מתרגל: ניר קרת
סמסטר ב' תשפ"א

חישוב סטטיסטי - תרגיל 3

- יש לענות על כל השאלות.
- הגשת התרגילים היא **בזוגות קבועים**.
- 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 ורצוי להגיש הכל יחד בקובץ אחד, למשל על ידי שימוש ב Rmarkdown.

1) בשאלה זו תדגמו מהתפלגות beta(3,b) על ידי rejection sampling כאשר b=0.5, 4.

א. חשבו מהו c עבור התפלגות מציעה אחידה סטנדרטית לכל אחת מהאפשרויות ל-b.

ב. חשבו מהו c עבור התפלגות Kumaraswamy (ההתפלגות מסעיף 1(3) מתרגיל הבית הקודם), כאשר הפרמטרים שלה זהים לאלו של התפלגות המטרה, עבור כל אחת מהאפשרויות ל-*b*.

ג. עבור *b=0.5,4* הגרילו 10000 תצפיות מהתפלגות בטא, לפי השיטה העדיפה מבין סעיפים א ו-ב לעיל. הדפיסו היסטוגרמה ואת פלט ה*summary*.

2) אנו מעוניינים לדגום מהתפלגות נורמלית סטנדרטית. קבעו מאיזו התפלגות נעדיף לדגום משיקולי תוחלת מספר הדגימות הנדרשות. ההתפלגויות האפשריות הן קושי *Cauchy(location=0,scale=s)* וכן *double exponential(location=0,scale=s)*

(עליכם למצוא את *s* האופטימלי לכל התפלגות).

3) כפי שלמדנו בכיתה, כאשר דוגמים בשיטת *acc-rej*, הבחירה האופטימלית עבור *C* הינה $C^* = sup_x \frac{f(x)}{g(x)}$. עם זאת, לעיתים מקסום יחס הצפיפויות $\frac{f(x)}{g(x)}$ היא פעולה קשה אנליטית. פתרון אחד הוא שיטות נומריות (נלמד כמה בהמשך הקורס). פתרון חלופי הוא הפתרון הבא:

נתחיל עם C=1, אשר הינו החסם התחתון עבור C*. כעת נתחיל לבצע acc-rej כרגיל: נדגום נקודות מההתפלגות המציעה G, נחשב את היחס $\frac{f(x)}{g(x)}$ ונקבל נקודות בהן היחס הנ"ל גדול יותר מ-U*C (U מתפלג אחיד סטנדרטי). עם זאת, ברור שמכיוון שאנו משתמשים ב C קטן מדיי, נשמרים ערכים שאסור היה לשמור, ולכן נקבל מדגם שאינו מההתפלגות הרצויה. לאחר דגימת תצפית חדשה, נקבל שיחס הצפיפויות בה יהיה גדול יותר מ-1, ובכך מצאנו לנו חסם תחתון חדש עבור C*. ככל שנמשיך לדגום עוד נקודות, נתקרב למציאת C בסביבת C* האופטימלי. מדי פעם נרצה לבצע תיקון על המדגם ששמרנו, בהתאם לחסם התחתון המעודכן לאותו רגע. הסבירו איזה תיקון יש לבצע על וקטור התצפיות אם:

1. שמרנו את ערכי u בהם השתמשנו לצורך ביצוע מבחן קבלה-דחייה.

2. לא שמרנו את ערכי u אלו.

בסופו של התהליך נקבל התפלגות מקורבת להתפלגות המטרה אך לא זהה לה, כיוון שנוכל רק לשאוף אל *C מלמטה, ולכן השיטה אינה שמרנית, במיוחד אם לא דוגמים מספר רב של תצפיות. אתם יכולים לנסות לחשוב בעצמכם על דרכים חלופיות אשר יהיו יותר שמרניות מחד, אך יעילות מאידך. (זהו תרגיל מחשבתי ואינכם נדרשים לענות עליו במסגרת התרגיל).

**4)

א. עבור התפלגות מטרה רציפה, הסבירו מדוע לא ניתן להשתמש בהתפלגות מציעה בדידה ללא עריכת תיקון לאחת ההתפלגויות.

ב. עבור התפלגות מטרה בדידה, הסבירו מדוע לא ניתן להשתמש בהתפלגות מציעה רציפה ללא עריכת תיקון לאחת ההתפלגויות.

ג. בסעיף זה תדגמו ערכים מהתפלגות בינומית על ידי התפלגות *Cauchy* מתוקנת. על מנת לעשות כן, עליכם לבצע מספר פעולות:

1. ראשית, עליכם לקטום את ההתפלגות. התומך של התפלגות קושי הינו כל הישר הממשי, בעוד שההתפלגות הבינומית מקבלת ערכים מ-0 ועד *n*. אמנם שלב זה אינו הכרחי, אבל הוא יחסוך הגרלות מיותרות של ערכים מחוץ לתומך של ההתפלגות הבינומית. מצאו את פונקציית הצפיפות והפה"מ של ההתפלגות הקטומה. (שימו שאופן הקטימה הרצוי יהיה תלוי באופן שבו תבחרו לעגל בסעיף הבא). (הבהרה: בהתפלגות קטומה הכוונה, שאם הקטימה היא בין a ל-b, אז למעשה מדובר בהתפלגות המותנית $X|a < X < b$).

2. כעת עליכם לעשות דיסקרטיזצייה להתפלגות קושי. כלומר, בהינתן ערך כלשהו מההתפלגות הרציפה, אנו נעגל אותו (לבחירתכם אם כלפי מטה/מעלה/לערך השלם הקרוב). מצאו מה פונקציית ההסתברות של התפלגות קושי קטומה ודיסקרטית.

3. בשאלה זו, המקסום של יחס ההסתברויות עלול להיות מסובך אנליטית. מאחר והתומך של ההתפלגות הבינומית הינו סופי, נוכל לבדוק מהו יחס ההסתברויות בכל הערכים האפשריים (0 ועד *n*) ולקחת את *C* להיות המקסימלי שאנו רואים.

4. עבור התפלגות קושי (ההתפלגות המציעה), השתמשו בתוחלת של ההתפלגות הבינומית בתור פרמטר המיקום, ובסטיית התקן של ההתפלגות הבינומית בתור פרמטר הכיול. הגרילו 10000 תצפיות מהתפלגות בינומית עם (*n=20,p=0.75*) בשיטת קבלה-דחייה. מהו *C*? מהי פרופורציית הדגימות ש"קיבלתם" מסך כל הדגימות? הדפיסו היסטוגרמה ואת פלט ה *summary*. מהי ההתפלגות של מספר הדגימות מהתפלגות קושי שדגמתם, על מנת לקבל 10000 תצפיות מהתפלגות בינומית?