מתרגל: ניר קרת
סמסטר ב' תשפ"א

חישוב סטטיסטי - תרגיל 2

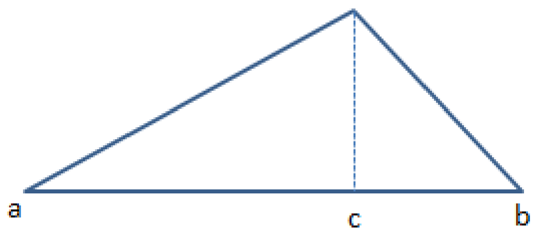
- יש לענות על כל השאלות.
- הגשת התרגילים היא **בזוגות**.
- שאלות המסומנות ב-** הן שאלות רשות.
- את התרגילים יש להגיש לתיבת ההגשה במודל.
- יש לצרף שני קבצים: קובץ קוד R וקובץ עם חישובים ופתרונות לשאלות, לפי הנדרש (אם בכתב יד, אז **ברור**). ניתן גם להגיש הכל יחד בקובץ אחד, למשל על ידי שימוש ב Rmarkdown.
- בתרגיל זה אין להגריל משתנים מאף התפלגות פרט להתפלגות אחידה סטנדרטית, אלא אם צוין אחרת..
- בתרגיל זה אין להשתמש בשיטת accept-reject.

1) צרו פונקצייה ב-R הדוגמת מכל אחת מההתפלגויות הבאות, ופרטו חישוביכם. בנוסף, הגרילו 10000 תצפיות מכל התפלגות, צרו היסטוגרמה והדפיסו את פלט פונקציית summary של וקטור התצפיות.

1. $f(x) = \frac{1}{-lnp} \cdot \frac{\beta(1-p)e^{-\beta x}}{1-(1-p)e^{-\beta x}} \quad x \in [0,\infty), \beta > 0, p \in (0,1)$

For sampling: $\beta = 2, p = 0.5$

2. The plot of the density function is:



$a \in (-\infty,+\infty), a < b, \ c \ge a, c \le b$

For sampling: a=3, c=4, b=5

3. $f(x) = abx^{a-1}(1-x^a)^{b-1} \quad a > 0, b > 0, x \in [0,1]$

For sampling: a=3, b=3

4. $f(x) = \frac{1}{\pi\sqrt{x(1-x)}} \quad x \in (0,1)$

5.

א .Fréchet distribution:

$$F(x) = e^{-\left(\frac{x-m}{s}\right)^{-\alpha}} \quad \alpha, s \in (0,\infty), m \in (-\infty,+\infty)$$

For sampling: $\alpha = 3, s = 2, m = 5$

S - פרמטר כיול (scale), m פרמטר מיקום, α פרמטר צורה.

ב. בהינתן דגימה מ Fréchet(α,1,0) כיצד נוכל להשיג דגימה מ Fréchet(α,s,m) מבלי לדגום מחדש?

2) יהיו $X\sim\exp(\mu)$ ו - $Y\sim\exp(\theta)$ שני מ"מ ב"ת. נגדיר $Z = X - Y$.

א. הגרילו מדגם בן 10000 תצפיות מהתפלגות *Z* בשיטה "נאיבית" (על ידי הגרלה של מ"מ מעריכיים וחישוב ההפרש ביניהם).

ב. הגרילו מדגם בן 10000 תצפיות מהתפלגות *Z* על ידי *ITS*. לצורך כך יהיה עליכם למצוא את ה*CDF* של *Z*.

רמז: כאשר תחשבו את $P(Z \le t)$ הפרידו למקרים $t \le 0$ ו- $t > 0$.

ג. השוו באמצעות qqplot בין המדגמים שקיבלתם בסעיפים א' ו-ב'.

לצורך הדגימות בשאלה זו קחו את הפרמטרים להיות $\mu = 0.5, \theta = 0.2$.

3)

א. כתבו פונקצייה ב *R* הדוגמת מהתפלגות המקסימום של *n* מ"מ מהתפלגות מעריכית עם פרמטר θ כלשהו. פרטו חישוביכם.

דגמו 10000 תצפיות מהתפלגות המקסימום של 30 תצפיות מהתפלגות מעריכית עם פרמטר 2. צרו הסטוגרמה והדפיסו את פלט ה *summary*.

ב. <u>הניחו שיש לכם דרך פשוטה לדגום מהתפלגות בטא</u>. הציעו דרך כללית לדגום מההתפלגות של סטטיסטי הסדר ה- *k* ממדגם בן *n* תצפיות של התפלגות <u>כלשהי</u>, אשר ממנה ניתן לדגום על ידי *Inverse transformation*. נמקו.

ג. האם התוצאה בסעיף א' עולה בקנה אחד עם הדרך שהצעתם בסעיף ב' (השוו מבחינת הביטוי המתמטי שקיבלתם בסעיף א')?

4)

א. יהי $Y\sim\exp(\lambda)$ ויהי $Z = \lceil Y \rceil$ (פונקציית *ceiling*, עיגול כלפי מעלה). הוכיחו:

$Z\sim geom\left(1 - e^{-\lambda}\right)$. (פתחו ביטוי מתמטי עבור $P(Z = z)$ והראו כי ההתפלגות היא גיאומטרית).

ב. הסבירו על סמך תוצאה כזו, כיצד הייתם דוגמים מ"מ גיאומטרי עם פרמטר *p*.

ג. השוו לתוצאה שראיתם בכיתה עבור דגימה מהתפלגות גיאומטרית (שקף 22 מצגת 1) והסבירו את הקשר.

5**) התפלגות Lomax היא מקרה פרטי של התפלגות *pareto*. פונקציות הצפיפות (*PDF*) והפה"מ (*CDF*) של ההתפלגות הן:

$$PDF(x; scale = k, shape = \theta) = \frac{\theta}{k}[1 + \frac{x}{k}]^{-(\theta+1)}$$

$$CDF(x; scale = k, shape = \theta) = 1 - [1 + \frac{x}{k}]^{-\theta}$$

א. כתבו פונקצייה ב-*R* הדוגמת מהתפלגות זו על ידי *inverse transformation sampling*. דגמו 10000 תצפיות מההתפלגות עם הפרמטרים *scale=3*, *shape=4*. צרו הסטגורמה והדפיסו את פלט ה *summary*.

ב. הוכיחו כי התפלגות זו מתקבלת כ *compound distribution* של התפלגות מעריכית עם פרמטר קצב λ, כך ש:

$$\lambda \sim gamma(scale = \frac{1}{k}, shape = \theta)$$

השתמשו בתוצאה שהוכחתם על מנת לדגום 10000 משתנים נוספים מהתפלגות זו. הדפיסו את ההיסטוגרמה ואת פלט ה *summary*. לצורך סעיף זה, מותר לכם להשתמש בפונקצייה *rgamma*.

6**)

בשאלה זו נראה כיצד ניתן לדגום מכל התפלגות בדידה בעלת תומך סופי על ידי התפלגות תערובת מתאימה.

וביתר פירוט: נניח ו-*X* מ"מ בדיד עם תומך של *n* ערכים שונים אשר יוסמנו כ $\{x_1, \dots, x_n\}$. נראה כי נוכל להביע את וקטור מסות ההסתברות של *X*, אשר יוסמן *P*, באופן הבא:

$$P = \frac{1}{n-1}\sum_{k=1}^{n-1} Q^{(k)}$$

כאשר כל *Q* הינו וקטור של מסות הסתברות הנותן מסה חיובית לשני ערכים בלבד מתוך *n* הערכים בתומך של *X*, ו-0 ליתר.

נסמן ב $p_1, \dots, p_n$ את הערכים בוקטור *P* (כלומר את ההסתברויות של *X* לקבל את הערכים $x_1, \dots, x_n$ בהתאמה).

א. הסבירו מדוע בהכרח קיים *i* כלשהו $1 \le i \le n$ כך שמתקיים $p_i \le \frac{1}{n-1}$.
ב. הסבירו מדוע בהכרח עבור אותו *i* מסעיף א' קיים בהכרח $j \neq i, j$ המקיים: $p_i + p_j \ge \frac{1}{n-1}$.

כעת נגדיר את $Q^{(1)}$ להיות התפלגות אשר נותנת מסה רק לשני ערכים: x_i, x_j. בנוסף נדרוש כי כל המסה של הערך x_i תהיה מרוכזת אך ורק ב-$Q^{(1)}$, ולכן חייב להתקבל ש:

$Q_i^{(1)} = (n-1)p_i$ (לפי סעיף א' לעיל מובטח לנו כי ערך זה לא יהיה גדול מ-1).

מכיוון שכל Q נותנת מסה רק לשני ערכים, חייב להתקבל: $Q_j^{(1)} = 1 - (n-1)p_i$.

בשלב הבא, נבטא את הוקטור P באופן הבא:

$$P = \frac{1}{n-1}Q^{(1)} + \frac{n-2}{n-1}P^{*(1)}$$

כאשר $P^{*(1)}$ הינו וקטור הסתברויות אשר יחושב כך שזהות זו תהיה נכונה, כלומר:

$$P_i^{*(1)} = 0, \qquad P_j^{*(1)} = \frac{n-1}{n-2}\left(P_j - \frac{1}{n-1}Q_j^{(1)}\right) = \frac{n-1}{n-2}\left(P_j + P_i - \frac{1}{n-1}\right)$$

$$P_k^{*(1)} = \frac{n-1}{n-2}P_k \quad k \neq i,j$$

ג. הסבירו מדוע $P^{*(1)}$ שהתקבלה בהכרח יוצרת התפלגות תקנית (כלומר הראו ש סכום ההסתברויות מסתכם ל-1 וכן הסבירו מדוע לא ייתכן שתתקבל הסתברות שלילית).

כעת באופן רקורסיבי נוכל להשתמש באותו פירוק עבור $P^{*(1)}$, אשר הינה התפלגות הנותנת מסה חיובית ל *n-2* ערכים בלבד (שכן לא נשארה מסה לx_i). נוכל אם כן להמשיך באופן הבא:

$$P = \frac{1}{n-1}Q^{(1)} + \frac{n-2}{n-1}\left(\frac{1}{n-2}Q^{(2)} + \frac{n-3}{n-2}P^{*(2)}\right)$$

וכן הלאה עד אשר בסופו של דבר יתקבל:

$$P = \frac{1}{n-1}Q^{(1)} + \cdots + \frac{1}{n-1}Q^{(n-1)}$$

המסקנה: נוכל לבחור באופן אחיד את אחת מההתפלגויות Q, ואז נדגום ערך מהתפלגות זו (הנותנת מסה לאחד משני ערכים), כדי לקבל דגימה מ-P.

ד. ישמו שיטה זו על ההתפלגות הבינומית *B(n=3,p=0.4)*, ודיגמו 10000 תצפיות. בנוסף, דיגמו מהתפלגות זו בשיטה אשר למדתם בכיתה (סכום של מ"מ ברנולי), והשוו את שני המדגמים על ידי השוואת הפרופורציות של הערכים שהתקבלו.