IML -EX1
Amitay Sicherman -203449004

Warm-up - Algebra Recap

$\boxed{\text{שאלה 1}}$

נשתמש בנוסחת ההיטל:

$$P = \|v\|\cos\theta \cdot \frac{u}{\|u\|} = \frac{\langle v,u\rangle}{\|u\|^2}\cdot u$$

$$= \frac{\left\langle \begin{pmatrix}1\\2\\3\\4\end{pmatrix}, \begin{pmatrix}0\\-1\\1\\2\end{pmatrix} \right\rangle}{\left\| \begin{pmatrix}0\\-1\\1\\2\end{pmatrix} \right\|^2} \cdot \begin{pmatrix}0\\-1\\1\\2\end{pmatrix}$$

$$= \frac{-2+3+8}{1+1+4} \cdot \begin{pmatrix}0\\-1\\1\\2\end{pmatrix} = \begin{pmatrix}0\\-\frac{3}{2}\\\frac{3}{2}\\3\end{pmatrix} = \boxed{\frac{3}{2}\begin{pmatrix}0\\-1\\1\\2\end{pmatrix}}$$

$\boxed{\text{שאלה 2}}$

לפי אותה נוסחה ונקבל:

$$\frac{\left\langle \begin{pmatrix}1\\2\\3\\4\end{pmatrix}, \begin{pmatrix}1\\0\\1\\-1\end{pmatrix} \right\rangle}{\left\| \begin{pmatrix}1\\0\\1\\-1\end{pmatrix} \right\|^2} \cdot \begin{pmatrix}1\\0\\1\\-1\end{pmatrix} = \frac{1+3-4}{1+1+1}\begin{pmatrix}1\\0\\1\\-1\end{pmatrix} = \boxed{\frac{0}{3}\begin{pmatrix}1\\0\\1\\-1\end{pmatrix} = 0}$$

שאלה 3:

נוכיח את שני הצדדים:

נניח ש $\langle v,w\rangle = 0$ ונוכיח כי הזווית ביניהם היא $90°$ –

מתרגול הוכחנו כי

$$\langle v,w\rangle = \|v\|\cdot\|w\|\cdot\cos\theta$$

ובמקרה שלנו:

$$\langle v,w\rangle = \|v\|\cdot\|w\|\cdot\cos\theta = 0$$

מכיוון ש ~~[illegible]~~ v, w הם וקטורים ה"ס- ולפי הגדרת הנורמה $\|x\|, \|y\| \neq 0$ ולכן מתקיים ש

$$\cos\theta = 0 \Rightarrow \boxed{\theta = 90°}$$

צד שני:

נניח כי $\theta = 90$, אז נוכיח כי לפי ההגדרה

$$\langle v,w\rangle = \|v\|\cdot\|w\|\cdot\cos 90 = 0$$

מ"ש

שאלה 4

$$\|Ax\|^2 = \langle Ax, Ax\rangle = (Ax)^T Ax$$

$$= x^T A A x = \|x\|_2^2 = \|x\|_2^2$$

$$\boxed{\Rightarrow \|Ax\|_2 = \|x\|_2}$$

SVD

שאלה 2:

<u>טענת עזר 1:</u>

חישוב מטריצה הפיכה (מטריצה אלכסונית) יכול להעשות בצורה יעילה.

כמו $O=diag(a_1 \dots a_n) \Rightarrow O^{-1} = diag(\frac{1}{a_1} \dots \frac{1}{a_n})$

עלות החישוב היא $O(n)$

<u>טענת עזר 2:</u>

חישוב מטריצה הפיכה (מטריצה אורתוגונלית) יכול להעשות בצורה יעילה - A = ortogonal matrix:

$$A \cdot A^{-1} = A \cdot A^T = I_n \Rightarrow A^{-1} = A^T$$

מאחר וחישוב A^T יכול להעשות ב- $O(n^2)$

$$A_{ij} = A_{ji} \quad \forall i,j \in \{1, n\}$$

<u>הצעת פתרון:</u>

$A = UDV^T$ | SVD decomposition

$$\Rightarrow A^{-1} = (UDV^T)^{-1} = (V^T)^{-1} \quad D^{-1} \quad U^{-1}$$

[אורתוגונלית, טענה 2] אורתוגונלית טענה 2

$$\boxed{= V \quad D^{-1} \quad U^T}$$

כאן חישוב - טענה 1

$\boxed{\text{שאלה 6:}}$

I)

$$C \cdot C^T = \begin{bmatrix} 50 & 30 \\ 30 & 50 \end{bmatrix}$$

II) לפי הלמה מהתרגול מתקיים ש

$$A \overset{SVD}{=} U O V^T$$

$$A A^T \overset{EVD}{=} U \underbrace{O \cdot O^T}_{\theta} U^T$$

32

ובפרט)

$$\begin{bmatrix} 50 & 30 \\ 30 & 50 \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} d_1^2 & 0 \\ 0 & d_2^2 \end{bmatrix} \begin{bmatrix} a & c \\ b & d \end{bmatrix}$$

$$1)\ \begin{bmatrix} 50 & 30 \\ 30 & 50 \end{bmatrix} = \begin{bmatrix} a^2 d_1^2 + c b d_2^2 & a b d_1^2 + b d d_2^2 \\ c a d_1^2 + d c d_2^2 & a b d_1^2 + b d d_1^2 \end{bmatrix}$$

$$\Rightarrow d_1^2 = 80,\ d_2^2 = 20 \qquad a = b = c = -d = \sqrt{\tfrac{1}{2}}$$

כלומר:

$$O = \begin{bmatrix} \sqrt{80} & \\ & \sqrt{20} \end{bmatrix} \quad U = \sqrt{\tfrac{1}{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

III) נתון כי ש $C = U O V^T$ נכפול ב V ונקבל

$$C V = U O V^T V = U O$$

ולכן

$$V^T = \begin{pmatrix} \sqrt{\frac{1}{10}} & \sqrt{\frac{9}{10}} \\ \sqrt{\frac{9}{10}} & -\sqrt{\frac{1}{10}} \end{pmatrix}$$

$\boxed{\text{שאלה 1:}}$

נאתחל את [illegible] כרגיל:

$$b_0 = \sum_{i=0}^{n} a_i V_i$$

~~מכיוון ש [illegible] b_0 [illegible]:~~

$$\cancel{b_0 = \sum_{i=0}^{n}}$$

נקבל:

$$A^k b_0 = \sum_{i=1}^{n} a_i A^k v_i$$

$$= \sum_{i=1}^{n} a_i \lambda_i^k v_i$$

[illegible]

$$= a_1 \lambda_1^k \sum_{i=1}^{n} \left(\frac{a_i}{a_1}\right)\left(\frac{\lambda_i}{\lambda_1}\right)^k v_i$$

ולכן ככל [illegible] נקבל שהוקטור [illegible] יותר ויותר

מהאיבר הראשון שהוא V_1

Multivariate Calculus

שאלה ק:

ראשית נכתוב את הפונקציה כסכום (בעזרת הפירוק הספקטרלי)

$$f(\bar\sigma) = \sum_{i=1}^{n} \sigma_i u_i u_i^T \bar x$$

u_i היא העמודה ה-i ב-U.

ולכן

$$f(\bar\sigma)_j = \left(\sum_{j=1}^{n} \sigma_i u_i u_i^T \bar x\right)_j$$

~~ולכן~~

וכעת נחשב את היעקוביאן:

$$\frac{\partial f(\bar\sigma)_j}{\partial \sigma_k} = \frac{\partial}{\partial \sigma_k}\left(\sum_{j=1}^{n} \sigma_i u_i u_i^T \bar x\right)_j$$

נשים לב ברור ש אם $j = k$ נקבל ש $\frac{\partial \sigma_i}{\partial \sigma_k}=1$ ואחרת נקבל 0, כלומר של הסכום נקבל רק איבר אחד, כלומר:

$$\boxed{\frac{\partial f(\bar\sigma)_j}{\partial \sigma_k} = \left(u_k u_k^T \bar x\right)_j}$$

אם נכתוב את התשובה בצורה של וקטור העמודה ה-k היעקוביאן יהיה $u_k u_k^T \bar x$

ובסה"כ נקבל ש $\boxed{J_\sigma(f) = U \operatorname{diag}(U^T \bar x)}$

שאלה 6:

נשתמש בכלל השרשרת:

נגדיר

$$x = f(\sigma) - y$$

$$y = \frac{1}{2}\|\mathbf{X}\|^2$$

[illegible]

$$h = \frac{1}{2}\|x(\sigma)\|^2$$

$$\nabla h = \frac{\partial}{\partial x}\left(\frac{1}{2}\|x(\sigma)\|^2\right) \cdot \frac{\partial x(\sigma)}{\partial \sigma}$$

$$= (\sigma) \cdot (f(\sigma) - y)^T \cdot \nabla f(\sigma)$$

שאלה 10

$$g(S)_j = \frac{e^{s_j}}{\sum_i e^{s_i}}$$

$$\frac{\partial g_i}{\partial s_j} = \frac{\partial}{\partial s_j}\left(\frac{e^{s_i}}{\sum_k e^{s_k}}\right)$$

נגזור לפי כלל המנה:

$$= \left(\sum_j e^{s_j}\right)^{-2} \cdot \left(\frac{\partial}{\partial s_j} e^{s_j} - e^{s_i} \cdot \frac{\partial}{\partial s_j} \sum_k e^{s_k}\right)$$

if $i = j$: $\Rightarrow \left(\sum_k e^{s_k}\right)^{-2} \cdot \left(e^{s_i} - e^{s_j} \cdot e^{s_i}\right)$

$$= g(s)_1 \cdot (1 - g(s_i))$$

if $j \neq j$: $\Rightarrow \left(\sum_k e^{s_k}\right)^{-2} \left(0 + e^{s_i} e^{s_j}\right)$

$$= g(s)_i \cdot (-) g(s_j)$$

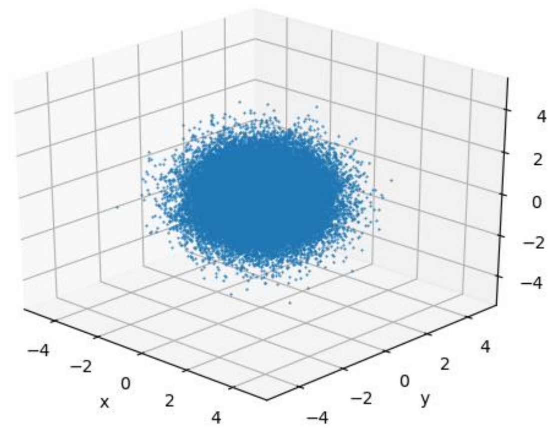
ובסה"כ נקבל:

$$\boxed{\frac{\partial g_i}{\partial s_j} = g(s)_i \left(\delta_{ij} - g(s)_j\right)}$$

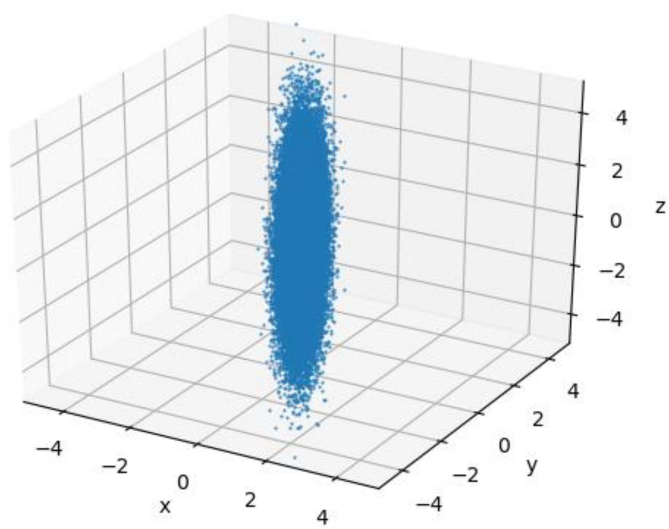
[illegible]

Multivariate Gaussian- practical question

שאלה 11:



שאלה 12:



מטריצת ה COV

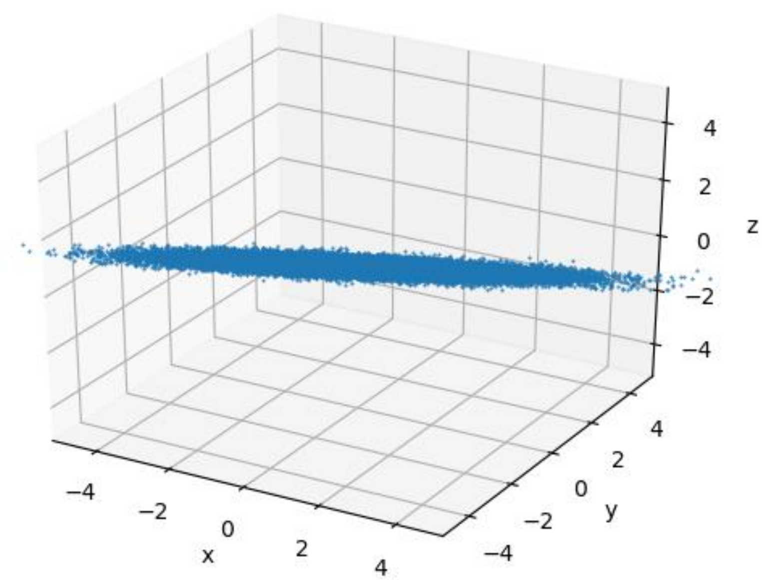
[[0.01 0.   0.]

 [0.   0.25 0.]

[0. 0. 4.]]

ניתן לראות כי עדיין אין קורלציה בין הצירים (רק האיברים על האלכסון קיימים) אבל יש מתיחה – כמו שניתן לראות בסרטוט. יש מתיחה אבל אין סיבוב. עדיין מקביל לצירים.

שאלה 13:



מטריצת הCOV:

[[0.2012325 0.77954016 0.22139559]

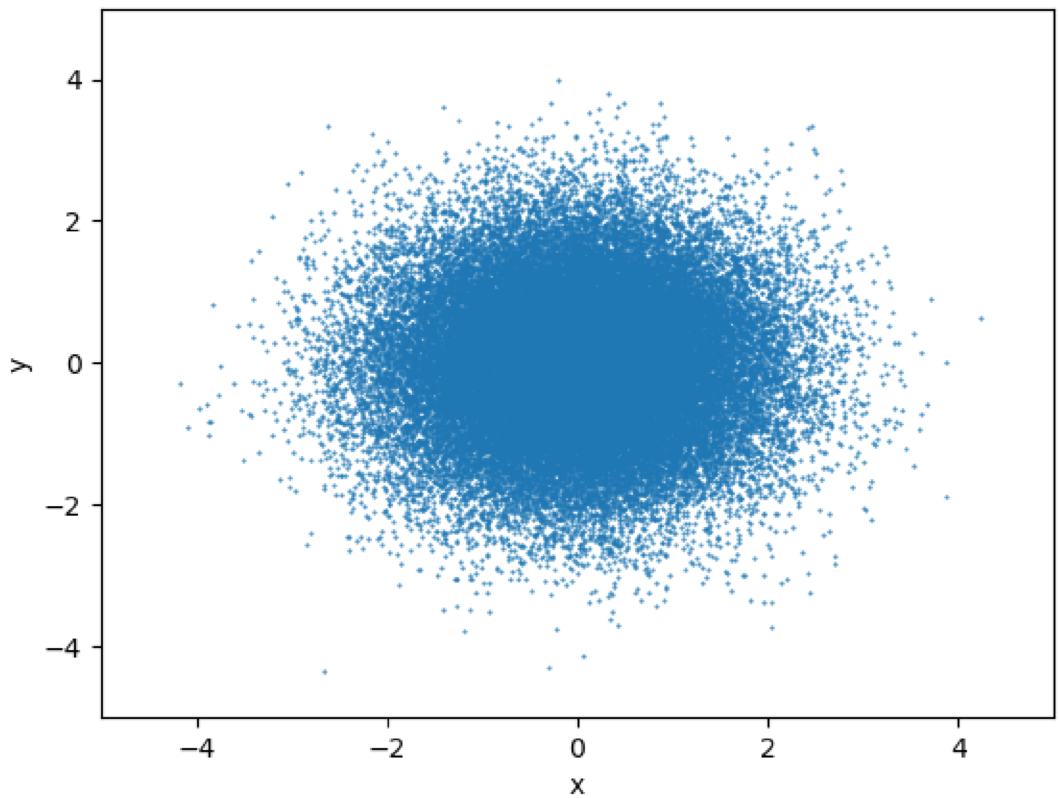
 [0.77954016 3.38267323 1.18514263]

 [0.22139559 1.18514263 0.67609427]]

המטריצה לא אלכסונית- יש גם מתיחה וגם סיבוב (כמו שרואים בסרטוט)

שאלה 14:

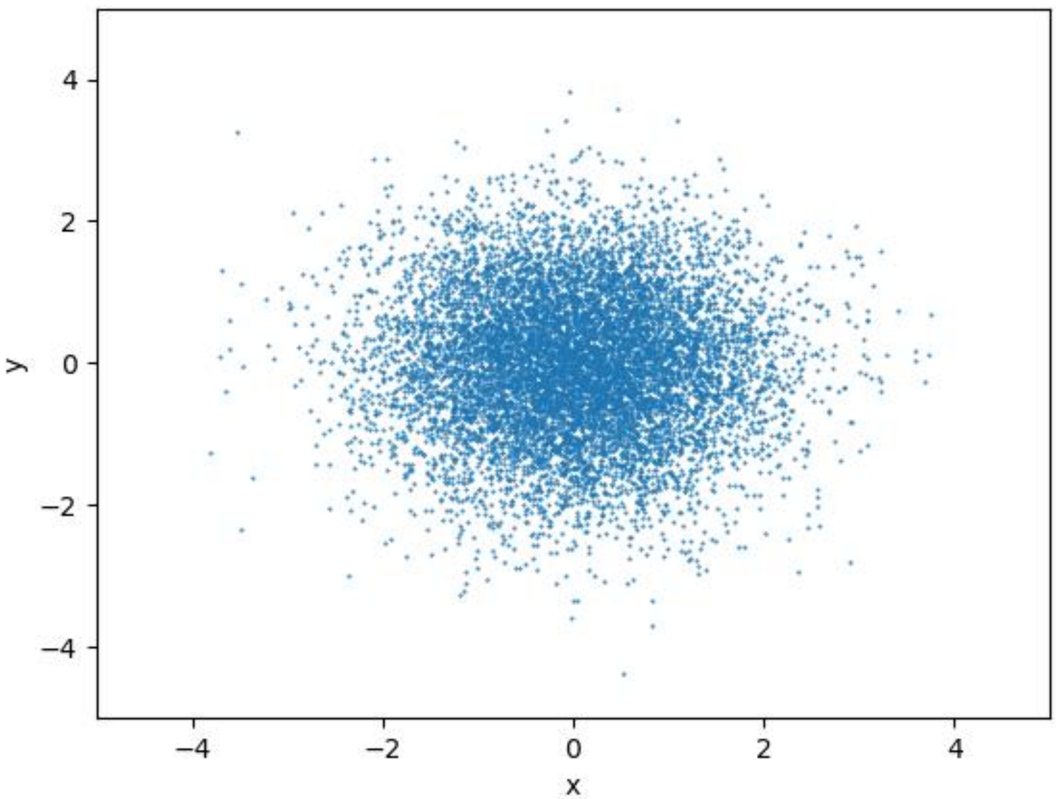
נעריך כי הסרטוט יראה כמו גאוסיאן דו מימד סביב אותם ערכים, וזה אכן מה שקיבלנו:



שאלה 15:

נצפה לראות גאוסיאן בעל אותם ערכים, אבל בצפיפות נמוכה יותר. כלומר צפיפות הנקודות דלילה יותר אבל

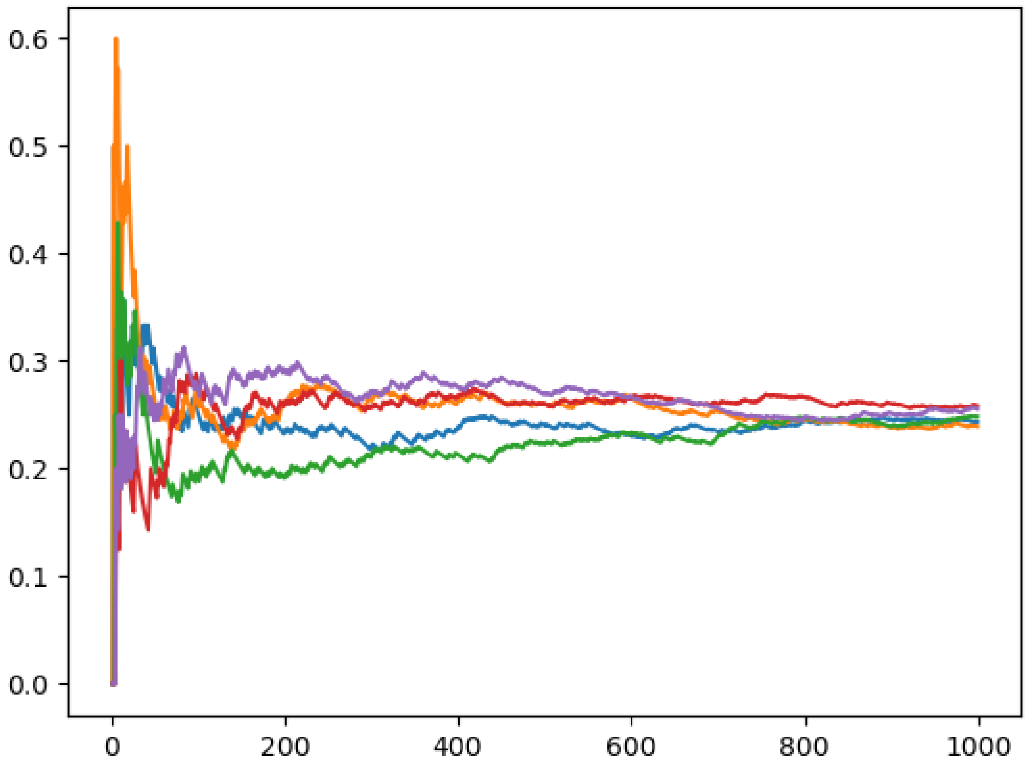
הצורה והערכים זהים- וזה אכן מה שקיבלנו:



Concentration inequalities - practical question

שאלה 16 A:

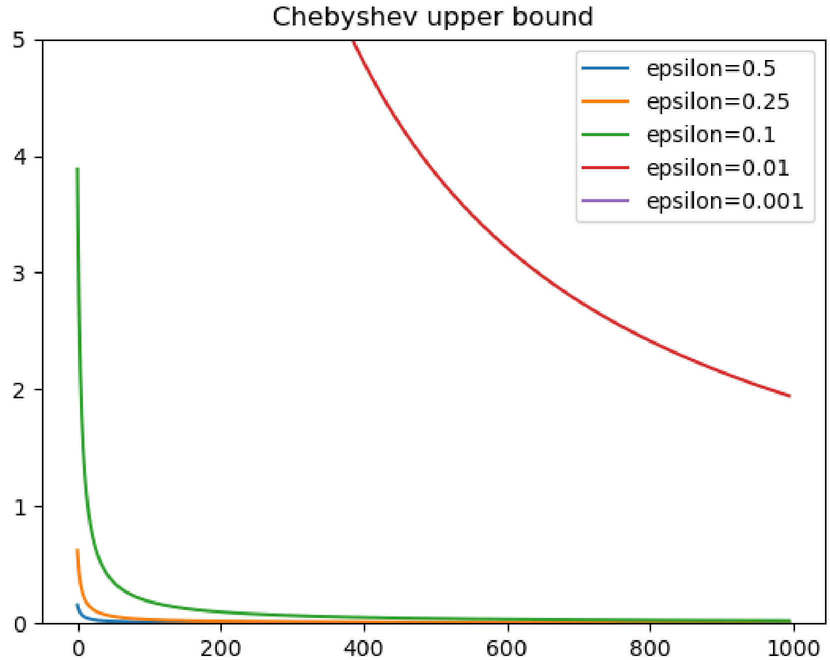
נצפה לראות ככל שM גדל, שאיפה אסימפטומטית לממוצע האמיתי (אשר מוערך להיות 0.25)



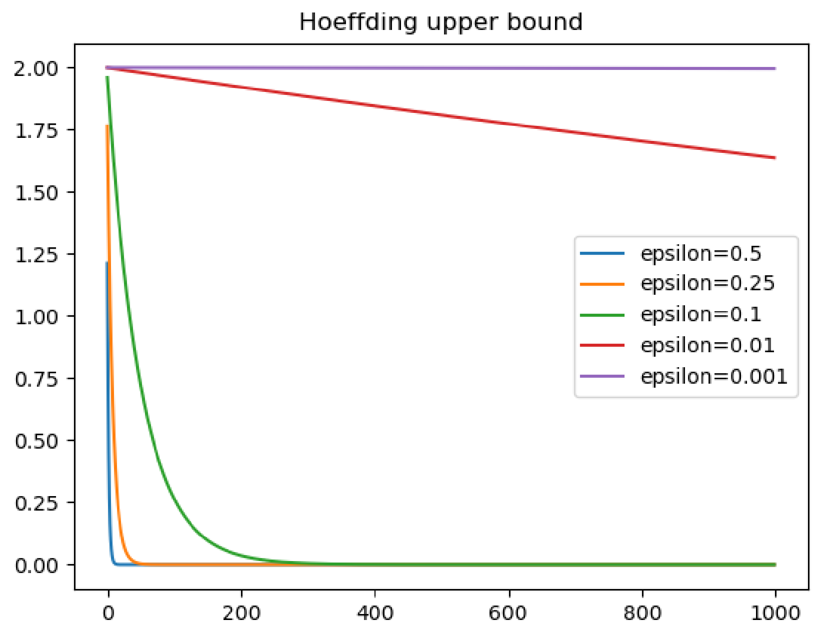
ניתן לראות כי כל השורות מתכנסות לערך מסוים ושכל הערכים הם "קרובים" לערך שציפינו לראות. כולם מתכנסים לאזור ה0.25 ("מתקרבים" זה לא מוגדר – ובהמשך השאלה נדייק יותר)

שאלה 16 B:

* בגרף הראשון חתכתי את ציר הY בנקודה 5 כאשר הקו הסגול מעליה על מנת לקבל גרף ברור יותר



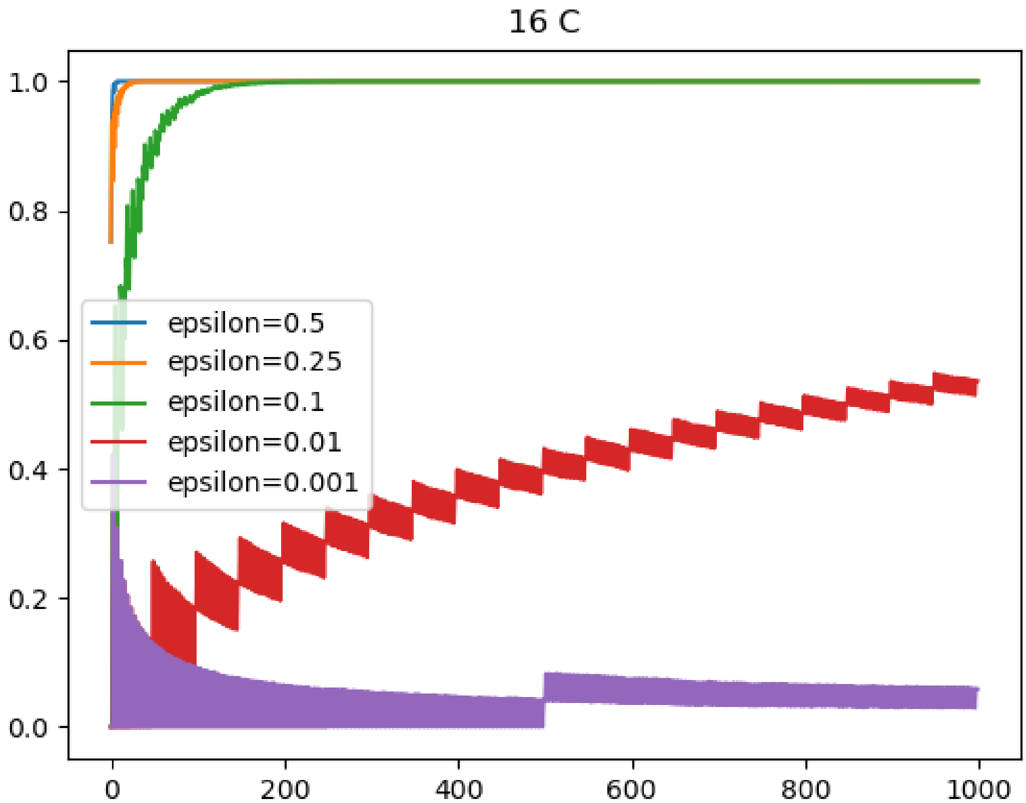
Chebyshev upper bound
epsilon=0.5
epsilon=0.25
epsilon=0.1
epsilon=0.01
epsilon=0.001



Hoeffding upper bound
epsilon=0.5
epsilon=0.25
epsilon=0.1
epsilon=0.01
epsilon=0.001

שאלה C16:

נצפה לראות שככל שM גדל נשאף למאה אחוז (1) – כאשר ככל שאפסילון גדול יותר נשאף מהר יותר



וזה אכן מה שקיבלנו- כאשר בעבור אפסילון קטן מאוד מספר הדגימות לא מספיק

Python Code

Python CODE

```
def conditional(x_y_z):
    temp=x_y_z[0:2,:]
    temp2=x_y_z[2,:]
    # cond=temp2>-0.4 and temp2<0.1
    return temp[:,np.logical_and(temp2>-0.4 , temp2<0.1)]

if __name__ == "__main__":
    mean = [0, 0, 0]
    cov = np.eye(3)
    x_y_z = np.random.multivariate_normal(mean, cov, 50000).T
    # plot_3d(x_y_z);
    s=np.diag([0.1,0.5,2])
    newData=np.matmul(s,x_y_z)
    newCov=np.matmul(np.matmul(s,cov),s.transpose())
    print("scale cove:")
    print(newCov)
    # plot_3d(newData)
    ortoM=get_orthogonal_matrix(3)
    ortoData=np.matmul(ortoM,newData)
    # plot_3d(ortoData)
    print("orto cove:")
    print(np.matmul(np.matmul(ortoM,newCov),ortoM.transpose()))

    # plot_2d(x_y_z[0:2,:])
    # plot_2d(newData[0:2,:])
    # plot_2d(ortoData[0:2,:])

    plot_2d(conditional(x_y_z))
    plot_2d(conditional(newData))
    plot_2d(conditional(ortoData))
```

```
import numpy as np
import matplotlib.pyplot as plt
import math
row=1000
col=100000
data = np.random.binomial(1, 0.25, (col,row))
epsilons=(0.5, 0.25, 0.1, 0.01, 0.001)

# part a:
av=np.zeros(row)
for i in range(5):
```

```
    for m in range(1,row):
        av[m]=np.mean(data[i,1:m])
    plt.plot(av)
plt.show()

# part b:
# Chebyshev
i=1
upperBoundChebyshev=np.zeros(row)
fig = plt.figure()
ax = fig.add_subplot(111)

for epsilon in epsilons:
    var=np.var(data[i,:])
    for m in range(1,row):

        upperBoundChebyshev[m]=var/(m*epsilon*epsilon)
    plt.plot(upperBoundChebyshev[5:,],label='epsilon='+str(epsilon))
legend = ax.legend()
plt.ylim([0,5])
plt.title("Chebyshev upper bound")

plt.show()

# Hoeffding
i=1
upperBoundHoeffding=np.zeros(row)
fig = plt.figure()
ax = fig.add_subplot(111)

for epsilon in epsilons:
    var=np.var(data[i,:])
    for m in range(1,row):

        upperBoundHoeffding[m]=2*math.exp(-2*m*epsilon*epsilon)

    plt.plot(upperBoundHoeffding[1:,],label='epsilon='+str(epsilon))
legend = ax.legend()
plt.title("Hoeffding upper bound")
plt.show()

# part c
fig = plt.figure()
ax = fig.add_subplot(111)
pres=np.zeros(row)
for epsilon in epsilons:
    print(epsilon)
    for m in range(1,row):
        temp=np.mean(data[:,0:m],axis=1)
        pres[m]=(col-np.count_nonzero(temp > 0.25+epsilon) -np.count_nonzero(temp
< 0.25-epsilon))/col
    plt.plot(pres[1:],label='epsilon='+str(epsilon))
```

```
legend = ax.legend()
plt.title("16 C")

plt.show()
```