ירון גרשון

308350503

שיטות חישוביות - תרגיל בית 1

1)

$$Proj_{\underline{w}}(\underline{v}) = \frac{(\underline{v}\cdot\underline{w})}{|\underline{w}|^2}\,\underline{w}$$

$$\underline{v}\cdot\underline{w} = \begin{pmatrix}1\\2\\3\\4\end{pmatrix}\cdot\begin{pmatrix}0\\-1\\1\\2\end{pmatrix} = 1\cdot 0 + 2\cdot(-1) + 3\cdot 1 + 4\cdot 2 = -2+3+8 = 9$$

$$|\underline{w}|^2 = 0^2 + (-1)^2 + 1^2 + 2^2 = 1+1+4 = 6$$

$$Proj_{\underline{w}}(\underline{v}) = \frac{9}{6}\begin{pmatrix}0\\-1\\1\\2\end{pmatrix} = \begin{pmatrix}0\\-3/2\\3/2\\3\end{pmatrix}$$

2) בהמשך לשאלה הקודמת:

$$\underline{v}\cdot\underline{w} = \begin{pmatrix}1\\2\\3\\4\end{pmatrix}\cdot\begin{pmatrix}1\\0\\1\\-1\end{pmatrix} = 1\cdot 1 + 2\cdot 0 + 3\cdot 1 + 4\cdot(-1) = 1+3-4 = 0$$

מכיוון וקיבלנו כי המכפלה הסקלרית בין הוקטורים הנתונים
שווה לאפס אזי הוקטורים אורתוגונלים ומכאן ההטלה
אחד על השני שווה לאפס גם-כן!

3) בתרגול הצגנו את המכפלה הסקלרית בין שני וקטורים כך:

$$\langle \underline{v}, \underline{w} \rangle = \|\underline{v}\|\cdot\|\underline{w}\|\cdot\cos\alpha$$

כעת, אם ידוע לנו כי הוקטורים אינם שווים לאפס אזי
המכפלה הסקלרית בניהם שווה לאפס אמ"מ $\cos\alpha = 0$
וזה קורה אמ"מ הזווית α שווה $\pm 90^\circ$, כנדרש.

4) ראשית, אם המטריצה A היא אורתונורמלית אזי מתקיים $A^TA = I$
ולכן עבור כל $x \in V$ מתקיים:

$$\|Ax\|^2 = (Ax)^T(Ax) = x^T\underbrace{(A^TA)}_{A\text{ אורתונורמלית}}x = x^Tx = \|x\|^2$$

לבסוף, מכיוון ששני צדי המשוואה ישנם מספרים חיוביים ולכן נוכל להוציא שורש ונקבל כי לכל $x \in V$ מתקיים:

$$\|Ax\| = \|x\|$$

כנדרש.

5) הפלט ה-SVD מאפשר לנו, בצורה קלה, לחשב את ההופכי של מטריצה A. לשם כך, אנו משתמשים בעובדה שהמטריצות U,V הן אורתונורמליות ולכן המטריצות ההופכיות שלהן הן הטרנספוז שלהן, או במילים אחרות:

$$U^TU = UU^T = I \quad , \quad V^TV = VV^T = I$$

כעת, בהינתן כי A מטריצה הפיכה, ההופכי שלה בעזרת SVD יהיה:

$$A^{-1} = VD^{-1}U^T$$

כאשר:

$$D^{-1} = \begin{bmatrix} 1/D_1 & & & \\ & 1/D_2 & & \\ & & \ddots & \\ & & & 1/D_n \end{bmatrix}$$

הרעיון שעומד מאחורי הפלט הוא שאנו מבצעים את שלושת הטיפולים (בקצרה) שהיינו עושים בפלט SVD על המטריצה A רק בסדר הפוך. נוכל בהכפלה במטריצה U^T לבטל את הסיבוב האחרון שעשינו, אחר-כך נבטל את המתיחה ע"י הכפלה באיבר $1/D_i$ לאורך כל ציר ולאחר מכן נבטל את הסיבוב הראשון שעשינו למטריצה A בעזרת הכפלה במטריצה V.

נוודא שהפלט נכון, נבדוק:

$$AA^{-1} = A^{-1}A = (VD^{-1}U^T)(UDV^T) = VD^{-1}(U^TU)DV^T = V(D^{-1}D)V^T = VV^T = I$$

ט) נעזר ע"י הנחיות התרגיל:

נחשב תחילה את C^TC:

$$C^TC = VD^TU^TUDV^T = VD^TDV^T$$

U מטריצה אורתוגונלית

$$U^TU = I$$

$$C^TC = \begin{pmatrix} 5 & -1 \\ 5 & 7 \end{pmatrix}\begin{pmatrix} 5 & 5 \\ -1 & 7 \end{pmatrix} = \begin{pmatrix} 5\cdot5+(-1)\cdot(-1) & 5\cdot5+(-1)\cdot7 \\ 5\cdot5+7\cdot(-1) & 5\cdot5+7\cdot7 \end{pmatrix}$$

$$= \begin{pmatrix} 26 & 18 \\ 18 & 74 \end{pmatrix}$$

נחשב את הערכים העצמיים בצורה רגילה:

$$\det(C^TC-\lambda I) = \det\begin{pmatrix} 26-\lambda & 18 \\ 18 & 74-\lambda \end{pmatrix} = \lambda^2-100\lambda+1600$$

$$= (\lambda-20)(\lambda-80)$$

לכן הערכים העצמיים הינם 20,80.

נמצא וקטורים עצמיים: נבחר כך שנוכל לקבל מטריצה V אורתוגונלית

$$C^TC-20I = \begin{pmatrix} 6 & 18 \\ 18 & 54 \end{pmatrix} \Rightarrow v_1 = \begin{pmatrix} -3/\sqrt{10} \\ 1/\sqrt{10} \end{pmatrix}$$

דרוש מאונכות ונרמול

$$C^TC-80I = \begin{pmatrix} -54 & 18 \\ 18 & -6 \end{pmatrix} \Rightarrow v_2 = \begin{pmatrix} 1/\sqrt{10} \\ 3/\sqrt{10} \end{pmatrix}$$

לכן,

$$V = \begin{pmatrix} -3/\sqrt{10} & 1/\sqrt{10} \\ 1/\sqrt{10} & 3/\sqrt{10} \end{pmatrix}, \quad D = \begin{pmatrix} 2\sqrt{5} & 0 \\ 0 & 4\sqrt{5} \end{pmatrix}$$

כעת נשתמש במשוואה $CV=UD$ ונמצא את U:

$$\begin{pmatrix} 5 & 5 \\ -1 & 7 \end{pmatrix}\begin{pmatrix} -3/\sqrt{10} & 1/\sqrt{10} \\ 1/\sqrt{10} & 3/\sqrt{10} \end{pmatrix} = \begin{pmatrix} -\sqrt{10} & 2\sqrt{10} \\ \sqrt{10} & 2\sqrt{10} \end{pmatrix} \Rightarrow U = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$$

כנדרש.

7) מפירוק סינגולרי נקבל כי המטריצה A מוגדרת כך: $A=UDV^T$

וכן אנו מגדירים את: $C_0=A^TA=(VD^TU^T)(UDV^T)=VD^TDV^T$

כעת, בהינתן $v_1,\dots,v_n$ וקטורים עצמיים של המטריצה C_0

עם הערכים העצמיים $\lambda_1\geq\dots\geq\lambda_n$ המתאימים להם בהתאמה

אם-כן את b_0 נכתוב כצירוף לינארי של כל הוקטורים העצמיים כך:

$$b_0=a_1v_1+a_2v_2+\dots+a_nv_n$$

נשתמש בהגדרת b_{k+1} ונראה כי:

$$b_1=\frac{C_0b_0}{\|C_0b_0\|}=\frac{1}{\|C_0b_0\|}\cdot C_0(a_1v_1+\dots+a_nv_n)$$

כעת נשתמש בערכים העצמיים כאשר אנו יודעים כי הכפלת כל וקטור עצמי במטריצה C_0 תגרום לו רק כאשר מוכפל בערך העצמי המתאים לאותו וקטור וכן כל השאר מצטמצמים ונקבל:

$$=\frac{1}{\|C_0b_0\|}\cdot(a_1\lambda_1v_1+\dots+a_n\lambda_nv_n)$$

באופן דומה,

$$b_2=\frac{C_0b_1}{\|C_0b_1\|}=\frac{1}{\|C_0b_0\|}\cdot\frac{1}{\|C_0b_1\|}\cdot C_0(a_1\lambda_1v_1+\dots+a_n\lambda_nv_n)$$

$$=\frac{1}{\|C_0b_0\|\cdot\|C_0b_1\|}\cdot(a_1\lambda_1^2v_1+\dots+a_n\lambda_n^2v_n)$$

לכן, באופן כללי:

$$b_k=\frac{1}{\prod_{i=0}^{k-1}\|C_0b_i\|}\cdot(a_1\lambda_1^kv_1+a_2\lambda_2^kv_2+\dots+a_n\lambda_n^kv_n)$$

$$=\frac{1}{\prod_{i=0}^{k-1}\|C_0b_i\|}\cdot\lambda_1^k\cdot\left(a_1v_1+a_2\left(\frac{\lambda_2}{\lambda_1}\right)^kv_2+\dots+a_n\left(\frac{\lambda_n}{\lambda_1}\right)^kv_n\right)$$

נשים-לב כי אנו יודעים כי $\lambda_1\geq\lambda_2\geq\dots\geq\lambda_n$ ולכן לכל $1<i$

מתקיים $\left(\frac{\lambda_i}{\lambda_1}\right)<1$ ולכן:

$$\lim_{k\to\infty}\left(\frac{\lambda_i}{\lambda_1}\right)^k=0$$

כעת, לבסוף:

$$\lim_{k\to\infty} b_k = \frac{1}{\prod_{i=0}^{k-1}\|C_0 b_i\|}\cdot \lambda_1^k \cdot a_1 \cdot v_1$$

לסיכום, שיטת החזקות הזו מנורמלת בכל שלב את הוקטור (בהנחה כי מדובר בנורמה הרצויה) ולכן בגבול זה מתכנס לוקטור העצמי v_1, כנדרש.

g) ראשית, נשים-לב כי המטריצה $diag(\sigma)$ הינה אלכסונית ומן U,U^T אורתוגונליות מהנתון ולכן $U\,diag(\sigma)\,U^T$ מהווה פירוק EVD של מטריצה כלשהיא $A\in\mathbb{R}^{n\times n}$ שערכיה העצמיים הינם σ_i^2 ומכאן $f(\sigma)=Ax$.
בתרגול ראינו כי היעקוביאן של פונקציה מהצורה $f(w)=Aw$ הינו $J_f(w)=A$ ולכן במקרה שלנו, כאשר x קבוע, מתקיים:

$$J_f(\sigma)=Ax$$

h) בתרגול ראינו כי $\nabla\|x\|_2^2=2x$ ומכאן $J_{\|\cdot\|_2^2}(x)=2x^T$.
נשתמש כעת בכלל השרשרת ונקבל:

$$J_h(\sigma)=\frac{1}{2}\|\cdot\|^2\circ(f(x)-y)(\sigma)=J_{\frac{1}{2}\|\cdot\|^2}(f(\sigma)-y)\cdot J_{f(x)-y}(\sigma)$$

$$=\frac{1}{2}\cdot 2(f(\sigma)-y)\cdot J_{f(x)-y}(\sigma)=(f(\sigma)-y)\cdot J_{f(x)-y}(\sigma)$$

לבסוף, בהמשך לשאלה g והפונקציה f, ומכיוון ש-y קבוע מתקיים כי $J_{f(x)-y}(\sigma)=Ax$ ולכן:

$$J_{h(\sigma)}=(Ax-y)Ax=A^2x^2-y\cdot Ax$$

כנדרש.

10) נחשב במפורש את הפונקציה $S:\mathbb{R}^n \to \mathbb{R}^n$ המוגדרת ע"י:

$$S(a) = \frac{e^{a}}{\sum_{k=1}^{n} e^{a_k}}$$

לצורך שנוכל לגזור את השבר נגדיר את הפונקציות:

$$g_i = e^{a_i} \quad ; \quad h = \sum_{k=1}^{n} e^{a_k}$$

חשוב כי מתקיים:

$$\frac{\partial h}{\partial a_j} = g_j \quad ; \quad \frac{\partial g_i}{\partial a_j} = \begin{cases} e^{a_i} = g_i & , \ i=j \\ 0 & , \ i\neq j \end{cases}$$

מכאן, עבור $i=j$ מתקיים:

$$\frac{\partial S_i}{\partial a_j} = \frac{\partial \frac{g_i}{h}}{\partial a_j} = \frac{g_i \cdot h - g_i g_i}{h^2} = \frac{g_i}{h} \cdot \frac{h - g_i}{h} = \underline{\underline{S_i(1-S_i)}}$$

כעת עבור $i\neq j$ נקבל:

$$\frac{\partial S_i}{\partial a_j} = \frac{\partial \frac{g_i}{h}}{\partial a_j} = \frac{1}{\partial a_j} \cdot \frac{\partial g_i \cdot h - g_i \cdot \partial h}{h^2} =$$

$$= \frac{\frac{\partial g_i}{\partial a_j} \cdot h - g_i \cdot \frac{\partial h}{\partial a_j}}{h^2} = \frac{0 \cdot h - g_i \cdot \frac{\partial h}{\partial a_j}}{h^2} = \frac{-e^{a_i} \cdot \frac{\partial \sum_{k=1}^{n} e^{a_k}}{\partial a_j}}{\left(\sum_{i=1}^{n} e^{a_k}\right)^2}$$

$$= -\frac{e^{a_i} \cdot e^{a_j}}{\left(\sum_{i=1}^{n} e^{a_k}\right)^2} = -\frac{e^{a_i}}{\sum_{i=1}^{n} e^{a_k}} \cdot \frac{e^{a_j}}{\sum_{i=1}^{n} e^{a_k}} = \underline{\underline{-S_i \cdot S_j}}$$

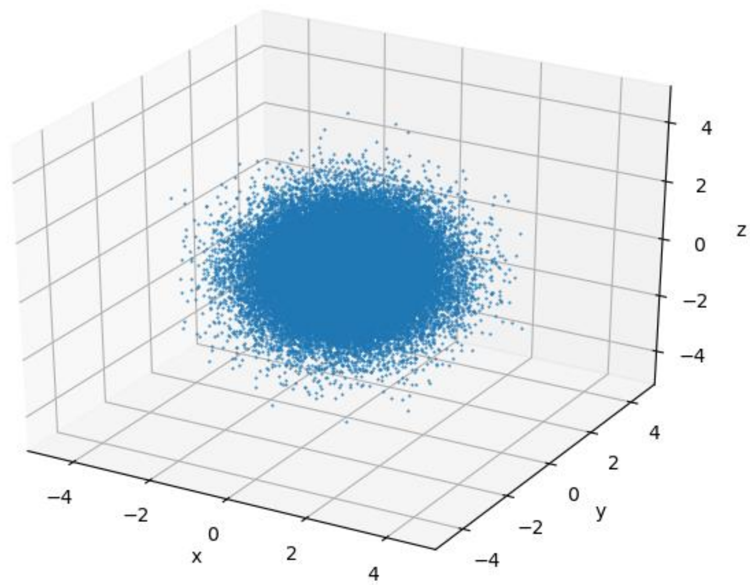
לכן קיבלנו:

$$J_S(a)_{i,j} = \begin{cases} S_i(1-S_i) & , \ i=j \\ -S_i S_j & , \ i\neq j \end{cases}$$

לירון גרשוני
308350503

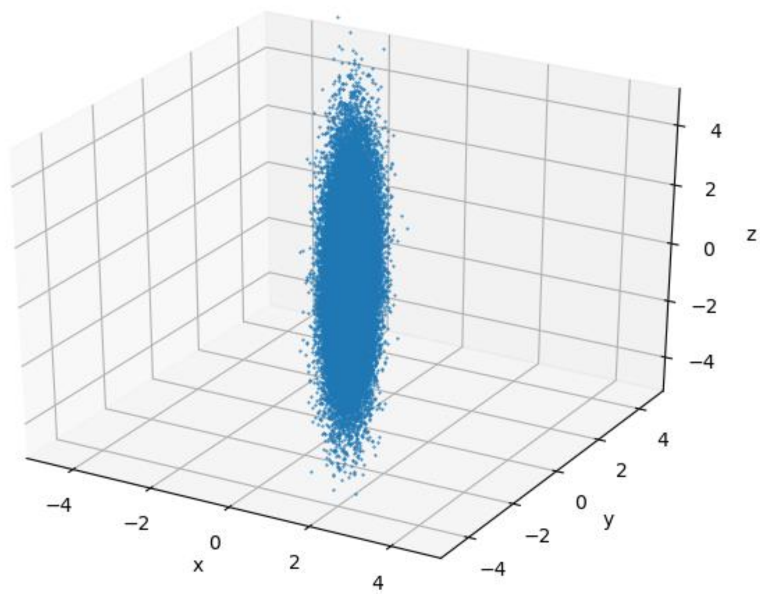
מערכות לומדות – המשך תרגיל בית 1

(11



התקבל פיזור נקודות המאפיין התפלגות נורמלית שכן הנקודות מרוכזות סביב הממוצע (0) בכל אחד מן הצירים.

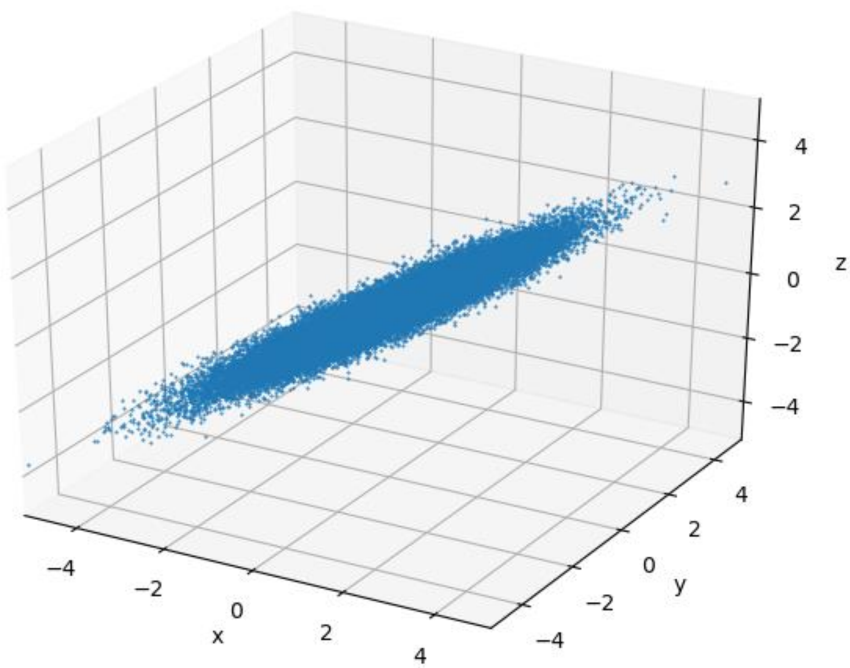
(12



מבחינה אנליטית, הכפלנו את המטריצה במטריצה S שהינה אלכסונית ובעלת ערכים שונים בכל ציר על-כן ציפינו לקבל שינוי בפיזור הנקודות בכל ציר- בצירים X ו-Y הכפלנו במספר שקטן מ-1 ולכן קיבלנו פעולה של "כיווץ".
לעומת-זאת, את ציר ה-Z הכפלנו במספר שגדול מ-1 ולכן קיבלנו פעולת "מתיחה".
מבחינה נומרית, מכיוון ואין תלות בין הצירים והוקטורים הינם בת"ל (ולכן השונות המשותפת ביניהם שווה לאפס) מטריצת השונויות תהא גם היא אלכסונית לאחר הכפלה במטריצה S וכן על איברי האלכסון נקבל את השונויות החדשות- במטריצה המקורית על האלכסון היה את הערך 1 כיוון ומדובר בהתפלגות נורמאלית ולכן לאחר הכפלה במטריצה S נקבל על האלכסון את הערכים של אלכסון S בריבוע:

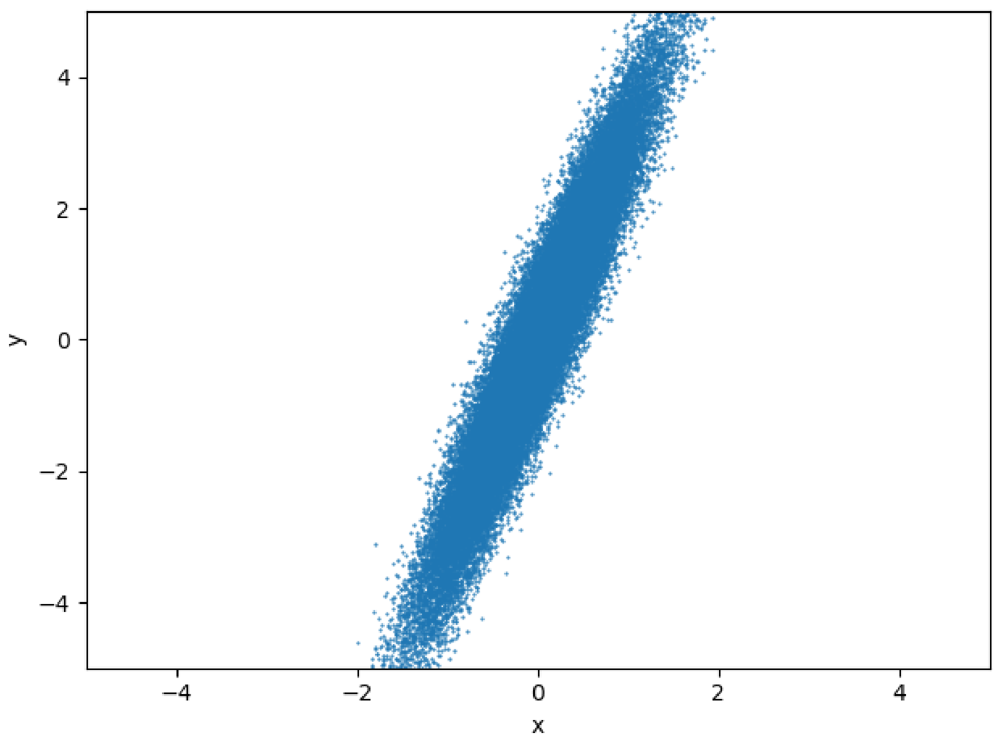
$$\begin{pmatrix} 0.01 & 0 & 0 \\ 0 & 0.25 & 0 \\ 0 & 0 & 4 \end{pmatrix}$$

(13



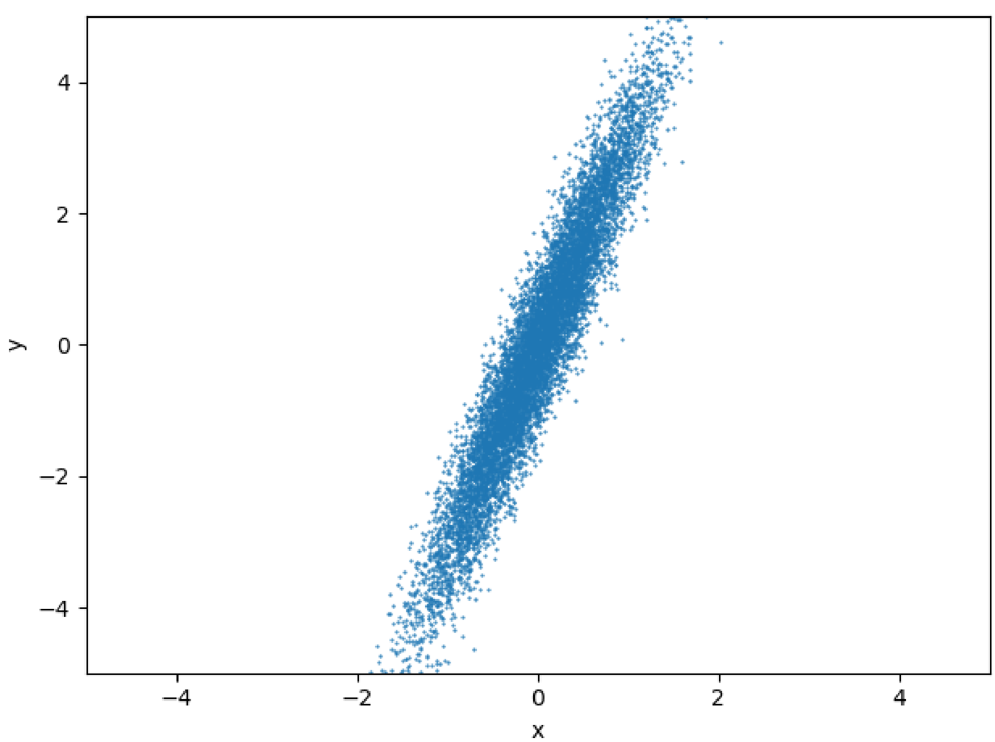
מהקורס באלגברה לינארית אנו יודעים שהכפלת מטריצה במטריצה אחרת אורתוגונלית מתבצעת פעולת "סיבוב" ולכן, כפי שציפינו, התקבל גרף המראה את אותו הפיזור משאלה 12 אך לאחר סיבוב. על-כן, מטריצת השונויות המשותפות כבר אינה אלכסונית.

(14



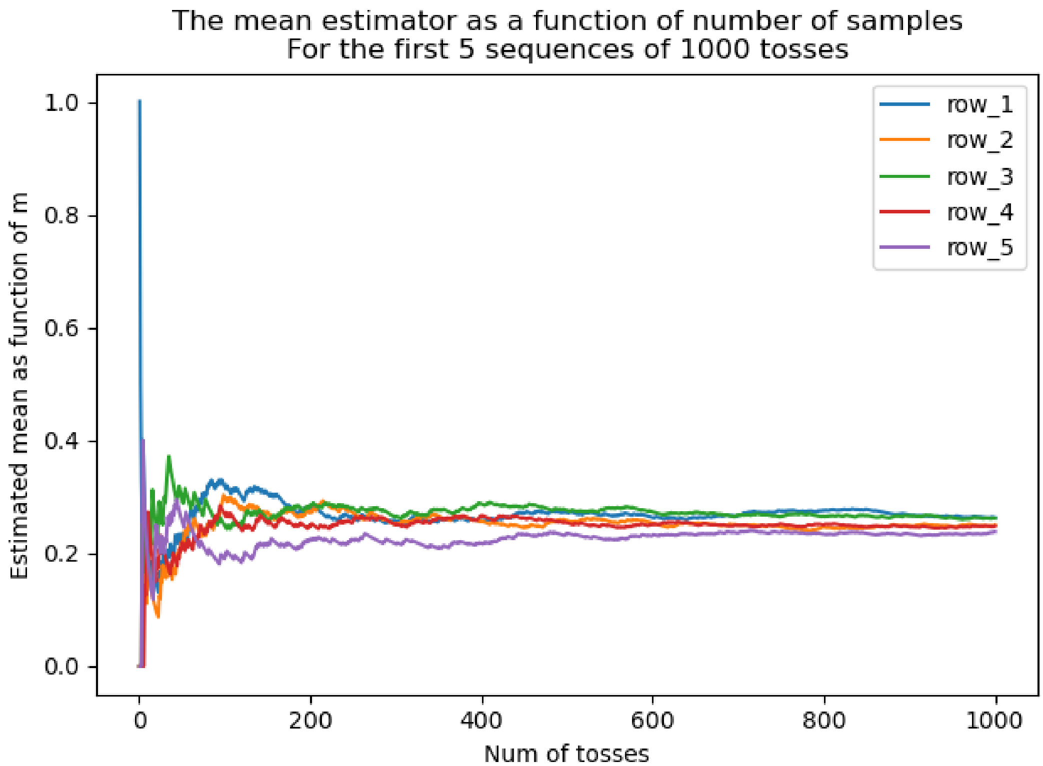
כעת הכפלנו את המטריצה מהשאלה הקודמת במטריצה אלכסונית שעל האלכסון שלה יש את הערך 1 בשתי השורות הראשונות ואת הערך 0 בשורה האחרונה ולכן קיבלנו את ההטלה של המטריצה מהשאלה הקודמת על הצירים X ו-Y וכפי שנאמר, לאחר הטלה על צירים מסוימים עדיין מתקבלת אותה ההתפלגות שהיתה לפני ההטלה.

(15



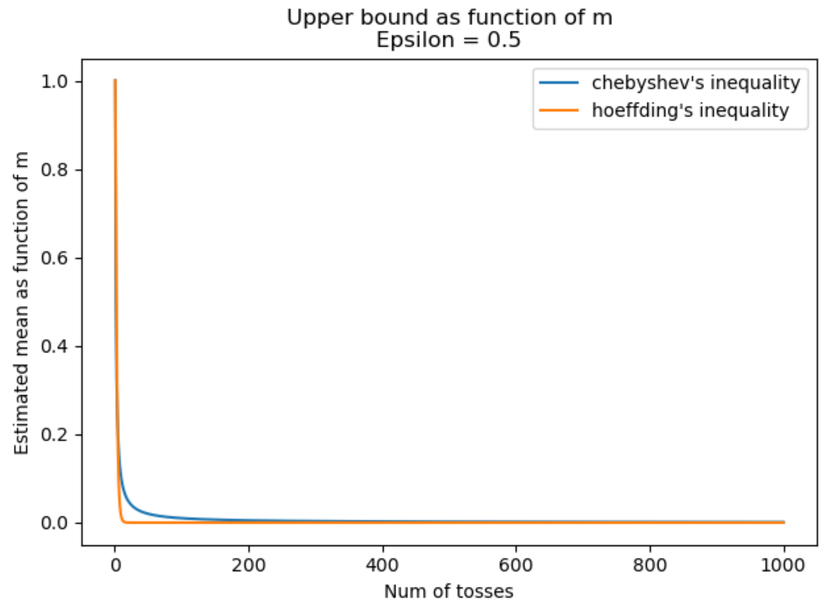
נשים-לב שבשאלה הזו ביצענו פילטור של הנקודות לערכים מסוימים בלבד בציר ה-Z ולכן, כפי שציפינו, קיבלנו גרף הדומה בצורתו לשל זה מהשאלה הקודמת אך בעל צפיפות נקודות פחותה שכן מחקנו נקודות שלא עמדו בקריטריון שנקבע.

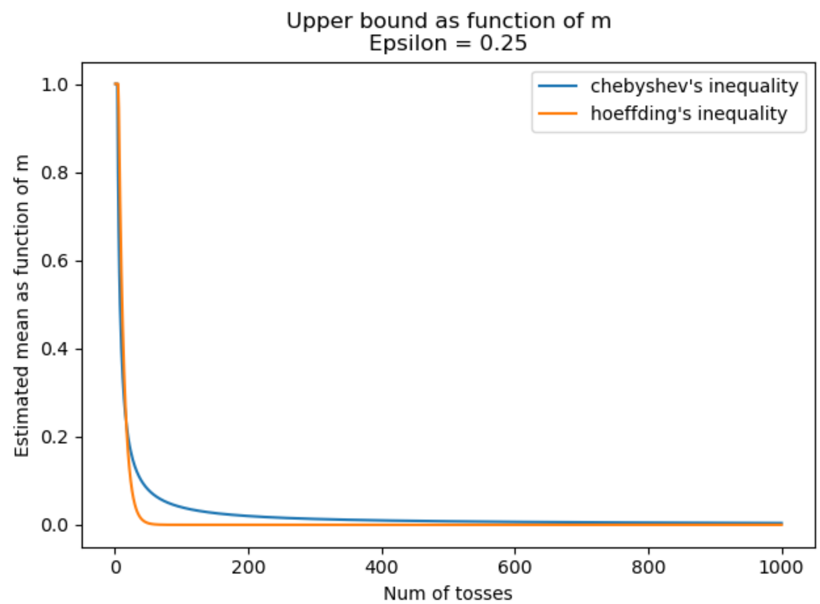
16) סעיף א':

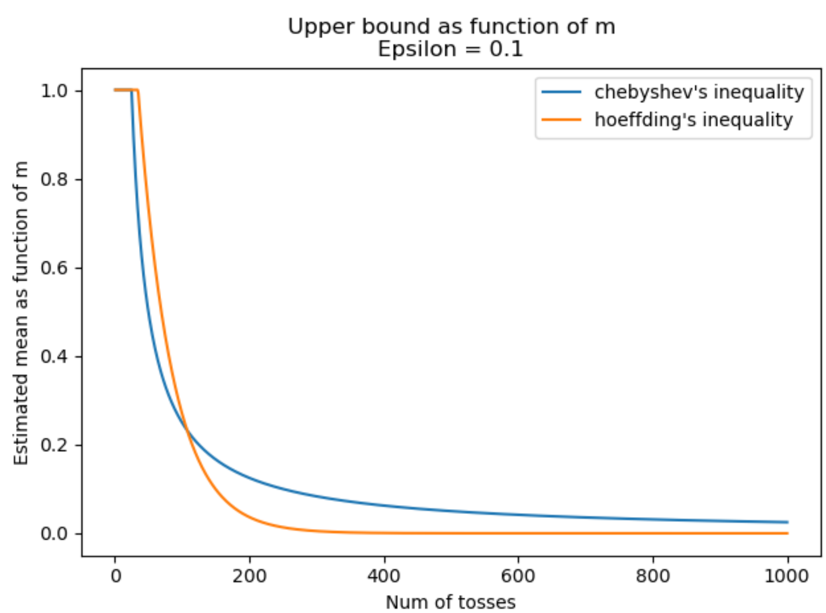


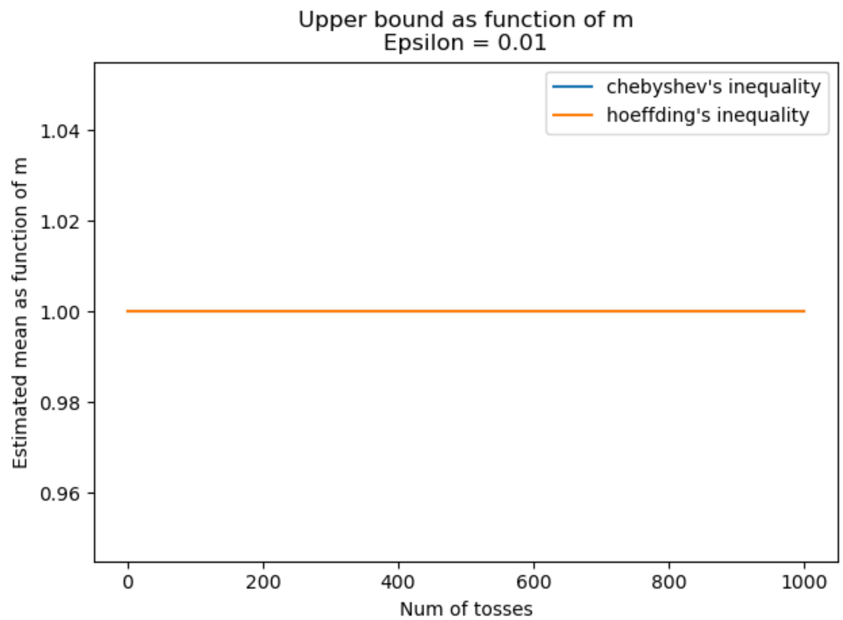
בהמשך לקורס בהסתברות וע"פ החוק החלש של המספרים הגדולים, ציפינו שככל וגודל המדגם הולך וגדל כך הממוצע ישאף לתוחלת. נשים-לב שהגרף לעיל מראה זאת בדיוק שכן ככל שמספר ההטלות גדל כך אנו שואפים יותר ויותר לסיכוי לקבלת עץ במטבע כלומר, 0.25.

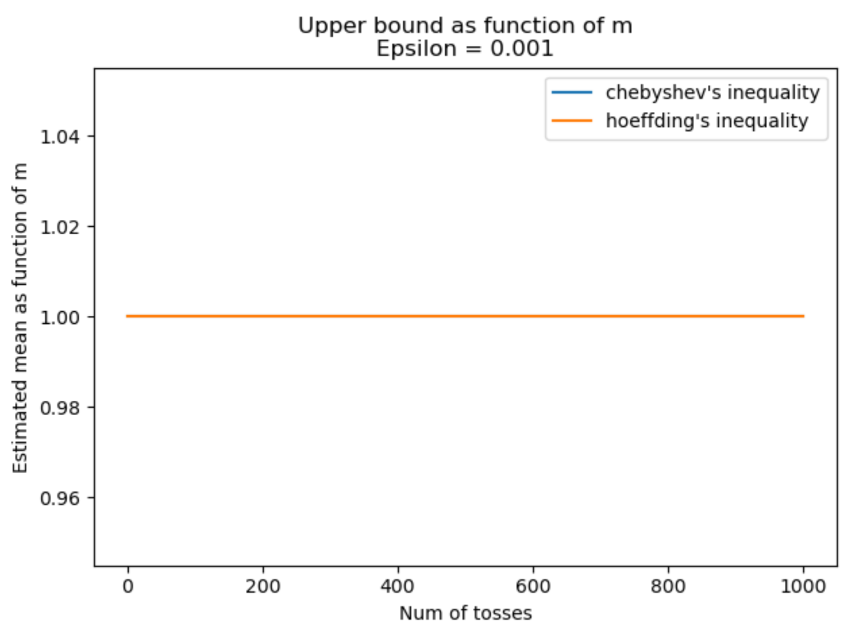
סעיף ב':


Upper bound as function of m
Epsilon = 0.5
chebyshev's inequality
hoeffding's inequality
Estimated mean as function of m
Num of tosses
0.0
0.2
0.4
0.6
0.8
1.0
0
200
400
600
800
1000


Upper bound as function of m
Epsilon = 0.25
chebyshev's inequality
hoeffding's inequality
Estimated mean as function of m
Num of tosses
0.0
0.2
0.4
0.6
0.8
1.0
0
200
400
600
800
1000

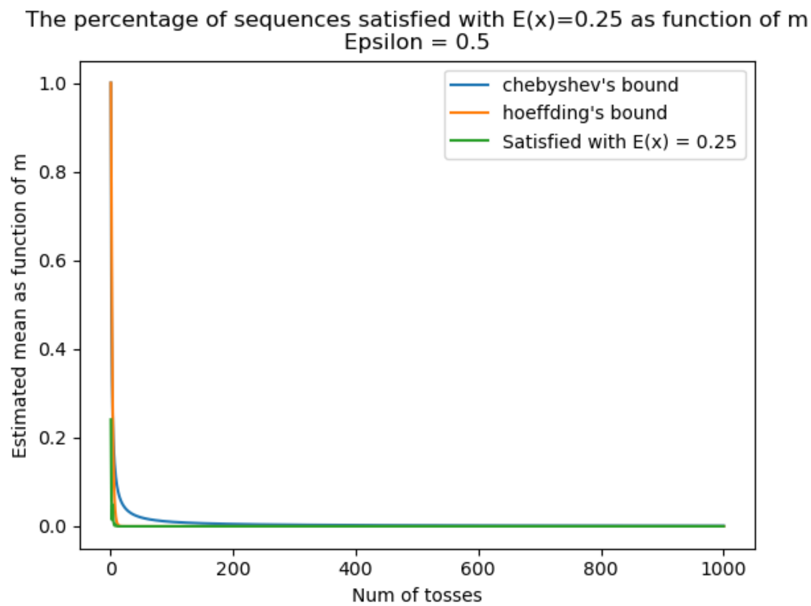

Upper bound as function of m
Epsilon = 0.1
chebyshev's inequality
hoeffding's inequality
Estimated mean as function of m
Num of tosses
0.0
0.2
0.4
0.6
0.8
1.0
0
200
400
600
800
1000

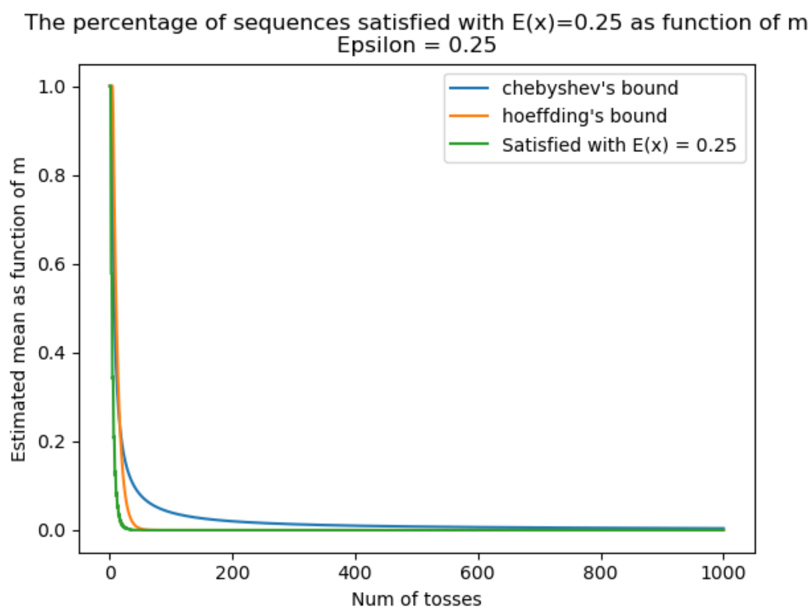


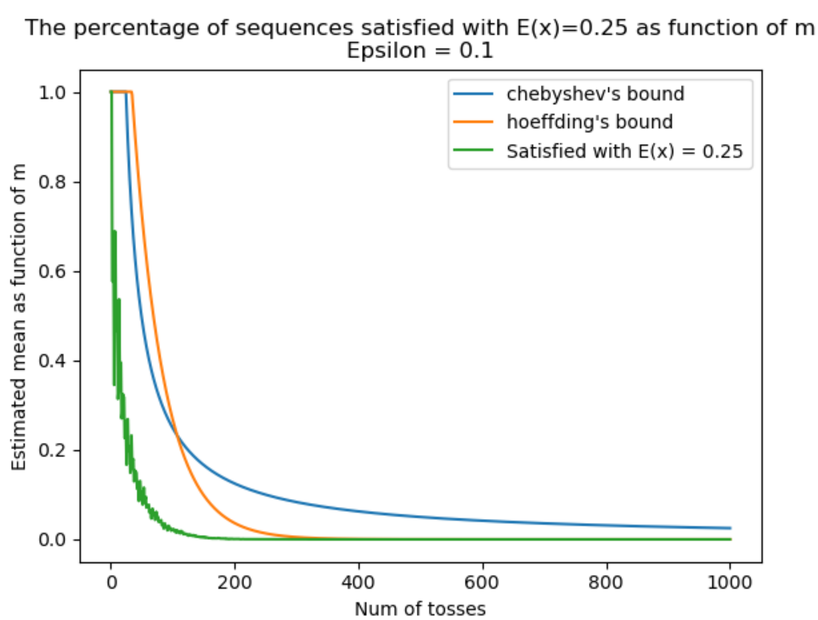


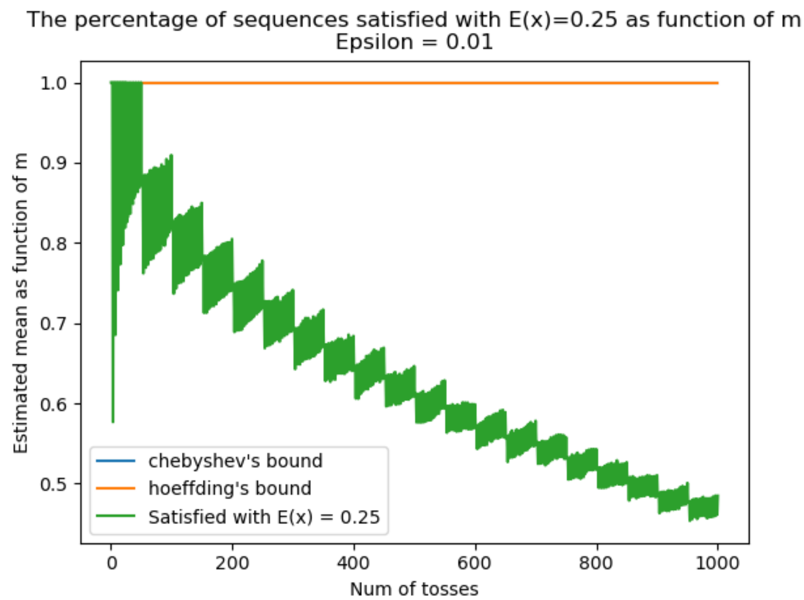
בהמשך לקורס בהסתברות, ניתן להבחין בשתי עובדות אותן ניתן להסיק מהגרפים לעיל- האחת, אי-שיוויון הופדינג נותן חסם טוב יותר משל צ'בישב שכן הינו חסם מעריכי ועל-כן הדוק יותר. השנייה, ניתן לראות כי ככל ומספר הדגימות עולה כך קטן ההפרש בין אפסילון לממוצע ולכן, עבור אפסילון קטן, נדרשות יותר דגימות על-מנת להבחין בשוני בין החסמים.

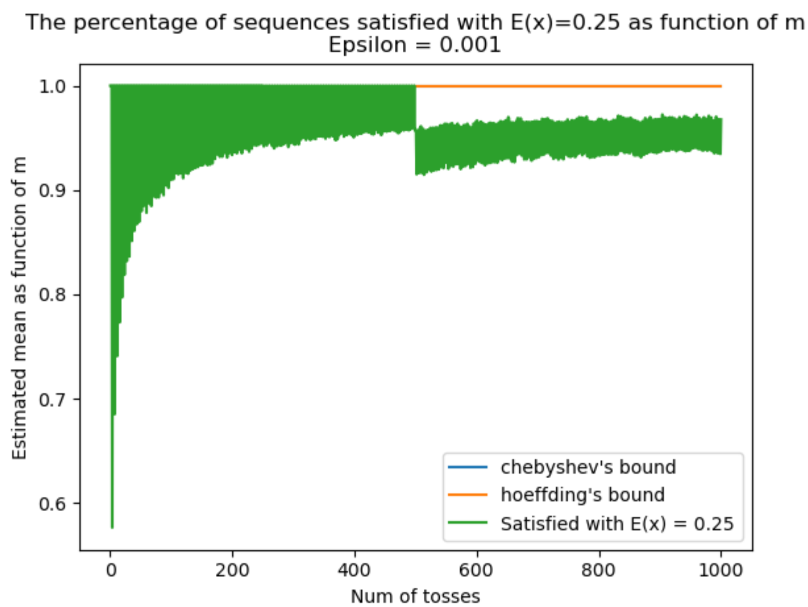
סעיף ג':


The percentage of sequences satisfied with E(x)=0.25 as function of m
Epsilon = 0.5
chebyshev's bound
hoeffding's bound
Satisfied with E(x) = 0.25
Estimated mean as function of m
Num of tosses
1.0
0.8
0.6
0.4
0.2
0.0
0
200
400
600
800
1000


The percentage of sequences satisfied with E(x)=0.25 as function of m
Epsilon = 0.25
chebyshev's bound
hoeffding's bound
Satisfied with E(x) = 0.25
Estimated mean as function of m
Num of tosses
1.0
0.8
0.6
0.4
0.2
0.0
0
200
400
600
800
1000


The percentage of sequences satisfied with E(x)=0.25 as function of m
Epsilon = 0.1
chebyshev's bound
hoeffding's bound
Satisfied with E(x) = 0.25
Estimated mean as function of m
Num of tosses
1.0
0.8
0.6
0.4
0.2
0.0
0
200
400
600
800
1000





כפי שציפינו, הגרפים לעיל מראים שככל ונגדיל את מספר הדגימות כך אחוז הדגימות שרחוק מהתוחלת ביותר מאפסילון נתון יורד. כמו-כן, כאשר אפסילון גדול יחסית אז הדעיכה יותר מהירה כיוון ויותר נקודות נמצאות במרחק מהתוחלת שקטן מאפסילון זה.

ישנה מסקנה ברורה מהגרפים הללו והיא שככל ונגדיל את כמות הדגימות כך נוכל להקטין את השגיאה.