

פתרון אפשרי למבחן סיום לקורס "הסתברות וסטטיסטיקה":
תכנית רוטשילד ויצמן,

ענה על כל השאלות. לרשותך שלוש שעות, אך קח בחשבון כי אין צורך להשתמש בכל הזמן המוקצב. **המבחן לא בודק רק את היכולת להציב נתונים בנוסחה.** תשובה מלאה אמורה להתייחס (בקיצור) **הן למודל שלפיו עובדים והן להנחות** המתבקשות לפני שמשמשים בנוסחה. אם השימוש בנוסחה מחייב ידיעה של פרמטר אותו אינך יודע אתה יכול להשתמש בסימן מתמטי תוך הסבר איזה פרמטר חסר. בכל שאלה הסבר מה אתה עושה. תשובה לקונית, למשל, המספר הנכון בלבד, תזכה רק בחלק קטן של הניקוד.

1. הנה בדיחה. אדם נתפס בעלייה למטוס נושא פצצה בתיק שלו. היה לו הסבר. "כוונתי לא הייתה לפוצץ את המטוס", אמר, "אלא ששמעתי שהסיכוי ששני אנשים שאינם מכירים אחד את השני יעלו בו-זמנית פצצות למטוס קטן בהרבה מן הסיכוי שרק אדם אחד יעלה עם פצצה. הבאתי את הפצצה רק כדי להקטין את הסיכוי שהמטוס יפוצץ".
האם הטענה של אדם מתקבלת על הדעת? נמק את טיעונך תוך שימוש מפורש ומדויק של מושגים מתורת ההסתברות.

תשובה: הטענה אינה מתקבלת על הדעת. הסיבה האינטואיטיבית היא שאין קשר בין הסיכוי לפצצה שיביא נוסע אחר לבין העובדה שאותו אדם הביא עמו פצצה. כדי לנסח זאת במסגרת תורת ההסתברות נבחר מרחב מדגם שיבטא את ההסתברות של המאורעות השונים, שנסמן אותה p . נקרא לאדם שבו מדובר אדם. נסמן ב- A את המאורע: אדם מביא פצצה למטוס. נסמן ב- B את המאורע: נוסע אחר מביא פצצה למטוס. הטענה של אדם היא $p(A \cap B) < p(B)$. אי-השוויון נכון אך מכיוון שאדם הביא עמו פצצה, הנתונים להשוואה הם $P(B)$ וההסתברות המותנית $p(B|A) = \frac{P(A \cap B)}{P(A)}$. ההבנה האינטואיטיבית שהזכרנו קודם לכן מתבטאת בכך ש- A ו- B הם מאורעות בלתי תלויים ואת זאת נניח. הניסוח המתמטי של אי-תלות זו הוא $p(A \cap B) = p(A)p(B)$. הצבה של הביטוי הזה ל- $p(A \cap B)$ בביטוי להסתברות המותנית מראה כי ההסתברות הזו שווה ל- $p(B)$, כלומר, הסיכוי שתהייה שתי פצצות במטוס אם אדם מביא פצצה שווה לסיכוי שתהיה פצצה אחת במטוס אם אדם אינו מביא עמו פצצה.

2. מכונית ובה שלושה אנשים: שני גברים ואישה, הייתה מעורבת בתאונה. למשטרה לא ידוע מי הנהג ולכן, מלכתחילה, מעריכה את הסיכויים של כל אחד מהשלושה להיות הנהג כ- $\frac{1}{3}$. בדיקה נוספת של המשטרה העלתה ממצאים נוספים. התגלה כי הנהג נפצע וסוג הדם שלו הוא O . התגלה כי לגבר הראשון סוג דם O . לא ידוע סוג הדם של הגבר השני ושל האישה אבל סטטיסטיקה מהימנה אומרת כי 10% מהגברים ו- 8% מהנשים הם בעלי סוג דם O . האם תוכל ליעץ למשטרה כיצד לעדכן את הערכת הסיכויים של כל אחד מהשלושה להיות הנהג?

תשובה: במרחב מדגם מתאים נסמן ב- A , B ו- C את המאורעות: גבר א' נהג, גבר ב' נהג והאישה נהגה. ההערכה המקורית של המשטרה אומרת $p(A)=p(B)=p(C)=\frac{1}{3}$. נסמן ב- D את המאורע: נמצא כי לנהג יש סוג דם O . ההסתברויות המעודכנות יהיו לכן $p(A|D)$, $p(B|D)$ ו- $p(C|D)$. כדי לחשב אותם נשתמש בחוק בייס. האומר

$$\frac{p(D|A)p(A)}{p(D|A)p(A)+p(D|B)p(B)+p(D|C)p(C)} = p(A|D)$$

וכדומה לגבי ההסתברויות המותנות האחרות כלומר

$$\frac{p(D|B)p(B)}{p(D|A)p(A)+p(D|B)p(B)+p(D|C)p(C)} = p(B|D)$$

$$\frac{p(D|C)p(C)}{p(D|A)p(A)+p(D|B)p(B)+p(D|C)p(C)} = p(C|D)$$

מכאן צריך רק להציב מספרים, כלומר להשתמש בשוויון $p(A)=p(B)=p(C)=\frac{1}{3}$ ובנתונים $p(D|A)=1$, $p(D|B)=0.1$ ו- $p(D|C)=0.08$. התוצאות הן: נהג א' 0.847, נהג ב' 0.0847, האישה 0.0677 (ההסתברויות לא מסתכמות ל- 1 בגלל שגיאת עיגול).

3. א. נסח במדויק את החוק החלש ואת החוק החזק של המספרים הגדולים.
 ב. הראה כי חוק המספרים הגדולים (החלש או החזק) אינו נכון ללא הנחת אי-תלות בין המשתנים המקריים.

תשובה:

א. יהיה Ω, Σ, p מרחב הסתברות ויהיו $X_1, X_2, X_3, \dots, X_i$ משתנים מקריים שווים התפלגות ובלתי תלויים מהמרחב לישר הממשי כך שהתוחלת של כל אחד מהם מוגדרת. מכיוון שהם שווים התפלגות נסמן את התוחלת המשותפת ב- μ . נסמן $Y_n = \frac{1}{n}(X_1 + \dots + X_n)$ כלומר, Y_n הוא הממוצע של n המשתנים המקריים הראשונים. החוק החלש של המספרים הגדולים אומר כי לכל $\alpha > 0$ ההסתברות של המאורע $\{|Y_n - \mu| > \alpha\}$ שואפת לאפס כאשר n שואף לאינסוף. החוק החזק של המספרים הגדולים אומר כי פרט לקבוצת מדגם של אברים ω ב- Ω שמידתה אפס, הערכים $Y_n(\omega)$ שואפים ל- μ כאשר n שואף לאינסוף.

ב. נבחר מרחב מדגם של שני אברים ω_1, ω_2 ומשתנה מקרי X המוגדר $X(\omega_1) = 0$ ו- $X(\omega_2) = 1$. נגדיר X_i להיות שווה ל- X לכל $i = 1, 2, 3, \dots$. אז הממוצע שהוגדר בחוק המספרים הגדולים מקיים $Y_n = X$ לכל n ובפרט לא שואף (גם במובן החלש וגם בחזק) לתוחלת של X שהיא $\frac{1}{2}$. זו דוגמה נגדית למשפט כאשר מורידים את אי התלות.

4. בישראל, לה כ- 8 מיליון תושבים, מדגם של 500 אנשים מספיק כדי לאמוד את שיעור התמיכה במפלגה גדולה עם טעות של $\pm 2\%$, וזאת ברמת מובהקות של 95% (כלומר ב- 95% מהפעמים האומדן יהיה נכון).
 א. מה גודל המדגם הנדרש כדי להשיג אותה רמת טעות ואותה רמת מובהקות בארצות הברית (לה כ- 320 מיליון אנשים).
 ב. מה גודל המדגם הנדרש בישראל כדי להשיג רמת טעות של $\pm 1\%$ ברמת מובהקות של 99%?

תשובה:

א. גודל המתאם הדרוש אינו תלוי כלל בגודל האוכלוסייה (כפי שניתן לראות גם מהתשובה לחלק ב. של השאלה). לכן גם בארצות הברית יספיקו 500 איש.
 ב. רמת הטעות נקבעת על ידי $\pm z \frac{\sigma}{\sqrt{n}}$ כאשר z היא סטיית התקן של המשתנה הנורמלי $N(0,1)$ המביאה לרמת המובהקות הדרושה ו- σ היא סטיית התקן של הדגימה הבודדת (המקורבת על ידי סטיית התקן של המדגם אבל למטרות השימוש אפשר להניח כי אלה שוות). לפי המספרים שנתנו בכיתה סטיות התקן של המשתנה הנורמלי $N(0,1)$ המביאה לרמת מובהקות של 95% היא 1.96 וזו המביאה לרמת מובהקות של 99% היא 2.324. אנו רוצים לקבל מחצית מרמת הטעות על ידי הגדלת גודל המדגם n נאמר ל m תוך העלאת רמת המובהקות. כלומר צריך להתקיים $\frac{1.96}{\sqrt{500}} = 2 \frac{2.324}{\sqrt{m}}$. מכאן חשבון פשוט אומר כי $m = 2812$.

5. ניסיון רב שנים מראה כי מתוך 40,000 הצופים המגיעים למגרש הכדורגל של קבוצת ארסנל מדי שבועיים, מספר המבקשים לקנות לחמנייה עם נקניקיה מתפלג קרוב מאוד להתפלגות נורמלית, עם ממוצע 2000 ושונות 100. הזכייין של אספקת הלחמניות עם נקניקיה התחייב לספק את הביקוש ב- 95% של המשחקים.
 א. כמה לחמניות עליו להכין מדי שבועיים כדי לעמוד בהתחייבות?
 ב. כמה לחמניות יהיה עליו להכין כדי לספק את הביקוש ב- 100% של המשחקים?
 ג. האם סביר כי הזכייין יניח שהביקוש ללחמניות נעשה על ידי כל צופה וצופה באופן בלתי תלוי אחד בשני ובהסתברות אחידה?

תשובה:

א. מספר הלחמניות, נסמן אותו ב- x , הדרוש כדי להשיג את היעד הוא כזה שאחרי שההתפלגות מנורמלת על ידי הביטוי $\frac{x-2000}{\sqrt{100}}$ מקבלים סטיית תקן z כזו שההסתברות לקבל ערך גדול מ- z במשתנה $N(0,1)$ היא רק 5%. הערך הזה ניתן בכיתה והוא 1.6449. מכאן אפשר לחשב מהו x והתוצאה היא שמספיק להכין 2017 לחמניות כדי לספק את הביקוש ב- 95% של המשחקים.

- ב. כדי לספק את הביקוש ב- 100% של המפגשים יש להכין 40,000 לחמניות כל פעם (מכיוון שיתכן, אמנם בהסתברות נמוכה, כי כל הצופים ירצו לקנות לחמנייה עם נקניקיה).
- ג. אם הביקוש ללחמניות היה מתבצע על ידי כך שהצופים היו מחליטים באופן בלתי תלוי לקנות או לא לקנות לחמנייה באותה הסתברות הרי, מכיוון שהסתברות זו קובעת את התוחלת, ההסתברות הזו צריכה הייתה להיות שווה ל- $\frac{2000}{40,000}$ כלומר $\frac{1}{20}$. השוואת אזה הייתה $1900 = 40,000 \left(\frac{1}{20} \left(1 - \frac{1}{20} \right) \right)$ ולא $\sigma^2 = 100$ ננתן. לכן המנגנון המוצע אינו סביר.

6. א. הגדר מהי התפלגות פואסונית.
- ב. הסבר את הקשר שבין מספר גדול של ניסיונות ברנולי שלכל אחד סיכוי הצלחה נמוך לבין התפלגות פואסונית.
- ג. הסבר מדוע התפלגות פואסונית לה תוחלת קטנה אינה קרובה להתפלגות נורמלית בעוד שאם התוחלת שלה גדולה היא קרובה להתפלגות נורמלית.
- ד. תן דוגמא מהחיים למשתנה מקרי שסביר כי יתפלג פואסונית.

תשובה:

- א. התפלגות פואסונית עם פרמטר λ נותנת הסתברות חיובית למספרים הטבעיים $0, 1, 2, 3, \dots$ כאשר ההסתברות של k היא $e^{-\lambda} \frac{\lambda^k}{k!}$.
- ב. ההתפלגות של המשתנה המקרי המתקבל על ידי סיכום של n חזרות על ניסיון ברנולי של קבלת 1 ו-0 בהסתברות p ו- $(1-p)$ בהתאמה שואפת – כאשר n שואף לאינסוף ו- p שואף ל-0 ו- np שואף ל- λ – להתפלגות פואסון עם אותו λ .
- ג. כאשר מתסכלים על הביטוי $e^{-\lambda} \frac{\lambda^k}{k!}$ המציג את ההסתברות של k רואים שהמקסימום שלו מתקבל פחות או יותר כאשר k קרוב לפרמטר λ , עד אז ערך ההסתברות עולה ואחר כך יורד באופן מונוטוני. כך שאם λ קטן ההתפלגות רחוקה מלהיות סימטרית סביב התוחלת, בעוד שאם λ גדול יש קרבה גדולה יותר להתפלגות נורמלית. הפרמטר λ הוא גם התוחלת של המשתנה הפואסוני.
- ד. בהנחה שתאונות מטוסים מתרחשות באופן בלתי תלוי זו בזו, יש אלפי טיסות כל יום וההסתברות לתאונה קטנה, הרי שלמשתנה המקרי: מספר תאונות המטוס בכל יום (ואתמול ארעה כזו תאונה, מטוס קוריאני התרסק בנחיתה בשדה התעופה בסן פרנסיסקו), סביר שיהיה משתנה פואסוני.

נקווה שהייתה הצלחה