מתרגל: ניר קרת
סמסטר ב' תשפ"א

חישוב סטטיסטי - תרגיל 4

- 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 ומומלץ גם להגיש הכל יחד בקובץ אחד, למשל על ידי שימוש ב Rmarkdown.

ראשית, צרו שתי פונקציות ב-R המבצעות קירוב לאינטגרל לפי שיטות הטרפז וסימפסון. הפונקציות מקבלות כארגומנט את הפונקצייה עליה נחשב את האינטגרל, את גבולות האינטגרל וכן את מספר הנקודות.

1) א. חשבו קירוב לכל אחד מהאינטגרלים/תוחלות הבאים לפי שיטת הטרפז, לפי שיטת סימפסון ולפי monte carlo integration עם רווח סמך בר"ס 95%. על ידי שימוש ב10000 נקודות/תצפיות (בסימפסון – **כולל** נקודות האמצע. ניתן להוסיף נקודה נוספת כדי שיהיה מספר אי-זוגי של נקודות).

<u>הערה:</u> שימו לב שבחלק מן המקרים לא תוכלו להפעיל את שיטות הטרפז/סימפסון ישירות בעקבות נקודות אי-הגדרה/גבול אינסופי. נסו לחשוב על החלפת משתנים כזו אשר תפתור את הבעייה.

ב. השוו את תוצאות השיטות השונות, ורשמו מסקנותיכם.

a. $\int_0^1 \frac{\cos(x)}{\sqrt{1-x^2}}\,dx$

b. $\int_1^\infty \frac{1}{(1+x^2)^2}\,dx$

c. $E\left(\frac{1}{1+\sqrt{X}}\right)\quad X\sim beta(2,3)$

d. $E(X^2\sin(X))\quad X\sim \exp(0.25), X\leq 5\ (truncated\ exponential)$

2) נניח ואנו צופים ב-15 תצפיות מהתפלגות פואסון, ואנו מעוניינים לבדוק את ההשערות:

$$H_0: \lambda = 3; \;\; H_1: \lambda > 3$$

א. מיצאו ערך קריטי עבור סטטיסטי מבחן $\bar{X}$ המבוסס על הקירוב הנורמלי, לבדיקת ההשערות לעיל, בר"מ $\alpha = 0.05$.

עם זאת, לא בטוח ש-15 תצפיות יספיקו לנו כדי להפעיל את משפט הגבול המרכזי.

ב. אמדו את α של המבחן הנ"ל בשיטת *monte carlo* על ידי 5000 חזרות, ובנו רווח סמך ל-α בר"ס 95%.

ג. אמדו את α של המבחן הנ"ל בשיטת *importance sampling* על ידי 5000 חזרות, ובנו רווח סמך, כאשר התפלגות ה *importance* הינה פואסון שהפרמטר שלה הוא אותו ערך שחישבתם בסעיף א'. האם אכן ר"מ של המבחן היא 0.05?

ד. איזו מן השיטות עדיפה משיקולי שונות?

3) חשבו אומד עבור $E\left(e^{X^2}\right)$ $X \sim U(0,1)$ על ידי 1000 תצפיות:

א. באמצעות monte carlo integration "נאיבי" (על ידי דגימה מהתפלגות אחידה).

ב. באמצעות importance sampling על ידי התפלגות שהצפיפות שלה פרופורציונאלית לפונקציה $(x+1)^2$ בקטע $0 \le X \le 1$. מה המוטיבציה לשימוש בהתפלגות זו במקום ההתפלגות האחידה?

ג. השוו את האומדים לשונויות האומדים.

4) ברצוננו לחשב את הביטוי $E(\cos(X^3))$ כאשר $X \sim N(0,1)$.

א. הגרילו 10000 דגימות מהתפלגות *double-exponential* (סטנדרטית)

ב. על ידי שיטת *acc-rej*, דגמו מהתפלגות נורמלית מתוך המדגם מהסעיף הקודם. חשבו אומדן לביטוי לעיל על סמך התצפיות שקיבלתם.

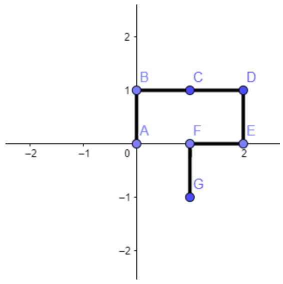
ג. חשבו אומדן לביטוי לעיל בשיטת *importance sampling* על ידי התצפיות מסעיף א'.

ד. איזה מן האומדים עדיף מבחינת כמות הדגימות מהתפלגות אחידה? איזה מן האומדים עדיף מבחינת השונות?

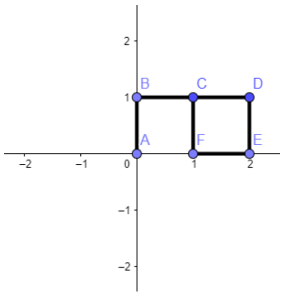
5**) (שאלת רשות עם בונוס + 15 נקודות)

SAW ‏(Self-Avoiding Walk) הינו רצף של הילוכים על מערכת צירים (כאשר תחילתו בראשית הצירים) כך שהרצף אינו מבקר באותה הנקודה פעמיים. נתייחס בשאלה זו ל-SAW על מערכת צירים דו-מימדית, כאשר כל הילוך הינו תזוזה באורך 1 בצורה אופקית או אנכית מהנקודה הקודמת.

לדוגמא, המסלול המתואר בגרף הבא הינו SAW מאורך 6 מכיוון שאינו חותך את עצמו.



לעומת זאת, המסלול המתואר בגרף הבא **אינו** SAW מאורך 6 מכיוון שהוא חותך פעמיים את הנקודה C.



ל-SAW ישנם שימושים בכימיה בתור מודל המחקה מבנה של פולימר.

נניח וחוקר מעוניין בתוחלת של תכונה מסויימת של SAW מאורך n, לדוגמא: מה תוחלת המרחק בין שני הקצוות של SAW, בהנחה שכל אחד מה-SAW האפשריים באורך n מתקבל בהסתברות אחידה.

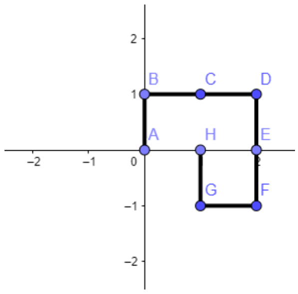
יצירת כל ה- SAW האפשריים מאורך n, במיוחד כאשר n גדול, היא משימה קשה, ובמקום זאת החליט החוקר לאמוד את התוחלת על ידי סימולציות.

1) דרך נאיבית ליצור SAW, הינה בכל שלב לבחור בהסתברות 0.25 את אחד הכיוונים (ימינה/שמאלה/למטה/למעלה) וללכת צעד בכיוון זה. אם המסלול מבקר בנקודה שכבר היה בה, פוסלים את הSAW ומתחילים מחדש.
הסבירו מה הבעייתיות בשיטה זו, ומדוע היא גרועה במיוחד כאשר n (אורך השרשרת) גדול.

2) החוקר הציע במקום זאת לבחור בכל פעם באקראי את אחד מהכיוונים **האפשריים לאותו שלב**. לשיטה זו נקרא look-one-ahead.
הסבירו (אפשר על ידי דוגמא) מדוע דגימה של SAW בשיטה זו לא בהכרח מניבה מסלולים בעלי הסתברות זהה.

3) עם זאת, לא הכל אבוד! מכיוון שהחוקר מעוניין באמידת תוחלת, הסבירו כיצד ניתן להשתמש בשיטה מהסעיף הקודם, אך לתקן אותה על ידי Importance Sampling.
נניח ו- s_i הינה התצפית ה-i של SAW שנדגם בשיטת look-one-ahead, ו$h(s)$ הינה פונקצייה המקבלת SAW כארגומנט ומחזירה את התכונה בה מעוניין החוקר (למשל המרחק בין הקצוות). הסבירו כיצד ייראה האומד על סמך m תצפיות לפי Importance Sampling, **וכן** הדגימו את הצעתכם על הגרף הראשון המופיע למעלה.
(כלומר, הניחו שגרף זה הינו תצפית SAW מאורך 6 שנבנה בשיטת look-one-ahead, והסבירו כיצד היא תבוא לידי ביטוי באומד הסופי, שהינו ממוצע של m ערכים).

4) גם בשיטת Look-one-ahead יכולים להיווצר מקרים שהדגימה תיכשל – למשל בדוגמא הבאה:



בשלב האחרון בוצע צעד מ-G ל-H (כי צעד זה היה אפשרי), אולם מ-H אין אפשרות להמשיך.

הסבירו כיצד נדגום על ידי שיטת look-two-ahead, ובכללי, על ידי שיטת look-k-ahead, ומהם היתרונות והחסרונות המתקבלים כאשר מגדילים את k.