפיפות ההסתברות של המשתנה המקרי ומקיימת:

$$F_X(k)=\int_{-\infty}^{k} f_X(t)dt \quad \forall k\in\mathbb{R}$$

בנוסף, פונקצית הצפיפות מקיימת - $f_X(k)=\frac{d}{dk}F_X(k)$.

ישנו תנאי הכרחי ומספיק לכך שפונקציה תהיה פונקצית צפיפות של משתנה מקרי רציף והוא

$$\int_{-\infty}^{\infty} f_X(t)dt=1 \quad , \quad f_X(t)\geq 0 \;\; \forall t\in\mathbb{R} \; .$$

תרגיל 1: אחוז התשובות הנכונות X של סטודנט במבחן באסטרולוגיה מפולג לפי פונקצית הצפיפות הבאה:

$$f_X(t)=\begin{cases} ct(100-t) & c_1\leq t\leq c_2 \\ 0 & otherwise \end{cases} \quad , \quad c>0$$

א. אילו ערכים יכולים לקבל c_1, c_2 ?

ב. בהינתן $c_1 = 0\,,\, c_2 = 100$ חשבו את c.

ג. מצאו את $F_X(t)$.

ד. חשבו את ההסתברות שתלמיד ייכשל (יקבל מתחת ל-55).

ה. חשבו את ההסתברות להיכשל אם ידוע שהציון שהסטודנט קיבל גבוה מ-40.

ו. 5 סטודנטים ניגשו למבחן. מה ההסתברות שכולם יעברו?

<u>פתרון תרגיל 1</u>:

א. בגלל שאנו מעוניינים שהפונקציה תהיה פונקצית צפיפות של משתנה מקרי רציף אזי היא חייבת לקיים - $f_X(t) \ge 0 \quad \forall t \in \mathbb{R}$. מכאן נובע כי התנאי הבא חייב להתקיים:

$$ct(100-t) \ge 0 \quad for \quad c_1 \le t \le c_2 \quad \Leftrightarrow \quad 0 \le t \le 100$$

לכן מה שחייב לקרות זה $0 \le c_1 \le c_2 \le 100$.

ב. נדרוש כי האינטגרל של פונקצית הצפיפות על כל הציר הממשי יהיה שווה ל-1.

$$1 = \int_{-\infty}^{\infty} f_X(t)\,dt = \int_0^{100} ct(100-t)\,dt = c\left[50t^2 - \frac{t^3}{3}\right]_0^{100} = c100^2\left[50 - \frac{100}{3}\right]$$

$$= c100^2\left[\frac{100}{6}\right] = c\frac{10^6}{6} \quad \Rightarrow \quad c = 6\cdot 10^{-6}$$

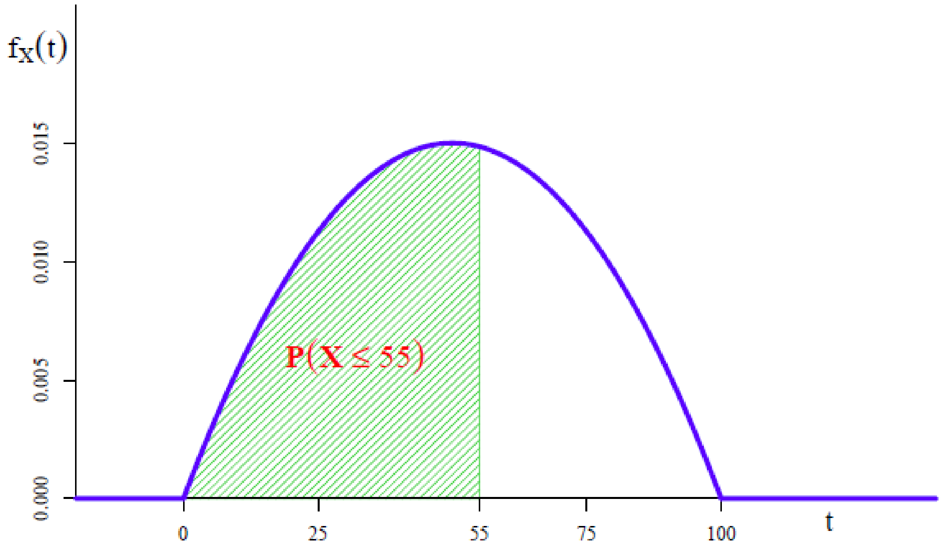
ג. נחשב את $F_X(t)$ לפי $F_X(k) = \int_{-\infty}^{k} f_X(t)\,dt \quad \forall k \in \mathbb{R}$.

$$F_X(k) = \int_{-\infty}^{k} f_X(t)\,dt = \begin{cases} 0 & k < 0 \\ 2\cdot 10^{-6} k^2 (150-k) & 0 \le k \le 100 \\ 1 & k > 100 \end{cases}$$

ד. נחשב את ההסתברות ישירות מפונקצית ההתפלגות המצטברת שמצאנו –

$$P(X < 55) = F_X(55) = 0.57475.$$

להן גרף הפונקציה שקיבלנו.



ה. נתון לנו כי הסטודנט קיבל מעל 40 וכעת שואלים אותנו מה ההסתברות שהוא נכשל.

$$P(X<55 \mid X>40)=\frac{P(\{X<55\}\cap\{X>40\})}{P(X>40)}=\frac{P(40<X<55)}{1-P(X\le 40)}$$
$$=\frac{F_X(55)-F_X(40)}{1-F_X(40)}\approx 0.484$$

ו. אנו מניחים כי הסטודנטים הם בלתי תלויים ונסמן את המאורע שכל החמישה עברו בתור A ושכל סטודנט עבר ב- A_i עבור הסטודנט ה-i.

$$P(A)=P\left(\bigcap_{i=1}^{5}A_i\right)=\prod_{i=1}^{5}P(A_i)=\left[P(X\ge 55)\right]^5$$
$$=\left[1-P(X<55)\right]^5\approx 0.014$$

<u>תרגיל 2</u>: נתונה הפונקציה הבאה

$$F(t)=\begin{cases}0 & t<0\\ \dfrac{t}{1+t} & otherwise\end{cases}$$

א. האם פונקציה זו הינה פונקצית התפלגות מצטברת של מ"מ רציף X?

ב. מצאו את הצפיפות של המ"מ X.

ג. מהו האחוזון ה-50 (החציון) של X?

<u>פתרון תרגיל 2:</u>

א. נבדוק שהפונקציה מקיימת את הדרישות של פונקצית התפלגות מצטברת.

ברור כי כאשר $t \to -\infty$ אז הפונקציה שווה ל-0 ובכיוון ההפוך כאשר $t \to \infty$

הפונקציה שווה ל- $\lim_{t\to\infty} F(t) = \lim_{t\to\infty} \frac{t}{1+t} = \lim_{t\to\infty} \frac{1}{\frac{1}{t}+1} = 1$ כנדרש.

בנוסף הפונקציה היא מונוטונית לא יורדת מאחר והנגזרת שלה גדולה מ-0 לכל t חיובי.

כל אלו גורמים לכך שהפונקציה יכולה להיות פונקצית התפלגות מצטברת של מ"מ X.

ב. נגזור את הפונקציה בכדי למצוא את הצפיפות.

$$f_X(t) = \frac{d}{dt}F(t) = \begin{cases} 0 & t \le 0 \\ \frac{1}{(1+t)^2} & t > 0 \end{cases}$$

ג. האחוזון ה-50 , t, מקיים את התכונה הבאה - $P(X \le t) = 0.5$

נחשב אותו ישירות.

$$0.5 = P(X \le t) = \frac{t}{1+t} \quad \Leftrightarrow \quad 0.5(1+t) = t$$
$$\Leftrightarrow \quad 0.5 = 0.5t \quad \Leftrightarrow \quad t = 1$$

בדיון על משתנים מקריים רציפים, נרצה לתת תשומת לב מיוחדת לשתי משפחות ספציפיות

- משתנים מקריים אחידים.
- משתנים מקריים מעריכיים.

<u>משתנה מקרי אחיד רציף</u>

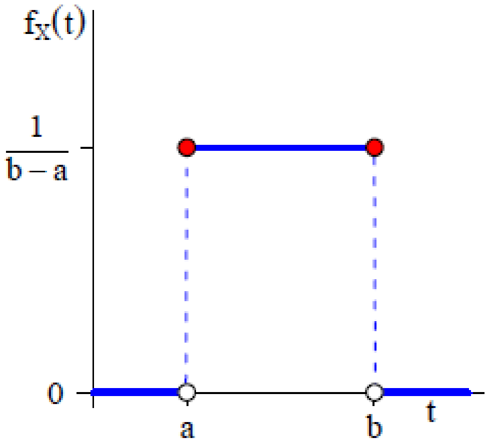
משתנה מקרי אחיד זה משתנה מקרי המקבל רצף של ערכים כאשר ההסתברות לקבל ערכים מקטע מסוים תלוי אך ורק באורך הקטע. אומרים כי X הוא מתפלג אחיד בקטע $[a,b]$, $X \sim U[a,b]$, אם

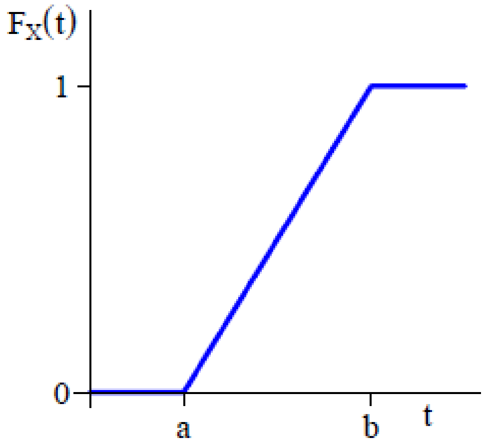
$$f_X(t) = \begin{cases} \frac{1}{b-a} & a \le t \le b \\ 0 & otherwise \end{cases}$$

או לחילופין,

$$F_X(t)=\begin{cases}0 & t<b\\ \dfrac{t-a}{b-a} & a\le t\le b\\ 1 & b<t\end{cases}$$

להלן גרפי הפונקציות הללו.





<u>תרגיל 3 (מ"מ רציף אחיד ללא צורך בתוחלת)</u>: בן הוא מהנדס בכיר. זמן היציאה של בן מהעבודה בערב, X , מתפלג $X \sim U[7,9]$ ומשך הנסיעה שלו הביתה הינו $Y=1+\dfrac{1}{X}$. מצאו את התפלגות וצפיפות Y.

<u>פתרון תרגיל 3</u>: נתחיל במציאת פונקצית ההתפלגות המצטברת של זמן הנסיעה.

ישנם מספר דברים ודאיים שנוכל להתחיל מהם, כמו:

$$F_Y(t)=\begin{cases}0 & t<\dfrac{10}{9}=1+\dfrac{1}{9}\\ ? & \dfrac{10}{9}\le t\le\dfrac{8}{7}\\ 1 & 1+\dfrac{1}{7}=\dfrac{8}{7}<t\end{cases}$$

הסיבה שאנחנו ודעים זאת היא כי X מקבל ערכים בין 7 ל-9 בלבד ולכן כל דבר שמחוץ לטווח זה מתקבל בהסתברות 0. עבור $\dfrac{10}{9}\le t\le\dfrac{8}{7}$ נקבל

$$F_Y(t)=P(Y\le t)=P\left(1+\frac{1}{X}\le t\right)=P\left(\frac{1}{t-1}\le X\right)=$$

$$=1-\begin{cases}0 & \frac{1}{t-1}<7\\ \frac{\frac{1}{t-1}-7}{9-7} & 7\le\frac{1}{t-1}\le 9\\ 1 & \frac{1}{t-1}>9\end{cases}=\begin{cases}1 & \frac{8}{7}<t\\ \frac{9t-10}{2(t-1)} & \frac{10}{9}\le t\le\frac{8}{7}\\ 0 & \frac{10}{9}>t\end{cases}$$

שימו לב שהמעבר האחרון בשורה הראשונה נכון רק בגלל שאנחנו יכולים להתייחס רק למקרים בהם t גדול מ-1.

נוכל כעת לגזור את פונקצית ההתפלגות המצטברת ולקבל את צפיפות של Y.

$$f_Y(t)=\frac{d}{dt}F_Y(t)=\begin{cases}0 & \frac{8}{7}<t\\ \frac{1}{2(t-1)^2} & \frac{10}{9}\le t\le\frac{8}{7}\\ 0 & \frac{10}{9}>t\end{cases}$$

<u>תרגיל 4 (מ"מ רציף אחיד ובדיד מעורב ללא צורך בתוחלת)</u>: אלון המהנדס יוצא מביתו אל התחנה הקרובה. בהסתברות של 2/3 יחכה לו שם בן עם רכבו החדש ואז הם ייסעו יחד לעבודה. במידה ובן לא יגיע, אז ידוע כי המונית הבאה מגיעה בעוד T דקות כאשר $T\sim U[0,10]$ ובכל מקרה אוטובוס מגיע בעוד 5 דקות בדיוק. אלון ייסע עם מי שיגיע קודם. מצאו את פונקצית ההתפלגות המצטברת של משך ההמתנה של אלון.

<u>פתרון תרגיל 4</u>: נסמן את משך ההמתנה של אלון ב-X. המצב בו אלון נמצא הוא מיוחד מבחינה הסתברותית. מצד אחד, ישנה אפשרות (הסתברות חיובית ממש) שזמן ההמתנה יהיה אפס בדיוק. מצד שני זמן ההמתנה גם מתפלג באופן רציף.

נתחיל במציאות הערכים הברורים של פונקצית ההתפלגות המצטברת. אנחנו יודעים בודאות שאלון לא יכול לחכות פחות מ-0 דקות ולכן עבור כל t הקטן ממש מ-0 נקבל שפונקצית ההתפלגות המצטברת היא 0. כמו כן, האוטובוס בודאות מגיע תוך 5 דקות ולכן בודאות אלון יחכה לכל היותר 5 דקות. מכאן נקבל:

$$F_X(t)=\begin{cases}0 & t<0\\ \dfrac{2}{3} & 0=t\\ ? & 0\le t\le 5\\ 1 & 5\le t\end{cases}$$

כעת צריכים לנתח את החלק הרציף של הבעיה. מה קורה בזמן שבין 0 לבין 5 דקות?

נגדיר את המאורע A כמאורע בו בן מחכה לאלון בתחנה. עבור כל t בין 0 ל-5 (שאינו שווה ל-0 או 5) נקבל

$$F_X(t)=P(X\le t)=P(X\le t\mid A)P(A)+P\left(X\le t\mid \overline{A}\right)P\left(\overline{A}\right)$$
$$=1\cdot\frac{2}{3}+P\left(X\le t\mid \overline{A}\right)\cdot\frac{1}{3}=\frac{2}{3}+P\left(X\le t\mid \overline{A}\right)\cdot\frac{1}{3}$$

נמצא את $P\left(X\le t\mid \overline{A}\right)$. נגדיר את B להיות המאורע שמונית מגיעה בזמן t.

$$P\left(X\le t\mid \overline{A}\right)=P\left(X\le t\mid \overline{A},B\right)P(B)+P\left(X\le t\mid \overline{A},\overline{B}\right)P\left(\overline{B}\right)=$$
$$=P\left(X\le t\mid \overline{A},B\right)\frac{t}{10}+P\left(X\le t\mid \overline{A},\overline{B}\right)\frac{10-t}{10}=1\cdot\frac{t}{10}+0\cdot\frac{10-t}{10}=\frac{t}{10}$$

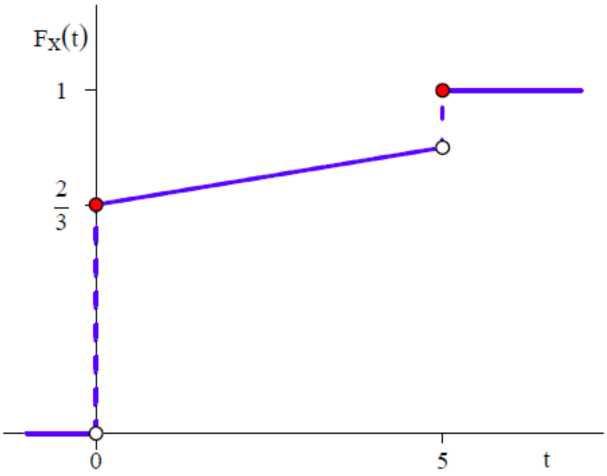
לכן,

$$F_X(t)=\frac{2}{3}+P\left(X\le t\mid \overline{A}\right)\cdot\frac{1}{3}=\frac{2}{3}+\frac{t}{30}=\frac{t+20}{30}$$

ו-

$$F_X(t)=\begin{cases}0 & t<0\\ \dfrac{2}{3} & 0=t\\ \dfrac{t+20}{30} & 0\le t\le 5\\ 1 & 5\le t\end{cases}$$

נשים לב שגרף הפונקציה שקיבלנו אינו רציף (!) וזאת מאחר וישנם אטומים בהתפלגות זאת, ז"א שהיא שילוב של התפלגות בדידה ורציפה.



<u>תהליך פואסון</u>

כפי שראינו משתנה מקרי פואסוני כמ"מ בדיד, ניתן להסתכל על תהליך פואסוני בזמן רציף. תהליך זה מתאר מספר אירועים שקוראים עד זמן מסוים והוא מקיים תכונות דומות מאוד למה שראינו במקרה הבדיד.

תהליך פואסון מוגדר כרצף של משתנים מקריים $\{X_t\}$ כאשר $t \geq 0$ וכל X_t הוא משתנה מקרי הסופר את מספר האירועים שקרו עד זמן t.

תהליך זה מקיים מספר תכונות חשובות:

1. מספר האירועים שנצפו במקטעי זמן זרים הם בלתי תלויים. בכתיבה מתמטית המשמעות של דבר זה היא שהמשתנים המקריים $X_t, X_s - X_t \ \forall s > t$ הם בלתי תלויים.
2. מספר האירועים שהתרחשו בקטע זמן באורך l תלוי באורך קטע הזמן ומתפלג $X_{t+l} - X_t \sim X_l \sim P(\lambda \cdot l)$ כאשר λ זה הפרמטר של התהליך.

ההתפלגות מבחינת נוסחה זהה למה שראינו עבור מ"מ בדיד וזה אומר ש-

$$P(X_{t+s} - X_t = k) = P(X_s = k) = e^{-\lambda s}\frac{(\lambda s)^k}{k!}$$

ההנחה הבסיסית קובעת כי $X_0 = 0$.

<u>משתנה מקרי מעריכי</u>

משתנה מקרי מעריכי הוא משתנה מקרי הקשור בצורה ישירה לתהליך פואסון. משתנה מקרי מעריכי עם פרמטר λ מודד את הזמן בין האירועים המתרחשים לפי תהליך פואסון עם

קצב λ. למעשה מה שניתן להראות זה שאם יש לנו תהליך בו הזמן בין אירועים מתפלג לפי משתנה מקרי מעריכי עם פרמטר λ אז תהליך זה הוא תהליך פואסון.

משתנה מקרי מעריכי X מסומן ע"י $X \sim Exp(\lambda)$ והפונקציות המתארות אותו הן:

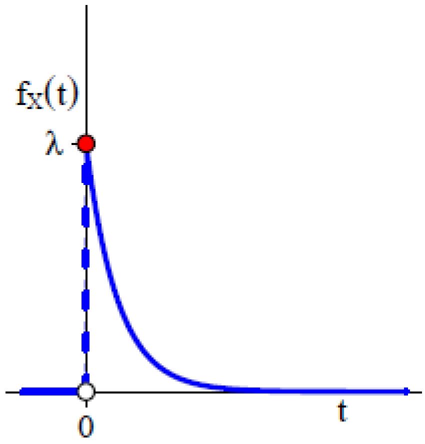
$$f_X(t) = \begin{cases} \lambda e^{-\lambda t} & 0 \le t \\ 0 & t < 0 \end{cases}$$

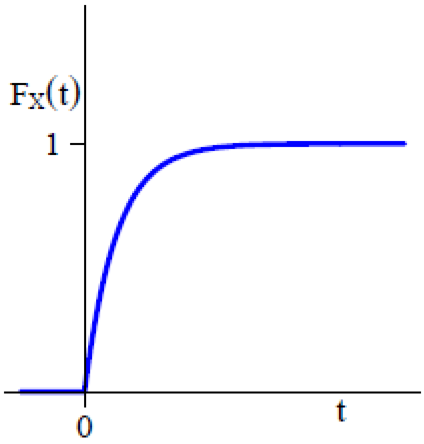
או לחילופין,

$$F_X(t) = \begin{cases} 1 - e^{-\lambda t} & 0 \le t \\ 0 & t < 0 \end{cases}$$

תכונה מאוד חשובה של משתנה מקרי מעריכי זו תכונת חוסר הזיכרון! תכונה זו קובעת כי עבור כל משתנה מקרי מעריכי $X \sim Exp(\lambda)$ מתקיים:

$$P(X > t + s \mid X > t) = P(X > s)$$





<u>תרגיל 5</u>: מספר התקלות, X_t, ברשת תקשורת באינטרוול זמן מ-0 עד t הוא תהליך פואסוני $\{X_t, t \ge 0\}$ עם קצב ממוצע של $\lambda = \frac{1}{4}$ תקלות לשעה.

א. מה ההסתברות לתקלה אחת לכל היותר ב-4 שעות הראשונות ו-2 תקלות לכל הפחות ב-8 שעות שלאחר מכן?

ב. מאז התקלה האחרונה עבור 3 שעות. מה ההסתברות שנשלים 5 שעות רצופות ללא תקלה?

ג. ב-12 השעות הראשונות התרחשה תקלה אחת. מה ההסתברות שהיא התרחשה בין השעה החמישית לתשיעית?

ד. ב-20 השעות הראשונות התרחשו 10 תקלות. מה ההסתברות ש-2 מתוכן התרחשו בין השעה העשירית ל-12?

<u>פתרון תרגיל 5</u>:

א. תחילה נרשום את מה שדורשים מאיתנו למצוא בשפה מתמטית.

$$P\left(X_4 \le 1, X_{12} - X_4 \ge 2\right) = P\left(X_4 \le 1\right)P\left(X_{12} - X_4 \ge 2\right) =$$
$$=\left[P\left(X_4 = 0\right) + P\left(X_4 = 1\right)\right]\left[1 - P\left(X_{12} - X_4 = 0\right) - P\left(X_{12} - X_4 = 1\right)\right] =$$
$$=$$
$$= e^{-\frac{1}{4}\cdot 4}\left(\frac{\left(0.25\cdot 4\right)^0}{0!} + \frac{\left(0.25\cdot 4\right)^1}{1!}\right)\left[1 - e^{-\frac{1}{4}\cdot 8}\left(\frac{\left(0.25\cdot 8\right)^0}{0!} + \frac{\left(0.25\cdot 8\right)^1}{1!}\right)\right] = 0.437$$

שימו לב שיכולנו לבצע את המעבר בשורה הראשונה בדיוק בגלל תכונת האי-תלות של משתנה מקרי פואסוני. בגלל שמספר האירועים שקרו ב-4 שעות הראשונות וב-8 שעות שלאחר מכן הם בלתי תלויים אז ניתן להפריד את ההסתברות של חיתוך של המאורעות למכפלת הסתברויות.

ב. נזכר בכך שהזמן בין המופעים הוא משתנה מקרי מעריכי עבור תהליך פואסון, ולכן אם נסמן ב-Y את משך הזמן שבין התקלה האחרונה לתקלה הבאה אחריה נקבל ש- $Y \sim Exp\left(\frac{1}{4}\right)$. נרשום את הבעיה בסימונים מתמטיים באופן הבא:

$$P\left(Y \ge 5 \mid Y \ge 3\right) = P\left(Y \ge 2\right) = 1 - F_Y\left(2\right) = 1 - \left(1 - e^{-0.25\cdot 2}\right) = 0.607$$

ג. נתון כי $X_{12} = 1$ ורוצים לדעת מה ההסתברות ש- $5 \le Y \le 9$ או לחילופין ש- $X_9 - X_5 = 1$.

$$P\left(9 \ge Y \ge 5 \mid X_{12} = 1\right) = P\left(X_9 - X_5 = 1 \mid X_{12} = 1\right) =$$
$$= \frac{P\left(X_5 = 0, X_9 - X_5 = 1, X_{12} - X_9 = 0\right)}{P\left(X_{12} = 1\right)} =$$
$$\frac{P\left(X_5 = 0\right)P\left(X_9 - X_5 = 1\right)P\left(X_{12} - X_9 = 0\right)}{P\left(X_{12} = 1\right)}$$

ולכן,

$$P\left(9\ge Y\ge 5\mid X_{12}=1\right)=\frac{e^{-0.25\cdot 5}\frac{\left(0.25\cdot 5\right)^{0}}{0!}\cdot e^{-0.25\cdot 4}\frac{\left(0.25\cdot 4\right)^{1}}{1!}\cdot e^{-0.25\cdot 3}\frac{\left(0.25\cdot 3\right)^{0}}{0!}}{e^{-0.25\cdot 12}\frac{\left(0.25\cdot 12\right)^{1}}{1!}}=\frac{1}{3}$$

ובאופן כללי, כאשר יש לנו תהליך פואסון ונתון שהתרחש עד זמן T מופע יחיד אזי הזמן עד למאורע הראשון מתפלג אחיד בקטע 0 עד T ולכן ההסתברות שהאירוע יתרחש דווקא בקטע $\left[x,y\right]\subseteq\left[0,t\right)$ היא $\frac{y-x}{t}$.

ד. מתכונת האי תלות אנחנו יודעים שישנה הסתברות של $\frac{2}{20}=0.1$ לכל אחת מהתקלות שקרו עד זמן 20 לקרות בקטע זמן $\left[10,12\right]$. לכן בהינתן שהיו 20 תקלות עד שעה 20, $X_{20}=10$, אז מספר התקלות שהתרחשו בקטע זמן $\left[10,12\right]$ הוא מ"מ בינומי $X_{12}-X_{10}\mid X_{20}=10\sim Bin\left(10,0.1\right)$.

$$P\left(X_{12}-X_{10}=2\mid X_{20}=10\right)=\binom{10}{2}\cdot 0.1^{2}\cdot 0.9^{8}$$

<u>תרגיל 6</u>: מספר הפונים לנציב תלונות הציבור ברעננה מתנהג בקירוב כמו תהליך פואסון עם ממוצע של 2 פניות ביום.

א. מה התפלגות מספר הימים בשבוע בודד בהם הייתה לפחות פניה אחת?

ב. מה ההסתברות שהן ביום ב' והן ביום ג' יהיו 3 פניות בכ"א בהינתן שביתר השבוע היו 5 פניות סה"כ?

<u>פתרון תרגיל 6</u>:

א. בכלל תכונת האי-תלות של תהליך פואסון כל יום הוא בלתי תלוי בימים הקודמים ובימים הבאים ולכן צריך להסתכל על כל יום בנפרד בתור ניסוי. ההסתברות להצלחה בניסוי (קרי שהייתה לפחות פנייה אחת באותו היום) היא

$$P\left(X_i\ge 1\right)=1-P\left(X_i=0\right)=1-e^{-2}\frac{2^{0}}{0!}$$

מאחר ומספר הפניות ביום מסוים, X_i , מתפלג פואסונית עם פרמטר 2. מכאן נסיק כי התפלגות מספר הימים בשבוע בודד בהם הייתה לפחות פנייה אחת מתפלגת בינומית, $Bin\left(7,1-e^{-2}\right)$.

ב. למעשה מה ששואלים אותנו זה,

$$P\left(X_2-X_1=3, X_3-X_2=3 \mid X_1+X_7-X_3=5\right)=?$$

ובכן זו למעשה שאלה הרבה יותר פשוטה ממה שניתן היה לחשוב מאחר וקטעי הזמן של יום ב' ו-ג' אל מול יתר השבוע הם בלתי תלויים ולכן הביטוי לעיל שווה ל-

$$P\left(X_2-X_1=3, X_3-X_2=3\right)$$

בגלל האי תלות. ולכן,

$$\begin{aligned}&P\left(X_2-X_1=3, X_3-X_2=3 \mid X_1+X_7-X_3=5\right)=\\&=P\left(X_2-X_1=3, X_3-X_2=3\right)=P\left(X_2-X_1=3\right)P\left(X_3-X_2=3\right)\end{aligned}$$

וכל אחד מהביטויים בצד ימין הוא מ"מ פואסוני עם פרמטר 2. נשאיר לכם להשלים את החסר... ☺

<u>תוחלת של משתנה מקרי רציף</u>

תוחלת של משתנה מקרי רציף לא שונה בהרבה מתוחלת של משתנה מקרי בדיד, אלא שבמקרה של מ"מ רציף מבצעים אינטגרציה במקום סכימה.

יהיה X מ"מ רציף עם פונקצית צפיפות $f_X(t)$, אזי התוחלת של X היא

$$E(x)=\int_{-\infty}^{\infty} x f_X(x)\,dx$$

ועבור פונקציה "טובה" כלשהי (רציפה למקוטעין וכו') $f_X(t)$ מתקיים כי

$$E\left(g(x)\right)=\int_{-\infty}^{\infty} g(x)\, f_X(x)\,dx$$

תכונות הלינאריות נשמרות כמו שראינו במקרה הבדיד וכן השונות מוגדרת באותו האופן.

<u>תרגיל 7</u>: יהי X מ"מ בעל הצפיפות

$$f_X(x)=\begin{cases}0 & x<1\\ \dfrac{3}{x^4} & x\geq 1\end{cases}$$

חשבו את התוחלת והשונות של X.

<u>פתרון תרגיל 7</u>: נתחיל בחישוב התוחלת

$$E(x)=\int_{-\infty}^{\infty} xf_X(x)\,dx=\int_{1}^{\infty} x\frac{3}{x^4}dx=\int_{1}^{\infty}\frac{3}{x^3}dx=$$

$$=\left.\frac{3x^{-2}}{-2}\right|_1^\infty=\frac{3}{2}$$

כעת נמצא את השונות בשלבים.

$$E\left(x^2\right)=\int_{-\infty}^{\infty} x^2 f_X(x)\,dx=\int_{1}^{\infty} x^2\frac{3}{x^4}dx=\int_{1}^{\infty}\frac{3}{x^2}dx=$$

$$=\left.\frac{3x^{-1}}{-1}\right|_1^\infty=3$$

ולכן נקבל כי השונות שווה ל-

$$V(X)=E\left(x^2\right)-E(x)^2=3-\left(\frac{3}{2}\right)^2=\frac{3}{4}$$

<u>התפלגות נורמאלית</u>

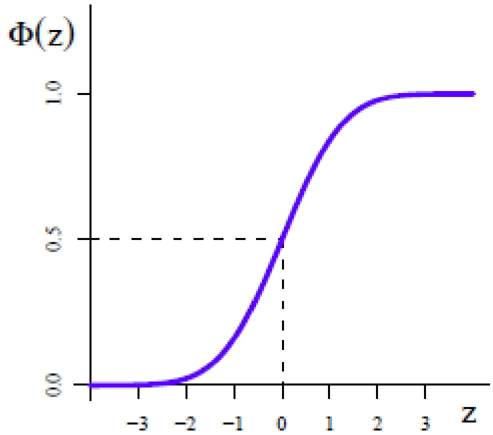
משתנה מקרי X המתפלג נורמאלית מסומן ב- $X \sim N\left(\mu,\sigma^2\right)$ כאשר $\mu=E(X)$ היא התוחלת של X ו-$\sigma^2=V(X)$ היא השונות של X. מכאן אנו מסיקים כי בכדי להגדיר משתנה מקרי נורמלי בצורה מלאה כל מה שצריך לדעת זה את השונות והתוחלת של המשתנה המקרי המדובר.

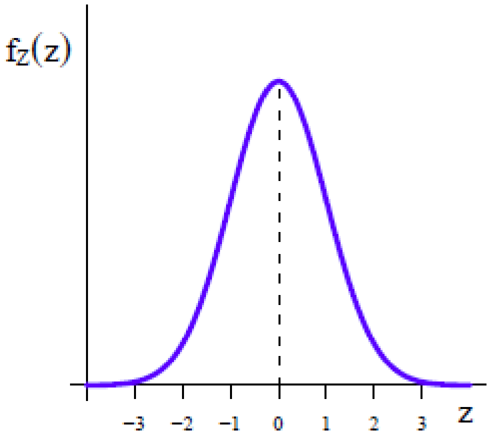
פונקצית הצפיפות של משתנה מקרי נורמאלי היא –

$$f_X(t)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(t-\mu)^2}{2\sigma^2}}$$

והיא מוגדרת על כל הישר הממשי.

במקרים רבים נדון במשתנה מקרי נורמלי סטנדרטי כאשר הכוונה במקרה זה היא ש- $X \sim N(0,1)$. כאשר משרטטים את פונקצית הצפיפות של משתנה מקרי מקבלים צורת פעמון משתנה בהתאם לפרמטרים של המשתנה המקרי.





מאחר ולא ניתן לפתור מפורשות אנליטית את האינטגרל לקבלת פונקצית ההתפלגות המצטברת של משתנה מקרי נורמאלי סטנדרטי, $\Phi(t)$, אנו נאלצים לבצע חישוב נומרי בכדי למצוא את ההסתברות המצטברת. מה שכן ידוע הוא, שבשל תכונת הסימטריה של מ"מ נורמאלי סטנדרטי, אז

$$\Phi(-t)=1-\Phi(t)$$

סטיית התקן של מ"מ נורמאלי סטנדרטי – סטיית התקן באופן כללי מוגדרת להיות שורש השונות ומסומנת על ידי, $\sigma_X=\sqrt{Var(X)}$.

<u>טענות</u>:

1. אם $X \sim N(\mu,\sigma^2)$ אז $\frac{X-\mu}{\sigma} \sim N(0,1)$. זאת תכונה מאוד חשובה בכדי לפשט חישובים בגלל שכל מה שנצטרך זה לדעת כיצד מתפלג מ"מ נורמאלי סטנדרטי ואז נוכל לחשב הסתברויות עבור כל מ"מ נורמאלי אחר.
2. אם $X \sim N(\mu,\sigma^2)$ אז האחוזון ה- $100p$ של X, x_p , שווה ל- $x_p=\mu+\sigma\cdot z_p$, כאשר z_p זה ה- $100p$ של התפלגות נורמאלית סטנדרטית.
3. עבור התפלגות נורמאלית סטנדרטית - $z_p=-z_{1-p}$.

<u>תרגיל 8</u>: אלון ובן מתרגלים קורס הסתברות-אלגברית. מניסיון העבר, ציוני התלמידים בקורס זה מתפלגים נורמאלית עם ממוצע 68 וסטיית תקן 10.

א. מה ההסתברות שסטודנט יקבל בין 78 ל-58? בין 88 ל-48? בין 38 ל-98?
ב. מה אחוז התלמידים הנכשלים (ציון נמוך מ-60) בקורס זה?מהו אחוז הסטודנטים המצטיינים (בעלי ציון מ-95 ומעלה)?

ג. מצאו את האחוזון ה-95 וה-10 של התפלגות הציונים.

ד. ראשי הפקולטה רוצים כי בשנה זו כ- 10% מהתלמידים ייכשלו בבחינה ושלא יותר מ-5% יקבלו ציון מעולה. לשם כך על המתרגלים לשנות את קנה המידה של הציון. מהי נוסחת השינוי שבה ישתמשו? מה יהיו הממוצע וסטיות התקן החדשות?

<u>פתרון תרגיל 8</u>:

א. נפתור בעיה אפילו יותר כללית בשאלה זו. נניח כי $X \sim N\left(\mu,\sigma^2\right)$ ונראה מה ההסתברות שעבור מ"מ נורמאלי כללי הציונים יחרגו במידות מסוימות מסטיית התקן.

הסעיף הראשון שואל מה ההסתברות ש-X יהיה במרחק של סטיית אחת מהממוצע / תוחלת.

$$P\left(\left|X-\mu\right|\le\sigma\right)=P\left(\sigma\le X-\mu\le\sigma\right)=P\left(-1\le\frac{X-\mu}{\sigma}\le1\right)=P\left(-1\le Z\le1\right)$$

כאשר Z הוא מ"מ נורמאלי סטנדרטי. נחשב את הביטוי מימין.

$$P\left(-1\le Z\le1\right)=P\left(Z\le1\right)-P\left(Z\le-1\right)=\Phi\left(1\right)-\Phi\left(-1\right)=$$
$$=\Phi\left(1\right)-\left(1-\Phi\left(1\right)\right)=2\Phi\left(1\right)-1=0.6826$$

ולכן,

$$P\left(\left|X-\mu\right|\le2\sigma\right)=P\left(-2\le Z\le2\right)=2\Phi\left(2\right)-1=0.9544$$
$$P\left(\left|X-\mu\right|\le3\sigma\right)=P\left(-3\le Z\le3\right)=2\Phi\left(3\right)-1=0.9974$$

ב. אחוז הסטודנטים שנכשלים שקול לסיכוי שסטודנט ספציפי ייכשל ולכן עבור $X \sim N\left(68,100\right)$ נקבל,

$$P\left(X<60\right)=P\left(X-68<-8\right)=P\left(\frac{X-68}{10}<-0.8\right)=$$
$$=P\left(Z<-0.8\right)=\Phi\left(-0.8\right)=1-\Phi\left(0.8\right)=0.2119$$

ולכן יהיו בערך 21% נכשלים. נבצע חישוב דומה עבור המצטיינים,

$$P\left(X\ge95\right)=P\left(\frac{X-68}{10}\ge2.7\right)=P\left(Z\ge2.7\right)=$$
$$=1-P\left(Z<2.7\right)=1-\Phi\left(2.7\right)=0.0035$$

זאת אומרת, שיהיה בערך שליש אחוז מצטיינים (מה שנקרא קורס לא קל... ☺).

ג. כעת אנחנו צריכים לבצע מספר חישובים מורכבים. נגדיר את x_t להיות האחוזון ה-t של משתנה מקרי X, ז"א ש- $P(X \le x_t) = \frac{t}{100}$. נתחיל במציאת האחוזונים המתאימים שדורשים עבור X, x_{10}, x_{95}.

$$P(X \le x_{10}) = 0.1 \Leftrightarrow P\left(\frac{X-68}{10} \le \frac{x_{10}-68}{10}\right) = 0.1 \Leftrightarrow P\left(Z \le \frac{x_{10}-68}{10}\right) = 0.1$$

$$\Leftrightarrow \Phi\left(\frac{x_{10}-68}{10}\right) = 0.1 \Leftrightarrow 1-\Phi\left(\frac{68-x_{10}}{10}\right) = 0.1 \Leftrightarrow \Phi\left(\frac{68-x_{10}}{10}\right) = 0.9$$

$$\Leftrightarrow \frac{68-x_{10}}{10} = 1.282 \Leftrightarrow x_{10} = 55.18$$

ועבור האחוזון ה-95 נקבל,

$$P(X \le x_{95}) = 0.95 \Leftrightarrow P\left(\frac{X-68}{10} \le \frac{x_{95}-68}{10}\right) = 0.95 \Leftrightarrow P\left(Z \le \frac{x_{95}-68}{10}\right) = 0.95$$

$$\Leftrightarrow \Phi\left(\frac{x_{95}-68}{10}\right) = 0.95 \Leftrightarrow \frac{x_{95}-68}{10} = 1.645 \Leftrightarrow x_{95} = 84.45$$

ד. בדרישות השאלה רוצים שנמצא התפלגות ציונים חדשה Y כך ש- $y_{10} = 60, y_{95} = 95$, ז"א ש- $P(Y < 60) = \frac{10}{100} = 0.1$ ו- $P(Y < 95) = \frac{95}{100} = 0.95$.

אנחנו רוצים לבצע טרנספורמציה ליניארית ל-X כך ש- $Y = aX + b$ והאחוזונים של Y יקיימו את נתוני השאלה (ע"י טרנספורמציה לינארית שומרים על ההתפלגות נורמאלית). לשמחתנו, בגלל ש-Y היא טרנספורמציה ליניארית אזי גם האחוזונים מקיימים את אותה הטרנספורמציה ולכן,

$$y_{95} = ax_{95} + b = 95$$
$$y_{10} = ax_{10} + b = 60$$

בגלל ש-

$$P(X \le x_{10}) = 0.1 \Leftrightarrow P(aX + b \le ax_{10} + b) = 0.1$$
$$\Leftrightarrow P(Y \le ax_{10} + b) = 0.1 \Leftrightarrow y_{10} = ax_{10} + b$$

וכל שנותר לנו הוא לפתור 2 משוואות ב-2 נעלמים ולקבל - $a = 1.2, b = -6.216$

ולכן נקבל ש-

$$Y = 1.2X - 6.216 \Rightarrow Y \sim N\left(1.2 \cdot 68 - 6.216, (1.2 \cdot 10)^2\right) = N(75.4, 144)$$

<u>תרגיל 9</u>: יהא X מ"מ נורמאלי עם תוחלת 1 וסטיית תקן 2, ז"א $X \sim N(1,4)$.

א. מה הסיכוי ש- Xיהיה בין 0 ל-4?

ב. מהו האחוזון ה-25 של X?

<u>פתרון תרגיל 9</u>:

א. נפתור בעזרת חישוב ישיר.

$$P(0 \le X \le 4) = P\left(\frac{0-1}{2} \le \frac{X-1}{2} \le \frac{4-1}{2}\right) = P\left(-\frac{1}{2} \le Z \le \frac{3}{2}\right) =$$
$$= \Phi\left(\frac{3}{2}\right) - \Phi\left(-\frac{1}{2}\right) = \Phi\left(\frac{3}{2}\right) - \left(1 - \Phi\left(\frac{1}{2}\right)\right) =$$
$$= 0.6247$$

ב. נראה משהו כללי יותר. נניח $X \sim N(\mu, \sigma^2)$ ונמצא את האחוזון ה-100p של התפלגות זו.

$$p = P(X \le X_{100p}) = P\left(\frac{X-\mu}{\sigma} \le \frac{X_{100p}-\mu}{\sigma}\right) =$$
$$= P\left(Z \le \frac{X_{100p}-\mu}{\sigma}\right) \Rightarrow \frac{X_{100p}-\mu}{\sigma} = Z_{100p}$$
$$\Rightarrow X_{100p} = \mu + \sigma Z_{100p}$$
$$\Rightarrow X_{25} = 1 + 2Z_{25} = 1 + 2(1 - Z_{75}) = 1 - 2 \cdot 0.675 = -0.35$$

שימו לב שאנחנו הוכחנו את הנוסחה הכללית לאחוזונים של התפלגות נורמאלית

$$X_{100p} = \mu + \sigma Z_{100p}$$

וכל מה שנצטרך מעתה זה לדעת את האחוזונים של התפלגות נורמאלית סטנדרטית מהטבלה.

<u>קירוב נורמאלי למשתנה מקרי בינומי</u>

ראינו עד כה קירוב של משתנה מקרי בינומי על ידי מ"מ פאוסוני, כעת נראה שישנם מצבים בהם ניתן לקרב בינומי על ידי נורמאלי.

נניח כי $X \sim Bin(n,p)$, אזי עבור n מספיק גדול מתקיים כי:

$$P(X \le k) = \Phi\left(\frac{k+0.5-np}{\sqrt{np(1-p)}}\right)$$

נשים לב לפקטור 0.5 שנמצא בפונקצית ההתפלגות המצטברת. הפקטור 0.5 נכנס בשם "תיקון רציפות" והוא מתקבל בכך שמבצעים קירוב של משתנה מקרי בדיד על ידי משתנה מקרי רציף. למעשה משפט דה-מואבר-לפלס אומר ש-

$$\lim_{n\to\infty} P\left(\frac{X-np}{\sqrt{np(1-p)}} \le t\right) = \Phi(t)$$

אולם מה שעושים זה לרשום את המאורע $\{X \le k\}$ כ- $\left\{X \le k+\frac{1}{2}\right\}$ ואז משתמשים בקירוב הנ"ל, באופן הבא:

$$P(X \le k) = P\left(X \le k+\frac{1}{2}\right) = P\left(\frac{X-np}{\sqrt{np(1-p)}} \le \frac{k+0.5-np}{\sqrt{np(1-p)}}\right) \cong \Phi\left(\frac{k+0.5-np}{\sqrt{np(1-p)}}\right)$$

<u>תרגיל 10</u>: המשטרה פתחה במבצע לבדיקת תקינות כל רכב לחורף. ניסיון העבר מלמד כי-40% מהרכבים הנוסעים בארץ לא תקינים.

א. מה ההסתברות בקירוב שמתוך 80 מכוניות הנבדקות ביום אחד יותר מחציין תימצאנה בלתי תקינות?

ב. מה ההסתברות שבמשך שבוע עבודה (6 ימים), לפחות ביום אחד תמצאנה יותר ממחצית מהמכוניות הנבדקות בלתי תקינות?

<u>פתרון תרגיל 10</u>:

א. נתחיל בכך שנזהה את התפלגות מספר המכוניות הבלתי תקינות מתוך ה-80 שנבדקו. נסמן את מספר המכוניות הבלתי תקינות ב-X ונקבל כי $X \sim Bin(80,0.4)$. כעת, מאחר ו-n מספיק גדול וההסתברות לא קטנה מידי, נוכל לבצע קירוב נורמאלי ולקבל

$$P(X > 40) = 1 - P(X \le 40) \cong 1 - \Phi\left(\frac{40+0.5-80\cdot 0.4}{\sqrt{80\cdot 0.4(1-0.4)}}\right) = 1-\Phi(1.94) = 0.0262$$

ב. נעזר בסעיף א'. נגדיר יום מוצלח בתור יום בו יותר מחצי מהמכוניות נמצאו לא תקינות. נגדיר את Y להיות מספר הימים המוצלחים. בשל הנחת אי תלות בין הימים נובע ש- $Y \sim Bin(6, 0.0262)$ ולכן,

$$P(Y \geq 1) = 1 - P(Y = 0) = 1 - \binom{6}{0} 0.0262^0 (1 - 0.0262)^6 = 0.147.$$

טבלת התפלגות נורמאלית

להלן טבלת התפלגות נורמאלית למציאת ההסתברויות בתרגיל.

0.09	0.08	0.07	0.06	0.05	0.04	0.03	0.02	0.01	0.00	
0.5359	0.5319	0.5279	0.5239	0.5199	0.5160	0.5120	0.5080	0.5040	0.5000	**0**
0.5753	0.5714	0.5675	0.5636	0.5596	0.5557	0.5517	0.5478	0.5438	0.5398	**0.1**
0.6141	0.6103	0.6064	0.6026	0.5987	0.5948	0.5910	0.5871	0.5832	0.5793	**0.2**
0.6517	0.6480	0.6443	0.6406	0.6368	0.6331	0.6293	0.6255	0.6217	0.6179	**0.3**
0.6879	0.6103	0.6808	0.6772	0.6736	0.6700	0.6664	0.6628	0.6591	0.6554	**0.4**
0.7224	0.7190	0.7157	0.7123	0.7088	0.7054	0.7019	0.6985	0.6950	0.6915	**0.5**
0.7549	0.7517	0.7486	0.7454	0.7422	0.7389	0.7357	0.7324	0.7291	0.7257	**0.6**
0.7852	0.7823	0.7794	0.7764	0.7734	0.7704	0.7673	0.7642	0.7611	0.7580	**0.7**
0.8133	0.8106	0.8078	0.8051	0.8023	0.7995	0.7967	0.7939	0.7910	0.7881	**0.8**
0.8389	0.8365	0.8340	0.8315	0.8289	0.8264	0.8238	0.8212	0.8186	0.8159	**0.9**
0.8621	0.8599	0.8577	0.8554	0.8531	0.8508	0.8485	0.8461	0.8438	0.8413	**1**
0.8830	0.8810	0.8790	0.8770	0.8749	0.8729	0.8708	0.8686	0.8665	0.8643	**1.1**
0.9015	0.8997	0.8980	0.8962	0.8944	0.8925	0.8907	0.8888	0.8869	0.8849	**1.2**
0.9177	0.9162	0.9147	0.9131	0.9115	0.9099	0.9082	0.9066	0.9049	0.9032	**1.3**
0.9319	0.9306	0.9292	0.9279	0.9265	0.9251	0.9236	0.9222	0.9207	0.9192	**1.4**
0.9441	0.9429	0.9418	0.9406	0.9394	0.9382	0.9370	0.9357	0.9345	0.9332	**1.5**
0.9545	0.9535	0.9525	0.9515	0.9505	0.9495	0.9484	0.9474	0.9463	0.9452	**1.6**
0.9633	0.9625	0.9616	0.9608	0.9599	0.9591	0.9582	0.9573	0.9564	0.9554	**1.7**
0.9706	0.9699	0.9693	0.9686	0.9678	0.9671	0.9664	0.9656	0.9649	0.9641	**1.8**
0.9767	0.9761	0.9756	0.9750	0.9744	0.9738	0.9732	0.9726	0.9719	0.9713	**1.9**
0.9817	0.9812	0.9808	0.9803	0.9798	0.9793	0.9788	0.9783	0.9778	0.9772	**2**
0.9857	0.9854	0.9850	0.9846	0.9842	0.9838	0.9834	0.9830	0.9826	0.9821	**2.1**
0.9890	0.9887	0.9884	0.9881	0.9878	0.9875	0.9871	0.9868	0.9864	0.9861	**2.2**
0.9916	0.9913	0.9911	0.9909	0.9906	0.9904	0.9901	0.9898	0.9896	0.9893	**2.3**
0.9936	0.9934	0.9932	0.9931	0.9929	0.9927	0.9925	0.9922	0.9920	0.9918	**2.4**
0.9952	0.9951	0.9949	0.9948	0.9946	0.9945	0.9943	0.9941	0.9940	0.9938	**2.5**
0.9964	0.9963	0.9962	0.9961	0.9960	0.9959	0.9957	0.9956	0.9955	0.9953	**2.6**
0.9974	0.9973	0.9972	0.9971	0.9970	0.9969	0.9968	0.9967	0.9966	0.9965	**2.7**
0.9981	0.9980	0.9979	0.9979	0.9978	0.9977	0.9977	0.9976	0.9975	0.9974	**2.8**
0.9986	0.9986	0.9985	0.9985	0.9984	0.9984	0.9983	0.9982	0.9982	0.9981	**2.9**
0.9990	0.9990	0.9989	0.9989	0.9989	0.9988	0.9988	0.9987	0.9987	0.9987	**3**
0.9992	0.9992	0.9992	0.9992	0.9991	0.9991	0.9991	0.9990	0.9990	0.9990	**3.1**
0.9994	0.9994	0.9994	0.9994	0.9994	0.9994	0.9993	0.9993	0.9993	0.9993	**3.2**
0.9996	0.9996	0.9996	0.9996	0.9995	0.9995	0.9995	0.9995	0.9995	0.9995	**3.3**

פרק 7

נושאים:

- התפלגויות דו-מימדיות.
- אי תלות בהתפלגויות דו מימדיות.
- שונות משותפת ומקדם מתאם.

התפלגות דו מימדית

לאחר שמיצינו את העבודה במימד אחד אנחנו מתקדמים לעבר משתנים מקריים ב-2 מימדים. הגדרות המשתנים המקריים עבור 2 מימדים לא משתנה בצורה משמעותית ממה שראינו עד כה.

יהיו X ו-Y זוג משתנים מקריים בדידים המוגדרים על מרחב הסתברות כלשהו. נגדיר את פונקצית ההתפלגות המשותפת שלהם כ-

$$P_{XY}(x,y)=P(X=x,Y=y)$$

עבור כל הערכים האפשריים בטווח.

להלן רשימת הגדרות ומושגים עבור משתנים מקריים בדידים ב-2 מימדים:

- ההסתברות השולית של X היא - $P_X(x)=\sum_y P(X=x,Y=y)$ (באופן סימטרי עבור Y).
- ההסתברות המותנית של X בהינתן $Y=y$ - $P_{X|Y}(x|y)=\dfrac{P(X=x,Y=y)}{P_Y(y)}$.
- התוחלת המותנית של X בהינתן $Y=y$ - $E(X|Y=y)=\sum_x xP_{X|Y}(x|y)$.

להלן רשימת הגדרות ומושגים עבור משתנים מקריים רציפים ב-2 מימדים:

- עבור מאורע $A\subseteq\mathbb{R}^2$ נגיד ש- $P((X,Y)\in A)=\iint_A f_{XY}(x,y)dxdy$ כאשר $f_{XY}(x,y)$ היא פונקצית צפיפות משותפת ל-X ו-Y.
- פונקצית הצפיפות חייבת להיות אי-שלילית פי שראינו במימד 1 וכן $\iint_{\mathbb{R}^2} f_{XY}(x,y)dxdy=1$.
- הצפיפות השולית של X היא - $f_X(x)=\int_{-\infty}^{\infty} f_{XY}(x,y)dy$.

- הצפיפות המותנית של X בהינתן $Y=y$ - $f_{X|Y}\left(x|y\right)=\dfrac{f_{XY}\left(x,y\right)}{f_Y\left(y\right)}$.
- התוחלת המותנית של X בהינתן $Y=y$ - $E\left(X|Y=y\right)=\int_{-\infty}^{\infty}xf_{X|Y}\left(x|y\right)dx$.
- התוחלת המותנית של X בהינתן $Y=y$ היא למעשה פונקציה של y, ולכן ניתן לדבר גם על התוחלת הבאה - $E\left(X|Y\right)$ שזה בעצם משתנה מקרי שמקבל את הערך $E\left(X|Y=y\right)$ בהסתברות $P\left(Y=y\right)$.

אי תלות בין משתנים מקריים

זוג משתנים מקריים הם בלתי תלויים אם לכל זוג מאורעות $A,B\subseteq\mathbb{R}^2$ מתקיים

$$P\left(X\in A,Y\in B\right)=P\left(X\in A\right)P\left(Y\in B\right)$$

באותה מידה ניתן להגיד כי זוג משתנים מקרים X, Y בדידים הם בלתי תלויים אם

$$P_{XY}\left(x,y\right)=P_X\left(x\right)P_Y\left(y\right)$$

לכל זוג נקודות x,y וזאת בגלל ש-

$$\begin{aligned}P_{XY}\left(x,y\right)&\equiv P\left(X=x,Y=y\right)=P\left(X=x\right)P\left(Y=y\right)=\\&=\sum_y P\left(X=x,Y=y\right)\sum_x P\left(X=x,Y=y\right)\equiv P_X\left(x\right)P_Y\left(y\right)\end{aligned}$$

בצורה דומה נובע כי זוג משתנים מקריים רציפים הם בלתי תלויים אם

$$f_{XY}\left(x,y\right)=f_X\left(x\right)f_Y\left(y\right)$$

לכל זוג נקודות x,y.

שונות משותפת ומקדם מתאם

השונות המשותפת של זוג משתנים מקריים X ו-Y מוגדרת באופן הבא:

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

למעשה לשונות המשותפת ישנה חשיבות גדולה בבדיקת ההתאמה בין שני משתנים מקריים. אם השונות המשותפת חיובית אז זה אומר שיש קשר של עליה בין המשתנים המקריים (כשאחד עולה גם השני עולה). אם היא קטנה מאפס אז הקשר הוא יורד, ככל שאחד גדול יותר כך השני קטן יותר.

משתנים מקריים נקראים בלתי מתואמים אם –

$$Cov(X,Y)=0=E(XY)-E(X)E(Y)$$
$$\Rightarrow E(XY)=E(X)E(Y)$$

<u>תכונות</u>:

1. סימטריה - $Cov(X,Y)=Cov(Y,X)$, $Cov(X,X)=V(X)$, $V(X\pm Y)=V(X)+V(Y)\pm 2Cov(X,Y)$.
2. אדפטיביות - $Cov(X,Y+Z)=Cov(X,Y)+Cov(X,Z)$.
3. הומוגניות - $Cov(aX+b,cY+d)=acCov(X,Y)$.
4. בילינאריות - $Cov\left(\sum_{i=1}^{n}a_iX_i+b_i,\sum_{j=1}^{m}c_jY_j+d_j\right)=\sum_{i=1}^{n}\sum_{j=1}^{m}a_ic_jCov(X_i,Y_j)$
5. עבור משתנים מקריים כלשהם - $V\left(\sum_{i=1}^{n}X_i\right)=\sum_{i=1}^{n}V(X_i)+2\sum_{1\le i<j\le n}Cov(X_i,X_j)$ ועבור משתנים מקריים בלתי מתואמים למעשה שונות הסכום שווה לסכום השונויות.

הערה חשובה: שימו לב, משתנים מקריים בלתי תלויים הם בהכרח בלתי מתואמים, אך לא להיפך!

<u>מקדם מתאם</u> – ההגדרה למקדם מתאם היא:

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

ואפשר לראות כי מקדם המתאם נע בין 1- ל-1 (כאשר בקצוות המשמעות היא שיש קשר ליניארי בין המשתנים המקריים).

<u>תרגיל 1</u>: משטרת התנועה מציבה K מכמונות מהירות על כביש 6 כאשר $K\sim U[1,3]$. בן החליט לנסוע על כביש 6 ולא שם לב שהוא חרג מהמהירות המותרת לכל אורך הנסיעה. ישנה הסתברות של $1/3$ שבן יצולם על ידי כל מצלמה. יהי N מספר הדו"חות שיקבל בן.

א. כיצד מתפלג $N\mid K=k$?

ב. מצאו את ההתפלגות המשותפת של K ו-N.

ג. אם קנס על מהירות הוא 270 ש"ח כמה צפוי לשלם יאיר בתום הנסיעה?

<u>פתרון תרגיל 1</u>:

א. מאחר ונתון מספר המצלמות בכביש, אזי $N\mid K=k$ מתפלג בינומית כאשר ישנם k ניסויים וההסתברות ל-"הצלחה" (קבלת דו"ח) בכל ניסוי היא $1/3$.

ב. במקרה זה ישנו קשר מאוד ברור של סיבה ותוצאה בין מספר המצלמות לכמות הדו"חות ולכן נוכל להתנות על מספר המצלמות באופן הבא:

$$P(N=n,K=k)=P(N=n\,|\,K=k)\cdot P(K=k)=P(N=n\,|\,K=k)\cdot\frac{1}{3}=$$

$$=\binom{k}{n}\left(\frac{1}{3}\right)^{n}\left(\frac{2}{3}\right)^{k-n}\cdot\frac{1}{3}=\binom{k}{n}\left(\frac{1}{3}\right)^{n+1}\left(\frac{2}{3}\right)^{k-n}\qquad k=1,2,3\,;\,n=0,1,2,3$$

ניתן להציג תוצאות אלו בטבלה הבאה:

$N \setminus K$	1	2	3	$P(N=n)$
0	2/9	4/27	8/81	38/81
1	1/9	4/27	4/27	11/27
2	0	1/27	2/27	1/9
3	0	0	1/81	1/81
$P(K=k)$	1/3	1/3	1/3	

ג. מאחר ועל כל דו"ח הוא צפוי לשלם 270 ש"ח צריך למצוא את התוחלת של $270\cdot N$,

$$E(270\cdot N)=270E(N)=270\left[\sum_{n=0}^{3}nP(N=n)\right]=$$

$$=270\left[0\cdot\frac{38}{81}+1\cdot\frac{11}{27}+2\cdot\frac{1}{9}+3\cdot\frac{1}{81}\right]=180$$

<u>תרגיל 2</u>: בן מגיע לעשות קניות אצל הירקן בסוף היום ומגלה שנשארו רק 8 פירות: 3 אגסים (גדול, בינוני וקטן), 3 בננות (גדולה, בינונית וקטנה) ו-2 תפוחים (גדול ובינוני). בן ממהר ולכן קונה 2 פירות באקראי. יהי X מספר התפוחים שבן קנה ו-Y מספר הפירות הקטנים שקנה.

א. רשמו את טבלת ההתפלגות המשותפת של (X,Y) כולל ההתפלגויות השוליות. האם X ו-Y תלויים?

ב. חשבו את $E(XY)$ ואת התוחלת והשונות של X ו-Y.

ג. אגס עולה 5 שקל, תפוח עולה 4 שקל ובננה עולה 5 שקל. יש הנחה של 2.5 שקל על כל פרי קטן. מצאו את התוחלת של ההוצאה של רותי על הפירות.

פתרון תרגיל 2:

א. נבנה את הטבלה בשלבים. נתחיל בכך שנשים לב כי ישנן סה"כ $\binom{8}{2}=28$ אפשרויות במרחב המדגם.
נתחיל בתאים בטבלה שהם 0 - לא יכול להיות שהיו 2 פירות קטנים ו-2 תפוחים ובאופן כללי אם $Y=2$ אז בן בוודאות לא קנה תפוחים ולכן נסמן 0 בטבלה במקומות הרלוונטיים. ההסתברות שבן קנה 2 פירות קטנים (ולכן 0 תפוחים) היא בדיוק $1/28$ כי ישנה האפשרות היחידה שהוא אגס ובננה קטנים.

$Y \setminus X$	0	1	2	$P(Y=k)$
0	6/28	8/28	1/28	15/28
1	8/28	4/28	0	12/28
2	1/28	0	0	1/28
$P(X=k)$	15/28	12/28	1/28	

$Y=1, X=2$: הסתברות 0, כי לא יכול להיות שבן קנה פירות קטנים ותפוחים בו-זמנית.

$Y=0, X=0$: הסתברות $\binom{4}{2}/28$, כי בן צריך היה לקחת רק פירות לא קטנים וללא תפוחים.

$Y=0, X=1$: הסתברות $8/28$, בוחרים תפוח אחד ופרי אחד מיתר ה-4 שאינם תפוחים ואינם קטנים.

$Y=0, X=2$: הסתברות $1/28$, ישנה רק אפשרות אחת (תפוח גדול ובינוני).

$Y=1, X=0$: הסתברות $8/28$, בוחרים פרי אחד מתוך ה-2 הקטנים ובנוסף בוחרים אחד מ-4 הנותרים שאינם קטנים או תפוחים.

$Y=1, X=1$: הסתברות $4/28$, בוחרים תפוח 1 ופרי קטן 1 (שניהם מתוך 2).

בנוסף, X ו-Y בוודאי תלויים מאחר ואנחנו רואים שישנם אפסים בטבלה כאשר כל ההתפלגויות השוליות חיוביות ולכן ישנה תלות.

ב. נתחיל בכך שנשים לב כי כ"א מהמשתנים המקריים הוא היפרגאומטרי בפני עצמו כאשר הם גם מתפלגים אותו הדבר. $X \sim Y \sim HG(8,2,2)$ ולכן

$$E(X)=E(Y)=2\cdot\frac{2}{8}=\frac{1}{2}$$

$$V\left(X\right)=V\left(Y\right)=D\cdot\frac{S}{N}\left(1-\frac{S}{N}\right)\frac{N-D}{N-1}$$

$$V\left(X\right)=V\left(Y\right)=2\cdot\frac{2}{8}\left(1-\frac{2}{8}\right)\frac{8-2}{8-1}=\frac{9}{28}$$

כעת נוכל לחשב גם את $E\left(XY\right)$. החישוב הזה הינו יחסית קל ממש מתוך ההגדרה של תוחלת.

$$E\left(XY\right)=\sum_{k=0}^{4}kP\left(XY=k\right)=1\cdot\frac{4}{28}=\frac{1}{7}$$

ג. נסמן ב-Z את סכום הכסף שהוציא בן.

$$Z=4X+5\cdot\left(2-X\right)-2.5\cdot Y$$

למה? ובכן, על כל פרי שאינו תפוח, בן משלם 5 ש"ח ולכן הכפלנו זאת במספר הפירות שאינם תפוחים. בנוסף, שיקללנו את ההנחה על כל קנייה של פרי קטן.
התוחלת של Z היא:

$$E\left(Z\right)=E\left(4X+5\cdot\left(2-X\right)-2.5\cdot Y\right)=4E\left(X\right)+10-5E\left(X\right)-2.5E\left(Y\right)=$$

$$=-E\left(X\right)+10-2.5E\left(Y\right)=10-0.5-\frac{1}{2}\cdot 2.5=8.25$$

<u>תרגיל 3</u>: אלון נוסע לקזינו בלאס וגאס ומהמר על קוביות. הוא מהמר על קובייה המוטלת שוב ושוב. יהיו – X מספר ההטלות עד שמתקבל לראשונה 6 ו-Y מספר ההטלות עד שמתקבל לראשונה מספר זוגי קטן מ-6.

א. רשמו את ההסתברות המשותפת של $\left(X,Y\right)$.

ב. חשבו $P\left(X+Y=4\right)$.

ג. מהי התפלגות $W=\min\left(X,Y\right)$, חשבו את $E\left(W\right),V\left(W\right)$.

ד. מצאו כיצד מתפלג $Y\,|\,X=2$.

<u>פתרון תרגיל 3</u>:

א. צריך לנתח שני מקרים – כאשר X קטן מ-Y ולהיפך.

$$P\left(X=k,Y=l\right)=\begin{cases}\left(\frac{3}{6}\right)^{k-1}\cdot\frac{1}{6}\cdot\left(\frac{4}{6}\right)^{l-k-1}\cdot\frac{2}{6} & k<l\\ \left(\frac{3}{6}\right)^{l-1}\cdot\frac{2}{6}\cdot\left(\frac{5}{6}\right)^{k-l-1}\cdot\frac{1}{6} & k>l\end{cases}\quad when\left\{\begin{matrix}k\neq l\\ k,l=1,2,\ldots\end{matrix}\right\}$$

אם $k<l$ זה אומר שהיו k-1 הטלות בהם יצאו רק 1,3,5 ולאחר מכן בהטלה ה-k יצא 6, ואז עברו עוד l-k-1 הטלות בהם יכול היה לצאת הכול למעט 2 או 4 ולבסוף בהטלה ה-l ייצא 4. על כן, רשמנו את ההסתברות לכל האירועים הללו כמכפלה. המקרה השני מתקבל על ידי אותם שיקולים.

ב.

$$P\left(X+Y=4\right)=P\left(X=3,Y=1\right)+P\left(X=1,Y=3\right)=$$
$$=\left(\frac{3}{6}\right)^{0}\cdot\frac{1}{6}\cdot\left(\frac{4}{6}\right)^{3-1-1}\cdot\frac{2}{6}+\left(\frac{3}{6}\right)^{0}\cdot\frac{2}{6}\cdot\left(\frac{5}{6}\right)^{3-1-1}\cdot\frac{1}{6}=\frac{1}{12}$$

ג. בסעיף זה אנחנו אפילו לא חייבים לנתח את X ו- Y מאחר ו-W סופר את מספר הפעמים שיצא מספר זוגי ולכן $W\sim G\left(\frac{1}{2}\right)$. מכאן נקבל כי –

$$E\left(W\right)=\frac{1}{0.5}=2,\quad V\left(W\right)=\frac{1-0.5}{0.5^{2}}=2$$

ד. נמצא את הנדרש בשלבים: תחילה עבור $Y=1$, מאחר וזו התפלגות ספציפית בהינתן ש- $X=2$,

$$P\left(Y=1\mid X=2\right)=\frac{P\left(Y=1,X=2\right)}{P\left(X=2\right)}=\frac{\frac{2}{6}\cdot\frac{1}{6}}{\frac{5}{6}\cdot\frac{1}{6}}=\frac{2}{5}$$

כעת עבור $Y>2$:

$$P\left(Y=l\mid X=2\right)=\frac{P\left(Y=l,X=2\right)}{P\left(X=2\right)}=\frac{\left(\frac{3}{6}\right)^{2-1}\cdot\frac{1}{6}\cdot\left(\frac{4}{6}\right)^{l-3}\cdot\frac{2}{6}}{\frac{5}{6}\cdot\frac{1}{6}}=$$
$$=\frac{\frac{3}{6}\cdot\frac{1}{6}\cdot\left(\frac{4}{6}\right)^{l-3}\cdot\frac{2}{6}}{\frac{5}{6}\cdot\frac{1}{6}}=\frac{1}{5}\cdot\left(\frac{4}{6}\right)^{l-3}$$

<u>תרגיל 4</u>: בקערה יש n אגסים, m בננות ו-k תפוחים $\left(n,m,k\geq 10\right)$. מן הקערה מוציאים באקראי וללא החזרה 10 פירות. נגדיר:

X – מספר האגסים שהוצאו. Y- מספר הבננות שהוצאו ו-Z מספר התפוחים שהוצאו.

א. חשבו את $V\left(X+Y+Z\right)$.

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

ג. עבור $n=m=k=10$, חשבו $\rho(X,Y)$.

<u>פתרון תרגיל 4</u>:

א. מכיוון ש- $X+Y+Z=10$ וזאת משום שישנם רק אגסים בננות ותפוחים בקערה והוצאנו 10 פירות לכן הסכום הוא פשוט 10. לכן נובע ש-
$V(X+Y+Z)=V(10)=0$.

ב. נחשב את ההסתברות מפורשות

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

ג. נמצא את מקדם המתאם בעזרת השונות המשותפת. נתחיל בכך ש-
X, Y ו-Z הם משתנים מקריים היפרגאומטריים כך ש-
$X\sim HG(30,10,10), Y\sim HG(30,10,10), Z\sim HG(30,10,10)$ ולכן
$V(X)=V(Y)=V(Z)$.

בנוסף נוכל להסתכל על Z בתור $X+Y=10-Z$, ולכן

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

על כן נקבל ש-

$$V(X+Y)=V(X)+V(Y)+2Cov(X,Y)=2V(X)+2Cov(X,Y)$$
$$\Rightarrow Cov(X,Y)=\frac{1}{2}V(X+Y)-V(X)=\frac{1}{2}V(Z)-V(X)$$
$$\Rightarrow Cov(X,Y)=-\frac{1}{2}V(X)$$

וכעת מאוד קל לחשב את מקדם המתאם בגלל

$$\rho(X,Y)=\frac{Cov(X,Y)}{\sqrt{V(X)V(Y)}}=\frac{-\frac{1}{2}V(X)}{\sqrt{V(X)V(X)}}=\frac{-\frac{1}{2}V(X)}{V(X)}=-0.5$$

<u>תרגיל 5</u>: בידיכם 5 קלפים ממוספרים בין 1 ל-5 בסדר מקרי. נגדיר X – מספר הקלף העליון, Y מספר הקלף התחתון ו-W סכום הקלפים.

א. חשבו את ההתפלגות המשותפת וההתפלגויות השוליות של X ו-Y. האם הם תלויים.

ב. חשבו את $E(X+Y)$.

ג. חשבו את $\rho(X,Y)$.

ד. חשבו את $\rho(X,W)$.

פתרון תרגיל 5:

א. אנחנו צריכים למצוא את ההסתברות המשותפת של Y,X. נשים לב שישנה הסתברות 0 שהם יקבלו את אותו הערך ובנוסף ההסתברות של הקלף העליון להיות בעל ערך מסויים היא $\frac{1}{5}$ ואז לקלף האחרון ישנה הסתברות $\frac{1}{4}$ לקבל את אחד הערכים האחרים שנותרו ולכן

$$P(X=k,Y=l)=P(Y=l\,|\,X=k)P(X=k)=\begin{cases}\frac{1}{20} & k\neq l\\ 0 & k=l\end{cases}$$

למעשה אנחנו רואים מכאן שההסתברות השולית עבור כל ערך היא $\frac{1}{5}$ מאחר וסכימה על כל הערכים של המשתנה המקרי השני תיתן

$\frac{1}{20}+\frac{1}{20}+\frac{1}{20}+\frac{1}{20}+0=\frac{4}{20}=\frac{1}{5}$ ולכן $X,Y\sim U[1,5]$.

המשתנים המקריים הללו מן הסתם תלויים מאחר וההסתברות ששניהם יהיו 1 בו זמנית היא 0 אבל כל אחד בנפרד יכול להיות שווה ל-1 בהסתברות חיובית ממש.

ב. $E(X+Y)=E(X)+E(Y)=\frac{5+1}{2}+\frac{5+1}{2}=3$.

ג. נמצא את ההסתברות של - $P(XY=k)$ עבור כל k אפשרי.

$$P(XY=1)=0$$

$$P(XY=2)=P(X=2,Y=1)+P(X=1,Y=2)=\frac{2}{20}=\frac{1}{10}$$

$$P(XY=3)=P(X=3,Y=1)+P(X=1,Y=3)=\frac{2}{20}=\frac{1}{10}$$

$$P(XY=4)=P(X=4,Y=1)+P(X=1,Y=4)+P(X=2,Y=2)=\frac{2}{20}+0=\frac{1}{10}$$

$$P(XY=5)=P(X=5,Y=1)+P(X=1,Y=5)=\frac{2}{20}=\frac{1}{10}$$

$$P(XY=6)=P(X=3,Y=2)+P(X=2,Y=3)=\frac{2}{20}=\frac{1}{10}$$

ממשיכים באותו אופן עבור הערכים האפשריים של המכפלה ומקבלים ש-

$$P(XY=k)=\frac{1}{10}\,,\ k=2,3,4,5,6,8,10,12,15,20$$

ולכן,

$$E(XY)=\frac{1}{10}(2+3+4+5+6+8+10+12+15+20)=8.5$$
$$Cov(X,Y)=E(XY)-E(X)E(Y)=8.5-9=-0.5$$

נזכיר שהשונות של משתנה מקרי אחיד בדיד $Z\sim U[a,b]$ היא

$V(Z)=\frac{(b-a+1)^2-1}{12}$ נציב זאת עבור המשתנים שלנו ונקבל:

$$V(X)=V(Y)=\frac{(5-1+1)^2-1}{12}=2$$
$$\Rightarrow \rho(X,Y)=\frac{Cov(X,Y)}{\sqrt{V(X)V(Y)}}=\frac{-0.5}{\sqrt{2\cdot 2}}=-0.25$$

שימו לב כי תוצאה זו היא הגיונית מאחר ואם הקלף העליון גבוה אזי הסיכוי שהקלף התחתון יהיה גבוה יורדת ולכן המתאם בינהם שלילי.

ד. $\rho(X,W)=\rho(X,X+Y)$ משיקולי סימטריה ולכן נמצא את $\rho(X,X+Y)$.

$$Cov(X,X+Y)=Cov(X,X)+Cov(X,Y)$$
$$=V(X)+Cov(X,Y)=2-\frac{1}{2}=\frac{3}{2}$$

$$V(X+Y)=V(X)+V(Y)+2Cov(X,Y)=2+2+2\cdot\left(-\frac{1}{2}\right)=3$$

ולכן,

$$\rho(X,W)=\rho(X,X+Y)=\frac{Cov(X,X+Y)}{\sqrt{V(X)V(X+Y)}}=\frac{1.5}{\sqrt{2}\cdot\sqrt{3}}=\frac{3}{2\sqrt{6}}\approx 0.61$$

גם תוצאה זו היא הגיונית מאחר ואם אחד מהקלפים גבוה אזי ככל הנראה גם הסכום שלהם יהיה גבוה בהתאם ולכן המתאם חיובי.

<u>תרגיל 6</u>: יהיו X, Y משתנים מקריים רציפים שפונקצית הצפיפות המשותפת שלהם נתונה על ידי

$$f_{X,Y}(x,y)=cx(y-x)e^{-y},\quad 0\le x\le y<\infty$$

א. מהו c?

ב. מצאו את $f_{Y|X}(y\mid x),E(Y\mid X)$.

<u>פתרון תרגיל 6</u>:

א. תחילה נשים לב שבאמת הצפיפות מקיימת את הדרישות שיש לנו מפונקצית צפיפות. היא אי-שלילית מאחר ו-x תמיד גדול שווה ל-0 ו-y גדול מ-x, ובנוסף האינטגרל בכל התחום על הצפיפות צריך להיות שווה ל-1.

$$1=\iint_{\mathbb{R}^2} f_{X,Y}(x,y)\,dxdy=\int_0^\infty\int_x^\infty cx(y-x)e^{-y}dxdy=\int_0^\infty\int_0^y cx(y-x)e^{-y}dxdy$$

$$=\int_0^\infty\int_0^y cxye^{-y}-cx^2e^{-y}dxdy=\int_0^\infty\left[c\frac{x^2}{2}ye^{-y}-c\frac{x^3}{3}e^{-y}\right]_0^y dy=$$

$$=\int_0^\infty\frac{cy^3e^{-y}}{6}dy=\int_0^\infty\frac{cy^3e^{-y}}{6}dy=\frac{c}{6}\int_0^\infty y^3e^{-y}dy=...=c$$

כאשר השתמשנו בזהות, $\int_0^\infty y^n e^{-y}dy=n!$ (נובע מאינטגרציה בחלקים המון פעמים...). ולכן קיבלנו ש- $f_{X,Y}(x,y)=x(y-x)e^{-y},\quad 0\le x\le y<\infty$ ו-0 אחרת.

ב. כעת אנחנו צריכים למצוא את הצפיפות המותנית של Y המוגדרת על ידי –

$$f_{Y|X}(y\mid x)=\frac{f_{XY}(x,y)}{f_X(x)}$$

ולשם כך אנחנו צריכים למצוא תחילה את הצפיפות השולית של X.

$$f_X(x)=\int_{-\infty}^\infty f_{XY}(x,y)dy=\int_x^\infty x(y-x)e^{-y}dy=xe^{-x}\int_x^\infty(y-x)e^xe^{-y}dy=$$

$$=xe^{-x}\int_x^\infty(y-x)e^{x-y}dy=xe^{-x}\int_0^\infty ze^{-z}dz=xe^{-x}$$

כאשר ביצענו החלפת משתנים בשורה האחרונה של - $z=y-x$, ולכן

$$f_{Y|X}(y\mid x)=\frac{f_{XY}(x,y)}{f_X(x)}=\frac{x(y-x)e^{-y}}{xe^{-x}}=(y-x)e^{x-y}\quad\forall\infty>y\ge x$$

כעת דורשים מאיתנו למצוא את $E(Y\mid X)$. כפי שציינו קודם לכן, מדובר בפונקציה של X שמקבלת את הערך $E(Y\mid X=x)$ בהסתברות $P(X=x)$.

כאן נסיק כי צריך למצוא למעשה את $E(Y\mid X=x)$ לכל x.

$$E(Y\mid X=x)=\int_{-\infty}^\infty yf_{Y|X}(y\mid x)dy=\int_x^\infty y(y-x)e^{x-y}dy=$$

$$=\int_0^\infty(x+z)ze^{-z}dz=x\int_0^\infty ze^{-z}dz+\int_0^\infty z^2e^{-z}dz=x+2$$

ולכן קיבלנו ש- $E(Y|X)=X+2$ כי $X=k-2 \Leftrightarrow E(Y|X)=k$.

<u>תרגיל 7</u>: יהיו שני משתנים מקריים מעריכיים $X \sim Exp(1), Y \sim Exp(1)$ בלתי תלויים.

מצאו את התפלגות $Z=\frac{X}{X+Y}$.

<u>פתרון תרגיל 7</u>: נחשב את הביטוי לפי הגדרה. נתחיל בכך שנשים לב כי

$0 \le Z=\frac{X}{X+Y}=\frac{1}{1+\frac{Y}{X}} \le 1$ תמיד ולכן נצטרך להסתכל על $0 \le t \le 1$.

$$F_Z(t)=P(Z \le t)=?$$

ננסה למצוא דרך אלטרנטיבית לבטא את המאורע $Z \le t$.

$$\{Z \le t\}=\left\{\frac{X}{X+Y} \le t\right\}=\{X \le t(X+Y)\}=\left\{X\frac{(1-t)}{t} \le Y\right\}=:A$$

ולכן נוכל להשתמש במשתנה מקרי דו מימדי לחישוב הנ"ל.

$$F_Z(t)=P(Z \le t)=P\left(X\frac{(1-t)}{t} \le Y\right)=\int_A f_{X,Y}(x,y)\,dxdy=\int_A e^{-x-y}dxdy=$$

$$=\int_0^\infty e^{-x}dx\int_{x\left(1-t/t\right)}^\infty e^{-y}dy=\int_0^\infty e^{-x}e^{-x\left(\frac{1-t}{t}\right)}dx=\int_0^\infty e^{-x}e^{-x\frac{1}{t}+x}dx=\int_0^\infty e^{-x\frac{1}{t}}dx=t$$

ובאופן די מוזר קיבלנו ש-

$$F_Z(t)=\begin{cases}0 & t<0\\ t & 0 \le t \le 1\\ 1 & t>0\end{cases}$$

מה שאומר ש-Z מתפלג אחיד רציף בין 0 ל-1, $Z \sim U(0,1)$.

<u>תרגיל 8</u>: יהיו $\{X_i\}_{i=1}^n$ מ"מ ב"ת שווי התפלגות עם פונקצית התפלגות מצטברת $F(t)$. מצאו

את ההתפלגות מצטברת של $U=\min\{X_1,\dots,X_n\}$

<u>פתרון תרגיל 8</u>: נבצע חישוב ישיר כפי שעשינו בדרך כלל עבור פונקצית התפלגות מצטברת,

$$F_U(t)=P\left(\min\{X_1,\dots,X_n\} \le t\right)=1-P\left(\min\{X_1,\dots,X_n\}>t\right)=$$

$$=1-P(X_1>t,\dots,X_n>t)=1-P(X_1>t)\cdots P(X_n>t)=1-\prod_{i=1}^n P(X_i>t)=$$

$$=1-\left[P(X_1>t)\right]^n=1-\left[1-P(X_1 \le t)\right]^n=1-\left[1-F(t)\right]^n$$

<u>תרגיל 9</u>: X ו-Y מתפלגים אחיד בטרפז שקודקודיו הם: $(0,1),(0,2),(1,0),(1,3)$ (קרי, X נע בין 0 ל-1 ו-Y נע בהתאם לגבולות הטרפז על פי X).

א. מצאו את הצפיפות המשותפת.

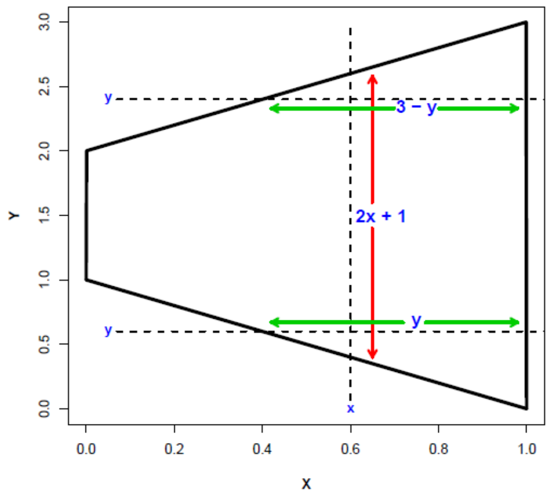
ב. מצאו את הצפיפויות השוליות והתוחלת של כל מ"מ.

ג. מהו מקדם המתאם בין X ל-Y?

ד. האם X ו-Y ב"ת? נמקו.

<u>פתרון תרגיל 9</u>:

א. הצפיפות של המשתנה המקרי הדו מימדי היא חיובית בתוך הטרפז, ז"א בתחום הבא - $D=\{(x,y):0\le x\le 1,1-x\le y\le x+2\}$.



כאשר Y למעשה נע בין 2 קווים ישרים שהם צלעות הטרפז. נחשב את שטח הטרפז ונקבל –

$$S=\frac{[1+3]\cdot 1}{2}=2$$

ולכן הצפיפות המשותפת היא –

$$f_{X,Y}(x,y)=\begin{cases}\frac{1}{2} & (x,y)\in D\\ 0 & otherwise\end{cases}$$

ב. ישנן 2 דרכים לחשב את הצפיפויות השוליות. נתחיל בכך ש-

$$f_X(x)=\int_{-\infty}^{\infty} f_{X,Y}(x,y)dy=\int_{1-x}^{x+2}\frac{1}{2}dy=\frac{1}{2}[x+2-1+x]=\frac{2x+1}{2}=x+\frac{1}{2}$$

עבור $0\le x\le 1$ כמובן.

ובנוסף,

$$f_Y(y)=\int_{-\infty}^{\infty} f_{X,Y}(x,y)dx=\begin{cases}\int_{1-y}^{1}\frac{1}{2}dx=\frac{1}{2}[1-1+y]=\frac{y}{2} & 0\le y<1\\ \int_{0}^{1}\frac{1}{2}dx=\frac{1}{2}0=\frac{1}{2} & 1\le y<2\\ \int_{y-2}^{1}\frac{1}{2}dx=\frac{1}{2}[1-y+2]=\frac{3-y}{2} & 2\le y<3\\ 0 & otherwise\end{cases}$$

ניתן גם באותה מידה לחשב את הצפיפויות השוליות בעזרת הצפיפויות המותנות באופן הבא.

בהינתן ש- $X=x$ (בטרפז) אז Y מתפלג אחיד בקטע $[1-x,x+2]$ שאורכו $2x+1$ ולכן,

$$f_{Y|X}(y|x)=\begin{cases}\frac{1}{2x+1} & 0\le x\le 1\\ 0 & otherwise\end{cases}$$

ומתוך הנוסחה ש- $f_{Y|X}(y|x)=\frac{f_{X,Y}(x,y)}{f_X(x)}$ נקבל ש-

$$f_X(x)=\frac{f_{X,Y}(x,y)}{f_{Y|X}(y|x)}=\frac{\frac{1}{2}}{\frac{1}{2x+1}}=\frac{2x+1}{2}=x+\frac{1}{2};\quad 0\le x\le 1$$

כעת נוכל לבצע אותו דבר עבור ההתפלגות המותנה השנייה אבל מדובר בחישוב ארוך ומיותר... ☺ בגדול הדרך זהה פשוט צריך לחלק למצבים כפי שעשינו קודם לכן ולמצוא את הצפיפות המותנה של X לכל ערך של Y.

נחשב כעת את התוחלת של המשתנים המקריים:

$$E(X)=\int_{-\infty}^{\infty} x\cdot f_X(x)dx=\int_0^1 x\cdot\frac{2x+1}{2}dx=\int_0^1 x^2+\frac{x}{2}dx=\frac{1}{3}+\frac{1}{4}=\frac{7}{12}$$

$$E(Y)=\int_{-\infty}^{\infty} y\cdot f_Y(y)dy=\int_0^1 y\cdot\frac{y}{2}dy+\int_1^2 y\cdot\frac{1}{2}dy+\int_2^3 y\cdot\frac{3-y}{2}dy=$$

$$=\frac{y^3}{6}\Big|_0^1+\frac{y^2}{4}\Big|_1^2+\left[\frac{3y^2}{4}-\frac{y^3}{6}\right]\Big|_2^3=\ldots=1.5$$

שימו לב כי באותה מידה יכולנו להגיד כי קיימת צפיפות אחידה ולכן משיקולי סימטריה התוחלת של Y חייבת לצאת 1.5.

ג. נחשב תחילה את ההתפלגות של המכפלה ואז נמצא את מקדם המתאם.

$$E\left(XY\right)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xy\cdot f_{X,Y}\left(x,y\right)dxdy=\frac{1}{2}\int_{0}^{1}x\int_{1-x}^{x+2}ydydx=$$
$$=\frac{3}{4}\int_{0}^{1}x\left(2x+1\right)dx=\frac{7}{8},$$

ולכן נקבל כי,

$$.Cov\left(X,Y\right)=E\left(XY\right)-E\left(X\right)E\left(Y\right)=\frac{7}{8}-\frac{7}{12}\cdot\frac{3}{2}=0$$

ד. קל לראות כי X ו-Y תלויים כבר מהצפיפויות המותנות בהן ישנה תלות של משתנה אחד בערכו של אחר.

התפלגות דו-נורמאלית

התפלגות מרכזית שראינו בפרק קודם היא ההתפלגות הנורמאלית ובהמשך בהחלט נראה כיצד התפלגות זו היא נפוצה ביותר. כעת , לאחר שלמדנו על משתנים דו מימדים, נוכל להרחיב את ההתפלגות הנורמאלית למקרה הדו-מימדי.

התפלגות נקראית דו נורמאלית אם יש לה את הצפיפות הבאה:

$$f_{XY}(x,y)=\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}\cdot e^{-\frac{1}{2(1-\rho^2)}\cdot\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2-2\rho\cdot\left(\frac{x-\mu_X}{\sigma_X}\right)\cdot\left(\frac{y-\mu_Y}{\sigma_Y}\right)+\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]}$$

כאשר σ_i,μ_i הן התוחלת והשונות של מ"מ i ו- ρ הוא מקדם המתאם בין X ל-Y.

כאשר דנים בהתפלגות דו נורמאלית סטנדרטית אזי המפלצת הנ"ל הופכת למשהו קצת יותר פשוט והוא –

$$f_{XY}(x,y)=\frac{1}{2\pi\sqrt{1-\rho^2}}\cdot e^{-\frac{1}{2(1-\rho^2)}\cdot\left[x^2-2\rho xy+y^2\right]}$$

וכמובן אנחנו רואים שכאשר X ו-Y הם בלתי תלויים (אזי מקדם המתאם יהיה 0), נקבל ש-

$$f_{XY}(x,y)=\frac{1}{2\pi\sigma_X\sigma_Y}\cdot e^{-\frac{1}{2}\cdot\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2+\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]}=$$
$$=\frac{1}{\sqrt{2\pi}\sigma_X}\cdot e^{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2}\frac{1}{\sqrt{2\pi}\sigma_Y}\cdot e^{-\frac{1}{2}\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2}$$

וקיבלנו בדיוק את צורת המכפלה של שתי התפלגויות נורמאליות כפי שצריך לקבל עבור מצב בו צמד המשתנים המקריים הם בלתי תלויים.

הערה חשובה! דרך טובה לזהות האם התפלגות היא דו-נורמאלית זה בעזרת הצפיפות שלה. בהמשך נראה כי למשפחת ההתפלגויות הנורמאליות ישנה תכונת חיבוריות מאוד פשוטה אבל יחד עם זאת, אסור להניח שהתפלגות כלשהי היא נורמאלית בלי לוודא זאת! (לדוגמא פונקציה של התפלגות נורמאלית אינה בהכרח נורמאלית).

פרק 8

נושאים:

- פונקציה יוצרת מומנטים.

פונקציה יוצרת מומנטים

הפונקציה יוצרת המומנטים של מ"מ X מוגדרת באופן הבא:

$$M_X(t)=E\left(e^{tX}\right)$$

על כן, עבור המקרי הבדיד נקבל כי - $M_X(t)=\sum_k e^{tx_k}P(X=x_k)$ ועבור המקרה הרציף נקבל - $M_X(t)=\int_{-\infty}^{\infty}e^{tx}f_X(x)dx$.

שימו לב כי פונקצית יוצרת המומנטים לעיל היא פונקציה של t ולא של המשתנה המקרי! לכל משתנה מקרי קיימת פונקציה משלו.

המומנט ה-k של משתנה מקרי X מוגדר להיות $E\left(X^k\right)$.

ישנו קשר משמעותי בין המומנט ה-k לפונקצית יוצרת המומנטים והוא מבוטא באופן הבא –

$$E\left(X^k\right)=\frac{d^k}{dt^k}M_X(t)\bigg|_{t=0}$$

לכן אם למשתנה מקרי כלשהו X יש מומנטים מכל סדר נוכל לרשום את הזהות הבאה:

$$M_X(t)=\sum_{k=0}^{\infty}\frac{E\left(X^k\right)}{k!}t^k$$

משפט – אם פונקציה יוצרת המומנטים מוגדרת בקטע פתוח הכולל את 0 אזי היא מגדירה ביחידות את ההתפלגות של המשתנה המקרי.

תרגיל 1: מצאו את פונקציה יוצרת המומנטים של $X\sim Exp(\lambda)$ ומצאו למה שווה כל מומנט.

פתרון תרגיל 1: נחשב את פונקצית יוצרת המומנטים ישירות מההגדרה עבור משתנה מקרי רציף.

$$M_X(t)=\int_{-\infty}^{\infty}e^{tx}f_X(x)dx=\int_0^{\infty}e^{tx}\lambda e^{-\lambda x}dx=\lambda\int_0^{\infty}e^{x(t-\lambda)}dx=\frac{\lambda}{t-\lambda}e^{x(t-\lambda)}\bigg|_0^{\infty}=\frac{1}{1-\frac{t}{\lambda}}$$

וחשוב לשים לב לתנאי ש- $t-\lambda<0$ (אחרת האינטגרל מתבדר!).

נשתמש בתנאי הגזירה עבור מציאת המומנטים של X.

$$E\left(X^k\right)=\frac{d^k}{dt^k}M_X\left(t\right)\bigg|_{t=0}=\frac{d^k}{dt^k}\left[1-\frac{t}{\lambda}\right]^{-1}\bigg|_{t=0}=\frac{d^{k-1}}{dt^{k-1}}\left[\frac{1}{\lambda}\left[1-\frac{t}{\lambda}\right]^{-2}\right]\bigg|_{t=0}$$

$$=\frac{1}{\lambda}\cdot\frac{d^{k-2}}{dt^{k-2}}\left[\frac{2}{\lambda}\left[1-\frac{t}{\lambda}\right]^{-3}\right]\bigg|_{t=0}=\frac{1}{\lambda}\cdot\frac{2}{\lambda}\cdot\frac{d^{k-2}}{dt^{k-2}}\left[\left[1-\frac{t}{\lambda}\right]^{-3}\right]\bigg|_{t=0}=\ldots=\frac{k!}{\lambda^k}$$

<u>תרגיל 2</u>: יהי משתנה מקרי פואסוני $X \sim P\left(\lambda\right)$.

א. מצאו את הפונקציה יוצרת המומנטים של X.

ב. מצאו את פונקצית יוצרת המומנטים של $X+c$ עבור $c\in\mathbb{R}$.

ג. מצאו את התוחלת והשונות של $X+c$ בעזרת פונקצית יוצרת המומנטים.

<u>פתרון תרגיל 2</u>:

א. נתחיל בחישוב מפורש של הפונקציה לפי

$$M_X\left(t\right)=\sum_{k=0}^{\infty}e^{tk}P\left(X=k\right)=\sum_{k=0}^{\infty}e^{tk}e^{-\lambda}\frac{\lambda^k}{k!}=e^{-\lambda}\sum_{k=0}^{\infty}\frac{\left(e^t\lambda\right)^k}{k!}=e^{-\lambda}e^{\lambda e^t}=\exp\left(\lambda\left(e^t-1\right)\right)$$

ב. למעשה נוכל לדלג על חישוב מפורש מאחר ו-

$$M_{X+c}\left(t\right)=E\left(e^{t(X+c)}\right)=E\left(e^{tX+tc}\right)=E\left(e^{tX}e^{tc}\right)=e^{tc}E\left(e^{tX}\right)=e^{tc}M_X\left(t\right)=\exp\left(\lambda e^t+tc-\lambda\right)$$

ג. נראה למצוא את התוחלת והשונות של $X+c$ בעזרת הפונקציה יוצרת המומנטים. בשביל לעשות זאת צריכים למצוא את המומנטים הראשון והשני. נשים לב כי במידה ולא היו דורשים את החישוב בעזרת פונקצית יוצרת המומנטים אזי היינו יכולים להגיד כי

$$E\left(X+c\right)=E\left(X\right)+c=\lambda+c$$
$$V\left(X+c\right)=V\left(X\right)=\lambda$$

נראה אם נקבל אותה תשובה בדרך האחרת.

$$E\left(X+c\right)=\frac{d}{dt}M_{X+c}\left(t\right)\bigg|_{t=0}=\frac{d}{dt}\exp\left(\lambda e^t+tc-\lambda\right)\bigg|_{t=0}=\exp\left(\lambda e^t+tc-\lambda\right)\left(\lambda e^t+c\right)\bigg|_{t=0}$$

$$=\exp\left(\lambda e^0+0\cdot c-\lambda\right)\left(\lambda e^0+c\right)=\exp\left(\lambda-\lambda\right)\left(\lambda+c\right)=\lambda+c$$

$$E\left((X+c)^2\right)=\frac{d^2}{dt^2}M_{X+c}(t)\Big|_{t=0}=\frac{d}{dt}\left[\exp\left(\lambda e^t+tc-\lambda\right)\left(\lambda e^t+c\right)\right]\Big|_{t=0}=$$
$$=\left[\exp\left(\lambda e^t+tc-\lambda\right)\left(\lambda e^t\right)+\exp\left(\lambda e^t+tc-\lambda\right)\left(\lambda e^t+c\right)\left(\lambda e^t+c\right)\right]\Big|_{t=0}=$$
$$=\left[\exp\left(\lambda e^0+0\cdot c-\lambda\right)\left(\lambda e^0\right)+\exp\left(\lambda e^0+0\cdot c-\lambda\right)\left(\lambda e^0+c\right)\left(\lambda e^0+c\right)\right]=$$
$$=\left[\exp(\lambda-\lambda)(\lambda)+\exp(\lambda-\lambda)(\lambda+c)(\lambda+c)\right]=\lambda+(\lambda+c)^2$$

ולכן,

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

<u>תרגיל 3</u>:

א. מצאו את הפונקציה יוצרת המומנטים של משתנה מקרי נורמלי סטנדרטי ורשמו את המומנטים מכל סדר.

ב. הרחיבו את תוצאת סעיף א' לכל משתנה מקרי נורמלי כלשהו.

<u>פתרון תרגיל 3</u>:

נתחיל במציאת פונקציה יוצרת המומנטים של $Z\sim N(0,1)$.

$$M_Z(t)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{tx}e^{-\frac{x^2}{2}}dx=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{x^2-2tx}{2}}dx=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{x^2-2tx+t^2-t^2}{2}}dx$$
$$=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{(x-t)^2}{2}+\frac{t^2}{2}}dx=\frac{1}{\sqrt{2\pi}}e^{\frac{t^2}{2}}\int_{-\infty}^{\infty}e^{-\frac{z^2}{2}}dz=e^{\frac{t^2}{2}}$$
$$M_Z(t)=e^{\frac{t^2}{2}}=\sum_{k=0}^{\infty}\frac{\left(t^2/2\right)^k}{k!}=\sum_{k=0}^{\infty}\frac{t^{2k}}{2^k k!}$$
$$\Rightarrow E\left(X^{2k}\right)=\frac{2k!}{2^k k!},\ E\left(X^{2k+1}\right)=0$$

עבור המקרה הכללי $X\sim N\left(\mu,\sigma^2\right)$ נוכל לרשום כי $X=\sigma Z+\mu$ ולכן נקבל ש-

$$M_X(t)=M_{\sigma Z+\mu}(t)=E\left(e^{(\sigma Z+\mu)t}\right)=E\left(e^{\sigma Zt+\mu t}\right)=e^{\mu t}E\left(e^{\sigma Zt}\right)$$
$$=e^{\mu t}E\left(e^{Z\sigma t}\right)=e^{\mu t}M_Z(\sigma t)=e^{\mu t}e^{\frac{(\sigma t)^2}{2}}=\exp\left\{\mu t+\tfrac{\sigma^2t^2}{2}\right\}$$

<u>תרגיל 4</u>: מצאו את פונקציה יוצרת המומנטים של $X\sim G(p)$.

<u>פתרון תרגיל 4</u>: נחשב את פונקצית יוצרת המומנטים ישירות מההגדרה עבור משתנה מקרי גיאומטרי.

$$M_X(t)=E\left(e^{xt}\right)=\sum_{k=1}^{\infty}e^{kt}P\left(X=k\right)=\sum_{k=1}^{\infty}e^{kt}\left(1-p\right)^{k-1}p=$$

$$=\sum_{k=1}^{\infty}e^{kt}p\frac{\left(1-p\right)^{k}}{1-p}=\frac{p}{1-p}\sum_{k=1}^{\infty}e^{kt}\left(1-p\right)^{k}=\frac{p}{1-p}\sum_{k=1}^{\infty}\left(e^{t}-e^{t}p\right)^{k}=$$

$$=\frac{p}{1-p}\sum_{k=1}^{\infty}\left[e^{t}\left(1-p\right)\right]^{k}=\frac{p}{1-p}\left[\frac{e^{t}\left(1-p\right)}{1-e^{t}\left(1-p\right)}\right]=\frac{pe^{t}}{1-\left(1-p\right)e^{t}}$$

פרק 9

נושאים:

- משפט התוחלת השלמה.
- סכום ו/או הפרש משתנים מקריים בלתי תלויים (נוסחת קונבולוציה).

משפט התוחלת השלמה

יהיו X ו-Y משתנים מקריים (X בעל תוחלת), אזי ניתן לחשב את התוחלת של X באופן הבא:

- אם Y מ"מ בדיד: $E(X)=E\left(E(X\mid Y)\right)=\sum_k E(X\mid Y=k)P(Y=k)$.
- אם Y מ"מ רציף: $E(X)=E\left(E(X\mid Y)\right)=\int_{-\infty}^{\infty}E(X\mid Y=k)f_Y(k)dk$.

משפט זה יכול להיות מאוד שימושי כאשר ישנה תלות קלה בין משתנים מקריים ואז חישוב התוחלת המותנה יותר קל ממציאת התוחלת בצורה ישירה.

נוסחת קונבולוציה וסכום / הפרש משתנים מקריים

יהיו X ו-Y מ"מ בלתי תלויים המוגדרים על אותו מרחב הסתברות, אזי ניתן להציג את התפלגות הסכום שלהם באופן הבא:

- במקרה הבדיד: $P_{X+Y}(t)=\sum_k P(X=k)P(Y=t-k)$.
- אם Y מ"מ רציף: $f_{X+Y}(t)=\int_{-\infty}^{\infty}f_X(k)f_Y(t-k)dk$.

שימו לב, כי הנוסחה הנ"ל נכונה גם עבור הפרש של משתנים מקריים מאחר ו-

$X-Y=X+(-Y)$ ואם נגדיר משתנה מקרי חדש $Z=-Y$ אזי $X-Y=X+Z$ ולכן

$$P_{X-Y}(t)=P_{X+Z}(t)=\sum_k P(X=k)P(Z=t-k)=\sum_k P(X=k)P(Y=k-t)$$

ובאופן דומה עבור המקרה הרציף,

$$f_{X-Y}(t)=f_{X+Z}(t)=\int_{-\infty}^{\infty}f_X(k)f_Z(t-k)dk=\int_{-\infty}^{\infty}f_X(k)f_Y(k-t)dk.$$

ישנן עוד תכונות רבות שאפשר לקבל מסכום משתנים מקריים בלתי תלויים, לדוגמא, מכפלת פונ' יוצרות מומנטים.

אם X ו-Y זוג משתנים מקריים עם פו' יוצרת מומנטים, אזי פונ' יוצרת המומנטים של $X+Y$ היא:

$$M_{X+Y}(t)=E\left(e^{(X+Y)t}\right)=E\left(e^{Xt}e^{Yt}\right)=E\left(e^{Xt}\right)E\left(e^{Yt}\right)=M_X(t)M_Y(t)$$

כאשר הפרדת התוחלות אפשרית מאחר והמשתנים המקריים בלתי תלויים ולכן בפרט בלתי מתואמים.

<u>תרגיל 1</u>: בן ואלון משחקים בקלפים. בן מערבב 6 קלפים (עם ערכים של 1 עד 6) ומשאיר אותם הפוכים.

- אם אלון ייבחר ב-1 הוא יקבל 10 שקלים.
- אם ייבחר ב-2 או 3 יפסיד שקל.
- אם ייבחר ב-4 עד 6 הוא יפסיד 2 שקלים.

אלון יכול לפרוש מהמשחק (והוא יעשה זאת) רק אחרי שהוא מנצח בסיבוב אחד. מה תוחלת הרווח של אלון? האם לפרוש מיד אחרי הניצחון הראשון זו החלטה נכונה?

<u>פתרון תרגיל 1</u>: נסמן ב-X את תוחלת הרווח של אלון וב-Y את מספר הקלף בו הוא ייבחר.

נשים לב כי התכונה הבאה מתקיימת: $X|_{Y=2}=X|_{Y=3}=Z-1$ כאשר Z הוא משתנה מקרי המתפלג בדיוק כמו X (מאחר ואחרי כל סיבוב שבו אלון מפסיד הוא מאבד שקל וחוזר לנקודת ההתחלה שלו), $X|_{Y=4}=X|_{Y=5}=X|_{Y=6}=Z-2$ ובנוסף $X|_{Y=1}=10$.

לכן נקבל ש-

$$E(X\mid Y=1)=10$$
$$E(X\mid Y=2)=E(X\mid Y=3)=E(Z-1)=E(Z)-1=E(X)-1$$
$$E(X\mid Y=4)=E(X\mid Y=5)=E(X\mid Y=6)=E(X)-2$$
$$\Downarrow$$
$$E(X)=E(E(X\mid Y))=\sum_{k=1}^{6}E(X\mid Y=k)P_Y(k)=$$
$$=\frac{1}{6}\left[2(E(X)-1)+3(E(X)-2)+10\right]=\frac{1}{6}\left[5E(X)+2\right]$$
$$\Downarrow$$
$$6E(X)=5E(X)+2$$

ולכן תוחלת הרווח היא 2.

נשים לב כי תוחלת הרווח בכל סיבוב היא חיובית ממש $\frac{1}{3}=\frac{1}{6}[10-1-1-2-2-2]$ ולכן אם אלון יחליט להישאר במשחק, אז, בסופו של דבר, הוא יכול להרוויח סכום אינסופי ולכן לא

בהכרח כדאי לו לפרוש מייד (כמובן שאנו מתעלמים מנושא של "אהבת סיכון" / "שנאת סיכון", אבל זה כבר קשור לתחום אחר).

<u>תרגיל 2</u>: אלון גר בת"א באזור ללא חנייה. לפעמים, כאשר אין לו כוח לחנות רחוק מהבית בחניון ציבורי, הוא מחנה באזור אסור ומקבל דו"ח של 100 ש"ח בהסתברות $\frac{1}{2}$. בהנחה שמספר הימים בחודש שאין לו כוח לחנות רחוק הוא $N \sim Bin\left(30, \frac{2}{3}\right)$.

א. מה תוחלת הקנס שייקבל?

ב. הניחו כי אלון מחליט להחנות כל יום ליד הבית במקום אסור ומצאו את הפונקציה יוצרת מומנטים של הקנס החודשי כעת.

<u>פתרון תרגיל 2</u>:

א. אם X זה מספר הדו"חות שיקבל בחודש אז $100X$ זה הקנס החודשי של אלון. נשים לב כי $X | N \sim Bin(N, 0.5)$ ולכן,

$$E(X) = E(E(X|N)) = E\left(\tfrac{1}{2}N\right) = \tfrac{1}{2}E(N) = \tfrac{1}{2}\cdot\tfrac{2}{3}\cdot 30 = 10$$
$$\Rightarrow E(100X) = 1000$$

ב. נוכל לראות מלכתחילה כי X הוא מ"מ בינומי אך בכל זאת נרצה לחשב זאת על ידי סכום אינדיקטורים. נקבע כי - $X = \sum_{i=1}^{30} X_i$ כאשר כל אינדיקטור מתייחס ליום אחר ובוחן האם אלון קיבל דו"ח באותו היום או לא. על כן, $P(X_i = 1) = 0.5$ ולכן,

$$e^{tX_i} = \begin{cases} e^t & w.p\ \ 0.5 \\ 1 & w.p\ \ 0.5 \end{cases} \quad \Rightarrow \quad M_{X_i}(t) = E\left(e^{tX_i}\right) = \frac{1}{2}\left[1 + e^t\right]$$

ועבור X עצמו נקבל,

$$M_X(t) = M_{\sum_{i=1}^{30} X_i}(t) = \prod_{i=1}^{30} \frac{1}{2}\left[1+e^t\right] = \left[\frac{1}{2}\left(1+e^t\right)\right]^{30}$$

$$M_{100X}(t) = E\left(e^{100Xt}\right) = M_X(100t) = \left[\frac{1}{2}\left(1+e^{100t}\right)\right]^{30}$$

<u>תרגיל 3</u>:

א. מצאו את הצפיפות של $X+Y$ עבור $X, Y \sim Exp(\lambda)$ ב"ת.

ב. מצאו את ההתפלגות של $X+Y$ עבור $X \sim P(\lambda_X), Y \sim P(\lambda_Y)$ ב"ת.

ג. מצאו את ההתפלגות / הצפיפות של $X+Y$ עבור $X \sim N(0,1), Y \sim N(0,1)$ בלתי תלויים.

ד. פתרו שוב את סעיף ב' בעזרת פונקציה יוצרת מומנטים.

<u>פתרון תרגיל 3:</u>

- נחשב בעזרת נוסחת קונבולוציה.

$$f_{X+Y}(t) = \int_{-\infty}^{\infty} f_X(k)\, f_Y(t-k)dk$$

נרשום את שתי הצפיפויות בנפרד –

$$f_X(k) = \begin{cases} 0 & k<0 \\ \lambda e^{-\lambda k} & k \ge 0 \end{cases}$$

$$f_Y(t-k) = \begin{cases} 0 & t-k<0 \\ \lambda e^{-\lambda(t-k)} & t-k \ge 0 \end{cases} = \begin{cases} 0 & t<k \\ \lambda e^{-\lambda(t-k)} & t \ge k \end{cases}$$

$$\Downarrow$$

$$f_{X+Y}(t) = \int_0^t \lambda e^{-\lambda k} \lambda e^{-\lambda(t-k)} dk = \lambda^2 e^{-\lambda t} \int_0^t dk = \lambda^2 t e^{-\lambda t} \text{ , } t \ge 0$$

- נחשב שוב בעזרת נוסחת קונבולוציה.

$$P(X+Y=n) = \sum_{k=0}^{n} P(X=k, Y=n-k) = \sum_{k=0}^{n} P(X=k)P(Y=n-k) =$$

$$= \sum_{k=0}^{n} e^{-\lambda_X} \frac{\lambda_X{}^k}{k!} e^{-\lambda_Y} \frac{\lambda_Y{}^{n-k}}{n-k!} = e^{-\lambda_Y} e^{-\lambda_X} \sum_{k=0}^{n} \frac{\lambda_X{}^k \cdot \lambda_Y{}^n}{\lambda_Y{}^k \cdot k! \cdot (n-k!)} =$$

$$= \frac{e^{-(\lambda_Y+\lambda_X)} \cdot \lambda_Y{}^n}{n!} \sum_{k=0}^{n} \left(\frac{\lambda_X}{\lambda_Y}\right)^k \frac{n!}{k! \cdot (n-k!)} =$$

$$= \frac{e^{-(\lambda_Y+\lambda_X)} \cdot \lambda_Y{}^n}{n!} \sum_{k=0}^{n} \left(\frac{\lambda_X}{\lambda_Y}\right)^k \binom{n}{k} = \frac{e^{-(\lambda_Y+\lambda_X)} \cdot \lambda_Y{}^n}{n!} \left(1+\frac{\lambda_X}{\lambda_Y}\right)^n$$

$$= e^{-(\lambda_Y+\lambda_X)} \frac{(\lambda_Y+\lambda_X)^n}{n!} \Rightarrow \quad X+Y \sim P(\lambda_X+\lambda_Y)$$

- נחשב בעזרת נוסחת קונבולוציה.

$$f_{X+Y}(t)=\int_{-\infty}^{\infty} f_X(k)\, f_Y(t-k)dk=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{k^2}{2}}\frac{1}{\sqrt{2\pi}}e^{-\frac{(t-k)^2}{2}}dk=$$

$$=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-\frac{2k^2+t^2-2tk}{2}}dk=\frac{1}{2\pi}e^{-\frac{t^2}{4}}\int_{-\infty}^{\infty}e^{\frac{t^2}{4}}e^{-\frac{2k^2+t^2-2tk}{2}}dk=$$

$$=\frac{1}{2\pi}e^{-\frac{t^2}{4}}\int_{-\infty}^{\infty}e^{-\frac{4k^2-4tk+t^2}{4}}dk=\frac{1}{2\pi}e^{-\frac{t^2}{4}}\int_{-\infty}^{\infty}e^{-\left(\frac{2k-t}{2}\right)^2}dk=$$

$$=\frac{1}{2\pi}e^{-\frac{t^2}{4}}\int_{-\infty}^{\infty}e^{-z^2}dz=\frac{1}{2\pi}e^{-\frac{t^2}{4}}\sqrt{2\pi}=\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{4}}$$

וקיבלנו כי סכום של מ"מ נורמאליים ב"ת מתפלג גם הוא נורמאלי כאשר התוחלת היא סכום התוחלות והשונות היא סכום השונויות. את התוצאה הנ"ל ניתן להכליל למשתנים מקריים נורמאליים ב"ת כלשהם.

- נרשום את הפונקציה יוצרת מומנטים של הסכום

$$M_{X+Y}(t)=E\left(e^{(X+Y)t}\right)=E\left(e^{Xt}\cdot e^{Yt}\right)=E\left(e^{Xt}\right)E\left(e^{Yt}\right)=$$

$$=M_X(t)\cdot M_Y(t)=e^{\lambda_X\left(e^t-1\right)}e^{\lambda_Y\left(e^t-1\right)}=e^{\lambda_X\left(e^t-1\right)+\lambda_Y\left(e^t-1\right)}=$$

$$=e^{(\lambda_X+\lambda_Y)\left(e^t-1\right)}$$

וקיבלנו פונקציה יוצרת מומנטים של מ"מ פואסוני עם פרמטר $\lambda_X+\lambda_Y$ ולכן סכום המשתנים המקריים מתפלג גם הוא פואסוני כאשר הפרמטר החדש הינו סכום הפרמטרים של X ו-Y.

פרק 10

נושאים:

- אי שיוויון מרקוב ואי שיוויון צ'בישב.
- חוקי המספרים הגדולים.
- משפט הגבול המרכזי.

אי שיוויונות מרקוב וצ'בישב

אי שיוויון מרקוב וצ'יבישב הם כלים לבצע הערכות של ההסתברות של מ"מ לקבל ערכים מסויימים.

אי שיוויון מרקוב – יהי X מ"מ בעל תוחלת, אזי לכל $a>0$ מתקיים

$$P\left(\left|X\right|\geq a\right)\leq\frac{E\left(\left|X\right|\right)}{a}$$

כמובן שאם X מקבל רק ערכים אי שליליים אזי ניתן להשמיט את הערך המוחלט בכל מקום באי השיוויון.

אי שיוויון צ'בישב – יהי X מ"מ בעל תוחלת ושונות, אזי לכל $\varepsilon>0$ מתקיים

$$P\left(\left|X-E\left(X\right)\right|\geq\varepsilon\right)\leq\frac{V\left(X\right)}{\varepsilon^2}$$

תזכורות – אנחנו אומרים שלמ"מ ישנה תוחלת או שונות אמ"מ $\sum_k |k| P_X\left(k\right)$ מתכנס או

ש- $\int_{-\infty}^{\infty} k\cdot f_X\left(k\right)dk$ מתכנס בהחלט במקרה הרציף.

חוקי מספרים גדולים

ישנם שני חוקי מספרים גדולים – החלש והחזק. חוקי אלו מדבר על התכנסות של סדרות של מ"מ בעלי אותה תוחלת אל תוחלת בסופו של דבר. ההבדל בינהם הוא לא פשוט ומבוסס על ההבדל בסוגי ההתכנסות האפשריים של משתנים מקריים.

לכל חוק ישנן מספר גרסאות ותנאים ובפרק זה נציג רק את המרכזיים שבהם.

<u>החוק החלש</u> – יהיו $X_1, X_2, \ldots$ סדרת מ"מ ש"ה (שווי התפלגות) בלתי מתואמים עם תוחלת μ ושונות חסומה, אזי לכל $\varepsilon>0$ מתקיים

$\lim_{n\to\infty} P\left(\left|\overline{X_n}-\mu\right|\leq\varepsilon\right)=1$ או לחילופין $\lim_{n\to\infty} P\left(\left|\overline{X_n}-\mu\right|>\varepsilon\right)=0$,

כאשר $\overline{X_n}=\frac{1}{n}\sum_{i=1}^{n}X_i$.

הערות:

- ניתן להחליף את הדרישה לשונות חסומה בכך שהמ"מ יהיו ב"ת.
- בעזרת אי שיוויון צ'בישב קל להראות כי כל סדרת מ"מ בלתי מתואמים ש"ה בעלי שונות מקיימת את החוק החלש.

<u>החוק החזק</u>- יהיו $X_1, X_2, \ldots$ סדרת מ"מ ש"ה (שווי התפלגות) בלתי תלויים עם תוחלת μ, אזי $P\left(\lim_{n\to\infty}\overline{X_n}=\mu\right)=1$.

הערות:

- בחוק החזק ניתן להחליף את הדרישה לכך שהמ"מ הם ש"ה בכך שהטור $\sum_n \frac{V(X_n)}{n^2}$ מתכנס (תנאי קולמוגורוב).
- החוק החזק, כשמו כן הוא, חזק ולכן גורר את החוק החלש. משמע – אם סדרת מ"מ מקיימת את החוק החזק, אזי היא בהכרח מקיימת את החוק החלש, אולם ההיפך אינו נכון.

מה ההבדל בין החוקים?

ובכן, ההבדל בין החוקים הוא די גדול אבל, יחד עם זאת, עדין. החוק החלש קובע כי לכל $\varepsilon>0$ אנחנו יכולים להמשיך ולבצע ניסויים כאשר מנקודה כלשהי ההסתברות שהממוצע יהיה קרוב עד כדי ε לתוחלת בהסתברות מאוד גבוהה (אבל לא בהכרח 1). לעומת זאת, החוק החזק קובע כי עבור כל $\varepsilon>0$ קיימת נקודה כלשהי סופית (תיתכן מאוד רחוקה) כך שממנה והלאה ממוצע הניסויים יהיה קרוב עד כדי ε לתוחלת בהסתברות 1!

<u>משפט הגבול המרכזי</u>

תהיה $X_1, X_2, \ldots$ סדרת מ"מ בלתי תלויים וש"ה בעלי תוחלת μ ושונות σ^2. נגדיר את Z_n להיות $Z_n=\frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}X_i-\sqrt{n}\mu}{\sigma}=\frac{\overline{X_n}-\mu}{\frac{\sigma}{\sqrt{n}}}$, אזי לכל $t\in\mathbb{R}$,

$$F_{Z_n}(t)\underset{n\to\infty}{\to}\Phi(t).$$

המשפט למעשה אומר כי הממוצעים של התפלגויות (באשר הן) מתפלגים בקירוב נורמאלית וזה בלי קשר לסוג ההתפלגות במקור.

<u>תרגיל 1</u>: יהי X ו-Y מ"מ כך ש- $X \sim U(0,1), Y \mid X \sim Bin(20, X)$. מצאו חסם עליון להסתברויות הבאות:

א. $P(Y \geq 5)$ ו- $P(Y \geq 15)$.

ב. חשבו את ההסתברות $P(Y \geq 15)$ בצורה מדוייקת.

<u>פתרון תרגיל 1</u>:

א. בכדי למצוא חסם עליון כנדרש ומאחר וY הוא מ"מ אי שלילי אז יהיה נוח להשתמש באי שיוויון מרקוב, $P(Y \geq 15) \leq \frac{E(Y)}{15}, P(Y \geq 5) \leq \frac{E(Y)}{5}$ ולכן כל שנותר לעשות זה למצוא את התוחלת של Y.

$$E(Y) = E(E(Y \mid X)) = E(20X) = 20E(X) = 20\frac{1}{2} = 10$$

ולכן,

$$P(Y \geq 5) \leq 2 \,,\, P(Y \geq 15) \leq \tfrac{2}{3}$$

ואנו רואים כי החסם הראשון הוא חסר משמעות ובסעיף הבא נראה שהשני ממש לא קרוב לערך האמיתי.

ב. נחשב את התפלגות Y מפורשות. נתחיל בכך שנשים לב כי Y מקבל ערכים שלמים אי שליליים בלבד.

$$P(Y = y) = \int_{-\infty}^{\infty} P(Y = y \mid X = k) f_X(k) dk = \int_0^1 \binom{20}{y} k^y (1-k)^{20-y} dk =$$

$$= \ldots = \frac{1}{21}, \quad y = 0,1,\ldots,20$$

כאשר מי שרוצה לחשב את האינטגרל האחרון בעצמו מוזמן (זה לוקח איזה יום / יומיים ☺ ובכל מקרה עדיף לבחור כמה ערכים של Y ולראות מה יוצא). על כן קיבלנו כי Y מתפלג אחיד בין 0 ל-20 ולכן

$$P(Y \geq 15) = P(Y \in \{15,16,17,18,19,20\}) = \frac{6}{21} = \frac{2}{7}$$

משמע החסם של מרקוב גדול פי $\frac{7}{3}$ מהערך המדויק.

<u>תרגיל 2</u>: גובה של גבר מתפלג נורמאלית $Y \sim N(178,3)$. השתמשו באי שיוויון מרקוב בכדי להעריך את ההסתברות ש-5 גברים הנבחרים באקראי יהיו כולם מעל 180 ס"מ. השתמשו בפונקציה יוצרת המומנטים בכדי לשפר את הדיוק של ההערכה הקודמת.

<u>פתרון תרגיל 2</u>: אם נסמן את מספר הגברים מתוך חמישה שגובהם הוא מעל 1.8 מטר ב-X
אזי ברור כי X הוא משתנה מקרי בינומי עם הסתברות,

$$P(Y>180)=P\left(\frac{Y-178}{\sqrt{3}}>\frac{180-178}{\sqrt{3}}\right)=1-\Phi\left(\frac{2}{\sqrt{3}}\right)\approx 0.124$$

ולכן $X \sim Bin(5,0.124)$ ובעזרת מרקוב נקבל,

$$P(X=5)=P(X\geq 5)\leq\frac{E(X)}{5}=\frac{5\cdot 0.124}{5}=0.124$$

נרצה לשפר חסם זה ונוכל בעזרת פונקציה יוצרת המומנטים.

$$P(X\geq 5)=P\left(e^{Xt}\geq e^{5t}\right)\leq\frac{E\left(e^{Xt}\right)}{e^{5t}}=\frac{M_X(t)}{e^{5t}}=$$

$$=\frac{\left(1-0.124+0.124e^t\right)^5}{e^{5t}}=\left(e^{-t}-e^{-t}0.124+0.124\right)^5$$

ניקח מינימום על t מאחר והחסם נכון לכל t עבור בגבול שהוא שואף לאינסוף נקבל כי

$$P(X\geq 5)\leq(0.124)^5$$

שזה בדיוק ההסתברות למצב בו כל חמשת הגברים מעל 180 ס"מ.

<u>תרגיל 3</u>: משחקי מזל!

א. מטילים מטבע שוב ושוב ובכל פעם שיוצא 'פלי' מקבלים שקל. מה ההסתברות שממוצע הזכיות ייסטה ביותר מ-0.1 מחצי שקל לאחר 100 הטלות?

ב. ברולטה אמריקאית יש 38 מספרים – 18 אדומים, 18 שחורים ו-2 ירוקים. ניתן להמר עם שקל אחד על צבעים אדום ושחור וכל זכייה מזכה בשקל נוסף. כמה הימורים כאלה צריכים להתרחש בכדי שהקזינו ירוויח בהסתברות של לפחות 95%?

ג. האם סדרות ההימורים הנ"ל מקיימות את חוקי המספרים הגדולים?

<u>פתרון תרגיל 3</u>:

א. נסמן ב- $\overline{X_{100}}=\frac{\sum_{i=1}^{100}X_i}{100}$ את ממוצע הזכיות ב-100 ההטלות וב- X_i את הרווח בסיבוב ה- i. X_i הוא משתנה מקרי ברנולי עם הסתברות 0.5 וכל משתנה מקרי כזה בלתי תלוי באחרים ולכן,

$$E\left(\overline{X_{100}}\right)=E\left(\frac{\sum_{i=1}^{100}X_i}{100}\right)=\frac{1}{100}E\left(\sum_{i=1}^{100}X_i\right)=\frac{1}{100}\cdot100\cdot\frac{1}{2}=0.5$$

$$V\left(\overline{X_{100}}\right)=V\left(\frac{\sum_{i=1}^{100}X_i}{100}\right)=\frac{1}{100^2}V\left(\sum_{i=1}^{100}X_i\right)=\frac{1}{100^2}\sum_{i=1}^{100}V\left(X_i\right)=\frac{1}{100^2}100\cdot\frac{1}{4}=\frac{1}{400}$$

כעת נוכל להשתמש באי שיוויון צ'בישב ולקבל

$$P\left(\left|\overline{X_{100}}-0.5\right|\geq0.1\right)\leq\frac{V\left(\overline{X_{100}}\right)}{0.1^2}=\frac{1/400}{1/100}=0.25$$

ב. נסמן ב- $\overline{Y_n}=\frac{\sum_{i=1}^{n}Y_i}{n}$ את ההפסד הממוצע של הקזינו ב-n סיבובים ואת Y_i כהפסד הקזינו בסיבוב ה- i. נחשב את התוחלת והשוניות לכל המשתנים המקריים הללו:

$$P\left(Y_i=1\right)=\frac{18}{38}=\frac{9}{19}\quad;\quad P\left(Y_i=-1\right)=1-\frac{9}{19}=\frac{10}{19}$$

$$E\left(Y_i\right)=-\frac{1}{19}\quad;\quad V\left(Y_i\right)=E\left(Y_i^2\right)-E\left(Y_i\right)^2=1-\frac{1}{361}=\frac{360}{361}$$

$$E\left(\overline{Y_n}\right)=E\left(\frac{\sum_{i=1}^{n}Y_i}{n}\right)=-\frac{1}{19}\quad;\quad V\left(\overline{Y_n}\right)=V\left(\frac{\sum_{i=1}^{n}Y_i}{n}\right)=\frac{360}{n\cdot361}$$

נשים לב כי כדאי שהקזינו ייצא מרוויח בהסתברות 0.95 הוא צריך שההסתברות להפסד תהיה קטנה מ-0.05.

אם נבחן את ההפסד הממוצע ב-n סיבובים, אזי נדרוש ש- $P\left(\overline{Y_n}>0\right)\leq0.05$ או

לחילופין, $P\left(\overline{Y_n}\geq\frac{1}{n}\right)\leq0.05$ מאחר וההפסד המינימאלי ב-n סיבובים הוא $\frac{1}{n}$ אזי

המאורעות הנ"ל שווים $\left\{\overline{Y_n}\geq\frac{1}{n}\right\}=\left\{\overline{Y_n}>0\right\}$ ולכן,

$$P\left(\overline{Y_n}>0\right)=P\left(\overline{Y_n}\ge\frac{1}{n}\right)=P\left(\overline{Y_n}+\frac{1}{19}\ge\frac{1}{n}+\frac{1}{19}\right)\le P\left(\left|\overline{Y_n}+\frac{1}{19}\right|\ge\frac{19+n}{19n}\right)\le$$

$$\le\frac{V\left(\overline{Y_n}\right)}{\left(\frac{19+n}{19n}\right)^2}=\frac{360/n\cdot 361}{(19+n)^2/n^2\cdot 361}=\frac{360n}{\left(19+n\right)^2}$$

כעת נדרוש שהביטוי האחרון יהיה קטן מ-0.05 ונקבל את האי שיוויון הבא:

$$\frac{360n}{\left(19+n\right)^2}\le 0.05 \quad\Leftrightarrow\quad 7200n\le\left(19+n\right)^2 \quad\Leftrightarrow\quad 0\le n^2-7162n+361$$

$$\Leftrightarrow\quad 7162\le n$$

ולכן 7162 זה מספר הסיבובים הדרושים בכדי שהקזינו ירוויח בהסתברות 95%.

ג. בשני המקרים מדובר בסדרות של מ"מ ב"ת ש"ה (עם אותה תוחלת ושונות, כמובן) ולכן צמד חוקי המספרים הגדולים חלים עליהם.

<u>תרגיל 4</u>: מספר המכוניות העוברות ביום i בכביש 6 הוא מ"מ X_i כך ש- $E\left(X_i\right)=V\left(X_i\right)=2400$. ישנו מחשב המודד בעזרת מצלמות את מספר הרכבים העוברים בכביש ומחשב את ממוצע הרכבים היומי מאז הקמת הכביש.

א. האם קיים מספר ימים n שהחל ממנו בהסתברות 0.99 לפחות הממוצע היומי הוא מעל 2399?

ב. מאתחלים את המחשב ומבצעים מדידות מחדש במשך 150 ימים. מה הסיכוי שהממוצע היומי קטן מ-2412?

<u>פתרון תרגיל 4</u>:

א. אנחנו אומנם לא יודעים מה התפלגות מספר המכוניות היומית אבל ממשפט הגבול המרכזי עדיין נוכל לחשב את הנדרש בצורה יחסית קלה. נשים לב כי לפי משפט הגבול המרכזי מתקיים ש- $\frac{\overline{X_n}-E\left(X_i\right)}{\sigma\left(X_i\right)/\sqrt{n}}=\frac{\overline{X_n}-2400}{\sqrt{2400}/\sqrt{n}}\sim N\left(0,1\right)$.

אנחנו צריכים למצוא את מספר הניסויים n כך שיתקיים - $P\left(\overline{X_n}\ge 2399\right)\ge 0.99$.

$$P\left(\overline{X_n}\geq 2399\right)=P\left(\frac{\overline{X_n}-2400}{\sqrt{2400}/\sqrt{n}}\geq\frac{2399-2400}{\sqrt{2400}/\sqrt{n}}\right)\approx P\left(Z\geq\frac{2399-2400}{\sqrt{2400}/\sqrt{n}}\right)$$
$$=P\left(Z\geq-\sqrt{\frac{n}{2400}}\right)=1-\Phi\left(-\sqrt{\frac{n}{2400}}\right)=\Phi\left(\sqrt{\frac{n}{2400}}\right)$$

כעת נוכל ללכת לטבלת התפלגות נורמאלית ולמצוא את הערך t המינימאלי כך ש-

$\Phi(t)\geq 0.99$ ונמצא כי זה מתקיים עבור $t=2.326$ ולכן

$n\geq(2.326)^2\cdot 2400\approx 13{,}000$.

ב. נבצע חישוב מפורש של הנדרש בשאלה באופן הבא –

$$P\left(\overline{X_{150}}<2412\right)=P\left(\frac{\overline{X_{150}}-2400}{\sqrt{2400}/\sqrt{150}}<\frac{2412-2400}{\sqrt{2400}/\sqrt{150}}\right)\approx$$
$$\approx\Phi\left(12\sqrt{\frac{150}{2400}}\right)=\Phi(3)=0.9987$$

<u>תרגיל 5</u>: שיכור מבצע הילוך מקרי על ציר X. בזמן 0 הוא נמצא בנקודת ה-0. בכל יחידת זמן הוא זז יחידת מרחק שלמה אחת ימינה או שמאלה בסיכוי שווה ובאופן בלתי תלוי בזמנים אחרים.

א. מה בקירוב ההסתברות שבשלב ה-900 הוא יהיה ממוקם מימין לנקודה 44?

ב. נניח שבשלב ה-100 השיכור היה בנקודה 10+, מה הסיכוי שהתנועה הראשונה שלו הייתה ימינה?

<u>פתרון תרגיל 5</u>:

א. נגדיר את X להיות מיקומו של השיכור בשלב ה-900. ניתן לפרק את X ל-900 מ"מ שמקבלים את הערכים ± 1 בהסתברות שווה. נמצא את התוחלת והשונות של כל אינדיקאטור בפירוק –

$$E(X_i)=1\cdot\frac{1}{2}-1\cdot\frac{1}{2}=0$$
$$V(X_i)=E\left({X_i}^2\right)-E(X_i)^2=(1)^2\cdot\frac{1}{2}+(-1)^2\cdot\frac{1}{2}-0^2=1$$
$$E(X)=E\left(\sum_{i=1}^{900}X_i\right)=\sum_{i=1}^{900}E(X_i)=900\left(1\cdot\frac{1}{2}-1\cdot\frac{1}{2}\right)=0$$
$$V(X)=V\left(\sum_{i=1}^{900}X_i\right)=\sum_{i=1}^{900}V(X_i)=900\left((1)^2\cdot\frac{1}{2}+(-1)^2\cdot\frac{1}{2}-0^2\right)=900$$

נוכל כעת להשתמש במשפט הגבול המרכזי ולקבל ש-

$$P\left(X \geq 45\right) = P\left(\frac{X - n\mu}{\sigma\sqrt{n}} \geq \frac{45 - n\mu}{\sigma\sqrt{n}}\right) = P\left(\frac{X - \sqrt{900}\cdot 0}{1\cdot\sqrt{900}} \geq \frac{45 - \sqrt{900}\cdot 0}{1\cdot\sqrt{900}}\right)$$

$$\approx P\left(Z \geq \frac{3}{2}\right) = 1 - \Phi\left(\frac{3}{2}\right) \approx 0.07$$

כאשר הצבנו $\sigma = \sqrt{V\left(X_i\right)} = 1$, $\mu = E\left(X_i\right) = 0$, $n = 900$.

ב. נוכל לנסות לפתור זאת בדרך מסובכת על ידי חישובים ארוכים ומורכבים או, לחילופין, נוכל לנסות למצוא דרך יותר מתוחכמת לפתרון. אם נתון לנו שהוא נמצא כרגע במיקום של 10+ זה אומר שהוא עשה סה"כ 55 צעדים ימינה ו-45 צעדים שמאלה. ז"א שיחס הצעדים שלו זה 55 ל-45 עם עדיפות לצעדים ימינה. מכאן נוכל בעצם להסיק שההסתברות של כל צעד להיות ימינה היא 0.55 וההסתברות של כל צעד להיות שמאלה היא 0.45 וזה בגלל שנתון לנו בדיוק כמה צעדים הוא ביצע ולאן (הדבר היחיד שנותר לנו לחשב זה את סידור הצעדים). על כן משיקולים אלו יחד עם שיקולי סימטריה נובע כי ההסתברות שהצעד הראשון היה ימינה הוא 0.55.

פרק 11

נושאים:

- אמידה נקודתית וחוסר הטיה
- אומד נראות מקסימאלית.
- רווח בר-סמך.

אמידה נקודתית וחוסר הטיה

אנו מעוניינים להעריך גודלו של פרמטר מסוים המאפיין אוכלוסיה כלשהי, לדוגמא – תוחלת, חציון, שונות וכד'. הנטייה הטבעית היא לבצע הערכה כזאת על בסיס מדגם מקרי מתוך האובייקטים באוכלוסיה.

עבור ערך / פרמטר θ מסוים אותו אנו מעוניינים לאמוד, נוכל להגדיר פונקציה T שתקרא ***האומד*** והיא תשמש אותו לטובת הערכת גודלו של θ (פעולה הנקראת 'אמידה').

לדוגמא: אם אנו מעוניינים לאמוד את ההסתברות למוצר פגום בפס ייצור אנחנו יכולים להניח שמספר המוצרים הפגומים במדגם מתפלג בינומית (מאחר וישנם מוצרים רבים, אנו מניחים בקירוב כאילו המדגם בוצע עם החזרות) ולכן מספר המוצרים הפגומים X יתפלג לפי $X \sim Bin(n,p)$ כאשר אנו קובעים את מספר המוצרים במדגם.

נוכל כעת להגדיר אומד T באופן הבא: $T=\frac{X}{n}$. ז"א שאנחנו מעריכים כי מספר המוצרים הפגומים במדגם חלקי הגודל של המדגם מייצג את ההסתברות לקבלת מוצר פגום בפס הייצור.

לאומד שכזה אנחנו קוראים אומד בלתי מוטה מאחר ו-

$$E(T)=E\left(\frac{X}{n}\right)=\frac{1}{n}E(X)=\frac{1}{n}\cdot np=p.$$

זאת אומרת, שהתוחלת של האומד שווה לערך שאנו מעוניינים לאמוד. בכל פעם בה $E(T)=\theta$ לכל ערך של הפרמטר θ, אנו קוראים לאומד 'בלתי מוטה'.

הגדרות:

- ***ההטיה*** של אומד T מוגדרת להיות ההפרש בין התוחלת של T לערך הנאמד, $E(T)-\theta$.
- ***ההפסד / הסיכון*** מוגדר להיות ריבוע הטעות של האומד ונרשם באופן הבא:

$$MSE(T)=E\left[(T-\theta)^2\right].$$

אומד בלתי מוטה שקול לכך שההפסד / הסיכון של האומד שווה לשונות שלו.

$$MSE(T)=E\left[(T-\theta)^2\right]=Var(T)+\left(E(T)-\theta\right)^2$$
$$\Rightarrow MSE(T)=Var(T)\Leftrightarrow E(T)=\theta$$

אם אנו דוגמים על פי תצפיות בלתי תלויות אזי האומד הבלתי מוטה עבור התוחלת הוא

ממוצע המדגם $\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$ והאומד הבלתי מוטה עבור השונות הוא ריבוע סטיית התקן

המדגמית (אם התוחלת אינה ידועה)

$$S^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\frac{1}{n}\sum_{i=1}^{n}X_i\right)^2.$$

אם התוחלת ידועה אזי

$$\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$$

אומד נראות מקסימאלית

אומד נראות מקסימאלית הוא אומד שמתקבל מאופטימיזציה של פונקצית הנקראת פונקצית נראות.

פונקצית נראות היא פונקציה של האומד ומוגדרת באופן הבא:

$$L(\theta)=L(\theta,x_1,\ldots,x_n)=\begin{cases}\prod_{i=1}^{n}f_\theta(x_i) & Cont.\,dist.\\ \prod_{i=1}^{n}P_\theta(X_i=x_i) & else\end{cases}$$

תחת ההנחה שהדגימות שלנו בלתי תלויות ונקבעות לפי התפלגות כלשהי, פונקצית הנראות המקסימאלית המוגדרת לעיל מקבל ערכים מסויימים עבור כל פרמטר שנבחר θ. לאחר כתיבת הפונקציה נוכל לגזור אותה (או את הלוגריתם שלה, מאחר ומדובר בפונקציה מונוטונית עולה) ולהשוות ל-0 על מנת למצוא נקודת מקסימום $\widehat{\theta}$. נקודה זו (שהיא פונקציה של תוצאות המדגם) נקראת ***אומד נראות מקסימאלית*** והיא מקיימת את תנאי המקסימאליות $L\left(\widehat{\theta}\right)\geq L(\theta)\ \ \forall\theta$ (מאחר והיא נקודת המקסימום של פונקצית הנראות).

אינטואיטיבית, נוכל לחשוב על $\widehat{\theta}$ כאומד אשר ממקסם את ההסתברות לקבל תוצאות מדגם כלשהן.

רווח בר-סמך

רווח בר סמך ברמת סמך $1-\alpha$ הוא קטע המתבסס על מדגם מקרי כך שההסתברות שהפרמטר יהיה בתוך קטע זה היא $1-\alpha$.

נגדיר זאת פורמלית עבור התוחלת. נניח כי יש לנו מדגם מקרי $X_1,\dots,X_n$ שהוא תצפיות בלתי תלויות התלויות בהתפלגות כלשהי עם תוחלת לא ידועה μ ושונות ידועה σ^2. מחוק המספרים הגדולים אנחנו יודעים כי אם נבצע מספר רב של דגימות (מדגם מספיק גדול) אזי הוא יתפלג נורמאלית בקירוב ולכן מתקיים ש-

$$P\left(z_{\alpha/2}\le \frac{\overline{X}_n-\mu}{\sigma/\sqrt{n}}\le z_{1-\alpha/2}\right)\cong 1-\alpha$$

או לחילופין,

$$1-\alpha\cong P\left(z_{\alpha/2}\le \frac{\overline{X}_n-\mu}{\sigma/\sqrt{n}}\le z_{1-\alpha/2}\right)=P\left(-z_{1-\alpha/2}\le \frac{\overline{X}_n-\mu}{\sigma/\sqrt{n}}\le z_{1-\alpha/2}\right)=$$
$$=P\left(\overline{X_n}-\frac{\sigma}{\sqrt{n}}z_{1-\alpha/2}\le\mu\le\overline{X_n}+\frac{\sigma}{\sqrt{n}}z_{1-\alpha/2}\right)$$

ואז הרווח הבא מוגדר להיות ***רווח בר-סמך לתוחלת ברמת סמך*** $1-\alpha$:

$$.\left[\overline{X_n}-\frac{\sigma}{\sqrt{n}}z_{1-\alpha/2},\overline{X_n}+\frac{\sigma}{\sqrt{n}}z_{1-\alpha/2}\right]$$

<u>תרגיל 1 (א.נ.מ)</u>: יהיו n תצפיות בלתי תלויות מהתפלגות גיאומטרית $X_i\sim G(p),\ i=1,\dots n$.

א. רשמו את פונקצית הנראות ומצאו אומד נראות מקסימאלית עבור p.

ב. עבור n=1, הוכיחו כי מדובר באומד מוטה.

<u>פתרון תרגיל 1</u>:

א. נתחיל ברישום פונקצית הנראות עבור ההסתברות p.

$$L(p)=L(p,x_1,\dots x_n)=\prod_{i=1}^{n}P_\theta(X_i=x_i)=\prod_{i=1}^{n}p(1-p)^{x_i-1}=$$
$$=p^n(1-p)^{\sum_{i=1}^{n}[x_i-1]}=p^n(1-p)^{\sum_{i=1}^{n}x_i-n}$$

כעת נשים לב כי אם $\sum_{i=1}^{n}x_i=n$ אז $L(p)=p^n$ וזו פונקציה מונוטונית ב-p שמקבלת מקסימום בערך 1. אחרת, $\sum_{i=1}^{n}x_i>n$ ואז

$$L(p)=p^n(1-p)^{\sum_{i=1}^{n}x_i-n}\Leftrightarrow \ln(L(p))=n\ln(p)+\left[\sum_{i=1}^{n}x_i-n\right]\ln(1-p)$$

$$\Rightarrow \frac{d}{dp}\ln(L(p))=0=\frac{n}{p}-\frac{\sum_{i=1}^{n}x_i-n}{1-p}\Rightarrow p=\frac{n}{\sum_{i=1}^{n}X_i}$$

זאת אומרת שקיבלנו שאומד נראות מקסימאלית הוא מספר התצפיות חלקי מספר הניסויים בכל תצפית.

ב. במקרה ש-n=1 אומד הנראות המקסימאלית הוא $p=\frac{1}{X}$. נבדוק האם הוא מוטה.

$$E(p)=E\left(\frac{1}{X}\right)=\sum_{k=1}^{\infty}\frac{1}{k}P_p(X=k)=\sum_{k=1}^{\infty}\frac{1}{k}p(1-p)^{k-1}=$$
$$=p+\sum_{k=2}^{\infty}\frac{1}{k}p(1-p)^{k-1}\geq p$$

כאשר האי שיוויון האחרון הופך להיות גדול ממש אם ההסתברות היא בין 0 ל-1 ולכן האומד הזה הוא מטה כלפי מעלה.

<u>תרגיל 2 (א.נ.מ, רווח בר-סמך)</u>: גובה של סטודנטית הוא משתנה מקרי המתפלג נורמאלית עם פרמטרים לא ידועים. במדגם של 100 סטודנטיות נמצא ממוצע מדגמי של 167 ס"מ וסטיית תקן מדגמית $S=5.1$.

א. מצאו אומד נראות מקסימאלית לתוחלת ולסטיית התקן של גובה סטודנטיות באוכלוסיה. מה הערכים שהם מקבלים במדגם שלנו?

ב. נניח שגובה הסטודנטיות באוכלוסיה מתפלג נורמאלית עם שונות 25. בנו רווח סמך ברמת סמך 98% עבור המדגם הנ"ל כדי לאמוד את תוחלת גובה הסטודנטיות באוכלוסיה.

ג. מהו רווח הסמך שיתקבל אם נסתפק במדגם של 25 סטודנטיות (בהנחה שממוצע המדגם יצא זהה)?

ד. לאיזה גודל מדגם נזדקק ע"מ להבטיח בהסתברות של 98% שממוצע המדגם סוטה לכל היותר ס"מ 1 מן הממוצע באוכלוסיה?

<u>פתרון תרגיל 2</u>:

א. נתחיל בבניית פונקצית הנראות עבור מדגם $X_1,\dots,X_n \sim N(\mu,\sigma^2)$.

$$L(\mu,\sigma)=L(\mu,\sigma,X_1,...,X_n)=\prod_{i=1}^{n}f_{X_i}(x_i)=\prod_{i=1}^{n}\tfrac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(x_i-\mu)^2}{2\sigma^2}\right\}=$$

$$=(2\pi\sigma^2)^{-n/2}\exp\left\{-\tfrac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$$

כעת נוכל לקחת לוגריתם לפונקציה הנ"ל ולגזור אותה ולקבל אומדן עבור התוחלת והשונות.

$$\ln(L(\mu,\sigma))=-\tfrac{n}{2}\left[\ln(2\pi)+\ln(\sigma^2)\right]-\tfrac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\frac{\partial}{\partial\mu}\ln(L(\mu,\sigma))=0=\tfrac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)\Rightarrow\mu=\frac{1}{n}\sum_{i=1}^{n}x_i$$

$$\frac{\partial}{\partial\sigma^2}\ln(L(\mu,\sigma))=0=-\tfrac{n}{2\sigma^2}+\tfrac{1}{2\sigma^4}\sum_{i=1}^{n}\left(x_i-\mu\right)\Rightarrow\sigma^2=\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\mu\right)^2$$

ולכן קיבלנו כי

$$\mu=\frac{1}{n}\sum_{i=1}^{n}x_i=167$$

$$\sigma=\sqrt{\sigma^2}=\sqrt{\frac{n-1}{n}\cdot\frac{1}{(n-1)}\sum_{i=1}^{n}\left(x_i-\mu\right)^2}=\sqrt{\frac{n-1}{n}\cdot S^2}=\sqrt{\frac{99}{100}\cdot 5.1^2}\approx 5.074$$

במעבר האחרון השתמשנו בהגדרת סטיית התקן המדגמית הנתונה לנו בשאלה.

ב. כזכור, מחוק המספרים הגדולים נובע כי

$$P\left(z_{\alpha/2}\le\tfrac{\overline{X_n}-\mu}{\sigma/\sqrt{n}}\le z_{1-\alpha/2}\right)=P\left(-z_{1-\alpha/2}\le\tfrac{\overline{X_n}-\mu}{\sigma/\sqrt{n}}\le z_{1-\alpha/2}\right)\cong 1-\alpha$$

ולכן נוכל לקבל לאחר העברת אגפים כי,

$$.1-\alpha\cong P\left(\overline{X_n}-\tfrac{\sigma}{\sqrt{n}}z_{1-\alpha/2}\le\mu\le\overline{X_n}+\tfrac{\sigma}{\sqrt{n}}z_{1-\alpha/2}\right)$$

על כן, הקטע $\left[\overline{X_n}-\tfrac{\sigma}{\sqrt{n}}z_{1-\alpha/2},\overline{X_n}+\tfrac{\sigma}{\sqrt{n}}z_{1-\alpha/2}\right]$ נותן לנו רווח סמך לתוחלת ברמת סמך $1-\alpha$ (ז"א, ההסתברות שהתוחלת תהיה בקטע זה היא $1-\alpha$). עבור מדגם של 100 סטודנטיות ודרישה לרמת סמך 98% נקבל את הקטע

$$\left(\overline{X_{100}}-\tfrac{5}{\sqrt{100}}z_{1-0.02/2},\overline{X_{100}}+\tfrac{5}{\sqrt{100}}z_{1-0.02/2}\right)=\left(167-\tfrac{1}{2}z_{0.99},167+\tfrac{1}{2}z_{0.99}\right)=$$

$$=\left(167-\tfrac{1}{2}\cdot 2.326,167+\tfrac{1}{2}\cdot 2.326\right)=$$

$$=(165.837,168.163)$$

ג. נציב 25 במקום 100 בנוסחה שקיבלנו לעיל,

$$\left(\overline{X_{25}}-\tfrac{5}{\sqrt{25}}z_{1-0.02/2},\overline{X_{25}}+\tfrac{5}{\sqrt{25}}z_{1-0.02/2}\right)=(164.674,169.236)$$

וקיבלנו כי הגדלת המדגם מאפשרת לנו להקטין את רווח הסמך ולדייק את האומדן.

ד. מהנוסחה לרווח סמך עבור התוחלת אנו למדים כי שאורך רווח הסמך הוא

$$l=2\tfrac{\sigma}{\sqrt{n}}z_{1-\alpha/2}.$$

נשים לב כי דורשים בשאלה שממוצע המדגם סוטה לכל היותר בס"מ 1 מן הממוצע באוכלוסיה. ממוצע המדגם נמצא במרכז המרווח וממוצע האוכלוסיה נמצא בהסתברות 98% בתוך הקטע ולכן למעשה דורשים שאורך הקטע יהיה קטן מ-2 ס"מ (מרחק כל קצה מהמרכז יהיה לפחות ס"מ 1). נציב זאת ונקבל,

$$2\geq 2\tfrac{5}{\sqrt{n}}z_{0.99}\Leftrightarrow n\geq(5\cdot 2.326)^2=135.26$$

ולכן צריך לפחות 136 סטודנטיות בכדי להבטיח את הנדרש בסעיף.

<u>תרגיל 3 (א.נ.מ)</u>: יהיו n תצפיות בלתי תלויות מהתפלגות מעריכית $X_i \sim Exp(\lambda),\ i=1,\dots n$.

א. רשמו את פונקצית הנראות ומצאו אומד נראות מקסימאלית עבור λ.

ב. עבור n=1, קבעו האם מדובר באומד מוטה או לא.

ג. פתרו את צמד הסעיפים הקודמים עבור $X_i \sim P(\lambda),\ i=1,\dots n$.

ד. מצאו אומד נראות מקסימאלית עבור התוחלת והשונות בהינתן $X_i \sim N(\mu,\sigma^2),\ i=1,\dots n$.

<u>פתרון תרגיל 3</u>:

א. נתחיל ברישום פונקצית הנראות:

$$L(\lambda)=L(\lambda,x_1,\dots x_n)=\prod_{i=1}^{n}f_\lambda(x_i)=\prod_{i=1}^{n}\lambda e^{-\lambda x_i}=\lambda^n e^{-\lambda\sum_{i=1}^{n}x_i}$$

$$\ln(L(\lambda))=n\ln(\lambda)-\lambda\sum_{i=1}^{n}x_i$$

ולכן אם נגזור את פונקצית הנראות נקבל כי

$$\frac{d}{d\lambda}\ln(L(\lambda))=0=\frac{n}{\lambda}-\sum_{i=1}^{n}x_i\Rightarrow\lambda=\frac{n}{\sum_{i=1}^{n}X_i}$$

ב. עבור $n=1$ מתקיים ש-

$$E\left(\hat{\lambda}\right)=E\left(\tfrac{1}{\bar{X}}\right)\neq\tfrac{1}{E(\bar{X})}=\lambda$$

ולכן אומד זה מוטה.

ג. פונקצית הנראות עבור דגימות ממ"מ מהתפלגות פואסונית היא

$$L(\lambda)=L(\lambda,x_1,\ldots x_n)=\prod_{i=1}^{n}P_\lambda\left(X_i=x_i\right)=\prod_{i=1}^{n}e^{-\lambda}\frac{\lambda^{x_i}}{x_i!}=e^{-n\lambda}\frac{\lambda^{\sum_{i=1}^{n}x_i}}{\prod_{i=1}^{n}x_i!}$$

$$\ln\left(L(\lambda)\right)=-n\lambda+\sum_{i=1}^{n}x_i\ln(\lambda)-\ln\left(\prod_{i=1}^{n}x_i!\right)$$

ולכן אם נגזור את פונקצית הנראות נקבל כי

$$\frac{d}{d\lambda}\ln\left(L(\lambda)\right)=0=-n+\frac{1}{\lambda}\sum_{i=1}^{n}x_i\Rightarrow\hat{\lambda}=\frac{1}{n}\sum_{i=1}^{n}X_i.$$

ומבחינת הטייה נוכל לראות בקלות כי

$$E\left(\hat{\lambda}\right)=E\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\frac{1}{n}\sum_{i=1}^{n}E\left(X_i\right)=\frac{1}{n}\sum_{i=1}^{n}\lambda=\frac{1}{n}n\lambda=\lambda$$

על כן זהו אומד חסר הטייה.

ד. נתחיל בבניית פונקצית הנראות עבור מדגם $X_1,\ldots,X_n\sim N\left(\mu,\sigma^2\right)$.

$$L(\mu,\sigma)=L\left(\mu,\sigma,X_1,\ldots,X_n\right)=\prod_{i=1}^{n}f_{X_i}\left(x_i\right)=\prod_{i=1}^{n}\tfrac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{\left(x_i-\mu\right)^2}{2\sigma^2}\right\}=$$

$$=\left(2\pi\sigma^2\right)^{-n/2}\exp\left\{-\tfrac{1}{2\sigma^2}\sum_{i=1}^{n}\left(x_i-\mu\right)^2\right\}$$

כעת נוכל לקחת לוגריתם לפונקציה הנ"ל ולגזור אותה ולקבל אומדן עבור התוחלת והשונות.

$$\ln\left(L(\mu,\sigma)\right)=-\tfrac{n}{2}\left[\ln(2\pi)+\ln\left(\sigma^2\right)\right]-\tfrac{1}{2\sigma^2}\sum_{i=1}^{n}\left(x_i-\mu\right)^2$$

$$\frac{\partial}{\partial\mu}\ln\left(L(\mu,\sigma)\right)=0=\tfrac{1}{\sigma^2}\sum_{i=1}^{n}\left(x_i-\mu\right)\quad\Rightarrow\quad\hat{\mu}=\frac{1}{n}\sum_{i=1}^{n}x_i$$

$$\frac{\partial}{\partial\sigma^2}\ln\left(L(\mu,\sigma)\right)=0=-\tfrac{n}{2\sigma^2}+\tfrac{1}{2\sigma^4}\sum_{i=1}^{n}\left(x_i-\mu\right)^2\Rightarrow\sigma^2=\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\mu\right)^2$$

$$\Rightarrow\quad\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\hat{\mu}\right)^2=\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\overline{x_n}\right)^2$$

נשים לב כי קיבלנו אומד בלתי מוטה עבור התוחלת.

תרגיל 4 (אומד חסר הטייה): במפעל יש 2 מכונות המייצרות את אותו החלק. המכונות מייצרות את החלקים באותו הקוטר בממוצע אך מכונה א' (הישנה) מייצרת אותם עם סטיית תקן כפולה מזו של החדשה, מכונה ב'. הממוצע יסומן ב- μ וסטיית התקן של מכונה ב' תסומן ב- σ^2. בכדי לאמוד פרמטרים אלו לוקחים מדגמים, $Y_1,\ldots,Y_n,X_1,\ldots,X_m$ ממכונות א' ו-ב' בהתאמה באופן ב"ת.

א. מצאו את ערכי $\alpha\in[0,1]$ עבורם $T=\alpha\overline{X_m}+(1-\alpha)\overline{Y_n}$ הוא אומד בלתי מוטה?

ב. מצאו את הערך $\alpha\in[0,1]$ עבורו הנדק / ההפסד של T, $MSE(T)$, הוא מינימאלי ומצאו הפסד זה.

ג. חשבו במפורש את הערך של $\alpha\in[0,1]$ עבור המקרה הפרטי בו $m=n$. מהו T במקרה זה ומה ההפסד / הנזק שלו?

פתרון תרגיל 4:

א. בכדי שאומד T יהיה בלתי מוטה עליו לקיים $E[T]=\mu$ לכן נבדוק עבור המקרה הכללי את התוחלת של T ונקבל,

$$E(T)=E\left(\alpha\overline{X_m}+(1-\alpha)\overline{Y_n}\right)=\alpha E\left(\overline{X_m}\right)+(1-\alpha)E\left(\overline{Y_n}\right)=\alpha\mu+(1-\alpha)\mu=\mu$$

ואנו למעשה רואים כי עבור כל ערך של $0\le\alpha\le1$ האומד הנ"ל הוא אומד חסר הטייה.

ב. נתחיל בחישוב ההפסד של האומד T. נתחיל מהגדרת ההפסד כתוחלת ריבוע ההפרש בין האומד לפרמטר הנאמד -

$$MSE(T)=E\left[(T-\mu)^2\right]=E\left[T^2-2T\mu+\mu^2\right]=$$
$$=E\left[T^2\right]-2\mu E[T]+\mu^2=E\left(T^2\right)-E(T)^2+E(T)^2-2\mu E[T]+\mu^2=$$
$$=Var(T)+\left(E(T)-\mu\right)^2$$

מאחר ו-T הוא אומד חסר הטייה עבור כל $0\le\alpha\le1$ אזי $MSE(T)=Var(T)$.

נחשב את השונות של T.

$$Var(T)=Var\left(\alpha\overline{X_m}+(1-\alpha)\overline{Y_n}\right)=\alpha^2Var\left(\overline{X_m}\right)+(1-\alpha)^2Var\left(\overline{Y_n}\right)=$$
$$=\alpha^2\tfrac{(2\sigma)^2}{m}+(1-\alpha)^2\tfrac{\sigma^2}{n}=\tfrac{\sigma^2}{mn}\left[4n\alpha^2+m(1-\alpha)^2\right]$$

את היבטוי שקיבלנו נרצה למזער, כאשר m ו-n קבועים ולכן נגזור ונשווה ל-0.

$$\tfrac{\sigma^2}{mn}\tfrac{d}{d\alpha}\left[4n\alpha^2+m(1-\alpha)^2\right]=\tfrac{\sigma^2}{mn}\left[8n\alpha-2m(1-\alpha)\right]=0$$
$$\Rightarrow\quad\alpha(8n+2m)-2m=0\quad\Rightarrow\quad\alpha=\frac{m}{4n+m}$$

נשים לב שהנגזרת השנייה היא חיובית ולכן זו נקודת מינימום. נציב את הערך שקיבלנו בפונקצית ההפסד ונקבל

$$\min_{0\le\alpha\le1} Var(T)=\tfrac{\sigma^2}{mn}\left[4n\left(\frac{m}{4n+m}\right)^2+m\left(1-\frac{m}{4n+m}\right)^2\right]=$$

$$=\tfrac{\sigma^2}{mn}\left[\frac{4nm^2+16mn^2}{(4n+m)^2}\right]=\sigma^2\left[\frac{4m+16n}{(4n+m)^2}\right]=\frac{4\sigma^2}{4n+m}$$

ג. במקרה בו יש שיוויון בין מספר הדגימות שנקלחו מכל מכונה נקבל ש- $\alpha=\frac{m}{4n+m}=\frac{1}{5}$, האומד הבלתי מוטה הטוב ביותר יהיה $T=\frac{1}{5}\overline{X_m}+\frac{4}{5}\overline{Y_n}$ והשונות שלו תהיה $\frac{4\sigma^2}{5m}$.

<u>תרגיל 5 (א.נ.מ)</u>: אלון רוצה לבדוק האם בן הוא נהג טוב ולשם כך הוא משתמש במודל הבא. זמן התגובה – הזמן שעובר (בשניות) עד שהנהג לוחץ על דוושת הבלם מהרגע שראה את אורות הבלם של הרכב שלפניו, מתפלג מעריכית $X\sim Exp(\theta)$ (הפרמטר תלוי באופי בנהג). נסמן ב- β את הסיכוי שזמן התגובה של נהג מסוים גבוה מ-2 שניות. נהג גרוע הוא זה שעבורו $\beta\ge\frac{1}{2}$. אלון דגם n פעמים את זמן התגובה של בן.

א. מצאו אומד נראות מקסימאלית עבור β. האם הוא בלתי מוטה?

ב. מצאו א.נ.מ עבור תוחלת זמן התגובה. האם הוא בלתי מוטה?

ג. מהם האומדנים המתקבלים אם ב-4 זמני התגובה שאלון מדד לבן הוא קיבל את התוצאות הבאות: 1.3, 0.6, 2.4, 0.5?

<u>פתרון תרגיל 5</u>:

א. יש בידי אלון n תצפיות של בן והוא יודע כי זמן התגובה מתפלג מעריכיות ולכן הוא יוכל לחשב א.נ.מ מקסימאלית לזמן התגובה של בן. ראינו בתרגילים קודמים כי א.נ.מ למקרה זה הוא $\theta=\frac{n}{\sum_{i=1}^{n}X_i}=\frac{1}{\overline{X_n}}$, ז"א, ממוצע הזמנים בחזקת 1-. נשים לב כי $\beta=P(X>2)$ ולכן $\beta=1-P(X\le2)=1-\left(1-e^{-2\theta}\right)=e^{-2\theta}$. כעת, במקום לחשב מההגדרה א.נ.מ עבור β נוכל להיעזר במשפט שאומר שאם θ הוא א.נ.מ עבור θ אז $g(\theta)$ הוא א.נ.מ עבור $g(\theta)$. על כן, $\beta=e^{-2\theta}=e^{-\frac{2}{\overline{X_n}}}$ ונשים לב שאומד זה הוא אכן אומד מוטה (זו לחלוטין לא פונקציה לינארית ולכן התוחלת לא תתקבל בהתאם).

ב. נמצא תחילה את הביטוח עבור תוחלת זמן התגובה. $\mu = E\left(X_i\right) = E\left(\overline{X_n}\right) = \frac{1}{\theta}$

ולכן $\mu = \frac{1}{\theta} = \overline{X_n}$ וזהו כמובן אומד בלתי מוטה מאחר ו-$\mu = E\left(\overline{X_n}\right) = E\left(\mu\right)$.

ג. על פי הנתונים שהתקבלו, ממוצע המדגם הוא $\overline{X_4} = 1.2$ ולכן $\beta = e^{-\frac{2}{1.2}} = 0.1889$

ולכן קיבלנו כי בן הוא נהג טוב.

<u>תרגיל 6 (א.נ.מ במדגם קטן)</u>: בן קונה לאלון חבילת זיקוקים ליום הולדתו. הזיקוקים מיוצרים במפעל האורז אותם בחבילות של 4 כאשר חלקם אדומים וחלקם ירוקים. לאחר שאלון הוציא 2 זיקוקים והדליק אותם, הוא מעוניין לאמוד את מספר הזיקוקים האדומים θ בחבילה כולה.

נגדיר את המשתנים המקריים המציינים X_1, X_2 באופן הבא:

$X_i = \begin{cases} 1 & the\ i^{th}\ firework\ was\ red \\ 0 & otherwise \end{cases}$. מצאו אומד נראות מקסימאלית עבור כל מדגם אפשרי שיכול היה אלון לקבל.

<u>פתרון תרגיל 6</u>:

במקרה זה מדובר במספר מצומצם של מקרים אפשריים $\theta \in \{0,1,2,3,4\}$ ולכן אין טעם לבצע את החישובים המסורבלים שעשינו קודם לכן. נתחיל בבניית טבלה התפלגויות עבור כל מדגם אפשרי:

$(x_1, x_2) \setminus \theta$	0	1	2	3	4
$(0,0)$	1	$\frac{1}{2}$	$\frac{1}{6}$	0	0
$(1,0)$	0	$\frac{1}{4}$	$\frac{1}{3}$	$\frac{1}{4}$	0
$(0,1)$	0	$\frac{1}{4}$	$\frac{1}{3}$	$\frac{1}{4}$	0
$(1,1)$	0	0	$\frac{1}{6}$	$\frac{1}{2}$	1

נחשב עבור כל מקרה בפני עצמו:

$$P_{\theta=1}\left(x_1 = 0, x_2 = 0\right) = P_{\theta=1}\left(x_1 = 0 \mid x_2 = 0\right) P_{\theta=1}\left(x_2 = 0\right) = \tfrac{3}{4}\tfrac{2}{3} = \tfrac{1}{2}$$
$$P_{\theta=1}\left(x_1 = 0, x_2 = 1\right) = P_{\theta=1}\left(x_1 = 0 \mid x_2 = 1\right) P_{\theta=1}\left(x_2 = 1\right) = \tfrac{1}{4}\tfrac{3}{3} = \tfrac{1}{4}$$

וכן הלאה עבור כל המקרים האחרים. כעת נוכל פשוט לבחון את המקרים השונים עבור כל מדגם ולבחור את ההסתברות המקסימאלית.

אם המדגם היה $(0,0)$ אז האומדן לפי שיטת הנראות המקסימאלית הוא 0.

אם המדגם היה $(1,0)$ או $(0,1)$ אז האומדן לפי שיטת הנראות המקסימאלית הוא 2.

אם המדגם היה $(1,1)$ אז האומדן לפי שיטת הנראות המקסימאלית הוא 4.

<u>תרגיל 7 (רווח סמך להתפלגות בינומית)</u>: מפלגת "אחינו, יהיה בסדר" שזה עתה הקומה מעוניינת לאמוד את אחוז התמיכה בה בציבור.

א. מה גודל המדגם שעל המפלגה לבחור בכדי שבסקר שיתקבל היא תדע את אחוז התמיכה בה בסדר גודל של עד 2% טעות, עבור רמת סמך של 90%?

ב. כיצד ניתן להקטין את הוצאות הסקר אם ידוע שאין סיכוי לתמיכה של יותר מ-7% מהציבור?

ג. בהינתן שבמדגם שמצאתם בסעיף ב' נמצאו 4.5% התומכים במפלגה, תנו רווח סמך 90% עבור פרופורציית התומכים בה באוכלוסיה כולה.

<u>פתרון תרגיל 7</u>:

א. נשים לב כי אנו דנים במודל בינומי בו כל אדם יכול לתמוך או לא לתמוך במפלגה.

נניח כי $X_i \sim B(p)$ זה מ"מ ברנולי המעריך את תמיכת אדם I במפלגה ובהתאם

$X = \sum_{i=1}^{n} X_i \sim Bin(n,p)$ ו- $p = \overline{X_n} = \frac{X}{n}$ זה א.נ.מ לאחוז התמיכה במפלגה

(פרופורציית התומכים במפלגה בכלל האוכלוסיה). הגדרת רווח הסמך ברמת סמך 0.9 היא הקטע המקיים

$$P\left(-z_{1-\alpha/2} \le \frac{\overline{X_n}-\mu}{\sigma/\sqrt{n}} \le z_{1-\alpha/2}\right) = P\left(-z_{1-0.05} \le \frac{\overline{X_n}-p}{\sigma/\sqrt{n}} \le z_{1-0.05}\right) \cong 0.9 = 1-0.1$$

$$\Rightarrow \overline{X_n} - z_{0.95}\sqrt{\frac{p(1-p)}{n}} \le p \le \overline{X_n} + \sqrt{\frac{p(1-p)}{n}} z_{0.95}$$

מאחר ו-p אינו ידוע נוכל להשתמש באומד שיש לנו ברווח סמך עצמו ולקבל

$$\left[\overline{X_n} - \sqrt{\frac{p(1-p)}{n}} z_{0.95}, \overline{X_n} + \sqrt{\frac{p(1-p)}{n}} z_{0.95}\right] = \left[\overline{X_n} - \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} z_{0.95}, \overline{X_n} + \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} z_{0.95}\right] =$$

$$= \left[\overline{X_n} - \sqrt{\frac{\overline{X_n}(1-\overline{X_n})}{n}} z_{0.95}, \overline{X_n} + \sqrt{\frac{\overline{X_n}(1-\overline{X_n})}{n}} z_{0.95}\right]$$

אורך רווח הסמך הוא

$$l=2\sqrt{\frac{\overline{X_n}\left(1-\overline{X_n}\right)}{n}}\cdot 1.646 \Rightarrow n=4\frac{\overline{X_n}\left(1-\overline{X_n}\right)}{l^2}\cdot\left(1.646\right)^2$$

אנו מעוניינים שחצי אורך הקטע יהיה קטן מ-0.02 ולכן

$$n\geq\frac{\overline{X_n}\left(1-\overline{X_n}\right)}{0.02^2}\cdot\left(1.646\right)^2$$

נשים לב כי הפונקציה $\overline{X_n}\left(1-\overline{X_n}\right)$ (פולינום ממעלה שנייה) חסומה על ידי 0.25 ונקבל

$$n\geq\frac{1}{4\cdot 0.02^2}\cdot\left(1.646\right)^2\cong 1693$$

ב. נתון כי $0.07\geq\overline{X_n}$ ולכן

$$n\geq\frac{0.07(1-0.07)}{0.02^2}\cdot\left(1.646\right)^2\geq\frac{\overline{X_n}\left(1-\overline{X_n}\right)}{0.02^2}\cdot\left(1.646\right)^2$$
$$\Rightarrow n\geq 440.4$$

כלומר, יש צורך רק ב-441 אנשים בכדי לאמוד את התמיכה במפלגה.

ג. בהינתן שבמדגם שמצאתם בסעיף ב' נמצאו 4.5% התומכים במפלגה, תנו רווח סמך 90% עבור פרופורציית התומכים בה באוכלוסיה כולה. הרווח סמך לפי תמיכה בשיעור של 4.5% מתוך 441 אנשים הוא

$$\left[\overline{X}_n-\sqrt{\frac{p(1-p)}{n}}z_{0.95},\overline{X}_n+\sqrt{\frac{p(1-p)}{n}}z_{0.95}\right]=\left[\overline{X}_n-\sqrt{\frac{\overline{X_n}\left(1-\overline{X_n}\right)}{n}}z_{0.95},\overline{X}_n+\sqrt{\frac{\overline{X_n}\left(1-\overline{X_n}\right)}{n}}z_{0.95}\right]=$$
$$=\left[0.045-\sqrt{\frac{0.045(1-0.045)}{441}}1.645, 0.045+\sqrt{\frac{0.045(1-0.045)}{441}}1.645\right]=\left[0.0288, 0.0612\right]$$

ז"א, אחוזי התמיכה במפלגה ינועו בין 3% ל-6%.

פרק 12

נושאים:

- בדיקת השערות.
- רמת מובהקות של מבחן ועוצמה של מבחן
- מובהקות תוצאה (P-value).
- הסקה במודל הנורמלי עם שונות ידועה.
- מבחני עוצמה מקסימאלית

בדיקת השערות

נניח כי קיבלנו תצפיות מסויימות מתוך מדגם של האוכלוסיה הנבדקת. כעת יש לנו צמד השערות הנוגעות לאופי האוכלוסיה, השערת האפס H_0 וההשערה האלטרנטיבית H_1.

השערת האפס היא לרוב ההנחה "השמרנית", ההנחה שמבוססת על המידע שיש עד כה וניסוי שיאשר אותה רק יישמר מצב קיים. לעומת זאת, ההשערה האלטרנטיבית היא השערה אשר מהווה את החידוש במחקר, וזו ההשערה אותה החוקר רוצה, למעשה, לאשר (ובכך לשנות את המצב הקיים).

את השערות אלו אנו נוכל לדחות או לקבל על בסיס המדגם הנתון ולכן חשוב לנסח מבחן כלשהו לבחינת ההשערות (*מבחן* – כלל החלטה הפותר את בעיית בדיקת השערות בהינתן מדגם). כיצד נעשה זאת?

ניקח את המדגם שקיבלנו ונחלק את התוצאות שלנו לאזור דחייה ואזור קבלה (ביחס להשערת האפס). נהוג לסמן את אזור הדחיה ב- C ואת אזור הקבלה ב- $\overline{C}$.

לאחר שביצענו זאת, אנחנו מגיעים לשלב הטעויות. ישנן 2 טעויות אפשריות:

1. טעות מן הסוג הראשון – דחיית השערת האפס שלא בצדק (קרי, השערת האפס נכונה) בהתאם לתוצאות המדגם ואזור הדחייה. ההסתברות לכך תסומן ב- $P_{H_0}(C)$.
2. טעות מן הסוג השני – קבלת השערת האפס בהתאם למדגם למרות שהיא איננה נכונה (אבסולוטית). ההסתברות לכך תסומן ב- $P_{H_1}(\overline{C})$.

המצב האמיתי - אבסולוטית	ההחלטה בעקבות המחקר: דחיית השערת האפס	ההחלטה בעקבות המחקר: קבלת השערת האפס
H_0 נכונה	טעות מסוג I	החלטה נכונה
H_1 נכונה	החלטה נכונה	טעות מסוג II

לרוב, טעויות מהסוג הראשון נחשבות לבעיתיות / מסוכנות יותר מאחר והן הביאו לשינוי מצב קיים ולכן הנזק שנגרם הינו ממשי (השקעת משאבים לשווא וכו'). לעומת זאת, טעויות מהסוג השני גורמות לנזק, אך נזק זה הינו תיאורטי מאחר ולא בוצע שינוי במערכת ולכן ההפסד הוא במידת אי-היעילות שלאורה מתנהלים.

כמובן שניתן לטעון כי הנזק התיאורטי הוא משמעותי יותר מאשר הנזק הממשי, אך בנקודה זו אנו נכנסים לתחום הפסיכולוגיה / כלכלה התנהגותית (ובתחומים הללו רואים פעמים רבות כי הפסד ממשי הוא קשה יותר לאנשים מאשר הפסד תיאורטי).

הערה – לרוב נבחר את השערת האפס בתור ההשערה אשר הנזק שנגרם מדחייתה בטעות או החמור ביותר.

טעויות מן הסוג הראשון נסמן ב- $\alpha = P_{H_0}\left(C\right)$ וטעויות מן הסוג השני נסמן ב- $\beta = P_{H_1}\left(\overline{C}\right)$.

רמת מובהקות ועוצמה של מבחן

הגדרה: **רמת המובהקות של מבחן** היא מספר α כך ש- $P_{H_0}\left(C\right) < \alpha$. ז"א, זה מספר שאנחנו קובעים כך שההסתברות שנבצע טעות מסוג I, דחיית השערת אפס שלא בצדק, היא קטנה מ- α.

הסיבה למינימיזציה זאת היא שלא ניתן לשלוט בצורה מוחלטת בו-זמנית על ההסתברות לצמד הטעויות ויכול להיות שמזעור אחת תוביל למצב מאוד בעייתי בטעות השנייה ולכן אנחנו קובעים באופן ברור כי אנחנו מעוניינים תחילה למזער טעויות מן הסוג הראשון.

כיצד מבצעים זאת? בכדי למנוע טעות מן הסוג הראשון צריך למזער את אזור הדחייה בכדי להגיע לרמת מובהקות רצויה.

מונח נוסף שמגדירים הוא העוצמה של המבחן. ***עוצמה של מבחן*** היא ההסתברות לדחיית השערת האפס כאשר, למעשה, ההשערה האלטרנטיבית היא הנכונה $\pi = P_{H_1}\left(C\right)$.

<u>תרגיל 1 (בדיקת השערות במודל בינומי)</u>: בן הוא שופט כדורגל שרוצה לבדוק האם המטבע שלו הוגן על בסיס 100 הטלות.

א. רשמו את המודל ואת ההשערות הנבדקות.
ב. חשבו את ההסתברות לטעות מסוג I כאשר השופט דוחה את המטבע כ-'לא מאוזן' אם יקבל פחות מ-40 או יותר מ-60 'עצים'.
ג. מהי ההסתברות לטעות מסוג II ומהי העוצמה של המבחן שהוצע בחלק ב', אם למעשה המטבע נופל על 'עץ' בהסתברות: 0.35, 0.45, 0.55, 0.65. הסבירו!

ד. הציעו מבחן בו רמת המובהקות לא תעלה על 0.01.

<u>פתרון תרגיל 1</u>:

א. נסמן ב- p את ההסתברות ל-'עץ' במטבע הנבדק וב-X את מספר ה-'עצים' ב-100 הטלות. על כן, $X \sim Bin(100,p)$. ההשערות הנבדקות הן:

$$H_0 : p = \frac{1}{2}; H_1 : p \neq \frac{1}{2}$$

קרי, השערת האפס היא שהמטבע הוגן וההשערה האלטרנטיבית שוללת זאת.

ב. נרשום תחילה את אזור הדחייה - $C = \{X < 40\} \cup \{X > 60\}$ ואת אזור הקבלה - $\overline{C} = \{40 \le X \le 60\}$.

ההסתברות לטעות מסוג I היא –

$$P_{H_0}(C) = P_{p=\frac{1}{2}}(X < 40) + P_{p=\frac{1}{2}}(X > 60) = 2P_{p=\frac{1}{2}}(X \le 39) =$$

$$\cong 2\Phi\left(\frac{39+0.5-50}{\sqrt{100\cdot\frac{1}{2}\left(1-\frac{1}{2}\right)}}\right) = 0.0358$$

ג. כמובן שההסתברות לטעות מסוג II הניתנת על ידי $P_{H_1}(\overline{C})$ (הסתברות לקבלת השערת האפס למרות שאיננה נכונה) תלויה במודל האלטרנטיבי ובהסתברות שנבחר לו. נבצע את החישוב הכללי ולאחר מכן נציב את הערכים:

$$P_{H_1}(\overline{C}) = P_{p=p_1}(\overline{C}) = P_{p=p_1}(40 \le X \le 60) =$$

$$= P_{p=p_1}(X \le 60) - P_{p=p_1}(X \le 39) =$$

$$= \Phi\left(\frac{60+0.5-100p_1}{\sqrt{100\cdot p_1(1-p_1)}}\right) - \Phi\left(\frac{39+0.5-100p_1}{\sqrt{100\cdot p_1(1-p_1)}}\right)$$

נזכיר שהעוצמה היא ההסתברות $P_{H_1}(C)$ ולכן שווה ל- $P_{H_1}(C) = 1 - P_{H_1}(\overline{C})$.

כעת כל שנותר זה להציב את ערכי ההסתברות האלטרנטיביים ולקבל:

הסתברות	$P_{H_1}(\overline{C})$ הסתברות לטעות מסוג II	$P_{H_1}(C)$ העוצמה של המבחן
0.35	0.1736	0.8264
0.45	0.8656	0.1344
0.55	0.8656	0.1344
0.65	0.1736	0.8264

כאשר צמד השורות התחתונות יכולות להתקבל משיקולי סימטריה לאחר חישוב צמד הראשונות. קיבלנו כי אם המטבע מאוד לא מאוזן (הסתברות 0.35) אז העוצמה של המבחן די גבוהה ונגלה זאת בהסתברות 0.82 בערך, לעומת זאת במקרים בהם המטבע לא מאוזנת במידה מועטה, אזי ישנה הסתברות גבוהה שלא נזהה זאת.

ד. נגדיר מבחן חדש בתור $C'=\{c_1 \le X \le c_2\}$ (ז"א שאזור הדחייה הוא $\overline{C'}$) ונדרוש שהמובהקות של מבחן תהיה 0.01, קרי שההסתברות לטעות מהסוג הראשון לא תעלה על 0.01 (נזכיר שההסתברות הזאת מחושבת תחת השערת האפס).

$$P_{p=\frac{1}{2}}\left(\overline{C'}\right)=2P_{p=\frac{1}{2}}\left(X<c_1\right)\cong 2\Phi\left(\tfrac{c_1+0.5-50}{\sqrt{25}}\right)\le 0.01$$
$$\Leftrightarrow\ \Phi\left(\tfrac{c_1+0.5-50}{\sqrt{25}}\right)\le 0.005\ \Leftrightarrow\ \tfrac{c_1+0.5-50}{\sqrt{25}}\le -2.576$$
$$\Leftrightarrow\ c_1\le 37.12$$

מובהקות התוצאה (P-value)

מובהקות התוצאה היא ההסתברות לקבלת תוצאה קיצונית לפחות כמו התוצאה שהתקבלה בניסוי, בהנחה ש- H_0 נכונה.

כיצד זה עוזר לנו? במקום לחשב מראש את הערך הקריטי של מבחן ברמת מובהקות α ניתן לבחון את השערת האפס תוך שימוש במובהקות P של התוצאה שהתקבלה, באופן הבא:

אם $P\le\alpha$ דוחים את השערת האפס ואם $P>\alpha$ לא דוחים אותה. נהוג לקבוע את רמת המובהקות ב-5%.

<u>הערה חשובה!</u> שימו לב שמה שאנו עושים על בסיס מובהקות התוצאה זה לפסול או לא לפסול את H_0. אם המובהקות של התוצאה גדולה מ-5% אנחנו בסה"כ לא יכולים לפסול את השערת האפס, אך זה לא אומר שהיא נכונה ולכן חשוב לנסח היטב את ההשערות כך שנקבל את התוצאה הרצויה מכך שההשערה לא נפסלה!

<u>תרגיל 2 (בדיקת השערות ומובהקות תוצאה במודל פואסוני)</u>: אלון היה רגיל לקבל כארבע הזמנות טלפוניות מנערות לבילוי משותף בליל שבת. לאחר שהוא הסתפר וצבע את שערו לירוק, התחיל, ככל הנראה, זרם ההזמנות לרדת.

א. רשמו את המודל ואת הבעיה של אלון כבעיית השערות.

ב. האם ניתן לומר שהירידה היא משמעותית, אם בשבוע הראשון לאחר התספורת החדשה אלון לא קיבל אפילו הזמנה אחת? (הניחו רמת מובהקות של $\alpha=0.05$).

ג. האם ניתן לומר שהירידה היא משמעותית אם במשך 5 שבועות רצופים לאחר התספורת הוא קיבל סה"כ 9 הזמנות (הניחו אותה רמת מובהקות כמו קודם לכן).

ד. האם ניתן להסיק מסעיף ג', ללא חישוב נוסף, גם מהי המסקנה עבור רמת מובהקות 0.1? 0.01?

<u>פתרון תרגיל 2</u>:

א. נניח שמספר ההזמנות של אלון לבילוי הן משתנה מקרי פואסוני עם פרמטר 4, $P(\lambda)$. ההשערות הנבדקות הן:

$$H_0 : \lambda = 4; H_1 : \lambda < 4$$

כאשר השערת האפס של אלון היא שהתספורת שלו אטרקטיבית ולעומת זאת ההשערה השנייה היא שהוריו צודקים ואסור היה עליו להסתפר כך.

ב. הירידה היא משמעותית אם ורק אם השערת האפס תוכח כלא נכונה. בכדי לעשות זאת עלינו למצוא מבחן / אזור דחייה המקיים $C = \{X \le c\}$ עבור ערך כלשהו שיקיים רמת מובהקות $\alpha = 0.05$ כאשר X הוא המדגם שלנו (תצפית בודדת).

נפתור זאת ב-2 דרכים:

1) נחפש ערך קריטי שיקיים רמת מובהקות מתאימה –

$$P_{H_0}\left(X \le c\right) = P_{\lambda=4}\left(X \le c\right) = \alpha = 0.05 \Rightarrow$$
$$P_{\lambda=4}\left(X \le 0\right) = e^{-4} = 0.0183\,;\, P_{\lambda=4}\left(X \le 1\right) = e^{-4} \cdot 4 = 0.0733$$

ולכן ברמת מובהקות זו נאלץ לקבוע את אזור הדחייה כ- $C = \{X \le 0\}$ כי הוספת הערך 1 תביא לכך שההסתברות לדחייה היא גבוה מידי ביחס ל-0.05. נשים לב שתוצאה שקיבלנו היא באזור הדחייה (אף הזמנה) ולכן השערת האפס נדחת.

2) נחשב את מובהקות התוצאה שקיבלנו –

$$P = P_{H_0}\left(X \le 0\right) = P_{\lambda=4}\left(X \le 0\right) = e^{-4} = 0.0183 < 0.05$$

וזו תוצאה מובהקת על כן עלינו לדחות את השערת האפס.

ג. כעת יש ברשותנו 5 תצפיות ולכן נשתמש בקירוב נורמלי עבור הסכום שלהן. גם את סעיף זה נפתור בשתי דרכים:

1) נחפש ערך קריטי שיקיים רמת מובהקות מתאימה –

$$P_{\lambda=4}\left(\sum_{i=1}^{5} X_i \le c\right) = 0.05 \Rightarrow P\left(\frac{\sum_{i=1}^{5} X_i - 4 \cdot 5}{\sqrt{5} \cdot \sqrt{4}} \le \frac{c + 0.5 - 4 \cdot 5}{\sqrt{5} \cdot \sqrt{4}}\right) = \Phi\left(\frac{c - 19.5}{\sqrt{20}}\right) = 0.05$$

$$\Rightarrow 1 - \Phi\left(\frac{c - 19.5}{\sqrt{20}}\right) = \Phi\left(\frac{19.5 - c}{\sqrt{20}}\right) = 0.95 \Rightarrow \frac{19.5 - c}{\sqrt{20}} = 1.645 \Rightarrow C = 12.14334$$

מאחר ואלון קיבל פחות מערך זה (9 הזמנות), עליו לדחות את השערת האפס.

2) נחשב את מובהקות התוצאה שקיבלנו –

$$P = P_{\lambda=4}\left(\sum_{i=1}^{5} X_i \le 9\right) = P\left(\frac{\sum_{i=1}^{5} X_i - 4\cdot 5}{\sqrt{5}\cdot\sqrt{4}} \le \frac{9.5-4\cdot 5}{\sqrt{5}\cdot\sqrt{4}}\right) = \Phi\left(\frac{9.5-20}{\sqrt{20}}\right) =$$

$$= \Phi(-2.35) = 0.0094 < 0.05$$

וזו תוצאה מובהקת על כן עלינו לדחות את השערת האפס.

ד. קל להסיק עבור רמת מובהקות 0.1 מאחר וזה מגדל את אזור הדחייה ולכן בכל מקרה השערת האפס תידחה. אם נדרוש רמת מובהקות 0.01 אזי בשיטה הראשונה נצטרך לחשב מחדש את הערך הקריטי ולראות מה מתקבל, לעומת זאת מהתוצאה של מובהקות התוצאה בשיטה השנייה בסעיף ג' רור כי 0.0094 קטן מ-0.01 ולכן גם ברמת מובהקות זו השערת האפס תידחה.

בסעיף האחרון ראינו נקודה חשובה והיא – חישוב מובהקות התוצאה מסייע לנו עבור מדגם נתון להעיד על טיב התוצאה שבמדגם ללא חישוב נוסף. אם אנו מציינים את ה- Pvalue אזי ניתן כבר להסיק מיד שעבור רמות מובהקות יותר קיצוניות, התוצאה שלנו תישמר.

<u>תרגיל 3 (מובהקות תוצאה וגודל מדגם במודל בינומי)</u>: בעיתון פורסם כי על פי סקר שנערך בציבור הסטודנטים מסתבר ש"הרוב של הסטודנטים לא מרוצה מכך שלומדים סטטיסטיקה עם הסתברות". על פי נתוני הסקר, 52% מהנשאלים ענו שאינם מרוצים.

א. רשמו את המודל ואת ההשערות הנבדקות.

ב. בעיתון לא פורסם מה היה גודל המדגם. הסיקו אם התוצאה אומנם מובהקת (קרי, אם ניתן לומר שיש באוכלוסיה כולה רוב לדעה השלילית) אם ההשערה נבדקה ברמת מובהקות של 0.05 ובמדגם השתתפו 500 איש. פתרו גם עבור מדגם של 1000 איש.

ג. איזה מדגם דרוש כדי שאם באוכלוסיה אכן ישנם 52% לא מרוצים, נצליח לגלות זאת בהסתברות 0.9 לפחות?

<u>פתרון תרגיל 3</u>:

א. נסמן ב-X את מספר הסטודנטים הלא מרוצים. $X \sim Bin(n, p)$. ההשערות הן:

$$H_0 : p \le \frac{1}{2}; H_1 : p > \frac{1}{2}$$

קרי, השערת האפס היא שאין רוב ללא מרוצים וההשערה האלטרנטיבית אומרת שיש רוב.

הסיבה לבחירת ההשערות בצורה הזו היא שהדבר שאנו עושים בעזרת מובהקות התוצאה היא לפסול את השערת האפס ולכן אם לא נצליח לפסול את השערת האפס הנ"ל אז נוכל להגיד שהקביעה *"הרוב של הסטודנטים לא מרוצה"* איננה מדויקת, מאחר ולהשערת האפס ישנה מובהקות גבוהה.

ב. אם המדגם מכיל 500 סטודנטים אזי 260 ענו שאינם מרוצים ותחת המודל של $p=0.5$ מתקיים ש- $E(X)=250, V(X)=125$. נחשב את מובהקות התוצאה שהתקבלה (נשים לב שהמדגם גדול דיו בכדי לבצע קירוב נורמאלי)

$$P=P_{p=\frac{1}{2}}(X\ge 260)\cong 1-\Phi\left(\tfrac{259.5-250}{\sqrt{125}}\right)=1-\Phi(0.85)=0.1977$$

וקיבלנו למעשה שישנה מובהקות מאוד גבוהה (בוודאי מעל 5%) ולכן השערת האפס נשארת (אין רוב ללא מרוצים).

<u>הערה חשובה</u>!!! למה בחרנו את 0.5 בתור ההסתברות? ובכן, מסתבר שכאשר יש לנו צמד השערות המורכבות מתחומים (ז"א שהשערת האפס הינה ערך מוגדר, אלא טווח וכנ"ל עבור ההשערה האלטרנטיבית), אזי המבחנים הם בעלי עוצמה מקסימאלית ברמת מובהקות מסוימת בערך הגבולי שבין ההשערות).

עבור מדגם של 1000 איש נקבל

$$P=P_{p=\frac{1}{2}}(X\ge 520)\cong 1-\Phi(1.23)=0.1093$$

ולכן התוצאה שקיבלנו קודם לכן עומדת בעינה.

ג. הדרישה היא שאם $p=0.52$ עוצמת המבחן תהיה לפחות 0.9 ובמדוייק זה אומר ש- $P_{H_1}(C)\ge 0.9$. מה זה C? זהו אזור הדחייה של מבחן עם רמת מובהקות $\alpha=0.05$ שזה בקירוב $C=\left\{\frac{X-n\cdot p_0}{\sqrt{n\cdot p_0(1-p_0)}}>z_{1-\alpha}\right\}$.

$$P_{H_1}(C)=P_{0.52=p>\frac{1}{2}}(C)=P_{p=0.52}\left(\frac{X-n\cdot p_0}{\sqrt{n\cdot p_0(1-p_0)}}>z_{1-\alpha}\right)=$$
$$=P_{p=0.52}\left(\frac{X-n\cdot 0.5}{\sqrt{n\cdot 0.25}}>z_{1-\alpha}\right)\ge 0.9$$

נפתח את הביטוי האחרון שקיבלנו בכדי שנכל לפתור אותו מפורשות, נגדיר $p_1:=p=0.52$ ונקבל

$$P_{p_1=0.52}\left(\frac{X-n\cdot 0.5}{\sqrt{n\cdot 0.25}}>z_{1-\alpha}\right)=P_{p_1}\left(X>z_{1-\alpha}\cdot\sqrt{n\cdot 0.25}+n\cdot 0.5\right)=$$
$$=P_{p_1}\left(\frac{X-np_1}{\sqrt{np_1(1-p_1)}}>\frac{z_{1-\alpha}\cdot\sqrt{n\cdot 0.25}+n\cdot 0.5-np_1}{\sqrt{np_1(1-p_1)}}\right)=$$
$$\cong P_{p_1}\left(Z>\frac{z_{1-\alpha}\cdot\sqrt{n\cdot 0.25}+n\cdot 0.5-np_1}{\sqrt{np_1(1-p_1)}}\right)=$$
$$=1-\Phi\left(\frac{z_{1-\alpha}\cdot\sqrt{n\cdot 0.25}+n\cdot 0.5-np_1}{\sqrt{np_1(1-p_1)}}\right)\geq 0.9$$

ולבסוף נקבל ש-

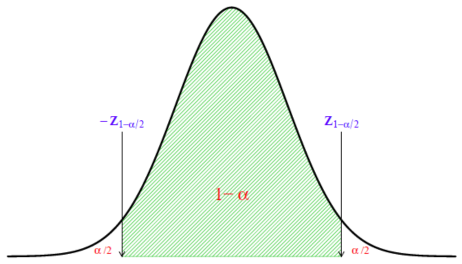
$$P_{H_1}(C)\geq 0.9 \quad\Leftrightarrow\quad 0.1\geq\Phi\left(\frac{z_{1-\alpha}\cdot\sqrt{n\cdot 0.25}+n\cdot 0.5-np_1}{\sqrt{np_1(1-p_1)}}\right)$$
$$\Leftrightarrow\quad z_{0.1}\geq\frac{z_{1-\alpha}\cdot\sqrt{n\cdot 0.25}+n\cdot 0.5-np_1}{\sqrt{np_1(1-p_1)}}=\frac{z_{1-\alpha}\cdot\sqrt{0.25}+\sqrt{n}\cdot(0.5-p_1)}{\sqrt{p_1(1-p_1)}}$$
$$\Leftrightarrow\quad z_{0.1}\sqrt{p_1(1-p_1)}\geq z_{0.95}\cdot 0.5-\sqrt{n}\cdot(p_1-0.5)$$
$$\Leftrightarrow\quad n\geq 5351$$

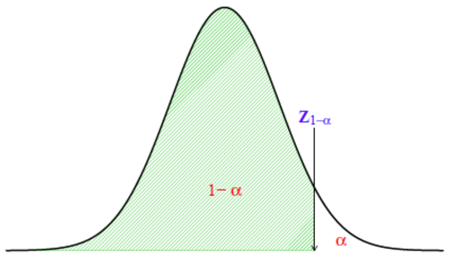
<u>הסקה במודל הנורמלי עם שונות ידועה</u>

יהיו $X_1,\dots,X_n$ תצפיות בלתי תלויות מהתפלגות נורמאלית $N(\mu,\sigma^2)$. בודקים את השערת האפס $H_0:\mu=\mu_0$. סטטיסטי המבחן הוא $Z_0=\frac{\overline{X}_n-\mu_0}{\sigma/\sqrt{n}}$ תחת השערת האפס, $Z_0\sim N(0,1)$ ונניח שבבדיקה התקבל ש- $Z_0=z_0$. המבחן המתאים ברמת מובהקות α הוא:

האלטרנטיבה H_1	אזור הדחייה	מובהקות התוצאה
$\mu<\mu_0$	$Z_0<-z_{1-\alpha}$	$P(Z\leq z_0)=\Phi(z_0)$
$\mu>\mu_0$	$Z_0>z_{1-\alpha}$	$P(Z\geq z_0)=1-\Phi(z_0)$
$\mu\neq\mu_0$	$\lvert Z_0\rvert>z_{1-\alpha/2}$	$P(\lvert Z\rvert\geq\lvert z_0\rvert)=2(1-\Phi(\lvert z_0\rvert))$

להלן שרטוטים של התפלגויות נורמאליות והאזורים המתאימים בתור אזורי דחייה. השרטוט הימני מציג מבחן דו-צדדי והשרטוט השמאלי מציג מבחן דחייה חד-צדדי ימני.





<u>תרגיל 4 (המודל הנורמאלי – חד צדדי עם עוצמה)</u>: משך זמן הייבוש בדקות של צבע מסויים מתפלג נורמאלית $N\left(90,10^2\right)$. בן מציע תוסף אשר ייקצר את משך זמן הייבוש. בכדי לבדוק זאת נלקחו 14 דגימות שנצבעו עם התוסף. הוחלט שאם משך הזמן הממוצע יהיה נמוך מ-85.6 דקות אז יוכנס השימוש בתוסף.

א. רשמו את המודל ואת ההשערות הנבדקות (הניחו שסטיית התקן נשמרת).
ב. מהי רמת המובהקות של המבחן שהוצע?
ג. אם אומנם התוסף מקצר את משך זמן הייבוש ב-4 דקות, מה ההסתברות שהדבר לא יתגלה בניסוי?
ד. כמה דגימות צריך בכדי שאם תוחלת זמן הייבוש אכן מתקצרת ב-4 דקות זה יתגלה בניסוי בהסתברות של 0.9 לפחות?

<u>פתרון תרגיל 4</u>:

א. הנחות המודל הן כי $X_i \sim N\left(\mu,\sigma^2\right), i=1,\dots,14$ כאשר כל גורם הוא דגימה בלתי תלויה. ההשערות הן:

$$H_0:\mu=90;\ H_1:\mu<90$$

ב. אזור הדחייה בו משתמשים בניסוי הוא $C=\left\{\overline{X_{14}}<85.6\right\}$, רמת המובהקות זו ההסתברות לטעות מסוג ראשון ולכן,

$$\alpha=P_{\mu=90}\left(\overline{X_{14}}<85.6\right)=\Phi\left(\tfrac{85.6-90}{10/\sqrt{14}}\right)=\Phi\left(-1.646\right)=0.05$$

לכן המבחן נערך ברמת מובהקות סטנדרטית של 0.05.

ג. מה שדורשים מאיתנו בסעיף זה, זה לחשב את ההסתברות לטעות מסוג II, שזה אומר שמשך הזמן אכן קוצר ב-4 דקות אך לא זיהינו זאת בניסוי.

$$\beta=P_{\mu=86}\left(\overline{X_{14}}\geq 85.6\right)=1-\Phi\left(\tfrac{85.6-86}{10/\sqrt{14}}\right)=0.5596$$

קרי, ההסתברות לטעות מסוג שני היא כמעט 56%!!! באותה מידה יכולנו לחשב את העוצמה של המבחן מאחר ו-

$$\beta = 1-\pi = 1-P_{\mu=86}\left(\overline{X_{14}} < 85.6\right) = 1-\Phi\left(\tfrac{85.6-86}{10/\sqrt{14}}\right) = 0.5596$$

ד. למעשה דורשים מאיתנו בשאלה שהעוצמה של המבחן תהיה לפחות 0.9. נרשום את הנתונים שברשותינו במסודר: $\pi > 0.9, \alpha = 0.05, \mu = 86$. נתחיל במציאת ביטוי כללי עבור העוצמה של המבחן

$$\pi = P_{\mu=86}\left(C\right) = P_{\mu=86}\left(\tfrac{\overline{X_n}-\mu_0}{\sigma/\sqrt{n}} < -z_{1-\alpha}\right) = P_{\mu=86}\left(\tfrac{\overline{X_n}-\mu_0}{\sigma/\sqrt{n}} < -z_{1-\alpha}\right) =$$

$$= P_{\mu=86}\left(\overline{X_n} < \mu_0 - \tfrac{\sigma}{\sqrt{n}} z_{1-\alpha}\right)$$

$$\Rightarrow \pi = P_{\mu=86}\left(\tfrac{\overline{X_n}-86}{\sigma/\sqrt{n}} < \tfrac{\mu_0 - \frac{\sigma}{\sqrt{n}} z_{1-\alpha} - 86}{\sigma/\sqrt{n}}\right) = P_{\mu=86}\left(\tfrac{\overline{X_n}-86}{10/\sqrt{n}} < \tfrac{90-86}{10/\sqrt{n}} - 1.646\right) =$$

$$= \Phi\left(\tfrac{2\sqrt{n}}{5} - 1.646\right) > 0.9 \Rightarrow \tfrac{2\sqrt{n}}{5} - 1.646 > 1.285 \Rightarrow n > 53.69$$

לכן נצטרך לפחות 54 דגימות בכדי לשמור על רמת מובהקות 0.05 ועוצמה של לפחות 0.9.

<u>תרגיל 5 (המודל הנורמאלי דו צדדי)</u>: במפעל מייצרים ברגים בקוטר μ מ"מ עם סטיית תקן של 0.5 מ"מ כאשר כל בוקר מכיילים את המכונה לערך μ הנדרש. בוקר אחד כיילו את המכונה ל-5 מ"מ וייצרו 100 ברגים ומדדו את הקוטר שלהם.

א. מה הכלל שתציע למנהל המפעל בכדי לבדוק ברמת מובהקות 0.05 שהמכונה כוילה כראוי?

ב. המדידה הניבה תוצאה ממוצעת של 4.901 מ"מ. מה תהיה ההחלטה ומה מובהקות התוצאה?

<u>פתרון תרגיל 5</u>:

א. נתחיל ברישום ההשערות:

$$H_0: \mu = 5; H_1: \mu \neq 5$$

כאשר אנו רוצים רמת מובהקות 0.05, אנו צריכים שאזור הדחייה יהיה דו-צדדי מהצורה

$$C = \left\{\left|\tfrac{\overline{X_n}-5}{0.5/\sqrt{100}}\right| > z_{1-0.05/2}\right\} = \left\{\left|\overline{X_n}-5\right| > \tfrac{1}{20} z_{0.975}\right\} = \left\{\left|\overline{X_n}-5\right| > \tfrac{1}{20} 1.96\right\}$$

$$\Rightarrow C = \left\{\left|\overline{X_n}-5\right| > 0.098\right\} = \left\{\overline{X_n} > 5.098\right\} \cup \left\{\overline{X_n} < 4.902\right\}$$

לכן הכלל יהיה – כל ממוצע מעל 5.098 או מתחת ל-4.902 יגרור דחייה של השערת האפס ויוביל למסקנה שהמכונה לא כוילה כראוי.

ב. מאחר והמדידה היא בתוך אזור הדחייה, אזי המשמעות היא שהמכונה לא כוילה כראוי. מובהקות התוצאה היא ההסתברות לקבל תוצאה קיצונית לפחות כמו שקיבלנו תחת השערת האפס ולכן

$$P = 2P_{\mu=5}\left(\overline{X_n} < 4.901\right) = 2P_{\mu=5}\left(\frac{\overline{X_n}-5}{0.5/\sqrt{100}} < \frac{4.901-5}{0.5/\sqrt{100}}\right) = 2\Phi\left(-1.98\right) =$$
$$= 2\left(1-\Phi\left(1.98\right)\right) = 2\left(1-0.9761\right) = 0.0478$$

מבחנים בעלי עוצמה מקסימאלית

הגדרה: ***מבחן בעל עוצמה מקסימאלית*** הוא מבחן עם העוצמה הגבוהה ביותר מבין כל המבחנים שהם בעלי רמת מובהקות α.

<u>תרגיל 6</u>: אלון ובן חזרו מהודו ומנסים לקבוע האם כעת יש בישראל קיץ או חורף. ההסתברות ליום גשום בקיץ היא 0.2 ובחורף היא 0.5.

א. אלון ספר את הימים הגשומים בפרק זמן של 3 ימים בעוד בן ספר את הימים עד ליום הגשם הראשון בפרק זמן של 3 ימים.רשמו את פונקצית הנראות עבור כל אחד מהם וכל השערה.

ב. נניח שלאחר 2 ימי שמש הגיע יום גשם. בדקו את ההשערה כי אנחנו בחורף לפי מבחן הנראות המקסימאלית של כל אחד מהטיילים.

ג. חזרו על הסעיף הקודם אם היום השני הוא היום היחיד בו היה גשם.

<u>פתרון תרגיל 6</u>:

א. נגדיר כי ההשערה שאנו נמצאים בתקופת קיץ היא H_0 ובתקופת חורף היא H_1.

כעת נעבור לחישוב פונקצית הנראות. אם X הוא מספר ימי הגשם במדגם אזי עבור אלון, שבחר במודל הבינומי, נקבל

$$P\left(X=0\,|\,H_0\right) = 0.8^3 = 0.512$$
$$P\left(X=1\,|\,H_0\right) = \binom{3}{1}0.8^2\cdot 0.2 = 0.384$$
$$P\left(X=2\,|\,H_0\right) = \binom{3}{2}0.8^1\cdot 0.2^2 = 0.096$$
$$P\left(X=3\,|\,H_0\right) = 0.2^3 = 0.008$$

$$P(X=0|H_1)=0.5^3=0.125$$

$$P(X=1|H_1)=\binom{3}{1}0.5^2\cdot 0.5=0.375$$

$$P(X=2|H_1)=\binom{3}{2}0.5^1\cdot 0.5^2=0.375$$

$$P(X=3|H_1)=0.5^3=0.125$$

וסיכום התוצאות בטבלה:

	X=0	X=1	X=2	X=3
קיץ	**0.512**	**0.384**	0.096	0.008
חורף	0.125	0.375	**0.375**	**0.125**

עבור בן, שבחר במודל הגיאומטרי נקבל כי –

$$P(Y=0|H_0)=0.2$$

$$P(Y=1|H_0)=0.8\cdot 0.2=0.16$$

$$P(Y=2|H_0)=0.8^2\cdot 0.2=0.128$$

$$P(Y\geq 3|H_0)=0.8^3=0.512$$

$$P(Y=0|H_1)=0.5=0.5$$

$$P(Y=1|H_1)=0.5\cdot 0.5=0.25$$

$$P(Y=2|H_1)=0.5^2\cdot 0.5=0.125$$

$$P(Y\geq 3|H_1)=0.5^3=0.125$$

וסיכום התוצאות בטבלה:

	Y=0	Y=1	Y=2	Y=3
קיץ	0.2	0.16	**0.128**	**0.512**
חורף	**0.5**	**0.25**	0.125	0.125

ב. אם נלך לפי אלון נקבל ש-

$$\frac{P(X=1|H_0)}{P(X=1|H_1)}=\frac{0.384}{0.375}>1$$

ולכן אלון יניח שאנחנו בקיץ.

אם נלך לפי בן נקבל ש-

$$\frac{P(Y=2|H_0)}{P(Y=2|H_1)}=\frac{0.128}{0.125}>1$$

וגם בן יגיד שאנחנו בקיץ.

ג. לעומת הסעיף הקודם, הפעם אנחנו מקבלים כי אלון קובע שאנחנו נמצאים בקיץ לעומת בן שחשוב שאנו בעונת החורף:

$$\frac{P(X=1|H_0)}{P(X=1|H_1)}=\frac{0.384}{0.375}>1>\frac{0.16}{0.25}=\frac{P(Y=1|H_0)}{P(Y=1|H_1)}$$

פרק 13

נושאים:

- בדיקת השערות על תוחלת עם שונות לא ידועה.
- התפלגות t-student
- רווח סמך עבור התוחלת עם שונות לא ידועה.

בדיקת השערות עם שונות לא ידועה

עד כה פתרנו בעיות בבדיקת השערות כאשר ההתפלגות הייתה נורמאלית והשונות הייתה ידועה, או לחילופין, כאשר המודל היה בינומי ו/או פואסוני וכו'. כאשר המודל לא היה נורמאלי, אזי יכולנו להשתמש באומדנים עבור השונות, אבל במקרה של התפלגות נורמאלית ישנה בעיה עקרונית והיא שלהתפלגות נורמאלית ישנם 2 פרמטרים בלתי תלויים (תוחלת ושונות). ההתפלגות יכולה להיות עם כל סטיית תקן יחד עם כל תוחלת לעומת המקרים בהם ישנה לדוגמא התפלגות ברנולי ואז התוחלת והשונות נגמרות בצורה יחידה עבור הפרמטר של ההתפלגות (כנ"ל עבור המקרה הפואסוני, המעריכי וכן הלאה).

מה עושים כאשר סטיית התקן לא ידועה?

ובכן, ישנה אומד פשוט עבור סטיית התקן (אותו מצאנו בפרק על אמידה נקודתית) והוא

$$S^2=\tfrac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\overline{X}_n\right)^2=\tfrac{1}{n-1}\sum_{i=1}^{n}\left(X_i^2-2X_i\overline{X}_n+\overline{X}_n^2\right)=\tfrac{1}{n-1}\sum_{i=1}^{n}\left(X_i^2+\overline{X}_n^2\right)-$$

$$-\tfrac{2}{(n-1)n}\sum_{i=1}^{n}X_i\sum_{j=1}^{n}X_j=\tfrac{1}{n-1}\sum_{i=1}^{n}\left(X_i^2+\overline{X}_n^2\right)-\tfrac{2}{(n-1)n}\overline{X}_n^2=\tfrac{1}{n-1}\sum_{i=1}^{n}\left(X_i^2-\overline{X}_n^2\right)$$

אומד זה הוא בלתי מוטה, כפי שניתן לראות בחישוב הבא –

$$E\left(S^2\right)=\tfrac{1}{n-1}E\left(\sum_{i=1}^{n}\left(X_i^2-\overline{X}_n^2\right)\right)=\tfrac{1}{n-1}\sum_{i=1}^{n}\left(E\left(X_i^2\right)-E\left(\overline{X}_n^2\right)\right)=$$

$$=\tfrac{1}{n-1}\sum_{i=1}^{n}\left(\sigma^2+\mu^2-\tfrac{1}{n^2}E\left(\sum_{k,j=1}^{n}X_kX_j\right)\right)=\tfrac{n}{n-1}\left(\sigma^2+\mu^2\right)-\tfrac{1}{(n-1)n^2}\sum_{i=1}^{n}\left(E\left(\sum_{k,j=1}^{n}X_kX_j\right)\right)=$$

$$=\tfrac{n}{n-1}\left(\sigma^2+\mu^2\right)-\tfrac{1}{(n-1)n}\sum_{k,j=1}^{n}E\left(X_kX_j\right)=\tfrac{n}{n-1}\left(\sigma^2+\mu^2\right)-\tfrac{1}{(n-1)n}\left[n\left(\sigma^2+\mu^2\right)+n\left(n-1\right)\mu^2\right]=$$

$$=\tfrac{n}{n-1}\left(\sigma^2+\mu^2\right)-\tfrac{1}{(n-1)}\left(\sigma^2+\mu^2\right)-\tfrac{n(n-1)}{(n-1)n}\mu^2=\tfrac{n-1}{n-1}\left(\sigma^2+\mu^2\right)-\mu^2=\sigma^2$$

נכיר כעת התפלגות חדשה שתשמש אותנו במקרים בהם השונות איננה ידועה.

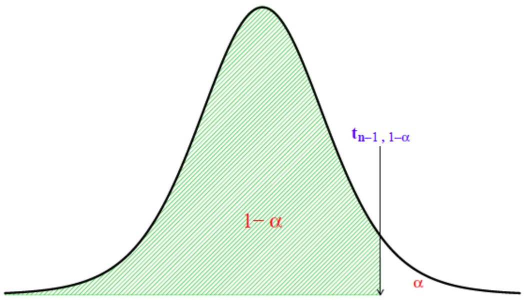
התפלגות t-סטודנט

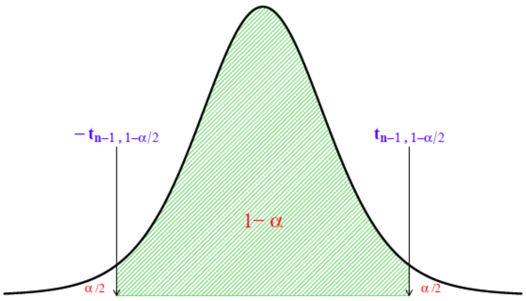
בפרקים הקודמים עבדנו עם התפלגות נורמאלית ומשפט הגבול המרכזי. לטובת המשפט היה עלינו לדעת את השונות של המ"מ. בפעמים בהם נצטרך לאמוד את סטיית התקן נשתמש בקירוב של התפלגות t.

<u>הגדרה</u>: המשתנה T בעל התפלגות t מוגדר באופן הבא. אם Z הוא משתנה מקרי נורמאלי סטנדרטי $Z \sim N(0,1)$ ו-V מתפלג כמו χ_k^2 קרי $V=\sum_{i=1}^{k} Z_i^2$ (כאשר כל איבר בסכום מתפלג כמו נורמאלי סטנדרטי בחזקה שנייה). אז $T=\frac{Z}{\sqrt{V/k}} \sim t_k$, כאשר t_k מציין התפלגות t עם k דרגות חופש.

הקירוב בו אנו נשתמש הוא $T=\frac{\overline{X}_n-\mu}{S/\sqrt{n}} \sim t_{n-1}$ עבור $X_i \sim N(\mu,\sigma^2)$ בלתי תלויים.

הקירוב הנ"ל הוא הקירוב בו נשתמש כאשר סטיית התקן לא ידועה בהתפלגות נורמאלית. התפלגות t היא קרובה מאוד בצורתה להתפלגות הנורמאלית וככל ש-n גדל כך היא מתקרבת להתפלגות הנורמאלית.





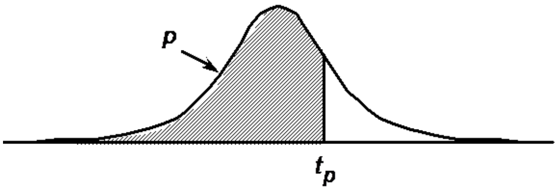
באיורים הנ"ל מוצגים האחוזונים עבור ההתפלגות, כאשר נשים לב שהיא סימטרית סביב ה-0 בדיוק כמו ההתפלגות הנורמאלית.

נסמן ב- $t_{n-1,1-\alpha}$ את האחוזון ה- $1-\alpha$ של התפלגות t עם n-1 דרגות חופש (או לחילופין ב- $t_{1-\alpha,n-1}$).

את ערכי ההתפלגות אנו מוצאים בטבלה (ניתן לחפש על הרשת) כמו בהתפלגות נורמאלית כאשר הטבלה מתחלקת על פי כמות דרגות החופש והאחוזון הרלוונטי (כפי שניתן לראות בעמוד הבא).

הערכים המוצגים בטבלה (בכל שורה) אלו הערכים עבורם ההסתברות של הזנב השמאלי של ההתפלגות יהיה בעל ההסתברות המצויינת בראש הטבלה.

טבלת התפלגות t של סטודנט – ערכי החלוקה t_p



דרגת חופש / הסתברות	0.75	0.9	0.95	0.975	0.99	0.995	0.9995
1	1	3.078	6.314	12.709	31.821	63.657	636.619
2	0.816	1.886	2.92	4.303	6.965	9.925	31.598
3	0.765	1.638	2.353	3.182	4.541	5.841	12.941
4	0.741	1.533	2.132	2.776	3.747	4.604	8.61
5	0.727	1.476	2.015	2.571	3.365	4.032	6.859
6	0.718	1.44	1.943	2.447	3.143	3.707	5.959
7	0.711	1.415	1.895	2.365	2.998	3.499	5.405
8	0.706	1.397	1.86	2.306	2.896	3.355	5.041
9	0.703	1.383	1.833	2.262	2.821	3.25	4.781
10	0.7	1.372	1.812	2.228	2.764	3.169	4.587
11	0.697	1.363	1.796	2.201	2.718	3.106	4.437
12	0.695	1.356	1.782	2.179	2.681	3.055	4.318
13	0.694	1.35	1.771	2.16	2.65	3.012	4.221
14	0.692	1.345	1.761	2.145	2.624	2.977	4.14
15	0.691	1.341	1.753	2.131	2.602	2.947	4.073
16	0.69	1.337	1.746	2.12	2.583	2.921	4.015
17	0.689	1.333	1.74	2.11	2.567	2.898	3.965
18	0.688	1.33	1.734	2.101	2.552	2.878	3.922
19	0.688	1.328	1.729	2.093	2.539	2.861	3.883
20	0.687	1.325	1.725	2.086	2.528	2.845	3.85
21	0.686	1.323	1.721	2.08	2.518	2.831	3.819
22	0.686	1.321	1.717	2.074	2.508	2.819	3.792
23	0.685	1.319	1.714	2.069	2.5	2.807	3.767
24	0.685	1.318	1.711	2.064	2.492	2.797	3.745
25	0.684	1.316	1.708	2.06	2.485	2.787	3.725
26	0.684	1.315	1.706	2.056	2.479	2.779	3.707
27	0.684	1.314	1.703	2.052	2.473	2.771	3.69
28	0.683	1.313	1.701	2.048	2.467	2.763	3.674
29	0.683	1.311	1.699	2.045	2.462	2.756	3.659
30	0.683	1.31	1.697	2.042	2.457	2.75	3.646
40	0.681	1.303	1.684	2.021	2.423	2.704	3.551
60	0.679	1.296	1.671	2	2.39	2.66	3.46
120	0.677	1.289	1.658	1.98	2.358	2.617	3.373
∞	0.674	1.282	1.645	1.96	2.326	2.576	3.291

רווח סמך עבור התוחלת עם שונות לא ידועה

עבור רווח סמך לתוחלת ברמת סמך $1-\alpha$ נקבל באופן דומה את הנוסחה הבאה:

$$\left[\overline{X_n}-\tfrac{S}{\sqrt{n}}t_{n-1,1-\alpha/2},\overline{X_n}+\tfrac{S}{\sqrt{n}}t_{n-1,1-\alpha/2}\right]$$

כאשר הפיתוח הוא זהה כמו במקרה של שונות ידועה:

$$P\left(-t_{n-1,1-\alpha/2}\le\tfrac{\overline{X_n}-\mu}{S/\sqrt{n}}\le t_{n-1,1-\alpha/2}\right)=1-\alpha$$

$$\Rightarrow P\left(\overline{X_n}-\tfrac{S}{\sqrt{n}}t_{n-1,1-\alpha/2}\le\mu\le\overline{X_n}+\tfrac{S}{\sqrt{n}}t_{n-1,1-\alpha/2}\right)=1-\alpha$$

<u>תרגיל 1</u>: ידוע כי תשבץ מסויים דורש בממוצע 5 דקות לפתרון. נותנים את התשבץ ל-9 סטודנטים אקראיים וזמני הפתרון שנמדדו הם: 5.1, 4.9, 5.2, 4.6, 4.6, 4.4, 4.8, 4.3, 5.3.

א. האם ברמת מובהקות של 0.1 תוכלו לקבוע כי סטודנטים פותרים מהר יותר מהאוכלוסיה הכללית? ענו גם עבור ר"מ 0.05.

ב. בנו רווח סמך ברמת סמך של 0.9 לתוחלת זמן הפתרון של תלמידי רפואה.

ג. האם תוכלו לומר כי בר"מ 0.1 משך הפתרון של הסטודנטים שונה מזה של אדם אקראי?

<u>פתרון תרגיל 1</u>:

א. נתחיל ברישום ההשערות. אנו מעוניינים לבדוק האם זמן הפתרון של הסטודנטים קצר מזה של אדם אקראי ולכן נקבע השערות חד צדדיות:

$$H_0:\mu\ge 5\ ;\ H_1:\mu<5$$

נוכל לחשב את סטיית התקן והממוצע המדגמי ונקבל $\overline{X_9}=4.8, S^2=\frac{1}{8}$. כעת נחשב את המובהקות של התוצאה Pvaule ובכך נדע האם התוצאה מובהקת או לא:

$$Pv=P\left(\tfrac{\overline{X_9}-5}{S/\sqrt{n}}<\tfrac{4.8-5}{\frac{1}{\sqrt{8}}/3}\right)=P\left(t_{9-1}<-1.697\right)=1-P\left(t_8<1.697\right)\cong 1-0.925=0.075$$

לכן נוכל לקבוע כי הסטודנטים פותרים מהר יותר בר"מ 0.1, אך לא בר"מ 0.05.

ב. נשתמש בנוסחה שפיתחנו קודם לכן עבור רווח סמך לתוחלת ונקבל

$$\left[4.8-\tfrac{1/\sqrt{8}}{\sqrt{9}}t_{8,0.95},4.8+\tfrac{1/\sqrt{8}}{\sqrt{9}}t_{8,0.95}\right]=\left[4.8-\tfrac{1}{3\sqrt{8}}1.86,4.8+\tfrac{1}{3\sqrt{8}}1.86\right]=\left[4.581,5.019\right]$$

ג. בסעיף הנוכחי שואלים אותנו לגבי האפשרות שזמן הפתרון של הסטודנטים ארוך או קצר מהזמן הכללי באוכלוסיה ולשם כך עלינו לשנות את ההשערות שברשותנו.

$$H_0:\mu=5\ ;\ H_1:\mu\ne 5$$

כעת נבצע את החישוב שעשינו קודם לכן למובהקות התוצאה עבור מבחן דו-צדדי.

$$Pv = 2P\left(\overline{X_9} < 4.8\right) = 2P\left(t_{9-1} < -1.697\right) \cong 0.15 > 0.1$$

ותוצאה זו איננה מובהקות בר"מ 0.1. באותה מידה יכולנו להסתכל על התוצאה בסעיף הקודם ולראות שבכדי להימנע מטעות מהסוג הראשון (קרי, כדי להעריך את התוחלת ברמת ודאות של 0.9 בקטע מסוים) עלינו לכלול את הערך 5 באזור הקבלה. רווח הסמך הוא קטע שהערך נמצא בו, תחת השערת האפס, בהסתברות 0.9 ולכן $\mu = 5$ זו תוצאה סבירה בהסתברות של מעל 0.9 ביחס לממוצע שקיבלנו.

<u>תרגיל 2</u>:

א. בדקו את ההשערה שתוחלת המשקל של שפופרת משחת שיניים היא 10 גרם, אם משקלן של 10 שפופרות שנבחרו באקראי היה: 9.3, 10.2, 9.7, 10.1, 10.3, 10.1, 9.8, 9.9, 10.4, 10.3.

ב. מצאו רווח סמך ברמת סמך 0.95 לתוחלת המשקל ע"ב המדגם הנ"ל.

<u>פתרון תרגיל 2</u>:

א. נתחיל ברישום ההשערות כאשר במקרה הנוכחי יש לנו מבחן דו-צדדי:

$$H_0 : \mu = 10 \;\; ; \;\; H_1 : \mu \neq 10$$

נוכל לחשב את סטיית התקן והממוצע המדגמי ונקבל $\overline{X_{10}} = 10.01, S = 03381$.

נמצא את הפתרון בעזרת מובהקות התוצאה ובעזרת אזור הדחייה. נשים לב כי $\frac{\overline{X_{10}}-10}{S/\sqrt{n}} \sim t_9$.

$$Pv = 2P\left(\overline{X_{10}} > 10.01\right) = 2P\left(\frac{\overline{X_{10}}-10}{S/\sqrt{n}} > \frac{10.01-10}{0.3381/\sqrt{10}}\right) = 2P\left(t_9 > 0.0935\right)$$
$$= 2\left(1 - P\left(t_9 \le 0.0935\right)\right)$$

מאחר ו- $0.5 < P\left(t_9 \le 0.0935\right) < 0.6$ נובע כי $Pv > 0.8 > 0.01$ ולכן התוצאה איננה מובהקת ולא דוחים את השערת האפס.

בעזרת אזור הדחייה:

$$C = \left\{\left|\frac{\overline{X_{10}}-10}{S/\sqrt{n}}\right| > t_{9,1-\frac{0.01}{2}}\right\} = \left\{\left|\frac{\overline{X_{10}}-10}{S/\sqrt{n}}\right| > t_{9,0.995}\right\} = \left\{\left|\overline{X_{10}} - 10\right| > \frac{0.3381}{\sqrt{10}} t_{9,0.995}\right\} =$$
$$= \left\{\overline{X_{10}} > 10 + \frac{0.3381}{\sqrt{10}} t_{9,0.995}\right\} \cup \left\{\overline{X_{10}} < 10 - \frac{0.3381}{\sqrt{10}} t_{9,0.995}\right\} =$$
$$= \left\{\overline{X_{10}} > 10 + \frac{0.3381}{\sqrt{10}} 3.25\right\} \cup \left\{\overline{X_{10}} < 10 - \frac{0.3381}{\sqrt{10}} 3.25\right\} =$$
$$= \left\{\overline{X_{10}} > 10.347\right\} \cup \left\{\overline{X_{10}} < 9.652\right\}$$

ואנו רואים כי התוצאה היא לא באזור הדחייה ולכן לא ניתן לדחות את השערת האפס.

ב. . נשתמש בנוסחה שפיתחנו קודם לכן עבור רווח סמך לתוחלת ונקבל

$$\left[10.01-\tfrac{0.3381}{\sqrt{10}}t_{9,0.975},10.01+\tfrac{0.3381}{\sqrt{10}}t_{9,0.975}\right]=\left[10.01-\tfrac{0.3381}{\sqrt{10}}2.262,10.01+\tfrac{0.3381}{\sqrt{10}}2.262\right]=$$
$$=\left[9.768,10.251\right]$$

<u>תרגיל 3</u>: משקל של פיתה הנאפת במאפייה מתפלג נורמאלית עם תוחלת 100 גרם. צרכנים התלוננו בפני בעל המאפיה כי משקל הפיתות נמוך מהממוצע המופיע על האריזה. לשם הבדיקה נקלחו 25 פיתות ונמצא כי המשקל הממוצע היה 96 גרם וסטית התקן 11 גרם. בדקו את טענות הצרכנים ברמת מובהקות של 5%.

<u>פתרון תרגיל 3</u>:

ההשערות שלנו בבעיה זו הן $H_0:\mu\geq 100\ ;\ H_1:\mu<100$ ונוכל לחשב את הסטטיסטי של הניסוי על ידי $\frac{\overline{X_{25}}-100}{S/\sqrt{n}}=\frac{\overline{X_{25}}-100}{11/\sqrt{25}}\sim t_{24}$. נחשב את מובהקות התוצאה שקיבלנו:

$$Pv=P_{H_0}\left(\overline{X_{25}}\leq 96\right)=P_{H_0}\left(\frac{\overline{X_{25}}-100}{S/\sqrt{n}}\leq\frac{96-100}{11/\sqrt{25}}\right)=P_{H_0}\left(t_{24}\leq -1.818\right)=$$
$$=1-P_{H_0}\left(t_{24}\leq 1.818\right)<1-0.95=0.05\ \Rightarrow Pv=<0.05$$

ולכן התוצאה היא מובהקת ונצטרך לדחות את השערת האפס.

<u>תרגיל 4</u>: הציון הממוצע במבחנים בסטטיסטיקה הוא 80. במסגרת ניסוי, לקחו 16 סטודנטים ונתנו להם לאכול שוקולד מריר לפני הבחינה. החוקרת שיערה כי השוקולד יסייע לסטודנטים בבחינה. סכום הציונים במדגם היה 1299 וסכום ריבוע הציונים היה 107,225.

א. האם ההשערה נכונה של החוקרת נכונה? בדקו בר"מ 0.05 וחשבו מובהקות תוצאה.

ב. מצאו רווח סמך ברמת סמך 95% עבור התוחלת של הציון במבחן. האם ניתן להסיק מן הרווח משהו לגבי ההשערה שנבדקה בחלק א'?

<u>פתרון תרגיל 4</u>:

א. נתחיל ברישום ההשערות וחישוב הערכים בניסוי.

$$H_0:\mu=80\ ;\ H_1:\mu>80$$

נחשב את ממוצע המדגם וסטיית התקן המדגמית:

$$\overline{X_{16}} = \frac{1299}{16} = 81.19$$

$$S^2 = \tfrac{1}{n-1}\sum_{i=1}^{n}\left(X_i^{\,2} - \overline{X_n}^{\,2}\right) = \tfrac{1}{n-1}\sum_{i=1}^{n}X_i^{\,2} - \tfrac{1}{n-1}\sum_{i=1}^{n}\overline{X_n}^{\,2} = \tfrac{1}{15}\sum_{i=1}^{16}X_i^{\,2} - \tfrac{16}{15}\overline{X_n}^{\,2} =$$

$$= \tfrac{107225}{15} - \tfrac{16}{15}\left(\tfrac{1299}{16}\right)^2 = 117.4958$$

מתוך נתונים אלו נוכל לחשב את מובהקות התוצאה:

$$Pv = P_{H_0}\left(\overline{X_{16}} > 81.19\right) = P_{H_0}\left(\frac{\overline{X_{16}}-80}{S/\sqrt{n}} > \frac{81.19-80}{\sqrt{117.5}/\sqrt{16}}\right) = P_{H_0}\left(t_{15} > 0.439\right) =$$

$$= 1 - P_{H_0}\left(t_{15} \le 0.439\right) > 1 - 0.75 = 0.25 > 0.05$$

לכן התוצאה איננה מובהקת ולא ניתן לשלול את השערת האפס.

ב. רווח הסמך שהתקבל מהניסוי הוא:

$$\left[81.19 - \tfrac{S}{\sqrt{n}}t_{15,0.975}, 81.19 + \tfrac{S}{\sqrt{n}}t_{15,0.975}\right] = \left[81.19 - \tfrac{\sqrt{117.5}}{\sqrt{16}}\cdot 2.13\,,\, 81.19 + \tfrac{\sqrt{117.5}}{\sqrt{16}}\cdot 2.13\right] =$$

$$= \left[75.42\,,\, 86.96\right]$$

ושוב אנו רואים כי רווח הסמך כולל את הערך 80 ולכן לא ניתן להסיק מסקנות מרחיקות לכת מחישוב זה.

פרק 14

נושאים:

- בדיקת השערות עבור השונות.
- התפלגות חי בריבוע.
- רווח סמך לשונות.

בדיקת השערות עבור שונות

נניח כי רוצים לבדוק השערה על השפעה של טיפול חדש. אם אפשר היה להניח שטיפול חדש משפיע רק על התוחלת אבל אינו משפיע על סטיית התקן, היינו יכולים להמשיך כאילו סטיית התקן ידועה, כפי שעשינו קודם לכן לצערנו,. ההנחה הזו לא תמיד סבירה, ולמעשה ניתן לבדוק אותה. אפשר לנסח השערות לגבי השונות החדשה:

$$H_0 : \sigma = \sigma_0 \;\; ; \;\; H_1 : \sigma \neq \sigma_0$$

בדיקת השערה זו תהיה מבוססת על האומד חסר הטיה עבור השונות שראינו קודם לכן:

$$.S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

על מנת להחליט האם חל שינוי בגודל השונות יש לבדוק עד כמה הערך של S^2 רחוק מהערך המשוער ולשם כך, עלינו לדעת כיצד מתפלג המשתנה S^2.

התפלגות חי-בריבוע

ידוע כי עבור $X \sim N\left(\mu,\sigma^2\right)$ אזי $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$ וכעת נגדיר התפלגות חדשה המתקבלת ע"י העלאה בריבוע של ההתפלגות הנורמלית הסטנדרטית:

$$Z^2 = \left(\frac{X-\mu}{\sigma}\right)^2 \sim \chi_1^2$$

להתפלגות זו (כפי שבוודאי ראיתם בשלב מסויים בתרגילי הבית בהסתברות) קוראים ***חי-בריבוע עם דרגת חופש אחת***.

דרגת-חופש אחת, כיוון שהיא מבוססת על תצפית אחת מהתפלגות נורמלית סטנדרטית.

<u>הגדרה</u>: סכום של ריבועים של שתי תצפיות ב"ת מתוקננות מתפלג לפי התפלגות חי-בריבוע עם שתי דרגות חופש.

$$Z_1^2+Z_2^2=\left(\frac{X_1-\mu}{\sigma}\right)^2+\left(\frac{X_2-\mu}{\sigma}\right)^2\sim\chi_1^2+\chi_1^2=\chi_2^2$$

ובאותו אופן, אם נחבר n ריבועים של תצפיות נורמליות מתוקננות נקבל התפלגות חי-בריבוע עם n דרגות חופש.

$$\sum_{i=1}^n Z_i^2=\sum_{i=1}^n\left(\frac{X_i-\mu}{\sigma}\right)^2\sim\chi_n^2$$

במקרה שלנו התוחלת אינה ידועה, ולכן נחליף אותה באומד שלה – ממוצע המדגם. באופן הזה מקבלים תקנון לא מדוייק, ונשארים עם דרגת חופש אחת פחות, כלומר:

$$\sum_{i=1}^n\left(\frac{X_i-\overline{X}}{\sigma}\right)^2\approx\chi_{n-1}^2$$

ולכן

$$\frac{n-1}{\sigma^2}S^2=\frac{n-1}{\sigma^2}\cdot\left\{\frac{1}{n-1}\sum_{i=1}^n\left(X_i-\overline{X}\right)^2\right\}=\sum_{i=1}^n\frac{\left(X_i-\overline{X}\right)^2}{\sigma^2}\approx\chi_{n-1}^2$$

כמו במקרה של ההתפלגות הנורמלית, אין נוסחה סגורה להתפלגות פיכך יש טבלה המכילה ערכים קריטיים להסתברויות ספציפיות.

<u>תרגיל 1</u>: חשבו את אזור הדחייה עבור $H_0:\sigma=\sigma_0, H_1:\sigma\neq\sigma_0$.

<u>פתרון תרגיל 1</u>: עבור מבחן ברמת מובהקות α, אזור הדחייה יהיה

$$C=\left\{S^2>k_1\right\}\cup\left\{S^2<k_2\right\}$$

ונדרוש כי

$$P_{H_0}(C)=P_{H_0}\left(S^2>k_1\right)+P_{H_0}\left(S^2<k_2\right)=\alpha.$$

$$\alpha=P_{H_0}\left(S^2>k_1\right)+P_{H_0}\left(S^2<k_2\right)=P\left(\tfrac{n-1}{\sigma_0^2}S^2>\tfrac{n-1}{\sigma_0^2}k_1\right)+P\left(\tfrac{n-1}{\sigma_0^2}S^2<\tfrac{n-1}{\sigma_0^2}k_2\right)=$$

$$=P\left(\chi_{n-1}^2>\tfrac{n-1}{\sigma_0^2}k_1\right)+P\left(\chi_{n-1}^2<\tfrac{n-1}{\sigma_0^2}k_2\right)=\tfrac{\alpha}{2}+\tfrac{\alpha}{2}$$

$$\Rightarrow\ \tfrac{n-1}{\sigma_0^2}k_2=\chi^2_{n-1,\frac{\alpha}{2}}\ ;\ \tfrac{n-1}{\sigma_0^2}k_1=\chi^2_{n-1,1-\frac{\alpha}{2}}\ \Rightarrow\ k_2=\tfrac{\sigma_0^2}{n-1}\chi^2_{n-1,\frac{\alpha}{2}}\ ;\ k_1=\tfrac{\sigma_0^2}{n-1}\chi^2_{n-1,1-\frac{\alpha}{2}}$$

$$\Rightarrow\ C=\left\{S^2>\tfrac{\sigma_0^2}{n-1}\chi^2_{n-1,1-\frac{\alpha}{2}}\right\}\cup\left\{S^2<\tfrac{\sigma_0^2}{n-1}\chi^2_{n-1,\frac{\alpha}{2}}\right\}$$

עבור המקרים החד-צדדיים נקבל נוסחה דומה עם האחוזונים המתאימים.

תרגיל 2: כדי למדוד את הרגלי הצפייה של בנות בטלוויזיה, מדדו את משך זמן הצפייה של חמש בנות וקיבלו את התוצאות: 2.5, 4.5, 5.5, 6.5, 8.5. הניחו שמשך זמן הצפייה של בנות מתפלג נורמלית._על-סמך המדגם, האם ניתן להסיק ברמת-מובהקות של 0.05 שסטיית התקן של משך זמן הצפייה גדולה מ – 1.45 שעות. מצאו גם את מובהקות התוצאה.

פתרון תרגיל 2: ההשערות הן: $H_0:\sigma^2=1.45^2;H_1:\sigma^2>1.45^2$ ולכן אזור הדחיה הוא (בדומה למה שמצאנו בתרגיל 1 אך עבור המקרה החד צדדי)

$$C=\left\{S^2>\frac{\sigma_0{}^2}{n-1}\chi^2_{n-1,1-\alpha}\right\}=\left\{S^2>\frac{1.45^2}{4}\chi^2_{4,0.95}\right\}=\left\{S^2>\frac{1.45^2}{4}9.49\right\}=\left\{S^2>4.99\right\}$$

נחשב את השונות המדגמית S^2:

$$\bar{X}=\frac{8.5+6.5+5.5+4.5+2.5}{5}=5.5$$

$$\sum_{i=1}^{n}X_i{}^2=8.5^2+6.5^2+5.5^2+4.5^2+2.5^2=171.25$$

מכאן:

$$S^2=\frac{\sum_{i=1}^{n}X_i{}^2-n\bar{X}^2}{n-1}=\frac{171.25-5\cdot5.5^2}{4}=5$$

ונאלץ לדחות את השערת האפס מאחר ו-5 נמצא באזור הדחייה. תוצאה דומה נקבל גם בחישוב מובהקות התוצאה.

$$Pv=P_{H_0}\left(S^2>5\right)=P\left(\frac{4}{1.45^2}S^2>\frac{4}{1.45^2}5\right)=P\left(\chi_4{}^2>9.5\right)=1-P\left(\chi_4{}^2\le9.5\right)$$

$$0.95<P\left(\chi_4{}^2\le9.5\right)<0.975\Rightarrow0.025<pv<0.05$$

רווח סמך עבור סטיית התקן

תרגיל 3: מצאו נוסחה לרווח סמך עבור סטיית התקן ברמת סמך $1-\alpha$.

פתרון תרגיל 3: נתחיל מהנוסחה שהגענו אליה קודם לכן והיא:

$$1-\alpha=P\left(\chi^2_{n-1,\frac{\alpha}{2}}<\frac{n-1}{\sigma_0{}^2}S^2<\chi^2_{n-1,1-\frac{\alpha}{2}}\right)=P\left(\frac{\sigma_0{}^2}{n-1}\chi^2_{n-1,\frac{\alpha}{2}}<S^2<\frac{\sigma_0{}^2}{n-1}\chi^2_{n-1,1-\frac{\alpha}{2}}\right)=$$

לכן נקבל כי השיוויון הבא מתקיים עבור סטיית התקן

$$1-\alpha = P\left(\frac{n-1}{\chi^2_{n-1,1-\frac{\alpha}{2}}}S^2 < {\sigma_0}^2 < \frac{n-1}{\chi^2_{n-1,\frac{\alpha}{2}}}S^2\right)$$

מכאן נסיק כי הנוסחה לרווח סמך ברמת סמך $1-\alpha$ לסטיית התקן היא

$$\left[\sqrt{\frac{n-1}{\chi^2_{n-1,1-\frac{\alpha}{2}}}}S, \sqrt{\frac{n-1}{\chi^2_{n-1,\frac{\alpha}{2}}}}S\right].$$

פרק 15

נושאים:

- השוואת שתי אוכלוסיות – מבחן t מזווג.
- מבחן t למדגמים בלתי תלויים.

השוואת שתי אוכלוסיות – מבחן t מזווג

עד כה התמודדנו עם השערות ובדיקות עבור אוכלוסיה ספציפית שנלקח ממנה מדגם. אך,כעת, באופן טבעי עולה השאלה לגבי שתי אוכלוסיות? מה עושים כאשר מעוניינים להשוות בין 2 אוכלוסיות וכיצד מתמודדים עם צמד מדגמים?

נתחיל בהתמודדות עם מצב בו ישנה תלות בין זוגות של תצפיות.

יהיו שני מדגמים של n תצפיות בלתי תלויות - $X_1,\ldots,X_n\ ;Y_1,\ldots,Y_n$ כך שכל מ"מ מתפלג נורמאלית עם תוחלות μ_X,μ_Y בהתאמה (שימו לב כי תיתכן תלות בין X_i ל- Y_i). נגדיר את המשתנים המקריים $D_i=X_i-Y_i, i=1,\ldots,n$.

אנו מעוניינים לבדוק האם ישנו הבדל בין התוחלות ולכן נקבע את ההשערות שלנו באופן הבא:

$$H_0:\mu_D=0\ ;\ H_1:\mu_D\neq 0 .$$

מאחר ויש ברשותנו משתנה מקרה חדש נצטרך גם אומד עבור השונות שלו וזה יהיה (בדומה לאומדים שראינו בפרקים קודמים):

$$S_D{}^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(D_i{}^2-\overline{D_n}^2\right)$$

ולכן נקבל כי הסטטיסטי שלנו הוא:

$$\frac{\overline{D_n}-\mu_D}{S_D/\sqrt{n}}\sim t_{n-1}$$

וברוב המקרים, כאמור, $\mu_D=0$ תחת השערת האפס.

מבחן t למדגמים בלתי תלויים

מקרה נוסף אפשרי הוא שיש ברשותנו 2 מדגמים בלתי תלויים $X_1,\ldots,X_n\ ;Y_1,\ldots,Y_m$, לא בהכרח מאותו הגודל. אנחנו נניח כי יש להם את אותה השונות. במקרה זה סטיית התקן תחושב לפי האומד

$$S^2=\frac{(n-1)S_X^2+(m-1)S_Y^2}{n+m-2}$$

כאשר $S_X^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i^{\ 2}-\overline{X_n}^{\ 2}\right)$ ובהתאם עבור Y והסטטיסטי שלנו (עבור $H_0:\mu_x=\mu_Y$) יהיה

$$\frac{\left(\overline{X_n}-\overline{Y_m}\right)-\left(\mu_X-\mu_Y\right)}{S\sqrt{\frac{1}{n}+\frac{1}{m}}}\sim t_{n+m-2}$$

רווח סמך עבור $\mu_x-\mu_Y$ ברמת סמך $1-\alpha$ יהיה

$$\left[\left(\overline{X_n}-\overline{Y_m}\right)-S\sqrt{\tfrac{1}{n}+\tfrac{1}{m}}t_{n+m-2,1-\alpha/2},\left(\overline{X_n}-\overline{Y_m}\right)+S\sqrt{\tfrac{1}{n}+\tfrac{1}{m}}t_{n+m-2,1-\alpha/2}\right].$$

<u>תרגיל 1 (מדגם מזווג)</u>: בכדי לבחון יעילות של סדנת גמילה מעישון, נלקח מדגם של 9 מעשנים ועבור כל אחד נבדקה צריכת הסיגריות היומיתלפני ואחרי הסדנה. להלן התוצאות:

לפני	40	22	18	28	20	25	30	30	5
אחרי	36	16	13	30	15	22	33	20	3

א. נסחו את ההשערות ובדקו ברמת מובהקות של 5%.

ב. מצאו רווח סמך לממוצע ההפחתה בכמות הסיגריות היומיות ברמת סמך 95%.

<u>פתרון תרגיל 1</u>:

א. נגדיר $\mu_D=\mu_{prior}-\mu_{after}$ וננסח את ההשערות באופן הבא:

$H_0:\mu_D\le 0\ ;\ H_1:\mu_D>0$ (מבחן חד צדדי ימני). נחשב את הערכים הרלוונטיים ונקבל - $\overline{D}=2.22,S_D=5.28$. נחשב את מובהקות התוצאה לפי הנתונים לעיל,

$$Pv=P_{H_0}\left(\overline{D}>3.33\right)=P\left(\tfrac{\overline{D}-0}{S_D/\sqrt{n}}>\tfrac{2.22-0}{5.28/\sqrt{9}}\right)=P\left(t_8>1.26\right)>0.1>0.05$$

וקיבלנו כי אין לדחות את השערת האפס ונסיק כי השערת האפס עובדת.

ב. נשתמש בנוסחה לרווח סמך עם התפלגות t.

$$\left[\overline{X_n}-\tfrac{S}{\sqrt{n}}t_{n-1,1-\alpha/2},\overline{X_n}+\tfrac{S}{\sqrt{n}}t_{n-1,1-\alpha/2}\right]=\left[2.22-\tfrac{5.28}{\sqrt{9}}t_{8,0.975},2.22+\tfrac{5.28}{\sqrt{9}}t_{8,0.975}\right]$$

<u>תרגיל 2 (מדגמים בלתי תלויים)</u>: נלקח מדגם של 7 סטודנטים ו-5 סטודנטיות ונמדדו מנות המשכל של כ"א. להלן התוצאות:

סטודנטים	104	125	103	105	13	117	110
סטודנטיות	107	110	102	110	116		

בדקו האם יש הבדל בין מנת המשכל הממוצעת של הסטודנטים לבין זו של הסטודנטיות ברמת מובהקות 0.05 כאשר מניחים שהשונויות שוות.

<u>פתרון תרגיל 2</u>: נשים לב כי קיבלנו מדגמים בלתי תלויים והמבחן שלנו יהיה דו-צדדי:

$$H_0: \mu_D = \mu_M - \mu_W = 0 \;\; ; \;\; H_1: \mu_D \neq 0$$

נחשב את כל הערכים הרלוונטיים עבור החישוב של מובהקות התוצאה.

$$\overline{X_7} = 111 \;\; ; \;\; \overline{Y_5} = 109$$

$$S_M^2 = \frac{1}{7-1}\sum_{i=1}^{7}\left(X_i^2 - 111^2\right) = 64.33$$

$$S_W^2 = \frac{1}{5-1}\sum_{i=1}^{5}\left(Y_i^2 - 109^2\right) = 26$$

$$S = \sqrt{\frac{(7-1)S_M^2 + (5-1)S_W^2}{7+5-2}} = 7$$

$$Pv = 2P_{H_0}\left(\overline{X_7} - \overline{Y_5} > 111 - 109\right) = 2P\left(\frac{\overline{X_7} - \overline{Y_5}}{S\sqrt{\frac{1}{7}+\frac{1}{5}}} > \frac{111-109}{7\sqrt{\frac{1}{7}+\frac{1}{5}}}\right) =$$

$$= 2P\left(t_{10} > 0.487\right) = 0.5 > 0.05$$

ולכן לא נוכל להצביע על שוני כלשהו במנת המשכל בין הקבוצות.

<u>תרגיל 3</u>: עקב תלונות מצד הסטודנטים על פערים באיכות ההוראה בין בן ואלון החליטו להשוות בין ציוני הסטודנטים שלמדו אצל המרצים השונים. לצורך השוואה נבחנו 9 סטודנטים מכל קבוצה ואלו התוצאות שהם הציגו:

אלון	88	73	87	65	46	67	92	34	69
בן	92	75	90	68	49	70	94	33	71

השוו בין הציונים אצל המרצים השונים ברמת מובהקות 0.05 בהנחה ש-

א. 2 הקבוצות נבחרו באופן בלתי תלוי זו בזו.

ב. כנגד כל תלמיד של אחד נבחר תלמיד בעל הישגי עבר זהים שלמד אצל האחר.

הסבירו את ההבדל בין הסעיפים.

<u>פתרון תרגיל 3</u>:

א. בסעיף זה ההנחה היא שהמדגמים ב"ת. נתיחל בהגדרת ההשערות שלנו:

$$H_0: \mu_D = \mu_B - \mu_A \leq 0 \;\; ; \;\; H_1: \mu_D > 0$$

(הנחנו כי אנחנו בוחנים האם בן עדיף על אלון, ניתן להגדיר את ההשערות בצורות אחרות גם כן). נחשב את הערכים הרלוונטיים לטובת מציאת מובהקות התוצאה,

$$\overline{X_A}=69\ ;\ \overline{X_B}=71\tfrac{1}{3}\ ;\ S_A^2=375.5\ ;\ S_B^2=410.5$$
$$S^2=393$$
$$Pv=P_{H_0}\left(\overline{X_B}-\overline{X_A}>71\tfrac{1}{3}-69\right)=P\left(\frac{\overline{X_B}-\overline{X_A}}{S\sqrt{\frac{1}{9}+\frac{1}{9}}}>\frac{2\frac{1}{3}}{\sqrt{393\cdot\frac{2}{9}}}\right)\cong P\left(t_{16}>0.25\right)=$$
$$=1-P\left(t_{16}\le 0.25\right)>0.25>0.05$$

לכן לא נפסול את הנחת האפס ונבטל את הטענות לגבי הפערים. שימו לב כי יכולנו לעשות את אותו החישוב עבור מבחן דו -צדדי ולקבל שרמת המובהקות היא 0.5 ולכן גם במבחן הדו-צדדי הנחת האפס הייתה נשמרת.

ב. כעת עלינו להניח כי המדגמים מזווגים, ולכן נחשב כמו מקודם:

$$H_0:\mu_D=\mu_B-\mu_A\le 0\ ;\ H_1:\mu_D>0$$

נחשב את כל הערכים הרלוונטיים:

$$\overline{D}=2\tfrac{1}{3}\ ;\ S_D^2=2$$
$$Pv=P_{H_0}\left(\overline{D}>2\tfrac{1}{3}\right)=P\left(\frac{\overline{D}}{S/\sqrt{9}}>\frac{2\frac{1}{3}}{\sqrt{2}/\sqrt{9}}\right)=P\left(t_8>\tfrac{7}{\sqrt{2}}\right)\cong P\left(t_8>4.95\right)=$$
$$=1-P\left(t_8\le 4.95\right)<0.005<0.05$$

והפעם אנו מקבלים כי התוצאה היא מובהקת ואכן בן הוא מרצה יותר טוב מאלון.

ממה נובע הפער?

באופן כללי ניתן לראות כי הישגי הסטודנטים של בן גבוהים מאלו של אלון אבל כאשר מניחים כי ישנה אי-תלות אז השונות בציונים גדולה מספיק בכדי לכפר על ההבדלים בינהם וניתן להניח כי השוני נובע פשוט מהשונות הגדולה בציונים ונובעת מאקראיות ותו לא. לעומת זאת, אם ישנה תלות, אז אנחנו רואים שבצורה עיקבית סטודנטים של בן מקבלים ציונים יותר גבוהים מאלו של אלון (למרות שההנחה במודל היא שהם בעלי אותן היכולות). השונות במקרה הזה היא קטנה מאוד ביחס למקרה הראשון ולכן התזוזה בציונים הופכת להיות משמעותית בהרבה.

פרק 16

נושאים:

- שאלות חזרה לפני מבחן.

תרגיל 1: בחדר יש n זוגות נשואים המסתדרים בשורה באופן אקראי. נגדיר את X להיות מספר הזוגות שבני הזוגות עומדים צמוד זה לזה. מצאו את התוחלת והשונות של X.

פתרון תרגיל 1: הדרך הקלה והמהירה לחישוב הנ"ל היא באמצעות פירוק לאינדיקטורים.

נניח כי $X=\sum_{i=1}^{n}X_i,\ X_i=\begin{cases}1 & \text{couple } i \text{ are side by side}\\ 0 & \text{otherwise}\end{cases}$.

$$P\left(X_i=1\right)=\frac{\left(2n-1\right)!\cdot 2!}{\left(2n\right)!}=\frac{1}{n}$$

ולכן נקבל ש-

$$\Rightarrow E\left(X_i\right)=\frac{1}{n},V\left(X_i\right)=\frac{1}{n}\left(1-\frac{1}{n}\right)=\frac{\left(n-1\right)}{n^2}$$

ולכן התוחלת של X שווה ל-

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

$$V\left(X\right)=V\left(\sum_{i=1}^{n}X_i\right)=\sum_{i=1}^{n}V\left(X_i\right)+2\sum_{i<j}Cov\left(X_i,X_j\right)=n\cdot\frac{\left(n-1\right)}{n^2}+2\binom{n}{2}Cov\left(X_1,X_2\right)$$

כאשר המעבר האחרון בוצע על ידי הצבת השונות שמצאנו קודם לכן וגם בגלל שכל השונויות המשותפת שוות וישנם סה"כ $\binom{n}{2}$ איברים כאלה בטור. נחשב את השונות המשותפת ונקבל:

$$Cov\left(X_1,X_2\right)=E\left(X_1X_2\right)-E\left(X_1\right)E\left(X_2\right)=P\left(X_1=1,X_2=1\right)-\frac{1}{n^2}=$$

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

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

תרגיל 2: נניח שמספר האירועים המתרחשים בזמן נתון הוא $X\sim P\left(\lambda\right)$. מצאו את התפלגות מספר האירועים שמבחינים בהם בפרק הזמן הזה, אם בכל אירוע מבחינים בהסתברות p ללא תלות באירועים האחרים.

<u>פתרון תרגיל 2</u>: נגדיר את מספר האירועים שמבחינים בהם כ-Y. נוכל לרשום כי $Y|X \sim Bin(X,p)$ ולכן לפי נוסחת התוחלת השלמה נקבל $E(Y|X)=X\cdot p$ ולכן התוחלת של Y היא $E(Y)=E(E(Y|X))=E(X\cdot p)=pE(X)=p\lambda$.

כעת נמצא כיצד מתפלג Y בעזרת נוסחת ההסתברות השלמה:

$$P(Y=k)=\sum_{m=k}^{\infty}P(Y=k|X=m)P(X=m)=\sum_{m=k}^{\infty}\binom{m}{k}p^k(1-p)^{m-k}\cdot e^{-\lambda}\tfrac{\lambda^m}{m!}=$$
$$=e^{-\lambda}p^k\sum_{m=k}^{\infty}\binom{m}{k}(1-p)^{m-k}\cdot\tfrac{\lambda^m}{m!}=e^{-\lambda}p^k\sum_{n=0}^{\infty}\binom{n+k}{k}(1-p)^n\cdot\tfrac{\lambda^{n+k}}{n+k!}=$$
$$=\frac{e^{-\lambda}(\lambda p)^k}{k!}\sum_{n=0}^{\infty}(1-p)^n\cdot\tfrac{\lambda^n}{k!}=\frac{e^{-\lambda}(\lambda p)^k}{k!}\sum_{n=0}^{\infty}\frac{((1-p)\lambda)^n}{n!}=\frac{e^{-\lambda}(\lambda p)^k e^{\lambda(1-p)}}{k!}=e^{-\lambda p}\cdot\frac{(\lambda p)^k}{k!}$$

ולכן $Y \sim P(\lambda p)$.

<u>תרגיל 3</u>: נתונה פונקצית הצפיפות הבאה

$$f_{X,Y}(x,y)=\begin{cases}\frac{1}{8}xe^{-\frac{x+y}{2}} & x,y>0\\ 0 & otherwise\end{cases}$$

א. חשבו את $P(X>1,Y>1),P(Y>2)$.

ב. האם X ו-Y ב"ת.

ג. חשבו את $E\left(\frac{Y}{X}\right)$.

<u>פתרון תרגיל 3</u>:

א.
$$P(X>1,Y>1)=\int_1^{\infty}\int_1^{\infty}\tfrac{1}{8}xe^{-\frac{x+y}{2}}dxdy=\tfrac{1}{8}\int_1^{\infty}xe^{-\frac{x}{2}}dx\int_1^{\infty}e^{-\frac{y}{2}}dy=\tfrac{1}{8}\int_1^{\infty}xe^{-\frac{x}{2}}2e^{-\frac{1}{2}}dx=$$
$$=\tfrac{1}{4}e^{-\frac{1}{2}}\int_1^{\infty}xe^{-\frac{x}{2}}dx=\tfrac{1}{4}e^{-\frac{1}{2}}\left[-2xe^{-\frac{x}{2}}\right]_1^{\infty}+\tfrac{1}{2}e^{-\frac{1}{2}}\int_1^{\infty}e^{-\frac{x}{2}}dx=$$
$$=\tfrac{1}{2}e^{-1}+e^{-1}=\tfrac{3}{2}e^{-1}$$

.

$$f_Y(y)=\int_{-\infty}^{\infty}f_{X,Y}(x,y)dx=\int_0^{\infty}\tfrac{1}{8}xe^{-\frac{x+y}{2}}dx=\tfrac{1}{8}e^{-\frac{y}{2}}\int_0^{\infty}xe^{-\frac{x}{2}}dx=\tfrac{1}{8}e^{-\frac{y}{2}}\int_0^{\infty}xe^{-\frac{x}{2}}dx=\tfrac{1}{2}e^{-\frac{y}{2}}$$
$$f_X(x)=\int_{-\infty}^{\infty}f_{X,Y}(x,y)dy=\int_0^{\infty}\tfrac{1}{8}xe^{-\frac{x+y}{2}}dy=\tfrac{1}{8}xe^{-\frac{x}{2}}\int_0^{\infty}e^{-\frac{y}{2}}dy=\tfrac{1}{4}xe^{-\frac{x}{2}}$$

ולכן נוכל לחשב את ההסתברות ש-Y גדול מ-2 ישירות ולקבל:

$$P\left(Y>2\right)=\int_2^{\infty}\tfrac{1}{2}e^{-\frac{y}{2}}dy=\left[-\tfrac{2}{2}e^{-\frac{y}{2}}\right]_2^{\infty}=e^{-1}$$

ב. האם X ו-Y ב"ת? כן וזאת אנו רואים כי הצפיפות המשותפת שלהם היא פשוט מכפלת הצפיפויות השוליות ולכן הם ב"ת לפי הגדרה.

ג.

$$E\left(\tfrac{Y}{X}\right)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\tfrac{y}{x}f_{X,Y}\left(x,y\right)dxdy=\int_0^{\infty}\int_0^{\infty}\tfrac{y}{x}\cdot\tfrac{1}{8}xe^{-\frac{x+y}{2}}dxdy=\int_0^{\infty}\int_0^{\infty}y\tfrac{1}{8}e^{-\frac{x+y}{2}}dxdy=$$

$$=\int_0^{\infty}\int_0^{\infty}x\tfrac{1}{8}e^{-\frac{y+x}{2}}dydx=1$$

כאשר המעבר האחרון בוצע על ידי החלפת X ב-Y ולהיפך (החלפת השמות של המשתנים המקריים / החלפת משתנים) והשיוויון ל-1 נובע מאחר וקיבלנו אינטגרל על פונקצית הצפיפות המקורית על כל המרחב ולכן הוא שווה ל-1.

<u>תרגיל 4</u>: ההסתברות שיאיר יהיה שר בממשלה היא 0.6. ההסתברות שנפתלי יהיה שר בממשלה היא 0.4. ההסתברות שציפי תהיה שרה בממשלה היא 0.2. ציפי תהיה שרה בממשלה רק אם יאיר יהיה שר בממשלה. ההסתברות שאף אחד מהשלושה לא יהיה שר בממשלה היא 0.1. המאורעות "נפתלי יהיה שר בממשלה" ו"ציפי תהיה שרה בממשלה" הם בלתי תלויים. יהי X_1 משתנה מקרי המקבל את הערך "1" אם יאיר יהיה שר ו-"0" אחרת. יהי X_2 משתנה מקרי המקבל את הערך "1" אם נפתלי יהיה שר ו-"0" אחרת. יהי X_3 משתנה מקרי המקבל את הערך "1" אם ציפי תהיה שרה ו-"0" אחרת.

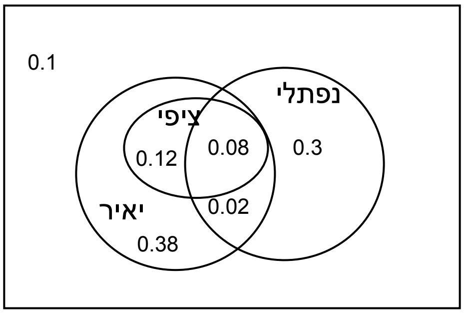
א. חשבו $COV(X_1,X_2)$ והסבירו את סימנו.

ב. האם X_1,X_3 בלתי מתואמים? בלתי תלויים? נמקו!

ג. חשבו $COV(X_1+X_3,X_2)$.

<u>פתרון תרגיל 4</u>:

א. יש להתחיל בשרטוט דיאגרמת ואן ומציאת ההסתברות לכל מאורע. נתחיל בכך שנשים לב שהמאורע "ציפי שרה בממשלה" מוכל במאורע "יאיר חבר בממשלה" (הפירוש המתמטי של הנתון "ציפי תהיה שרה בממשלה רק אם יאיר יהיה שר בממשלה" <u>איננו שהמאורעות 'ציפי שרה בממשלה' ו'יאיר שר בממשלה' זהים</u>, ולכן אין שום סיבה שההסתברות שציפי תהיה בממשלה תהיה 0.6!) ולכן נצטרך לשים מאורע אחד בתוך אחר בדיאגרמה. בנוסף, ההסתברות שציפי ונפתלי יהיו יחדיו בממשלה היא 0.08 (בגלל אי-תלות, $0.2\cdot0.4=0.08$).



את ההסתברות של החיתוך של יאיר ונפתלי שלא כולל את ציפי מוצאים ע"י סכימת כל ההסתברויות השונות ל-1.

לאחר מכן מבצעים חישוב ישיר באופן הבא:

$$COV\left(X_1,X_2\right)=E\left(X_1\cdot X_2\right)-E\left(X_1\right)E\left(X_2\right)=$$
$$=P\left(X_1=1,X_2=1\right)-P\left(X_1=1\right)E\left(X_2=1\right)=$$
$$=0.1-0.6\cdot 0.4=-0.14$$

ב. כמובן שהמ"מ המקריים תלויים מאחר ואם יאיר לא שר בממשלה אז ציפי בוודאות גם לא. ניתן גם להראות זאת מפורשות על ידי:

$$.P\left(X_1=0,X_2=1\right)=0\neq P\left(X_1=0\right)P\left(X_2=1\right)$$

בכדי לקבוע אם הם מתואמים או לא חייבים לחשב מפורשות, מאחר ומשתנים מקריים יכולים להיות תלויים ויחד עם זאת לא מתואמים (תזכרו שהם מ"מ הם בלתי תלויים אז הם בהכרח בלתי מתואמים אבל ההיפך אינו נכון!!! ולכן חייבים לבדוק אם הם מתואמים בצורה מפורשת ע"י חישוב).

$$COV\left(X_1,X_3\right)=E\left(X_1X_3\right)-E\left(X_1\right)E\left(X_3\right)=P\left(X_1=1,X_3=1\right)-P\left(X_1=1\right)P\left(X_3=1\right)=$$
$$=0.2-0.2\cdot 0.6=0.08$$

ולכן המ"מ הם מתואמים.

ג. פשוט משתמשים בתכונת הלינאריות של שונות משותפת:

$$COV(X_1+X_3,X_2)=COV\left(X_1,X_2\right)+COV\left(X_3,X_2\right)=-0.14+0=-0.14$$

השונות המשותפת של X_2,X_3 היא 0 מאחר והם בלתי תלויים ולכן בלתי מתואמים, ואת $COV\left(X_1,X_2\right)$ כבר חישבנו בסעיף א'.

<u>תרגיל 5</u>: מחפיסת קלפים טרופה היטב הוצאו 13 קלפים בזה אחר זה ללא החזרה למשחק פוקר. יהי X – מספר הקלפים ב-"יד" מסוג "תמונה". (בחבילה ישנם 52 קלפים, 12 תמונות שכולל נסיך, מלכה ומלך מכל צורה והיתר הם מספרים, 1 עד 10).

א. הציגו פירוק של X לסכום של משתנים מצביעים $X_1,\dots,X_{13}$.

ב. רשמו את ההתפלגות המשותפת של $\left(X_i,X_j\right)$.

ג. חשבו את $COV\left(X_i,X_j\right),\rho\left(X_i,X_j\right)$ עבור $i\neq j$ והסבירו את הסימן.

ד. חזרו על צמד הסעיפים עבור $i=j$.

ה. הסיקו תוחלת ושונות של X והשוו לתוחלת והשונות של המשתנה Y המתקבל מהוצאה של 13 קלפים עם החזרה מחפיסה. הסבירו את השווה ואת השונה.

ו. כיצד מתפלג המשתנה X אם ידוע לכם ש-$X_1+X_2+X_3=0$. נמקו. מהו הניבוי הטוב ביותר במקרה הזה (הפסד ריבועי)?

ז. מהו הניבוי הטוב ביותר (הפסד ריבועי) בעבור המשתנה X, אם ידוע לכם ש-$X_1+X_2+X_3=k$, $k=0,1,2,3$. האם התקבלה פונקציה ניבוי ליניארית ב-k?

<u>פתרון תרגיל 5</u>:

א. נגדיר $X_i=\begin{cases}1 & Card\ i\ is\ an\ image\\ 0 & Otherwise\end{cases}$ ולכן $X=\sum_{i=1}^{13}X_i$.

ב. נרשום את ההתפלגות המשותפת בעזרת טבלה.

$X_i \setminus X_j$	0	1	
0	$\frac{40}{52}\cdot\frac{39}{51}$	$\frac{40}{52}\cdot\frac{12}{51}$	$\frac{10}{13}$
1	$\frac{40}{52}\cdot\frac{12}{51}$	$\frac{12}{52}\cdot\frac{11}{51}$	$\frac{3}{13}$
	$\frac{10}{13}$	$\frac{3}{13}$	

ג. נמצא בעזרת חישוב ישיר -

$$COV\left(X_i,X_j\right)=E\left(X_i\cdot X_j\right)-E\left(X_i\right)\cdot E\left(X_j\right)=\tfrac{12}{52}\cdot\tfrac{11}{51}-\tfrac{3}{13}\cdot\tfrac{3}{13}=$$
$$=\tfrac{3}{13}\left(\tfrac{11\cdot 13-3\cdot 51}{51\cdot 13}\right)=\tfrac{-10}{2873}$$

$$\rho\left(X_i,X_j\right)=\frac{COV\left(X_i,X_j\right)}{\sqrt{V\left(X_i\right)V\left(X_j\right)}}=\frac{\tfrac{-10}{2873}}{\sqrt{\tfrac{3}{13}\left(1-\tfrac{3}{13}\right)\cdot\tfrac{3}{13}\left(1-\tfrac{3}{13}\right)}}=-\frac{1}{51}$$

ד. עבור $i=j$ הפתרונות הם מיידים מאחר ואת טבלת ההתפלגות המשותפת ניתן לרשום באופן הבא,

$X_i \setminus X_i$	**0**	**1**	
0	$\frac{10}{13}$	0	$\frac{10}{13}$
1	0	$\frac{3}{13}$	$\frac{3}{13}$
	$\frac{10}{13}$	$\frac{3}{13}$	

בנוסף, נשים לב כי $\frac{3}{13}\cdot\frac{10}{13}$ $.\rho\left(X_i,X_i\right)=1, COV\left(X_i,X_j\right)=V\left(X_i\right)=\frac{3}{13}\cdot\frac{10}{13}$

ה. X הוא משתנה מקרי היפרגאומטרי עם אוכלוסיה בגודל 52, מספר מיוחדים 12 וגודל מדגם 13 בעוד Y הוא משתנה מקרי בינומי עם 13 ניסויים ב"ת והסתברות $\frac{3}{13}$(בכדי למצוא את התוחלת והשונות ניתן פשוט לקחת מדף הנוסחאות את התוחלת והשונות של משתנים מקריים שכאלו עם הפרמטרים הנתונים).

ו. למעשה נתון לנו כי שלושת הקלפים הראשונים הם אינם תמונות ולכן כל שנותר לבחור זה 10 קלפים מתוך 49 שעדיין בחפיסה כאשר ישנם 12 תמונות, ניתן להסיק מכך ש- $X \mid X_1+X_2+X_3=0$ מתפלג היפרגאומטרית עם אוכלוסיה של 49 קלפים, 12 מיוחדים ומדגם בגודל 10,
$X \mid X_1+X_2+X_3=0 \sim HG\left(49,12,10\right)$. התוחלת היא כמובן $10\cdot\frac{12}{49}$.

ז. נוכל להרחיב את תוצאת הסעיף הקודם באופן הבא – אם הסכום שווה ל-K זה אומר שישנם K תמונות לפחות ואת היתר עוד נותר לבחור מתוך הקלפים שנותרו ולכן התוחלת תהיה $k+10\cdot\frac{12-k}{49}$ שזה k שבהכרח יש לנו ובנוסף תוחלת של משתנה מקרי היפרגאומטרי עם אוכלוסיה בגודל 49, 12-k פריטים מיוחדים (קלפים של תמונה) ומדגם בגודל 10.