PROBABILITY AND STATISTICS

הסתברות וסטטיסטיקה מאת יוג'ין קנציפר

הרצאה 4

משתנה מקרי חד ממדי בדיד

• משתנה מקרי • משתנה מקרי בדיד • פונקצית הסתברות • פונקצית התפלגות מצטברת •
• טבלת הסתברות • דיאגרמת מקלות • הסטוגרמה • תוחלת • שונות • סטיית תקן • חציון •

מטרת ההרצאה – לפתח שיטת תיאור של ניסוים אקראיים על ידי **פונקצית הסתברות**.

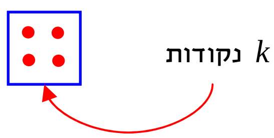
4.1 הגדרות בסיסיות

הגדרה D4.1

משתנה מקרי X [random variable] על מרחב מדגם Ω הוא <u>פונקציה ממשית</u> המוגדרת על Ω כך שלכל תוצאה ω של ניסוי אקראי מתאים ערך מספרי ממשי $X(\omega)$.

דוגמא E4.1

בניסוי "הטלת קוביית שש-בש", מרחב המדגם הוא $\Omega=\{\omega_1,\omega_2,\omega_3,\omega_4,\omega_5,\omega_6\}$; התוצאה האלמנטרית ω_k הוא המאורע במשמעות "פאה עם k נקודות":



א. נגדיר משתנה מקרי X על ידי הפונקציה $X(\omega_k)=k$ כך ש—

$$X(\omega_1)=1,\ X(\omega_2)=2,\ X(\omega_3)=3,\ X(\omega_4)=4,\ X(\omega_5)=5,\ X(\omega_6)=6\ .$$

ב. אפשר להגדיר כמה משתנים מקריים על אותו מרחב מדגם. למשל, פונקציה $Y(\omega_k)=10\cdot k-5$ מגדירה משתנה מקרי Y על אותו מרחב המדגם Ω.

דוגמא E4.2

בניסוי "הטלת מטבע שלוש פעמים", מרחב המדגם הוא

$$\Omega=\{HHH,HHT,HTH,HTT,THH,THT,TTH,TTT\}.$$

א. נגדיר משתנה מקרי X כפונקציה אשר שווה ל—1 אם בהטלה הראשונה מתקבל <u>עץ</u> ושווה ל—0 במקרים אחרים.

תוצאה	HHH	HHT	HTH	HTT	THH	THT	TTH	TTT
X	1	1	1	1	0	0	0	0

ב. עבור אותו הניסוי אפשר להגדיר משתנה מקרי בדרך אחרת. למשל, נגדיר משתנה מקרי Y כפונקציה אשר שווה ל—1 אם <u>פלי</u> נתקבל באמצע ושווה ל—0 במקרים אחרים.

תוצאה	HHH	HHT	HTH	HTT	THH	THT	TTH	TTT
Y	0	0	1	1	0	0	1	1

הגדרה D4.2

משתנה מקרי בדיד X [discrete random variable] על מרחב מדגם Ω הוא משתנה מקרי המקבל רק מספר סופי או בן מניה של ערכים (למשל, מספרים שלמים – אפשר למנות אותם למרות כמותם אינסופי).

4.2 פונקצית הסתברות ותכונותיה

פונקצית הסתברות מכילה את כל האינפורמציה על תופעה מקרית. היא מחושבת ממודל פיסיקלי—מתמטי של התופעה הנחקרת.

4.2.1 פונקצית הסתברות

בסעיף זה, נתבונן במשפחות בנות 2 ילדים. יהיה X משתנה מקרי המתאר מספר בנות במשפחה שנבחרה מקרית. נסמן בת ב— F ובן ב— M. כל התוצאות האפשריות של הניסוי מביאות את מרחב המדגם. מבחינת מין הילדים, מרחב המדגם הוא $\Omega = \{MF, FM, MM, FF\}$.

לכל תוצאה אלמנטרית ממרחב המדגם מתאים ערך מסוים של משתנה מקרי X:

$$X(MM) = 0,\ X(MF) = X(FM) = 1,\ X(FF) = 2 .$$

אזי, הערכים האפשריים של המשתנה X הם 0, 1, 2. נחשב את ההסתברות לקבלת כל אחד מן הערכים האלה:

$$P(X = 0) = P(MM) = \frac{1}{4},$$

$$P(X = 1) = P(FM) + P(MF) = \tfrac{1}{4} + \tfrac{1}{4} = \frac{1}{2},$$

$$P(X = 2) = P(FF) = \frac{1}{4}.$$

כרגע חישבנו את פונקצית ההסתברות של משתנה מקרי X.

הגדרה D4.3

פונקצית הסתברות [probability function] $P_X(x)$ של משתנה מקרי **בדיד** X מצמידה לכל ערך אפשרי x של המשתנה X את ההסתברות $P_X(x) = P(X = x)$.

הפונקציה הפורמאלית $P_X(x)$ מכילה את כל האינפורמציה ההסתברותית על משתנה מקרי X. בעזרתה נוכל לחשב כל ההסתברויות הדרושות. למשל,

— הסתברות שישנן בנות במשפחה היא $P(X \geq 1) = P(X = 1) + P(X = 2) = \tfrac{3}{4}$

— הסתברות שישנן יותר בנות מאשר בנים היא $P(X > 1) = P(X = 2) = \tfrac{1}{4}$

— וכו'

4.2.2 תכונות של פונקציית הסתברות

תכונות של פונקציית הסתברות נובעות משלוש אקסיומות של קולמוגורוב (ראה/י את ההרצאה הראשונה). הן:

א. פונקציית הסתברות אינה שלילית: $P_X(x) \geq 0$ לכל ערך x.

ב. עבור פונקציית הסתברות כלשהי מתקיימת **תכונת הנרמול**: $\sum_x P_X(x) = 1$. כאן, הסכום עובר על כל הערכים האפשריים של המשתנה X.

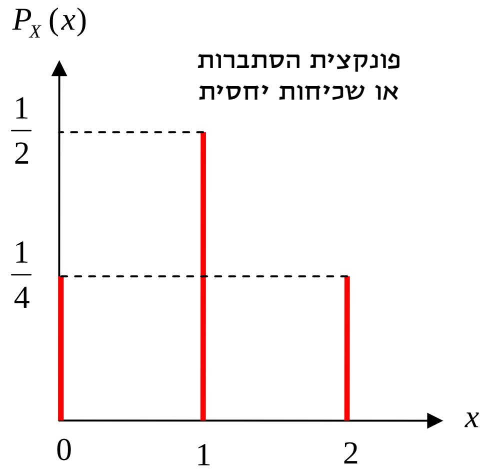
4.2.3 הצגות שונות של פונקציית הסתברות

נתבונן בדוגמא של סעיף 4.2.1 כדי לייצג שלוש דרכי הצגה של פונקציית הסתברות.

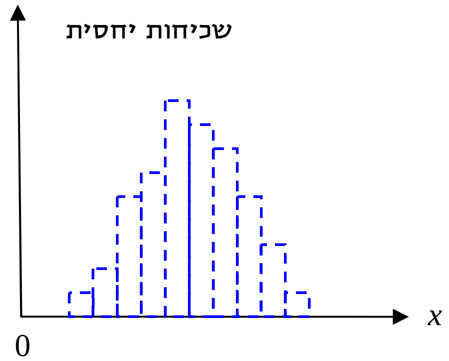
א. טבלת הסתברות

X	0	1	2	נרמול $\sum_x P_X(x)$
$P_X(x)$	1/4	1/2	1/4	1

ב. דיאגרמת מקלות (משתמשים גם בניסוי וגם בתיאוריה)



ג. הסטוגרמה (משתמשים בניסוי)



4.3 פונקצית התפלגות מצטברת ותכונותיה

4.3.1 פונקצית התפלגות מצטברת

הגדרה D4.4

פונקציית התפלגות מצטברת [cumulative distribution function] $F_X(t)$ של משתנה מקרי בדיד X מוגדרת על ידי הנוסחא

$$F_X(t) = P(X \le t) = \sum_{x \le t} P_X(x).$$

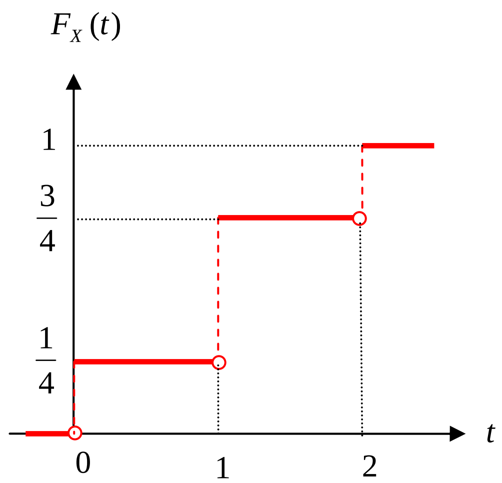
כאן, הסכום רק על ערכים של משתנה מקרי אשר לא עולים על ערך נתון $-\infty < t < +\infty$. הערך t הוא ארגומנט של פונקצית התפלגות מצטברת $F_X(t)$.

דוגמא E4.3

חישוב והצגה של פונקצית ההתפלגות מצטברת עבור הדוגמא של סעיף 4.2.1. נוח להתחיל את החישוב מ**דיאגרמת מקלות** עבור פונקצית הסתברות $P_X(x)$. מהגרף הלוו כי קיימות 3 נקודות חשובות: $t = 0,1,2$. מקבלים:

$$\begin{aligned}
&F_X(t<0) = 0,\\
&F_X(0 \le t < 1) = P(X=0) = 1/4,\\
&F_X(1 \le t < 2) = P(X=0) + P(X=1) = 1/4 + 1/2 = 3/4,\\
&F_X(t \ge 2) = P(X=0) + P(X=1) + P(X=2) = 1/4 + 1/2 + 1/4 = 1.
\end{aligned}$$

גראף עבור $F_X(t)$:



4.3.2 תכונות של פונקצית התפלגות מצטברת

א. $\lim_{t \to -\infty} F_X(t) = 0$

ב. $\lim_{t \to +\infty} F_X(t) = 1$

ג. עבור כל $t_1 < t_2$ מתקיים: $F_X(t_1) \le F_X(t_2)$

ד. אם $F_X(t_0) = 0$ עבור t_0 כלשהו, אזי $F_X(t) = 0$ עבור כל $t \le t_0$

ה. $P(x_1 < X \le x_2) = F_X(x_2) - F_X(x_1)$ עבור $x_1 < x_2$

הוכחות:

א. על פי הגדרה, $F_X(t)=P(X\le t)$. אזי, $\lim_{t\to-\infty}F_X(t)=P(X\le-\infty)=0$.
סוף הוכחה.

ב. על פי הגדרה, $F_X(t)=P(X\le t)$. אזי, $\lim_{t\to+\infty}F_X(t)=P(X\le+\infty)=1$.
סוף הוכחה.

ג. כיוון ש— $F_X(t)=P(X\le t)$, מקבלים:
$F_X(t_2)-F_X(t_1)=P(X\le t_2)-P(X\le t_1)=P(t_1<X\le t_2)$. עבור $t_1<t_2$ מתקיים:
$P(t_1<X\le t_2)\ge 0$. כתוצאה, $F_X(t_2)-F_X(t_1)\ge 0$.
סוף הוכחה.

ד. עבור כל $t\le t_0$ מתקיים: $F_X(t_0)-F_X(t)\ge 0$ (ראה/י סעיף ג'). כתוצאה,
$F_X(t)\le F_X(t_0)=0$. מצד שני, ההגדרה של $F_X(t)=P(X\le t)$ אומרת כי $F_X(t)$ אינה
שלילית. האופציה היחידה שנשארת היא $F_X(t)=0$.
סוף הוכחה.

ה. ראה/י סעיף ג'.

4.4 מדדים של משתנה מקרי

נלמד מספר מדדים חשובים של משתנה מקרי: תוחלת, שונות, חציון, שכיח וכו', ותכונותיהם.

4.4.1 תוחלת של משתנה מקרי ותכונותיה

הגדרה D4.5

תוחלת [expectation value=mean] של משתנה מקרי X נתונה על ידי הנוסחא

$$\mu=E[X]=\sum_x x\cdot P_X(x).$$

כאן, סכום רץ על כל הערכים האפשריים x של משתנה מקרי X.

הערה R4.1

<u>משמעות של תוחלת</u>: תוחלת של משתנה מקרי היא ממוצע של אותו המשתנה המחושב על סמך מספר אינסופי של חזרות הניסוי. אנחנו נוכיח את הטענה בהמשך הקורס.

דוגמא E4.4

תוחלת עבור המשתנה X = {מספר בנות במשפחה בת 2 ילדים} היא

$$E[X]=0\cdot\tfrac{1}{4}+1\cdot\tfrac{1}{2}+2\cdot\tfrac{1}{4}=1.$$

על פי **הערה R4.1**, משמעות התשובה: בממוצע (המחושב על סמך מספר רב של משפחות—מספר אינסופי) תימצא בת אחת במשפחות בנות 2 ילדים.

הגדרה D4.6

תוחלת של פונקציה כלשהי של משתנה מקרי X מחושבת על פי הנוסחא:

$$E[g(X)]=\sum_x g(x)\cdot P_X(x).$$

תכונות התוחלת

א. תוחלת של משתנה המקבל ערך קבוע a בוודאות, היא קבוע עצמו: $E[a]=a$.

הוכחה: $E[a]=\sum_x a\cdot P_X(x)=a\cdot\underbrace{\sum_x P_X(x)}_{=1}=a$.

ב. עבור כל קבוע a, מתקיים: $E[X+a]=E[X]+a$.

הוכחה: $E[X+a]=\sum_x (x+a)\cdot P_X(x)=\sum_x x\cdot P_X(x)+a\cdot\underbrace{\sum_x P_X(x)}_{=1}=E[X]+a$.

ג. עבור כל קבוע b, מתקיים: $E[b\cdot X]=b\cdot E[X]$.

הוכחה: $E[b\cdot X]=\sum_x b\cdot x\cdot P_X(x)=b\cdot\sum_x x\cdot P_X(x)=b\cdot E[X]$.

ד. אם X_1 ו–X_2 הם שני משתנים מקריים, $E[X_1+X_2]=E[X_1]+E[X_2]$.

הוכחה: בהמשך הקורס.

4.4.2 שונות של משתנה מקרי ותכונותיה

הגדרה D4.7

שונות [variance] של משתנה מקרי X נתונה על ידי הנוסחא

$$\sigma^2=\mathrm{var}[X]=E\left[(X-\mu)^2\right]=\sum_x (x-\mu)^2\cdot P_X(x)$$

כאן, $\mu=E[X]$.

הערה R4.2

הנוסחא המקבילה לחישוב השונות היא $\mathrm{var}[X]=E\left[X^2\right]-(E[X])^2$.

הוכחה.

$$\begin{aligned}\mathrm{var}[X]&=E\left[(X-\mu)^2\right]=\sum_x (x-\mu)^2\cdot P_X(x)=\sum_x \left(x^2-2\mu\, x+\mu^2\right)\cdot P_X(x)\\&=\sum_x x^2\cdot P_X(x)-2\mu\underbrace{\sum_x x\cdot P_X(x)}_{\mu}+\mu^2\underbrace{\sum_x P_X(x)}_{1}=E\left[X^2\right]-\mu^2\\&=E\left[X^2\right]-(E[X])^2.\end{aligned}$$

סוף הוכחה.

הערה R4.3

מתקיים:

$$\mathrm{var}[X]=\sum_x x^2\cdot P_X(x)-\left(\sum_x x\cdot P_X(x)\right)^2.$$

הגדרה D4.7

סטיית התקן [standard deviation] של משתנה מקרי X היא שורש השונות: $\sigma = \sqrt{\mathrm{var}[X]}$.

הערה R4.4

<u>משמעות של שונות וסטיית התקן</u>: לשונות עצמה אין משמעות. משמעות של <u>סטיית התקן</u> היא סדר גודל הרוחב של פונקציית הסתברות ברמה של מחצית הגובה. במילים אחרות, סטיית התקן היא מידת הפיזור של ערכי המשתנה סביב מרכזם.

<u>תכונות השונות</u>

א. שונות אינה שלילית: $\mathrm{var}[X] \ge 0$ לכל משתנה מקרי X.

הוכחה: על פי הגדרה, $\mathrm{var}[X] = E\left[(X-\mu)^2\right] = \sum_x \underbrace{(x-\mu)^2}_{\ge 0} \cdot \underbrace{P_X(x)}_{\ge 0} \ge 0$.

סוף הוכחה.

ב. $\mathrm{var}[X] = 0$ אך ורק כאשר X הוא קבוע בוודאות.

הוכחה: על פי הגדרה, $\mathrm{var}[X] = E\left[(X-\mu)^2\right] = \sum_x \underbrace{(x-\mu)^2}_{\ge 0} \cdot \underbrace{P_X(x)}_{\ge 0} = 0$. השוויון אפשרי אך ורק כאשר משתנה מקרי X מקבל רק ערך אפשרי אחד השווה ל— μ. מכיוון ש— μ הוא קבוע, התכונה הוכחה.

ג. עבור כל קבוע a מתקיים: $\mathrm{var}[X+a] = \mathrm{var}[X]$.

הוכחה.

$$\begin{aligned}
\mathrm{var}[X+a] &= E\left[(X+a)^2\right] - (E[X+a])^2 = E\left[X^2 + 2aX + a^2\right] - (E[X]+a)^2 \\
&= E\left[X^2\right] + 2aE[X] + a^2 - (E[X])^2 - 2aE[X] - a^2 \\
&= E\left[X^2\right] - (E[X])^2 = \mathrm{var}[X]
\end{aligned}$$

סוף הוכחה.

ד. עבור כל קבוע b, מתקיים: $\mathrm{var}[b \cdot X] = b^2 \cdot \mathrm{var}[X]$.

הוכחה.

$$\begin{aligned}
\mathrm{var}[b \cdot X] &= E\left[(b \cdot X)^2\right] - (E[b \cdot X])^2 = E\left[b^2 \cdot X^2\right] - (b \cdot E[X])^2 \\
&= b^2 \cdot E\left[X^2\right] - b^2 \cdot (E[X])^2 = b^2 \cdot \mathrm{var}[X]
\end{aligned}$$

ה. עבור שני משתנים מקריים X ו— Y **בלתי תלויים**, מתקיים:

$$\mathrm{var}[X+Y] = \mathrm{var}[X] + \mathrm{var}[Y].$$

הוכחה. בהמשך הקורס.

4.4.3 חציון וערך החלוקה ה—p

הגדרה D4.8

חציון [median] עבור משתנה מקרי X הוא ערך M המקיים את שני התנאים הבאים:

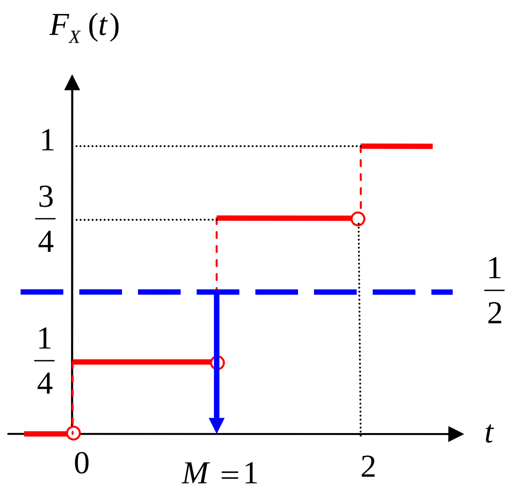
א. $P(X < M) \le \frac{1}{2}$ **ב.** $P(X \le M) \ge \frac{1}{2}$.

הערה R4.5

חציון הוא הערך הראשון, שעבורו פונקצית ההתפלגות המצטברת היא לפחות $\frac{1}{2}$.

דוגמא E4.5

חציון M עבור המשתנה X = {מספר בנות במשפחה בת 2 ילדים שנבחרה מקרית} הוא $M = 1$ (ראה/י ציור).



הגדרה D4.9

ערך החלוקה ה- p עבור משתנה מקרי X הוא ערך M_p המקיים את שני התנאים הבאים:

א. $P(X < M_p) \le p$ **ב.** $P(X \le M_p) \ge p$.

כאן, $0 < p < 1$.

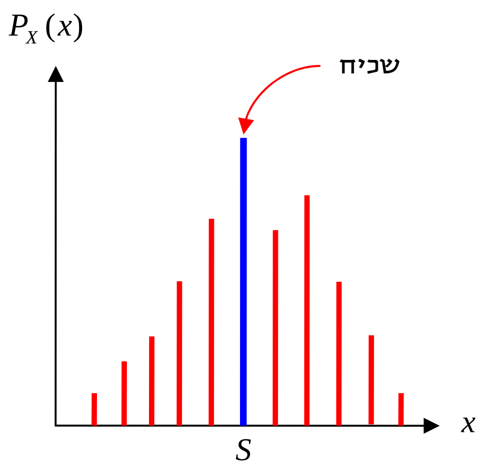
הערה R4.6

ערך החלוקה ה- p הוא הערך הראשון, שעבורו פונקצית ההתפלגות המצטברת היא לפחות $0 < p < 1$.

4.4.4 שכיח

הגדרה D4.10

שכיח S הוא ערך של משתנה מקרי X בוא פונקצית ההסתברות היא גבוהה ביותר:
$P_X(x \neq S) < P_X(x = S)$.



הערה R4.7

לפעמים, ישנם שני (או יותר) שכיחים: S_1 ו—S_2. עבורם מתקיים:

$$P_X(x \neq S_1, S_2) < P_X(x = S_1) = P_X(x = S_2).$$