

פתרון אפשרי למבחן סיום שניתן בעבר בקורס "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 שמשתמשים 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. הצגת עובדות לא רלבנטיות לא תעלה, ועלולה אף להוריד, את הציון.

1. נתון כי הסיכויים להולדת בן ולהולדת בת שווים. מספרים לך כי למשפחה מסוימת שנבחרה באופן מקרי יש שני ילדים וביניהם בת. שואלים אותך מה הסיכוי שיש לה גם בן? וכן שואלים האם הסיכוי ישתנה אם צלצלת בפעמון הדירה של המשפחה ופותחת את הדלת בת?
לגבי כל אחת מהשאלות הראה כי הנחות שונות לגבי המאורע האקראי מביאות לתוצאות שונות. נסח את ההנחות האלה במדויק תוך התייחסות למרחב מדגם ו/או למאורעות מקריים.

תשובה: בכל המקרים מרחב המדגם יהיה כל הזוגות של ילד-ילדה במשקל שווה ובכל זוג כל בן-בת במשקל שווה, סך הכול 16 אברים. המשתנים המקריים נקבעים לפי הגרסאות הבאות. לשאלה הראשונה: גרסה א': אם למשפחה יש בת מספרים זאת. קל לראות שהסיכוי שאז השני הוא בן הוא $\frac{2}{3}$. גרסה ב': בוחרים ילד מהמשפחה באקראי ומודיעים את מינו. הסיכוי שהשני הוא בן הוא $\frac{1}{2}$. לשאלה השנייה: גרסה א': תמיד פותח את הדלת הילד המבוגר יותר. הסיכוי כי השני הוא בן הוא $\frac{1}{2}$. גרסה ב': אם יש בת בבית תפקידה לפתוח את הדלת (ואם יש שתיים אחת מהן פותחת). הסיכוי שאז הילד שלא פתח הוא בן הוא $\frac{2}{3}$.

2. בצומת דרכים עוברות אלפי מכוניות כל יום ומדי פעם יש תאונות. מצאו כי התפלגות מספר התאונות לשבוע היא פואסונית וממוצע התאונות לשבוע הוא 1. מה הסיכוי כי בשבוע מסוים מספר התאונות יהיה גדול (או שווה) משתי סטיות תקן מעל הממוצע?

תשובה: התפלגות פואסונית מוגדרת על ידי פרמטר λ ונותנת הסתברות חיובית למספרים הטבעיים (כולל 0) כאשר ההסתברות של k היא $e^{-\lambda} \frac{\lambda^k}{k!}$. הפרמטר הוא גם התוחלת וגם השונות של ההתפלגות. אצלנו $\lambda = 1$ ולכן הממוצע הוא 1 וסטיות התקן שווה ל-1. גדול או שווה לשתי סטיות תקן פירושו הסיכוי ש- k גדול או שווה ל-3. כדי לחשב את הסיכוי הזה צריך לסכם את האברים $e^{-1} \frac{1}{k!}$ עבור k גדול או שווה ל-3 ו לחילופין לחשב

$$0.08 \text{ (בערך)} = 1 - e^{-1} \left(1 + 1 + \frac{1}{2}\right)$$

3. א. הסבר מדוע כדי לטעון ששני משתנים מקריים הם בלתי תלויים עליהם להיות מוגדרים על אותו מרחב מדגם.
ב. הסבר מדוע כדי לטעון כי לשני משתנים מקריים עם ערכים ממשיים יש אותה התפלגות אין צורך כי יהיו מוגדרים על אותו מרחב מדגם.

תשובה: א. שני משתנים מקריים X ו- Y הם בלתי תלויים אם לכל שני אינטרולים ממשיים I ו- J מתקיים $P(X^{-1}(I) \cap Y^{-1}(J)) = P(X^{-1}(I))P(Y^{-1}(J))$ כאשר P היא ההסתברות על מרחב המדגם. לפי צד שמאל של ההגדרה צריכים הם להיות מוגדרים על אותו מרחב מדגם.
ב. ההתפלגות של משתנה מקרי מוגדרת על מרחב הטווח שלו, ואפשר לכן להשוות התפלגויות של שני משתנים מקריים אפילו אם הם מוגדרים על מרחבי מדגם שונים.

4. יהיה (p, Σ, Ω) מרחב הסתברות ויהיו $X_i, i = 1, 2, 3, \dots$ משתנים מקריים שווי התפלגות ובלתי תלויים מהמרחב לישר הממשי כך שהתוחלת של כל אחד מהם מוגדרת. נסמן את התוחלת המשותפת באות μ . נסמן $Z_n = \frac{1}{n}(X_1 + \dots + X_n)$ כלומר, Z_n הוא הממוצע של n המשתנים המקריים הראשונים.

- א. נסח במדויק (תוך התייחסות לנתון לעיל) את החוק החלש ואת החוק החזק של המספרים הגדולים והסבר באיזה מובן החוק החלש חלש יותר מאשר החוק החזק.
 ב. האם המשתנים המקריים Z_n הם בלתי תלויים?

תשובה: א. החוק החלש של המספרים הגדולים אומר כי לכל $\alpha > 0$ ההסתברות של המאורע $\{|Z_n - \mu| > \alpha\}$ שואפת לאפס כאשר n שואף לאינסוף (כאשר μ היא התוחלת). החוק החזק של המספרים הגדולים אומר כי פרט לקבוצת מדגם של אברים ω ב- Ω שמידתה אפס, הערכים $Z_n(\omega)$ שואפים ל- μ כאשר n שואף לאינסוף. סדרת משתנים מקריים מקיימת את התנאי של החוק החזק מקיימת גם את התנאי של החוק החלש. זה קל לראות כאשר בוחנים את סדרת המאורעות $\{|Z_n - \mu| > \alpha\}$ שהיא סדרה עולה של קבוצות שאיחוד האברים בה הוא כל המרחב. אבל סדרת המקיימת את תנאי החוק החלש לא מקיימת בהכרח את התנאי של החוק החזק. דוגמא נגדית תהיה הסדרה Y_n על אינטרוול היחידה שערכיה הם 0 למעט על אינטרוול מאורך $\frac{1}{n}$ והמתחיל בקצה הימני (מודולו 1) של האינטרוול שלפניו.

ב. למעט במקרה שהסדרה היא של משתנים מקריים קבועים סדרת הממוצעים אינה בלתי תלויה, כפי שקל לראות כבר במקרה של משתני ברנולי. שם המאורעות $A_n = \{1 = Z_n\}$ תלויים זה בזה כי הסיכוי של A_n בהינתן $A_{(n-1)}$ הוא חצי, הרבה יותר מאשר הסיכוי של A_n עצמו.

5. א. הצג במדויק (תוך התייחסות למרחב מדגם) את אי-שוויון צ'בישב.
 ב. הראה כי באי-שוויון צ'בישב אי אפשר להשיג הערכה טובה יותר כל עוד אי-השוויון מתייחס להתפלגות כל שהיא.
 ג. הראה כי כאשר מתייחסים להתפלגויות מסוימות, ההערכות שאי-שוויון צ'בישב נותן ניתנות לשיפור.

תשובה: א. יהיה (p, Σ, Ω) מרחב ההסתברות ועליו משתנה מקרי X , עם תוחלת μ וסטיית תקן s . אי שוויון צ'בישב הוא $p(\{|X - \mu| \geq ks\}) \leq 1/k^2$
 ב. דוגמא לכך שאי-אפשר לשפר את אי-השוויון: מרחב מדגם ועליו משתנה מקרי ממשי שהתפלגותו:
 $p(1) = 1/2a^2, p(-1) = 1/2a^2$ ושאר המשקל בנקודה 0 (הדוגמא ניתנה בכיתה).
 ג. במשתנה מקרי שערכו מספר קבוע צד ימין באי-השוויון הוא אפס ולא חיובי.

6. תכשיט יקר נגנב ממוזיאון בעיר שבה מיליון וחצי תושבים וידוע כי הגנב הוא תושב העיר. הגנב לא היה זהיר והשאיר טביעת DNA כזו שידוע כי ההסתברות שלאזרח תהיה טביעת DNA זהה היא 10^{-7} . המשטרה בדקה 100.000 תושבים באופן אקראי אחד אחרי השני עד שמצאה אזרח עם טביעת DNA זהה. מה ההסתברות שזה אמנם הגנב?

תשובה: תוחלת מספר אלה בעלי ה- DNA המבוקש מבין 1.400.000 התושבים שלא נבדקו היא 1.400.00 $\cdot 10^{-7}$ חלקי 10, או 0.14. לכן הסיכוי שהאדם בעל ה- DNA שכבר זוהה הוא הגנב שווה רק ל- $1/1.14$ (שווה בערך ל- 87 אחוזים)

מקווה שהייתה בהצלחה
 שנה טובה