תקציר

אולי מפני התאבון להימורים אשר גרם לפיתוח של התורת ההסתברות. במאמץ להגדיל את זכיותיהם, מהמרים פנו למתמטיקאים לספק אסטרוטגיות הכי טובות במשחקי סיכוי. חלק של המתמטיקאים האלה כולל ליבניץ (Leibniz) פרמט (Fermat) ברנולי (Bernoulli) ועוד. כתוצאה של ההתפתחות של תורת הסתברות, עם כל התחזיות וכללים, מסקנה סטטיסטית התפשט בהרבה לתחומים מעבר משחקי סיכוי להכליל מגוון תחומים קשורים למקרי סיכוי, למשל פוליטיקה, שוק ההון, תחזית מזג האוויר, מחקר מדע, ועוד.

מבוא להסתברות וסטטיסטיקה

תוכן העניינים

[block]שיעור 0
1em

[block]1em

[block]0 1em

1 מבוא להסתברות ולוגיקה 27-6

הגדרת המושג הסתברות

מה המשמעות של המשפט "שמואל קרוב לוודא ינצח את המשחק כדור-רגל " או " יש לי סיכוי של 50-50 לקבל מספר זוגי כאשר אני זורק קוביה" או " יש סיכוי קטן שאני אנצח בלוטו" או " רוב התלמידים בכיתה יעבור את המבחן". בכל אחד של המצבים האלה אנחנו מבטאים במילים את **ההסתברות** אשר תוצאה מסויימת יתרחש. בקורס הזה אנחנו נותנים שיטות של נוסחאות כדי לחשב את ההסתברות של מקרה מדובר כמספר.

תורת הסתברות נותן מספר לסבירות של תוצאה של ניסוי. לדוגמה, ההסתברות לזרוק מספר 2 בהטלת קוביה הוגנת. יש 6 אפשרויות, ואין שום סיבה להניח שלאחד מהתוצאות יש סיכוי יותר ופחות מהשני. לכן ההסתברות לזרוק 2 היא $\frac{1}{6}$ יש דרך אחר להסתכל על זה, אם נבצע אותו ניסוי הרבה פעמים, נגלה ש ב אחד מתוך שש פעמים נקבל 2.

הקבוצה של כל התוצאות של הניסוי היא

$$\Omega = \{1, 2, 3, 4, 5, 6\} . \tag{1.1}$$

(1.1) הוא דוגמה של דבר אשר נקרא מרחב מדגם. אם על המטבע נמצאים הסימונים המוכרים של T ו- H (H מציין את התוצאה "עץ" בעוד T מתייחס לתוצאה "פלי") אזי מרחב המדגם יהיה

$$\Omega = \{H, T\} . \tag{1.2}$$

מרחב המדגם זו גם מורכב מתוצאות אשר יש להם סבירויות שוות, אזי לכל תוצאה ב Ω יש סבירות של $\frac{1}{2}$. או אחרי לבצע אותו ניסוי הרבה פעמים, חצי של הזמן נקבל H וחצי של התוצאות יהיו T.

1.1 הגדרה. (מרחב מדגם)

מרחב מדגם זו קבוצה, המסומנת לרוב באות היוונית Ω (אומגה), המכילה את כל התוצאות האפשריות של הניסוי.

בקורס הזה נתעניין במרחבי מדגם המורכבים ממספר תוצאות סופי. הדוגמאות לעיל הם של מרחב מדגם אשר בו תוצאות של סבירויות שוות, אבל באופן כללי לתוצאות יש סבירויות אי-שוות. יש מספר תכונות אשר מאפיינות את המושג של סבירות של תוצאה של ניסוי:

- לכל תוצאה במרחב מדגם נותנים מספר בין 0 ל1 אשר מאפיין את הסבירות שלו.
- ככל שהסיכוי של תוצאה יותר גדול אז הסבירותו שלו יותר קרוב ל 1.
- ככל שהסיכוי של תוצאה יותר קטן אז הסבירותו שלו יותר קרוב ל 0.
- לתוצאה אי-אפשרית, יש הסתברות של 0.

- לתוצאה שיהתרחש בוודאות, יש הסתברות של 1.
- הסכום של הסבירויות של כל התוצאות בשום מרחב מדגם שווה ל 1 (נרמול הסתברות).

1.2 הגדרה. (מאורע)

מאורע הוא תת קבוצה של מרחב מדגם.

1.3 דוגמא. (מאורע) נתון המרחב המדגם $\Omega = \{HH, TH, HT, TT\}$ של תוצאות של זריקת מטבע פעמיים ברצף, המאורע של לקבל H לפחות פעם אחת הוא $A = \{HH, TH, HT\}$. דרך פורמלי לבטא זה הוא

$$A \subset \Omega.$$

תורת הקבוצות לוגיקה

ניקח את האיבר 2 ואת המאורע $A = \{1, 2, 3\}$. נבחין כי האיבר 2 הוא איבר במאורע A. על כן "2" **הוא שייך למאורע** A. נסמן זאת ב

$$2 \in A.$$

שונה לזה "4" אינו נמצא ב-A, ולכן **אינו שייך ל-A** :

$$4 \notin A.$$

1.4 הגדרה. (ההכלה) מאורע A מוכל במאורע אם כל האיברים של A שייכים גם ל-B.

במילים פשוטות, אם כל האיברים של A נמצאים ב-B אז המאורע A מוכל במאורע B:

$$A \subseteq B.$$

דוגמא. אם הניסוי הוא הטלת קוביה ומאורע $A = \{6\}$ מציין את התוצאה 6 בעוד מאורע $B = \{4, 5, 6\}$ מציין שהתוצאה גדולה או שווה ל 4, אז ברגע ש A מתרחש גם B מתרחש, ו

$$A \subseteq B.$$

נשים לב כי $B = A$ זה מקרי פרטי של הכלה, כי כל איברי A נמצאים ב-B ולהפך:

$$A \subseteq B \quad \text{ו} \quad B \subseteq A \quad \Leftrightarrow A = B.$$

ברגע שיחס ההכלה הנ"ל $B \subseteq A$ מתקיים נאמר כי מאורע A הוא תת קבוצה של B.

1.5 הגדרה. (המשלים המאורע)

המשלים של המאורע A ביחס ל Ω הוא התת קבוצה של כל האיברים שנמצאים ב A אך אינם נמצאים ב A.

דוגמא. אם R הוא המאורע שקלף אדום נלקח מחבילת רגילה של 52 קלפים, ואם Ω הוא המרחב המדגם המורכב מהחבילה של כל הקלפים. לכן המשלים המאורע R' הוא המאורע כי הקלף הנלקח מהחבילה אינו אדום, אלא קלף שחור.

דוגמא. נתון המרחב המדגם

$$\Omega = \{\text{Dropbox}, \text{f}, \text{G}, \text{in}, \text{WhatsApp}, \text{YouTube}\}$$

והתת קבוצה

$$A = \{\text{Dropbox}, \text{G}, \text{in}, \text{YouTube}\}.$$

הקבוצה המשלימה של A היא

$$A' = \{\text{f}, \text{WhatsApp}\}.$$

1.6 הגדרה. (החיתוך בין מאורעות)

החיתוך $A\cap B$ של צמד המאורעות זו קבוצה שמכילה את כל האיברים שנמצאים הן ב- A והן ב- B.

דוגמא. החיתוך בין המאורעות

$$A=\{\text{א},\text{ש},\text{ק},\text{ל},\text{ו},\text{ן}\} \quad \text{ו} \quad B=\{\text{י},\text{ר},\text{ו},\text{ש},\text{ל},\text{י},\text{ם}\}$$

הוא

$$A\cap B=\{\text{ש},\text{ל},\text{ו}\}\,.$$

דוגמא. החיתוך בין המאורעות

$$A=\{1,2,3,5,7,11,13,17\} \quad \text{ו} \quad B=\{2,4,6,8,10\}$$

הוא

$$A\cap B=\{2\}\,.$$

בעוד מרחב המדגם הוא המאורע אשר מכיל את כל התוצאות האפשריות בניסוי, עולה השאלה כיצד נראה מאורע קיצוני אחר אשר לא מכיל אף תוצאה? עבור מצבים כאלה אנו מגדירים את **הקבוצה הריקה** או לחילופין **המאורע הריק** והיא מסומנת ב- ϕ:

1.7 הגדרה. (הקבוצה הריקה או המאורע הריק)

הקבוצה הריקה זו הקבוצה

$$\phi=\{\ \}\,.$$

המאורע הריק הוא מקרה של הקבוצה הריקה, אשר הוא מאורע שאין בו תוצאות אפשריות.

1.8 הגדרה. (מאורעות זרים)

מאורעות A ו-B נקראים זרים זה לזה אם אין להם איברים משותפים, ולכן

$$A\cap B=\phi.$$

דוגמא. אם O הוא הקבוצה המורכב מן המספרים האי זוגיים מ1 עד 10 וE הקבוצה המורכב מהמספרים הזוגיים מ0 עד 10, כלומר

$$E=\{2,4,6,8,10\} \qquad O=\{1,3,5,7,9\}\,,$$

לכן קל לראות שהמאורעות האלה הם מאורעות זרים זה לזה בגלל

$$E\cap O=\phi.$$

1.9 הגדרה. (האיחוד)

האיחוד של שתי המאורעות A ו-B המסומן ב $A\cup B$ זו המאורע הכולל את כל האיברים אשר שייכים או ל A או לB או גם לA וגם לB.

דוגמא. אם

$$A=\{a,b,c\} \quad \text{ו} \quad \{b,c,d,e\}$$

אזי

$$A\cup B=\{a,b,c,d,e\}\,.$$

דוגמא. אם

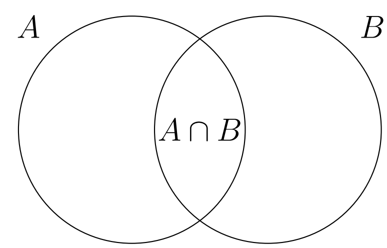
$$M = \{x \mid 3 < x < 9\} \quad \text{ו} \quad N = \{y \mid 5 < y < 12\}$$

אזי

$$M \cup N = \{z \mid 3 < z < 12\}\,.$$

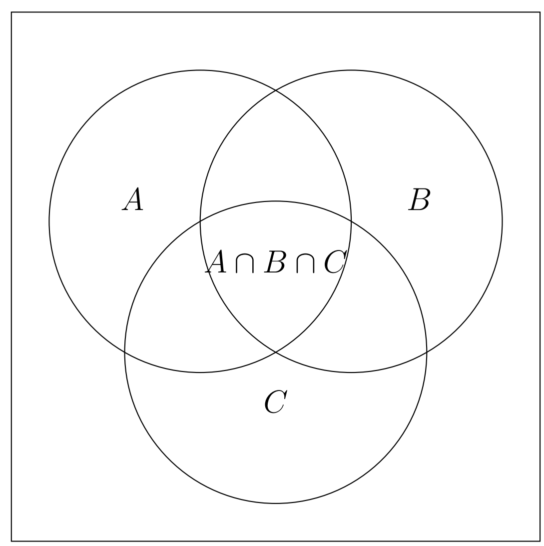
1.10 חוק. (חוק הקיבוץ) חוק הקיבוץ קובע כי סדר כתיבת המאורעות באיחוד או בחיתוך אינו משפיע על התוצאה סופית:

$$\begin{aligned} A \cap B =& B \cap A, \\ A \cup B =& B \cup A. \end{aligned}$$



1.11 חוק. (חוק החילוף) חוק החילוף קובע כי לכל שלושה מאורעות A, B, C מתקיים

$$\begin{aligned} (A \cap B) \cap C =& A \cap (B \cap C), \\ (A \cup B) \cup C =& A \cup (B \cup C). \end{aligned}$$



1.12 חוק. (חוק הפילוג) חוק הפילוג קובע כי לכל שלושה מאורעות A, B, C מתקיים

$$\begin{aligned} (A \cup B) \cap C =& (A \cap C) \cup (B \cap C), \\ (A \cap B) \cup C =& (A \cup C) \cap (B \cup C). \end{aligned}$$

1.13 הגדרה. (הפרש בין מאורעות) ההפרש בין מאורעות הוא הפעולה אשר לוקחת צמד מאורעות A ו-B ולוקחת את כל איברי A ומורידים מהם את האיברים המשותפים ל A ו-B:

$$A/B.$$

במידה ואין איברים משותפים,

$$A/B = A.$$

דוגמא. אם $A = \{1, 2\}$ ו $B = \{1, 3, 6\}$,

$$A/B = \{2\}$$

1.14 מסקנה. (ההפרש בין מאורע ומדגם מרחב)

$$\bar{A} = S/A.$$

דוגמא. בן מטיל קוביה הוגנת.

1. רשמו את מרחב המדגם
2. רשמו את המאורעות הבאים:
   (א) A התוצאות קטנה מ 4 ,
   (ב) B התוצאות גדולה או שווה ל 3,
   (ג) C התוצאות זוגית,
   (ד) D התוצאות אי זוגית,
3. האם $4 \in A$? האם $3 \in B$?
4. רשמו מפורשות את המאורעות הבאים:
   (א) $A \cap B$,
   (ב) $A \cup B$,
   (ג) $C \cap B$,
   (ד) $(A \cap B) \cup C$.

פיתרון. 1. $S = \{1, 2, 3, 4, 5, 6\}$.

2. (א) $A = \{1, 2, 3\}$.
   (ב) $B = \{3, 4, 5, 6\}$.
   (ג) $C = \{2, 4, 6\}$.
   (ד) $D = \{1, 3, 5\}$.
3. $A \notin 4$ לא .
   $B \in 3$ כן.
4. (א) $A \cap B = \{3\}$,
   (ב) $A \cup B = \{1, 2, 3, 4, 5, 6\}$,
   (ג) $C \cap B = \{3, 4, 6\}$,
   (ד) $(A \cap B) \cup C = A \cap (B \cap C) = \{1, 2, 3\} \cap \{4, 6\} = \phi$.

■

דוגמא. ניקח את קבוצות הסטודנטים הנמצאים בכיתה ונגדיר את המאורעות הבאים:

- הסטודנטים שאוהבים חתולים $= C$,
- הסטודנטים שאוהבים כלבים $= D$,
- הסטודנטים שאוהבים דגים $= F$.

רשמו את המאורעות הבאים:

1. A_1 - הסטודנטים שאוהבים לפחות חיה אחת.
2. A_2 - הסטודנטים שלא אוהבים אף חיה.
3. A_3 - הסטודנטים שאוהבים רק חתולים.
4. A_4 - הסטודנטים שאוהבים את כל החיות.
5. A_5 - הסטודנטים שאוהבים בעל חיים אחד בלבד.
6. A_6 - הסטודנטים שאוהבים לפחות 2 בעלי חיים.

פיתרון. 1. $A_1 = C \cup D \cup F$.

2. $A_2 = \bar{A}_1 = \bar{C} \cap \bar{D} \cap \bar{F}$.

3. $A_3 = C \cap \bar{D} \cap \bar{F} = (C/D)/F$.
4. $A_4 = C \cap D \cap F$.
5. $A_5 = (C \cap \bar{D} \cap \bar{F}) \cup (\bar{C} \cap D \cap \bar{F}) \cup (\bar{C} \cap \bar{D} \cap F)$.
6. $A_6 = (C \cap D) \cup (D \cap F) \cup (C \cap F)$.

■

חישובים של הסתברות

כדי לחשב את ההסתברות של מאורע A, לוקחים את הסכום של כל דגימה בA. הסכום הזו נקרא ההסתברות של A והוא מסומן ב $P(A)$:

1.15 הגדרה. (הסתברות של מאורע)

ההסתברות של מאורע A זו הסכום של המשקלות של כל האיברים ב A. לכן

$$0 \leq P(A) \leq 1, \qquad P(\phi) = 0, \qquad P(\Omega) = 1.$$

עוד, אם $A_1, A_2, A_3, \ldots$ הוא סידרה של מאורעות זרים, אזי

$$P(A_1 \cup A_2 \cup A_3 \cup \ldots) = P(A_1) + P(A_2) + P(A_3) + \cdots \ .$$

1.16 דוגמא. מהו הסיכוי שבהטלת כפולה של מטבע הוגן נקבל לפחות H אחת?

פיתרון. המרחב המדגם לניסוי הזה הוא

$$S = \{HH, HT, TH, TT\}\,.$$

נניח שהמטבע הוגן. לכן לכל אחת מהתוצאות האלה יש סיכוי שווה. נסמן את ההסתברות של כל תוצאה ב ω. אזי

$$4\omega = 1 \qquad \Rightarrow \omega = \frac{1}{4}.$$

נסמן את המאורע שנקבל לפחות H אחת ב- A.

$$A = \{HH, HT, TH\}\,,$$
$$P(A) = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4}.$$

■

דוגמא. קוביה משוקלת באופן כך שיש סיכוי לזרוק מספר זוגי פי שתיים מהסיכוי לזרוק מספר אי זוגי. נסמן ב- E את המאורע לזרוק מספר פחות מ 4. מהי $P(E)$?

פיתרון. המרחב המדגם הוא

$$S = \{1, 2, 3, 4, 5, 6\}\,.$$

נותנים הסתברות של w לכל מספר אי-זוגי והסתברות $2w$ לכל מספר זוגי. הסכום של ההסתברויות שווה ל 1. לכן

$$3(2w) + 3w = 6w + 3w = 9w = 1. \qquad \Rightarrow \qquad w = \frac{1}{9}.$$

למאורע E יש את האיברים

$$E = \{1, 2, 3\}\,.$$

אזי

$$P(E) = w + 2w + w = 4w = \frac{4}{9}.$$

■

דוגמא. אם A זו המאורע לזרוק מספר זוגי ו B המאורע לזרוק מספר אשר מתחלק ב-3. חפשו $P(A\cap B)$ ו $P(A\cup B)$.

פיתרון.

$$A=\{2,4,6\}\,,\qquad B=\{3,6\}\,.$$

$$A\cup B=\{2,3,4,6\}\,.$$

$$A\cap B=\{6\}\,.$$

לכל מספר זוגי יש הסתברות של $w=\frac{2}{9}$ ולכל מספר אי-זוגי יש הסתברות של $w=\frac{1}{9}$. אזי

$$P(A\cup B)=\frac{2}{9}+\frac{1}{9}+\frac{2}{9}+\frac{2}{9}=\frac{7}{9}.$$

$$P(A\cap B)=\frac{2}{9}.$$

■

1.17 חוק. (הסתברויות שוות) אם יש לניסוי N תוצאות ויש לכל תוצאה סיכויים שווים, ויש למאורע- A n תוצאות. אזי ההסתברות של A הוא

$$P(A)=\frac{n}{N}.$$

דוגמא. רשמו את המרחב מדגם והמאורעות המצויינים בהמשך במונחי תורת הקבוצות:

1. הוצאה של כדור מתוך כד בו יש 2 כדורים לבנים, וכדור שחור. A- הוצא כדור שחור.
2. הוצאה של כדור מתוך כד בו יש 2 כדורים לבנים הממוספרים ב $1,2$ וכדור שחור. A- הוצא כדור לבן.
3. הוצאה עם החזרה של 2 כדורים מתוך כד בו יש 2 כדורים לבנים, ו-1 שחור. A- הכדור השני שהוצא איננו לבן.
4. הוצאה ללא החזרה של כדור מתוך כד בו יש 2 כדורים לבנים, ו-1 שחור. A- הכדור השני שהוצא הוא לבן.

פיתרון.

1. מרחב המדגם בהוצאה של כדור מתוך כד בו יש 2 כדורים לבנים, וכדור שחור.

$$\Omega=\{w,b\}\,.$$

המאורע נתון על ידי

$$A=\{b\}\,.$$

2. מרחב המדגם בהוצאה של כדור מתוך כד בו יש 2 כדורים לבנים הממוספרים ב $1,2$ וכדור שחור הוא

$$\Omega=\{w_1,w_2,b\}$$
$$A=\{w_1,w_2\}\,.$$

3.

$$\Omega=\{(w,w),(w,b),(b,w),(b,b)\}$$
$$A=\{(w,b),(b,b)\}\,.$$

.4

$$\Omega = \{(w,w),(w,b),(b,w)\}$$
$$A = \{(w,w),(b,w)\}\,.$$

■

תכונות בסיסיות של ההסתברות

(קולמוגורוב, תחילת המאה ה-20).

1. **האקסיומה אי-שלילית** קובעת כי הסתברות איננה שלילית:

$$P(A) \geq 0. \tag{1.3}$$

2. **האקסיומה הנרמול** קובעת כי

$$P(\Omega) = 1 \tag{1.4}$$

כאשר Ω המרחב המדגם במלואו.

3. **תכונה החיבורית**

$$\sum_{\omega \in A} P(\omega) = P(A). \tag{1.5}$$

במילים פשוטות, ההסתברות של מאורע נתונה על ידי סכום ההסתברויות של האפשריות הנמצאות באותו המאורע.

1.18 מסקנה. (לקבוצה הריקה יש הסתברות 0) מתכונת החיבורית (1.5) אנחנו מסיקים כי ההסתברות של הקבוצה הריקה היא אפס:

$$P(\phi) = 0. \tag{1.6}$$

1.19 דוגמא. () נתון מרחב המדגם

$$\Omega = \{1,2,\ldots,10\}$$

ונתון ההסתברויות

$$P(i) = ci^2 \qquad \forall\ i \in S,$$

כאשר c הוא קבוע כלשהו. מצאו את ערכו של c וחשבו את ההסתברויות של תוצאות הניסוי תתחלק ב- 3.

פיתרון. נוכל למצוא את הקבוע c מתנאי הנרמול. בפרט אנחנו יודעים שסכום ההסתברויות חייב להיות 1. לכן

$$1 = \sum_{i=1}^{10} P(i) = \sum_{i=1}^{10} c\ i^2 = \frac{c(10)(21)(11)}{6} = 385c.$$

מכאן נובע ש

$$c = \frac{1}{385}.$$

המאורע המבקש בשאלה (תוצאות הניסוי מתחלקת ב- 3) הוא

$$A = \{3,6,9\}$$

מחיבוריות נקבל

$$P(A) = P(3) + P(6) + P(9) = \frac{3^2+6^2+9^2}{385} = \frac{126}{385}.$$

■

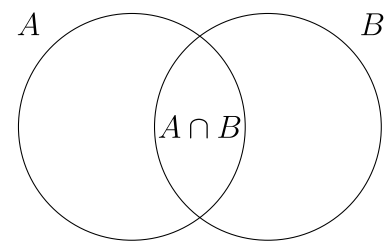
חוקי הסתברות בסיסיים

נתפנה להוכיח מספר תכונות בסיסיות של הסתברות.

1.20 חוק. (נוסחת ההכלה וההפרדה) נוסחת ההכלה וההפרדה, או לעיתים נקרא **חוק החיבורית**. הנוסחה קובעת כי לכל צמד מאורעות A-B מתקיים

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) . \tag{1.7}$$

הוכחה. הדרך הכי הקלה להמחיש את נוסחת ההכלה וההפרדה היא בעזרת הדיאגרמת ואן. הסתכלו אל הדיאגרמת ואן להלן:



$P(A \cup B) =$ סכום של הסתברויות של נקודות דגימה ב $A \cup B$.

$P(A) + P(B) =$ סכום של ההסתברויות ב-A פלוס סכום של ההסתברויות ב-B.

לכן, הוספנו את ההסתברויות ב $A \cap B$ פעמיים, ולכן יש צורך להפחית $P(A \cap B)$ מהסכום כדי להגיע אל משוואה (1.7). ■

דוגמא. בניסוי הטלת קוביה הוגנת נגדיר המאורעות

$$A = \{1, 2, 3\} \qquad B = \{2, 3, 4\} .$$

אזי

$$P(A \cap B) = P(\{2, 3\}) = \frac{2}{6},$$

$$P(A) = \frac{3}{6}, \qquad P(B) = \frac{3}{6}.$$

$$P(A \cup B) = P(\{1, 2, 3, 4\}) = \frac{4}{6},$$

$$P(A) + P(B) - P(A \cap B) = \frac{3}{6} + \frac{3}{6} - \frac{2}{6} = \frac{4}{6}.$$

מסקנה. () אם A ו- B זרים אז

$$P(A \cap B) = P(\phi) = 0.$$

לכן, נוסחת ההכלה וההפרדה היא הרחבה ישירה של תכונה החיבורית:

$$P(A \cup B) = P(A) + P(B) .$$

מסקנה. () אם

$$A_1, A_2, \ldots, A_n$$

זרים, אז

$$P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + \ldots + P(A_n).$$

או

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i).$$

הגדרה. (מדגם הפרדה) אוסף המאורעות $\{A_1,\dots,A_n\}$ של מרחב מדגם הוא הפרדה של Ω אם ורק אם $A_1,\dots,A_n$ זרים, כלומר

$$A_1 \cup A_2 \cup \ldots \cup A_n = \Omega \qquad \left(\bigcup_{i=1}^{n} A_i = \Omega\right).$$

לכן

1.21 מסקנה. () אם $A_1,\dots,A_n$ זו הרפדה של מרחב מדגם Ω, אזי

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P\left(A_i\right) = P(\Omega) = 1.$$

1.22 תורת. () עבור שלושה מאורעות A, B, C נוסחת ההכלה וההפרדה קובעת כי

$$\begin{aligned}
P(A\cup B\cup C) =&P(A\cup B)+P(C)-P((A\cup B)\cap C)\\
=&P(A)+P(B)-P(A\cap B)+P(C)-P((A\cap C)\cup(B\cap C))\\
=&P(A)+P(B)+P(C)-P(A\cap B)-P(A\cap B)-P(A\cap C)-P(B\cap C)\\
&+P(A\cap C\cap B\cap C)\\
=&P(A)+P(B)+P(C)-P(A\cap B)-P(A\cap B)-P(A\cap C)-P(B\cap C)+P(A\cap B\cap C).
\end{aligned}$$

הוכחה. ■

דוגמא. בניסוי הטלת שתי קוביות הוגנות, מהי ההסתברות לזרוק 7 או 11?

פיתרון. נסמן ב- A המאורע לזרוק 7 ונסמן ב- B המאורע לזרוק 11. זריקת 7 מתרחש ב 6 מתוך 36 של הנקודות במרחב מדגם. 11 מתרחש ב 2 מתוך 36 של הנקודות במרחב מדגם. מכיוון שלכל הנקודות יש סכוי שווה, אזי

$$P(A) = \frac{6}{36} = \frac{1}{6}, \qquad P(B) = \frac{2}{36} = \frac{1}{18}.$$

המאורעות A ו-B זרים, (אי-אפשר לזרוק 7 באותו זמן של לזרוק 11). לכן,

$$P(A\cup B) = P(A)+P(B) = \frac{1}{6}+\frac{1}{18} = \frac{2}{9}.$$

■

2 29-6

עוד חוקי הסתברות בסיסיים

2.1 חוק. (נוסחת ההכלה וההפרדה) לכל צמד מאורעות A ו- B מתקיים

$$P(A\cup B)=P(A)+P(B)-P(A\cap B). \tag{2.1}$$

חוק. (הסתברות של איחוד של שלושה מאורעות)
אם A ו-B זרים אז

$$P(A\cap B)=P(\phi)=0. \tag{2.2}$$

חוק. (הסתברות של מאורעות זרים) עבור שלושה מאורעות A, B, C מתקיים

$$P(A\cup B\cup C)=P(A)+P(B)+P(C)-P(A\cap B)-P(B\cap C)-P(C\cap A)+P(A\cap B\cap C). \tag{2.3}$$

חוק. (הסתברויות המשלימות) אם $\bar{A}$ המאורע משלים של A אז

$$P(A)+P(\bar{A})=1, \qquad \Leftrightarrow \qquad P(\bar{A})=1-P(A). \tag{2.4}$$

דוגמא. בהצגה לכבוד פורים מככבים אסתר, מרדכי והמן הרשע. לכל אחד מכוכבי ההצגה יש שחקן מחליף למקרה של היעדרות. אסתר נעדרת ב- 40% מההופעות, מרדכי ב 50% מההופעות והמן ב 35% מההופעות. ידוע שכל שניים מהכוכבים נעדרים יחדיו ב 15% מההופעות וכולם ביחד נעדרים ב 5%. מההופעות. חשבו את ההסתברות שבהצגה אליה קניתם כרטיסים לא יהיה אף שחקן מחליף.

פיתרון. נסמן:

$A=$ אסתר נעדרת
$B=$ מרדכי נעדר
$C=$ המן נעדר

נתון כי

$$\begin{aligned}P(A)=&0.4,\\ P(B)=&0.5,\\ P(C)=&0.35,\\ P(A\cap B)=&P(B\cap C)=P(C\cap A)=0.15, P(A\cap B\cap C)=0.05.\end{aligned}$$

המאורע שלא יהיה אף שחקן הוא $\bar{A}\cap\bar{B}\cap\bar{C}$.

חוקי דה מורגן

$$\begin{aligned}\overline{A\cup B}=&\bar{A}\cap\bar{B},\\ \overline{A\cap B}=&\bar{A}\cup\bar{B},\\ \overline{\bigcup_{i=1}^{n}A_i}=&\bigcap_{i=1}^{n}\bar{A}_i,\\ \overline{\bigcap_{i=1}^{n}A_i}=&\bigcup_{i=1}^{n}\bar{A}_i.\end{aligned} \tag{2.5}$$

לפי חוקי דה מורגן

$$\bar{A}\cap\bar{B}\cap\bar{C}=\overline{A\cup B\cup C}.$$

לכן

$$P(\bar{A}\cap\bar{B}\cap\bar{C})=P(\overline{A\cup B\cup C})=1-P(A\cup B\cup C).$$

לפי (2.3),

$$\begin{aligned}P(A\cup B\cup C)=&P(A)+P(B)+P(C)-P(A\cap B)-P(B\cap C)-P(C\cap A)+P(A\cap B\cap C)\\=&0.4+0.5+0.35-0.15-0.15-0.15+0.05\\=&0.85.\end{aligned}$$

לכן

$$P(\bar{A}\cap\bar{B}\cap\bar{C})=1-0.85=0.15.$$

■

דוגמא. בסקר שנערך בעיר מסוימת נמצא ש־ 60% מהתושבים מגדלים כלב. בנוסף, 30% מהתושבים מגדלים חתולים ו־ 15% מגדלים גם כלב וגם חתול. חחפשו את ההסתברות שתושב מקרי

1. מגדל לפחות בעל חיים אחד
2. לא מגדל 2 בעלי חיים
3. מגדל כלב, אך לא חתול
4. מגדל חתול, אך לא כלב.

פיתרון.

$C=$ המאורע של בעלי כלבים,

$D=$ המאורע של בעלי חתולים.

.1

$$\begin{aligned}P(C\cup D)=&P(C)+P(D)-P(C\cap D)\\=&0.6+0.3-0.15\\=&0.75.\end{aligned}$$

.2

$$\begin{aligned}P(\overline{C\cap D})=&1-P(C\cap D)\\=&1-0.15\\=&0.85.\end{aligned}$$

.3 יש צורך להשתמש בהחוקים

$$P(B)=P(B\cap A)+P(B\cap\bar{A}), \tag{2.6}$$

$$P(B\cap\bar{A})=P(B)-P(B\cap A). \tag{2.7}$$

אזי

$$\begin{aligned}P(C\cap\bar{D})=&P(C)-P(C\cap D)\\=&0.6-0.15\\=&0.45.\end{aligned}$$

.4

$$\begin{aligned}P(\bar{C}\cap D)=&P(D)-P(C\cap D)\\=&0.3-0.15\\=&0.15.\end{aligned}$$

■

2.1 הגדרה. (מרחב מדגם אחיד) מרחב מדגם נקרא אחיד אם לכל איבר $\omega \in \Omega$ במרחב מתקיים

$$P(\omega) = \frac{1}{|\Omega|}, \tag{2.8}$$

כאשר

$|\Omega| = $ # נקודות ב- Ω.

2.2 דוגמא. (מרחב מדגם אחיד) לדוגמא לניסוי הטלת קוביה הוגנת יש מרחב מדגם Ω כאשר $\Omega = \{1, \ldots, 6\}$ ומתקיים כי $P(\omega \in \Omega) = \frac{1}{6}$. על כן Ω הוא מרחב מדגם אחיד.

2.3 דוגמא. (מרחב מדגם לא סימטרי) בכד נמצא 12 כדורים: שחור אחת, לבן 2, כחול 3, אדום 4 וירוק-5. המרחב מדגם הוא

$$\Omega = \{bk, w, r, bl, g\}.$$

ומתקיים כי

$$P(bk) = \frac{1}{12}, \quad P(w) = \frac{2}{12}, \quad P(r) = \frac{3}{12}, \quad P(bl) = \frac{4}{12}, \quad P(g) = \frac{5}{12}.$$

על כן Ω הוא מרחב מדגם אי-סימטרי.

עקרון הכפל

כדי לחשב ההסתברות של מאורע, יש צורך לדעת כיצד לספור את כמות האפשרויות בכל מאורע. עקרון הכפל מסייע לנו לקבוע את המספר האפשרויות כאשר ישנו תהליך המתבצע בשלבים. בכדי להבין את עיקרון הכפל כראוי, נעבור על כמה דוגמאות:

- מספר התוצאות האפשריות בהטלת מטבע פעמיים הוא 4 $(\Omega = \{HH, HT, TH, TT\})$ 2 אפשרויות בשלב הראשון ועבור כל אפשרות בשלב הראשון יש 2 אפשרויות בשלב השני:

$$|\Omega| = 2.2 = 2^2 = 4.$$

- מספר התוצאות האפשריות בהטלת מטבע 3 פעמים הוא $2^3 = 8$.
- מספר התוצאות האפשריות בהטלת קוביה פעמים הוא $36 = 6.6$ (6 תוצאות אפשריות בהטלה הראשונה ו־6 תוצאות בהטלה השנייה).
- מספר התוצאות האפשריות בהטלת קוביה ולאחר מכן מטבע הוא $12 = 6.2$.

  מספר המילים השונות בנות 5 אותיות שניתן ליצור מהא"ב העברי הוא 22^5.

> באופן כללי, אם בכל שלב בניסוי ה-k-שלבי יש n אפשרויות, אזי ישנן n^k תוצאות אפשריות סה"כ.

דוגמא. בבית ספר מקצים באקראי 4 מורים ל-8 כיתות בלי הגבלה על המספר הכיתות שכל מורה ילמד. מהי המספר החלוקות האפשרי?

פיתרון. כל כיתה צריך לבחור את המורה שילמד אותו. הכיתות בוחרים את המורים, ולכן,
לכיתה הראשון יש 4 אפשרויות.
לכיתה השני יש 4 אפשרויות,
וכן הלאה.
סה"כ יש 4^8 חלוקות אפשרויות. ■

דוגמא. אלון בוחר סיסמא למחשב באורך 5 תווים. שני התווים הראשונים חייבים להיות אותיות לטיניות ($A, B, C, \ldots$), שלושת התווים הבאים נקבעים בעזרת הטלת קובייה. סיסמא נחשבת לחוקית אם היא לא מכילה 2 אותיות לטיניות זהות. מה ההסתברות שהסיסמא שנבחרה היא חוקית?

פיתרון. עבור תו ראשון יש 26 אפשרויות.
עבור תו שני יש 26 אפשרויות.
עבור כל אחד מהתווים הנאים יש 6 אפשריות סה"כ.

$$|\Omega| = 26.26.6.6.6 = 26^2 6^3.$$

נסמן את המאורע בו יש 2 אותיות זהות בסיסמא ב A. לתו ראשון יש 26 אפשרויות, התו השני חייב להיות זהה לתו ראשון. לכן מספר אפשרויות לבחור את צמד התווים הראשונים הוא

$$26.1 = 26.$$

אזי

$$P(A) = \frac{|A|}{|\Omega|} = \frac{26.1.6.6.6}{26^2 6^3} = \frac{1}{26},$$

על כן

$$P(\bar{A}) = 1 - P(A) = \frac{25}{26}.$$

■

מדגם סדור ללא החזרה

נניח ויש ברשותנו סלסלה עם 3 פירות: בננה, תפוח, ואגס. אנחנו רוצים לחלק אותם באקראי לאלון, בן, ובר. כמה אפשרויות חלוקה יש בפנינו? הדרך הקלה לדמיין את הבעיה היא להניח שאלון, בן, ובר מסודרים בשורה ואנחנו עוברים ומחלקים להם את הפירות. עבור אלון יש לנו 3 אפשרויות, עבור בן יוותרו לנו 2 אפשרויות, ועבור בר נשארה אפשרות אחת בלבד, לכן מספר האפשרויות הכללי, לפי עקרון הכפל הוא

$$3.2.1 = 3!$$

למעשה סידרנו את הפירות בשורה. הגרלנו את הפרי הראשון שקיבל אלון, לאחר מכן הגרלנו את הפרי השני שעבר לבן, ולבסוף הגרלנו את הפרי האחרון שבר קיבלה. בעזרת הדוגמא הזאת נוכל לנסח את המסקנה הכללית לגבי סידור איברים שונים בשורה.

2.4 חוק. (מדגם סדור ללא החזרה) מספר הדרכים לסדר n איברים שונים בשורה הוא

$$n! = 1.2.3 \ \ldots \ n \ .$$

עכשיו ונניח שאנו רוצים לשלוף 2 פירות מתוך ה־3, ולחלק אותם לאלון ובן. מספר האפשרויות לביצוע החלוקה דומה למקרה הקודם. ישנן 3 אפשרויות לבחירת הפרי עבור אלון ו־2 אפשרויות לבחירה עבור בן. נשים לב שעברנו על כל האפשרויות שכן כל אחד מהילדים יכול לקבל כל פרי מן הסלסלה. ננסח את המסקנה הכללית.

2.5 חוק. (מדגם סדור ללא החזרה) מספר הדרכים לדגום k איברים מתוך n איברים שונים, עם חשיבות לסדר וללא החזרה, הוא

$$n.(n-1) \cdots (n-k+1) = \frac{n!}{(n-k)!}.$$

במילים אחרות, עבור האיבר הראשון ישנן n אפשרויות, עבור האיבר השני ישנן $n-1$ אפשרויות, וכן הלאה עד האיבר- k.

3 יסודות הקומבינטוריקה 4-7

מדגם ללא החזרה: סדור ולא סדור

3.1 חוק. (מדגם סדור ללא החזרה) מספר הדרכים לסדר n איברים שונים בשורה הוא

$$n! = n(n-1)\dots(2)(1) \ . \tag{3.1}$$

מספר הדרכים לדגום k איברים מתוך n איברים שונים עם חשיבות לסדר וללא החזרה הוא

$$n(n-1)\dots(n-k+1) = \frac{n!}{(n-k)!} \ . \tag{3.2}$$

3.2 חוק. (מדגם לא סדור ללא החזרה) מספר הדרכים לדגום k איברים מתוך n איברים שונים בלי חשיבות לסדר וללא החזרה הוא

$$\frac{n(n-1)\dots(n-k+1)}{k!} = \frac{n!}{(n-k)!k!}. \tag{3.3}$$

3.3 הגדרה. (מקדם הבינומיאלי) הסימון

$$\binom{n}{k} := \frac{n!}{(n-k)!k!}, \tag{3.4}$$

או לעיתים

$$_nC_k := \frac{n!}{(n-k)!k!}. \tag{3.5}$$

גם וגם נקראים המקדם הבינומיאלי.

3.4 הגדרה. ( ) הסימון

$$_nP_k := \frac{n!}{(n-k)!} \ . \tag{3.6}$$

דוגמא. לוקחים 3 אותיות שונות מתוך ה 6 אותיות (a,b,c,d,e,f). הצירופים האפשריים מפורטים להלן:

$r1$	(a,b,c)	(a,c,b)	(b,a,c)	(b,c,a)	(c,a,b)	(c,b,a)
$r2$	(a,b,d)	(a,d,b)	(b,a,d)	(b,d,a)	(d,a,b)	(d,b,a)
$r3$	(a,b,e)	(a,e,b)	(b,a,e)	(b,e,a)	(e,a,b)	(e,b,a)
$r4$	(a,b,f)	(a,f,b)	(b,a,f)	(b,f,a)	(f,a,b)	(f,b,a)
$r5$	(a,c,d)	(a,d,c)	(c,a,d)	(c,d,a)	(d,a,c)	(d,c,a)
$r6$	(a,c,e)	(a,e,c)	(c,a,e)	(c,e,a)	(e,a,c)	(e,c,a)
$r7$	(a,c,f)	(a,f,c)	(c,a,f)	(c,f,a)	(f,a,c)	(f,c,a)
$r8$	(a,d,e)	(a,e,d)	(d,a,e)	(d,e,a)	(e,a,d)	(e,d,a)
$r9$	(a,d,f)	(a,f,d)	(d,a,f)	(d,f,a)	(f,a,d)	(f,d,a)
$r10$	(a,e,f)	(a,f,e)	(e,a,f)	(e,f,a)	(f,a,e)	(f,e,a)
$r11$	(b,c,d)	(b,d,c)	(c,b,d)	(c,d,b)	(d,b,c)	(d,c,b)
$r12$	(b,c,e)	(b,e,c)	(c,b,e)	(c,e,b)	(e,b,c)	(e,c,b)
$r13$	(b,c,f)	(b,f,c)	(c,b,f)	(c,f,b)	(f,b,c)	(f,c,b)
$r14$	(b,d,e)	(b,e,d)	(d,b,e)	(d,e,b)	(e,b,d)	(e,d,b)
$r15$	(b,d,f)	(b,f,d)	(d,b,f)	(d,f,b)	(f,b,d)	(f,d,b)
$r16$	(b,e,f)	(b,f,e)	(e,b,f)	(e,f,b)	(f,b,e)	(f,e,b)
$r17$	(c,d,e)	(c,e,d)	(d,c,e)	(d,e,c)	(e,c,d)	(e,d,c)
$r18$	(c,d,f)	(c,f,d)	(d,c,f)	(d,f,c)	(f,c,d)	(f,d,c)
$r19$	(c,e,f)	(c,f,e)	(e,c,f)	(e,f,c)	(f,c,e)	(f,e,c)
$r20$	(d,e,f)	(d,f,e)	(e,d,f)	(e,f,d)	(f,d,e)	(f,e,d)

באף סדרה שום תוו לא יכול להופיע יותר מפעם אחת (זו דווקא התכונה של ללא החזרה). ישנן 120 אפשרויות עבור המדגם סדור:

$$\begin{array}{ccc} \square & \square & \square \\ \uparrow & \uparrow & \uparrow \\ 6 & 5 & 4 \end{array} = 6.5.4 = \frac{6.5.4.3.2.1}{3.2.1} = \frac{6!}{(6-3)!} = {}_6C_3. \tag{3.7}$$

שימו לב שכל סדרה שבאותה שורה כוללת אותם שלושה תווים, אך הם מופיעים בסדרים שונים. בנגוד לרשימה לעיל, הרשימה להלן מופיע את כל האפשרויות אך ללא חשיבות לסדר:

$$\begin{array}{ll} r1 & (a,b,c) \\ r2 & (a,b,d) \\ r3 & (a,b,e) \\ r4 & (a,b,f) \\ r5 & (a,c,d) \\ r6 & (a,c,e) \\ r7 & (a,c,f) \\ r8 & (a,d,e) \\ r9 & (a,d,f) \\ r10 & (a,e,f) \\ r11 & (b,c,d) \\ r12 & (b,c,e) \\ r13 & (b,c,f) \\ r14 & (b,d,e) \\ r15 & (b,d,f) \\ r16 & (b,e,f) \\ r17 & (c,d,e) \\ r18 & (c,d,f) \\ r19 & (c,e,f) \\ r20 & (d,e,f) \end{array}$$

(לדוגמה, בשורה $r2$ כל סדרה כוללת רק התווים (a,b,d) בצירופים שונים.) בכל שורה ישנן $3!$ סדרות בנות אותם תווים:

$$\begin{array}{ccc} \square & \square & \square \\ \uparrow & \uparrow & \uparrow \\ 3 & 2 & 1 \end{array}$$

לכן, כדי לחשב את המספר הדרכים לסדר 6 תווים שונים ב 3 מקומות אך ללא חשיבות לסדר (או במילים אחרות האורך של המדגם לא סדור ללא החזרה) יש צורך לחלק את המספר (3.7) בהמספר של צירופים של אותם שלושה תווים בכל שורה (אשר הוא דווקא $3!$). דרך זה מקבלים

$$\frac{6!}{(6-3)!3!} = {}_6P_3\ . \tag{3.8}$$

*סימנים בקומבינטוריקה

3.5 הגדרה. (עוד סימון שכיח לקומבינטוריקה)

$$\begin{aligned}
n^{[r]} &= n(n+1)\dots(n+r-1) && = \frac{(n+r-1)!}{(n-1)!}\,, \\
n_{[r]} &= \frac{n^{[r]}}{r!} && = \frac{(n+r-1)!}{(n-1)!r!}\,, \\
n^{(r)} &= (n-r+1)^{[r]} && = (n-r+1)(n-r+2)\dots(n) \quad = \frac{n!}{(n-r)!}\,, \\
n_{(r)} &= \frac{n^{(r)}}{r!} && = \frac{n!}{(n-1)!r!} \quad = {}_nC_r\,.
\end{aligned}$$

שימו לב שלפי הגדרות האלה,

$$n^{[r]} = (n+r-1)(n+r-2)\dots(n) = \frac{(n+r-1)!}{(n+1)!} = (n+r-1)^{(r)}\,.$$

תרגיל: מדגם לא סדור ללא החזרה

דוגמא. כיתה עם 30 סטודנטים צריכים לבחור ועד של 3 סטודנטים. יתר התלמידים שלא נבחרו לועד יעסקו בפעילות אקדמית אחרת. כמה אפשרויות שונות יש לבחירת ועד? כמה אפשרויות שונות יש לבחירת התלמידים שיעסקו בפעילות אקדמית אחרת?

פיתרון.

$$\binom{30}{3} = \frac{30!}{3!27!}, \qquad \binom{30}{27} = \frac{30!}{27!3!}.$$

■

הניתוח של מדגם סדור עם החזרה זהה לעיקרון הכפל, כאשר בכל שלב יש n אפשרויות, לכן בתהליכי באורך k נקבל סה"כ n^k אפשרויות.

מדגם עם החזרה: סדור ולא סדור

3.6 חוק. (מדגם סדור עם החזרה) מספר הדרכים לדגום k איברים מתוך n איברים שונים עם חשיבות לסדר ועם החזרה הוא

$$\begin{matrix} \square & \square & \dots & \square & = n^k. \\ \uparrow & \uparrow & \dots & \uparrow & \\ n & n & \dots & n & \end{matrix} \tag{3.9}$$

3.7 חוק. (מדגם לא סדור עם החזרה) המספר הדרכים לדגום r תווים מתוך n תווים שונים ללא חשיבות לסדר ועם החזרה הוא

$$\frac{n^{[r]}}{r!} = \frac{(n+r-1)!}{(n-1)!r!} \tag{3.10}$$

דוגמא. (מדגם לא סדור עם החזרה)
כמה צירופים יש בסה"כ, של קבוצות של שתי מספרים $(1,1)$ $(4,3)$, $\ldots$, $(1,2)$ עד $(6,6)$, מניסוי של לזרוק שתי קוביות, כאשר $(1,2)$ ו $(2,1)$ נחשבים כאותה תוצאות, זאת אומרת ללא חשיבות לסדר?

פיתרון. האפשרויות הן מפורטות להלן:

$$
\begin{array}{llllll}
(1,1), & (1,2), & (1,3), & (1,4), & (1,5), & (1,6), \\
\cancel{(2,1)}, & (2,2), & (2,3), & (2,4), & (2,5), & (2,6), \\
\cancel{(3,1)}, & \cancel{(3,2)}, & (3,3), & (3,4), & (3,5), & (3,6), \\
\cancel{(4,1)}, & \cancel{(4,2)}, & \cancel{(4,3)}, & (4,4), & (4,5), & (4,6), \\
\cancel{(5,1)}, & \cancel{(5,2)}, & \cancel{(5,3)}, & \cancel{(5,4)}, & (5,5), & (5,6), \\
\cancel{(6,1)}, & \cancel{(6,2)}, & \cancel{(6,3)}, & \cancel{(6,4)}, & \cancel{(6,5)}, & (6,6).
\end{array}
$$

זו בעיה של מדגם לא סדור עם החזרה ויש צורך להשתמש בהנוסחא (3.10) כאשר $n=6$ ו $r=2$:

$$\frac{6^{[2]}}{2!}=\frac{(6+2-1)!}{(6-1)!2!}=\frac{7!}{5!2!}=21\ .$$

■

תרגילים: קומבינטוריקה והסתברות

דוגמא. בבית ספר ובו 7 מורים מלמדים המורים 5 ימים בשבוע. כל מורה בחר מקרית יום בו לא יעבוד (א־ו). חשבו את ההסתברויות הבאות:

1. המורה לאנגלית והמורה למתמטיקה לא יהיו חופשיים באותו יום,
2. כל המורים בחרו ביום ו' כיום החופשי שלהם,
3. אף מורה לא בחר את יום ו' כיום החופשי שלו,
4. כל יום בשבוע נבחר כיום חופשי של לפחות אחד המורים.

פיתרון. לכל מורה יש 6 אפשרויות לבחירת היום החופשי שלו לכן כמות הצירופים הכוללת היא

$$|\Omega|=6^7.$$

1. למורה לאנגלית יש 6 אפשרויות לבחור את היום החופשי שלו. לאחר מכן, למורה למתמטיקה יש 5 אפשרויות עבור היום החופשי 6. סה"כ $30=6.5$ אפשרויות. לכל המורים האחרים יש 6^5 אפשרויות. על כן צריך לכפול ב- 6^5 לתת 30.6^5. לכן
$$P=\frac{30.6^5}{6^7}=\frac{5}{6}.$$

2. יש רק דרך אחד לייצר המאורע הדרוש, לכן

$$P = \frac{1}{6^7}.$$

3. לכל מורה יש רק 5 אפשרויות. לכן כמות הצירופים הוא 5^7. לכן

$$P = \frac{5^7}{6^7} = \left(\frac{5}{6}\right)^7.$$

4. במאורע המבוקש חייב להיות 2 מורים שיבחרו את אותו היום ורק הם. יש $\binom{7}{2}$ אדרכים לבחור המורים מתוך 7. לאחר שבחרנו אותם יש 6 פריטים: (5 מורים ועוד זוג מוים) שצריכים לסדר בשורה. (כל מקום הוא היום החופשי) ולכך יש $6!$ אפשרויות. לכן ההסתברות תהיה

$$P = \frac{6!\binom{7}{2}}{6^7}.$$

■

דוגמא. מרכיבים באקראי מילה בת 4 תווים מן האותיות א' עד ו'. חשבו את ההסתברויות של המאורעות הבאים.

1. מאורע A: א' מופיע פעם אחת לפחות,
2. מאורע B: א' מופיע בדיוק פעם אחת,
3. מאורע C: אין אות שחוזרת בסיסמא.

פיתרון. לכל תו יש 6 אפשרויות ולכן

$$|\Omega| = 6^4.$$

1. $\bar{A} =$ אות א' לא נבחרה כלל. זאת אומרת כל אפשרויות מורכבת אך ורק מן התווים ב' - ו'.

$$\Rightarrow \qquad P(A) = 1 - P(\bar{A}) = 1 - \frac{5^4}{6^4}.$$

2. $B_i =$ מאורע ש א' מופיע במקום i ($i = 1, \ldots, 4$).

$$\Rightarrow \qquad B = \bigcup_{i=1}^{4} B_i,$$

והאיחוד הוא על פני מאורעות זרים. על כן

$$P(B) = P\left(\bigcup_{i=1}^{4} B_i\right) = \sum_{i=1}^{4} P(B_i).$$
$$P(B_i) = \frac{5^3}{6^4}.$$
$$\therefore \qquad P(B) = \sum_{i=1}^{4} P(B_i) = \frac{4(5^3)}{6^4}.$$

3.
- לתו הראשון יש 6 אפשרויות,
- לתו שני יש 5 אפשרויות,

- לתו שלישי יש 4 אפשרויות,
- לתו רביעי יש 3 אפשרויות,

$$\therefore \qquad P=\frac{6.5.4.3}{6^4}=\frac{6!}{6^4(6-4)!}\ .$$

■

דוגמא. במחלקה לכלכלה 25 חברי סגל ־ 9 דוקטורים ו־16 פרופסורים. בוחרים באקראי ועדת הוראה בת 4 חברי סגל. מצאו את ההסתברות ש

1. בועדה יש 2 דוקטורים ו־2 פרופסורים,
2. בועדה יש לפחות 3 פרופסורים,
3. בועדה יש לפחות דוקטור אחד ולפחות פרופסור אחד.

פיתרון.

$$|\Omega|=\binom{25}{4}\ .$$

1. $$P(A)=\frac{\binom{9}{2}\binom{16}{2}}{\binom{25}{4}}$$
2. $$P(B)=\frac{\binom{9}{1}\binom{16}{3}+\binom{9}{0}\binom{16}{4}}{\binom{25}{4}}$$
3. $$\begin{aligned}P(C)=&1-P(\bar{C})\\ P(\bar{C})=&\frac{\binom{9}{4}}{\binom{25}{4}}+\frac{\binom{16}{4}}{\binom{25}{4}}\ .\end{aligned}$$

■

דוגמא. בכיתה n סטודנטים. מהי ההסתברות שיש לפחות זוג אחד של סטודנטים עם אותו תאריך יום הולדת (מבלי להתייחס לשנת הלידה)?

פיתרון. גודל מרחב המדגם שלנו הוא 365^n משום שלכל סטודנט יש 365 ימים אפשריים ליום הולדת וישנם n סטודנטים. נסמן ב A המאורע בו לפחות זוג אחד של סטודנטים נולד באותו היום, ונבחן את המאורע המשלים $\bar{A}$. המאורע בו לכל הסטודנטים ימי הולדת שונים. זו דוגמה של מדגם לא סדור וללא החזרה, ולפי זה עם העזקה של 3.2 נמצא ש מספר האפשרויות במאורע $\bar{A}$ הוא

$$|\bar{A}|={}_{365}C_n=\frac{365!}{(365-n)!}\ .$$

מכאן נקבל ש

$$P(\bar{A})=1-P(A)=1-\frac{|A|}{|\Omega|}=1-\frac{365!}{(365-n)!365^n}\ .$$

התוצאה המפתיעה היא שעבור $n=23$ ההסתברות גדולה מ 50% (0.507 בקירוב) ועבור $n=60$ סטודנטים ההסתברות היא 0.994.

■

*מדגם סדור של קבוצות בו תת קבוצו ת של איברים זההים

נניח שיש n תווים בקבוצה Ω, אשר בתוך זה יש את

- תת קבוצה A בת a תווים זההים,
- תת קבוצה B בת b תווים זהים,
- תת קבוצה C בת c תווים זה הים.

$$\Omega = \{\overbrace{\circ,\ldots,\circ}^{a},\ \overbrace{\square,\ldots,\square}^{b},\ \overbrace{\triangle,\ldots,\triangle}^{c}\}$$

בטח מתקיים

$$a+b+c=n.$$

מסמנים את המספר הדרכים לסדר את האיברים ב Ω ב

$$_{n}P_{(a,b,c)}\ .$$

כדי להגיע לנוסחא ל $_{n}P_{(a,b,c)}$ בפירוש, מחליפים את כל ה a תווים הזההים בתוך A בתווים שונים, זאת אומרת מחליפים את כל תוו של $\circ$ בתוו שונה. כך המספר הסדרים שונים הוא פי $a!$ פעמים יותר גדול מקודם. לאחר מכן מחליפים את כל ה b תווים הזההים בתוך B בתווים שונים. זאת אומרת מחליפים את כל תוו של $\square$ בתוו שונה. כך המספר הסדרים שונים הוא פי $b!$ פעמים יותר גדול מקודם. לאחר מכן מחליפים את כל ה c תווים הזההים בתוך C בתווים שונים. זאת אומרת מחליפים את כל תוו של $\triangle$ בתוו שונה. כך המספר הסדרים שונים הוא פי $c!$ פעמים יותר גדול מקודם. בסוף מקבלים סדר של n תווים שונים, אשר יש לו $_{n}P_{n}=n!$ צירופים שונים. לכן

$$\begin{aligned} a!b!c!\,_{n}P_{(a,b,c)} &= \ _{n}P_{n} &&= n! \\ \Rightarrow \quad _{n}P_{(a,b,c)} &= \frac{_{n}P_{n}}{a!b!c!} &&= \frac{n!}{a!b!c!}\ . \end{aligned}$$

3.8 חוק. (צירופים עבור קבוצה בת תת קבוצות של תווים זההים)
המספר הדרכים לסדר n איברים שונים מתוך קבוצה של n איברים, אשר המורכב מ

- תת קבוצה A בת a איברים זההים,
- תת קבוצה B בת b איברים זההים,
- תת קבוצה C בת c איברים זההים,

$\vdots$

הוא

$$_{n}P_{(a,b,c,\ldots)} = \frac{n!}{a!b!c!\ldots}\ . \tag{3.11}$$

דוגמא. כמה אפשרויות יש לסדר 6 מטבעות של ₪1 , 7 מטבעות של ₪5, 8 מטבעות של ₪10 ?

פיתרון.

$$_{21}P_{(6,7,8)} = \frac{21!}{6!7!8!} = 349,188,840.$$

■

3.9 חוק. (מדגם מתוך קבוצה בת תת קבוצות של תווים זההים עם החזרה)
נתון קבוצה Ω בו יש

- תת קבוצה A בת a איברים זההים,
- תת קבוצה B בת b איברים זההים,

- תת קבוצה C בת c איברים זההים,

$\vdots$.

המספר הדרכים לדגום u איברים מתוך A, v איברים מתוך B, w איברים מתוך C, עם החזרה עם חשיבות לסדר הוא הוא

$$ {}_nP_{(u,v,w,\dots)} a^u b^v c^w \cdots = \frac{n!}{u!v!w!\dots} a^u b^v c^w \ . \tag{3.12} $$

3.10 חוק. (מדגם מתוך קבוצה בת תת קבוצות של תווים זההים ללא החזרה)
נתון קבוצה Ω בו יש

- תת קבוצה A בת a איברים זההים,
- תת קבוצה B בת b איברים זההים,
- תת קבוצה C בת c איברים זההים,

$\vdots$.

המספר הדרכים לדגום u איברים מתוך A, v איברים מתוך B, w איברים מתוך C, ללא החזרה עם חשיבות לסדר הוא הוא

$$ {}_nP_{(u,v,w,\dots)} a^{(u)} b^{(v)} c^{(w)} \cdots = \frac{n!}{u!v!w!\dots} \frac{a!}{(a-u)!} \frac{b!}{(b-v)!} \frac{c!}{(c-w)!} \cdots \ . \tag{3.13} $$

4 חזרה ותרגילים, הפירמדיה של פסקל ותורת הבינומיאלי 6-7

סיכום נוסחאות בקומבינטוריקה

4.1 חוק. (מדגם סדור ללא החזרה) מספר הדרכים לסדר n איברים שונים בשורה הוא

$$n! = n(n-1)\dots(2)(1)\ . \tag{4.1}$$

מספר הדרכים לדגום k איברים מתוך n איברים שונים עם חשיבות לסדר וללא החזרה הוא

$$n(n-1)\dots(n-k+1) = \frac{n!}{(n-k)!} \equiv {}_nP_k\ . \tag{4.2}$$

4.2 חוק. (מדגם לא סדור ללא החזרה) מספר הדרכים לדגום k איברים מתוך n איברים שונים בלי חשיבות לסדר וללא החזרה הוא

$$\frac{n(n-1)\dots(n-k+1)}{k!} = \frac{n!}{(n-k)!k!} \equiv \binom{n}{k}. \tag{4.3}$$

4.3 חוק. (מדגם סדור ללא החזרה של של הפרדה) המספר הדרכים לסדר קבוצה בת n דברים, בו יש n_1 דברים שכולם של סוג אחד, n_2 דברים שכולם של סוג אחר, . . .
הוא

$$\frac{n!}{n_1!n_2!\dots} \equiv \binom{n}{n_1, n_2, \dots} \tag{4.4}$$

במילים אחרות, המספר הדרכים לסדר קבוצה של n איברים, להפרדה המורכב מתת קבוצות של n_1 איברים, n_2 איברים, . . .
הוא

$$\frac{n!}{n_1!n_2!\dots} \equiv \binom{n}{n_1, n_2, \dots} \tag{4.4}$$

4.4 הגדרה. (מקדם הבינומיאלי) המקדם הבינומיאלי הוא

$$\binom{n}{k} := \frac{n!}{(n-k)!k!}, \tag{4.5}$$

או לעיתים

$${}_nC_k := \frac{n!}{(n-k)!k!}\ . \tag{4.6}$$

4.5 חוק. (מדגם סדור עם החזרה) מספר הדרכים לדגום k איברים מתוך n איברים שונים עם חשיבות לסדר ועם החזרה הוא n^k .

$$\begin{array}{cccc} \#1 & \#2 & \dots & \#k \\ & & & \\ \square & \square & \dots & \square \\ \uparrow & \uparrow & \dots & \uparrow \\ n & n & \dots & n \end{array} \quad \to \overbrace{n \times n \times \dots \times n}^{\text{איברים } k} = n^k.$$

4.6 חוק. (מדגם לא סדור עם החזרה) המספר הדרכים לדגום k תווים מתוך n תווים שונים ללא חשיבות לסדר ועם החזרה הוא

$$\frac{(n+k-1)!}{(n-1)!k!} \tag{4.7}$$

תרגילים

דוגמא. 3 כדורים של 3 צבעים שונים נתונים ל3 תלמידים בכיתה בת 25 תלמידים. אם כל תלמיד מקבל כדור אחד, כמה אפשרויות ישנן לחלק את הכדורים?

פיתרון. בשל העובדה הכדורים באים ב3 צבעים שונים, אז ניתן להבחים בין הכדורים. על כן הבעיה זו בעיה של מדגם סדור ללא החזרה, ובמיוחד עם חשבות לסדר. אז התשובה היא ניתן על ידי הנוסחא (4.2):

$$ {}_{25}P_3 = \frac{25!}{(25-3)!} = \frac{25!}{22!} = 25.24.23 \ . $$

■

דוגמא. 3 כדורים של אותו צבע נתונים ל3 תלמידים בכיתה בת 25 תלמידים. אם כל תלמיד מקבל כדור אחד, כמה אפשרויות ישנן לחלק את הכדורים?

פיתרון. עכשיו לא ניתן להבחים בין הכדורים. על כן הבעיה זו בעיה של מדגם סדור ללא החזרה, ללא חשבות לסדר. אז התשובה היא ניתן על ידי הנוסחא (4.3):

$$ {}_{25}C_3 = \frac{25!}{(25-3)!3!} = \frac{25!}{22!3!} = \frac{25.24.23}{6} = 25.4.23 = 2300 \ . $$

■

דוגמא. מרכיבים באקראי מילה בת 4 תווים מן האותיות א' עד ו'. חשבו את ההסתברויות של המאורעות הבאים.

1. מאורע A: א' מופיע פעם אחת לפחות,
2. מאורע B: א' מופיע בדיוק פעם אחת,
3. מאורע C: אין אות שחוזרת בסיסמא.

פיתרון. לכל תו יש 6 אפשרויות ולכן

$$ |\Omega| = 6^4. $$

1. $\bar{A} =$ אות א' לא נבחרה כלל. זאת אומרת כל אפשרויות מורכבת אך ורק מן התווים ב' - ו'.

$$ \Rightarrow \qquad P(A) = 1 - P(\bar{A}) = 1 - \frac{5^4}{6^4}. $$

2. $B_i =$ מאורע ש א' מופיע במקום i ($i = 1, \dots, 4$).

$$ \Rightarrow \qquad B = \bigcup_{i=1}^{4} B_i, $$

והאיחוד הוא על פני מאורעות זרים. על כן

$$ \begin{aligned} P(B) =& P\left(\bigcup_{i=1}^{4} B_i\right) = \sum_{i=1}^{4} P(B_i). \\ P(B_i) =& \frac{5^3}{6^4}. \end{aligned} $$

$$ \therefore \qquad P(B) = \sum_{i=1}^{4} P(B_i) = \frac{4(5^3)}{6^4}. $$

3.
- לתו הראשון יש 6 אפשרויות,
- לתו שני יש 5 אפשרויות,
- לתו שלישי יש 4 אפשרויות,
- לתו רביעי יש 3 אפשרויות,

$$\therefore \qquad P = \frac{6.5.4.3}{6^4} = \frac{6!}{6^4(6-4)!} \ .$$

■

דוגמא. כמה אפשרויות יש לסדר 6 מטבעות של ₪1 , 7 מטבעות של ₪5, 8 מטבעות של ₪10 ?

פיתרון. זו היא בעיה של כמה אפשרויות ישנן לסדר 21 דברים בו יש תת קבוצות של 8 דברים של אותו סוג, 7 דברים של אותו הסוג, ... על כן התשובה ניתנת על ידי הנוסחא (4.4):

$$\frac{21!}{6!7!8!} = 349,188,840.$$

■

דוגמא. כמה דרכים יש לשים 7 אנשים ב 2 חדרים בת 2 מיטות וחדר אחד בת 3 מיטות?

פיתרון. הביעה זו היא בעיה למצוא את המספר הדרכים לסדר קבוצה של 7 אנשים להפרדה המורכב מתת קבוצות של 2, 2 ו 3, והתשובה ניתנת על ידי הנוסחא (4.4):

$$\frac{7!}{2!2!3!} = \frac{7.6.5.4}{2.2} = 210 \ .$$

■

דוגמא. בכיתה n סטודנטים. מהי ההסתברות שיש לפחות זוג אחד של סטודנטים עם אותו תאריך יום הולדת (מבלי להתייחס לשנת הלידה)?

פיתרון. גודל מרחב המדגם שלנו הוא 365^n משום שלכל סטודנט יש 365 ימים אפשריים ליום הולדת וישנם n סטודנטים. נסמן ב A המאורע בו לפחות זוג אחד של סטודנטים נולד באותו היום, ונבחן את המאורע המשלים $\bar{A}$. המאורע בו לכל הסטודנטים ימי הולדת שונים. זו דוגמה של מדגם **סדור** וללא החזרה, ולפי זה עם העזרה של 3.2 נמצא ש מספר האפשרויות במאורע $\bar{A}$ הוא

$$|\bar{A}| = {}_{365}C_n = \frac{365!}{(365-n)!} \ .$$

מכאן נקבל ש

$$P(A) = 1 - P(\bar{A}) = 1 - \frac{|\bar{A}|}{|\Omega|} = 1 - \frac{365!}{(365-n)!365^n} \ .$$

התוצאה המפתיעה היא שעבור $n = 23$ ההסתברות גדולה מ 50% (0.507 בקירוב) ועבור $n = 60$ סטודנטים ההסתברות היא 0.994.

■

המשולש של פסקל

$$
\begin{array}{ll}
n=0 & 1 \\
n=1 & 1 \quad 1 \\
n=2 & 1 \quad 2 \quad 1 \\
n=3 & 1 \quad 3 \quad 3 \quad 1 \\
n=4 & 1 \quad 4 \quad 6 \quad 4 \quad 1 \\
n=5 & 1 \quad 5 \quad 10 \quad 10 \quad 5 \quad 1 \\
n=6 & 1 \quad 6 \quad 15 \quad 20 \quad 15 \quad 6 \quad 1 \\
n=7 & 1 \quad 7 \quad 21 \quad 35 \quad 35 \quad 21 \quad 7 \quad 1 \\
n=8 & 1 \quad 8 \quad 28 \quad 56 \quad 70 \quad 56 \quad 28 \quad 8 \quad 1 \\
n=9 & 1 \quad 9 \quad 36 \quad 84 \quad 126 \quad 126 \quad 84 \quad 36 \quad 9 \quad 1 \\
n=10 & 1 \quad 10 \quad 25 \quad 120 \quad 210 \quad 252 \quad 210 \quad 120 \quad 25 \quad 10 \quad 1
\end{array}
$$

$$
\begin{array}{ll}
n=0 & \binom{0}{0} \\
n=1 & \binom{1}{0} \quad \binom{1}{1} \\
n=2 & \binom{2}{0} \quad \binom{2}{1} \quad \binom{2}{2} \\
n=3 & \binom{3}{0} \quad \binom{3}{1} \quad \binom{3}{2} \quad \binom{3}{3} \\
n=4 & \binom{4}{0} \quad \binom{4}{1} \quad \binom{4}{2} \quad \binom{4}{3} \quad \binom{4}{4} \\
n=5 & \binom{5}{0} \quad \binom{5}{1} \quad \binom{5}{2} \quad \binom{5}{3} \quad \binom{5}{4} \quad \binom{5}{5} \\
n=6 & \binom{6}{0} \quad \binom{6}{1} \quad \binom{6}{2} \quad \binom{6}{3} \quad \binom{6}{4} \quad \binom{6}{5} \quad \binom{6}{6}
\end{array}
$$

על ידי המשולש של פסקל, קל לראות שמתקיים

$$\binom{n}{k} = \binom{n}{n-k} \quad . \tag{4.8}$$

תורת הבינומיאלי

נניח שאנחנו רוצים לפתוח את הסוגריים של דו-איבר,

$$(p+q)^2 = (p+q)(p+q) \ .$$

הפעולה הזו היא קלה ומקבלים

$$(p+q)^2 = p^2 + 2pq + q^2 \tag{4.9}$$

אבל לעשות אותה הפעולה, כדי לפתוח את הסוגריים של $(p+q)^n$ כאשר $n = 3, 4, \ldots$ היא כבר לא פשוטה. מבטאים את $(p+q)^n$ כפי

$$(p+q)^n = \overbrace{(p+q)(p+q)(p+q)\ldots(p+q)}^{n \text{ איברים}} \ . \tag{4.10}$$

כאשר פותחים את הסוגריים אחד אחד אנחנו מקבלים סכום של איברים כמו

$$(p+q)^n = \sum_{k=0}^{n} a_{n,k} p^k q^{n-k} \ . \tag{4.11}$$

המקדם $a_{n,k}$ הוא שווה להמספר הדרכים לבחור k איברים של p מתוך n , אשר הוא דווקא המקדם הבינומיאלי

$$a_{n,k} = \binom{n}{k} = \frac{n!}{(n-k)!k!} \ .$$

או להפך, המקדם $a_{n,k}$ הוא שווה להמספר הדרכים לבחור $n-k$ איברים של q מתוך n, וזה דווקא המקדם הבינומיאלי

$$a_{n,k} = \binom{n}{n-k} = \frac{n!}{(n-k)!k!} = \frac{n!}{(n-k)!k!} \ .$$

לכן,

4.7 תורת. (תורת הבינומיאלי)

$$(p+q)^n = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} \ . \tag{4.12}$$

*הערה: מדגם סדור של קבוצות בו תת קבוצו ת של איברים זההים

נניח שיש n תווים בקבוצה Ω, אשר בתוך זה יש את

- תת קבוצה A בת a תווים זההים,
- תת קבוצה B בת b תווים זהים,
- תת קבוצה C בת c תווים זה הים.

$$\Omega = \{\overbrace{\circ, \ldots, \circ}^{a},\ \overbrace{\square, \ldots, \square}^{b},\ \overbrace{\triangle, \ldots, \triangle}^{c}\}$$

בטח מתקיים

$$a + b + c = n.$$

מסמנים את המספר הדרכים לסדר את האיברים ב Ω ב

$$_nP_{(a,b,c)}\ .$$

כדי להגיע לנוסחא ל $_nP_{(a,b,c)}$ בפירוש, מחליפים את כל ה a תווים הזההים בתוך A בתווים שונים, זאת אומרת מחליפים את כל תוו של $\circ$ בתוו שונה. כך המספר הסדרים שונים הוא פי $a!$ פעמים יותר גדול מקודם. לאחר מכן מחליפים את כל ה b תווים הזההים בתוך B בתווים שונים. זאת אומרת מחליפים את כל תוו של $\square$ בתוו שונה. כך המספר הסדרים שונים הוא פי $b!$ פעמים יותר גדול מקודם. לאחר מכן מחליפים את כל ה c תווים הזההים בתוך C בתווים שונים. זאת אומרת מחליפים את כל תוו של $\triangle$ בתוו שונה. כך המספר הסדרים שונים הוא פי $c!$ פעמים יותר גדול מקודם. בסוף מקבלים סדר של n תווים שונים, אשר יש לו $_nP_n = n!$ צירופים שונים. לכן

$$\begin{aligned} a!b!c!\,_nP_{(a,b,c)} &= \ _nP_n &= n! \\ \Rightarrow \qquad _nP_{(a,b,c)} &= \frac{_nP_n}{a!b!c!} &= \frac{n!}{a!b!c!}\ . \end{aligned}$$

4.8 חוק. (צירופים עבור קבוצה בת תת קבוצות של תווים זההים)
עיין חוק 4.3 לעייל כדי לראות אותו החוק במילים שונות. המספר הדרכים לסדר n איברים שונים מתוך קבוצה של n איברים, אשר המורכב מ

- תת קבוצה A בת a איברים זההים,
- תת קבוצה B בת b איברים זההים,
- תת קבוצה C בת c איברים זההים,

$\vdots$

הוא

$$\frac{n!}{a!b!c!\ldots} \equiv \binom{n}{a, b, c, \ldots}\ . \tag{4.13}$$

4.9 חוק. (מדגם מתוך קבוצה בת תת קבוצות של תווים זההים עם החזרה)
נתון קבוצה Ω בו יש

- תת קבוצה A בת a איברים זההים,
- תת קבוצה B בת b איברים זההים,
- תת קבוצה C בת c איברים זההים,

$\vdots$.

המספר הדרכים לדגום u איברים מתוך A, v איברים מתוך B, w איברים מתוך C, עם החזרה עם חשיבות לסדר הוא הוא

$$\frac{n!}{u!v!w!\ldots} a^u b^v c^w \cdots \equiv \binom{n}{u, v, w, \ldots} a^u b^v c^w \ldots\ . \tag{4.14}$$

4.10 חוק. (מדגם מתוך קבוצה בת תת קבוצות של תווים זההים ללא החזרה)
נתון קבוצה Ω בו יש

- תת קבוצה A בת a איברים זההים,
- תת קבוצה B בת b איברים זההים,
- תת קבוצה C בת c איברים זההים,

$\vdots$.

המספר הדרכים לדגום u איברים מתוך A, v איברים מתוך B, w איברים מתוך C, ללא החזרה עם חשיבות לסדר הוא הוא

$$\frac{n!}{u!v!w!\dots}\ \frac{a!}{(a-u)!}\frac{b!}{(b-v)!}\frac{c!}{(c-w)!}\cdots \equiv \binom{n}{u,v,w,\dots}\,{}_aP_u\;{}_bP_v\;{}_cP_w\cdots\ . \tag{4.15}$$

5 הסתברות מותנית 11-7

הסתברות מותנה ואי־תלות

נניח והטלנו קובייה וכעת חבר אומר לנו כי התוצאה איננה 1 ואיננה 6 .טרם קבלת המידע, ייחסנו הסתברות של $\frac{1}{6}$ לתוצאה 1 ולתוצאה 6, אבל לאחר קבלת המידע, ההסתברות הללו ירדו ל-0. מצד שני, תחת המידע החדש, מה הסיכוי שהקוביה נחתה על 4 ?או על 5 ?בכדי לענות על השאלה אנו נדרשים לעדכן את מרחב ההסתברות שלנו בהתאם למידע החדש. הסתברות מותנה זו הסתברות אשר עודכנה בהתאם למידע / מאורע ספציפי.

5.1 הגדרה. (הסתברות מותנית) לכל צמד מאורעות A ו־B כך ש־B בעל הסתברות חיובית, ההסתברות של מאורע A בהינתן מאורע B מוגדרת להיות

$$P(A|B)=\frac{P(A\cap B)}{P(B)}\ . \tag{5.1}$$

5.2 דוגמא. בהטלת מטבע ניקח את המאורע $A=\{2\}$ ו $B=\overline{\{1,6\}}$. ברור ש־ $P(A)=\frac{1}{6}$. אבל, אם B מתממש ותוצאת ההטלה איננה 1 או 6 הסיכוי לקבל את התוצאה 2 משתנה. אינטואיציה שלנו מרמזת כי הסיכוי המעודכן יהיה רבע, משיקולי סימטריה בין התוצאות הנותרות. נבחן זאת לפי ההגדרה.

$$P(\{2\}|B)=\frac{P(\{2\}\cap B)}{P(B)}=\frac{P(2)}{\frac{4}{6}}=\frac{\left(\frac{1}{6}\right)}{\left(\frac{4}{6}\right)}=\frac{1}{4},$$

כפי שחזינו.

5.3 דוגמא. בדוגמא הבאה נגדיר את המאורע $A=\{1,2,5,6\}$. ברור ש $P(A)=P(B)=\frac{4}{6}$. לאחר קבלת המידע שהתוצאה איננה 1 או 6 נבחין כי

$$P(A|B)=\frac{P(A\cap B)}{P(B)}=\frac{P(2,5)}{\frac{4}{6}}=\frac{\left(\frac{2}{6}\right)}{\left(\frac{4}{6}\right)}=\frac{1}{2}.$$

זאת אומרת שההסתברות של מאורע A ירדה מ$\frac{2}{3}$ ל $\frac{1}{2}$.

עדכון ההסתברות שביצענו לעיל נקרא **עדכון בייסיאני.**

5.4 דוגמא. בפונטי־פאנדי יש

- 45% גברים,
- 30% מעשנים, ו־
- 15% הם גברים מעשנים.

נבחר אדם באקראי. בהנחה ונבחר גבר,

1. מה הסיכוי שהוא מעשן?
2. מה הסיכוי שאינו מעשן?
3. כיצד התשובות תשתננה בהנחה ונבחרה אישה?

פיתרון. נסמן ב־ A את המאורע שנבחר גבר וב־ B את המאורע שהאדם שנבחר מעשן. באופן כללי, ההסתברויות של A ו־ B הן $P(A)=0.45$ ו־ $P(B)=0.3$.

1. נחשב את P שהאדם הנבחר מעשן, תחת ההנחה שמדובר בגבר.

$$P(B|A)=\frac{P(A\cap B)}{P(A)}=\frac{0.15}{0.45}=\frac{1}{3}.$$

זאת אומרת, שיעור המעשנים באוכלוסיית הגברים גבוה מן הממוצע באוכלוסייה הכוללת.

2. בכדי לבחון את P שאינו מעשן נוכל להשתמש בכלל המשלים, (הסתברות מותנה מקיימת את כל כללי פונקצית ההסתברות). לכן,
$$P(\bar{B}|A) = 1 - P(B|A) = \frac{2}{3}.$$
כנדרש.

3. נשים לב שמדובר בהתנייה שונה ולכן מדובר במרחב שונה לחלוטין. לא ניתן להמיר את התוצאות הקודמות בגלל ששינוי ההתנייה משנה את כל הנחות הבסיס של החישוב. בהתאם לזאת, נצטרך לבצע את החישובים הקודמים בשנית, וכעת נתנה על $\bar{A}$. נשים לב כי
$$B = (B \cap A) \cup (B \cap \bar{A})$$
והאיחוד הוא איחוד זר (של מאורעות זרים). לכן נקבל
$$P(B|\bar{A}) = \frac{P(B \cap \bar{A})}{P(\bar{A})} = \frac{P(B) - P(B \cap A)}{P(\bar{A})} = \frac{0.3 - 0.15}{0.55} \approx 0.27.$$
המעשנות נמוך יותר משיעור המעשנים באוכלוסיה. כמו כן,
$$P(\bar{B}|\bar{A}) = 1 - P(B| \cap A)1 - \frac{0.3 - 0.15}{0.55} \approx 0.73.$$

■

5.5 דוגמא. בכד 10 כדורים הממספרים מ־1 עד 10 .שולפים 4 ללא החזרה.

1. מה הסיכוי שהכדור עם המספר 7 בחוץ?
2. ידוע ש־9 בחוץ, מה הסיכוי שגם 7 בחוץ?
3. ידוע שאין מספר קטן מ־4 בחוץ, מה הסיכוי ש־7 בחוץ?
4. מה הסיכוי ש־7 לא בחוץ אם ידוע שאין מספר קטן מ־4 בחוץ?

פיתרון. ישנם 10 כדורים ומוציאים 4 ללא החזרה ולכן מרחב המדגם מורכב מכל האפשרויות לבחור 4 כדורים מתוך 10 , סה"כ
$$|\Omega| = \binom{10}{4}.$$

1. נגדיר את המאורע ־ "הכדור עם המספר 7 הינו אחד מהכדורים שהוצאו" בתור A .מאחר ומדובר במרחב מדגם שווה הסתברות, אזי
$$P(A) = \frac{1.\binom{9}{3}}{\binom{10}{4}} = \frac{9!6!4!}{10!6!3!} = \frac{4}{10} = \frac{2}{5}\ .$$

2. נסמן ב־ B את המאורע המציין כי הכדור עם המספר 9 הינו אחד מהכדורים שהוצאו.
$$\begin{aligned} P(A|B) = & \frac{P(A \cap B)}{P(B)} & = \frac{\left(\frac{1.\binom{8}{2}}{\binom{10}{4}}\right)}{\left(\frac{1.\binom{9}{3}}{\binom{10}{4}}\right)} \\ = & \frac{\binom{8}{2}}{\binom{9}{3}} & = \frac{8!6!3!}{9!6!2!} = \frac{1}{3}. \end{aligned}$$
האם היינו יכולים לדעת זאת מראש? כן! האינטואיציה היא מאוד פשוטה. נתון שהכדור עם המספר 9 יצא כבר ולכן נותרו 3 כדורים 1 שצריך להוציא מתוך 9 נותרים, הסיכוי ש־7 הוא אחד מהם הוא בדיוק $\frac{1}{3}$ (ע"ב אותו רעיון של סידור בשורה שראינו קודם לכן)

3. נגדיר מאורע C ־ כל המספרים שהוצאו גדולים או שווים ל־4 .

$$P(A|C) = \frac{P(A\cap C)}{P(C)} = \frac{\frac{\binom{6}{3}}{\binom{10}{4}}}{\frac{\binom{7}{4}}{\binom{10}{4}}}$$
$$= \frac{\binom{6}{3}}{\binom{7}{4}} = \frac{6!4!3!}{7!3!3!} = \frac{4}{7} \ .$$

4. אנו נדרשים לחשב את ההסתברות $P(\bar{A}|C)$. נזכר כי הסתברות מותנה מקיימת את כל הכללים של הסתברות ולכן
$$P(\bar{A}^c|C) = 1 - P(A|C) = 1 - \frac{4}{7} = \frac{3}{7} \ .$$

■

כלל הכפל

הסתברות מותנה מפשט את החישוב, שכן אנו מסוגלים להתייחס לכל שלב בנפרד תחת ההנחה שהתוצאות המתבקשות בשלבים הקודמים אכן התקיימו. הגרסה הבסיסית של כלל הכפל מתקיימת עבור צמד מאורעות A ו־ B כאשר B בעל הסתברות חיובית ($P(B) > 0$). נעזר בהגדרה להסתברות מותנה [עיין משוואה (5.1)], ועל ידי העברת אגפים נקבל

$$P(A\cap B) = P(A|B)P(B) \ . \tag{5.2}$$

זאת אומרת, הסיכוי ש־ A ו־ B יתרחשו, שווה לסיכוי ש־ B יתרחש, כפול הסיכוי ש־ A יתרחש, תחת הנחה ש־ B אכן קרה. ההרחבה למספר כללי של מאורעות היא כמפורט בחוק הבא:

5.6 חוק. (כלל כפל) לכל קבוצת מאורעות $A_1, \dots, A_n$ כך ש
$$P\left(\bigcap_{i=1}^{n} A_i\right) > 0$$
מתקיים
$$P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1)P(A_2|A_1)P(A_3|A_1\cap A_2)\dots P\left(A_n|\bigcap_{i=1}^{n-1} A_i\right) \ . \tag{5.3}$$

קל מאוד לבדוק מדוע טענה זאת נכונה. כל מה שצריך זה לעבור על האיברים באגף ימין ולהציב את ההגדרה להסתברות מותנה. האיברים המתאימים מצטמצמים ומקבלים חזרה את הביטוי באגף שמאל.

5.7 דוגמא. בכד כדור שחור אחד וכדור לבן אחד. שולפים כדור באקראי. לאחר שמוציאים כדור באקראי מחזירים אותו יחד עם עוד כדור אחד נוסף בעל אותו הצבע. מוציאים סה"כ 5 כדורים. מה P שכולם שחורים?

פיתרון. נגדיר את המאורעות B_i $i = 1, \dots, 5$ כמאורעות בהם הכדור ה־ i הוא שחור. נשתמש בחוק הכפל [עיין משוואה (5.3)]:

$$\begin{aligned} P\left(\bigcap_{i=1}^{5} B_i\right) &= P(B_1)P(B_2|B1)P(B_3|B1\cap B2)\dots P\left(B_5|\bigcap_{i=1}^{4} B_i\right) \\ &= \frac{1}{2}\cdot\frac{2}{3}\cdot\frac{3}{4}\cdot\frac{4}{5}\cdot\frac{5}{6} \\ &= \frac{1}{6} \ . \end{aligned}$$

■

נוסחת ההסתברות השלמה

איור 1: עץ הסתברות עבור דוגמה הפוליגרף

לשם המחשה ניקח סיטואציה של שימוש בפוליגרף. פוליגרף, מור לבחון האם אנשים הם דוברי אמת או שקר. הבעיה עם מכונת הפוליגרף היא שאיננה מדוייקת דיה, ולכן יש סיכוי מסויים לשקר ולא להיתפס ולהיפך. אנו יודעים כי אדם המשקר נתפס בשקר בסיכוי 80% ואדם הדובר אמת מזוהה כדובר אמת בסיכוי 90% (שימו לב שכל הנתונים הם למעשה הסתברויות מותנות!). הנחת העובדה כי 70% מן האנשים שמגיעים לבדיקת פוליגרף משקרים. מהי ההסתברות שאדם מקרי בבדיקה יימצא כדובר אמת?

נגדיר את המאורעות הבאים.

- T_1: אדם דובר אמת, L_1: אדם דובר שקר, T_2: אדם מוזהה כדובר אמת (על ידי הפוליגרף), L_2 אדם מזוהה כדובר שקר.

הנתונים בשאלה הם

$$P(T_2|T_1) = 0.9, \quad P(L_2|L_1) = 0.8, \quad P(L_1) = 0.7 \ .$$

אנחנו מעוניינים לחשב את ההסתברות שאדם יימצא דובר אמת, קרי $P(T_2)$. האיור לעיל מציג את הבעיה הנתונה כעץ הסתברויות. נעבור על הענפים המתאימים שנגמרים בתוצאה הרצויה בה האדם נמצא דובר אמת, T_2, ונחשב את ההסתברות של כל ענף בפני עצמו. תחילה,

$$P(T_2) = P(T_2 \cap T_1) + P(T_2 \cap L_1) \ .$$

לפי הסתברות מותנה משוואה [עיין נוסחא (5.2)]

$$P(T_1 \cap T_2) = P(T_1)P(T_1|T_1) = 0.3 \times 0.9 = 0.27 \ .$$

נבצע חישוב דומה עבור המחובר השני ונקבל

$$P(T_2 \cap L_1) = P(L_1)P(T_2|L_1) = 0.7 \times 0.2 = 0.14 \ .$$

כעת נסכום את צמד התוצאות יחד ונקבל

$$P(T_2) = P(T_2|T_1)P(T_1) + P(T_2|L_1)P(L_1) = 0.9 \times 0.3 + 0.2 \times 0.7 = 0.41 \ .$$

שימו לב כי הסיכוי שאדם יוכרז כדובר אמת הוא 41% בלבד!

5.8 חוק. (נוסחת ההסתברות השלמה) נניח כי המאורעות $B_1, \ldots, B_n$ הם זרים בזוגות, בעלי הסתברות חיובית כ"א, ומחלקים את מרחב המדגם באופן מלא. ז"א

$$B_i \cap B_j = \phi$$

לכל $i \neq j$ ובנוסף

$$\bigcup_{i=1}^{n} B_i = \Omega \ ,$$

כאשר $P(B_i) > 0$. אזי לכל מאורע

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i) \ . \qquad (5.4)$$

הוכחה. במילים פשוטות, אנחנו רוצים לדעת את ההסתברות למאורע A, בעוד העולם שלנו מחולק למספר סופי של מאורעות $B_1, \dots B_n$. נתייחס לכל אחד מהמאורעות הללו בנפרד. קרי, תחת ההנחה ש־ B_i התרחש, נבדוק את הסבירות של מאורע A, ונכפול בהסתברות ש־ B_i תתרחש. במידה ונעשה זאת עבור כל מאורע B_i בנפרד, נקבל את ההסתברות של A. נפרק את A לחלקים הבאים:

$$A \cap B_1, \quad A \cap B_2, \dots, A \cap B_n \ .$$

האיחוד של כל המאורעות הוא כל מרחב המדגם, מכאן אנו יודעים שלא פספסנו אף אפשרות של A ועל כן

$$A = \bigcup_{i=1}^{n} A \cap B_i.$$

בנוסף, המאורעות הללו זרים בזוגות בגלל ש־ $B_1, \dots B_n$ זרים בזוגות. לכן

$$P(A) = \sum_{i=1}^{n} P(A \cap B_i) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$$

כנדרש, כאשר במעבר האחרון השתמשנו בכלל הכפל עבור צמד מאורעות. ■

חוק Bayes

לפי ההגדרה של הסתברות מותנה

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

ו

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

נכפול את צמד המשוואות במכנה שלהן ונשווה בינהן בכדי להגיע לחוק בייס הבא.

5.9 חוק. (חוק בייס) עבור כל צמד מאורעות A ו B בעלי הסתברות חיובית מתקיים

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{5.5}$$

לשם המחשה, נחזור לדוגמא הקודמת של הפוליגרף.

5.10 דוגמא. כעת ננסה לחשב את ההסתברות שאדם מקרי דובר אמת, תחת ההנחה שהוא נמצא דובר אמת:

$$P(T_1|T_2).$$

הנתונים בשאלה מוצגים בתצורה של 'אם האדם עשה כך וכך, אז זה הסיכוי שהוא יוכרז כך וכך', אלו יחסי הסיבה־תוצאה בניסוי בפועל. אבל השאלה הנתונה הופכת את הסדר. השאלה מעוניינת לקבוע את הסיכוי שאדם דיבר אמת (סיבה) בהינתן הוא נמצא דובר אמת (תוצאה). במקרים הללו חוק בייס הוא יעיל במיוחד. אכן,

$$P(T_1|T_2) = \frac{P(T_2|T_1)P(T_1)}{P(T_2)} = \frac{0.8.0.3}{0.41} \approx 0.59 \ .$$

זאת אומרת, אדם שנמצא דובר אמת אכן אמר אמת בהסתברות 0.59 בלבד !

5.11 דוגמא. הרכב הסטודנטים בתואר ראשון הוא כדלקמן:

- 50% מהסטודנטים הם בשנה א' לתואר,
- 30% מהסטודנטים הם בשנה ב' לתואר,
- 20% מהסטודנטים הם בשנה ג' לתואר.

בנוסף, מצבם המשפחתי של הסטודנטים הנ"ל הוא:

- 20% מהסטודנטים שנמצאים בשנה א' ־ נשואים,
- 30% מהסטודנטים שנמצאים בשנה ב' ־ נשואים,
- 40% מהסטודנטים שנמצאים בשנה ג' ־ נשואים.

ענו על השאלות הבאות.

1. בהנחה שפגשנו באקראי את אלון, סטודנט מתואר ראשון, מה הסיכוי שהוא נשוי?
2. מסתבר שאלון אכן נשוי, מה הסיכוי שהוא משנה ג'?
3. פגשנו סטודנט אחר, בן, גם כן מתואר ראשון ומסתבר שהוא איננו נשוי, מה הסיכוי שהוא משנה ג'?

פיתרון. בשאלה זו נשתמש בעיקר בנוסחת ההסתברות השלמה וחוק בייז.

1. נסמן ב־ M את המאורע בו הסטודנט נשוי וב־ I ו II ן III את המאורעות בהם הסטודנט הוא בשנה א',ב' ו־ג' לתואר בהתאמה.

$$\begin{aligned}P(M) =&P(M|I)P(I) + P(M|II)P(II) + P(M|III)P(III)\\ &= \frac{2}{10}.\frac{5}{10} + \frac{3}{10}.\frac{3}{10} + \frac{4}{10}.\frac{2}{10} = 0.27\ .\end{aligned}$$

2. נשתמש בחוק בייז מסיבה פשוטה ־ נתוני השאלה מציגים הסתברויות להיות נשוי בהתאם לשנת הלימוד, אבל ההתנייה ההפוכה לא נתונה. לכן,

$$P(III|M) = \frac{P(M|III)P(III)}{P(M)} = \frac{0.4.0.2}{0.27} = \frac{0.08}{0.27} \approx 0.296\ .$$

3. נחשב באופן דומה לחישוב קודם, אלא שהפעם ההתנייה היא על מאורע שונה.

$$P(III|\bar{M}) = \frac{P(\bar{M}|III)P(III)}{P(\bar{M})} = \frac{0.6.0.2}{1-0.27} = \frac{0.12}{0.73} \approx 0.164\ .$$

■

אי-תלות בין מאורעות

תלות בין מאורעות הוא מושג שמבטא את מידת הקשר ההסתברותי בין צמד המאורעות. לדוגמא, ניקח חבילת קלפים תיקנית עם 52 קלפים ונשלוף קלף באקראי. כעת עלינו לנחש את הקלף שהוצא ־ צבע וצורה. אפריורית,

$$P(\text{king}) = \frac{1}{13}, \qquad P(\text{diamond}) = \frac{13}{52},$$

כעת נניח שראינו שהקלף שהוצא הוא אדום. כיצד המידע הנוסף השפיע על ההסתברות הקודמות? ובכן,

$$P(\text{diamond}|\text{red}) = \frac{13}{26}$$

ובעוד

$$P(\text{king}|\text{red}) = \frac{2}{26} = \frac{1}{13}\ .$$

זאת אומרת, המידע הנוסף לגבי צבע הקלף השפיע על הסתברות להוציא יהלום, אך לא השפיע על ההסתברות להוציא מלך. במילים אחרות, המאורע 'יהלום' תלוי במאורע 'קלף אדום' (ולהיפך, כמובן), בעוד המאורע 'מלך' אינו תלוי במאורע 'אדום'. ננסח זאת באופן מדוייק.

5.12 הגדרה. (תלות בין מאורעות)
צמד מאורעות A ו- B הם בלתי־תלויים אם

$$P(A\cap B)=P(A)P(B)\ .$$

מנגד, המאורעות A ו- B הם **תלויים** אם הם לא בלתי־תלויים ז"א אם

$$P(A\cap B)\neq P(A)P(B)\ .$$

ישירות מן ההגדרה נסיק כי אם המאורעות בלתי־תלויים, אזי קיום של אחד לא משפיע על ההסתברות של המאורע האחר. בכדי להוכיח זאת נבחן את ההסתברות המותנה $P(A|B)$. אם המאורעות הם בלתי תלויים, אזי

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

5.13 מסקנה. () אם המאורעות A ו- B בלתי תלויים, אזי

$$P(A|B)=P(A)\ .$$

לחילופין, במילים פשוטות, ההסתברות של מאורע A איננה מושפעת מקיומו או אי־קיומו של B.

5.14 דוגמא. בספרייה ישנם 10 ספרי הסתברות. 5 ספרים הכוללים פתרונות ו־5 ללא פתרונות. הספרנית איבדה ספר אחד מתוך ה־10 (כלשהו). אלון מגיע לקחת ספר, מקבל ספר מתוך התשעה הנותרים באקראי ומחזיר לאחר יומיים. לאחר שבועיים מגיע בן ולוקח גם כן ספר 1 מתוך ה־9 ומחזיר לאחר יומיים. נגדיר את המאורעות A־ הספר של אלון כלל פתרונות, ו־B ־ הספר של בן כלל פתרונות. האם A ו־B תלויים?

פיתרון. לכאורה, אין סיבה שתהיה תלות. מצד אחד, אלון לקח ספר והחזיר לאחר יומיים. שבועיים לאחר מכן, אחרי שאלון החזיר כבר את הספר, מגיע בן ולוקח ספר כלשהו, לכן למה שתהיה תלות?! אבל במקרה זה אכן המאורעות תלויים. נראה זאת בעזרת חישוב ישיר. נגדיר C ־ הספר שאבד כלל פתרונות. משיקולי סימטריה, $P(A)=P(B)$ ולכן נחשב רק אחד מהם.

$$\begin{aligned}P(A)=&P(A|C)P(C)+P(A|\bar{C})P(\bar{C})\\=&\frac{4}{9}.\frac{1}{2}+\frac{5}{9}.\frac{1}{2}\\=&\frac{1}{2},\end{aligned}$$

על כן נקבל ש־ $P(A).P(B)=\frac{1}{4}$. מצד שני,

$$\begin{aligned}P(A\cap B)=&P(A\cap B|C)P(C)+P(A\cap B|\bar{C})P(\bar{C})\\=&\left(\frac{4}{9}\right)^2.\frac{1}{2}+\left(\frac{5}{9}\right)^2.\frac{1}{2}\\\approx&0.253\neq P(A).P(B),\end{aligned}$$

לכן המאורעות תלויים. ההסבר הוא פשוט יחסית, ניקח את הסיטואציה למצב קיצון ונניח שבין אלון ובן היו מגיעים עוד 1000 איש ומבצעים את הניסוי, והיינו שמים לב שרק $\frac{4}{9}$ מתוכם קיבלו ספר עם פתרונות. במצב כזה היינו חושבים שככל הנראה אבד ספר עם פתרונות ■

6 הסתברות מותנית תרגילים 7-13

סיכום חוקים ונוסחאות עבור הסתברות מותנית

6.1 הגדרה. (הסתברות מותנית) לכל צמד מאורעות A ו־B כך ש־B בעל הסתברות חיובית, ההסתברות של מאורע A בהינתן מאורע B מוגדרת להיות

$$P(A|B)=\frac{P(A\cap B)}{P(B)}\ . \tag{6.1}$$

6.2 חוק. (כלל כפל) הגרסה הבסיסית של כלל הכפל מתקיימת עבור צמד מאורעות A ו־ B כאשר B בעל הסתברות חיובית ($P(B)>0$). נעזר בהגדרה להסתברות מותנה [עיין משוואה (**??**)], ועל ידי העברת אגפים נקבל

$$P(A\cap B)=P(A|B)P(B)\ . \tag{6.2א}$$

באופן כללי לכל קבוצת מאורעות $A_1,\dots,A_n$ כך ש

$$P\left(\bigcap_{i=1}^{n}A_i\right)>0$$

מתקיים

$$P\left(\bigcap_{i=1}^{n}A_i\right)=P(A_1)P(A_2|A_1)P(A_3|A_1\cap A_2)\dots P\left(A_n|\bigcap_{i=1}^{n-1}A_i\right)\ . \tag{6.2ב}$$

6.3 חוק. (נוסחת ההסתברות השלמה) נניח כי המאורעות $B_1,\dots,B_n$ הם זרים בזוגות, בעלי הסתברות חיובית כ"א, ומחלקים את מרחב המדגם באופן מלא. ז"א

$$B_i\cap B_j=\phi$$

לכל $i\neq j$ ובנוסף

$$\bigcup_{i=1}^{n}B_i=\Omega\ ,$$

כאשר $P(B_i)>0$. אזי לכל מאורע

$$P(A)=\sum_{i=1}^{n}P(A|B_i)P(B_i)\ . \tag{6.3}$$

6.4 חוק. (חוק בייס) עבור כל צמד מאורעות A ו B בעלי הסתברות חיובית מתקיים

$$P(A|B)=\frac{P(B|A)P(A)}{P(B)} \tag{6.4}$$

6.5 הגדרה. (תלות בין מאורעות)
צמד מאורעות A ו- B הם בלתי־תלויים אם

$$P(A\cap B)=P(A)P(B)\ . \tag{6.5א}$$

מנגד, המאורעות A ו- B הם **תלויים** אם הם לא בלתי־תלויים ז"א אם

$$P(A\cap B)\neq P(A)P(B)\ . \tag{6.5ב}$$

ישירות מן ההגדרה נסיק כי אם המאורעות בלתי־תלויים, אזי קיום של אחד לא משפיע על ההסתברות של המאורע האחר. בכדי להוכיח זאת נבחן את ההסתברות המותנה $P(A|B)$. אם המאורעות הם בלתי תלויים, אזי

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

6.6 מסקנה. () אם המאורעות A ו- B בלתי תלויים, אזי

$$P(A|B) = P(A)\ .$$

לחילופין, במילים פשוטות, ההסתברות של מאורע A איננה מושפעת מקיומו או אי־קיומו של B.

6.7 הגדרה. () קבוצה של מאורעות $A_1,\dots,A_n$
הם בלתי־תלויים אם

$$P\left(\bigcap_{i=1}^{n} B_i\right) = \prod_{i=1}^{n} P(B_i)$$

לכל $B_i \in \left\{A_i, \bar{A}_i\right\}$ זאת אומרת, ישנן 2^n משוואות אשר אמורות להתקיים בכדי שהמאורעות יהיו ב"ת.

הערה: נוסחאות עבור ניסויים של הטלת קוביות

המספר קומבינטוריות הקיימות לזרוק "n" קוביות, אשר לכל קוביה יש s פנים ולקבל "p":

$$N(p,n,s) = \sum_{k=0}^{\lfloor (p-n)/s\rfloor} (-1)^k \binom{n}{k}\binom{p-sk-1}{n-1}. \tag{6.6}$$

כאשר $\lfloor x\rfloor$ הוא פונקציה הריצפה אשר נותנת המספר שלם הכי קרוב ופחות מ "x". לדוגמה $\lfloor 3.6\rfloor = 3$ או $\lfloor 2.1\rfloor = 2$.

ההסתברות לזרוק "n" קוביות, אשר לכל קוביה יש S פנים ולקבל "p":

$$P(p,n,s) = \frac{1}{s^n} N(p,n,s). \tag{6.7}$$

מספר קומבינטוריות הקיימות לזרוק "n" קוביות רגילות (בנות 6 פנים) ולקבל "p":

$$N(p,n,6) = \sum_{k=0}^{\lfloor (p-n)/6\rfloor} (-1)^k \binom{n}{k}\binom{p-6k-1}{n-1} = \begin{cases} \binom{p-1}{n-1} & p < n+6 \\ \binom{p-1}{n-1} - n\binom{p-7}{n-1} & n+6 \le p < n+12 \\ \binom{p-1}{n-1} - n\binom{p-7}{n-1} - \frac{1}{2}n(n-1)\binom{p-13}{n-1} & n+12 \le p < n+18 \end{cases} \tag{6.8}$$

מספר קומבינטוריות הקיימות לזרוק "$n=2$" קוביות רגילות (בנות 6 פנים) ולקבל "p":

$$N(p,2,6) = \sum_{k=0}^{\lfloor (p-2)/6\rfloor} (-1)^k \binom{2}{k}\binom{p-6k-1}{1} = \begin{cases} (p-1) & p < 8 \\ (p-1) - 2(p-7) & 8 \le p < 14 \end{cases} \tag{6.9}$$

לדוגמה

$$N(7,2,6) = 6.$$

תרגילים: הסתברות מותנה ואי־תלות בין מאורעות

6.8 דוגמא. (שליפת קלפים ברצף) שולפים שלוש קלפים אחד אחרי השני ללא החזרה מחבילת קלפים תקינה. מהי ההסתברות שקלף הראשון הוא red ace, קלף השני הוא 10 או jack , וקלף שלישי הוא יותר גדול מ 3 ופחות מ 7.

פיתרון. השאלה היא, מהי ההסתברות של המאורע $A_1 \cap A_2 \cap A_3$ כאשר

- $A_1 =$ המאורע של הקלף הראשון הוא red ace,
- $A_2 =$ המאורע ש קלף השני הוא או 10 או jack,
- $A_3 =$ המאורע ש קלף שלישי הוא יותר גדול מ 3 ופחות מ 7.

$$P(A_1) = \frac{2}{52}, \quad P(A_2|A_1) = \frac{8}{51}, \quad P(A_3|A_1 \cap A_2) = \frac{12}{50} \ ,$$

לכן לפי נוסחת ההסתברות ההשלמה,

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) = \frac{8}{5525} \ .$$

■

6.9 דוגמא. (כדים של כדורים בנות צבעים שונים) בכד יש 4 כדורים לבנים, ו-3 כדורים שחורים. בכד שני 3 כדורים לבנים ו 5 כדורים שחורים. כדור אחד הוצא מכד ראשון ומונח בכד שני ללא לבדוק את הצבע שלו. מהי ההסתברות אשר כדור אשר הוצא מהכד שני הוא שחור?

פיתרון. נגדיר את המאורעות הבאים:

- $B_1 =$ כדור שחור הוצא מכד ראשון
- $B_2 =$ כדור שחור הוצא מכד שני
- $W_1 =$ כדור לבן הוצא מכד ראשון

השאלה היא מהי ההסתברות של האיחוד המאורעות $B_1 \cap B_2$ ו $W_1 \cap B_2$, כלומר

$$(B_1 \cap B_2) \cup (W_1 \cap B_2) \ ,$$

לקחת בחשבון שהם מאורעות זרים:

$$(B_1 \cap B_2) \cap (W_1 \cap B_2) = \phi \ .$$

איור 2: עץ הסתברות עבור דוגמה לעיל

$$\begin{aligned} P\left((B_1 \cap B_2) \cup (W_1 \cap B_2)\right) =& P(B_1 \cap B_2) + P(W_1 \cap B_2) \\ =& P(B_1)P(B_2|B_1) + P(W_1)P(B_2|W_1) \\ =& \frac{3}{7}.\frac{6}{9} + \frac{4}{7}.\frac{5}{9} \\ =& \frac{38}{63} \ . \end{aligned}$$

■

6.10 דוגמא. הרכב הסטודנטים בתואר ראשון הוא כדלקמן:

- 50% מהסטודנטים הם בשנה א' לתואר,
- 30% מהסטודנטים הם בשנה ב' לתואר,
- 20% מהסטודנטים הם בשנה ג' לתואר.

בנוסף, מצבם המשפחתי של הסטודנטים הנ"ל הוא:

- 20% מהסטודנטים שנמצאים בשנה א' ־ נשואים,
- 30% מהסטודנטים שנמצאים בשנה ב' ־ נשואים,
- 40% מהסטודנטים שנמצאים בשנה ג' ־ נשואים.

ענו על השאלות הבאות.

1. בהנחה שפגשנו באקראי את אלון, סטודנט מתואר ראשון, מה הסיכוי שהוא נשוי?
2. מסתבר שאלון אכן נשוי, מה הסיכוי שהוא משנה ג'?
3. פגשנו סטודנט אחר, בן, גם כן מתואר ראשון ומסתבר שהוא איננו נשוי, מה הסיכוי שהוא משנה ג'?

פיתרון. בשאלה זו נשתמש בעיקר בנוסחת ההסתברות השלמה וחוק בייז.

1. נסמן ב־ M את המאורע בו הסטודנט נשוי וב־ I ו II ן III את המאורעות בהם הסטודנט הוא בשנה א',ב' ו־ג' לתואר בהתאמה.

$$\begin{aligned} P(M) =& P(M|I)P(I) + P(M|II)P(II) + P(M|III)P(III) \\ =& \frac{2}{10}.\frac{5}{10} + \frac{3}{10}.\frac{3}{10} + \frac{4}{10}.\frac{2}{10} = 0.27 \ . \end{aligned}$$

2. נשתמש בחוק בייז מסיבה פשוטה - נתוני השאלה מציגים הסתברויות להיות נשוי בהתאם לשנת הלימוד, אבל ההתנייה ההפוכה לא נתונה. לכן,

$$P(III|M) = \frac{P(M|III)P(III)}{P(M)} = \frac{0.4.0.2}{0.27} = \frac{0.08}{0.27} \approx 0.296 \ .$$

3. נחשב באופן דומה לחישוב קודם, אלא שהפעם ההתנייה היא על מאורע שונה.

$$P(III|\bar{M}) = \frac{P(\bar{M}|III)P(III)}{P(\bar{M})} = \frac{0.6.0.2}{1-0.27} = \frac{0.12}{0.73} \approx 0.164 \ .$$

■

6.11 דוגמא. בספרייה ישנם 10 ספרי הסתברות. 5 ספרים הכוללים פתרונות ו-5 ללא פתרונות. הספרנית איבדה ספר אחד מתוך ה-10 (כלשהו). אלון מגיע לקחת ספר, מקבל ספר מתוך התשעה הנותרים באקראי ומחזיר לאחר יומיים. לאחר שבועיים מגיע בן ולוקח גם כן ספר 1 מתוך ה-9 ומחזיר לאחר יומיים. נגדיר את המאורעות A- הספר של אלון כלל פתרונות, ו-B - הספר של בן כלל פתרונות. האם A ו-B תלויים?

פיתרון. לכאורה, אין סיבה שתהיה תלות. מצד אחד, אלון לקח ספר והחזיר לאחר יומיים. שבועיים לאחר מכן, אחרי שאלון החזיר כבר את הספר, מגיע בן ולוקח ספר כלשהו, לכן למה שתהיה תלות?! אבל במקרה זה אכן המאורעות תלויים. נראה זאת בעזרת חישוב ישיר. נגדיר C - הספר שאבד כלל פתרונות. משיקולי סימטריה, $P(A)=P(B)$ ולכן נחשב רק אחד מהם.

$$\begin{aligned} P(A) =& P(A|C)P(C) + P(A|\bar{C})P(\bar{C}) \\ =& \frac{4}{9}.\frac{1}{2} + \frac{5}{9}.\frac{1}{2} \\ =& \frac{1}{2}, \end{aligned}$$

על כן נקבל ש- $P(A).P(B)=\frac{1}{4}$. מצד שני,

$$\begin{aligned} P(A\cap B) =& P(A\cap B|C)P(C) + P(A\cap B|\bar{C})P(\bar{C}) \\ =& \left(\frac{4}{9}\right)^2.\frac{1}{2} + \left(\frac{5}{9}\right)^2.\frac{1}{2} \\ \approx& 0.253 \neq P(A).P(B), \end{aligned}$$

לכן המאורעות תלויים. ההסבר הוא פשוט יחסית, ניקח את הסיטואציה למצב קיצון ונניח שבין אלון ובן היו מגיעים עוד 1000 איש ומבצעים את הניסוי, והיינו שמים לב שרק $\frac{4}{9}$ מתוכם קיבלו ספר עם פתרונות. במצב כזה היינו חושבים שככל הנראה אבד ספר עם פתרונות ■

6.12 דוגמא. (הכד של פוליה) בכד 5 כדורים שחורים ו-3 לבנים. בכל הוצאה של כדור מהכד, מחזירים אותו חזרה יחד עם 4 נוספים מאותו הצבע.

1. מה ההסתברות שהכדור הראשון שנוציא יהיה שחור?
2. מה ההסתברות שהכדור השני שנוציא יהיה שחור?
3. מה ההסתברות שהכדור ה-100 שנוציא יהיה שחור?

פיתרון. פתרון: נפתור את צמד הסעיפים הראשונים בעזרת חישובים קומבינטוריים ונוסחת ההסתברות השלמה.

1. הפתרון לסעיף הראשון הוא מיידי והתשובה היא $\frac{5}{8}$. (חלוקה של גודל המאורע במרחב המדגם).

2. נחשב בעזרת נוסחת ההסתברות השלמה. נסמן ב B_i, W_i $i = 1, 2$ את המאורעות בהם הכדור ה־ i הוא שחור או לבן, בהתאמה.

$$\begin{aligned} P(B_2) =& P(B_2|B_1)P(B_1) + P(B_2|W_1)P(W_1) \\ =& \frac{9}{12}.\frac{5}{8} + \frac{5}{12}.\frac{3}{8} = \frac{45+15}{96} = \frac{5}{8} \end{aligned}$$

באופן מפתיע קיבלנו שההסתברות להוציא שחור בשלב הראשון ובשלב השני שוות.

3. לפני שנענה על הסעיף הנוכחי, ננסה למצוא הוכחה יותר פשוטה שלא תצריך התניות רבות. נתחיל במצב יחסית פשוט: מה היה קורה לו היו לנו רק 2 כדורים בכד בהתחלה ־ 1 שחור ו־1 לבן? אזי משיקולי סימטריה ההסתברות שהכדור ה־ 100 היה שחור היא (מאחר והכדור ה־ 100 הוא שחור או לבן, וישנה סימטריה מלאה בין השחורים ללבנים, אזי ההסתברות שנוציא כל אחד מהם בפעם ה־ 100 היא שווה). מה לגבי מצב בו ישנם 8 כדורים בכד ללא צבעים אשר ממוספרים מ־ 1 עד 8 (לאחר הוצאה מחזירים את הכדורים עם אותו המספר של הכדור שהוצא)? במצב זה, גם כן יש סימטריה בין שמונת הכדורים ולכן ההסתברות להוציא כדור עם ערך 1 . כעת נוכל להניח שיש מספרים על הכדורים בבעיה המקורית, הסיכוי שנוציא את כדורים 1 עד 5) אלו הכדורים הצבועים בשחור) הוא $\frac{5}{8}$ כסכום ההסתברויות להוציא כל מספר מ־ 1 עד 5 בנפרד. ולסיכום, ההסתברות שנוציא כדור שחור בפעם ה 100 היא $\frac{5}{8}$ וזה נכון עבור כל שלב שנבחר

■

6.13 דוגמא. לבן יש מטבע כחול ולאלון מטבע אדום. למטבע הכחול הסתברות p לנחות על 'עץ' (H) ולמטבע האדום הסתברות q. .אלון ובן משחקים את המשחק הבא: בכל תור, מטילים שניהם את המטבע שלהם באופן ב"ת באחר או בהטלות הקודמות. אם התקבל פעמיים H , אז הם מחליפים ביניהם את המטבעות. אחרת, כל אחד נשאר עם המטבע שהתחיל איתו את הסיבוב. התהליך חוזר על עצמו שוב ושוב. לכל $n \geq 1$, נגדיר את המאורע A_n כמאורע שבתום n סיבובים אלון מחזיק את המטבע האדום. חשבו את $P(A_2)$, ומצאו עבור אילו ערכי p ו q המאורעות A_2 ו A_3 ב"ת.

פיתרון. ננסה לחשוב מתי A_2 אפשרי. זהו המאורע בו אחרי 2 סיבובים אלון מחזיק את המטבע האדום איתו התחיל. דבר זה יכול להתרחש אם ורק אם לא הוחלפו מטבעות כלל, או לחילופין, התבצעו שתי החלפות. ההסתברות להחלפה של מטבע בסיבוב נתון היא pq לכן נקבל ש־

$$P(A_2) = (1-pq)^2 + (pq)^2 \ .$$

דרך אחת לבחון אי-תלות היא בעזרת ההגדרה

$$P(A_2 \cap A_3) = P(A_2)P(A_3) \ .$$

מנגד, ישנה עוד דרך להראות אי־תלות וזאת על ידי השיוויון

$$P(A_3|A_2) = P(A_3) \ .$$

מדוע זה אפשרי? ובכן, לפי הגדרה, נובע ש־

$$P(A_3|A_2) = \frac{P(A_3 \cap A_2)}{P(A_2)},$$

ואם בנוסף אנו יודעים ש

$$P(A_3|A_2) = P(A_3)$$

מתרחשים אזי נובע ש

$$\frac{P(A_3 \cap A_2)}{P(A_2)} = p(A_3),$$

כנדרש. נחשב מפורשות את $P(A_3|A_2)$ ואת $P(A_3)$. נסמן $s = pq$ ונקבל

$$P(A_3) = (1-s)^3 + 3(1-s)s^2 \ ,$$

לכן, נבדוק מתי מתקיים

$$1 - s = (1-s)^3 + 3(1-s)s^2 \ .$$

פתרון ראשון הוא $s = 1$. אחרת

$$\begin{aligned} 1 =&(1-s)^2 + 3s^2, \\ 0 =& -2s + 4s^2, \\ 0 =& s(2s-1) \ . \end{aligned}$$

לכן נקבל את הפתרונות $s_1 = 0, s_2 = 0, s_3 = \frac{1}{2}$. נתרגם זאת למונחי p ו q באופן הבא:

$$p = \frac{1}{2q} \Leftrightarrow s = \frac{1}{2} \ ,$$

ואלו המקרים בהם ישנה אי-תלות. התוצאה שקיבלנו היא די הגיונית ואפילו אינטואיטיבית. כאשר $pq = 0$ או $pq = 1$ אז מקבלים שהמאורעות A_2 ו A_3 מתרחשים בהסתברות 0 או 1 ולכן הם ב"ת בכל מאורע אחר. בנוסף, כאשר $pq = \frac{1}{2}$ אז בכל שלב יש סיכוי שווה להיות בכל מצב ללא תלות בעבר. למעשה, ישנה סימטריה בבעיה מבחינת מיקום המטבעות אצל אלון ובן ולכן מתקיימת אי-תלות. ■

6.14 דוגמא. על השולחן מונחים 2 כדים: בכד הראשון (כד א') שלושה כדורים לבנים וחמישה כדורים אדומים; בכד השני (כד ב') ישנם שישה כדורים לבנים ושלושה כדורים אדומים. אלון מטיל מטבע הוגן. אם תוצאת ההטלה היא "עץ", הוא יוציא מכד א' 2 כדורים עם החזרה. אחרת, הוא יוצא מכד ב' 2 כדורים עם החזרה. נסמן ב־A את המאורע שהכדור הראשון לבן, ב־B את המאורע שהכדור השני לבן וב־C את המאורע בו נבחר כד א'. האם המאורעות A, B תלויים? האם המאורעות A, B תלויים בהינתן C?

פיתרון. נתחיל בבניית מרחב המדגם. נסמן את הכדורים בכד א' במספרים $1, \ldots, 8$ כאשר שלושת הראשונים יהיו הכדורים הלבנים והיתר אדומים. באופן דומה נסמן את הכדורים בכד ב' במספרים $1, \ldots, 9$ כאשר ששת הראשונים לבנים והיתר אדומים. מרחב המדגם מורכב משלשות כאשר הקוארדינטה הראשונה היא זהות הכד שנבחר והקוארדינטות האחרונות הן ערכי הכדורים שהוצאו מהכד שנבחר. נשים לב שמרחב המדגם מכיל

$$|\Omega| = 8^2 + 9^2 = 145$$

אפשרויות שונות. בנוסף נשים לב שמערכת המדגם אינו מרחב מדגם שווה הסתברות. לדוגמא: ההסתברות להוציא צמד כדורים כלשהם מכד א' היא

$$\frac{1}{2} . \frac{1}{6}$$

בעוד ההסתברות להוציא צמד כדורים מכד ב' היא

$$\frac{1}{2} . \frac{1}{81} .$$

75נחשב את ההסתברות של כל מאורע בנפרד.

$$P(A) = \frac{1}{2} . \frac{3.8}{64} + \frac{1}{2} + \frac{6.9}{81} = \frac{3}{16} + \frac{3}{9} = \frac{75}{144}$$

כאשר ביצענו התנניה על הכד שנבחר, הסתברות $\frac{1}{2}$ לכל כד, ולאחר מכן ספרנו את מספר האפשרויות בכל חלק ממרחב המדגם. באופן זהה לחישוב הקודם, נקבל ש־

$$P(B) = \frac{1}{2} . \frac{3.8}{64} + \frac{1}{2} . \frac{6.9}{81} = \frac{3}{16} + \frac{3}{9} = \frac{75}{144} ,$$

וכמובן

$$P(C)=\frac{1}{2}\ .$$

נחשב כעת ההסתברות למאורע $P(A\cap B)=A\cap B$

$$\frac{1}{2}.\frac{3/3}{64}+\frac{1}{2}.\frac{6.6}{81}=\frac{9}{128}+\frac{2}{9}=\frac{337}{1152}\ .$$

נשים לב ש

$$P(A)P(B)\neq P(A\cap B)$$

ולכן המאורעות תלויים. נעבור לחישוב השני בשאלה. נחשב את ההסתברות של כל מאורע מותנה בנפרד את ההסתברות של חיתוך המאורעות תחת ההתנייה

$$P(A|C)=\frac{3.8}{64}=\frac{3}{8}=P(B|C)$$

תלויים

$$P(A\cap B|C)=\frac{3.3}{64}=\frac{9}{64}=P(A|C)P(B|C)$$

לכן המאורעות הללו בלתי־תלויים בהינתן ■

6.15 דוגמא. נקבע $1\leq k\leq n$ מספרים שלמים. בספארי נולדו לקרנף n גורים אחד אחרי השני כאשר כל גור יכל להיוולד לכל מין בהסתברות שווה ובאופן ב"ת במין הגורים האחרים. דוליטול מסתובב בגן החיות ופוגש גור שנבחר באקראי באופן אחיד מבין הגורים. נסמן ב־A את המאורע בו k הגורים הראשונים שנולדו הם ממין זכר. נ סמן ב־B את המאורע שנולדו לפחות k גורים ממין זכר. נסמן ב־C את המאורע בו דוליטול פוגש פוגש זכר. עבור $k=1,n=2$ חשבו את $P(C|B)$. עבור $n=12,k=4$ חשבו חשבו $P(A|C)$.

פיתרון. נסמן ב־ D_i את המאורע בו נולדו i גורים ממין זכר. נעזר בפירוק של מרחב המדגם על פי המאורעות הללו.

$$\begin{aligned}
P(C|B)=&\frac{P(C\cap B)}{P(B)}\\
=&\frac{\sum_{i=0}^{2}P(C\cap B\cap D_i)}{P(B)}\\
=&\frac{P(C\cap B\cap D_1)+P(C\cap B\cap D_2)}{P(B)}\\
=&\frac{P(C\cap D_1)+P(C\cap D_2)}{1-P(B^c)}\\
=&\frac{P(C|D_1)P(D_1)+P(C|D_2)P(D_2)}{1-P(B^c)}\\
=&\frac{\frac{1}{2}.2.\left(\frac{1}{2}\right)^2+1.\left(\frac{1}{2}\right)^2}{1-\left(\frac{1}{2}\right)^2}\\
=&\frac{\frac{1}{2}}{\frac{3}{4}}\\
=&\frac{2}{3}\ .
\end{aligned}$$

נשתמש בחוק בייס לפתרון המשך התרגיל. נסמן ב־E את המאורע בו דוליטול פוגש אחד מארבעת הגורים הראשונים שנולדו.

$$\begin{aligned}
P(A|C) =& \frac{P(C|A)P(A)}{P(C)} \\
=& \frac{\left[P(C|A\cap E)P(E|A)+P(C|A\cap \bar{E})P(\bar{E}|A)\right]\left(\frac{1}{2}\right)^4}{\frac{1}{2}} \\
=& \frac{\left[1.\frac{4}{12}+\frac{1}{2}.\frac{8}{12}\right]\left(\frac{1}{2}\right)^4}{\frac{1}{2}} \\
=& \frac{1}{12}\ .
\end{aligned}$$

■

7 משתנה מקרי חד מימדי בדיד 7-20

7.1 הגדרה. (משתנה מקרי בדיד) משתנה מקרי (חד מימדי) בדיד X (מ"מ) הוא פונקציה על המרחב מדגם המתאימה ערך מספרי יחיד לכל תוצאה במרחב הסתברות,

$$X : \Omega \to \mathbb{R} \ .$$

הפונקציה משתנה בהתאם לתוצאות הניסוי, בעוד תוצאת הניסוי היא סטוכסטית (אקראית) ולכן השינוי הוא מקרי. בשלב הראשון אנחנו נעסוק רק במשתנים מקריים בדידים וחד-מימדים. זאת אומרת, משתנים מקריים שמקבלים ערכים בדידים (בניגוד למשתנים מקריים רציפים שיכולים לקבל רצף של ערכים), ובנוסף כל משתנה מחזיר מספר בודד $X(\omega) \in \mathbb{R}$ כאשר $\omega \in \Omega$ ו- Ω הינו המרחב המדגם.

7.2 דוגמא. מטילים 2 קוביות. נסמן ב- X את סכום התוצאות שהתקבלו. כמו כן

$$\begin{aligned}
&\omega = \{(1,1)\}, && X(\omega) = 2,\\
&\omega = \{(2,1),(1,2)\}, && X(\omega) = 3,\\
&\omega = \{(1,3),(2,2),(3,1)\}, && X(\omega) = 4,\\
&\vdots\\
&\omega = \{(6,6)\}, && X(\omega) = 12.
\end{aligned}$$

זו דוגמה של מ"מ כאשר X הוא המשתנה המקרי.

7.3 דוגמא. מטילים 3 קוביות הוגנות. נגדיר משתנה מקרי X להיות התוצאה המקסימלית מבין ההטלות. כמו כן

$$\begin{aligned}
&\omega = \{(1,1,1)\}, && X(\omega) = 1,\\
&\omega = \{(2,1,1),(1,2,1),(1,1,2)\}, && X(\omega) = 2,\\
&\vdots\\
&\omega = \{(3,2,1),(2,3,1),(1,2,3),\dots\}, && X(\omega) = 3,\\
&\vdots\\
&\omega = \{(6,6,6)\}, && X(\omega) = 6.
\end{aligned}$$

זו דוגמה של מ"מ כאשר X הוא המשתנה המקרי.

7.4 הגדרה. (תומך של מ"מ בדיד) התומך של משתנה מקרי בדיד X זו קבוצת המספרים ש-X יכול להיות שווה בהסתברות חיובית, קרי

$$\text{Supp}(X) = \{k \in \mathbb{R} : P(X = k) > 0\} \ .$$

7.5 הגדרה. (פונקצית הסתברות / פונקצית התפלגות) הפונקצית ההתפלגות של משתנה מקרי בדיד (מ"מ) X מוגדרת להיות הפונקציה $f_X(k)$ המקבלת המ"מ X ומחזירה את ההסתברות כי למ"מ X יש ערך k:

$$f_X(k) = P(X = k) \ ,$$

עם התכונות

1. $\sum_{k \in X} f_X(k) = 1$.
2. $f_X(k) \geq 0 \quad \forall k$.

זאת אומרת, ההתפלגות של X בנקודה k מציינת את ההסתברות של המאורע

$$\{X = k\} \ .$$

למעשה, פונקצית התפלגות ופונקצית הסתברות הן די דומות מבחינת התכונות, אבל כדאי לא להתבלבל בינהן. פ ונקצית הסתברות מוגדרת על מרחבי מדגם ולכן הקלט של פונקצית ההסתברות יכול להיות מאוד מורכב. מנגד, התפלגויות מקבלות מספרים ומחזירות מספרים.

7.6 דוגמא. נטיל קוביה הוגנת ונניח כי X מחזיר את תוצאת ההטלה. אזי

$$\text{Supp}(X) = \{1,2,3,4,5,6\}$$

וההתפלגות של X היא

$$f_X(k) = \begin{cases} \frac{1}{6} & k \in \text{supp}(X) \\ 0 & \text{אחרת} \end{cases}$$

במילים, $f_X(k) = \frac{1}{6}$ לכל k בתומך, ו- 0 אחרת. חשוב לזכור שהתפלגות היא פונקציה ולכן יש צורך לתאר אותה בכל נקודה. משמע, זה לא מספיק לציין את ערכי הפונקציה רק על התומך, אלא בכל נקודה $k \in \mathbb{R}$.

7.7 דוגמא. אלון הוא סוחר ממולח בשוק ההון. הוא משקיע סכום של $\$1000$ במכשיר פיננסי אשר בכל שנה מכפיל את הכסף בסיכוי $\frac{2}{3}$ או מפסיד חצי מההשקעה בסיכוי $\frac{1}{3}$. השנים הן בלתי תלויות. נגדיר את X להיות שווי ההשקעה של אלון לאחר שנתיים. התומך של X כעת הוא

$$\text{supp}(X) = \{250, 1000, 4000\}\ ,$$

על כן, ההתפלגות של X היא

$$f_X(1000) = \frac{4}{9}, \quad f_X(4000) = \frac{4}{9}, \quad f_X(250) = \frac{1}{9}$$

ו - 0 אחרת.

7.8 דוגמא. מטילים 2 קוביות. נסמן ב- X את סכום התוצאות שהתקבלו. מצאו את ההתפלגות של X.

פיתרון. נתחיל מהתומך. התוצאות האפשרויות בסכום של 2 קוביות הן המספרים הטבעיים מ- 2 עד 12. בכדי לחשב את ההסתברות לכל ערך נשתמש במרחב מדגם סימטרי: גודל מרחב המדגם הוא $6^2 = 36$, ככמות האפשרויות לבחור תוצאה לקוביה הראשונה מתוך 6 ותוצאה לקוביה השניה מתוך 6. נחשב את ההסתברות של כל תוצאה בנפרד:

$$\begin{aligned}
P(X=2) =& \frac{|\{(1,1)\}|}{6^2} &&= \frac{1}{36}, \\
P(X=3) =& \frac{|\{(2,1),(1,2)\}|}{6^2} &&= \frac{2}{36}, \\
P(X=4) =& \frac{|\{(1,3),(2,2),(3,1)\}|}{6^2} &&= \frac{3}{36}, \\
&\vdots \\
P(X=7) =& \frac{|\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}|}{6^2} &&= \frac{6}{36}, \\
P(X=8) =& \frac{|\{(2,6),(3,5),(4,4),(5,3),(6,2)\}|}{6^2} &&= \frac{5}{36}, \\
P(X=9) =& \frac{|\{(3,6),(4,5),(5,4),(6,3)\}|}{6^2} &&= \frac{4}{36}, \\
&\vdots \\
P(X=12) =& \frac{|\{(6,6)\}|}{6^2} &&= \frac{1}{36},
\end{aligned}$$

ובאופן דומה ממשיכים לחשב את האפשרויות לכל הערכים האפשריים בתומך. באופן כללי קל לראות שהנוסחא, כאשר p זו הסכום של ה 2 הקוביות, היא (עיין משוואה (6.9) לעייל)

$$P(X=p) = \frac{1}{36} \times \begin{cases} (p-1) & p < 8 \\ (p-1) - 2(p-7) & 8 \le p < 14 \end{cases}$$

■

7.9 הגדרה. (פונקציית התפלגות מצטברת) פונקציית התפלגות מצטברת מסומנת ע"י $F_X(x)$ של מ"מ בדיד X בעל פונקציית התפלגות $f_X(k)$ מוגדרת להיות

$$F_X(x) = \sum_{\substack{k\in X \\ k\leq x}} f_X(k) \ .$$

במילים אחרות, פונקציית התפלגות מצטברת $F_X(x)$ מוגדרת להיות הפונקציה שתחומה הוא הישר הממשי וטווחה הקטע $[0,1]$ ומקיימת

$$F_X(x) = P(X \leq x) \qquad \forall x \in X \ .$$

7.10 דוגמא. אלון הוא סוחר ממולח בשוק ההון. הוא משקיע סכום של $1000\$$ במכשיר פיננסי אשר בכל שנה מכפיל את הכסף בסיכוי $\frac{2}{3}$ או מפסיד חצי מההשקעה בסיכוי $\frac{1}{3}$. השנים הן בלתי תלויות. נגדיר את X להיות שווי ההשקעה של אלון לאחר שנתיים. מצאו את פונקציית ההתפלגות המצטברת של X וציירו אותה.

פיתרון. כפי שציינו בדוגמה 9.2, התומך של X כעת הוא

$$\text{supp}(X) = \{250, 1000, 4000\} \ .$$

על כן, פונקציית ההתפלגות המצטברת של X היא

$$F_X(k) = P(X \leq k) = \begin{cases} 0 & k < 250, \\ \frac{1}{9} & 250 \leq k < 1000, \\ \frac{5}{9} & 1000 \leq k < 4000, \\ 1 & 4000 \leq k. \end{cases}$$

■

תוחלת היא מדד אמצע, לכן היא מעניקה לנו אינדיקציה מאוד בסיסית בנוגע לערך הממוצע של משתנה מקרי נתון. התוחלת לוקחת את הערכים השונים שמקבל המשתנה המקרי ומשקללת אותם, בהתאם לסבירות של כל ערך, לכדי מספר בודד שמציין את הערך הממוצע של אותו המשתנה. קרי, בדומה למדד המחירים לצרכן אשר משקלל את עלות סל צריכה בהתאם לצריכה, התוחלת היא ממוצע משוקלל של ערכי המשתנה המקרי בהתאם לסבירות של הערכים. בפועל, החישוב של התוחלת הוא מיידי: לוקחים את כל הערכים שמקבל המ"מ X, מכפילים כל ערך בהסתברותו וסוכמים.

7.11 הגדרה. (תוחלת של משתנה נקרי בדיד)

יהי X משתנה מקרי בדיד בעל פונקציית התפלגות $f_X(k)$. התוחלת (expectation) של X היא

$$\mu_X \equiv E[X] = \sum_{k\in X} k f_X(k) \ .$$

7.12 דוגמא. התוחלת של הטלת קוביה היא

$$E[X] = \sum_{k=1}^{6} k\frac{1}{6} = \frac{7}{2} \ .$$

7.13 דוגמא. בן הוא סוחר מתוחכם ולכן הוא מציע לאלון להפקיד אצלו 100 שקלים בתחילת השנה. בסיכוי של $\frac{2}{5}$ הסכום יוכפל, ובסיכוי $\frac{2}{5}$ הסכום יצטמצם בחצי בסוף השנה. מה תהיה תוחלת הכסף של אלון לאחר קבלת ההצעה? נסמן ב- X את שווי כסף של אלון בסוף השנה.

$$E[X] = 200.\frac{2}{5} + 50.\frac{3}{5} = 110.$$

זאת אומרת שהצעתו של בן מגלמת תוחלת רווח של 10%!

7.14 מסקנה. (ליניאריות התוחלת)

1. עבור כל קבוע $a \in \mathbb{R}$ התוחלת של a היא

$$E[a] = a\ .$$

2. עבור כל משתנה מקרי X וקבוע $a \in \mathbb{R}$ התוחלת של aX היא

$$E[aX] = aE[X]\ .$$

3. עבור כל צמד משתנים מקריים X ו- Y התוחלת של $X + Y$ היא

$$E[X + Y] = E[X] + E[Y]\ .$$

4. כמובן שהתכונה הנ"ל ניתנת להכללה גם עבור מספר כללי של משתנים , שכן

$$E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E[X_i]\ .$$

5. עבור המספרים קבועים a_i ומ"מ X_i,

$$E[\sum_{i=1}^{n} a_i X_i] = \sum_{i=1}^{n} a_i E[X_i]\ .$$

7.15 דוגמא. בחנות ירקות קטנה ישנם 3 תפוחים ו-4 אגסים. אלון קונה 2 פירות באקראי. בן המוכר מרוויח על כל תפוח 2 ש"ח ועל כל אגס 4 ש"ח. מצאו את תוחלת הרווח של בן.

פיתרון. נגדיר מ"מ Q עבור הרווח של בן, ומשתנה מקרי X כמספר התפוחים שאלון קנה. על כן,

$$Q = 2X + 4(2 - X) = 8 - 2X$$

$$P(X = 1) = \frac{4}{7}, \qquad P(X = 2) = \frac{3}{7}.\frac{2}{6} = \frac{1}{7}, \qquad P(X = 0) = \frac{4}{7}.\frac{3}{6} = \frac{2}{7}\ .$$

מכאן נובע ש

$$E[X] = 0.\frac{2}{7} + 1.\frac{4}{7} + 2.\frac{1}{7} = \frac{6}{7}\ .$$

נשתמש בתכונת הליניאריות לחישוב התוחלת של

$$E[Q] = 8 - 2E[X] = 8 - \frac{12}{7} = \frac{44}{7}\ .$$

■

דוגמא. בוחרים קוד באורך 10 תווים, כאשר כל תו נבחר באקראי מתוך הספרות 9 – 0 ללא קשר לתווים האחרים שנבחרו. נסמן ב- X מ"מ הסופרים את מספר הרצפים 11 בקוד. חשבו את $E[X]$.

פיתרון. נגדיר מ"מ $X_i,\ i=1,\ldots,9$ כאשר X_i מקבל את הערך 1 אם ורק אם במקום ה- i מתחיל רצף של 11. לכן

$$X=\sum_{i=1}^{9}X_i\ .$$

$$\begin{aligned}E[X]=&E\left[\sum_{i=1}^{9}X_i\right]\\=&\sum_{i=1}^{9}E\left[X_i\right]\\=&\sum_{i=1}^{9}\left[1.P(X_i=1)+0.P(X_i=0)\right]\\=&\sum_{i=1}^{9}P(X_i=1)\\=&\sum_{i=1}^{9}\frac{1}{10}.\frac{1}{10}\\=&\frac{9}{100}\ .\end{aligned}$$

■

7.16 הגדרה. (תוחלת פונקציה של משתנה מקרי)

יהי X משתנה מקרי בדיד בעל פונקצית התפלגות $P_X(x)$. התוחלת של המשתנה מקרי $g(X)$ הוא

$$E[g(X)]=\sum_{k\in X}g(x)f_X(k)\ .$$

7.17 דוגמא. ניקח משתנה מקרי X בעל התפלגות

$$P(X=k)=\begin{cases}\frac{1}{3} & k=-1,\\ \frac{1}{6} & k=0,\\ \frac{1}{2} & k=1.\end{cases}$$

חשבו את $E[X]$ ואת $E[X^4]$.

פיתרון. נשתמש בהגדרת התוחלת ונקבל

$$\begin{aligned}E[X]=&(-1).\frac{1}{3}+0.\frac{1}{6}+1.\frac{1}{2} &&=\frac{1}{6},\\E[X^4]=&(-1)^4.\frac{1}{3}+0^4.\frac{1}{6}+1^4.\frac{1}{2} &&=\frac{5}{6}.\end{aligned}$$

■

7.18 דוגמא. נניח ש X הוא המספר של רכבים אשר עוברים דרך השתיפת רכב בין השעות $16:00$ ו- $17:00$. ל- X יש את ההתפלגות

k	4	5	6	7	8	9
$P(X=k)$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{6}$

הפונקציה $g(X)=2X-1$ מצייג את הרווח ב \$ עבור X. מצאו את תוחלת הרווח בין השעות $16:00$ ו- $17:00$.

פיתרון.

$$\begin{aligned}E[g(X)] =&E[2X-1]\\ =&\sum_{k=4}^{9}(2x-1)f_X(k)\\ =&7.\frac{1}{12}+9.\frac{1}{12}+11.\frac{1}{4}+13.\frac{1}{4}+15.\frac{1}{6}+17.\frac{1}{6}\\ =&\$12.67\ .\end{aligned}$$

■

7.19 הגדרה. (שונות של משתנה מקרי) יהי X משתנה מקרי בדיד בעל פונקצית התפלגות $f_X(k)$ ותוחלת μ. השונות (variance) של X הוא

$$V(X)=E[(X-\mu)^2]=\sum_{k\in X}(k-\mu)^2 f_X(k)\ .$$

הגדרה. (סטיית התקן) סטיית התקן (standard deviation) של משתנה מקרי X מוגדרת להיות שורש השונות ומסומנת ב-

$$\sigma_X=\sqrt{V(X)}\ .$$

7.20 מסקנה. (קיצור דרך לשונות)

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

7.21 דוגמא. בר היא בנקאית השקעות מתוחכמת המציעה ללקוחותיה את ההגרלה הבאה: בסיכוי $\frac{2}{5}$ תקבלו רווח של 10. בסיכוי $\frac{1}{5}$ לא תרוויחו ולא תפסידו דבר, ובסיכוי $\frac{3}{5}$ תפסידו 5. חשבו את התוחלת וסטיית התקן של X המייצג את הרווח מהשקעה אצל בר.

פיתרון. תוחלת הרווח היא חיובית שכן

$$\begin{aligned}E[X] =&10.\frac{2}{5}+0.\frac{1}{5}-5.\frac{2}{5} &&=2,\\ (E[X])^2 =&100.\frac{2}{5}+0^2.\frac{1}{5}+25.\frac{2}{5} &&=50,\\ V(X) =&E[X^2]-(E[X])^2 &&=50-4=46,\\ \Rightarrow\quad \sigma_X =&\sqrt{46}\ .\end{aligned}$$

■

7.22 מסקנה. (תכונות השונות)

1. השונות של מספר קובע היא אפס: ניקח את המספר הקבוע b. אזי

$$V(b) = E[(b-\mu_b)^2] = E[(b-b)^2] = 0 \ .$$

2. שונות היא אי-שלילית.

3. שונות של טרנספורמציה לינארית של משתנה מקרי:
עבור X מ"מ וקבועים $a, b \in \mathbb{R}$ נקבל b,

$$\begin{aligned} V(aX+b) =& E[(aX+b-E(aX+b))^2] \\ =& E[(aX+b-aE(X)-b)^2] \\ =& E[(aX-aE(X))^2] \\ =& E[a^2(X-E(X))^2] \\ =& a^2E[(X-E(X))^2] \\ =& a^2V(X) \ . \end{aligned}$$

כאשר השתמשנו שוב ושוב בתוכנת הלינאריות. נקבל

$$V(aX+b) = a^2V(X) \ .$$

הערה: שונות משותפת

7.23 הגדרה. (שונות משותפת) השונות המשותפת (covariance) של צמד מ"מ X ו- Y היא

$$\text{Cov}(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] \ .$$

(כאשר $E[X] \equiv \mu_X$ ו $E[Y] \equiv \mu_Y$)
במילים, השונות המשותפת מוגדרת ע"י תוחלת המכפלה $(X-\mu_X)(Y-\mu_Y)$.

7.24 מסקנה. (קיצור דרך לשונות משותפת)

$$\begin{aligned} \text{Cov}(X,Y) =& E((X-E(X))(Y-E(Y))) \\ =& E(XY-XE(Y)-YE(X)+E(X)E(Y)) \\ =& E(XY)-E(XE(Y))-E(YE(X))+E(E(X)E(Y)) \\ =& E(XY)-2E(X)E(Y)+E(X)E(Y) \\ =& E(XY)-E(X)E(Y). \end{aligned}$$

8 משתנה מקריים חד מימידיים מיוחדים 25-7

סיכום נוסחאות: פונקצית התפלגות (מצטברת), תוחלת, שונות.

8.1 הגדרה. (משתנה מקרי בדיד) משתנה מקרי (חד מימדי) בדיד X (מ"מ) הוא פונקציה על המרחב מדגם המתאימה ערך מספרי יחיד לכל תוצאה במרחב הסתברות,

$$X : \Omega \to \mathbb{R} \ .$$

8.2 דוגמא. מטילים 2 קוביות. נסמן ב- X את סכום התוצאות שהתקבלו. כמו כן

$$\begin{aligned}
\omega &= \{(1,1)\}, & X(\omega) &= 2,\\
\omega &= \{(2,1),(1,2)\}, & X(\omega) &= 3,\\
&\vdots \\
\omega &= \{(6,6)\}, & X(\omega) &= 12.
\end{aligned}$$

זו דוגמה של מ"מ כאשר X הוא המשתנה המקרי.

8.3 הגדרה. (תומך של מ"מ בדיד) התומך של משתנה מקרי בדיד X זו קבוצת המספרים ש-X יכול להיות שווה להם בהסתברות חיובית, קרי

$$\text{Supp}(X) = \{k \in \mathbb{R} : P(X=k) > 0\} \ .$$

8.4 הגדרה. (פונקצית הסתברות / פונקצית התפלגות) הפונקצית ההתפלגות של משתנה מקרי בדיד (מ"מ) X מוגדרת להיות הפונקציה $f_X(k)$ המקבלת המ"מ X ומחזירה את ההסתברות כי למ"מ X יש ערך k:

$$f_X(k) = P(X=k) \ ,$$

עם התכונות

1. $\sum_{k \in X} f_X(k) = 1$.
2. $f_X(k) \geq 0 \quad \forall k$.

8.5 דוגמא. אלון הוא סוחר ממולח בשוק ההון. הוא משקיע סכום של $\$1000$ במכשיר פיננסי אשר בכל שנה מכפיל את הכסף בסיכוי $\frac{2}{3}$ או מפסיד חצי מההשקעה בסיכוי $\frac{1}{3}$. השנים הן בלתי תלויות. נגדיר את X להיות שווי ההשקעה של אלון לאחר שנתיים. התומך של X כעת הוא

$$\text{supp}(X) = \{250, 1000, 4000\} \ ,$$

על כן, ההתפלגות של X היא

$$P_X(1000) = \frac{4}{9}, \quad P_X(4000) = \frac{4}{9}, \quad P_X(250) = \frac{1}{9}$$

ו - 0 אחרת.

8.6 הגדרה. (פונקצית התפלגות מצטברת) פונקצית התפלגות מצטברת מסומנת ע"י $F_X(x)$ של מ"מ בדיד X בעל פונקצית התפלגות $f_X(k)$ מוגדרת להיות

$$F_X(x) = \sum_{\substack{k \in X \\ k \leq x}} f_X(k) \ .$$

במילים אחרות, פונקצית התפלגות מצטברת $F_X(x)$ מוגדרת להיות הפונקציה שתחומה הוא הישר הממשי וטווחה הקטע $[0,1]$ ומקיימת

$$F_X(x) = P(X \leq x) \qquad \forall x \in X \ .$$

8.7 הגדרה. (תוחלת של משתנה נקרי בדיד) יהי X משתנה מקרי בדיד בעל פונקצית התפלגות $f_X(k)$. התוחלת (expectation) של X היא

$$\mu_X \equiv E[X] = \sum_{k\in X} k f_X(k) \ .$$

8.8 דוגמא. התוחלת של הטלת קוביה היא

$$E[X] = \sum_{k=1}^{6} k\frac{1}{6} = \frac{7}{2} \ .$$

8.9 מסקנה. (לינאריות התוחלת)

1. עבור כל קבוע $a \in \mathbb{R}$ התוחלת של a היא

$$E[a] = a \ .$$

2. עבור כל משתנה מקרי X וקבוע $a \in \mathbb{R}$ התוחלת של aX היא

$$E[aX] = aE[X] \ .$$

3. עבור כל צמד משתנים מקריים X ו- Y התוחלת של $X+Y$ היא

$$E[X+Y] = E[X] + E[Y] \ .$$

4. כמובן שהתכונה הנ"ל ניתנת להכללה גם עבור מספר כללי של משתנים , שכן

$$E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E[X_i] \ .$$

5. עבור המספרים קבועים a_i ומ"מ X_i,

$$E[\sum_{i=1}^{n} a_i X_i] = \sum_{i=1}^{n} a_i E[X_i] \ .$$

דוגמא. בוחרים קוד באורך 10 תווים, כאשר כל תו נבחר באקראי מתוך הספרות 9 – 0 ללא קשר לתווים האחרים שנבחרו. נסמן ב- X מ"מ הסופרים את מספר הרצפים 11 בקוד. חשבו את $E[X]$.

פיתרון. נגדיר מ"מ $X_i,\ i = 1,\dots,9$ כאשר X_i מקבל את הערך 1 אם ורק אם במקום ה- i מתחיל רצף של 11. לכן

$$X = \sum_{i=1}^{9} X_i \ .$$

$$
\begin{aligned}
E[X] =& E\left[\sum_{i=1}^{9} X_i\right] \\
=& \sum_{i=1}^{9} E\left[X_i\right] \\
=& \sum_{i=1}^{9} \left[1.P(X_i = 1) + 0.P(X_i = 0)\right] \\
=& \sum_{i=1}^{9} P(X_i = 1) \\
=& \sum_{i=1}^{9} \frac{1}{10}.\frac{1}{10} \\
=& \frac{9}{100} \ .
\end{aligned}
$$

■

8.10 הגדרה. (שונות של משתנה מקרי) יהי X משתנה מקרי בדיד בעל פונקצית התפלגות $f_X(k)$ ותוחלת μ. השונות (variance) של X הוא

$$V(X) = E[(X-\mu)^2] = \sum_{k\in X} (k-\mu)^2 f_X(k) \ .$$

הגדרה. (סטיית התקן) סטיית התקן (standard deviation) של משתנה מקרי X מוגדרת להיות שורש השונות ומסומנת ב-

$$\sigma_X = \sqrt{V(X)} \ .$$

8.11 מסקנה. (קיצור דרך לשונות)

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

8.12 דוגמא. בר היא בנקאית השקעות מתוחכמת המציעה ללקוחותיה את ההגרלה הבאה: בסיכוי $\frac{2}{5}$ תקבלו רווח של 10. בסיכוי $\frac{1}{5}$ לא תרוויחו ולא תפסידו דבר, ובסיכוי $\frac{3}{5}$ תפסידו 5. חשבו את התוחלת וסטיית התקן של X המייצג את הרווח מהשקעה אצל בר.

פיתרון. תוחלת הרווח היא חיובית שכן

$$
\begin{aligned}
\mu_X =& 10.\frac{2}{5} + 0.\frac{1}{5} - 5.\frac{2}{5} && = 2, \\
\mu_{X^2} =& 100.\frac{2}{5} + 0^2.\frac{1}{5} + 25.\frac{2}{5} && = 50, \\
V(X) =& \mu_{X^2} - \mu_X^2 && = 50 - 4 = 46, \\
\Rightarrow \quad \sigma_X =& \sqrt{46} \ .
\end{aligned}
$$

■

8.13 מסקנה. (תכונות השונות)

1. השונות של מספר קובע היא אפס: ניקח את המספר הקבוע b. אזי

$$V(b) = E[(b-\mu_b)^2] = E[(b-b)^2] = 0 \ .$$

2. שונות היא אי-שלילית.
3. שונות של טרנספורמציה לינארית של משתנה מקרי:
עבור X מ"מ וקבועים $a, b \in \mathbb{R}$ נקבל,

$$\begin{aligned} V(aX+b) =& E[(aX+b-E(aX+b))^2] \\ =& E[(aX+b-aE(X)-b)^2] \\ =& E[(aX-aE(X))^2] \\ =& E[a^2(X-E(X))^2] \\ =& a^2E[(X-E(X))^2] \\ =& a^2V(X) \ . \end{aligned}$$

כאשר השתמשנו שוב ושוב בתוכנת הלינאריות. נקבל

$$V(aX+b) = a^2V(X) \ .$$

משתנה ברנולי

ניסוי ברנולי הוא ניסוי עם תוצאה בינארית - כן או לא. 0 או 1 .הצלחה או כישלון. ניתן לחשוב על ניסוי ברנולי בתור הטלת מטבע, לא בהכרח הוגן. המטבע נוחת על עץ או פלי, כאשר אחד מהם מגדיר הצלחה = 1 ,והאחר מגדיר כישלון = 0 .הסיכוי להצלחה נתון על ידי הפרמטר

$$0 \le p \le 1 \ .$$

8.14 הגדרה. (משתנה ברנולי) משתנה ברנולי הוא משתנה המקבל את הערכים 0 ו-1 בהתבסס על ניסוי ברנולי עם סיכוי p להצלחה. המשתנה מסומן ב- $X \sim \mathrm{Ber}(p)$ (במילים, X מתפלג לפי התפלגות ברנולי עם פרמטר p). פורמאלית,

$$P(X=k) = \begin{cases} p, & k=1, \\ 1-p, & k=0 \ . \end{cases}$$

8.15 מסקנה. (תוחלת של משתנה ברנולי) התוחלת של משתנה ברנולי היא

$$E[X] = 1.p + 0.(1-p) = p \ ,$$

ו- נובע כי

$$E[X^2] = 1^2.p + 0^2.(1-p) = p \ ,$$

וכן הלאה,

$$E[X^n] = 1^n.p + 0^n.(1-p) = p \ ,$$

לכל n טבעי.

8.16 מסקנה. (שונות של משתנה ברנולי) השונות של משתנה ברנולי היא (עיין מסקנה (8.11) לעייל)

$$V(X) = E[X^2] - (E[X])^2 = p - p^2 = p(1-p) \ .$$

התפלגות הבינומית

התפלגות הבינומית מכלילה את התפלגות ברנולי למספר ניסויי ברנולי בלתי תלויים. נדמיין ניסוי ברנולי, כגון הטלת מטבע, אשר מבוצע שוב ושוב ובסה"כ n פעמים. השאלה הטבעית שעולה היא מה יהיו מספר ההצלחות בכלל הניסויים. לדוגמא, אם בוחנים השפעות של תרופה על n נבדקים, נרצה לדעת בכמה מקרים התרופה אכן הייתה אפקטיבית. המשתנה אשר מציין את מספר ההצלחות ב-n ניסויי ברנולי ב"ת עם פרמטר p הוא משתנה בינומי, בעל התפלגות בינומית. נסמן זאת ב-

$$X \sim \text{Bin}(n,p) \ .$$

התומך של משתנה מקרי בינומי הוא

$$\text{supp}(X) = \{0,1,\ldots,n\}$$

יכולות להתקבל 0 הצלחות, הצלחה אחת וכן הלאה, עד למקסימום של n הצלחות. ההתפלגות עצמה נתונה על ידי הנוסחה

$$P(X=k) = \binom{n}{k} p^k (1-p)^k \ .$$

עבור כל k בתומך. נסביר את הנוסחה. המאורע $X=k$ מציין k הצלחות. הניסויים עצמם בלתי תלויים ולכן הסיכוי שיהיה בדיוק k הצלחות ספציפיות ב n ניסויים הוא $p^k(1-p)^{n-k}$ (ההסתברות לקבל k הצלחות ו $n-k$ כשלונות מסויימים). בנוסף, צריך לקחת בחשבון את כלל הקומבינציות לקבלת אותן ההצלחות - ניתן לבחור באילו ניסויים, מתוך סה"כ הניסויים, יהיו הצלחות ומספר האפשרויות הללו הוא בחירה של k מתוך n, $\binom{n}{k}$ כידוע.

8.17 הגדרה. (משתנה מקרי בינומי) מבצעים n ניסויים בלתי תלויים. לכל ניסוי יש 2 תוצאות אפשרויות: הצלחה עם הסתברות p או כישלון עם הסתברות $q \equiv 1-p$. משתנה מקרי בינומי X סופר את מספר הצלחות ב n הניסויים.

חוק. (התפלגות בינומית) משתנה בינומי הוא משתנה המקבל את הערכים $0,1,\ldots,n$ עד ערך מקסימלי n בהתבסס על n ניסויים שבוצעו. ההתפלגות מוגדרת על ידי הסיכוי p^k עבור k הצלחות, יחד עם הסיכוי $(1-p)^{n-k}$ עבור $1-p$ כישלונות, ואלו יחד עם המספר הדרכים לבחור k מתוך n, אשר בדיוק שווה ל $\binom{n}{k}$, שכזה נובע להתפלגות בינימית:

$$f_X(k) \equiv P(X=k) = \binom{n}{k} p^k (1-p)^{n-k} \ .$$

8.18 דוגמא. הסיכוי אשר חולה מחלים ממחלת דם חריג הוא 0.4. אם 15 אנשים נדבקים עם המחלה זו, מהי ההסתברות אשר

1. לפחות 10 יחלימו
2. בין 3 עד 8 יחלימו
3. בדיוק 5 יחלימו?

פיתרון. נגדיר X להיות מספר החולים אשר יחלימו.

.1

$$\begin{aligned}
P(X \geq 10) =& 1 - P(X<10) \\
=& 1 - \sum_{k=0}^{9} f_{X\sim\text{Bin}(15,0.4)}(k) \\
=& 1 - \sum_{k=0}^{9} \binom{15}{k} 0.4^k (1-0.4)^{15-k} \\
=& 1 - 0.9662 = 0.0338 \ .
\end{aligned}$$

.2

$$\begin{aligned}
P(3 \leq X \leq 8) =& 1 - P(X < 10) \\
=& \sum_{k=3}^{8} f_{X\sim\text{Bin}(15,0.4)}(k) \\
=& \sum_{k=0}^{8} f_{X\sim\text{Bin}(15,0.4)}(k) - \sum_{k=0}^{2} f_{X\sim\text{Bin}(15,0.4)}(k) \\
=& \left(\sum_{k=0}^{8} - \sum_{k=0}^{2}\right)\binom{15}{k} 0.4^k (1-0.4)^{15-k} \\
=& 0.9050 - 0.0271 = 0.8779 \ .
\end{aligned}$$

.3

$$\begin{aligned}
P(X = 5) =& f_{X\sim\text{Bin}(15,0.4)}(k=5) \\
=& \binom{15}{5} 0.4^5 (1-0.4)^{10} \\
=& 0.1859 \ .
\end{aligned}$$

■

8.19 דוגמא. מחשב משתמש במילים באורך 8 ביטים. כאשר משדרים ביטים בקו התקשורת יש סיכוי של $.(1$ לשגיאה בביט בודד שמשודר. בכדי לשפר את איכות הקו, משודר כל ביט 3 פעמים ברצף והפענוח של כל שלשת ביטים מתבצע על ידי החלטת הרוב. למשל: אם רוצים לשדר 0, משדרים בקו "000" והמקלט מפענח את התשדורת כ-0 אם השלשה שהגיעה אליו מורכבת מלפחות שני אפסים.

1. חשבו את ההסתברות שמילת מחשב תפוענח נכון.
2. מצאו את התפלגות מספר המילים שפוענחו באופן שגוי בקובץ המכיל 1000 מילים.

פתרון: נשים לב שכמות הטעויות בשליחה של ביט בודד מתפלגת $X \sim \text{Bin}(3, 0.1)$ ולכן ההסתברות שתהיה שגיאה בפענוח ביט בודד היא

$$P(X \geq 2) = \binom{3}{2} 0.1^2 0.9 + \binom{3}{1} 0.1^3 = 0.028$$

1. כמות השגיאות במילה באורך 8 ביטים מתפלגת

$$Y \sim \text{Bin}(8, 0.028)$$

ולכן המילה תפוענח נכון בהסתברות

$$P(Y = 0) = \binom{8}{0} p^0 (1-p)^8 = (1 - 0.028)^8 = 0.796 \ .$$

2. לאור הסעיף הקודם, ההסתברות לשגיאה בפענוח מילה בודדת היא

$$1 - 0.796 = 0.203 \ ,$$

לכן מספר המילים שפוענחו באופן שגוי בקובץ המכיל 1000 מילים מתפלג $\text{Bin}(1000, 0.203)$.

8.20 מסקנה. (תוחלת של התפלגות בינומית) התוחלת של משתנה מקרי X בעל התפלגות בינומית, או במילים אחרות התוחלת של מספר ההצלחות ב-n ניסויי ברנולי בלתי תלויים היא

$$E[X] = np$$

הוכחה. הצבה ישירה להגדרה של תוחלת תניב

$$E[X] = \sum_{k=0}^{n} k f_X(k) = \sum_{k=0}^{n} k \binom{n}{k} p^k q^{n-k}$$

כאשר $q \equiv 1 - p$. יש לכתוב kp^k כ $p\frac{d}{dp}p^k$ וכמו כן

$$E[X] = \sum_{k=0}^{n} k f_X(k) = \sum_{k=0}^{n} \binom{n}{k} p\frac{d}{dp}p^k q^{n-k} = p\frac{d}{dp}\sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} \ .$$

קיים זיהוי אשר מציין כי $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$. לכן מקבלים

$$E[X] = p\frac{d}{dp}(p+q)^n = pn(p+q)^{n-1} \ .$$

אבל $p + q = p + 1 - p = 1$, על כן

$$E[X] = np \ .$$

■

8.21 מסקנה. (שונות של התפלגות בינומית) השונות משתנה מקרי X בעל התפלגות בינומית, או במילים אחרות השונות של מספר ההצלחות ב-n ניסויי ברנולי בלתי תלויים היא

$$V[X] = np(1-p)$$

הוכחה. נחשב את

$$E[X^2] = \sum_{k=0}^{n} k^2 f_X(k) = \sum_{k=0}^{n} k^2 \binom{n}{k} p^k q^{n-k}$$

כאשר $q \equiv 1 - p$. יש לכתוב k^2p^k כ $p\frac{d}{dp}\left(p\frac{d}{dp}p^k\right)$ וכמו כן

$$E[X] = \sum_{k=0}^{n} k f_X(k) = \sum_{k=0}^{n} \binom{n}{k} p\frac{d}{dp}\left(p\frac{d}{dp}p^k q^{n-k}\right) = p\frac{d}{dp}\left(p\frac{d}{dp}\sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k}\right) \ .$$

קיים זיהוי אשר מציין כי $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$. לכן מקבלים

$$E[X] = p\frac{d}{dp}\left(p\frac{d}{dp}(p+q)^n\right) = p\frac{d}{dp}\left(pn(p+q)^{n-1}\right) = np(p+q)^{n-1} + n(n-1)p^2(p+q)^{n-2} \ .$$

אבל $p + q = p + 1 - p = 1$, על כן

$$E[X^2] = np + n(n-1)p^2 \ .$$

על כן

$$V(X) = E[X^2] - E[X]^2 = np + n(n-1)p^2 - (np)^2 = np - np^2 = np(1-p) \ .$$

■

התפלגות גיאומטרית

נניח כי אנו מטילים צמד קוביות שוב ושוב עד שמתקבלת לראשונה התוצאה $(6,6)$. מה מספר ההטלות שנדרש לבצע? ובכן, התשובה לשאלה הזו היא משתנה מקרי כלשהו. ייתכן וכבר בסיבוב הראשון נקבל את התוצאה הרצויה, ייתכן ותתקבל בסיבוב השני, וגם קיימת האפשרות כי נאלץ לחכות מאה סיבובים (ואף יותר מכך) עד שנקבל את התוצאה המבוקשת. מספר הסיבובים הנדרש אינו חסום ויוכל להיות גדול כרצוננו. המשתנה המקרי המתאר את מספר הסיבובים נקרא משתנה מקרי גיאומטרי, והתפלגותו נקראת התפלגות גיאומטרית. פורמאלית,

8.22 הגדרה. (משתנה גיאומטרי) מבצעים ניסויים בלתי תלויים עם הסתברות להצלחה p וכישלון $q \equiv 1-p$. משתנה מקרי גיאומטרי X סופר את מספר הניסויים עד להצלחה הראשונה (כולל).

8.23 חוק. (והתפלגותו של משתנה גיאומטרי) משתנה מקרי גיאומטרי הוא משתנה מקרי המציין את מספר ניסויי הברנולי הבלתי תלויים שמתקיימים עד לקבלת הצלחה ראשונה. התפלגות גיאומטרית נתונה על ידי פרמטר יחיד p והוא הסיכוי לקבלת הצלחה בניסוי ברנולי בודד. נסמן את המשתנה ב- $X \sim G(p)$ והתפלגותו היא

$$f_X(k) = \begin{cases} p(1-p)^{k-1} & k = 1, 2, \ldots \\ 0 \\ \text{else} \end{cases}$$

עבור כל

$$k \geq 1$$

טבעי.

ההסבר להתפלגות הוא מיידי. בכדי להגיע להצלחה הראשונה בניסוי ה-k, צריכים להתרחש $k-1$ כשלונות רצופים ומיד לאחר מכן הצלחה. זאת בדיוק הנוסחה ההסתברותית שכתבנו. התומך של ההתפלגות הוא כל המספרים הטבעים, כאשר הערך המינימאלי הוא 1, כי צריך להתקיים לפחות ניסוי אחד בכדי להגיע להצלחה אחת.

חישובי תוחלת ושונות עבור התפלגות גיאומטרית מתבססים על סכומים גיאומטריים, סכום סדרה הנדסית.

8.24 מסקנה. (תוחלת של התפלגות גיאומטרית)

$$E[X] \equiv \sum_{k=1}^{\infty} k f_X(k) = \frac{1}{p} .$$

הוכחה. הצבה של הביטוי המפורש להפונקצית הסתברות של מ"מ גיאומטרי, קרי $f_X(k) = (1-p)^{k-1}p$ תניב

$$E[X] \equiv \sum_{k=1}^{\infty} k f_X(k) = \sum_{k=1}^{\infty} kp(1-p)^{k-1} = \sum_{k=0}^{\infty} kp(1-p)^{k-1} .$$

כאשר השוויון האחרון נובע בשל העובדה כי האיבר $k=0$ שווה אפס. במקום $k(1-p)^{k-1}$ אפשר להחליף עם $-\frac{d}{dp}(1-p)^k$ כך

$$E[X] = -\sum_{k=0}^{\infty} p\frac{d}{dp}(1-p)^k .$$

ניתן להעביר את כל האיברים האינם תלויים בk לחוץ הסכום:

$$E[X] = -p\frac{d}{dp}\sum_{k=0}^{\infty}(1-p)^k .$$

ישר מן הזיהוי $\frac{1}{1-x} = \sum_{k=0}^{\infty} x^k$ מסיקים כי

$$E[X] = -p\frac{d}{dp}\left(\frac{1}{1-[1-p]}\right) = -p\frac{d}{dp}\left(\frac{1}{p}\right) = -p\left(-\frac{1}{p^2}\right) = \frac{1}{p} .$$

■

8.25 מסקנה. **(שונות של התפלגות גיאומטרית)**

$$V(X) = E[X^2] - (E[X])^2 = \frac{1-p}{p^2} \ .$$

הוכחה. קודם כל יש צורך לגזור את $E[X^2]$. הצבה של הביטוי המפורש להפונקצית הסתברות של מ"מ גיאומטרי, קרי $f_X(k) = (1-p)^{k-1}p$ תניב

$$E[X^2] \equiv \sum_{k=1}^{\infty} k^2 f_X(k) = \sum_{k=1}^{\infty} k^2 p(1-p)^{k-1} = \sum_{k=0}^{\infty} k^2 p(1-p)^{k-1} \ .$$

כאשר השוויון האחרון נובע בשל העובדה כי האיבר $k=0$ שווה אפס. במקום $k^2(1-p)^{k-1}$ אפשר להחליף אותו עם $\frac{d}{dp}\left((1-p)\frac{d}{dp}(1-p)^k\right)$ כך

$$E[X^2] = \sum_{k=0}^{\infty} p\frac{d}{dp}\left((1-p)\frac{d}{dp}(1-p)^k\right) \ .$$

ניתן להעביר את כל האיברים האינם תלויים בk לחוץ הסכום:

$$E[X^2] = p\frac{d}{dp}\left((1-p)\frac{d}{dp}\sum_{k=0}^{\infty}(1-p)^k\right) \ .$$

ישר מן הזיהוי $\frac{1}{1-x} = \sum_{k=0}^{\infty} x^k$ מסיקים כי

$$\begin{aligned}
E[X^2] =& p\frac{d}{dp}\left((1-p)\frac{d}{dp}\left(\frac{1}{1-[1-p]}\right)\right) \\
=& p\frac{d}{dp}\left((1-p)\frac{d}{dp}\left(\frac{1}{p}\right)\right) \\
=& p\frac{d}{dp}\left(\frac{-(1-p)}{p^2}\right) \\
=& p\left(\frac{-1}{p^2} + \frac{2}{p^3}\right) \\
=& \frac{1}{p}\left(-1 + \frac{2}{p}\right) \\
=& \frac{1}{p}\left(\frac{2-p}{p}\right) \ .
\end{aligned}$$

לכן

$$V(X) = E[X^2] - E[X]^2 = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2} \ .$$

■

8.26 דוגמא. בכד 9 כדורים לבנים ו- 1 שחור. מושכים באקראי בזה אחר זה כדור מהכד. נסמן ב- X את מספר המשיכות עד לקבלת שחור וב- Y את מספר המשיכות עד לקבלת 3 שחורים.

מצאו את התפלגות X כאשר המשיכה היא עם החזרה.

פיתרון. X מקבל ערכים $\mathbb{N}/\{0\}$. בכל הוצאה ישנה הסתברות של $\frac{9}{10}$ לכדור לבן (כישלון) והסתברות של $p = \frac{1}{10}$ לכדור לבן (הצלחה). כלומר, קיבלנו מ"מ גאומטרי

$$X \sim G\left(\frac{1}{10}\right) \ .$$

■

התפלגות פואסונית

דוגמה של תהליך פאוסוני היא מספר שיחות טלפון נענו במוקד שרות של חברת סלולר. יהי

$$t = \text{מספר הלקוחות הממוצע שיתקשרו לשעה}$$

ו-

$$n = \text{המספר הלקוחות הכללי}$$

שכזה

$$p \equiv \frac{t}{n} = \text{ההסתברות ש } t \text{ לקוחות יתקשרו בשעה אחת}$$

לכן

$$np = \text{מספר הלקוחות הממוצע שיתקשרו בשעה כלשהי}$$

מניחים כי בכל שעה כל לקוח מתקשר בהסתברות קטנה, בלי תלות ביתר הלקוחות. מה המספר הלקוחות הממוצע שיתקשרו בשעתיים? מתכונות הלינאריות אנחנו יודעים ש

$$2np = \text{מספר הלקוחות הממוצע שיתקשרו בשעתיים}$$

וכן

$$knp = \text{מספר הלקוחות הממוצע שיתקשרו ב } k \text{ שעות הבאות}$$

באופן דומה המספר הלקוחות הממוצע אשר יתקשרו חצי שעה כלשה יִ? נובע מתכונת הלינאריות:

$$\frac{1}{2}np = \text{מספר הלקוחות הממוצע שיתקשרו בחצי שעה כלשהי}$$

מניחים כי הפרמטר n מייצג את מספר הלקוחות והוא יחסית גדול, בעוד הפרמטר p מייצג את הסיכוי של כל לקוח להתקשר והוא יחסית קטן. זאת אומרת ש-np הוא (בקירוב) גודל קבוע כלשהו:

$$\lambda \equiv np$$

כאשר λ הוא מספר קבוע. התכונות הבאות מאפיינות תהליך פואסון.

- מספר האירועים הממוצע ליחידת זמן כלשהי (או ליחידת שטח כלשהי) הוא ערך קבוע $\lambda > 0$ זהו הפרמטר המרכזי בהתפלגות פואסון.
- מספר האירועים בקטעי זמן (או שטח) זרים הם בלתי תלויים זה בזה. במילים אחרות, מספר השיחות שנקבל בשעה הראשונה הוא בלתי תלוי במספר השיחות שנקבל בשעה השנייה.
- הסיכוי לאירוע בקטע זמן (שטח) מסויים תלוי אך ורק באורך קטע הזמן (השטח). במילים אחרות, הסיכוי שתתקבל שיחה בשעה הראשונה שווה לסיכוי שתתקבל שיחה בשעה השנייה ושניהם שווים לסיכוי שתתקבל שיחה בשעה השלישית וכן הלאה. באופן דומה, הסיכוי לקבל שיחה בחצי השעה הראשונה שווה לסיכוי לקבל שיחה בחצי השעה השנייה וכו'. אורך קטע הזמן הוא המשפיע על ההסתברות, ולא המיקום של הקטע על ציר הזמן.

8.27 הגדרה. (משתנה פואסוני) משתנה פואסוני זה סופר את מספר האירועים שהתרחשו ביחידת זמן, או שטחף וכדומה. λ הוא הקצב הממוצע ליחידת זמן

8.28 חוק. (התפלגות פואסוני) נניח ש X הוא משתנה מקרי של תהליך פואסון. נסמן את זה ב $X \sim P(\lambda)$, כאשר λ הוא המספר האירועים ליחידה זמן, או יחידת שטח, וכדומה. התומך של X הוא

$$\operatorname{supp}(X) = \{0, 1, \dots, \}$$

כלומר כל מספר שלם $k \geq 0$. מספר האירועים אינו שלילי, ויכול להתקבל כל מספר אירועים, החל מ-0 ומעלה. ההתפלגות עצמה נתונה על ידי הנוסחה

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

עבור כל ערך בתומך.

הוכחה. כיצד הגענו לנוסחה הזאת? התשובה נעוצה בסיפור ממנו נפתח הדיון. לוקחים התפלגות בינומית ומניחים כי $\lambda = np$. כעת לוקחים את הגבול $n \to \infty$ מחשבים את הערך

$$\binom{n}{k} p^k (1-p)^{n-k} \ .$$

זאת ההסתברות של משתנה בינומי לקבל את הערך k. הגבול המתקבל תחת ההנחה ש $\lambda = np$ הוא:

$$\begin{aligned}\lim_{n\to\infty} \binom{n}{k} p^k (1-p)^{n-k} &= \lim_{n\to\infty} \frac{n!}{(n-k)!k!} \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k} = \lim_{n\to\infty} \left(1-\frac{\lambda}{n}\right)^n \frac{\lambda^k}{k!} \\ &= \frac{\lambda^k}{k!} \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^n\end{aligned}$$

אבל $\lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n$ מוגדר להיות e^x, לכן מקבלים

$$\lim_{n\to\infty} \binom{n}{k} p^k (1-p)^{n-k} = e^{-\lambda}\frac{\lambda^k}{k!} \ ,$$

וזה בדיוק הביטוי שכתבנו. ■

מסקנה. (תוחלת של התפלגות פואסון)

$$E[X] = \lambda$$

הוכחה.

$$E[X] = \sum_{k=0}^{\infty} kP(X=k) = \sum_{k=0}^{\infty} ke^{-\lambda}\frac{\lambda^k}{k!} = \sum_{k=1}^{\infty} ke^{-\lambda}\frac{\lambda^k}{k!} = \lambda \sum_{k=1}^{\infty} e^{-\lambda}\frac{\lambda^{k-1}}{(k-1)!} = e^{-\lambda}\lambda \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}$$

אבל $\sum_{k=0}^{\infty}\frac{x^k}{k!}$ מוגדר להיות e^x לכן

$$E[X] = \lambda e^{-\lambda} e^{\lambda} = \lambda \ .$$

■

8.29 מסקנה. (שונות של התפלגות פואסון)

$$V(X) = \lambda \ .$$

הוכחה.

$$\begin{aligned}E[X^2] &= \sum_{k=0}^{\infty} k^2 P(X=k) & &= \sum_{k=0}^{\infty} k^2 e^{-\lambda}\frac{\lambda^k}{k!} = \sum_{k=1}^{\infty} k^2 e^{-\lambda}\frac{\lambda^k}{k!} \\ &= \lambda \sum_{k=1}^{\infty} e^{-\lambda}\frac{k\lambda^{k-1}}{(k-1)!} & &= \lambda \sum_{k=1}^{\infty} e^{-\lambda}\frac{(k-1)\lambda^{k-1}}{(k-1)!} + \lambda \sum_{k=1}^{\infty} e^{-\lambda}\frac{\lambda^{k-1}}{(k-1)!} \\ &= \lambda \sum_{k=2}^{\infty} e^{-\lambda}\frac{(k-1)\lambda^{k-1}}{(k-1)!} + \lambda \sum_{k=1}^{\infty} e^{-\lambda}\frac{\lambda^{k-1}}{(k-1)!} & &= \lambda^2 \sum_{k=2}^{\infty} e^{-\lambda}\frac{\lambda^{k-2}}{(k-2)!} + \lambda \sum_{k=1}^{\infty} e^{-\lambda}\frac{\lambda^{k-1}}{(k-1)!} \\ &= \lambda^2 \sum_{k=0}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!} + \lambda \sum_{k=0}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!}\end{aligned}$$

$\sum_{k=0}^{\infty}\frac{x^k}{k!}$ מוגדר להיות e^x לכן

$$\begin{aligned}E[X^2] &= \lambda^2 + \lambda & &= \lambda + \lambda^2 \ . \\ \Rightarrow \quad V(X) &= E[X^2] - (E[X])^2 & &= \lambda + \lambda^2 - \lambda^2 = \lambda \ .\end{aligned}$$

■

8.30 דוגמא. חלקיקים נפלטים מחומר רדיואקטיבי בתהליך פואסון עם קצב של 0.5 חלקיקים לשנייה. חשבו את ההסתברויות הבאות:

1. ייפלט לפחות חלקיק אחד בשנייה נתונה.
2. ייפלטו יותר מ-3 חלקיקים ב-5 שניות.

פיתרון. התהליך בשאלה הוא תהליך פואסון בעל פרמטר $\lambda = 0.5$ לכן

.1

$$P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-0.5}\frac{0.5^0}{0!} = 0.393 \ .$$

2. כעת בשאלה הוא תהליך פואסון שונה בעל פרמטר 5λ:

$$\begin{aligned} P(X_5 > 3) =& 1 - P(X_5 = 0) - P(X_5 = 1) - P(X_5 = 2) - P(X_5 = 3) \\ =& 1 - e^{-5\lambda}\frac{(5\lambda)^0}{0!} - e^{-5\lambda}\frac{(5\lambda)^1}{1!} - e^{-5\lambda}\frac{(5\lambda)^2}{2!} - e^{-5\lambda}\frac{(5\lambda)^3}{3!} \\ =& 1 - e^{-5(0.5)}\frac{(5(0.5))^0}{0!} - e^{-5(0.5)}\frac{(5(0.5))^1}{1!} - e^{-5(0.5)}\frac{(5(0.5))^2}{2!} - e^{-5(0.5)}\frac{(5(0.5))^3}{3!} \\ =& 0.242 \end{aligned}$$

מאחר ו- X_5 מתפלג $P(0.5.5)$

■

תרגילים

8.31 דוגמא. (פונקצית התפלגות) מטילים 2 קוביות. נסמן ב- X את סכום התוצאות שהתקבלו. מצאו את ההתפלגות של X.

פיתרון. נתחיל מהתומך. התוצאות האפשרויות בסכום של 2 קוביות הן המספרים הטבעיים מ- 2 עד 12. בכדי לחשב את ההסתברות לכל ערך נשתמש במרחב מדגם סימטרי: גודל מרחב המדגם הוא $36 = 6^2$, ככמות האפשרויות לבחור תוצאה לקוביה הראשונה מתוך 6 ותוצאה לקוביה השניה מתוך 6. נחשב את ההסתברות של כל תוצאה בנפרד:

$$\begin{aligned} P(X = 2) =& \frac{|\{(1,1)\}|}{6^2} &= \frac{1}{36}, \\ P(X = 3) =& \frac{|\{(2,1),(1,2)\}|}{6^2} &= \frac{2}{36}, \\ P(X = 4) =& \frac{|\{(1,3),(2,2),(3,1)\}|}{6^2} &= \frac{3}{36}, \\ & \vdots \\ P(X = 7) =& \frac{|\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}|}{6^2} &= \frac{6}{36}, \\ P(X = 8) =& \frac{|\{(2,6),(3,5),(4,4),(5,3),(6,2)\}|}{6^2} &= \frac{5}{36}, \\ P(X = 9) =& \frac{|\{(3,6),(4,5),(5,4),(6,3)\}|}{6^2} &= \frac{4}{36}, \\ & \vdots \\ P(X = 12) =& \frac{|\{(6,6)\}|}{6^2} &= \frac{1}{36}, \end{aligned}$$

ובאופן דומה ממשיכים לחשב את האפשרויות לכל הערכים האפשריים בתומך. באופן כללי קל לראות שהנוסחא, כאשר p זו הסכום של ה 2 הקוביות, היא (עיין משוואה (6.9) לעייל)

$$P(X=p)=\frac{1}{36}\times\begin{cases}(p-1) & p<8\\ \\ (p-1)-2(p-7) & 8\le p<14\end{cases}$$

■

8.32 דוגמא. (פונקצית התפלגות מצטברת) אלון הוא סוחר ממולח בשוק ההון. הוא משקיע סכום של $\$1000$ במכשיר פיננסי אשר בכל שנה מכפיל את הכסף בסיכוי $\frac{2}{3}$ או מפסיד חצי מההשקעה בסיכוי $\frac{1}{3}$. השנים הן בלתי תלויות. נגדיר את X להיות שווי ההשקעה של אלון לאחר שנתיים. מצאו את פונקצית ההתפלגות המצטברת של X וציירו אותה.

פיתרון. כפי שציינו בדוגמה 9.2, התומך של X כעת הוא

$$\text{supp}(X)=\{250,1000,4000\}\ .$$

על כן, פונקצית ההתפלגות המצטברת של X היא

$$F_X(k)=P(X\le k)=\begin{cases}0 & k<250,\\ \\ \frac{1}{9} & 250\le k<1000,\\ \\ \frac{5}{9} & 1000\le k<4000,\\ \\ 1 & 4000\le k.\end{cases}$$

■

8.33 דוגמא. (תוחלת) בן הוא סוחר מתוחכם ולכן הוא מציע לאלון להפקיד אצלו 100 שקלים בתחילת השנה. בסיכוי של $\frac{2}{5}$ הסכום יוכפל, ובסיכוי $\frac{2}{5}$ הסכום יצטמצם בחצי בסוף השנה. מה תהיה תוחלת הכסף של אלון לאחר קבלת ההצעה? נסמן ב- X את שווי כסף של אלון בסוף השנה.

$$E[X]=200.\frac{2}{5}+50.\frac{3}{5}=110.$$

זאת אומרת שהצעתו של בן מגלמת תוחלת רווח של 10%!

8.34 דוגמא. (ליניאריות של תוחלת) בחנות ירקות קטנה ישנם 3 תפוחים ו-4 אגסים. אלון קונה 2 פירות באקראי. בן המוכר מרוויח על כל תפוח 2 ש"ח ועל כל אגס 4 ש"ח. מצאו את תוחלת הרווח של בן.

פיתרון. נגדיר מ"מ Q עבור הרווח של בן, ומשתנה מקרי X כמספר התפוחים שאלון קנה. על כן,

$$Q=2X+4(2-X)=8-2X$$

$$P(X=1)=\frac{4}{7},\qquad P(X=2)=\frac{3}{7}.\frac{2}{6}=\frac{1}{7},\qquad P(X=0)=\frac{4}{7}.\frac{3}{6}=\frac{2}{7}\ .$$

מכאן נובע ש

$$E[X]=0.\frac{2}{7}+1.\frac{4}{7}+2.\frac{1}{7}=\frac{6}{7}\ .$$

נשתמש בתכונת הליניאריות לחישוב התוחלת של

$$E[Q]=8-2E[X]=8-\frac{12}{7}=\frac{44}{7}\ .$$

■

8.35 דוגמא. (ליניאריות של תוחלת) בוחרים קוד באורך 10 תווים, כאשר כל תו נבחר באקראי מתוך הספרות 9 – 0 ללא קשר לתווים האחרים שנבחרו. נסמן ב- X מ"מ הסופרים את מספר הרצפים 11 בקוד. חשבו את $E[X]$.

פיתרון. נגדיר מ"מ 9 $X_i,\ i=1,\dots,9$ כאשר X_i מקבל את הערך 1 אם ורק אם במקום ה- i מתחיל רצף של 11. לכן

$$X=\sum_{i=1}^{9} X_i\ .$$

$$\begin{aligned}E[X]=&E\left[\sum_{i=1}^{9} X_i\right]\\ =&\sum_{i=1}^{9} E\left[X_i\right]\\ =&\sum_{i=1}^{9}\left[1.P(X_i=1)+0.P(X_i=0)\right]\\ =&\sum_{i=1}^{9} P(X_i=1)\\ =&\sum_{i=1}^{9}\frac{1}{10}.\frac{1}{10}\\ =&\frac{9}{100}\ .\end{aligned}$$

■

*העשרה: שונות משותפת

8.36 הגדרה. (שונות משותפת) השונות המשותפת (covariance) של צמד מ"מ X ו- Y היא

$$\mathrm{Cov}(X,Y)=E[(X-\mu_X)(Y-\mu_Y)]\ .$$

(כאשר $\mu_X\equiv E[X]$ ו $\mu_Y\equiv E[Y]$)
במילים, השונות המשותפת מוגדרת ע"י תוחלת המכפלה $(X-\mu_X)(Y-\mu_Y)$.

8.37 מסקנה. (קיצור דרך לשונות משותפת)

$$\begin{aligned}\mathrm{Cov}(X,Y)=&E((X-E(X))(Y-E(Y)))\\ =&E(XY-XE(Y)-YE(X)+E(X)E(Y))\\ =&E(XY)-E(XE(Y))-E(YE(X))+E(E(X)E(Y))\\ =&E(XY)-2E(X)E(Y)+E(X)E(Y)\\ =&E(XY)-E(X)E(Y).\end{aligned}$$

*העשרה: התפלגות אחידה

8.38 הגדרה. (התפלגות אחידה) התפלגות אחידה מוגדרת כהתפלגות כך ש

$$f_X(k)=p\quad \forall\ k\in X$$

במילים, לכל ערך $k\in X$ יש הסתברות שווה.

ניקח שני מספרים שלמים $a < b$ ונגיד שמשתנה מקרי X מתפלג אחיד על $[a, b]$. בסימונים

$$X \sim [a, b] \qquad \text{אם } P(X = k) = \frac{1}{b{-}a + 1}$$

לכל $a \leq k \leq b$ זאת אומרת התומך של משתנה אחיד שכזה הוא

$$\text{supp}(X) = a, a + 1, a + 2, \ldots, b$$

וכל הערכים מתקבלים בהסתברות שווה.

דוגמא מוכרת של התפלגות אחידה היא הטלת קוביה הוגנת. בהטלת קוביה הוגנת מקבלים ערכים מ-1 עד 6, וכל ערך מתקבל בהסתברות שווה ($\frac{1}{6}$). זה הרעיון שעומד בבסיס ההתפלגות האחידה. כמובן שנוסחה זאת עיקבית עם הדוגמא הבסיסית של קוביה הוגנת בעלת התפלגות אחידה מ-1 עד 6, עבורה כל ערך מתקבל בסיכוי $\frac{1}{6-1+1} = \frac{1}{6}$.

חישוב התוחלת והשונות של משתנה בעלת התפלגות אחידה מבוסס על נוסחאות של סכום סדרה חשבונית וסכום ריבועים של מספרים שלמים עוקבים:

8.39 מסקנה. (תוחלת של משתנה אחידה) ניקח שני מספרים שלמים $a < b$ ונגיד שמשתנה מקרי X מתפלג אחיד על $[a, b]$ בסימונים

$$X \sim [a, b] \quad \text{אם} \quad P(X = k) = \frac{1}{b{-}a + 1}$$

לכל $a \leq k \leq b$. התוחלת של משתנה בעלת התפלגות אחידה, מבוסס על נוסחאות של סכום סדרה חשבונית וסכום ריבועים של מספרים שלמים עוקבים, היא

$$\begin{aligned}
E[X] =& \sum_{k=a}^{b} k . \frac{1}{b - a + 1} \\
=& \frac{1}{b - a + 1} \sum_{k=a}^{b} k \\
=& \frac{1}{b - a + 1} . \frac{(b + a)(b - a + 1)}{2} \\
=& \frac{b + a}{2} \ ,
\end{aligned}$$

כאשר המעבר השלישי מבוסס על סכום סדרה חשבונית.

8.40 מסקנה. (שונות של משתנה אחידה) השונות של משתנה בעלת התפלגות אחידה, מבוסס על נוסחאות של סכום סדרה חשבונית ($\sum_{a}^{b} k = \frac{(b+a)(b-a+1)}{2}$) וסכום ריבועים של מספרים שלמים עוקבים ($\sum_{a}^{b} k^2 = -\frac{1}{6}(a - b - 1)\left(2a^2 + 2ab - a + 2b^2 + b\right)$), הוא

$$\begin{aligned}
E[X^2] =& \sum_{k=a}^{b} k^2 . \frac{1}{b - a + 1} \\
=& \frac{1}{b - a + 1} \sum_{k=a}^{b} k^2 \\
=& - \frac{1}{b - a + 1} . \frac{1}{6} (a - b - 1) \left(2a^2 + 2ab - a + 2b^2 + b\right) \\
=& \frac{1}{6} \left(2a^2 + 2ab - a + 2b^2 + b\right)
\end{aligned}$$

$$\begin{aligned}V(X) =& E[X^2] - (E[X])^2 \\ =& \frac{1}{6}\left(2a^2 + 2ab - a + 2b^2 + b\right) - \frac{(b+a)^2}{4} \\ =& \frac{1}{12}\left(4a^2 + 4ab - 2a + 4b^2 + 2b - 3b^2 - 3a^2 - 6ab\right) \\ =& \frac{1}{12}\left(a^2 - 2ab - 2a + b^2 + 2b\right) = \qquad \frac{1}{12}\left((b-a+1)^2 - 1\right)\end{aligned}$$

8.41 דוגמא. ריבית הלייבור במכללה שמעון נקבעת באקראי על ידי דגימה של אחד מעשרת הבנקים המסחריים. הבנקים ממספרים מ-1 עד 10 כאשר בנק i מעניק ריבית בשיעור i ללקוחותיו. לאחר זמן מה נסגרו בנקים 1 ו-2) בשל הענקת ריבית לא כדאית...). חשבו את הסיכוי שריבית הלייבור נמוכה מ-6 וכן את תוחלת ושונות הריבית. בצעו את החישובים עבור התקופה שקדמה לסגירת הבנקים ולאחר הסגירה.

פיתרון. נגדיר משתנה מקרי X להיות הריבית טרם הסגירה ו- Y להיות הריבית לאחר שנסגרו הבנקים המדוברים. לכן

$$X \sim U[1,0]$$

ו-

$$Y \sim U[3,10] \ .$$

כמו כן, לפי מסקנה 8.39

$$E[X] = \frac{10+1}{2} = 5.5$$

לפי מסקנה 8.40

$$V(X) = \frac{(10-1+1)^2-1}{12} = \frac{99}{12} \ .$$

באופן דומה עבור מאורע Y, לפי מסקנה 8.39

$$E[Y] = \frac{10+3}{2} = 6.5$$

לפי מסקנה 8.40

$$V(Y) = \frac{(10-3+1)^2-1}{12} = \frac{63}{12} \ .$$

$$P(Y \le 5) = \frac{3}{8} = 0.375 \ , \qquad P(X \le 5) = \frac{5}{10} = 0.5 \ .$$

■

9 התפלגות בינומית, גיאומטרית ופואסונית 27-7

סיכום פונקצית הסתברות, תוחלת, שונות והתפלגות בינומיאלית, גיאומטרית ופואסונית

9.1 הגדרה. (פונקצית הסתברות / פונקצית התפלגות) הפונקצית ההתפלגות של משתנה מקרי בדיד (מ"מ) X מוגדרת להיות הפונקציה $f_X(k)$ המקבלת המ"מ X ומחזירה את ההסתברות כי למ"מ X יש ערך k:

$$f_X(k) = P(X = k) \ ,$$

עם התכונות

1. $\sum\limits_{k \in X} f_X(k) = 1$.
2. $f_X(k) \geq 0 \quad \forall k$.

9.2 דוגמא. אלון הוא סוחר ממולח בשוק ההון. הוא משקיע סכום של $\$1000$ במכשיר פיננסי אשר בכל שנה מכפיל את הכסף בסיכוי $\frac{2}{3}$ או מפסיד חצי מההשקעה בסיכוי $\frac{1}{3}$. השנים הן בלתי תלויות. נגדיר את X להיות שווי ההשקעה של אלון לאחר שנתיים. התומך של X כעת הוא

$$\text{supp}(X) = \{250, 1000, 4000\} \ ,$$

על כן, ההתפלגות של X היא

$$P_X(1000) = \frac{4}{9}, \quad P_X(4000) = \frac{4}{9}, \quad P_X(250) = \frac{1}{9}$$

ו - 0 אחרת.

9.3 הגדרה. (פונקצית התפלגות מצטברת) פונקצית התפלגות מצטברת מסומנת ע"י $F_X(x)$ של מ"מ בדיד X בעל פונקצית התפלגות $f_X(k)$ מוגדרת להיות

$$F_X(x) = \sum_{\substack{k \in X \\ k \leq x}} f_X(k) \ .$$

במילים אחרות, פונקצית התפלגות מצטברת $F_X(x)$ מוגדרת להיות הפונקציה שתחומה הוא הישר הממשי וטווחה הקטע $[0, 1]$ ומקיימת

$$F_X(x) = P(X \leq x) \qquad \forall x \in X \ .$$

9.4 הגדרה. (תוחלת של משתנה נקרי בדיד) התוחלת של משתנה מיקרי X בעל פונקצית הסתברות $f_X(k)$ היא

$$E[X] = \sum_{k \in X} k f_X(k) \ .$$

9.5 דוגמא. התוחלת של הטלת קוביה היא

$$E[X] = \sum_{k=1}^{6} k \frac{1}{6} = \frac{7}{2} \ .$$

9.6 מסקנה. (תכונות של תוחלת) יהי c מספר קבוע. אזי

$$E(c) = c, \qquad E(cX) = cE(X), \qquad E(X_1 \pm X_2) = E(X_1) \pm E(X_2) \ .$$

9.7 הגדרה. (שונות של משתנה נקרי בדיד) השונות של משתנה מיקרי X בעל תוחלת $E[X]$ היא

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

9.8 מסקנה. (תכונות של שונות) יהי c מספר קבוע:

$$V(c)=0, \qquad V(cX)=c^2V(X), \qquad V(X\pm c)=V(X)\ .$$

9.9 הגדרה. (משתנה מקרי בינומי) מבצעים n ניסויים בלתי תלויים. לכל ניסוי יש 2 תוצאות אפשרויות: הצלחה עם הסתברות p או כישלון עם הסתברות $q\equiv 1-p$. משתנה מקרי בינומי X סופר את מספר הצלחות ב n הניסויים.

$$f_X(k)\equiv P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}, \qquad E[X]=np, \qquad V[X]=np(1-p)\ .$$

9.10 הגדרה. (משתנה גיאומטרי) מבצעים ניסויים בלתי תלויים עם הסתברות להצלחה p וכישלון $q\equiv 1-p$. משתנה מקרי גיאומטרי $X\sim G(p)$ סופר את מספר הניסויים עד להצלחה הראשונה (כולל).

$$f_X(k)\equiv P(X=k)=\begin{cases}(1-p)^{k-1}p & k1,2,\ldots\\ 0 & \text{else}\end{cases}, \qquad E[X]=\frac{1}{p}\ , \qquad V(X)=\frac{1-p}{p^2}\ .$$

9.11 הגדרה. (משתנה פואסוני) משתנה פואסוני זה סופר את מספר האירועים שהתרחשו ביחידת זמן, או שטחף וכדומה. λ הוא הקצב הממוצע ליחידת זמן.

$$f_X(k)\equiv P(X=k)=e^{-\lambda}\frac{\lambda^k}{k!}\ k=0,1,2,\ldots\ , \qquad E[X]=\lambda\ , \qquad V[X]=\lambda\ .$$

תרגילים

9.12 דוגמא. בוחרים קוד באורך 10 תווים, כאשר כל תו נבחר באקראי מתוך הספרות 9 – 0 ללא קשר לתווים האחרים שנבחרו. נסמן ב- X מ"מ הסופרים את מספר הרצפים 11 בקוד. חשבו את $E[X]$.

פיתרון. נגדיר מ"מ $X_i,\ i=1,\ldots,9$ כאשר X_i מקבל את הערך 1 אם ורק אם במקום ה- i מתחיל רצף של 11. לכן

$$X=\sum_{i=1}^{9}X_i\ .$$

$$\begin{aligned}E[X]=&E\left[\sum_{i=1}^{9}X_i\right]\\ =&\sum_{i=1}^{9}E\left[X_i\right]\\ =&\sum_{i=1}^{9}\left[1.P(X_i=1)+0.P(X_i=0)\right]\\ =&\sum_{i=1}^{9}P(X_i=1)\\ =&\sum_{i=1}^{9}\frac{1}{10}.\frac{1}{10}\\ =&\frac{9}{100}\ .\end{aligned}$$

■

9.13 דוגמא. ניקח משתנה מקרי X בעל התפלגות

$$P(X=k)=\begin{cases}\frac{1}{3} & k=-1,\\ \frac{1}{6} & k=0,\\ \frac{1}{2} & k=1.\end{cases}$$

חשבו את $E[X]$ ואת $E[X^4]$.

פיתרון. נשתמש בהגדרת התוחלת ונקבל

$$\begin{aligned}E[X]=&(-1).\frac{1}{3}+0.\frac{1}{6}+1.\frac{1}{2} &&=\frac{1}{6},\\ E[X^4]=&(-1)^4.\frac{1}{3}+0^4.\frac{1}{6}+1^4.\frac{1}{2} &&=\frac{5}{6}.\end{aligned}$$

■

9.14 דוגמא. נניח ש X הוא המספר של רכבים אשר עוברים דרך השתיפת רכב בין השעות $16:00$ ו- $17:00$. ל- X יש את ההתפלגות

x	4	5	6	7	8	9
$P(X=x)$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{6}$

הפונקציה $g(X)=2X-1$ מצייג את הרווח ב \$ עבור X. מצאו את תוחלת הרווח בין השעות $16:00$ ו- $17:00$.

פיתרון.

$$\begin{aligned}E[g(X)]=&E[2X-1]\\ =&\sum_{x=4}^{9}(2x-1)P_X(x)\\ =&7.\frac{1}{12}+9.\frac{1}{12}+11.\frac{1}{4}+13.\frac{1}{4}+15.\frac{1}{6}+17.\frac{1}{6}\\ =&\$12.67\ .\end{aligned}$$

■

9.15 דוגמא. בר היא בנקאית השקעות מתוחכמת המציעה ללקוחותיה את ההגרלה הבאה: בסיכוי $\frac{2}{5}$ תקבלו רווח של 10. בסיכוי $\frac{1}{5}$ לא תרוויחו ולא תפסידו דבר, ובסיכוי $\frac{3}{5}$ תפסידו 5. חשבו את התוחלת וסטיית התקן של X המייצג את הרווח מהשקעה אצל בר.

פיתרון. תוחלת הרווח היא חיובית שכן

$$\begin{aligned}\mu_X=&10.\frac{2}{5}+0.\frac{1}{5}-5.\frac{2}{5} &&=2,\\ \mu_{X^2}=&100.\frac{2}{5}+0^2.\frac{1}{5}+25.\frac{2}{5} &&=50,\\ V(X)=&\mu_{X^2}-\mu_X^2 &&=50-4=46,\\ \Rightarrow\quad \sigma_X=&\sqrt{46}\ .\end{aligned}$$

■

9.16 דוגמא. הסיכוי אשר חולה מחלים ממחלת דם חריג הוא 0.4. אם 15 אנשים נדבקים עם המחלה זו, מהי ההסתברות אשר

1. לפחות 10 יחלימו
2. בין 3 עד 8 יחלימו
3. בדיוק 5 יחלימו?

פיתרון. נגדיר X להיות מספר החולים אשר יחלימו.

.1

$$\begin{aligned} P(X \geq 10) =& 1 - P(X < 10) \\ =& 1 - \sum_{k=0}^{9} f_{X\sim \text{Bin}(15,0.4)}(k) \\ =& 1 - \sum_{k=0}^{9} \binom{15}{k} 0.4^k (1-0.4)^{15-k} \\ =& 1 - 0.9662 = 0.0338 \ . \end{aligned}$$

.2

$$\begin{aligned} P(3 \leq X \leq 8) =& 1 - P(X < 10) \\ =& \sum_{k=3}^{8} f_{X\sim \text{Bin}(15,0.4)}(k) \\ =& \sum_{k=0}^{8} f_{X\sim \text{Bin}(15,0.4)}(k) - \sum_{k=0}^{2} f_{X\sim \text{Bin}(15,0.4)}(k) \\ =& \left(\sum_{k=0}^{8} - \sum_{k=0}^{2}\right) \binom{15}{k} 0.4^k (1-0.4)^{15-k} \\ =& 0.9050 - 0.0271 = 0.8779 \ . \end{aligned}$$

.3

$$\begin{aligned} P(X = 5) =& f_{X\sim \text{Bin}(15,0.4)}(k=5) \\ =& \binom{15}{5} 0.4^5 (1-0.4)^{10} \\ =& 0.1859 \ . \end{aligned}$$

■

9.17 דוגמא. מחשב משתמש במילים באורך 8 ביטים. כאשר משדרים ביטים בקו התקשורת יש סיכוי של $.1$) לשגיאה בביט בודד שמשודר. בכדי לשפר את איכות הקו, משודר כל ביט 3 פעמים ברצף והפענוח של כל שלשת ביטים מתבצע על ידי החלטת הרוב. למשל: אם רוצים לשדר 0, משדרים בקו "000" והמקלט מפענח את התשדורת כ-0 אם השלשה שהגיעה אליו מורכבת מלפחות שני אפסים.

1. חשבו את ההסתברות שמילת מחשב תפוענח נכון.
2. מצאו את התפלגות מספר המילים שפוענחו באופן שגוי בקובץ המכיל 1000 מילים.

פתרון: נשים לב שכמות הטעויות בשליחה של ביט בודד מתפלגת $X \sim \text{Bin}(3, 0.1)$ ולכן ההסתברות שתהיה שגיאה בפענוח ביט בודד היא

$$P(X \geq 2) = \binom{3}{2} 0.1^2 0.9 + \binom{3}{1} 0.1^3 = 0.028$$

1. כמות השגיאות במילה באורך 8 ביטים מתפלגת

$$Y \sim \text{Bin}(8, 0.028)$$

ולכן המילה תפוענח נכון בהסתברות

$$P(Y=0) = \binom{8}{0} p^0 (1-p)^8 = (1-0.028)^8 = 0.796 \ .$$

2. לאור הסעיף הקודם, ההסתברות לשגיאה בפענוח מילה בודדת היא

$$1 - 0.796 = 0.203 \ ,$$

לכן מספר המילים שפוענחו באופן שגוי בקובץ המכיל 1000 מילים מתפלג $\text{Bin}(1000, 0.203)$.

9.18 דוגמא. בכד 9 כדורים לבנים ו- 1 שחור. מושכים באקראי בזה אחר זה כדור מהכד. נסמן ב- X את מספר המשיכות עד לקבלת שחור וב- Y את מספר המשיכות עד לקבלת 3 שחורים.

מצאו את התפלגות X כאשר המשיכה היא עם החזרה.

פיתרון. X מקבל ערכים $\mathbb{N}/\{0\}$. בכל הוצאה ישנה הסתברות של $\frac{9}{10}$ לכדור לבן (כישלון) והסתברות של $p = \frac{1}{10}$ לכדור לבן (הצלחה). כלומר, קיבלנו מ"מ גאומטרי

$$X \sim G\left(\frac{1}{10}\right) \ .$$

■

9.19 דוגמא. חלקיקים נפלטים מחומר רדיואקטיבי בתהליך פואסון עם קצב של 0.5 חלקיקים לשנייה. חשבו את ההסתברויות הבאות:

1. ייפלט לפחות חלקיק אחד בשנייה נתונה.
2. ייפלטו יותר מ-3 חלקיקים ב-5 שניות.

פיתרון. התהליך בשאלה הוא תהליך פואסון בעל פרמטר $\lambda = 0.5$ לכן

.1

$$P(X \geq 1) = 1 - P(X=0) = 1 - e^{-0.5}\frac{0.5^0}{0!} = 0.393 \ .$$

2. כעת בשאלה הוא תהליך פואסון שונה בעל פרמטר $\lambda = 5 \times 0.5 = 2.5$:

$$\begin{aligned} P(X_5 > 3) =& 1 - P(X_5=0) - P(X_5=1) - P(X_5=2) - P(X_5=3) \\ =& 1 - e^{-\lambda}\frac{(\lambda)^0}{0!} - e^{-\lambda}\frac{(\lambda)^1}{1!} - e^{-\lambda}\frac{(\lambda)^2}{2!} - e^{-\lambda}\frac{(\lambda)^3}{3!} \\ =& 1 - e^{-2.5}\frac{(2.5)^0}{0!} - e^{-2.5}\frac{(2.5)^1}{1!} - e^{-2.5}\frac{(2.5)^2}{2!} - e^{-2.5}\frac{(2.5)^3}{3!} \\ =& 0.242 \end{aligned}$$

■

9.20 דוגמא. (פונקצית התפלגות) מטילים 2 קוביות. נסמן ב- X את סכום התוצאות שהתקבלו. מצאו את ההתפלגות של X.

פיתרון. נתחיל מהתומך. התוצאות האפשרויות בסכום של 2 קוביות הן המספרים הטבעיים מ- 2 עד 12. בכדי לחשב את ההסתברות לכל ערך נשתמש במרחב מדגם סימטרי: גודל מרחב המדגם הוא $6^2 = 36$, ככמות האפשרויות לבחור תוצאה לקוביה הראשונה מתוך 6 ותוצאה לקוביה השניה מתוך 6. נחשב את ההסתברות של כל תוצאה בנפרד:

$$\begin{aligned}
P(X=2) &= \frac{|\{(1,1)\}|}{6^2} && = \frac{1}{36},\\
P(X=3) &= \frac{|\{(2,1),(1,2)\}|}{6^2} && = \frac{2}{36},\\
P(X=4) &= \frac{|\{(1,3),(2,2),(3,1)\}|}{6^2} && = \frac{3}{36},\\
&\vdots\\
P(X=7) &= \frac{|\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}|}{6^2} && = \frac{6}{36},\\
P(X=8) &= \frac{|\{(2,6),(3,5),(4,4),(5,3),(6,2)\}|}{6^2} && = \frac{5}{36},\\
P(X=9) &= \frac{|\{(3,6),(4,5),(5,4),(6,3)\}|}{6^2} && = \frac{4}{36},\\
&\vdots\\
P(X=12) &= \frac{|\{(6,6)\}|}{6^2} && = \frac{1}{36},
\end{aligned}$$

ובאופן דומה ממשיכים לחשב את האפשרויות לכל הערכים האפשריים בתומך. באופן כללי קל לראות שהנוסחא, כאשר p זו הסכום של ה 2 הקוביות, היא (עיין משוואה (6.9) לעייל)

$$P(X=p) = \frac{1}{36} \times \begin{cases} (p-1) & p < 8 \\ (p-1) - 2(p-7) & 8 \le p < 14 \end{cases}$$

■

9.21 דוגמא. (פונקצית התפלגות מצטברת) אלון הוא סוחר ממולח בשוק ההון. הוא משקיע סכום של $\$1000$ במכשיר פיננסי אשר בכל שנה מכפיל את הכסף בסיכוי $\frac{2}{3}$ או מפסיד חצי מההשקעה בסיכוי $\frac{1}{3}$. השנים הן בלתי תלויות. נגדיר את X להיות שווי ההשקעה של אלון לאחר שנתיים. מצאו את פונקצית ההתפלגות המצטברת של X וציירו אותה.

פיתרון. כפי שציינו בדוגמה 9.2, התומך של X כעת הוא

$$\text{supp}(X) = \{250, 1000, 4000\} \ .$$

על כן, פונקצית ההתפלגות המצטברת של X היא

$$F_X(k) = P(X \le k) = \begin{cases} 0 & k < 250, \\ \frac{1}{9} & 250 \le k < 1000, \\ \frac{5}{9} & 1000 \le k < 4000, \\ 1 & 4000 \le k. \end{cases}$$

■

9.22 דוגמא. (תוחלת) בן הוא סוחר מתוחכם ולכן הוא מציע לאלון להפקיד אצלו 100 שקלים בתחילת השנה. בסיכוי של $\frac{2}{5}$ הסכום יוכפל, ובסיכוי $\frac{2}{5}$ הסכום יצטמצם בחצי בסוף השנה. מה תהיה תוחלת הכסף של אלון לאחר קבלת ההצעה? נסמן ב- X את שווי כסף של אלון בסוף השנה.

$$E[X] = 200.\frac{2}{5} + 50.\frac{3}{5} = 110.$$

זאת אומרת שהצעתו של בן מגלמת תוחלת רווח של 10%!

9.23 דוגמא. (ליניאריות של תוחלת) בחנות ירקות קטנה ישנם 3 תפוחים ו-4 אגסים. אלון קונה 2 פירות באקראי. בן המוכר מרוויח על כל תפוח 2 ש"ח ועל כל אגס 4 ש"ח. מצאו את תוחלת הרווח של בן.

פיתרון. נגדיר מ"מ Q עבור הרווח של בן, ומשתנה מקרי X כמספר התפוחים שאלון קנה. על כן,

$$Q = 2X + 4(2 - X) = 8 - 2X$$

$$P(X = 1) = \frac{4}{7}, \qquad P(X = 2) = \frac{3}{7}.\frac{2}{6} = \frac{1}{7}, \qquad P(X = 0) = \frac{4}{7}.\frac{3}{6} = \frac{2}{7}\ .$$

מכאן נובע ש

$$E[X] = 0.\frac{2}{7} + 1.\frac{4}{7} + 2.\frac{1}{7} = \frac{6}{7}\ .$$

נשתמש בתכונת הליניאריות לחישוב התוחלת של

$$E[Q] = 8 - 2E[X] = 8 - \frac{12}{7} = \frac{44}{7}\ .$$

■

10 משתנה מקרי רציף, צפיפות והתפלגות המצטברת 1-8

משתנים מקריים רציפים אחידים

לסיכום, בשיעורים קודמים למדנו את התכונות של משתנה מקרי **בדיד**, קרי משתנה אשר שיש לו ערכים בדידים. למשל התוצאות של הטלת קוביה מרכיבות משתנה מקרי בדיד

$$X = (1,2,3,4,5,6)\,. \tag{*1}$$

למדנו גם כי הפונקציית הסתברות $f_X(k)$ של משתנה מקרי בדיד מתאימה הסתברות לכל ערך של משתנה מקרי בדיד X, לפי

$$f_X(k) = \begin{cases} \frac{1}{6} & k \in X \\ 0 & \text{אחרת} \end{cases} \tag{*2}$$

בשונה לזה, משתנה מקרי רציף הוא משתנה בעל ערכים רציפים בין שתי גבולות. למשל, גמל שותה מים על בסיס שבועי. בפעם שבוע הגמל בוא לשתות לתקופת זמן בין 30 דקות עד 90 דקות. התקופת זמן X מהווה משתנה מקרי רציף. X הוא משתנה מקרי רציף **אחיד**. הגמל שותה לתקופת זמן בין 30 עד 90 דקות. על כן הערכים של X הם רציפים בין 30 עד 60:

$$X = \begin{cases} x & 30 \le x \le 90 \\ 0 & \text{אחרת.} \end{cases} \tag{*3}$$

הפונקציית צפיפות של מ"מ רציף היא האנלוגית לפונקציית הסתברות של מ"מ בדיד. האורך של X הוא $90-30=60$ דקות ולכן הפונקציית צפיפות נותנת את המספר דקות שהגמל שותה כחלק של האורך של 60 דקות של X ומוגדרת להיות

$$f_X(x) = \frac{1}{90-30} = \frac{1}{60}\ . \tag{*4}$$

כמו כן,

$$f_X(30)=0, \qquad f_X(40)=\frac{1}{9}, \qquad f_X(60)=\frac{1}{2}, \qquad f_X(90)=1.$$

ניתן לחשב את ההסתברות אשר הגמל שותה בין a דקות עד b דקות על ידי

$$P(a \le X \le b) = \int_a^b f_X(x)dx\ . \tag{*5}$$

לדוגמה, ההסתברות שהגמל שותה בין 30 עד 50 דקות היא

$$P(30 \le X \le 50) = \int_{30}^{50} f_X(x)dx = \int_{30}^{50} \frac{1}{60}dx = \left[\frac{x}{60}\right]_{30}^{50} = \frac{50-30}{60} = \frac{1}{3}\ . \tag{*6}$$

באופן כללי נתון מ"מ רציף אחיד X בעל ערכים רציפים על הקטע $[A,B]$, כלומר

$$X = \begin{cases} x & A \le x \le B \\ 0 & \text{אחרת.} \end{cases} \tag{*7}$$

מסמנים X ב

$$X \sim U(A,B)\ .$$

10.1 הגדרה. (פונקציית צפיפות של מ"מ רציף אחיד) באופן כללי, ניקח קטע סופי $[A, B]$ ונגדיר ש- X מתפ[לג אחיד בקטע זה, $X \sim U(A, B)$ אם ורק אם

$$f_X(x) = \begin{cases} \dfrac{1}{B-A} & A \leq x \leq B, \\ 0 & \text{אחרת.} \end{cases}$$

במילים, **משתנה מקרי אחיד רציף הוא משתנה בעלת צפיפות קבועה** על קטע כלשהו $[A, B]$, וביתר הנקודות הצפיפות שווה לאפס.

כעת חוזרים לשאלה אשר הוביל למשוואה (6*): מהי ההסתברות אשר הגמל ישתה לתקופת עד 50 דקות? או באופן כללי, מהי ההסתברות אשר הגמל ישתה לתקופת עד k דקות כאשר $30 \leq k \leq 90$? בדיוק כמו (6*) התשובה ניתנת ע"י האינטגרל

$$P(A \leq k \leq A) = \int_A^k f_X(x) = \frac{k-A}{B-A} \tag{*8}$$

המשוואה זו היא בעצם **הפונקציית ההתפלגות המצטברת** $F_X(k)$. פורמאלית:

10.2 הגדרה. (פונקציית ההתפלגות המצטברת של מ"מ רציף אחיד) פונקציית ההתפלגות המצטברת של משתנה מקרי X מתפלג אחיד על הקטע $[A, B]$ היא ליניארית (קו ישר), ניתנת על ידי

$$F_X(k) = \begin{cases} 0, & k < A, \\ \dfrac{k-A}{B-A}, & 0 \leq k \leq B, \\ 1, & k > B. \end{cases}$$

עבור משתנה בדיד

$$F_X(b) - F_X(a) = P(a < X < b) = \sum_{k>a}^{b} P_X(k),$$

בעוד עבור משתנה רציף מתקבל השוויון הבא:

$$F_X(b) - F_X(a) = P(a < X < b) = \int_a^b f_X(x)\, dx \ .$$

זאת אומרת, בשני המיקרים ניתן לסכום על פונקצית ההתפלגות הצפיפות בכדי לקבל את פונקתיית ההתפלגות המצטברת. במקרה הבדיד מדובר בסכום ובמקרה הרציף מדובר באינגרל.

10.3 מסקנה. (תוחלת של מ"מ רציף אחיד) התוחלת של משתנה מקרי מתפלג אחיד על הקטע $[A, B]$ היא

$$E[X] = \frac{B+A}{2} \ .$$

הוכחה.

$$E[X] = \int_A^B x f_X(x) dx = \int_A^B x \left(\frac{1}{B-A} \right) dx = \frac{1}{B-A} \int_A^B x = \frac{1}{B-A} \left[\frac{x^2}{2} \right]_A^B = \frac{1}{B-A} \left(\frac{B^2 - A^2}{2} \right)$$
$$= \frac{A+B}{2} \ .$$

■

10.4 מסקנה. (שונות של מ"מ רציף אחיד) השונות של משתנה מקרי מתפלג אחיד על הקטע $[A,B]$ היא

$$V[X]=\frac{(B-A)^2}{12}\ .$$

הוכחה.

$$V[X]=E[X^2]-E[X]^2$$

קודם כל נחשב את $E[X^2]$:

$$E[X^2]=\int_A^B x^2 f_X(x)dx=\int_A^B x^2\left(\frac{1}{B-A}\right)dx=\frac{1}{B-A}\int_A^B x^2dx=\frac{1}{B-A}\left[\frac{x^3}{3}\right]_A^B=\frac{1}{B-A}\left(\frac{B^3-A^3}{3}\right).$$

ומ מסקנה 10.3, $E[X]^2=\left(\frac{A+B}{2}\right)^2$. לכן

$$\begin{aligned}V[X]=&E[X^2]-E[X]^2=\frac{1}{B-A}\left(\frac{B^3-A^3}{3}\right)-\frac{(A+B)^2}{4}\\=&\frac{1}{12(B-A)}\left[4(B^3-A^3)-3(B-A)(A+B)^2\right]\\=&\frac{1}{12(B-A)}\left[4B^3-4A^3-3BA^2-6AB^2-3B^3+3A^3+6A^2B+3B^2A\right]\\=&\frac{1}{12(B-A)}\left[B^3-A^3+3BA^2-3AB^2\right]\\=&\frac{(B-A)^3}{12(B-A)}\\=&\frac{(B-A)^2}{12}.\end{aligned}$$

■

10.5 דוגמא. אוטובוס מגיע לתחנה כל 30 דקות. בר מגיעה לתחנה באקראי, מבלי לתכנן מראש ומבלי לבדוק את לוחות הזמנים של האוטובוס. נסמן ב-T את זמן ההמתנה המדוייק לאוטובוס, בדקות. רשמו את פונקצית ההתפלגות המצטברת ואת פונקצית הצפיפות של T .חשבו את ההסתברויות

$$P(20\le T\le 40)$$

ו-

$$P(T>23)\ .$$

ענו שוב על השאלה עבור אוטובוס המגיע כל 40 דקות.

פיתרון. במקרה הנוכחי, זמן ההמתנה הוא אחיד משום שמועד ההגעה של בר הוא אחיד בזמן לכן,

$$F_T(t)=\begin{cases}0, & t<0\\ \frac{t}{30}, & 0\le t\le 30,\\ 1, & t\ge 30,\end{cases}$$

$$f_T(t)=F_T'(t)=\begin{cases}\frac{1}{30}, & 0\le t\le 30,\\ 0, & \text{אחרת},\end{cases}$$

$$P(20\le T\le 40)=F_T(40)-F_T(20)=1-\frac{20}{30}=\frac{1}{3},$$

$$P(T>23)=1-P(T\le 23)=1-F_T(23)=1-\frac{23}{30}=\frac{7}{30}.$$

עבור אוטובוס המגיע כל 40 דקות נקבל

$$F_T(t)=\begin{cases}0, & t<0\\ \frac{t}{40}, & 0\le t\le 40,\\ 1, & t\ge 40,\end{cases}$$

$$f_T(t)=F_T'(t)=\begin{cases}\frac{1}{40}, & 0\le t\le 40,\\ 0, & \text{אחרת},\end{cases},$$

$$P(20\le T\le 40)=F_T(40)-F_T(20)=1-\frac{20}{40}=\frac{1}{2},,$$

$$P(T>23)=1-P(T\le 23)=1-F_T(23)=1-\frac{23}{40}=\frac{17}{40}.,$$

■

10.6 דוגמא. מכונה במפעל ברגים אמורה לייצר ברגים בקוטר 3 מ"מ. בפועל, קוטר הברגים אינו מדויק והצפיפות שלו נתונה על ידי איור להלן אשר מראה את צפיפות קוטר הברגים.

רשמו במדויק את פונקצית הצפיפות. מה ההסתברות שקוטר בורג יעלה על 3 מ"מ? על 4 מ"מ?

פיתרון. יש צורך למצוא את פונקצית הצפיפות בצורה מדוייקת. נשתמש תחילה בתנאי הנרמול. אנחנו יודעים כי השטח שבין הגרף לציר x שווה ל-1 כי כלל השטח מייצג את כל ההסתברות,

$$\int_1^5 f_X(x)\,dx=P(1\le X\le 5)=1.$$

נשתמש בנוסחה לשטח משולש (שטח משולש שווה למכפלת הבסיס בחצי הגובה) ונקבל שגובה המשולש h מקיים

$$\frac{4h}{2}=1\qquad\Rightarrow\qquad h=0.5\ .$$

עם כן, יש לנו את הנקודות

$$(1,0),\quad (5,0),\quad (3,0.5)\ .$$

ונוכל למצוא את משוואות הקווים בהתאם. לכן הצפיפות נתונה על ידי

$$f_X(x)=\begin{cases}0, & x<1\\ \frac{1}{4}, & 1\le x\le 3,\\ -\frac{1}{4}(x-5), & 3\le x\le 5,\\ 0, & x>5.\end{cases},$$

בעזרת הצפיפות נוכל לחשב ישירות את ההסתברויות המבוקשות.

$$P(X>3)=\int_3^5\left(-\frac{1}{4}(x-5)\right)dx \quad =-\frac{1}{4}\left[\frac{x^2}{2}-5x\right]_3^5=\frac{1}{2},$$
$$P(X>4)=\int_4^5\left(-\frac{1}{4}(x-5)\right)dx \quad =-\frac{1}{4}\left[\frac{x^2}{2}-5x\right]_4^5=\frac{1}{8}.$$

■

משתנה מקרי רציף מעריכי

קודם למדנו מזה התפלגות פואסון. הגדרנו משתנה פואסוני X **כמשתנה מקרי בדיד אשר סופר את מספר האירועים** k **שהתרחשו ביחידת זמן**, או ביחידת שטח וכדומה. כאשר λ הוא הקצב הממוצע ליחידת זמן, אזי הפונקציית ההסתברות $f_X(k)$ היא מוגדרת להיות ההסתברות $P(X=k)$ אשר k אירועים התרחשו ביחידת זמן (k שיחות התקבלו למוקד שרות בשעה כלשהי כאשר λ הוא המספר שיחות לשעה הממוצע, או k חלקיקים נפלטו בשנייה כלשהי כשטר λ הוא המספר החלקיקים הממוצע הנפלטים לשעה) וניתנת ע"י

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

יחד עם התוחלת והשונות

$$E[X]=V[X]=\lambda.$$

במקום לחשוב על מספר השיחות הנכנסות בשעה (אשר הוא משתנה מקרי בדיד), נחשוב על **הזמן בין כל שיחה לשיחה, אשר הוא משתנה מקרי רציף.** לדוגמא, אם בממוצע נכנסות 5 שיחות בשעה, אזי הזמן הממוצע בין השיחות הוא $\frac{60}{5}=12$ דקות. הזמן הוא רציף ולכן הזמן בין השיחות השונות הוא משתנה מקרי רציף, הנקרא **משתנה מקרי מעריכי** . משתנה מקרי מעריכי מודד את הזמן (או המרחק) בין אירועים שונים המתרחשים לפי תהליך פואסון (התפלגות פואסונית). מסמנים משתנה מקרי מעריכי ב

$$X\sim\exp(-\lambda).$$

בדוגמה זו

$$\lambda=\frac{1}{12}\ \text{שיחות לדקה}$$

10.7 הגדרה. (פונקציית צפיפות של משתנה מקרי רציף מעריכי) משתנה מקרי מעריכי X מסומן ב

$$X\sim\exp(\lambda)$$

כאשר λ הוא הפרמטר של התפלגות פואסונ המתאימה, זאת אומרת קצב האירועים הממוצה ליחדת זמן (או שטח). הצפיפות של X היא

$$f_X(x)=\begin{cases}0, & x<0\\ \lambda e^{-\lambda x} & x\geq 0.\end{cases}$$

10.8 דוגמא. לדוגמה, נגיד שמוקד שרות פתוח בין השעות $18:00$ עד $22:00$, כלומר לאורך זמן של 240 דקות. ההסתברות ששיחה התקבלה בין $18:05$ עד $18:10$, כלומר תוך 5 דקות כלשהן, כאשר $\lambda=\frac{1}{12}$, היא

$$P(18:05–18:10)=P(X=5)=\int_0^5 f_X(x)dx=\int_0^5\lambda e^{-\lambda x}dx=\left[-e^{-\lambda x}\right]_0^5=1-e^{-5\lambda}=1-e^{-5/12}=0.34.$$

ההסתברות ששיחה אחת תגיע תוך 15 דקות היא

$$P(X=15)=\int_0^{15}f_X(x)dx=\int_0^{15}\lambda e^{-\lambda x}dx=\left[-e^{-\lambda x}\right]_0^{15}=1-e^{-15\lambda}=1-e^{-15/12}=0.71.$$

10.9 מסקנה. (פונקציית התפלגות המצטברת של מ"מ רציף מעריכי) פונקציית התפלגות המצטברת של משתנה נקרי רציף מעריכי X הוא

$$F_X(k) = \begin{cases} 0, & k < 0 \\ 1 - e^{-\lambda k} & k \geq 0 \end{cases}$$

הוכחה.

$$F_X(k) = \int_0^k f_X(x)\,dx = \int_0^k \lambda e^{-\lambda x} = 1 - e^{-\lambda k}\ .$$

■

10.10 דוגמא. ההסתברות אשר השיחה הראשונה תגיע בין הדקה השנייה לבין הדקה הרביעית היא

$$P(2 \leq X \leq 4) = F_X(4) - F_X(2) = 1 - e^{-4/12} - (1 - e^{-2/12}) = 0.13\ .$$

10.11 דוגמא. בהנחה שבחמש הדקות הראשונות לא התקבל שיחה, ההסתברות שנמתין לפחות עוד דקה עד לשיחה הראשונה ניתנת ע"י

$$P(X > 5{+}1|X > 5) = \frac{P(X > 6 \cap X > 5)}{P(X > 5)} = \frac{P(X > 6)}{P(X > 5)} = \frac{1 - F_X(6)}{1 - F_X(5)} = \frac{e^{-6/12}}{e^{-5/12}} = e^{-1/12} = P(X > 1).$$

10.12 דוגמא. מספר טיפות הגשם הנופלות ביחידת שטח נתונה מתפלגות על פי התפלגות פואסון. ניקח מטר אחד ונניח כי בממוצע נפלו לאורכו 10 טיפות:

$\lambda = 10$ טיפות למטר .

המרחק X (במטרים) בין טיפות מתפלג פואסונית

$$X \sim \exp(\lambda = 10).$$

ההסתברות שטיפה אחד תיפול לתוך 10 ס"מ (0.1 [m]) כלשהו היא

$$P(X \leq 2) = F_X(20) = 1 - e^{-0.1\lambda} = 1 - e^{-1} = 0.63\ .$$

10.13 מסקנה. (תוחלת של מ"מ רציף מעריכי) התוחלת של משתנה מקרי רציף מעריכי היא

$$E[X] = \frac{1}{\lambda}\ .$$

הוכחה.

$$\begin{aligned} E[X] &= \int_0^\infty x f_X(x)\,dx \\ &= \int_0^\infty x\lambda e^{-\lambda x} \\ &= \left[\frac{x\lambda e^{-\lambda x}}{-\lambda}\right]_0^\infty - \int_0^\infty e^{-\lambda x}\,dx \\ &= 0 - \left[\frac{e^{-\lambda x}}{-\lambda}\right]_0^\infty = \frac{1}{\lambda}\ . \end{aligned}$$

■

10.14 מסקנה. (שונות של מ"מ רציף מעריכי) השונות של משתנה מקרי רציף מעריכי הו א

$$V[X] = \frac{1}{\lambda^2}\ .$$

הוכחה.

$$
\begin{aligned}
E[X^2] =& \int_0^\infty x^2 f_X(x)\,dx \\
=& \int_0^\infty x^2 \lambda e^{-\lambda x} \\
=& \left[\frac{x^2\lambda e^{-\lambda x}}{-\lambda}\right]_0^\infty - \frac{2\lambda}{-\lambda}\int_0^\infty x e^{-\lambda x}\,dx \\
=& 0 + 2\left[\frac{x e^{-\lambda x}}{-\lambda}\right]_0^\infty - \frac{2}{-\lambda}\int_0^\infty e^{-\lambda x}\,dx \\
=& \frac{2}{\lambda}\left[\frac{e^{-\lambda x}}{-\lambda}\right]_0^\infty \\
=& \frac{2}{\lambda^2}\ ,
\end{aligned}
$$

ע ל כן

$$
V(X) = E[X^2] - E[X]^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}\ .
$$

■

11 תרגילים משתנה מקרי רציף 3-8

סיכום נוסחאות: משתני מקרי רציפים

11.1 חוק. () עבור משתנה מקרי בדיד כלשהו

$$F_X(b) - F_X(a) = P(a < X < b) = \sum_{k>a}^{b} P_X(k),$$

בעוד עבור משתנה מקרי רציף כלשהו מתקבל השוויון הבא:

$$F_X(b) - F_X(a) = P(a < X < b) = \int_a^b f_X(x)\,dx \ .$$

זאת אומרת, בשני המיקרים ניתן לסכום על פונקצית ההתפלגות הצפיפות בכדי לקבל את פונקציית ההתפלגות המצטברת. במקרה הבדיד מדובר בסכום ובמקרה הרציף מדובר באינטגרל.

11.2 הגדרה. (פונקציית צפיפות של מ"מ רציף אחיד) נתון קטע סופי $[A,B]$, משתנה מקרי רציף X מתפלג אחיד בקטע זה, $X \sim U(A,B)$ אם ורק אם

$$f_X(x) = \begin{cases} \dfrac{1}{B-A} & A \le x \le B, \\ 0 & \text{אחרת}. \end{cases}$$

במילים, **משתנה מקרי אחיד רציף הוא משתנה בעלת צפיפות קבועה** על קטע כלשהו $[A,B]$, וביתר הנקודות הצפיפות שווה לאפס.

11.3 הגדרה. (פונקציית ההתפלגות המצטברת של מ"מ רציף אחיד) פונקציית ההתפלגות המצטברת של משתנה מקרי X מתפלג אחיד על הקטע $[A,B]$ היא ליניארית (קו ישר), ניתנת על ידי

$$F_X(k) = \begin{cases} 0, & k < A, \\ \dfrac{k-A}{B-A}, & 0 \le k \le B, \\ 1, & k > B. \end{cases}$$

11.4 מסקנה. (תוחלת ושונות של מ"מ רציף אחיד) עבור משתנה מקרי מתפלג אחיד על הקטע $[A,B]$, כלומר $X \sim U(A,B)$:

$$f_X(x) \equiv P(X=x) = \begin{cases} \dfrac{1}{B-A} & A < x < B \\ 0 & \text{אחרת} \end{cases} , \qquad E[X] = \frac{A+B}{2} \ , \qquad V(X) = \frac{(B-A)^2}{12} \ .$$

11.5 הגדרה. (פונקציית צפיפות של מ"מ רציף מעריכי) משתנה מקרי מעריכי X מסומן ב

$$X \sim \exp(\lambda)$$

כאשר λ הוא הפרמטר של התפלגות פואסון המתאימה, זאת אומרת קצב האירועים הממוצה ליחדת זמן (או ליחידת שטח). הצפיפות של X היא

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

11.6 מסקנה. (פונקציית התפלגות המצטברת של מ"מ רציף מעריכי) פונקציית התפלגות המצטברת של משתנה מקרי רציף מעריכי X הוא

$$F_X(k) = \begin{cases} 0, & k < 0 \\ 1 - e^{-\lambda k} & k \geq 0 \end{cases}$$

11.7 הגדרה. (תוחלת ושונות של מ"מ רציף מעריכי) עבור משתנה מקרי מתפלג מעריכי עם פרמטר λ, כלומר $X \sim \exp(\lambda)$:

פונקציית צפיפות:
$$f_X(x) = \begin{cases} 0, & x < 0 \\ \lambda e^{-\lambda x} & x \geq 0. \end{cases}$$

פונקציית התפלגות מצטברת:
$$F_X(k) = \begin{cases} 0, & k < 0 \\ 1 - e^{-\lambda k} & k \geq 0 \end{cases}$$

תוחלת:
$$E(X) = \frac{1}{\lambda}$$

שונות:
$$V(X) = \frac{1}{\lambda^2}$$

תרגילים

11.8 דוגמא. מכונה במפעל ברגים אמורה לייצר ברגים בקוטר 3 מ"מ. בפועל, קוטר הברגים אינו מדויק והצפיפות שלו נתונה על ידי איור להלן אשר מראה את צפיפות קוטר הברגים.

רשמו במדויק את פונקצית הצפיפות. מה ההסתברות שקוטר בורג יעלה על 3 מ"מ? על 4 מ"מ?

פיתרון. יש צורך למצוא את פונקצית הצפיפות בצורה מדוייקת. נשתמש תחילה בתנאי הנרמול. אנחנו יודעים כי השטח שבין הגרף לציר x שווה ל-1 כי כלל השטח מייצג את כל ההסתברות,

$$\int_1^5 f_X(x)\,dx = P(1 \leq X \leq 5) = 1.$$

נשתמש בנוסחה לשטח משולש (שטח משולש שווה למכפלת הבסיס בחצי הגובה) ונקבל שגובה המשולש h מקיים

$$\frac{4h}{2} = 1 \qquad \Rightarrow \qquad h = 0.5\ .$$

עם כן, יש לנו את הנקודות

$$(1,0), \quad (5,0), \quad (3,0.5)\ .$$

ונוכל למצוא את משוואות הקווים בהתאם. לכן הצפיפות נתונה על ידי

$$f_X(x) = \begin{cases} 0, & x < 1 \\ \frac{1}{4}, & 1 \le x \le 3, \\ -\frac{1}{4}(x-5), & 3 \le x \le 5, \\ 0, & x > 5. \end{cases} ,$$

בעזרת הצפיפות נוכל לחשב ישירות את ההסתברויות המבוקשות.

$$P(X > 3) = \int_3^5 \left(-\frac{1}{4}(x-5)\right) dx \quad = -\frac{1}{4}\left[\frac{x^2}{2} - 5x\right]_3^5 = \frac{1}{2},$$
$$P(X > 4) = \int_4^5 \left(-\frac{1}{4}(x-5)\right) dx \quad = -\frac{1}{4}\left[\frac{x^2}{2} - 5x\right]_4^5 = \frac{1}{8}.$$

■

11.9 דוגמא. מספר טיפות הגשם הנופלות ביחידת שטח נתונה מתפלגות על פי התפלגות פואסון. ניקח מטר אחד ונניח כי בממוצע נפלו לאורכו 10 טיפות:

$$\lambda = 10 \text{ טיפות למטר} .$$

המרחק X (במטרים) בין טיפות מתפלג פואסונית

$$X \sim \exp(\lambda = 10).$$

ההסתברות שטיפה אחד תיפול לתוך 10 ס"מ (0.1 [m]) כלשהו היא

$$P(X \le 2) = F_X(20) = 1 - e^{-0.1\lambda} = 1 - e^{-1} = 0.63 .$$

11.10 דוגמא. משתנה מקרי רציף X בעל פונקציית צפיפות

$$f_X(x) = \begin{cases} cx, & 0 \le x \le 2, \\ 0, & \text{אחרת} \end{cases}$$

חשבו את c, ומצאו את פונקציית ההתפלגות המצטברת F_X.

פיתרון. בכדי למצוא את הקבוע c נשתמש בתנאי הנרמול. ידוע כי

$$1 = \int_{-\infty}^{\infty} dx\, f_X(x) = \int_0^2 dx\, cx = c\left[\frac{x^2}{2}\right]_0^2 = 2c,$$

לכן

$$c = \frac{1}{2}.$$

עבור $k < 0$, ההסתברות

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

כי המשתנה נע בין 0 ל- 2. עבור $k \geq 2$, ההסתברות

$$P(X \leq k) = 1$$

מאותה הסיבה בדיוק. כעת נותר למצוא את הערך עבור $k \in (0,2)$.

$$\begin{aligned} F_X(k) &= \int_{-\infty}^{k} f_X(x)\,dx \\ &= \int_{-\infty}^{0} f_X(x)\,dx + \int_{0}^{k} f_X(x)\,dx \\ &= \int_{-\infty}^{0} 0\,dx + \int_{0}^{k} \frac{x}{2}\,dx \\ &= \left[\frac{x^2}{4}\right]_0^k = \frac{k^2}{4}. \end{aligned}$$

לסיכום, פונקצית ההתפלגות המצטברת היא

$$F_X(k) = \begin{cases} 0, & k < 0 \\ \dfrac{k^2}{4}, & 0 \leq k \leq 2, \\ 1, & k > 2. \end{cases}$$

■

11.11 דוגמא. נתונה פונקציית הצפיפות הבאה:

$$f_X(x) = \begin{cases} -x & -1 \leq x \leq 0, \\ cx^2 & 0 < x \leq 1 \\ 0 & |x| > 1 \end{cases}$$

1. מצאו את ערכו של c.
2. חשבו את פונקציית ההתפלגות המצטברת של X.
3. חשבו את ההסתברויות:
   (א) $P(X \leq -0.5)$
   (ב) $P(X < -0.5)$
   (ג) $P(X \leq 0.5)$
   (ד) $P(-0.2 \leq X \leq 0.3)$

פיתרון. 1. מתכונת הנרמול נקבל

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} dx\, f_X(x) = \int_{-1}^{0} (-x)dx + \int_{0}^{1} cx^2\,dx \\ &= \frac{1}{2} + c\left(\frac{1}{3} - 0\right) \end{aligned}$$

ולכן

$$c = 1.5\ .$$

.2

$$F_X(k)=\int_{-\infty}^{k} f_X(x)\,dx=\begin{cases}0 & k<-1\\ \frac{1}{2}-\frac{k^2}{2} & -1\le k<0\\ \frac{1}{2}+\frac{k^3}{2} & 0\le k\le 1\\ 1 & k>1.\end{cases}$$

3. נחשב את ההסתברויות באמצעות הצבה בפונקציית ההתפלגות המצטברת:
   (א) $P(X\le -0.5)=F_X(-0.5)=0.375$
   (ב) $P(X<0.5)=P(X\le -0.5)=0.375$ שכן כזכור, נקודה אחת אינה משפיעה על תוצאת אינטגרל וההסתברות להיות שווה בדיוק ל- 0.5 היא אפס.
   (ג) $P(X\le 0.5)=F_X(0.5)=0.5625$
   (ד) $P(-0.2\le X\le 0.3)=F_X(0.3)-F_X(-0.2)=0.0335$

■

דוגמא. אספקת הדלק למאגר של תחנת דלק מתבצעת אחת לשבוע. כמות הדלק (באלפי ליטרים) שתחנה זו מוכרת בשבוע היא משתנה מקרי המתפלג בהתאם לפונקציית הצפיפות:

$$f_X(x)=\begin{cases}c(1-x)^4 & x\in[0,1]\\ 0 & \text{אחרת}\end{cases}$$

1. חשבו את c.
2. מה צריכה להיות קיבולת המאגר כדי שההסתברות שהוא יתרוקן בשבוע תהיה קטנה מ 5%?

פיתרון. 1. מתכונת הנרמול נקבל

$$1=\int_{-\infty}^{\infty} f_X(x)\ dx=c\int_0^1(1-x)^4\ dx=-\frac{c}{5}(1-x)^5\Big|_{x=0}^{x=1}=\frac{c}{5}\ ,$$

ולכן

$$c=5.$$

2. סמן את קיבולת המאגר ב M. אנו דורשים שההסתברות שצריכת הדלק תהיה גדולה מהקיבולת תהיה קטנה מ- 5%, כלומר

$$P(X>M)\le 5\%\ .$$

$$P(X>M)=\int_M^1 f_X(x)\,dx=\int_M^1 c(1-x)^4\,dx=-\frac{c}{5}(1-x)^5\Big|_M^1=(1-M)^5\le 0.05\ ,$$

ולכן

$$M\ge 0.4507\ .$$

המספר שקיבלנו הוא בדיוק האחוזון ה- 95% מהשבועות צריכת הדלק נמוכה ממספר זה.

■

11.12 דוגמא. תרבית חיידקים מפוזרת באופן אחיד על פני צלחת שרדיוסה 10 ס"מ. יהא R המרחק של חיידק אקראי ממרכז הצלחת.

1. מה ההסתברות שהמרחק הוא בדיוק 3 ס"מ?
2. מצאו את פונקציית ההתפלגות המצטברת של R.
3. מהי ההסתברות שהמרחק הוא 3 ס"מ לפחות אם ידוע שהוא 6 ס"מ לכל היותר?

4. מצאו את פונקציית הצפיפות של R.

פיתרון. 1. מאחר וזהו משתנה מקרי רציף, ההסתברות שיקבל בדיוק ערך מסוים היא אפס.

2. מאחר והחיידקים מפולגים באופן אחיד על הצלחת, הסיכוי שחיידק מסוים נמצא במרחק קטן מ-r היא היחס בין השטח של האזורים שנמצאים במרחק קטן מ-r ביחס לשטח הכולל. לכן

$$F_R(r)=\begin{cases}0, & r<0,\\ \frac{r^2}{10^2}, & 0\le r\le 10,\\ 1, & r>10.\end{cases}$$

3.
$$P(R>3|R>6)=\frac{P(3<R<6)}{P(R<6)}=\frac{F_R(6)-F_R(3)}{F_R(6)}=0.75\ .$$

4.
$$f_R(r)=\frac{dF_R}{dr}=\begin{cases}\frac{r}{50}, & 0\le r\le 10,\\ 0, & \text{אחרת}\ .\end{cases}$$

■

11.13 דוגמא. בתחנת כיבוי האש של אשדוד מתקבלת שיחה אחת בממוצע בכל 3 דקות.

1. מה הסיכוי שהשיחה הראשונה תגיע בין הדקה השנייה לדקה הרביעית?
2. מה הסיכוי שהשיחה הראשונה תגיע אחרי 5 דקות?
3. בהנחה שבחמש הדקות הראשונות לא התקבלה שיחה, מה הסיכוי שנמתין לפחות עוד דקה עד לשיחה הראשונה?

פיתרון. הזמן עד השיחה הראשונה מתפלג פואסונית $Y\sim \exp\left(\frac{1}{3}\right)$ כאשר הזמן נמדד בדקות.

1.
$$P(2\le Y\le 4)=F_Y(4)-F_Y(2)=1-e^{-4/3}-(1-e^{-2/3})\approx 0.25.$$

2.
$$P(Y>5)=1-P(Y\le 5)=1-F_Y(5)=1-(1-e^{-5/3})\approx 0.189.$$

3.
$$P(Y>5{+}1|Y>5)=\frac{P(Y>6\cap Y>5)}{P(Y>5)}=\frac{P(Y>6)}{P(Y>5)}=\frac{1-F_Y(6)}{1-F_Y(5)}=\frac{e^{-6/3}}{e^{-5/3}}=e^{-1/3}=P(Y>1)\ .$$

בחישוב האחרון קיבלנו תכונה מוכרת: חוסר זיכרון. הסיכוי שנמתין 6 דקות לפחות עד לשיחה הראשונה, בהנחה שכבר המתנו 5 דקות ללא שיחה, היא בדיוק הסיכוי שנמתין דקה אחת נוספת - זו תכונת חוסר הזיכרון שראינו עבור משתנה מקרי גיאומטרי. נסכם בטענה הבאה. ■

11.14 מסקנה. (תכונת חוסר זיכרון) עבור משתנה מקרי מעריכי $X\sim\exp(\lambda)$, וכל צמד מספרים $s,t>0$, מתקיים

$$P(X>s+t|X>s)=P(X>t).$$

הוכחה.

$$\begin{aligned}P(X>s+t|X>s)=&\frac{P(X>s+t\cap X>s)}{P(X>s)}\\=&\frac{P(X>s+t)}{P(X>s)}\\=&\frac{1-F_X(s+t)}{1-F_X(s)}\\=&\frac{1-(1-e^{-\lambda(s+t)})}{1-(1-e^{-\lambda s})}\\=&\frac{e^{-\lambda(s+t)}}{e^{-\lambda s}}\\=&e^{-\lambda t}\\=&1-(1-e^{-\lambda t})\\=&1-F_X(t)\\=&P(X>t)\end{aligned}$$

■

11.15 דוגמא. בשוק שני יצרני נורות. מפעל 1 שולט על 60% מהשוק ומפעל 2 על היתר. אורך חיים ממוצע של נורה ממפעל 1 הוא חודשיים, בעוד מפעל 2 מייצר נורות עם אורך חיים ממוצע של ארבעה חודשים. עבור נורה שנקנית בשוק (

1. מה הסיכוי שהנורה לא תתקלקל בחודש הראשון?
2. בהנחה והנורה שקנינו דולקת כבר חודש ללא תקלה, מה הסיכוי שמדובר בנורה ממפעל 2?

פיתרון. נסמן את זמן החיים של הנורות:

$$X_1\sim\exp\left(\frac{1}{2}\right),X_2\sim\exp\left(\frac{1}{4}\right).$$

נסמן את אורך החיים של הנורה שנקנתה ב- Y.

.1

$$\begin{aligned}P(Y>1)=&P(Y>1|Y=X_1)P(Y=X_1)+P(Y>1|Y=X_2)P(Y=X_2)\\=&P(Y>1|Y=X_1)0.6+P(Y>1|Y=X_2)0.4\\=&(1-F_{X_1}(1))0.6+(1-F_{X_2}(1))0.4\\=&e^{-\frac{1}{2}\cdot 1}.0.6+e^{-\frac{1}{4}\cdot 1}.0.4\\\approx&0.675\end{aligned}$$

.2

$$P(Y=X_2|Y>1)=\frac{P(Y>1|Y=X_2}{P(Y>1)}=\frac{e^{-\frac{1}{4}\cdot 1}.0.4}{0.675}\approx 0.462>0.4$$

ז"א שזמן החיים הארוך בפועל מגדיל את הסיכוי שמדובר בנורה ממפעל 2, שכן אורך החיים הממוצע של הנורות ממפעל2 הוא ארוך יותר מאלו של מפעל 1. ■

12 התפלגות נורמלית 8-8

התפלגות נורמלית

תפלגות נורמאלית היא אחת ההתפלגויות הנפוצות ביותר בעולם הסטטיסטי וההסתברותי. לדוגמא, המשקל או הגובה של אוכלוסיה מסויימת, לחץ הדם של קבוצת אנשים גדולה, ה אורך החיים של מכוניות במדינה כלשהי, ועוד.

התוצאה הממוצעת של סדרת ניסויים בלתי תלויים מתפלגת נורמאלית !

הדרך הטובה ביותר לתאר התפלגות נורמאלית היא באמצעות עקומת פעמון כמתואר באיור:

באיור מוצגים גרפים אשר נראים כמו פעמונים. הכל אחד מהם מייצג את הצפיפות של ההתפלגות הנורמאלית.

12.1 הגדרה. (צפיפות של התפלגות נורמאלית)

הנוסחה האלגברית של הפונקציית הצפיפות $f_X(x)$ של משתנה מקרי X אשר מתפלג נורמאלי היא מסומן ב $n(x,\mu,\sigma)$ כאשר

$$n(x,\mu,\sigma) := \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

פרמטרים של ההתפלגות הנורמאלית הם μ אשר מייצגת את התוחלת, מרכז הפעמון, ו- σ אשר מייצגת את סטיית התקן של ההתפלגות ובאה לידי ביטוי ברוחב הפעמון (מידת הפיזור). משתנה מקרי נורמאלי כזה נסמן ב

$$X \sim N(\mu,\sigma) \ .$$

באיור לעייל מוצגות מספר התפלגויות בעלות פרמטרים שונים. ניתן לראות כיצד שינוי של התוחלת וסטיית התקן משנות את מבנה ההתפלגות. צפיפות ההתפלגות הנורמאלית בעלת מספר תכונות חשובות. בראש ובראשונה הצפיפות היא תמיד חיובית ושואפת לאפס בגבולות כאשר $x \to \pm\infty$.

בנוסף, כמתואר באיור להלן, התפלגות נורמאלית היא סימטרית סביב μ (התוחלת), והיא קמורה בקטע $[\mu-\sigma,\mu+\sigma]$ ואחרת קעורה. מן הסתם, ובדומה לכל פונקצית צפיפות, השטח התחום בין גרף הפונקציה לציר ה x שווה ל-1 כנדרש מתנאי הנרמול.

12.2 חוק. (תוחלת של התפלגות נורמאלית) התוחלת של התפלגות נורמאלית $X\sim N(\mu,\sigma)$, בעל צפיפות $f_X(x)=n(x,u,\sigma):=\frac{1}{\sqrt{2\pi}\sigma}e^{-(x-\mu)^2/(2\sigma^2)}$, היא

$$E[X]=\mu\ .$$

הוכחה.

$$E[X]=\int_{-\infty}^{\infty}xf_X(x)\ dx=\int_{-\infty}^{\infty}dx\ x\ \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

מחליפים את המשתנה x ב

$$y:=\frac{(x-\mu)}{\sqrt{2\sigma^2}}\qquad\Leftrightarrow\qquad x=\sqrt{2\sigma^2}y+\mu$$

כך ש $y'(x)=\dfrac{1}{\sqrt{2\sigma^2}}$ ומקבלים

$$\begin{aligned}E[X]=&\frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty}dx\,y'(x)\ \left(\sqrt{2\sigma^2}y+\mu\right)\ e^{-y^2}\\=&\frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty}dy\ \left(\sqrt{2\sigma^2}y+\mu\right)\ e^{-y^2}\\=&\frac{\sqrt{2\sigma^2}}{\sqrt{\pi}}\int_{-\infty}^{\infty}dy\ ye^{-y^2}+\frac{\mu}{\sqrt{\pi}}\int_{-\infty}^{\infty}dy\ e^{-y^2}\end{aligned}$$

ישר מהתוצאות

$$\int_{-\infty}^{\infty}dt\ e^{-t^2}=\sqrt{\pi}\ ,\qquad\int_{-\infty}^{\infty}dt\ t\ e^{-t^2}=0,$$

מקבלים

$$E[X]=\mu\ .$$

כנדרש. ■

12.3 חוק. (שונות של התפלגות נורמאלית) השונות של התפלגות נורמאלית $X\sim N(\mu,\sigma)$, בעל צפיפות $f_X(x)=n(x,u,\sigma):=\frac{1}{\sqrt{2\pi}\sigma}e^{-(x-\mu)^2/(2\sigma^2)}$, היא

$$V(X)=\sigma^2\ .$$

הוכחה.

$$E[X^2]=\int_{-\infty}^{\infty}x^2f_X(x)\ dx=\int_{-\infty}^{\infty}dx\ x^2\ \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

מחליפים את המשתנה x ב

$$y:=\frac{(x-\mu)}{\sqrt{2\sigma^2}}\qquad\Leftrightarrow\qquad x=\sqrt{2\sigma^2}y+\mu\qquad\Leftrightarrow\qquad x^2=2\sigma^2y^2+\mu^2+2\sqrt{2\sigma^2}y$$

כך ש $y'(x) = \frac{1}{\sqrt{2\sigma^2}}$ ומקבלים

$$\begin{aligned} E[X^2] =& \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} dx\, y'(x) \, \left(2\sigma^2 y^2 + 2\sqrt{2\sigma^2} y + \mu^2\right) \, e^{-y^2} \\ =& \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} dy \; 2\sigma^2 y^2 \; e^{-y^2} + \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} dy \; 2\sqrt{2\sigma^2} y \; e^{-y^2} + \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} dy \; \mu^2 \; e^{-y^2} \end{aligned}$$

ישר מהתוצאות

$$\int_{-\infty}^{\infty} dt \; e^{-t^2} = \sqrt{\pi} \; , \qquad \int_{-\infty}^{\infty} dt \; t \; e^{-t^2} = 0, \qquad \int_{-\infty}^{\infty} dt \; t^2 \; e^{-t^2} = \frac{\sqrt{\pi}}{2},$$

מקבלים

$$E[X^2] = \frac{1}{\sqrt{\pi}} \; 2\sigma^2 \frac{\sqrt{\pi}}{2} + \frac{1}{\sqrt{\pi}} \mu^2 \sqrt{\pi} = \sigma^2 + \mu^2 \; .$$

לכן

$$V[X] = E[X^2] - E[X]^2 = \sigma^2 + \mu^2 - \mu^2 = \sigma^2$$

כנדרש. ■

שטח התחום בין גרף הפונקציית הצפיפות הנורמאלית לציר ה x

12.4 חוק. (שטח התחום בין גרף הפונקציית הצפיפות לציר ה x)

הגרף של פונקציית הצפיפות כלשהי $f_X(x)$ היא כך שהשטח התחום בין גרף הפונקציית הצפיפות לציר ה x בין הערכים $x = a$ ו- $x = b$ ($a < b$) הוא שווה להסתברות כי המשתנה מקרי X נמצא בין הערכים a ו- b:

$$P(a < X < b) = \int_a^b dx \;\, f_X(x) \; .$$

לכן, עבר הגרף של הצפיפות הנורמאלית (עקומת פעמון) להלן, ההסתברות כי $a < X < b$ הוא שווה ל

$$P(a < X < b) = \int_a^b dx \;\, n(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \int_a^b dx \;\, e^{-(x-\mu)^2/(2\sigma^2)} \; .$$

השטח זו מוצג ע"י השטח של האזור המוצל באיור.

12.5 הגדרה. (מ"מ נורמאלי סטנדרדי)
ההתפלגות של משתנה מקרי נורמאלי בעל תוחלת $\mu = 0$ ושונות $\sigma^2 = 1$ נקרא משתנה מקרי נורמאלי סטנדרדי.

12.6 חוק. (צימצום מ"מ נורמאלי לצורת נורמאלי סטנדרדי)
עבור משתנה מקרי נורמאלי כלשהו X, ניתן לבטא אותו במונחים של משתנה מקרי Z, בעל תוחלת $\mu = 0$, ושונות $\sigma^2 = 0$ ע"י היחס

$$Z = \frac{X - \mu}{\sigma} \; .$$

על כן, כאשר $X=x$ אזי $Z=\dfrac{x-\mu}{\sigma}$. לכן כאשר הערך של X נמצא בין הערכים $x=x_1$ ו- $x=x_2$ הערך של מ"מ Z נמצא בין הערכים $z_1=\dfrac{x_1-\mu}{\sigma}$ ו- $z_2=\dfrac{x_2-\mu}{\sigma}$. כתוצאה

$$\begin{aligned}P(a<X<b)=&\frac{1}{\sqrt{2\pi}\sigma}\int_a^b dx\ \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\\=&\frac{1}{\sqrt{2\pi}}\int_{z_1}^{z_2} dz\ e^{-z^2/2}\\=&\int_{z_1}^{z_2} dz\ n(z,0,1)=P(z_1<Z<z_2)\ .\end{aligned}$$

במילים, ההתפלגויות $n(x,\mu,\sigma)$ ו- $n(z,0,1)$ הם מתוארים באיור להלן. מאחר שיש לכל הערכים של X הנמצאים בין x_1 ו- x_2 ערכים המתאימים של Z הנמצאים בין z_1 ו- z_2, אזי השטח התחום בין גרף של הצפיפות של X בין x_1 ו- x_2 הוא שווה ל השטח התחום בין גרף של הצפיפות של Z בין z_1 ו- z_2.

12.7 חוק. (פונקציית ההתפלגות המצטברת של מ"מ נורמאלי) הפונקציית ההתפלגות המצטברת של משתנה מקרי המתפלג נורמאלי, $F_X(x)$, נתון ע"י הנוסחאה

$$F_X(x)=\int_{-\infty}^x dt\ n(t,\mu,\sigma)=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^x dt\exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^z dt\ e^{-t^2/2}$$

או

$$F_X(x)=\Phi(z)$$

כאשר $z=\dfrac{x-\mu}{\sigma}$ ו-

$$\Phi(z):=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^z dt\ e^{-t^2/2}\ .$$

אפשר לבטא את הפה"מ של מ"מ נורמאלי בצורה

$$\Phi(z)=\frac{1}{2}\left(1+\text{erf}\left(\frac{z}{\sqrt{2}}\right)\right)$$

כאשר הפונקצייה $\text{erf}(z)$ מוגדרת להיות האינטגרל $\text{erf}(z):=\dfrac{2}{\sqrt{\pi}}\int_0^z dt\ e^{-t^2}$.

הוכחה. שים לב ש

$$\begin{aligned}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^z dt\ e^{-t^2/2}=&\frac{1}{\sqrt{2\pi}}\int_0^z dt\ e^{-t^2/2}+\frac{1}{\sqrt{2\pi}}\int_{-\infty}^0 dt\ e^{-t^2/2}\\=&\frac{1}{\sqrt{\pi}}\int_0^{z/\sqrt{2}} dt\ e^{-t^2}+\frac{1}{\sqrt{\pi}}\int_{-\infty}^0 dt\ e^{-t^2}\\=&\frac{1}{2}\left(\text{erf}\left(\frac{z}{\sqrt{2}}\right)+1\right)\end{aligned}$$

לכן

$$F_X(x)=\Phi(z)=\frac{1}{2}\left(1+\text{erf}\left(\frac{z}{\sqrt{2}}\right)\right)$$

■

12.8 חוק. () נובע מחוק 12.7 כי נתון w כך ש ל- $w\%$ של אנשים מתוך אוכלוסיה נתונה יש $X<x_1$, אפשר לחשב את הערך x_1 ע"י

$$x_1=\sqrt{2}\ \text{erf}^{-1}(2w-1)\ .$$

שימוש בטבלת ההתפלגות הנורמלית הסטנדרטית

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.400000	0.000337	0.000325	0.000313	0.000302	0.000291	0.000280	0.000270	0.000260	0.000251	0.000242
−3.300000	0.000483	0.000466	0.000450	0.000434	0.000419	0.000404	0.000390	0.000376	0.000362	0.000349
−3.200000	0.000687	0.000664	0.000641	0.000619	0.000598	0.000577	0.000557	0.000538	0.000519	0.000501
−3.100000	0.000968	0.000935	0.000904	0.000874	0.000845	0.000816	0.000789	0.000762	0.000736	0.000711
−3.000000	0.001350	0.001306	0.001264	0.001223	0.001183	0.001144	0.001107	0.001070	0.001035	0.001001
−2.900000	0.001866	0.001807	0.001750	0.001695	0.001641	0.001589	0.001538	0.001489	0.001441	0.001395
−2.800000	0.002555	0.002477	0.002401	0.002327	0.002256	0.002186	0.002118	0.002052	0.001988	0.001926
−2.700000	0.003467	0.003364	0.003264	0.003167	0.003072	0.002980	0.002890	0.002803	0.002718	0.002635
−2.600000	0.004661	0.004527	0.004396	0.004269	0.004145	0.004025	0.003907	0.003793	0.003681	0.003573
−2.500000	0.006210	0.006037	0.005868	0.005703	0.005543	0.005386	0.005234	0.005085	0.004940	0.004799
−2.400000	0.008198	0.007976	0.007760	0.007549	0.007344	0.007143	0.006947	0.006756	0.006569	0.006387
−2.300000	0.010724	0.010444	0.010170	0.009903	0.009642	0.009387	0.009137	0.008894	0.008656	0.008424
−2.200000	0.013903	0.013553	0.013209	0.012874	0.012545	0.012224	0.011911	0.011604	0.011304	0.011011
−2.100000	0.017864	0.017429	0.017003	0.016586	0.016177	0.015778	0.015386	0.015003	0.014629	0.014262
−2.000000	0.022750	0.022216	0.021692	0.021178	0.020675	0.020182	0.019699	0.019226	0.018763	0.018309
−1.900000	0.028717	0.028067	0.027429	0.026803	0.026190	0.025588	0.024998	0.024419	0.023852	0.023295
−1.800000	0.035930	0.035148	0.034380	0.033625	0.032884	0.032157	0.031443	0.030742	0.030054	0.029379
−1.700000	0.044565	0.043633	0.042716	0.041815	0.040930	0.040059	0.039204	0.038364	0.037538	0.036727
−1.600000	0.054799	0.053699	0.052616	0.051551	0.050503	0.049471	0.048457	0.047460	0.046479	0.045514
−1.500000	0.066807	0.065522	0.064255	0.063008	0.061780	0.060571	0.059380	0.058208	0.057053	0.055917
−1.400000	0.080757	0.079270	0.077804	0.076359	0.074934	0.073529	0.072145	0.070781	0.069437	0.068112
−1.300000	0.096800	0.095098	0.093418	0.091759	0.090123	0.088508	0.086915	0.085343	0.083793	0.082264
−1.200000	0.115070	0.113139	0.111232	0.109349	0.107488	0.105650	0.103835	0.102042	0.100273	0.098525
−1.100000	0.135666	0.133500	0.131357	0.129238	0.127143	0.125072	0.123024	0.121000	0.119000	0.117023
−1.000000	0.158655	0.156248	0.153864	0.151505	0.149170	0.146859	0.144572	0.142310	0.140071	0.137857
−0.900000	0.184060	0.181411	0.178786	0.176186	0.173609	0.171056	0.168528	0.166023	0.163543	0.161087
−0.800000	0.211855	0.208970	0.206108	0.203269	0.200454	0.197663	0.194895	0.192150	0.189430	0.186733
−0.700000	0.241964	0.238852	0.235762	0.232695	0.229650	0.226627	0.223627	0.220650	0.217695	0.214764
−0.600000	0.274253	0.270931	0.267629	0.264347	0.261086	0.257846	0.254627	0.251429	0.248252	0.245097
−0.500000	0.308538	0.305026	0.301532	0.298056	0.294599	0.291160	0.287740	0.284339	0.280957	0.277595
−0.400000	0.344578	0.340903	0.337243	0.333598	0.329969	0.326355	0.322758	0.319178	0.315614	0.312067
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−0.300000	0.382089	0.378280	0.374484	0.370700	0.366928	0.363169	0.359424	0.355691	0.351973	0.348268
−0.200000	0.420740	0.416834	0.412936	0.409046	0.405165	0.401294	0.397432	0.393580	0.389739	0.385908
−0.100000	0.460172	0.456205	0.452242	0.448283	0.444330	0.440382	0.436441	0.432505	0.428576	0.424655
0.000000	0.500000	0.503989	0.507978	0.511966	0.515953	0.519939	0.523922	0.527903	0.531881	0.535856
0.100000	0.539828	0.543795	0.547758	0.551717	0.555670	0.559618	0.563559	0.567495	0.571424	0.575345

0.200000	0.579260	0.583166	0.587064	0.590954	0.594835	0.598706	0.602568	0.606420	0.610261	0.614092
0.300000	0.617911	0.621720	0.625516	0.629300	0.633072	0.636831	0.640576	0.644309	0.648027	0.651732
0.400000	0.655422	0.659097	0.662757	0.666402	0.670031	0.673645	0.677242	0.680822	0.684386	0.687933
0.500000	0.691462	0.694974	0.698468	0.701944	0.705401	0.708840	0.712260	0.715661	0.719043	0.722405
0.600000	0.725747	0.729069	0.732371	0.735653	0.738914	0.742154	0.745373	0.748571	0.751748	0.754903
0.700000	0.758036	0.761148	0.764238	0.767305	0.770350	0.773373	0.776373	0.779350	0.782305	0.785236
0.800000	0.788145	0.791030	0.793892	0.796731	0.799546	0.802337	0.805105	0.807850	0.810570	0.813267
0.900000	0.815940	0.818589	0.821214	0.823814	0.826391	0.828944	0.831472	0.833977	0.836457	0.838913
1.000000	0.841345	0.843752	0.846136	0.848495	0.850830	0.853141	0.855428	0.857690	0.859929	0.862143
1.100000	0.864334	0.866500	0.868643	0.870762	0.872857	0.874928	0.876976	0.879000	0.881000	0.882977
1.200000	0.884930	0.886861	0.888768	0.890651	0.892512	0.894350	0.896165	0.897958	0.899727	0.901475
1.300000	0.903200	0.904902	0.906582	0.908241	0.909877	0.911492	0.913085	0.914657	0.916207	0.917736
1.400000	0.919243	0.920730	0.922196	0.923641	0.925066	0.926471	0.927855	0.929219	0.930563	0.931888
1.500000	0.933193	0.934478	0.935745	0.936992	0.938220	0.939429	0.940620	0.941792	0.942947	0.944083
1.600000	0.945201	0.946301	0.947384	0.948449	0.949497	0.950529	0.951543	0.952540	0.953521	0.954486
1.700000	0.955435	0.956367	0.957284	0.958185	0.959070	0.959941	0.960796	0.961636	0.962462	0.963273
1.800000	0.964070	0.964852	0.965620	0.966375	0.967116	0.967843	0.968557	0.969258	0.969946	0.970621
1.900000	0.971283	0.971933	0.972571	0.973197	0.973810	0.974412	0.975002	0.975581	0.976148	0.976705
2.000000	0.977250	0.977784	0.978308	0.978822	0.979325	0.979818	0.980301	0.980774	0.981237	0.981691
2.100000	0.982136	0.982571	0.982997	0.983414	0.983823	0.984222	0.984614	0.984997	0.985371	0.985738
2.200000	0.986097	0.986447	0.986791	0.987126	0.987455	0.987776	0.988089	0.988396	0.988696	0.988989
2.300000	0.989276	0.989556	0.989830	0.990097	0.990358	0.990613	0.990863	0.991106	0.991344	0.991576
2.400000	0.991802	0.992024	0.992240	0.992451	0.992656	0.992857	0.993053	0.993244	0.993431	0.993613
2.500000	0.993790	0.993963	0.994132	0.994297	0.994457	0.994614	0.994766	0.994915	0.995060	0.995201
2.600000	0.995339	0.995473	0.995604	0.995731	0.995855	0.995975	0.996093	0.996207	0.996319	0.996427
2.700000	0.996533	0.996636	0.996736	0.996833	0.996928	0.997020	0.997110	0.997197	0.997282	0.997365
2.800000	0.997445	0.997523	0.997599	0.997673	0.997744	0.997814	0.997882	0.997948	0.998012	0.998074
2.900000	0.998134	0.998193	0.998250	0.998305	0.998359	0.998411	0.998462	0.998511	0.998559	0.998605
3.000000	0.998650	0.998694	0.998736	0.998777	0.998817	0.998856	0.998893	0.998930	0.998965	0.998999
3.100000	0.999032	0.999065	0.999096	0.999126	0.999155	0.999184	0.999211	0.999238	0.999264	0.999289
3.200000	0.999313	0.999336	0.999359	0.999381	0.999402	0.999423	0.999443	0.999462	0.999481	0.999499
3.300000	0.999517	0.999534	0.999550	0.999566	0.999581	0.999596	0.999610	0.999624	0.999638	0.999651
3.400000	0.999663	0.999675	0.999687	0.999698	0.999709	0.999720	0.999730	0.999740	0.999749	0.999758

12.9 דוגמא. נתון התפלגות נורמאלי סטנדרדי, מצאו את השטח התחום בין הגרף לציר ה x

1. לצד הימין של $z = 1.84$
2. בין הערכים $z = -1.97$ ו- $z = 0.86$.

פיתרון. .

1. השטח לצד הימין של $z = 1.84$ שווה ל- 1 פחות השטח לצד שמאול של $z = 1.84$, קרי

$$1 - 0.9671 = 0.0329.$$

2. השטח בין הערכים $z = -1.97$ ו- $z = 0.86$ שווה לשטח לצד שמאול של $z = 0.86$ פחות השטח לצד שמאול של $z = -1.97$, קרי

$$0.8051 - 0.0244 = 0.7807 .$$

■

12.10 דוגמא. נתון מ"מ X בעל התפלגות נורמאלי עם הפרמטרים

$$\mu = 50 , \quad \sigma = 10 ,$$

מצאו את ההסתברות אשר ל- X יש ערך בין 45 לבין 62.

פיתרון. הערכים של הz המתאים ל- $x_1 = 45$ ו- $x_2 = 62$ הם

$$z_1 = \frac{45-50}{10} = -0.5 , \qquad z_2 = \frac{62-50}{10} = 1.2 .$$

לכן

$$\begin{aligned} P(45 < X < 62) =& P(-0.5 < Z < 1.2) \\ =& P(Z < 1.2) - P(Z < -0.5) \\ =& 0.8849 - 0.3085 \\ =& 0.5764 . \end{aligned}$$

■

לעתים יש צורך למצוא את הערך של z המתאים להסתברות נתון אשר נמצא בין הערכים בהטבלה. בהתרגילים לעייל מצאנו ערך של z עם ערך של x נתון. עכשיו עושים החישוב ההפוך: נתון ערך של שטח שלתחום של הגרף, או נתון ערך של ההסתברות , מחפשים את הערך של z ולאחר מכן מחפשים את הערך של x על ידי הנוסחאה

$$z = \frac{x-\mu}{\sigma} \qquad \Leftrightarrow \qquad x = \sigma \; z + \mu .$$

12.11 דוגמא. נתון התפלגות נורמאלי עם הפרמטרים

$$\mu = 40, \quad \sigma = 6,$$

מצאו את הערך של x אשר יש לו את ההסתברות של

1. 45% של השטח לצד שמאול,
2. 14% של השטח לצד ימין.

פיתרון. 1. מחפשים ערך של z כך ש 0.45 של השטח כולו נמצא לצד שמאול שלו. מהטבלה נמצא ש

$$P(Z < -0.13) = 0.45 ,$$

לכן הz הנדרש הוא -0.13 ולכן

$$x = 6(-0.13) + 40 = 39.22 .$$

2. כעת מחפשים ערך של z כך ש 0.14 של השטח כולו נמצא לצד ימין שלו, ולכן 0.86 של השטח כולו נמצא לצד שמאול שלו. מהטבלה נמצא ש

$$P(Z < 1.08) = 0.86$$

ולכן הערך הנדרש של z הוא 1.08. על כן

$$x = 6(1.08) + 40 = 46.48 \ .$$

■

12.12 דוגמא. יש דגם של מצבר אשר יש לו אורך חיים ממוצע של 3 שנים עם סטיית התקן של 0.5 שנים. על בסיס שאורך חייפ של המצבר מתפלג נורמאלי, חפשו את ההסתברות אשר המצבר ישרוד לתקופת זמן פחות מ 2.3 שנים.

פיתרון. לחשב את $P(X < 2.3)$, יש צורך למצוא את השטח התחום של הגרף לצד שמאול של הערך $X = 2.3$. ניתן לחשב את זה ע"י לחשב את השטח לצד שמאול של הערך של הz המתאים:

$$z = \frac{2.3 - 3}{0.5} = -1.4 \ .$$

מהטבלה נמצא ש

$$P(X < 2.3) = P(Z < -1.4) = 0.0808 \ .$$

■

13 תרגילים על התפלגות נורמלית 10-8

הקירוב הנורמלי להתפלגות הבינומית

אנו זוכרים כי מ"מ בדיד המתפלג בינומיאלי, כלומר $X \sim \mathrm{Bin}(n,p)$, יש לה פונקציית התפלגות

$$f_X(k) = b(x,n,p) := \binom{n}{p} p^k q^{n-k} \ .$$

ההתפלגות הנורמאלי בעל תוחלת $\mu = np$ ו- $\sigma^2 = np(1-p)$ היא קירוב לההתפלגות הבינומיאלי לא רק כאשר n גדול ו- p לא קרוב מדיי ל- 0 או 1, אלא היא גם קירוב טובה כאשר n קטן ו- p קרוב ל $\frac{1}{2}$. כדי להמחיש את הקירוב של ההתפלגות המורמאלי לההתפלגות הבינומיאלי, ההיסטוגרמה של ההתפלגות הבינומיאלי $b(x,15,0.4)$ והגרף (עקומת פעמון) של ההתפלגות הנורמאלי מצויירים ביחד באיור להלן, כאשר יש לההתפלגות הנורמאלי תוחלת ושונות כך ש תוחלת

$$\mu = np = 15(0.4), \qquad \sigma^2 = npq = 15(0.4)(0.6) = 3.6 \ .$$

ההסתברות המדוייקת כי המ"מ בדיד X מקבל ערך נתון x היא שווה לשטח של המלבן אשר בסיסו מרוכז על x. למשל, ההסתברות ש X מקבל הערך של 4, היא שווה לשטח של המלבן אשר בסיסו מרוכז על $x = 4$:

$$P(X=4) = b(4,15,0.4) = \binom{15}{4} (0.4)^4 (0.6)^{15-4} = 0.1268 \ .$$

זו גם שווה בערך לשטח התחום של הגרף בין $x_1 = 3.5$ ו- $x_2 = 4.5$. במונחים של הערכים המתאימים של הz:

$$z_1 = \frac{3.5-6}{1.897} = -1.32 \ , \qquad z_2 = \frac{4.5-6}{1.897} = -0.79 \ ,$$

נמצא את השטח זו להיות

$$P(-1.32 < Z < -0.79) = P(Z < -0.79) - P(Z < -1.32) = 0.214764 - 0.093418 = 0.121346 \ ,$$

והמספר זו כמעט מסכים לגמרי עם הערך לעייל.

הקריוב הנורמאלי שימושי בלחשב סכום של הסתברויות של מ"מ בינומיאלי כאשר n הוא גדול. לדוגמה, נתון $X \sim \mathrm{Bin}(n,p)$ כאשר $n = 15$ ו- $p = 0.4$, מהי ההסתברות $P(7 < X < 9)$? התשובה היא

$$\sum_{k=7}^{9} \binom{n}{k} p^k (1-p)^{n-k} = 0.356354 \ .$$

אבל קשה לחשב סכום ארוך כזה. במקום נשתמש בקירוב של מ"מ נורמאלי בעל $\mu = np = 6$ ו- $\sigma^2 = npq = 3.6$: עבור $x_1 = 6.5$ ו- $x_2 = 9.5$, ה z_1 ו- z_2 המתאימים הם

$$z_1 = \frac{x_1 - \mu}{\sigma} = 0.2635, \qquad z_2 = \frac{x_2 - \mu}{\sigma} = 1.84466,$$

נמצא ש

$$P(z_1 < Z < z_2) = P(Z < 1.84466) - P(Z < 0.2635) = 0.9678 - \quad 0.6026 \quad = 0.3652.$$

פורמאלית:

13.1 חוק. (קירוב הנורמלי להתפלגות הבינומית) אם X הוא משתנה מקרי בדיד ומתפלג בינומיאלי, ובעל תוחלת $\mu = np$ ושונות $\sigma^2 = npq$, הגבול של ההתפלגות של

$$Z = \frac{X - np}{\sqrt{npq}}$$

כאשר $n \to \infty$ שואף להתפלגות הנורמאלי הסטנדרדי $n(z,0,1)$.

14 רווח $TZ_1 5 - 8$

סיכום של ההתפלגות נורמית וההתפלגות הנורמלית הסטנדרדית

14.1 הגדרה. (מ"מ רציף נורמאלי) משתנה מקרי X מתפלג נורמאלי מוגדר להיות כך שצפיפותו נתון ע"י הנוסחאה

$$n(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) ,$$

כאשר μ הוא התוחלת ו- σ הסטיית התקן. מסמנים מ"מ נורמאלי ב $X \sim N(\mu, \sigma)$.

14.2 הגדרה. (מ"מ רציף נורמאלי סטנדרדי) משתנה מקרי Z מתפלג נורמאלי סטנדרדי מוגדר להיות מ"מ נורמאלי בעל תוחלת $\mu = 0$ ו- סטיית התקן $\sigma = 1$, כלומר המ"מ נורמאלי עם צפיפות נתון ע"י

$$n(z, \mu = 0, \sigma = 0) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} .$$

14.3 הגדרה. (התפלגות המצטברת של מ"מ נורמאלי סטנדרדי) התפלגות המצטברת של מ"מ נורמאלי סטנדרדי Z מוגדרת להיות

$$\Phi(z) = \int_{-\infty}^{z} dt \ \frac{1}{\sqrt{2\pi}} e^{-t^2/2} = \frac{1}{2}\left[1 + \text{erf}\left(\frac{z}{\sqrt{2}}\right)\right] ,$$

כאשר

$$\text{erf}\,(z) := \frac{2}{\sqrt{\pi}} \int_0^z dt \ e^{-t^2} ,$$

(כלומר $\text{erf}\left(z/\sqrt{2}\right) = \frac{2}{\sqrt{\pi}} \int_0^{z/\sqrt{2}} dt \ e^{-t^2} = \frac{1}{\sqrt{2\pi}} \int_0^z dt \ e^{-t^2}$) והתכונה ש

$$\text{erf}(-z) = -\text{erf}(z)$$

נובע למסקנה

$$\Phi(-z) = 1 - \Phi(z) .$$

14.4 הגדרה. (הסתברות של מ"מ נורמאלי) ההסתברות שמ"מ נורמאלי X בעל תוחלת μ וסטיית התקן σ מקבל ערך פחות או שווה ל x_1 נתון ע"י

$$P(X \leq x_1) = \Phi(z_1)$$

כאשר

$$z_1 = \frac{x_1 - \mu}{\sigma} .$$

ההסתברות שמ"מ נורמאלי X בעל תוחלת μ וסטיית התקן σ מקבל ערים בין x_1 ו- x_2 נתון ע"י

$$P(x_1 \leq X \leq x_2) = \Phi(z_2) - \Phi(z_1)$$

כאשר

$$z_1 = \frac{x_1 - \mu}{\sigma} , \qquad z_2 = \frac{x_2 - \mu}{\sigma} .$$

14.5 חוק. (קירוב נורמלי להתפלגות בינומית) יהי $X \sim \text{Bin}\,(n, p)$. אז עבור $n \geq 30$,

$$X \sim N\,(np, nq) .$$

תיקון רציפות:

$$P(a \leq X \leq b)$$

כמשתנה בדיד שווה ל

$$P(a - 0.5 < X < b + 0.5)$$

כמשתנה רציף.

משפט הגבול המרכזי

14.6 חוק. (משפט הגבול המרכזי) יהי X משתנה מקרי בעל תוחלת μ וסטיית התקן σ, ויהי

$$x_1, \ldots, x_n$$

מדגם מקרי מתוך X. אזי,כאשר $n \geq 30$ או לכל n במקרה ש X מתפלג נורמלית, מתקיים

$$\sum_{i=1}^{n} X_i \sim N\left(n\mu, n\sigma^2\right) ,$$

או במילים אחרות, כאשר $\bar{X} := \frac{1}{n} \sum_{i=1}^{n} X_i$ הוא התוחלת של המדגם,

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \ .$$

14.7 מסקנה. (משפט הגבול המרכזי) יהי $\bar{X}$ התוחלת של מדגם מקרי של אורך n מתוך אוכלוסיה בעל תוחלת μ ושונות σ^2. אזי,כאשר $n \geq 30$, ההתפלגות של המשתנה

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

היא ההתפלגות הנורמלית הסטנדרדית $n(z, 0, 1)$, קרי

$$Z \sim N(0, 1) \ .$$

רווחי סמך לתוחלת של מדגם מקרי: שונות ידועה

נניח שיש אוכלוסיה בעל תוחלת μ אינה ידועה ושונות σ^2 ידועה. עבור מדגם X כלשהו מתוך האוכלוסיה זו, לפי המשפט הגבול המרכזי, ההתפלגות של $\bar{X}$ מתפלג בקירוב נורמלי עם תוחלת $\mu_{\bar{X}} = \mu$ וסטיית התקן $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$.
השאלה היא:

> נתון $\bar{X}$, מהו הטווח
>
> $$a \leq \mu \leq b$$
>
> כך כי יש הסתברות $(1 - \alpha)$ ל μ כן להיות נמצא בטווח זו, קרי
>
> $$P(a \leq \mu \leq b) = 1 - \alpha$$
>
> והסתברות α ל μ לא להיות נמצא בטווח זו, כלומר
>
> $$P(\mu \notin [a, b]) = \alpha \ .$$

לדוגמה, מהו הטווח של μ כך שיש הסתברות של 0.95 (או 95%) להיות נמצא בו? הנה $1 - \alpha = 0.95$ אזי $\alpha = 0.05$. אנו מתבקשים למצוא a ו- b כך ש $P(a \leq \mu \leq b) = 0.95$. נגלה להלן [עיין משוואה (2*)] שהערכים הנדרשים הם

$$a = \bar{X} - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \ , \qquad b = \bar{X} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \ ,$$

כאשר $\bar{X}$ הוא התוחלת המחושב של המדגם מקרי X, σ הסטיית התקן הידוע של האוכלוסיה כולה שממנה המדגם נלקח, n האורך של המדגם ו- $z_{1-\alpha/2}$ הוא הערך של המשתנה נורמלי סטנדרדי z המתאים ל $\bar{X}$ [עיין נוסחאה (1*)

להלן], מוגדר כך שהשטח התחום בטווח $[-z_{1-\alpha/2}, z_{1-\alpha/2}]$ של הגרף [עיין אייור להלן] שווה ל $(1-\alpha)$ והשטח בצד ימין של $z_{1-\alpha/2}$ שווה ל $\alpha/2$.

הטווח $[a,b] = [\ \bar{X} - z_{1-\alpha/2}\sigma/\sqrt{n}\ ,\ \bar{X} + z_{1-\alpha/2}\sigma/\sqrt{n}\ ]$ נקרא **רווח סמך** וההסתברות $1-\alpha$ נקרא **רמת מובהקות** .

כדי למצוא את הטווח בשאלה, אנחנו זוכרים כי השטח התחום בגרף של משתנה מקרי כלשהו הוא שווה דווקא להסתברות כי המ"מ נמצא בטווח זו. ז"א השטח התחום של הגרף של $\bar{X}$ בטווח $[a,b]$ הוא שווה ל $P(a \le \bar{X} \le b)$. אבל בשל העובדה ש $\bar{X}$ הוא מתפלג נורמלי (לפי המשפט הגבול המרכזי) אזי ניתן להגדיר משתנה מקרי נורמלי סטנדרדי המתאים, והשטח זו יהיה שווה לשטח התחום של הגרף של Z בטווח המתאים.

נגדיר המשתנה Z להיות

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\ . \tag{*1}$$

יהי

$$z_{1-\alpha/2}$$

הערך של Z אשר עבורו השטח התחום בטווח $[-z_{1-\alpha/2}, z_{1-\alpha/2}]$ של הגרף [עיין אייור] שווה ל $(1-\alpha)$ והשטח בצד ימין שלו הוא $\alpha/2$ כמתואר באייור להלן.

כמו כן הטווח הנדרש של μ כדי לתת רמת מובהקות $1-\alpha$ (הסתברות $(1-\alpha)$ ל- μ לפול בטווח זו) נמצא ע"י למצוא הטווח המתאים של Z כך שהשטח התחום שווה ל- $(1-\alpha)$. כמו כן לפי הגרף נמצא ש

$$P(-z_{1-\alpha/2} < Z < z_{1-\alpha/2}) = 1-\alpha\ ,$$

על כן

$$P\left(-z_{1-\alpha/2} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < z_{1-\alpha/2}\right) = 1-\alpha\ .$$

מכפילים אגף הימין ואגף השמאול ב $\sigma/\sqrt{n}$ ולוקחים $\bar{X}$ מכל אגף ואגף ומקבלים כי

$$P\left(\bar{X} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha\ .$$

במילים, לוקחים מדגם מקרי של אורך n מתוך אוכלוסיה בעל שונות σ^2 ידועה. רווח סמך של $100(1-\alpha)\%$ לתוחלת μ נתון ע"י

$$\bar{X} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \tag{*2}$$

כאשר $z_{1-\alpha/2}$ הוא הערך של Z אשר עבורו יש שטח של $\alpha/2$ בצד הימין שלו.
פורמלית:

14.8 חוק. (רווח סמך מדגם מקרי σ^2 ידועה)

אם $\bar{x}$ הוא התוחלת של מדגם מקרי של אורך n מתוך אוכלוסיה בעל שונות σ^2 ידועה, רווח סמך של $100(1-\alpha)\%$ להתוחלת μ נתון ע"י

$$\bar{x} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$$

כאשר $z_{1-\alpha/2}$ הוא הערך של z אשר עבורו יש שטח של $\alpha/2$ בצד הימין שלו.

14.9 דוגמא. (רווח סמך מדגם מקרי σ^2 ידועה)

הריכוז הממוצע של חמצן ממדגם של מדידות הנלקחות מ 36 מקומות שונים בנהר הוא 2.6 [gr/mm]. מהו הרווח סמך של הרמות המבהוקות של 95% ו 99% להתוחלת של הריכוז חמצן בהנהר בשאלה. יש להניח שהסטיית התקן של האוכלוסיה הוא 0.3 [gr/mm].

פיתרון.
הממוצע של המדגם מקרי הוא

$$\bar{x} = 2.6 ,$$

ו-

$$1-\alpha = 0.95 \qquad \Leftrightarrow \qquad \alpha = 0.05 .$$

הערך של z אשר עבורו השטח בצד הימין שלו הוא $\alpha/2 = 0.025$, הוא

$$z_{1-\alpha/2} = z_{0.975} = 1.96 .$$

מהטבלה. לכן הרווח סמך של 95%, לפי נוסחאה (2*), הוא

$$2.6 - (1.96)\left(\frac{0.3}{\sqrt{36}}\right) < \mu < 2.6 + (1.96)\left(\frac{0.3}{\sqrt{36}}\right) .$$

ניתן לצמצם זה ל

$$2.50 < \mu < 2.70 .$$

למצוא הרווח סמך של רמת מובהקות של 99%, שים לב ש

$$1-\alpha = 0.99 \qquad \Leftrightarrow \qquad \alpha = 0.01, \qquad \Leftrightarrow \qquad \alpha/2 = 0.005 , \qquad \Leftrightarrow \qquad 1-\alpha/2 = 0.995 .$$

יש לחפש את הערך של z כך שבצד הימין שלו יש שטח של $\alpha/2 = 0.005$ מהטבלה הערך הנדרש הוא $z_{1-\alpha/2} = z_{0.995} = 2.575$ ולכן הרווח סמך של רמת מובהקות של 99% הוא

$$2.6 - (2.575)\left(\frac{0.3}{\sqrt{36}}\right) < \mu < 2.6 + (2.575)\left(\frac{0.3}{\sqrt{36}}\right) ,$$

או

$$2.47 < \mu < 2.73 .$$

שימו לב יש צורך לרווח יותר ארוך כדי להשיג ערך של μ יותר מדוייק. ■

הרווח סמך נותן הדייוק של האומדן של μ. אם μ נמצא במרכז של הרווח, אז $\bar{x}$ מעריך את μ ללא שגיאה. רוב הזמן אבל, $\bar{x}$ לא יהיה שווה בדיוק ל μ, כך שיהיה שגיאה בין האומדן לבין הערך המדויק של μ. הגודל של השגיאה זו הוא שווה להערך מוחלט של ההפרש בין μ לבין $\bar{x}$, וניתן להיות $100(1-\alpha)$% בטוח כי ההפרש זו לא יעבור $z_{1-\alpha/2}\sigma/\sqrt{n}$. אפשר לראות את זה עם העזרה של האיור להלן.

14.10 מסקנה. (סמך באומדן של μ)
אם לוקחים $\bar{x}$ להיות אומדן של μ, אז יש רמת מובהקות של $100(1-\alpha)$% שהשגיאה לא יעבור $z_{1-\alpha/2}\sigma/\sqrt{n}$.

בדוגמה לעייל יש רמת מובהקות של 95% שההפרש בין התוחלת של המדגם $\bar{x} = 2.6$ ו μ לא יעבור $(1.96)(0.3)/\sqrt{36} = 0.13$. לעתים יש צורך לדעת את האורך הנדרש של המדגם כדי לוודע שהשגיאה באומדן של μ לא יעבור ערך נתון e. על ידי המסקנה 15.10 לעייל, יש צורך לבחור n כך ש

$$z_{1-\alpha/2}\sigma/\sqrt{n} = e .$$

פותרים את המשוואה זו כדי לקבל נוסחאה ל n:

14.11 מסקנה. (סמך באומדן של μ)
אם לוקחים $\bar{x}$ להיות אומדן של μ, אז יש רמת מובהקות של $100(1-\alpha)\%$ שהשגיאה לא יעבור e כאשר האורך של המדגם הוא

$$n = \left(\frac{z_{1-\alpha/2}\sigma}{e}\right)^2 \ .$$

14.12 דוגמא. מהו האורך הנדרש של המדגם להשיג רמת מובהקות של 95% שיש לאומדן של μ בדוגמה 15.9 שגיאה פחות מ 0.05?

פיתרון. הסטיית התקן הוא $\sigma = 0.3$. לכן לפי מסקנה 15.11:

$$n = \left(\frac{(1.96)(0.3)}{(0.05)}\right)^2 = 138.3 \ .$$

לכן אפשר להיות 95% בטוח שמדגם מקרי של אורך של $n = 139$ יתן אומדן $\bar{x}$ של μ עם שגיאה פחות מ 0.05. ■

טבלות של ערכים של התפלגויות

$\Phi(z)$	z
0.5000000	0.0000000
0.5500000	0.1256613
0.6000000	0.2533471
0.6500000	0.3853205
0.7000000	0.5244005
0.7500000	0.6744898
0.8000000	0.8416212
0.8500000	1.0364334
0.9000000	1.2815516
0.9100000	1.3407550
0.9200000	1.4050716
0.9300000	1.4757910
0.9400000	1.5547736
0.9500000	1.6448536
0.9600000	1.7506861
0.9700000	1.8807936
0.9800000	2.0537489
0.9900000	2.3263479
0.9950000	2.5758293
0.9990000	3.0902323
0.9995000	3.2905267
0.9999000	3.7190165
0.9999500	3.8905919
0.9999900	4.2648908
0.9999950	4.4171734
0.9999990	4.7534243
0.9999995	4.8916385
0.9999999	5.1993376

n	$t_{0.995}$	$t_{0.990}$	$t_{0.975}$	$t_{0.950}$	$t_{0.900}$	$t_{0.800}$	$t_{0.750}$	$t_{0.700}$	$t_{0.600}$	$t_{0.550}$
1.000	63.657	31.821	12.706	6.314	3.078	1.376	1.000	0.727	0.325	0.158
2.000	9.925	6.965	4.303	2.920	1.886	1.061	0.816	0.617	0.289	0.142
3.000	5.841	4.541	3.182	2.353	1.638	0.978	0.765	0.584	0.277	0.137
4.000	4.604	3.747	2.776	2.132	1.533	0.941	0.741	0.569	0.271	0.134
5.000	4.032	3.365	2.571	2.015	1.476	0.920	0.727	0.559	0.267	0.132
6.000	3.707	3.143	2.447	1.943	1.440	0.906	0.718	0.553	0.265	0.131
7.000	3.499	2.998	2.365	1.895	1.415	0.896	0.711	0.549	0.263	0.130
8.000	3.355	2.896	2.306	1.860	1.397	0.889	0.706	0.546	0.262	0.130
9.000	3.250	2.821	2.262	1.833	1.383	0.883	0.703	0.543	0.261	0.129
10.000	3.169	2.764	2.228	1.812	1.372	0.879	0.700	0.542	0.260	0.129
11.000	3.106	2.718	2.201	1.796	1.363	0.876	0.697	0.540	0.260	0.129
12.000	3.055	2.681	2.179	1.782	1.356	0.873	0.695	0.539	0.259	0.128
13.000	3.012	2.650	2.160	1.771	1.350	0.870	0.694	0.538	0.259	0.128
14.000	2.977	2.624	2.145	1.761	1.345	0.868	0.692	0.537	0.258	0.128
15.000	2.947	2.602	2.131	1.753	1.341	0.866	0.691	0.536	0.258	0.128
16.000	2.921	2.583	2.120	1.746	1.337	0.865	0.690	0.535	0.258	0.128
17.000	2.898	2.567	2.110	1.740	1.333	0.863	0.689	0.534	0.257	0.128
18.000	2.878	2.552	2.101	1.734	1.330	0.862	0.688	0.534	0.257	0.127
19.000	2.861	2.539	2.093	1.729	1.328	0.861	0.688	0.533	0.257	0.127
20.000	2.845	2.528	2.086	1.725	1.325	0.860	0.687	0.533	0.257	0.127
21.000	2.831	2.518	2.080	1.721	1.323	0.859	0.686	0.532	0.257	0.127
22.000	2.819	2.508	2.074	1.717	1.321	0.858	0.686	0.532	0.256	0.127
23.000	2.807	2.500	2.069	1.714	1.319	0.858	0.685	0.532	0.256	0.127
24.000	2.797	2.492	2.064	1.711	1.318	0.857	0.685	0.531	0.256	0.127
25.000	2.787	2.485	2.060	1.708	1.316	0.856	0.684	0.531	0.256	0.127
26.000	2.779	2.479	2.056	1.706	1.315	0.856	0.684	0.531	0.256	0.127
27.000	2.771	2.473	2.052	1.703	1.314	0.855	0.684	0.531	0.256	0.127
28.000	2.763	2.467	2.048	1.701	1.313	0.855	0.683	0.530	0.256	0.127
29.000	2.756	2.462	2.045	1.699	1.311	0.854	0.683	0.530	0.256	0.127
30.000	2.750	2.457	2.042	1.697	1.310	0.854	0.683	0.530	0.256	0.127
40.000	2.704	2.423	2.021	1.684	1.303	0.851	0.681	0.529	0.255	0.126
60.000	2.660	2.390	2.000	1.671	1.296	0.848	0.679	0.527	0.254	0.126
120.000	2.617	2.358	1.980	1.658	1.289	0.845	0.677	0.526	0.254	0.126

העשרה: הוכחה של המשפט הגבול מרכזי

14.13 חוק. (משפט הגבול המרכזי) יהי X משתנה מקרי בעל תוחלת μ וסטיית התקן σ, ויהי

$$x_1, \dots, x_n$$

מדגם מקרי מתוך X. אזי,כאשר $n \geq 30$ או לכל n במקרה ש X מתפלג נורמלית, מתקיים

$$\sum_{i=1}^{n} X_i \sim N\left(n\mu, n\sigma^2\right),$$

או במילים אחרות, כאשר $\bar{X} := \frac{1}{n}\sum_{i=1}^{n} X_i$ הוא התוחלת של המדגם,

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \ .$$

14.14 מסקנה. (משפט הגבול המרכזי) יהי $\bar{X}$ התוחלת של מדגם מקרי של אורך n מתוך אוכלוסיה בעל תוחלת μ ושונות σ^2. אזי,כאשר $n \geq 30$, ההתפלגות של המשתנה

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

היא ההתפלגות הנורמלית הסטנדרדית $n(z, 0, 1)$, קרי

$$Z \sim N(0, 1) \ .$$

הוכחה.

The central limit theorem (CLT) states that when independent random variables are added, their properly normalized sum tends toward a normal distribution (a bell curve) even if the original variables themselves are not normally distributed.

Let $\{X_1, \dots, X_n\}$ be a random sample of size n, that is, a sequence of independent and identically distributed (IID)[1] random variables drawn from a distribution of expected value μ and finite variance σ^2. The sample average

$$\bar{X}_n := \frac{X_1 + \cdots + X_n}{n},$$

by the law of large numbers[2] converges to the expected value μ as $n \to \infty$.

Assume $\{X_1, \dots, X_n\}$ are independent and identically distributed random variables, each with mean μ abd finite variance σ^2. The sum $X_1 + \cdots + X_n$ has mean $n\mu$ and variance $n\sigma^2$. Define the random variable

$$Z_n := \frac{X_1 + \cdots + X_n - n\mu}{\sqrt{n\sigma^2}} = \sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sqrt{n\sigma^2}}\right) = \sum_{i=1}^{n}\frac{1}{\sqrt{n}}Y_i, \qquad \text{(\#1)}$$

[1] A set of random variables is independent and identically distributed if each random variable has the same probability distribution as the others and all are mutually independent.

[2] The law of large numbers (LLN) posits that the average of the results obtained from a large number of trials should be close to the expected value, and will tend to become closer to the expected value as more trials are performed

where $Y_i := (X_i - \mu)/\sigma$, which has zero mean and unit variance: $\text{var}(Y) = E\left[(Y_i - \text{mean of } Y)^2\right] = E\left[(Y_i)^2\right] = \frac{1}{\sigma^2} E\left[(X_i - \mu)^2\right] = \sigma^2/\sigma^2 = 1$. The characteristic function of Z_n is given by

$$\varphi(t)_{Z_n} = \varphi(t)_{\sum_{j=1}^{n} n^{-1/2} Y_j} = E\left[\exp\left(it \sum_{j=1}^{n} n^{-1/2} Y_j\right)\right] = E\left[\prod_{j=1}^{n} \exp\left(itn^{-1/2} Y_j\right)\right]. \quad \textbf{(\#2)}$$

By assumption the Y_j are identically distributed, which means they each have the same expectation value, and it follows from **(#2)** that

$$\varphi(t)_{Z_n} = \prod_{j=1}^{n} E\left[\exp\left(itn^{-1/2} Y_j\right)\right] = \varphi\left(n^{-1/2} t\right)_{Y_1} \varphi\left(n^{-1/2} t\right)_{Y_2} \cdots \varphi\left(n^{-1/2} t\right)_{Y_n} = \left[\varphi\left(n^{-1/2} t\right)_{Y_1}\right]^n. \quad \textbf{(\#3)}$$

The characteristic function of Y_1 is, by Taylor's theorem, $E\left[e^{itn^{-1/2} Y_1}\right] = E[1] + E\left[in^{-1/2} Y_1\right] t + E\left[-\left(n^{-1/2} Y_1\right)^2\right] t^2/2 + O\left(t^3\right)$. Above it was established that Y_1 has zero expectation value ($E[Y_1] = 0$) and variance one ($E\left[(Y_1)^2\right] = 1$), which means that

$$\varphi\left(n^{-1/2} t\right)_{Y_1} = E\left[e^{itn^{-1/2} Y_1}\right] = 1 - \frac{t^2}{2n} + \frac{t^4}{4! n^2} + o\left(\frac{t^4}{n^2}\right), \quad \textbf{(\#4)}$$

where $o\left(t^4/n^2\right)$ means something that goes to zero more rapidly than t^4/n^2. Hence,

$$\begin{aligned} \varphi(t)_{Z_n} &= \left(\varphi\left(n^{-1/2} t\right)_{Y_1}\right)^n \\ &= \left(1 - \frac{t^2}{2n}\right)^n + n\left(1 - \frac{t^2}{2n}\right)^{n-1} \left(\frac{t^4}{4! n^2} + o\left(\frac{t^4}{n^2}\right)\right) + \cdots \end{aligned} .$$

The form of the exponential function as a limit is $e^x = \lim_{n\to\infty}(1 + x/n)^n$. It follows that

$$\lim_{n\to\infty} \varphi(t)_{Z_n} = e^{-t^2/2} \lim_{n\to\infty} \left(1 + n\left(1 - \frac{t^2}{2n}\right)^{-1} \left(\frac{t^4}{4! n^2} + o\left(\frac{t^4}{n^2}\right)\right) + \cdots\right) .$$

All of the higher order terms inside the brackets vanish in the limit $n \to \infty$. Therefore,

$$\lim_{n\to\infty} \varphi(t)_{Z_n} = e^{-t^2/2}. \quad \textbf{(\#5)}$$

The right hand side equals the characteristic function of a standard normal distribution $N(0,1)$ **(prove)**, which implies through Le□vy's continuity theorem **(state and prove)** that the distribution of Z_n will approach $N(0,1)$ as $n \to \infty$. Therefore, the sample average

$$\bar{X}_n = \frac{X_1 + \cdots + X_n}{n}$$

is such that

$$\frac{\sqrt{n}}{\sigma}\left(\bar{X}_n - \mu\right)$$

converges to the normal distribution $N(0,1)$, from which the central limit theorem follows.

■

$TZ_1 7-8$

סיכום של ההתפלגות נורמית וההתפלגות הנורמלית הסטנדרדית

15.1 הגדרה. (מ"מ רציף נורמאלי) משתנה מקרי X מתפלג נורמאלי מוגדר להיות כך שצפיפותו נתון ע"י הנוסחאה

$$n(x,\mu,\sigma)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right),$$

כאשר μ הוא התוחלת ו- σ הסטיית התקן. מסמנים מ"מ נורמאלי ב $X\sim N(\mu,\sigma)$.

15.2 הגדרה. (מ"מ רציף נורמאלי סטנדרדי) משתנה מקרי Z מתפלג נורמאלי סטנדרדי מוגדר להיות מ"מ נורמאלי בעל תוחלת $\mu=0$ ו- סטיית התקן $\sigma=1$, כלומר המ"מ נורמאלי עם צפיפות נתון ע"י

$$n(z,\mu=0,\sigma=0)=\frac{1}{\sqrt{2\pi}}e^{-z^2/2}\ .$$

15.3 הגדרה. (התפלגות המצטברת של מ"מ נורמאלי סטנדרדי) התפלגות המצטברת של מ"מ נורמאלי סטנדרדי Z מוגדרת להיות

$$\Phi(z)=\int_{-\infty}^{z}dt\ \frac{1}{\sqrt{2\pi}}e^{-t^2/2}=\frac{1}{2}\left[1+\text{erf}\left(\frac{z}{\sqrt{2}}\right)\right]\ ,$$

כאשר

$$\text{erf}\,(z):=\frac{2}{\sqrt{\pi}}\int_0^z dt\ e^{-t^2}\ ,$$

(כלומר $\text{erf}\left(z/\sqrt{2}\right)=\frac{2}{\sqrt{\pi}}\int_0^{z/\sqrt{2}}dt\ e^{-t^2}=\frac{1}{\sqrt{2\pi}}\int_0^z dt\ e^{-t^2}$) והתכונה ש

$$\text{erf}(-z)=-\text{erf}(z)$$

נובע למסקנה

$$\Phi(-z)=1-\Phi(z)\ .$$

15.4 הגדרה. (הסתברות של מ"מ נורמאלי) ההסתברות שמ"מ נורמאלי X בעל תוחלת μ וסטיית התקן σ מקבל ערך פחות או שווה ל x_1 נתון ע"י

$$P(X\leq x_1)=\Phi(z_1)$$

כאשר

$$z_1=\frac{x_1-\mu}{\sigma}\ .$$

ההסתברות שמ"מ נורמאלי X בעל תוחלת μ וסטיית התקן σ מקבל ערים בין x_1 ו- x_2 נתון ע"י

$$P(x_1\leq X\leq x_2)=\Phi(z_2)-\Phi(z_1)$$

כאשר

$$z_1=\frac{x_1-\mu}{\sigma}\ ,\qquad z_2=\frac{x_2-\mu}{\sigma}\ .$$

15.5 חוק. (קירוב נורמלי להתפלגות בינומית) יהי $X\sim\text{Bin}\,(n,p)$. אז עבור $n\geq 30$,

$$X\sim N\,(np,nq)\ .$$

תיקון רציפות:

$$P(a\leq X\leq b)$$

כמשתנה בדיד שווה ל

$$P(a-0.5<X<b+0.5)$$

כמשתנה רציף.

משפט הגבול המרכזי

15.6 חוק. (משפט הגבול המרכזי) יהי X משתנה מקרי בעל תוחלת μ וסטיית התקן σ, ויהי

$$x_1, \ldots, x_n$$

מדגם מקרי מתוך X. אזי,כאשר $n \geq 30$ או לכל n במקרה ש X מתפלג נורמלית, מתקיים

$$\sum_{i=1}^{n} X_i \sim N\left(n\mu, n\sigma^2\right),$$

או במילים אחרות, כאשר $\bar{X} := \frac{1}{n}\sum_{i=1}^{n} X_i$ הוא התוחלת של המדגם,

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) .$$

15.7 מסקנה. (משפט הגבול המרכזי) יהי $\bar{X}$ התוחלת של מדגם מקרי של אורך n מתוך אוכלוסיה בעל תוחלת μ ושונות σ^2. אזי,כאשר $n \geq 30$, ההתפלגות של המשתנה

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

היא ההתפלגות הנורמלית הסטנדרדית $n(z, 0, 1)$, קרי

$$Z \sim N(0,1) .$$

רווחי סמך לתוחלת של מדגם מקרי: שונות ידועה

נניח שיש אוכלוסיה בעל תוחלת μ אינה ידועה ושונות σ^2 ידועה. עבור מדגם X כלשהו מתוך האוכלוסיה זו, לפי המשפט הגבול המרכזי, ההתפלגות של $\bar{X}$ מתפלג בקירוב נורמלי עם תוחלת $\mu_{\bar{X}} = \mu$ וסטיית התקן $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$.
השאלה היא:

נתון $\bar{X}$, מהו הטווח

$$a \leq \mu \leq b$$

כך כי יש הסתברות $(1-\alpha)$ ל μ כן להיות נמצא בטווח זו, קרי

$$P(a \leq \mu \leq b) = 1 - \alpha$$

והסתברות α ל μ לא להיות נמצא בטווח זו, כלומר

$$P(\mu \notin [a,b]) = \alpha .$$

לדוגמה, מהו הטווח של μ כך שיש הסתברות של 0.95 (או 95%) להיות נמצא בו? הנה $1-\alpha = 0.95$ אזי $\alpha = 0.05$. אנו מתבקשים למצוא a ו- b כך ש $P(a \leq \mu \leq b) = 0.95$. נגלה להלן [עיין משוואה (2*)] שהערכים הנדרשים הם

$$a = \bar{X} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} , \qquad b = \bar{X} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} ,$$

כאשר $\bar{X}$ הוא התוחלת המחושב של המדגם מקרי X, σ הסטיית התקן הידוע של האוכלוסיה כולה שממנה המדגם נלקח, n האורך של המדגם ו- $z_{1-\alpha/2}$ הוא הערך של המשתנה נורמלי סטנדרדי z המתאים ל $\bar{X}$ [עיין נוסחאה (1*)

להלן], מוגדר כך שהשטח התחום בטווח $[-z_{1-\alpha/2}, z_{1-\alpha/2}]$ של הגרף [עיין אייור להלן] שווה ל $(1-\alpha)$ והשטח בצד ימין של $z_{1-\alpha/2}$ שווה ל $\alpha/2$.

הטווח $[a,b] = [\ \bar{X} - z_{1-\alpha/2}\sigma/\sqrt{n}\ ,\ \bar{X} + z_{1-\alpha/2}\sigma/\sqrt{n}\ ]$ נקרא **רווח סמך** וההסתברות $1-\alpha$ נקרא **רמת מובהקות** .

כדי למצוא את הטווח בשאלה, אנחנו זוכרים כי השטח התחום בגרף של משתנה מקרי כלשהו הוא שווה דווקא להסתברות כי המ"מ נמצא בטווח זו. ז"א השטח התחום של הגרף של $\bar{X}$ בטווח $[a,b]$ הוא שווה ל $P(a \le \bar{X} \le b)$. אבל בשל העובדה ש $\bar{X}$ הוא מתפלג נורמלי (לפי המשפט הגבול המרכזי) אזי ניתן להגדיר משתנה מקרי נורמלי סטנדרדי המתאים, והשטח זו יהיה שווה לשטח התחום של הגרף של Z בטווח המתאים.

נגדיר המשתנה Z להיות

$$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\ . \tag{*1}$$

יהי

$$z_{1-\alpha/2}$$

הערך של Z אשר עבורו השטח התחום בטווח $[-z_{1-\alpha/2}, z_{1-\alpha/2}]$ של הגרף [עיין אייור] שווה ל $(1-\alpha)$ והשטח בצד ימין שלו הוא $\alpha/2$ כמתואר באייור להלן.

כמו כן הטווח הנדרש של μ כדי לתת רמת מובהקות $1-\alpha$ (הסתברות $(1-\alpha)$ ל- μ לפול בטווח זו) נמצא ע"י למצוא הטווח המתאים של Z כך שהשטח התחום שווה ל- $(1-\alpha)$. כמו כן לפי הגרף נמצא ש

$$P(-z_{1-\alpha/2} < Z < z_{1-\alpha/2}) = 1-\alpha\ ,$$

על כן

$$P\left(-z_{1-\alpha/2} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < z_{1-\alpha/2}\right) = 1-\alpha\ .$$

מכפילים אגף הימין ואגף השמאול ב $\sigma/\sqrt{n}$ ולוקחים $\bar{X}$ מכל אגף ואגף ומקבלים כי

$$P\left(\bar{X} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha\ .$$

במילים, לוקחים מדגם מקרי של אורך n מתוך אוכלוסיה בעל שונות σ^2 ידועה. רווח סמך של $100(1-\alpha)\%$ לתוחלת μ נתון ע"י

$$\bar{X} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \tag{*2}$$

כאשר $z_{1-\alpha/2}$ הוא הערך של Z אשר עבורו יש שטח של $\alpha/2$ בצד הימין שלו.
פורמלית:

15.8 חוק. (רווח סמך מדגם מקרי σ^2 ידועה)

אם $\bar{x}$ הוא התוחלת של מדגם מקרי של אורך n מתוך אוכלוסיה בעל שונות σ^2 ידועה, רווח סמך של $100(1-\alpha)\%$ להתוחלת μ נתון ע"י

$$\bar{x} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$$

כאשר $z_{1-\alpha/2}$ הוא הערך של z אשר עבורו יש שטח של $\alpha/2$ בצד הימין שלו.

15.9 דוגמא. (רווח סמך מדגם מקרי σ^2 ידועה)

הריכוז הממוצע של חמצן ממדגם של מדידות הנלקחות מ 36 מקומות שונים בנהר הוא 2.6 [gr/mm]. מהו הרווח סמך של הרמות המבהוקות של 95% ו 99% להתוחלת של הריכוז חמצן בהנהר בשאלה. יש להניח שהסטיית התקן של האוכלוסיה הוא 0.3 [gr/mm].

פיתרון.
הממוצע של המדגם מקרי הוא

$$\bar{x} = 2.6 \ ,$$

ו-

$$1-\alpha = 0.95 \qquad \Leftrightarrow \qquad \alpha = 0.05 \ .$$

הערך של z אשר עבורו השטח בצד הימין שלו הוא $\alpha/2 = 0.025$, הוא

$$z_{1-\alpha/2} = z_{0.975} = 1.96 \ .$$

מהטבלה. לכן הרווח סמך של 95%, לפי נוסחאה (2*), הוא

$$2.6 - (1.96)\left(\frac{0.3}{\sqrt{36}}\right) < \mu < 2.6 + (1.96)\left(\frac{0.3}{\sqrt{36}}\right) \ .$$

ניתן לצמצם זה ל

$$2.50 < \mu < 2.70 \ .$$

למצוא הרווח סמך של רמת מובהקות של 99%, שים לב ש

$$1-\alpha = 0.99 \qquad \Leftrightarrow \qquad \alpha = 0.01, \qquad \Leftrightarrow \qquad \alpha/2 = 0.005 \ , \qquad \Leftrightarrow \qquad 1-\alpha/2 = 0.995 \ .$$

יש לחפש את הערך של z כך שבצד הימין שלו יש שטח של $\alpha/2 = 0.005$ מהטבלה הערך הנדרש הוא $z_{1-\alpha/2} = z_{0.995} = 2.575$ ולכן הרווח סמך של רמת מובהקות של 99% הוא

$$2.6 - (2.575)\left(\frac{0.3}{\sqrt{36}}\right) < \mu < 2.6 + (2.575)\left(\frac{0.3}{\sqrt{36}}\right) \ ,$$

או

$$2.47 < \mu < 2.73 \ .$$

שימו לב יש צורך לרווח יותר ארוך כדי להשיג ערך של μ יותר מדוייק. ■

הרווח סמך נותן הדייוק של האומדן של μ. אם μ נמצא במרכז של הרווח, אז $\bar{x}$ מעריך את μ ללא שגיאה. רוב הזמן אבל, $\bar{x}$ לא יהיה שווה בדיוק ל μ, כך שיהיה שגיאה בין האומדן לבין הערך המדויק של μ. הגודל של השגיאה זו הוא שווה להערך מוחלט של ההפרש בין μ לבין $\bar{x}$, וניתן להיות $100(1-\alpha)\%$ בטוח כי ההפרש זו לא יעבור $z_{1-\alpha/2}\sigma/\sqrt{n}$. אפשר לראות את זה עם העזרה של האייור להלן.

15.10 מסקנה. (סמך באומדן של μ)
אם לוקחים $\bar{x}$ להיות אומדן של μ, אז יש רמת מובהקות של $100(1-\alpha)\%$ שהשגיאה לא יעבור $z_{1-\alpha/2}\sigma/\sqrt{n}$.

בדוגמה לעייל יש רמת מובהקות של 95% שההפרש בין התוחלת של המדגם $\bar{x} = 2.6$ ו μ לא יעבור $(1.96)(0.3)/\sqrt{36} = 0.13$. לעתים יש צורך לדעת את האורך הנדרש של המדגם כדי לוודע שהשגיאה באומדן של μ לא יעבור ערך נתון e. על ידי המסקנה 15.10 לעייל, יש צורך לבחור n כך ש

$$z_{1-\alpha/2}\sigma/\sqrt{n} = e \ .$$

פותרים את המשוואה זו כדי לקבל נוסחאה ל n:

15.11 מסקנה. (סמך באומדן של μ)
אם לוקחים $\bar{x}$ להיות אומדן של μ, אז יש רמת מובהקות של $100(1-\alpha)\%$ שהשגיאה לא יעבור e כאשר האורך של המדגם הוא
$$n = \left(\frac{z_{1-\alpha/2}\sigma}{e}\right)^2 \ .$$

15.12 דוגמא. מהו האורך הנדרש של המדגם להשיג רמת מובהקות של 95% שיש לאומדן של μ בדוגמה 15.9 שגיאה פחות מ 0.05?

פיתרון. הסטיית התקן הוא $\sigma = 0.3$. לכן לפי מסקנה 15.11:
$$n = \left(\frac{(1.96)(0.3)}{(0.05)}\right)^2 = 138.3 \ .$$
לכן אפשר להיות 95% בטוח שמדגם מקרי של אורך של $n = 139$ יתן אומדן $\bar{x}$ של μ עם שגיאה פחות מ 0.05. ■

רווחי סמך לתוחלת של מדגם מקרי מתוך התפלות נורמלית: שונות אינה ידועה

לעתים יש צורך לאמוד התוחלת של אוכלוסיה נתונה כאשר השונות אינה ידועה. נתון מדגם מקרי X מתוך התפלגות נורמלית, יש להמשתנה מקרי
$$T = \frac{\bar{X}-\mu}{s/\sqrt{n}}$$
כאשר למשתנה T יש התפלגות t בעל $n-1$ דרגות החופש, כאשר s הוא הסטיית התקן של המדגם. במקרה זה כאשר σ אינו ידוע, ניתן להשתמש ב T כדי למצוא רווחי סמך של μ. השלבים הם אותם השלבים עבור σ ידוע לעייל, אלא במקום σ מציבים s ובמקום ההתפלגות הנורמלית הסטנדרדית יש את ההתפלגות t:
$$P(-t_{1-\alpha/2} < T < t_{1-\alpha/2}) = 1-\alpha \ ,$$
כאשר $t_{1-\alpha/2}$ הוא הערך של t בעל $n-1$ דרגות החופש, אשר בצד הימין שלו יש שטח של $\alpha/2$. מציבים בT ומקבלים
$$P\left(-t_{1-\alpha/2} < \frac{\bar{X}-\mu}{S/\sqrt{n}} < t_{1-\alpha/2}\right) = 1-\alpha \ ,$$
מכפילים אגף הימין ואגף השמאול ב $S/\sqrt{n}$ ולוקחים $\bar{X}$ מכל אגף ואגף ומקבלים כי
$$P\left(\bar{X} - t_{1-\alpha/2}\frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{1-\alpha/2}\frac{S}{\sqrt{n}}\right) = 1-\alpha \ .$$
במילים, עבור מדגם מקרי נתון של אורך n, מחשבים את התוחלת $\bar{x}$ וסטיית התקן s. אז מקבלים כי הרווח סמך של רמת מובהקות $100(1-\alpha)\%$ להתוחלת μ של האוכלוסיה נתון ע"י
$$\bar{x} - t_{1-\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{1-\alpha/2}\frac{s}{\sqrt{n}} \ .$$

פורמלית:

15.13 חוק. (רווח סמך מדגם מקרי: σ^2 אינה ידועה)
כאשר $\bar{x}$ ו- s הם התוחלת וסטיית התקן של מדגם מקרי הנלקח מתוך אוכלוסיה כלשהי, במבצ שהתוחלת μ וסטיית התקן σ של האוכלוסיה כולה אינם ידועים, הרווח סמך של רמת מובהקות $100(1-\alpha)\%$ להתוחלת μ נתון ע"י
$$\bar{x} - t_{1-\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{1-\alpha/2}\frac{s}{\sqrt{n}}$$
כאשר $t_{\alpha/2}$ הוא הערך של t אשר עבורו יש שטח של $\alpha/2$ בצד הימין שלו.

15.14 דוגמא. (רווח סמך מדגם מקרי σ^2 אינה ידועה)
התוכן של שבע מכולות הדומות של חומצה הם
9.8, 10.2, 10.4, 9.8, 10.0 10.2 ו- 9.6 ליטרים. חפשו רווח סמך של 95% להתוחלת של כל המכולות. יש להניח שהכמות בכל מכולה מתפלגת נורמלית.

פיתרון.
התוחלת וסטיית התקן של המדגם הם $\bar{x} = 10.0$ ו- $s = 0.283$. בדוגמה זו

$$1 - \alpha = 0.95, \qquad \Rightarrow \qquad \alpha = 0.05, \qquad \Rightarrow \quad 1 - \alpha/2 = 0.975 \ .$$

על ידי הטבלה, $t_{0.975} = 2.447$ עבור $n - 1 = 6$ דרגות החופש. לכן הרווח סמך של 95% להתוחלת μ הוא

$$10.0 - (2.447)\left(\frac{0.283}{\sqrt{7}}\right) < \mu < 10.0 + (2.447)\left(\frac{0.283}{\sqrt{7}}\right)$$

או

$$9.74 < \mu < 10.26 \ .$$

■

בדיקות השערות על התוחלת

15.15 חוק. (מבצ 1: רווח סמך ובדיקת השערות לתוחלת) באחד המקרים

א' σ ידוע ו $n \geq 30$,

ב' σ ידוע ו $n \gtrless 30$, ו- $\bar{X}$ מתפלג נורמאלית,

רווח סמך ברמת $(1-\alpha)100\%$ נתון ע"י

$$\bar{X} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \ .$$

15.16 חוק. (מבצ 1: בדיקת השערוצ ברמת מובהקות α)

$$\begin{cases} H_0: & \mu \leq \mu_0, \\ H_1: & \mu > \mu_0, \end{cases} \quad \bar{X} \geq \mu_0 + z_{1-\alpha}\frac{\sigma}{\sqrt{n}} \ \Rightarrow \ H_0 \text{ נדחית} \ , \quad \bar{X} < \mu_0 + z_{1-\alpha}\frac{\sigma}{\sqrt{n}} \ \Rightarrow \ H_0 \text{ מתקבלת} \ .$$

$$\begin{cases} H_0: & \mu \geq \mu_0, \\ H_1: & \mu < \mu_0, \end{cases} \quad \bar{X} \leq \mu_0 - z_{1-\alpha}\frac{\sigma}{\sqrt{n}} \ \Rightarrow \ H_0 \text{ נדחית} \ , \quad \bar{X} > \mu_0 - z_{1-\alpha}\frac{\sigma}{\sqrt{n}} \ \Rightarrow \ H_0 \text{ מתקבלת} \ .$$

$$\begin{cases} H_0: & \mu = \mu_0, \\ H_1: & \mu \neq \mu_0, \end{cases} \quad X \leq \mu_0 - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \text{ או } X \geq \mu_0 + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \quad \Rightarrow \ H_0 \ \text{ נדחית} \ ,$$

$$\mu_0 - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} < \bar{X} < \mu_0 + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \qquad \Rightarrow \ H_0 \text{ מתקבלת} \ .$$

15.17 חוק. (מבצ 2: רווח סמך ובדיקת השערות לתוחלת)
במקרה ש σ לא ידוע ו $n \geq 30$, ו- $\bar{X}$ מתפלג נורמאלית, רווח סמך ברמת $(1-\alpha)100\%$

$$\bar{X} - z_{1-\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{X} + z_{1-\alpha/2}\frac{s}{\sqrt{n}} \ .$$

15.18 חוק. (מבצ 2: בדיקת השערוץ ברמת מובהקות α)

$$\begin{cases} H_0: & \mu \le \mu_0, \\ H_1: & \mu > \mu_0, \end{cases} \quad \bar{X} \ge \mu_0 + z_{1-\alpha}\frac{s}{\sqrt{n}} \Rightarrow H_0 \text{ נדחית }, \quad \bar{X} < \mu_0 + z_{1-\alpha}\frac{s}{\sqrt{n}} \Rightarrow H_0 \text{ מתקבלת }.$$

$$\begin{cases} H_0: & \mu \ge \mu_0, \\ H_1: & \mu < \mu_0, \end{cases} \quad \bar{X} \le \mu_0 - z_{1-\alpha}\frac{s}{\sqrt{n}} \Rightarrow H_0 \text{ נדחית }, \quad \bar{X} > \mu_0 - z_{1-\alpha}\frac{s}{\sqrt{n}} \Rightarrow H_0 \text{ מתקבלת }.$$

$$\begin{cases} H_0: & \mu = \mu_0, \\ \\ H_1: & \mu \ne \mu_0, \end{cases} \quad X \le \mu_0 - z_{1-\alpha/2}\frac{s}{\sqrt{n}} \text{ או } X \ge \mu_0 + z_{1-\alpha/2}\frac{s}{\sqrt{n}}, \quad \Rightarrow H_0 \text{ נדחית },$$

$$\mu_0 - z_{1-\alpha/2}\frac{s}{\sqrt{n}} < \bar{X} < \mu_0 + z_{1-\alpha/2}\frac{s}{\sqrt{n}} \qquad \Rightarrow H_0 \text{ מתקבלת }.$$

15.19 חוק. (מבצ 3: רווח סמך ובדיקת השערות לתוחלת)

ב מקרה ש σ לא ידוע ו $n < 30$, ו- $\bar{X}$ מתפלג נורמאלית, רווח סמך ברמת $(1-\alpha)100\%$

$$\bar{X} - t^{n-1}_{1-\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{X} + t^{n-1}_{1-\alpha/2}\frac{s}{\sqrt{n}} .$$

15.20 חוק. (מבצ 3: בדיקת השערוץ ברמת מובהקות α)

$$\begin{cases} H_0: & \mu \le \mu_0, \\ H_1: & \mu > \mu_0, \end{cases} \quad \bar{X} \ge \mu_0 + {t^{n-1}}_{1-\alpha}\frac{s}{\sqrt{n}} \Rightarrow H_0 \text{ נדחית }, \quad \bar{X} < \mu_0 + {t^{n-1}}_{1-\alpha}\frac{s}{\sqrt{n}} \Rightarrow H_0 \text{ מתקבלת }.$$

$$\begin{cases} H_0: & \mu \ge \mu_0, \\ H_1: & \mu < \mu_0, \end{cases} \quad \bar{X} \le \mu_0 - {t^{n-1}}_{1-\alpha}\frac{s}{\sqrt{n}} \Rightarrow H_0 \text{ נדחית }, \quad \bar{X} > \mu_0 - {t^{n-1}}_{1-\alpha}\frac{s}{\sqrt{n}} \Rightarrow H_0 \text{ מתקבלת }.$$

$$\begin{cases} H_0: & \mu = \mu_0, \\ \\ H_1: & \mu \ne \mu_0, \end{cases} \quad X \le \mu_0 - {t^{n-1}}_{1-\alpha/2}\frac{s}{\sqrt{n}} \text{ או } X \ge \mu_0 + {t^{n-1}}_{1-\alpha/2}\frac{s}{\sqrt{n}}, \quad \Rightarrow H_0 \text{ נדחית },$$

$$\mu_0 - {t^{n-1}}_{1-\alpha/2}\frac{s}{\sqrt{n}} < \bar{X} < \mu_0 + {t^{n-1}}_{1-\alpha/2}\frac{s}{\sqrt{n}} \qquad \Rightarrow H_0 \text{ מתקבלת }.$$

15.21 דוגמא.

מדגם מקרי של 100 מיתות הרשומות בארצ"ה במהלך השנה שעברה מראה אורך חיים הממוצע של 71.8 שנים. מניחים שהסטיית התקן של האוכלוסיה כולה הוא 8.9 שנים. האם זאת אומרת שהתוחלת של ה חיים היום היא יותר מ 70 שנים? יש להשתמש בסמך של $\alpha = 0.05$.

פיתרון.

1. $H_0: \quad \mu = 70$ שנים.
2. $H_1: \quad \mu > 70$ שנים.
3. $\alpha = 0.05$.

4. האיזור הקבלה הוא $z > z_{1-\alpha}$ כאשר $1-\alpha = 1-0.05 = 0.95$, לכן $z > z_{0.95} = 1.645$, כאשר $z = \dfrac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$.
5. חישובים: $\bar{x} = 71.8$ שנים, $\sigma = 8.9$ שנים, ולכן

$$z = \frac{71.8-70}{8.9/\sqrt{100}} = 2.02 \ .$$

6. תוצאה: השערה H_0 נדחית ומסיקים כי האורך החיים הממוצע היום הוא יותר גדול מ 70 שנים.

■

טבלות של ערכים של התפלגויות

$\Phi(z)$	z
0.5000000	0.0000000
0.5500000	0.1256613
0.6000000	0.2533471
0.6500000	0.3853205
0.7000000	0.5244005
0.7500000	0.6744898
0.8000000	0.8416212
0.8500000	1.0364334
0.9000000	1.2815516
0.9100000	1.3407550
0.9200000	1.4050716
0.9300000	1.4757910
0.9400000	1.5547736
0.9500000	1.6448536
0.9600000	1.7506861
0.9700000	1.8807936
0.9800000	2.0537489
0.9900000	2.3263479
0.9950000	2.5758293
0.9990000	3.0902323
0.9995000	3.2905267
0.9999000	3.7190165
0.9999500	3.8905919
0.9999900	4.2648908
0.9999950	4.4171734
0.9999990	4.7534243
0.9999995	4.8916385
0.9999999	5.1993376

n	$t_{0.995}$	$t_{0.990}$	$t_{0.975}$	$t_{0.950}$	$t_{0.900}$	$t_{0.800}$	$t_{0.750}$	$t_{0.700}$	$t_{0.600}$	$t_{0.550}$
1.000	63.657	31.821	12.706	6.314	3.078	1.376	1.000	0.727	0.325	0.158
2.000	9.925	6.965	4.303	2.920	1.886	1.061	0.816	0.617	0.289	0.142
3.000	5.841	4.541	3.182	2.353	1.638	0.978	0.765	0.584	0.277	0.137
4.000	4.604	3.747	2.776	2.132	1.533	0.941	0.741	0.569	0.271	0.134
5.000	4.032	3.365	2.571	2.015	1.476	0.920	0.727	0.559	0.267	0.132
6.000	3.707	3.143	2.447	1.943	1.440	0.906	0.718	0.553	0.265	0.131
7.000	3.499	2.998	2.365	1.895	1.415	0.896	0.711	0.549	0.263	0.130
8.000	3.355	2.896	2.306	1.860	1.397	0.889	0.706	0.546	0.262	0.130
9.000	3.250	2.821	2.262	1.833	1.383	0.883	0.703	0.543	0.261	0.129
10.000	3.169	2.764	2.228	1.812	1.372	0.879	0.700	0.542	0.260	0.129
11.000	3.106	2.718	2.201	1.796	1.363	0.876	0.697	0.540	0.260	0.129
12.000	3.055	2.681	2.179	1.782	1.356	0.873	0.695	0.539	0.259	0.128
13.000	3.012	2.650	2.160	1.771	1.350	0.870	0.694	0.538	0.259	0.128
14.000	2.977	2.624	2.145	1.761	1.345	0.868	0.692	0.537	0.258	0.128
15.000	2.947	2.602	2.131	1.753	1.341	0.866	0.691	0.536	0.258	0.128
16.000	2.921	2.583	2.120	1.746	1.337	0.865	0.690	0.535	0.258	0.128
17.000	2.898	2.567	2.110	1.740	1.333	0.863	0.689	0.534	0.257	0.128
18.000	2.878	2.552	2.101	1.734	1.330	0.862	0.688	0.534	0.257	0.127
19.000	2.861	2.539	2.093	1.729	1.328	0.861	0.688	0.533	0.257	0.127
20.000	2.845	2.528	2.086	1.725	1.325	0.860	0.687	0.533	0.257	0.127
21.000	2.831	2.518	2.080	1.721	1.323	0.859	0.686	0.532	0.257	0.127
22.000	2.819	2.508	2.074	1.717	1.321	0.858	0.686	0.532	0.256	0.127
23.000	2.807	2.500	2.069	1.714	1.319	0.858	0.685	0.532	0.256	0.127
24.000	2.797	2.492	2.064	1.711	1.318	0.857	0.685	0.531	0.256	0.127
25.000	2.787	2.485	2.060	1.708	1.316	0.856	0.684	0.531	0.256	0.127
26.000	2.779	2.479	2.056	1.706	1.315	0.856	0.684	0.531	0.256	0.127
27.000	2.771	2.473	2.052	1.703	1.314	0.855	0.684	0.531	0.256	0.127
28.000	2.763	2.467	2.048	1.701	1.313	0.855	0.683	0.530	0.256	0.127
29.000	2.756	2.462	2.045	1.699	1.311	0.854	0.683	0.530	0.256	0.127
30.000	2.750	2.457	2.042	1.697	1.310	0.854	0.683	0.530	0.256	0.127
40.000	2.704	2.423	2.021	1.684	1.303	0.851	0.681	0.529	0.255	0.126
60.000	2.660	2.390	2.000	1.671	1.296	0.848	0.679	0.527	0.254	0.126
120.000	2.617	2.358	1.980	1.658	1.289	0.845	0.677	0.526	0.254	0.126