

הסתברות מותנית ונוסחת בייס

מה משתנה כשיודעים משהו – ולמה בדיקה "מדויקת ב-99%" יכולה לטעות בחצי מהתוצאות החיוביות שלה

כ־30 דקות קריאה

3 הוכחות לנוסחת בייס

הסתברות

4-5 יח"ל

כיתות ו'-י"א

מה כדאי לדעת קודם

- הסתברות כ"מספר התוצאות הרצויות חלקי מספר כל התוצאות".
- חיתוך של מאורעות $(A \cap B)$ ומאורע משלים $(\bar{A})$.
- חיבור הסתברויות של מאורעות זרים.
- עץ הסתברויות (ראו גם את המסמך "הבינום של ניוטון ונוסחת ברנולי" בסדרה).

מה תמצאו כאן

- אינטואיציה: מידע חדש "מצמצם את העולם".
- ההגדרה של הסתברות מותנית, ושני נימוקים שמצדיקים אותה.
- כלל הכפל, אי-תלות ונוסחת ההסתברות השלמה – עם הוכחות.
- נוסחת בייס: שלוש הוכחות.
- הפרדוקס של הבדיקה המדויקת, ו"כשל התובע".
- שאלות פתורות בסגנון הבגרות, ושאלות לבדיקה עצמית.

1 האינטואיציה: מצמצמים את העולם

האינטואיציה

בכיתה יש 30 תלמידים. במגמת פיזיקה לומדים 12 מהם, ובמגמת מדעי המחשב – 10. שמונה תלמידים לומדים בשתי המגמות. המורה מגרילה תלמיד אחד. מה ההסתברות שהוא לומד מדעי המחשב? התשובה: 10 מתוך 30, כלומר $\frac{1}{3}$.

עכשיו מישוה לוחש לנו: "התלמיד שעלה בהגרלה לומד פיזיקה". כל 18 התלמידים שלא לומדים פיזיקה כבר לא באים בחשבון, וכאילו ההגרלה נעשתה רק בין 12 תלמידי הפיזיקה. שמונה מהם לומדים מדעי המחשב, ולכן ההסתברות היא עכשיו 8 מתוך 12, כלומר $\frac{2}{3}$ – פי שניים ממקודם.

זה כל הרעיון של הסתברות מותנית: **מידע חדש מצמצם את העולם**. מחשבים שוב "כמה מתוך כמה", אבל ה"מתוך" הוא רק מה שעדיין אפשרי.

סך הכול	לא מדעי המחשב	מדעי המחשב	
12	4	8	פיזיקה
18	16	2	לא פיזיקה
30	20	10	סך הכול

את הנתונים נוח לסדר בטבלה דו-ממדית. לפני הלחישה מחלקים את סך עמודת מדעי המחשב במספר כל התלמידים: $\frac{10}{30}$. אחרי הלחישה נשארת רק שורת הפיזיקה, ומחלקים בסך השורה: $\frac{8}{12}$.



סרטוט 1: מימין: כל הכיתה, ובירוק תלמידי מדעי המחשב. משמאל: אחרי הלחישה "העולם" הוא רק עיגול הפיזיקה, ובתוכו 8 תלמידים לומדים גם מדעי המחשב.

2 ההגדרה

נסמן ב- A את המאורע "התלמיד לומד מדעי המחשב", וב- B את המאורע "התלמיד לומד פיזיקה". מה שחישובנו הוא **ההסתברות של A בהינתן B** : ההסתברות ש- A קרה, כשידוע ש- B קרה.

הגדרה: הסתברות מותנית

יהיו A ו- B שני מאורעות, ונניח ש- $P(B) > 0$. ההסתברות של A בהינתן B היא

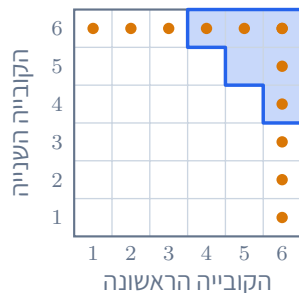
$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

במילים: איזה חלק מההסתברות של B "תופס" החיתוך שלו עם A .

על הסימון. כך, עם קו נטוי, כתובה ההסתברות המותנית בדף הנוסחאות של הבגרות (שם האותיות מוחלפות: $P(B/A) = \frac{P(A \cap B)}{P(A)}$ – זו אותה נוסחה). בספרים רבים בעולם כותבים $P(A|B)$, עם קו אנכי; המשמעות זהה. בדוגמת הכיתה ההגדרה נותנת בדיוק את מה שקיבלנו בספירה: $P(A/B) = \frac{8/30}{12/30} = \frac{8}{12} = \frac{2}{3}$. ולמה צריך $P(B) > 0$? אי אפשר "לדעת" שקרה מאורע שאינו יכול לקרות – וגם אי אפשר לחלק באפס.

3 למה דווקא ההגדרה הזאת?

הגדרה לא מוכיחים – מחליטים עליה. אבל הגדרה טובה חייבת להתאים למה שאנחנו כבר יודעים בוודאות. נראה שבשני מצבים שבהם ברור מה התשובה צריכה להיות, ההגדרה נותנת בדיוק אותה.



סרטוט 2: בכחול: B . בנקודות כתומות: A .

נימוק 1 תוצאות שוות הסתברות. כשכל התוצאות שוות הסתברות, "מספר התוצאות הרצויות חלקי מספר כל התוצאות". "בהינתן B " פירושו שבמכנה נשארות רק התוצאות של B .

דוגמה. מטילים שתי קוביות, זו אחר זו. יש 36 תוצאות שוות הסתברות (סרטוט 2). ידוע שסכום התוצאות הוא 10 לפחות – זה המאורע B , ויש בו 6 תוצאות. מה ההסתברות שלפחות אחת הקוביות יצא 6 (המאורע A)? ב- 5 מתוך 6 התוצאות של B מופיע 6, ולכן $P(A/B) = \frac{5}{6}$. בלי המידע, $P(A) = \frac{11}{36}$, רק בערך 0.31.

באופן כללי, נסמן ב- $n(E)$ את מספר התוצאות במאורע E , וב- $n(\Omega)$ את מספר כל התוצאות. המידע "קרה B " פוסל את התוצאות שמחוץ ל- B , אבל לא מעדיף אף תוצאה של B על פני אחרת. לכן בעולם המצומצם יש $n(B)$ תוצאות שוות הסתברות, ו- A קורה בדיוק ב- $n(A \cap B)$ מהן. נחלק את המונה ואת המכנה ב- $n(\Omega)$:

$$P(A/B) = \frac{n(A \cap B)}{n(B)} = \frac{n(A \cap B) / n(\Omega)}{n(B) / n(\Omega)} = \frac{P(A \cap B)}{P(B)}.$$



סרטוט 3: מגדילים את החלק של B . בתוכו, החלק של A הוא בערך $\frac{5}{6}$.

נימוק 2 שכיחויות. ההסתברות של מאורע היא בערך **השכיחות היחסית** שלו: החלק מהפעמים שבהן הוא קורה, כשחוזרים על הניסוי פעמים רבות. "בהינתן B " פירושו שמסתכלים רק על הפעמים שבהן B קרה.

דוגמה. נטיל את שתי הקוביות 3600 פעמים. סכום של 10 לפחות יתקבל בערך ב- $600 = 3600 \cdot \frac{6}{36}$ הטלות, ו- A יחד עם B – בערך ב- $500 = 3600 \cdot \frac{5}{36}$ הטלות. מבין 600 ההטלות של B , בערך $\frac{500}{600} = \frac{5}{6}$ הן גם הטלות של A .

באופן כללי, נחזור על הניסוי N פעמים, כש- N מספר גדול. B יקרה בערך $N \cdot P(B)$ פעמים, ו- A יחד עם B בערך $N \cdot P(A \cap B)$ פעמים. לכן השכיחות של A בין הפעמים שבהן B קרה היא בערך $\frac{N \cdot P(A \cap B)}{N \cdot P(B)} = \frac{P(A \cap B)}{P(B)}$, ו- N מצטמצם. ככל ש- N גדול יותר, הקירוב טוב יותר. שני הנימוקים מגיעים לאותה נוסחה, ובשניהם היא אומרת אותו דבר: $P(A/B)$ היא **החלק של A בתוך B** .

4 כלל הכפל ועץ ההסתברויות

נכפול את שני האגפים של ההגדרה ב- $P(B)$, ונקבל כלל שימושי במיוחד:

כלל הכפל

אם $P(B) > 0$, אז

$$P(A \cap B) = P(B) \cdot P(A/B)$$

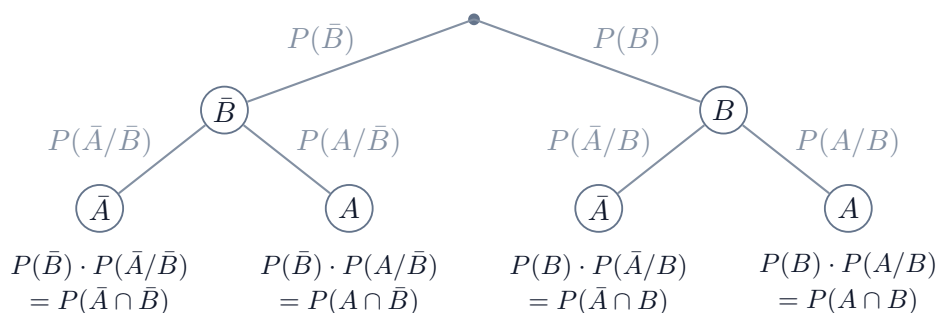
ובאותו אופן, אם $P(A) > 0$, אז $P(A \cap B) = P(A) \cdot P(B/A)$.

במילים: כדי ש- A ו- B יקרו שניהם, צריך ש- B יקרה, וגם ש- A יקרה **בעולם המצומצם** שבו B כבר קרה.

זה בדיוק מה שעושים כשכופלים הסתברויות לאורך מסלול בעץ הסתברויות (סרטוט 4). בשלב הראשון של העץ מתפצלים לפי B , וההסתברויות על הענפים הן $P(B)$ ו- $P(\bar{B})$. בשלב השני כבר "נמצאים" בתוך B או בתוך $\bar{B}$, ולכן ההסתברויות על הענפים שם הן **מותנות**. מכפלה לאורך המסלול מהשורש אל A דרך B היא כלל הכפל: $P(B) \cdot P(A/B) = P(A \cap B)$.

בדוגמת הכיתה: $P(B) = \frac{12}{30} = \frac{2}{5}$ ו- $P(A/B) = \frac{2}{3}$, ולכן $P(A \cap B) = \frac{2}{5} \cdot \frac{2}{3} = \frac{4}{15} = \frac{8}{30}$, כלומר 8 תלמידים מתוך 30, כמו בטבלה. ולמה ההסתברויות על הענפים שיוצאים מכל צומת מסתכמות ב-1? כי בעולם המצומצם B המאורע A קורה או לא קורה:

$$P(A/B) + P(\bar{A}/B) = \frac{P(A \cap B) + P(\bar{A} \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1.$$



סרטוט 4: עץ בשני שלבים. בשלב השני ההסתברויות מותנות, והמכפלה לאורך כל מסלול היא ההסתברות של החיתוך שבסופו.

5 מאורעות בלתי תלויים

לפעמים המידע לא משנה דבר. מטילים שתי קוביות. ההסתברות שהסכום יהיה 7 היא $\frac{1}{6} = \frac{6}{36}$. ואם ידוע שבקובייה הראשונה יצא 6? אז הסכום הוא 7 רק אם בשנייה יצא 1, ולכן ההסתברות היא שוב $\frac{1}{6}$. במקרה כזה אומרים שהמאורעות "הסכום הוא 7" ו"בראשונה יצא 6" **בלתי תלויים**. (ההסתברות שהסכום יהיה 8, לעומת זאת, משתנה: מ- $\frac{5}{36}$ ל- $\frac{1}{6}$.)

אי-תלות: שני ניסוחים שקולים

אם $P(B) > 0$, אז

$$P(A/B) = P(A) \iff P(A \cap B) = P(A) \cdot P(B)$$

הוכחה. אם $P(A/B) = P(A)$, אז לפי כלל הכפל $P(A \cap B) = P(B) \cdot P(A/B) = P(B) \cdot P(A)$. ובכיוון השני: אם $P(A \cap B) = P(A) \cdot P(B)$, אז לפי ההגדרה $P(A/B) = \frac{P(A \cap B)}{P(B)} = P(A)$.

מש"ל

בניסוח השני A ו- B ממלאים תפקיד שווה. לכן, אם גם $P(A) > 0$, אז

$$P(A/B) = P(A) \iff P(B/A) = P(B).$$

כלומר אם B לא מלמד דבר על A , גם A לא מלמד דבר על B . (בגלל הסימטריה הזאת נוה להגדיר אי-תלות בעזרת המכפלה.) בכיתה, לעומת זאת, $P(A/B) = \frac{2}{3}$ ואילו $P(A) = \frac{1}{3}$, ולכן A ו- B תלויים.

⚠ טעות נפוצה: "זרים" זה לא "בלתי תלויים"

מאורעות זרים לא יכולים לקרות יחד. אם $P(A) > 0$ ו- $P(B) > 0$ והמאורעות זרים, אז $P(A/B) = 0$, ולא $P(A)$: הידיעה ש- B קרה אומרת בוודאות ש- A לא קרה. כלומר מאורעות כאלה **תלויים** מאוד.

6 נוסחת ההסתברות השלמה

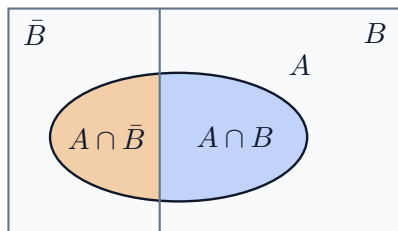
לעיתים קרובות יודעים מה קורה בכל ענף של העץ, ורוצים למצוא את ההסתברות הכוללת של A . למשל: איזה חלק מהכיתה לומד מדעי המחשב, אם יודעים איזה חלק מתלמידי הפיזיקה לומד את המקצוע, ואיזה חלק מהאחרים.

נוסחת ההסתברות השלמה

אם $0 < P(B) < 1$, אז

$$P(A) = P(B) \cdot P(A/B) + P(\bar{B}) \cdot P(A/\bar{B})$$

במילים: מחברים את כל המסלולים בעץ שמסתיימים ב- A .



סרטוט 5: A מתחלק לשני חלקים זרים.

הרעיון "חותכים" את A לשני חלקים זרים: החלק שבתוך B והחלק שמחוץ לו (סרטוט 5). את ההסתברות של כל חלק מחשבים בכלל הכפל.

בדוגמת הכיתה:

$$P(A) = \frac{2}{5} \cdot \frac{2}{3} + \frac{3}{5} \cdot \frac{1}{9} = \frac{4}{15} + \frac{1}{15} = \frac{1}{3},$$

כלומר 10 תלמידים מתוך 30. (מבין 18 התלמידים שלא לומדים פיזיקה, 2 לומדים מדעי המחשב, ולכן $P(A/\bar{B}) = \frac{2}{18} = \frac{1}{9}$).

נימוק

טענה

כל תוצאה של A נמצאת ב- B או ב- $\bar{B}$, ולא בשניהם

$$A = (A \cap B) \cup (A \cap \bar{B}) \quad (1)$$

ההסתברות של איחוד מאורעות זרים היא סכום ההסתברויות

$$P(A) = P(A \cap B) + P(A \cap \bar{B}) \quad (2)$$

כלל הכפל בכל מחובר: $P(B) > 0$ וגם $P(\bar{B}) > 0$

$$P(A) = P(B) \cdot P(A/B) + P(\bar{B}) \cdot P(A/\bar{B}) \quad (3)$$

מש"ל

אותה הוכחה עובדת גם כשמפצלים לשלושה מאורעות זרים או יותר, שמכסים יחד את כל האפשרויות (למשל שלוש מכונות במפעל): בנוסחה יש מחובר אחד לכל ענף.

7 נוסחת בייס

עד עכשיו הלכנו בעץ "קדימה": מ- B אל A . אבל לעיתים קרובות השאלה המעניינת הפוכה: רופא יודע מה ההסתברות שבדיקה של אדם חולה תצא חיובית, והנבדק רוצה לדעת מה ההסתברות שהוא חולה, כשהבדיקה שלו יצאה חיובית. נוסחת בייס **הופכת את ההתניה**.

נוסחת בייס

אם $0 < P(B) < 1$ ו- $P(A) > 0$ אז

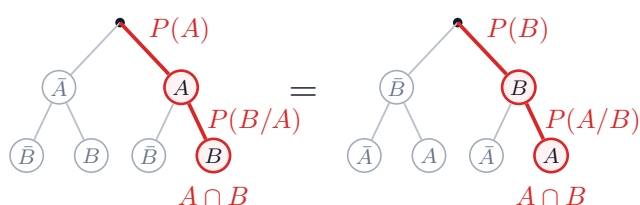
$$P(B/A) = \frac{P(B) \cdot P(A/B)}{P(B) \cdot P(A/B) + P(\bar{B}) \cdot P(A/\bar{B})}$$

ובכתיב מקוצר: $P(B/A) = \frac{P(B) \cdot P(A/B)}{P(A)}$

במילים: ההסתברות "להגיע ל- A דרך B ", חלקי ההסתברות להגיע ל- A בכל הדרכים.

ל- $P(B)$ קוראים לפעמים ההסתברות "לפני" (לפני שיודעים ש- A קרה), ול- $P(B/A)$ – ההסתברות "אחרי". נוסחת בייס מתרגמת "לפני" ל"אחרי".

8 הוכחה 1: את החיתוך כותבים בשתי דרכים



הרעיון את $P(A \cap B)$ אפשר לחשב בעץ שמתפצל קודם לפי B , או בעץ שמתפצל קודם לפי A (סרטוט 6). בעץ הימני מגיעים ל- $A \cap B$ דרך B , ובשמאלי – דרך A . זה אותו מאורע, ולכן שני החישובים חייבים לתת אותו מספר.

סרטוט 6: אותו מאורע, $A \cap B$, בשני עצים.

נימוק	טענה
כלל הכפל ($P(B) > 0$)	$P(A \cap B) = P(B) \cdot P(A/B)$ ①
כלל הכפל בסדר ההפוך ($P(A) > 0$)	$P(A \cap B) = P(A) \cdot P(B/A)$ ②
צעדים 1 ו-2: שני ביטויים לאותו מספר	$P(A) \cdot P(B/A) = P(B) \cdot P(A/B)$ ③
חילוק ב- $P(A)$, שהיא חיובית	$P(B/A) = \frac{P(B) \cdot P(A/B)}{P(A)}$ ④
נוסחת ההסתברות השלמה במכנה ($0 < P(B) < 1$)	$P(B/A) = \frac{P(B) \cdot P(A/B)}{P(B) \cdot P(A/B) + P(\bar{B}) \cdot P(A/\bar{B})}$ ⑤

מש"ל

עצרו ובדקו

מה נותנת נוסחת בייס כש- $P(A/\bar{B}) = 0$? הסבירו במילים למה התוצאה הגיונית.

9 הוכחה 2: סופרים אנשים במקום הסתברויות

הרעיון נדמיין קבוצה גדולה של N אנשים (למשל $N = 10,000$), ונסדר בטבלה **מספרי אנשים**, כמו בטבלת הכיתה בסעיף 1. בערך $N \cdot P(B)$ מהאנשים נמצאים ב- B , ומתוכם חלק $P(A/B)$ נמצא גם ב- A (כי $P(A/B)$ היא החלק של A בתוך B). באותו אופן סופרים את $\bar{B}$:

כמה אנשים בקבוצה?	כמה מהם ב- A ?	
$N \cdot P(B)$	$N \cdot P(B) \cdot P(A/B)$	ב- B
$N \cdot P(\bar{B})$	$N \cdot P(\bar{B}) \cdot P(A/\bar{B})$	ב- $\bar{B}$
N	$N \cdot P(B) \cdot P(A/B) + N \cdot P(\bar{B}) \cdot P(A/\bar{B})$	סך הכול

$P(B/A)$ היא החלק של B בתוך A : התא העליון בעמודה השמאלית, חלקי סך העמודה. נצמצם ב- N ונקבל את נוסחת בייס:

$$\begin{aligned} P(B/A) &= \frac{N \cdot P(B) \cdot P(A/B)}{N \cdot P(B) \cdot P(A/B) + N \cdot P(\bar{B}) \cdot P(A/\bar{B})} \\ &= \frac{P(B) \cdot P(A/B)}{P(B) \cdot P(A/B) + P(\bar{B}) \cdot P(A/\bar{B})}. \end{aligned}$$

מש"ל

הוכחה שונה באמת? לא לגמרי: זה החישוב של הוכחה 1, רק שכל הסתברות הוכפלה ב- N . מה שמרוויחים הוא הבנה: כל מספר בהוכחה הוא מספר של אנשים, ואת התשובה "רואים" בטבלה. גם הוכחה 3 נשענת על כלל הכפל, אבל את המכנה היא מקבלת בדרך אחרת: מכך ששתי ההסתברויות "אחרי" מסתכמות ב-1, בלי נוסחת ההסתברות השלמה.

10 הוכחה 3: היחס בין שתי האפשרויות (להעמקה)

הרעיון נשווה בין שתי ההסתברויות "אחרי", $P(B/A)$ ו- $P(\bar{B}/A)$, בלי לחשב את $P(A)$. נסמן

$$x = P(B) \cdot P(A/B), \quad y = P(\bar{B}) \cdot P(A/\bar{B}).$$

לפי ההגדרה וכלל הכפל, $P(B/A) = \frac{x}{P(A)}$ ו- $P(\bar{B}/A) = \frac{y}{P(A)}$. לשני השברים אותו מכנה, ולכן היחס ביניהם הוא $x : y$. בנוסף, סכומם 1, כי בעולם המצומצם A המאורע B קורה או לא קורה (סעיף 4). שני מספרים שהיחס ביניהם $x : y$ וסכומם 1 הם $\frac{x}{x+y}$ ו- $\frac{y}{x+y}$, ולכן $P(B/A) = \frac{x}{x+y}$ ו- $P(\bar{B}/A) = \frac{y}{x+y}$ — זאת נוסחת בייס.

מש"ל

מאותו יחס נובעת גם צורה של נוסחת בייס שאין בה $P(A)$ (אם $y > 0$):

$$\frac{P(B/A)}{P(\bar{B}/A)} = \frac{P(B)}{P(\bar{B})} \cdot \frac{P(A/B)}{P(A/\bar{B})}.$$

את היחס $\frac{P(B)}{P(\bar{B})}$ מבטאים בשפת היום-יום כ**סיכויים** (odds): "הסיכויים הם 1 ל-99", כלומר הסתברות של $\frac{1}{100}$ ולא של $\frac{1}{99}$. לפי הנוסחה, המידע ש- A קרה **נופל** את הסיכויים של B במספר $\frac{P(A/B)}{P(A/\bar{B})}$.

11 שלוש ההוכחות זו מול זו

הוכחה 1: החיתוך בשתי דרכים	הוכחה 2: סופרים אנשים	הוכחה 3: יחס
הרעיון	לחשב את $P(A \cap B)$ משני "כיוונים"	להשוות בין שתי האפשרויות "אחרי"
היתרון	קצרה, ומראה מאיפה בא כל איבר	רואים את התשובה בלי אלגברה
	טבלה של מספרי אנשים	בלי $P(A)$; המידע "כופל את הסיכויים"

12 הפרדוקס של הבדיקה המדויקת

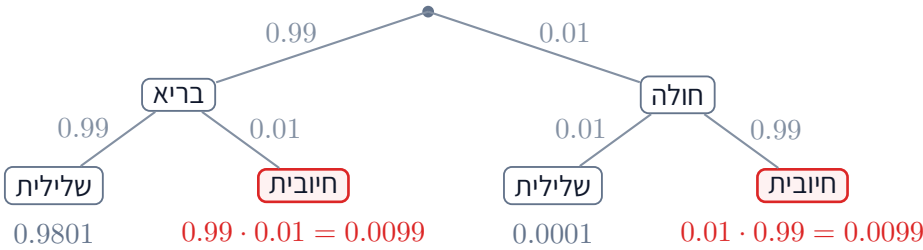
מחלה נדירה פוגעת ב־1% מהאוכלוסייה, ויש בדיקה מצוינת לגילוייה:

- אצל חולה הבדיקה יוצאת חיובית ב־99% מהמקרים (זאת **הרגישות** של הבדיקה).
- אצל בריא היא יוצאת שלילית ב־99% מהמקרים (זאת **הסגוליות** שלה). כלומר אצל 1% מהבריאים היא טועה ויוצאת חיובית.

דנה, שאין לה שום תסמין, נבדקה, והתוצאה יצאה חיובית. מה ההסתברות שהיא חולה? עצרו לרגע ונחשו. התשובה שקופצת לראש היא 99%, אבל התשובה הנכונה היא 50%.

נסמן ב־ B את המאורע "דנה חולה", וב־ A את המאורע "הבדיקה יצאה חיובית". לפי הנתונים, שכיחות המחלה היא $P(B) = 0.01$, הרגישות היא $P(A/B) = 0.99$, וההסתברות לטעות אצל בריא היא $P(A/\bar{B}) = 0.01$. מחפשים את $P(B/A)$, ונחשב אותה בשלוש דרכים.

דרך 1: עץ. לתוצאה חיובית מגיעים בשני מסלולים (סרטוט 7), ולשניהם אותה הסתברות, 0.0099. לכן ההסתברות של המסלול שעובר דרך "חולה" היא בדיוק חצי מההסתברות להגיע לתוצאה חיובית.



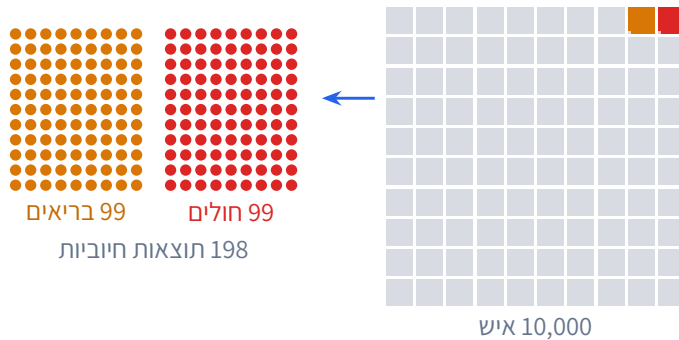
סרטוט 7: העץ של הבדיקה. המסלולים שמסתיימים בתוצאה חיובית מסומנים באדום.

דרך 2: טבלה של 10,000 איש. מתוך 10,000 איש יש 100 חולים, ו־9,900 בריאים. יש גם 9,900 בריאים, ו־1% מהם, כלומר 99, מקבלים גם הם תוצאה חיובית. סך הכול 198 תוצאות חיוביות, ומהן רק 99 של חולים: $\frac{99}{198} = \frac{1}{2}$.

דרך 3: נוסחת בייס.

$$P(B/A) = \frac{0.01 \cdot 0.99}{0.01 \cdot 0.99 + 0.99 \cdot 0.01} = \frac{1}{2}.$$

סך הכול	חיובית	שלילית	סך הכול
חולים	99	1	100
בריאים	99	9,801	9,900
סך הכול	198	9,802	10,000



סרטוט 8: כל ריבוע — 100 איש. אדום: חולים; כתום: בריאים עם תוצאה חיובית. משמאל: רק בעלי התוצאה החיובית.

איך ייתכן? הבדיקה באמת טובה: היא צודקת אצל $9,900 = 9,801 + 99$ מתוך 10,000 איש, כלומר אצל 99% מהנבדקים. אבל הבריאים רבים פי 99 מהחולים. לכן גם 1% בלבד מקבוצת הבריאים הענקית הם 99 אנשים עם תוצאה חיובית שגויה — בדיוק כמו מספר החולים שהבדיקה מגלה (סרטוט 8). מי שעונה 99% מתעלם מכך שהמחלה נדירה, כלומר **משיעור הבסיס** שלה (base rate).

וכמה זה תלוי בשכיחות המחלה? מאוד. עם אותה בדיקה בדיוק: אילו המחלה הייתה פוגעת ב־10% מהאוכלוסייה, מי שקיבל תוצאה חיובית היה חולה בהסתברות של כ־92%. ואם המחלה הייתה פוגעת רק ב־0.1%, ההסתברות הזאת הייתה רק כ־9%. גם אנשי מקצוע טועים כאן. ב־1978 הציגו חוקרים שאלה דומה ל־60 רופאים וסטודנטים לרפואה בבית הספר לרפואה של הרווארד: מחלה פוגעת באחד מכל 1,000 איש, ובדיקה לגילוייה יוצאת חיובית אצל 5% מהבריאים. התשובה הנפוצה ביותר הייתה 95%, ורק 11 ענו נכון: בערך 2% (אם הבדיקה מגלה כל חולה).

בדיקה שנייה. דנה נבדקת שוב, בבדיקה נוספת, וגם היא יוצאת חיובית. נניח שאצל חולים, וגם אצל בריאים, תוצאת הבדיקה השנייה בלתי תלויה בתוצאת הראשונה (זו הנחה חשובה: אם התוצאה השגויה נגרמת ממשוהו בגוף של הנבדק, היא עלולה לחזור גם בבדיקה השנייה). עכשיו ההסתברות "לפני" היא כבר $\frac{1}{2}$, ונוסחת בייס נותנת
$$0.99 = \frac{0.5 \cdot 0.99}{0.5 \cdot 0.99 + 0.5 \cdot 0.01}$$
 ובשפה של הוכחה 3: הסיכויים של דנה להיות חולה התחילו ב־1 ל־99, וכל תוצאה חיובית כופלת אותם ב־99 $\frac{0.99}{0.01}$. לכן אחרי הבדיקה הראשונה הם 1 ל־1, ואחרי השנייה — 99 ל־1, כלומר 99%.

13 מלכודת: $P(A/B)$ אינה $P(B/A)$

הפרדוקס נולד מבלבול בין שתי הסתברויות. הנתון 99% הוא $P(A/B)$, ההסתברות לתוצאה חיובית אצל חולה. השאלה הייתה על $P(B/A)$, ההסתברות להיות חולה אצל מי שקיבל תוצאה חיובית. בשתייה יש אותו חיתוך, אבל העולם המצומצם שונה: פעם מחלקים במספר החולים (100), ופעם במספר התוצאות החיוביות (198). דוגמה מוכרת: כמעט כל שחקני הכדורסל המקצוענים גבוהים מאוד, אבל רק חלק זעיר מהאנשים הגבוהים מאוד הם שחקני כדורסל מקצוענים.

כשל התובע (prosecutor's fallacy). בזירת פשע נמצא דנ"א שמתאים לאדם אקראי בהסתברות של אחת למיליון. החשוד מתאים, והתובע טוען: "ההסתברות שהוא חף מפשע היא אחת למיליון". אבל אחת למיליון היא ההסתברות להתאמה **בהינתן** שהאדם חף מפשע, ובית המשפט צריך את ההסתברות שהוא חף מפשע **בהינתן** ההתאמה. אם החשוד נמצא רק בגלל ההתאמה, בחיפוש בין 5 מיליון תושבי העיר, צפויים בעיר עוד בערך 5 חפים מפשע שמתאימים גם הם, ובלי ראיות נוספות ההסתברות שהחשוד הוא האשם היא רק בערך 1 מתוך 6.

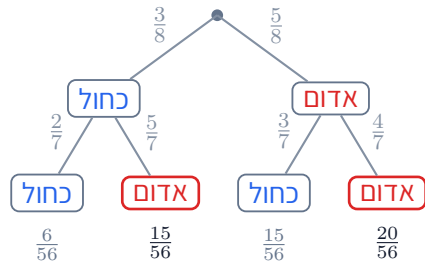
⚠ עוד שתי טעויות נפוצות

- **מחליפים בין $P(A/B)$ ל־ $P(A \cap B)$.** ההסתברות שהתלמיד לומד פיזיקה וגם מדעי המחשב היא $\frac{8}{30}$, אבל ההסתברות שהוא לומד מדעי המחשב, כשידוע שהוא לומד פיזיקה, היא $\frac{8}{12}$. מילים כמו "ידוע ש..." ו"בהינתן ש..." מסמנות הסתברות מותנית.
- **כותבים $P(A)$ על ענפי השלב השני.** שם כותבים הסתברויות מותנות, ובענפים שונים הן בדרך כלל שונות.

14 מה עושים עם הנוסחאות?

שאלה 1: "ידוע שהכדור השני אדום"

השאלה. בקופסה 5 כדורים אדומים ו-3 כחולים. מוציאים כדור באקראי, ובלי להחזיר אותו מוציאים כדור נוסף. **א.** מה ההסתברות שהכדור השני אדום? **ב.** ידוע שהכדור השני אדום. מה ההסתברות שגם הראשון היה אדום?



הפתרון. נסמן: R_1 – הראשון אדום, R_2 – "השני אדום". בשלב השני של העץ ההסתברויות מותנות: אחרי שיצא כדור אדום נשארו 4 אדומים מתוך 7.

א. מחברים את שני המסלולים שמסתיימים באדום:

$$P(R_2) = \frac{5}{8} \cdot \frac{4}{7} + \frac{3}{8} \cdot \frac{5}{7} = \frac{20}{56} + \frac{15}{56} = \frac{5}{8}.$$

$$\text{ב. } P(R_1/R_2) = \frac{P(R_1 \cap R_2)}{P(R_2)} = \frac{20/56}{35/56} = \frac{4}{7}.$$

שימו לב: בסעיף א יצא $\frac{5}{8}$, בדיוק ההסתברות שהכדור הראשון אדום – לפני שמסתכלים, אין הבדל בין הכדור הראשון לשני. ובסעיף ב שאלנו על ההוצאה הראשונה לפי מה שקרה בשנייה. זה מותר: הסתברות מותנית עוסקת במה **שיודעים**, לא בסדר שבו הדברים קרו.

שאלה 2: פרמטר לא ידוע

שאלה בסגנון הבגרות

יואב מגיע לבית הספר באוטובוס בהסתברות p , ואחרת הוא הולך ברגל. כשהוא נוסע באוטובוס, הוא מאחר בהסתברות 0.1, וכשהוא הולך ברגל – בהסתברות 0.3. ההסתברות שיואב מאחר היא 0.18. **א.** מצאו את p . **ב.** ביום מסוים יואב איחר. מה ההסתברות שהוא נסע באוטובוס?

א. לפי נוסחת ההסתברות השלמה, $0.1p + 0.3(1 - p) = 0.18$.

נפתח סוגריים ונקבל $0.3 - 0.2p = 0.18$, ולכן $p = 0.6$.

ב. לפי נוסחת בייס, ההסתברות היא $\frac{0.6 \cdot 0.1}{0.18} = \frac{0.06}{0.18} = \frac{1}{3}$. אף שברוב הימים יואב נוסע באוטובוס, רק שליש מהאיחורים שלו הם בימי אוטובוס.

הידעתם?

תומאס בייס (בערך 1701–1761) היה כומר פרסביטריאני ומתמטיקאי אנגלי. הוא לא פרסם את עבודתו על ההסתברות בחייו. אחרי מותו ערך ידידו ריצ'רד פרייס את החיבור שהשאיר, והחיבור הוקרא בפני החברה המלכותית בלונדון בדצמבר 1763 ופורסם בכתב העת שלה. בייס עצמו לא כתב את הנוסחה בצורה שבה היא מופיעה כאן: הוא פתר מקרה מיוחד שלה. ב-1774, כנראה בלי לדעת על עבודתו של בייס, הגיע המתמטיקאי הצרפתי פייר-סימון לפלס לאותו רעיון. לפלס פיתח אותו הרבה מעבר לבייס, וסיכם את עבודתו על ההסתברות בספר גדול שיצא לאור ב-1812.

15 בדקו את עצמכם

חמש שאלות ✓

1. מטילים שתי קוביות. ידוע שלפחות באחת מהן יצא 6. מה ההסתברות שהסכום הוא 10 לפחות? השוו לתוצאה $\frac{5}{6}$ בנימוק 1.
2. בכיתה הידיעה "לומד פיזיקה" הכפילה את ההסתברות ללמוד מדעי המחשב, וגם הידיעה "לומד מדעי המחשב" הכפילה את ההסתברות ללמוד פיזיקה ($\frac{2}{5}$ הפך ל- $\frac{4}{5}$). הוכיחו שזה לא מקרה: אם $P(A) > 0$ ו- $P(B) > 0$ אז $\frac{P(A/B)}{P(A)} = \frac{P(B/A)}{P(B)}$.
3. ההסתברויות של A ושל B הן $P(A) = 0.5$ ו- $P(B) = 0.4$, וההסתברות של האיחוד היא $P(A \cup B) = 0.7$. האם A ו- B בלתי תלויים? רמז: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
4. מחלה פוגעת ב-2% מהאוכלוסייה. בדיקה מגלה 90% מהחולים, ויוצאת חיובית גם אצל 5% מהבריאים. (א) מה ההסתברות שמי שקיבל תוצאה חיובית חולה? (ב) ומה ההסתברות שמי שקיבל תוצאה שלילית חולה? (רמז: טבלה של 1,000 איש).
5. ענו על השאלה מתיבת "עצרו ובדקו".

תשובות

1. ב- A (באחת הקוביות לפחות יצא 6) יש 11 תוצאות, ומהן 5 נמצאות גם ב- B (סכום 10 לפחות). לכן $P(B/A) = \frac{5}{11}$, הרבה פחות מ- $\frac{5}{6} = P(A/B)$: אותו חיתוך, עולם מצומצם אחר.
2. לפי ההגדרה, גם $\frac{P(A/B)}{P(A)}$ וגם $\frac{P(B/A)}{P(B)}$ שווים ל- $\frac{P(A \cap B)}{P(A) \cdot P(B)}$. בכיתה $P(A \cap B) = \frac{4}{15}$ ו- $P(A) \cdot P(B) = \frac{1}{3} \cdot \frac{2}{5} = \frac{2}{15}$.
3. $P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0.5 + 0.4 - 0.7 = 0.2$, ו- $P(A) \cdot P(B) = 0.5 \cdot 0.4 = 0.2$. המכפלה שווה להסתברות החיתוך, ולכן המאורעות בלתי תלויים.
4. מתוך 1,000 איש: 20 חולים, ו-18 מהם מקבלים תוצאה חיובית; 980 בריאים, ו-49 מהם מקבלים תוצאה חיובית. (א) $\frac{18}{18+49} = \frac{18}{67}$, בערך 27%. (ב) תוצאה שלילית מקבלים 2 חולים ו-931 בריאים, ולכן ההסתברות היא $\frac{2}{933}$, בערך 0.002. תוצאה שלילית כמעט שוללת את המחלה, ותוצאה חיובית רק מחייבת בדיקה נוספת.
5. כש- $P(A/\bar{B}) = 0$ המכנה שווה למונה, ולכן $P(B/A) = 1$. זה הגיוני: אם הבדיקה אף פעם לא יוצאת חיובית אצל בריא, תוצאה חיובית מוכיחה שהנבדק חולה.