תרגיל בית 6
התפלגויות דגימה

שאלה 1:

ידוע שגובה התושבים במדינת ליליפוט מתפלג נורמלית עם תוחלת ושונות כלשהן. כדי לאמוד את שונות גובה התושבים נלקח מדגם מקרי של 10 תושבים.

1. מהי ההסתברות שהסטייה **היחסית** של אומד נראות מירבית לשונות מהשונות האמיתית של גובה התושבים, אינה עולה על 0.5?
2. מהי ההסתברות שהסטייה **היחסית** של אומד חסר הטיה לשונות S^2 מהשונות האמיתית של גובה התושבים, אינה עולה על 0.5?
3. מצאו את הטעות הריבועית הממוצעת (MSE) של אומד נראות מירבית לשונות ושל S^2 כפונקציה של גודל המדגם n ושונות σ^2. האם אחד מבין האומדים הנ״ל עדיף על-פני השני משיקולי MSE? נמקו. רמז: היעזרו בנוסחה לתוחלת ושונות של התפלגות חי בריבוע.
4. השוו את התוצאות מכל הסעיפים ודונו בתוצאות. האם התוצאות שהתקבלו הן הגיוניות?

שאלה 2:

יהי $X_1,\ldots,X_n$ מדגם מקרי בגודל n מהתפלגות $N\left(\mu_X,\sigma^2\right)$ ויהי $Y_1,\ldots,Y_m$ מדגם מקרי בגודל m מהתפלגות $N\left(\mu_Y,\sigma^2\right)$. נתון כי המדגמים הם בלתי תלויים.

א. נגדיר: $S^2=\dfrac{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2+\sum_{i=1}^{m}\left(Y_i-\bar{Y}\right)^2}{K}$, כאשר K הינו קבוע שעבורו S^2 הוא אומד חסר הטיה ל- σ^2. מהו K?

ב. נגדיר $W=\dfrac{(n+m-2)S^2}{\sigma^2}$. כיצד מתפלג W? הוכיחו.

ג. נגדיר:

$$M=\frac{\left(\bar{Y}-\bar{X}\right)-\left(\mu_Y-\mu_X\right)}{S\sqrt{\frac{1}{n}+\frac{1}{m}}}$$

כיצד מתפלג M? הוכיחו.

בכל הסעיפים הבאים נניח $n=m$.

כעת הגרילו N=10000 מדגמים בגודל $n = 10$ מהתפלגות $N\left(\mu_X, \sigma^2\right)$ והגרילו N=10000 מדגמים בגודל $n = 10$ מהתפלגות $N\left(\mu_Y, \sigma^2\right)$. קבעו את μ_X, μ_Y ואת σ כרצונכם.

ד. עבור כל אחד מהמדגמים חשבו את W.
בנו היסטוגרמה של ערכי W שחישבתם. על גבי ההיסטוגרמה ציירו את פונקציית הצפיפות של W לפי ההתפלגות שמצאתם בסעיף ב׳. האם הסימולציה תומכת בתשובתכם?

עבור כל אחד מהמדגמים חשבו את M.
בנו היסטוגרמה של ערכי M שחישבתם. על גבי ההיסטוגרמה ציירו את פונקציית הצפיפות של M לפי ההתפלגות שמצאתם בסעיף ג׳. האם הסימולציה תומכת בתשובתכם?

ה. מהי ההתפלגות המקורבת של W כאשר n גדול? חזרו על הסימולציה בסעיף הקודם עם $n > 100$. בנו היסטוגרמה של ערכי W שחישבתם. על גבי ההיסטוגרמה ציירו את פונקציית הצפיפות המקורבת של W. האם הסימולציה תומכת בתשובתכם?

מהי ההתפלגות המקורבת של M כאשר n גדול? חזרו על הסימולציה בסעיף הקודם עם $n > 100$. בנו היסטוגרמה של ערכי M שחישבתם. על גבי ההיסטוגרמה ציירו את פונקציית הצפיפות המקורבת של M. האם הסימולציה תומכת בתשובתכם?

ו. חשבו את $\lim_{n\to\infty} P\left(\left|\frac{S^2-\sigma^2}{\sigma^2}\right| \le 0.5\right)$. השתמשו בתכונה האסימפטוטית של התפלגות חי-בריבוע שראיתם בהרצאה. האם התוצאה אינטואיטיבית? הסבירו.

חשבו את ההסתברות $P\left(\left|\frac{S^2-\sigma^2}{\sigma^2}\right| \le 0.5\right)$ עבור n=10. השוו את התוצאה עם תשובתכם לשאלה 1 סעיף ב׳2.

חשבו את ההסתברות $P\left(\left|\frac{S^2-\sigma^2}{\sigma^2}\right| \le 0.5\right)$ עבור n=100. האם התוצאות אינטואיטיביות? הסבירו.

ז. חשבו את ה-MSE של S^2 ביחס ל-σ^2. השוו ל-MSE של שני האומדים בסעיף 3 בשאלה הקודמת. איזה משלושת האומדים עדיף משיקולי MSE? האם התשובה היא אינטואיטיבית? נמקו.

שאלה 3:

נתונות שורות הקוד הבאות ב-R:

```
num_samples=10000
sample_size=10

vec=rep(0, num_samples)

for (j in 1:num_samples){
  sample=rnorm(sample_size,mean=1,sd=2)
  vec[j]=sum((sample-1)^2)
}
```

חשבו, ללא שימוש ב-R, למה שווה בערך var(vec)?