

תרגיל 7

- יש לענות על כל השאלות.
- הגשת התרגילים היא בזוגות קבועים.
- שאלות המסומנות ב-** הן שאלות רשות.
- את התרגילים יש להגיש לתיבת ההגשה במודל.
- יש להגיש הכל בקובץ PDF אחד, למשל על ידי שימוש ב Rmarkdown.

1. חוקרים במשרד הבריאות רצו לבדוק מה שיעור צריכת סמים קלים בקרב בני נוער. לצורך המחקר, 1500 נערים ונערות נסקרו ונשאלו כמה פעמים צרכו סמים בחודש האחרון. נסמן ב n_i את מספר המשיבים שענו כי עישנו i פעמים.

את תוצאות הסקר ניתן לסכם בטבלה שנמצאת במודל.

- (1) נהוג למדל משתנה הסופר אירועים על ידי התפלגות פואסון. הניחו שהנתונים אכן מגיעים מהתפלגות פואסון. מצאו אנ"מ לפרמטר של ההתפלגות ובדקו גראפית האם ההתפלגות הפואסונית עם האומדן שמצאתם אכן מתאר טוב את הנתונים.
- (2) יותר ריאלי להניח שהנסקרים נחלקים בגדול ל-3 קבוצות:
- קבוצה אחת, אשר תמיד תדווח 0 מקרים, גם אם זה אינו המצב האמיתי. הניחו כי נבדק שייך לקבוצה זו בהסתברות α .
 - קבוצה שנייה, אשר מייצגת התנהגות ממוצעת, ומאופיינת על ידי התפלגות פואסון עם פרמטר λ_1 . לנבדק הסתברות β להשתייך לקבוצה זו.
 - קבוצה שלישית, אשר מייצגת התנהגות חריגה, של בני נוער הצורכים סמים לעיתים תכופות. הניחו כי התנהגות זו מאופיינת על ידי התפלגות פואסון עם פרמטר λ_2 , וההסתברות של נבדק להשתייך לקבוצה זו הינה $1 - \alpha - \beta$.
- א. רשמו את פונקציית ההסתברות של נבדק אשר דיווח על i אירועים.
- ב. הביעו את פונקציית הנראות על ידי הנתונים הנצפים (שימו לב שאין ברשותכם הדיווח של כל נבדק ונבדק, אלא רק שכיחויות האירועים). הסבירו מדוע חישוב אנ"מ על ידי פונקצייה זו הינו קשה.
- (3) מצאו אנ"מים לארבעת הפרמטרים $\alpha, \beta, \lambda_1, \lambda_2$ על ידי אלגוריתם EM (כולל מימוש ב-R על הנתונים ex7q1). חישבו היטב מהם הנתונים החסרים המתאימים שבאמצעותם פונקציית הנראות של הנתונים המלאים מאפשרת אמידה יותר נוחה לפרמטרים.

2.

יהיו \vec{X}, \vec{Y} שני וקטורים ב"ת, כל אחד בן 10000 תצפיות, כך ש:

$$\forall i \ X_i \sim N(\mu_1, \sigma^2), Y_i \sim N(\mu_2, \sigma^2)$$

עם זאת, בפועל איננו צופים בוקטורים אלו, אלא אנו צופים בוקטור \vec{O} כך ש: $O_i = \max(X_i, Y_i)$.

בנוסף, ידוע לנו שעבור $i=1, \dots, m$ מתקיים $X_i \geq Y_i$ ועבור היתר $Y_i > X_i$.

לשם פשטות, הניחו כי σ^2 ידועה.

א. כתבו את פונקציית הנראות לפי הנתונים הנצפים (\vec{O}). האם ניתן לקבל ביטויים סגורים עבור הפרמטרים? מיצאו אנ"מ על ידי שימוש בניוטון-רפסון (כולל מימוש ב-R על הנתונים ex7q2, בנתונים אלו $m=7659$ וכן $\sigma^2 = 4$).

ב. מצאו אנ"מ לפרמטרים על ידי שימוש באלגוריתם EM (כולל מימוש על הנתונים). בפיתוחים שלכם היעזרו בתוצאות הידועות עבור התפלגות נורמלית קטומה, והשוו לתוצאות מסעיף א.

https://en.wikipedia.org/wiki/Truncated_normal_distribution

ג. מצאו אנ"מ בצורה נאיבית: על ידי כתיבת פונקציית הנראות כך ש- m התצפיות הראשונות מגיעות מהתפלגות נורמלית 1 והיתר מהתפלגות נורמלית 2. האם האומדים מוטים? לאיזה כיוון ההטייה? מדוע זה הכיוון?

**** (שאלת רשות)**

3.

נחזור אל הבעיה שבשאלה 2, אולם כעת:

$$X \sim \text{pois}\left(\frac{1}{p}\right), \quad Y \sim \text{geom}(p) \quad x, y \in \{0, 1, 2, \dots\}$$

ואנו צופים בוקטור \vec{O} המוגדר כך:

$$O_i = \begin{cases} X_i & \text{if } X_i \geq Y_i \\ Y_i & \text{if } X_i < Y_i \end{cases}$$

א. כתבו את פונקציית הנראות של הנתונים הנצפים אם:

(1) ידוע לנו שעבור $O_i = X_i \quad i=1,...,m$ ועבור יתר התצפיות $O_i = Y_i$.

(2) הידע שבסעיף (1) אינו זמין לנו.

(3) האם ניתן למצוא אנ"מ בצורה אנליטית תחת שני התרחישים לעיל?

ב. מיצאו אנ"מ ל- p על ידי EM תחת שני התרחישים המתוארים בסעיפים

א(1) ו-א(2).

לצורך מענה על השאלות, השתמשו בקובץ הנתונים **ex7q3** (בנתונים אלו $m=7252$).