IML תר 1 - חלק תאורטי:

1

① $V=(1,2,3,4)^T=\begin{pmatrix}1\\2\\3\\4\end{pmatrix}$, $W=(0,-1,1,2)^T=\begin{pmatrix}0\\-1\\1\\2\end{pmatrix}$

מהתרגול:

$$P=\frac{\langle V,W\rangle}{\|W\|^2}\cdot W$$

$$\langle V,W\rangle = 1\cdot 0+2\cdot(-1)+3\cdot 1+4\cdot 2=9$$

$$\|W\|=\sqrt{0^2+(-1)^2+1^2+2^2}=\sqrt{6}$$

$$\Rightarrow \boxed{P=\frac{9}{6}\cdot\begin{pmatrix}0\\-1\\1\\2\end{pmatrix}}$$

② $V=\begin{pmatrix}1\\2\\3\\4\end{pmatrix}$, $W=\begin{pmatrix}1\\0\\1\\-1\end{pmatrix}$:באופן דומה

$$\langle V,W\rangle = 1+2\cdot 0+3+4\cdot(-1)=0$$

כיוון שאבר המונה של ההטלה מוכפל ב $\langle V,W\rangle$ הכל מתאפס

$$\Rightarrow \boxed{P=0}$$

③ עבור וקטורים $V,W\in\mathbb{R}^m$ ← $V^TW=\langle V,W\rangle$,

(מהתרגול) $\langle V,W\rangle=\|V\|\cdot\|W\|\cdot\cos(\theta)\overset{!}{=}0$

כיוון ש $V,W\neq\bar{0}$ ← $\|V\|,\|W\|>0$ ⇐ $\cos(\theta)=0$, $\pi \ge \theta \ge 0$

$$\Rightarrow |\theta|=\left|\frac{\pi}{2}\right|=\pm 90$$

④ עבור מטריצה אורתונורמלית A ווקטור $x\in V$:

$$\|Ax\|=\langle Ax,Ax\rangle=\langle A^TAx,x\rangle=\langle x,x\rangle=\|x\|$$

(כאשר $\|\|$ הינו נורמה אוקלידית)

IML תר' 1 השלמה

(5) ראינו כי עבור מטריצה הפיכה A ניתן לפרק: $A = U\Sigma V^T$

כאשר Σ מטריצה אלכסונית, ועבור מטריצה שכזו:

$$\Sigma^{-1} = \begin{bmatrix} \frac{1}{\Sigma_{11}} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\Sigma_{nn}} \end{bmatrix} \Rightarrow \boxed{A^{-1} = U \cdot \Sigma^{-1} V^T}$$

כמובן שבהינתן SVD של מטריצה, ניתן לחשב את ההופכית שלה בעלות חישוב יותר קל מעלות הפירוק כפי שהודגם.

(6)

$$A = \begin{bmatrix} 5 & 5 \\ -1 & 7 \end{bmatrix} \rightarrow A^T = \begin{bmatrix} 5 & -1 \\ 5 & 7 \end{bmatrix}$$

$$A^T \cdot A = \begin{bmatrix} 26 & 18 \\ 18 & 74 \end{bmatrix} \rightarrow \lambda_1 = 80,\ V_1 = (1,3),\ \lambda_2 = 20,\ V_2 = (3,-1)$$

$$\Rightarrow \hat{V}_1 = \left(\sqrt{\tfrac{1}{10}}, \sqrt{\tfrac{9}{10}}\right),\ \hat{V}_2 = \left(\sqrt{\tfrac{9}{10}}, -\sqrt{\tfrac{1}{10}}\right)$$

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

$$AV \overset{!}{=} U\Sigma,\quad \Sigma = \begin{bmatrix} \sqrt{80} & 0 \\ 0 & \sqrt{20} \end{bmatrix}$$

$$\Rightarrow U = \begin{bmatrix} \sqrt{1/2} & \sqrt{1/2} \\ \sqrt{1/2} & -\sqrt{1/2} \end{bmatrix}$$

(7)

$$b_0 = \sum_i a_i v_i \rightarrow a_0^K b_0 = a_0^K \sum_i a_i v_i = \sum_i a_0^K a_i v_i =$$

$$= \sum_i a_i \lambda_i^K v_i = a_1 \lambda_1^K \left(v_1 + \sum_i \frac{a_i}{a_1}\left(\frac{\lambda_i}{\lambda_1}\right)^K v_i\right),$$

$$\forall i:\ \left(\frac{\lambda_i}{\lambda_1}\right) < 1 \Rightarrow \frac{a_i}{a_1}\left(\frac{\lambda_i}{\lambda_1}\right)^K v_i \xrightarrow[K\to\infty]{} 0$$

נישאר עם $a_1 \lambda_1^K v_1$, ונשים לב כי התלות היחידה על K

אנו מנרמלים את הוקטור לפי גודלו בנורמה $(a_1 \lambda_1^K)$

נישאר עם הוקטור היחידה v_1 לפי סימנו ובסופו של דבר:

$$\rightarrow \lim_{K\to\infty} b_K = \pm v_1$$

(8)

$$J_\sigma(f)_{ij} = \frac{\partial f_i}{\partial \sigma_j} = \frac{\partial}{\partial \sigma_j}\left(U \cdot diag(\sigma)_{ij}\right) = \sum_k U_{ik}\, diag(\sigma)_{kj} =$$

$$= \sigma_j U_{ij} \Rightarrow \left(U \cdot diag(\sigma) \cdot U^T\right)_{ij} = \sum_k \left(U \cdot diag(\sigma)\right)_{ik} \cdot U^T_{kj} =$$

$$= \sum_k \sigma_k U_{ik} U^T_{kj}$$

$$\Rightarrow f_i(\sigma) = \left(U \cdot diag(\sigma) \cdot U^T \cdot x\right)_i = \sum_k \left(U \cdot diag(\sigma) \cdot U^T\right)_{ik} \cdot x_k =$$

$$= \sum_k x_k \cdot \sum_m \sigma_m U_{im} U^T_{mk}$$

$$\Rightarrow J_\sigma(f)_{ij} = \frac{\partial f_i}{\partial \sigma_j} = \frac{\partial}{\partial \sigma_j}\left(\sum_k x_k \cdot \sum_m \sigma_m U_{im} U_{mk}\right) = \sum_k x_k U_{ij} U^T_{jk} =$$

$$= U_{ij} \sum_k x_k U^T_{jk} = U_{ij} \cdot (U^T \cdot x)_j \Rightarrow \boxed{J_\sigma(f) = U diag(U^T \cdot x)}$$

(9)

$$h(\sigma) = \frac{1}{2} \| f(\sigma) - y \|^2$$

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

(10)

$$g(z)_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}}$$

עבור $i = j$:

$$\frac{\partial g_i}{\partial z_j} = g_i(1 - g_i)$$

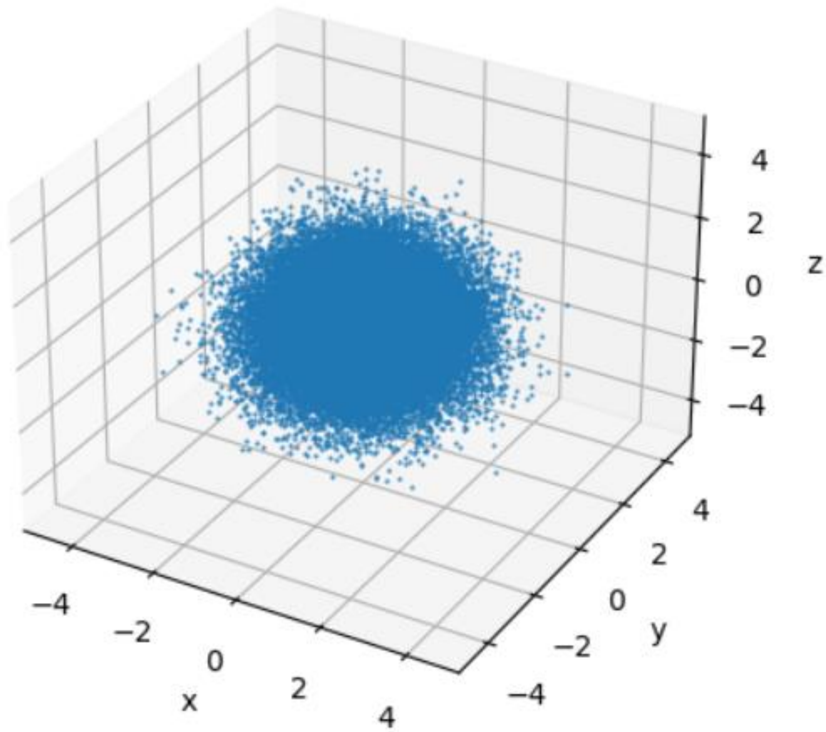
ועבור $i \neq j$:

$$\frac{\partial g_j}{\partial z_j} = \frac{\partial}{\partial z_j} \frac{e^{z_i}}{\sum_k e^{z_k}} = e^{z_i} \frac{\partial}{\partial z_j} \frac{1}{\sum_k e^{z_k}} = e^{z_i} \cdot \left(-\frac{e^{z_j}}{(\sum_k e^{z_k})^2}\right) =$$

$$= -\left(\frac{e^{z_j}}{\sum_k e^{z_k}}\right)\left(\frac{e^{z_i}}{\sum_k e^{z_k}}\right) = \underline{\underline{-g_i g_j}}$$

<u>IML – Ex1- Practical:</u>

<u>11.</u>



<u>12.</u>

Remembering that the cov matrix for a diagonal scaling matrix:

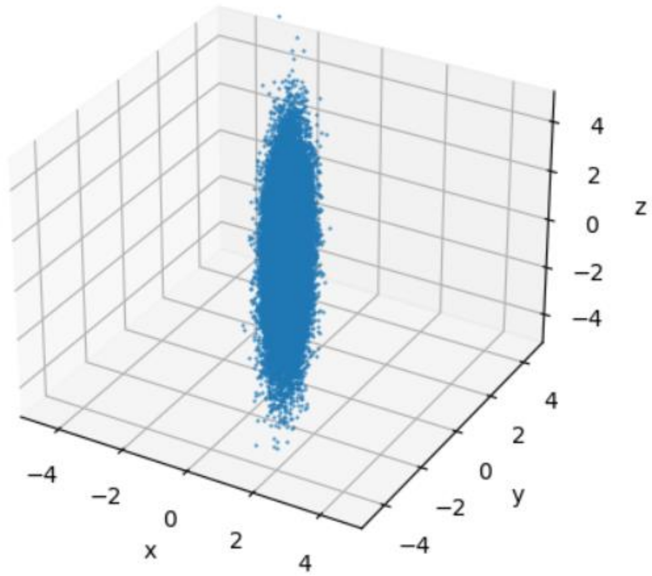
Cov = S*I*S=

S_{11}^2		
	S_{22}^2	
		S_{33}^2

=

0.01		
	0.25	
		4

And the data after transformation:



13.

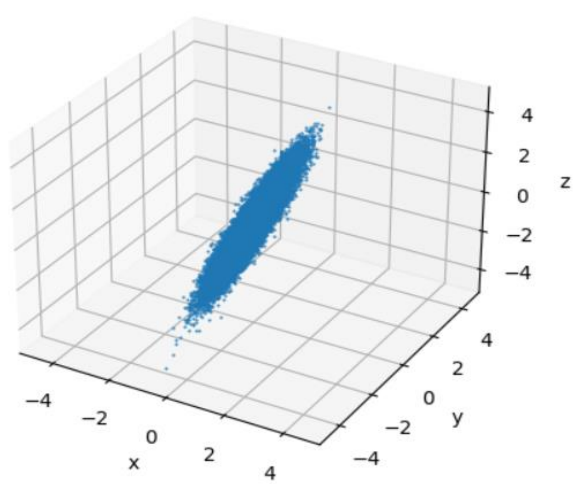
Remembering that the cov for a matrix scaled by an orthogonal matrix:

Cov = U*(S^2)* U^T =

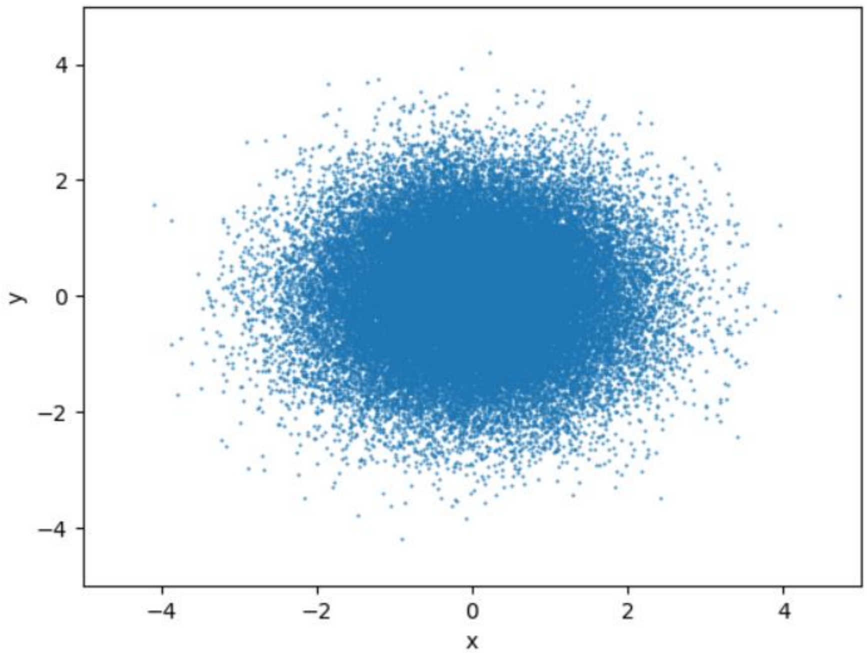
$u_{11}{s^2}_1$	$u_{12}{s^2}_2$	$u_{13}{s^2}_3$
$u_{21}{s^2}_1$	$u_{22}{s^2}_2$	$u_{23}{s^2}_3$
$u_{31}{s^2}_1$	$u_{32}{s^2}_2$	$u_{33}{s^2}_3$

* Since the numeric matrix is randomly generated it is visible when running the attached code.

And the scaled data multiplied by a random orthogonal matrix:

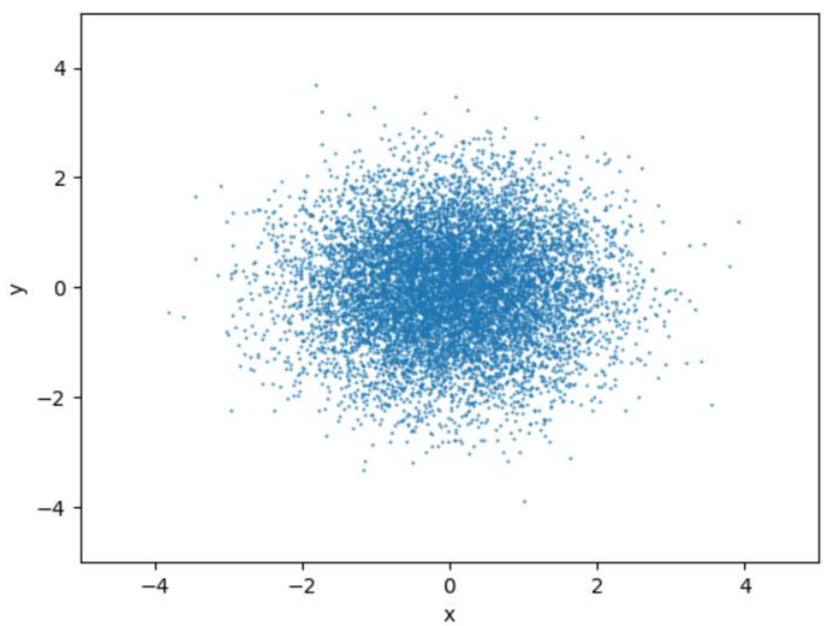


14.



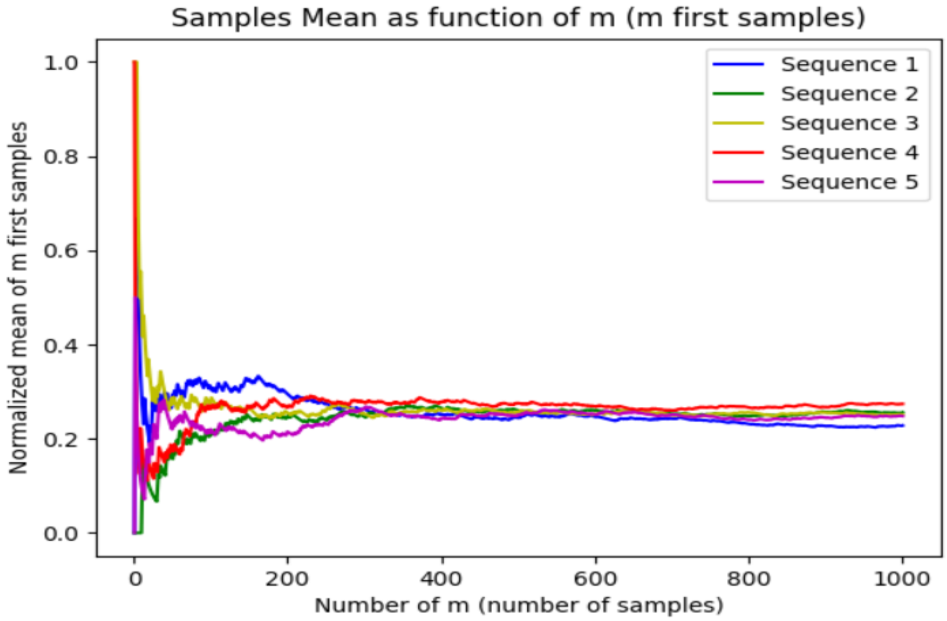
Indeed, the marginal distribution looks like a 2D gaussian.

15.



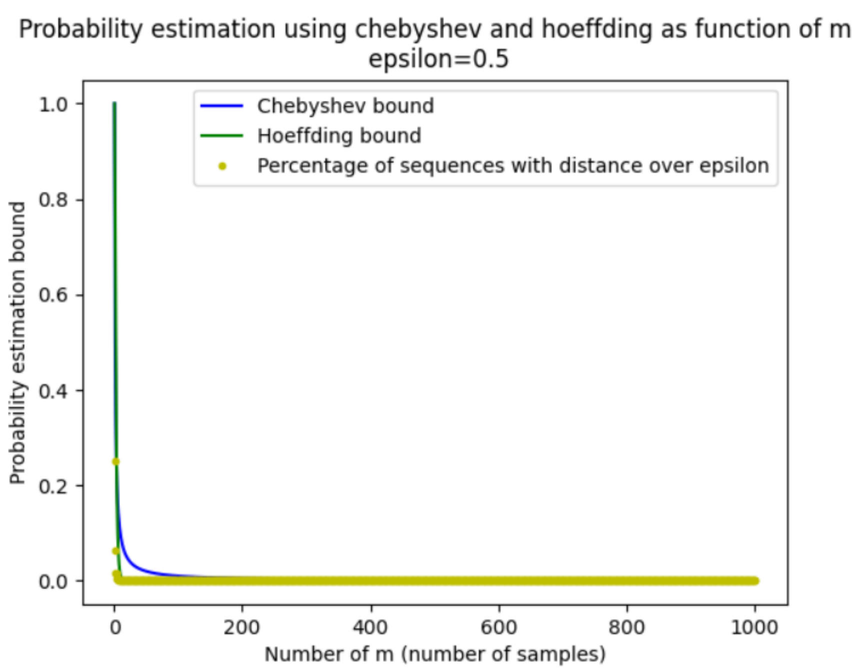
Indeed, the conditional distribution looks like a 2D gaussian.

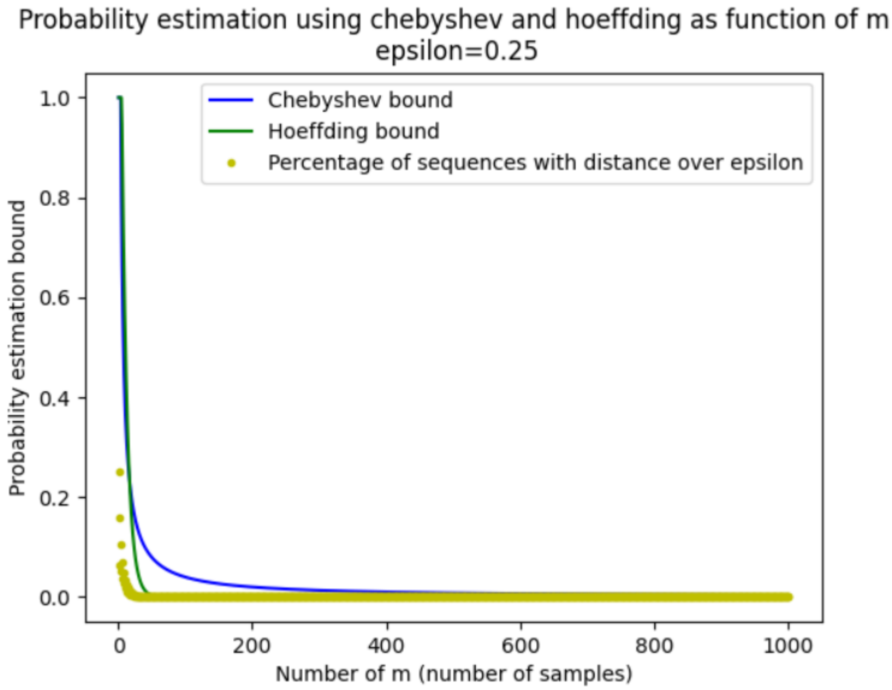
16. A

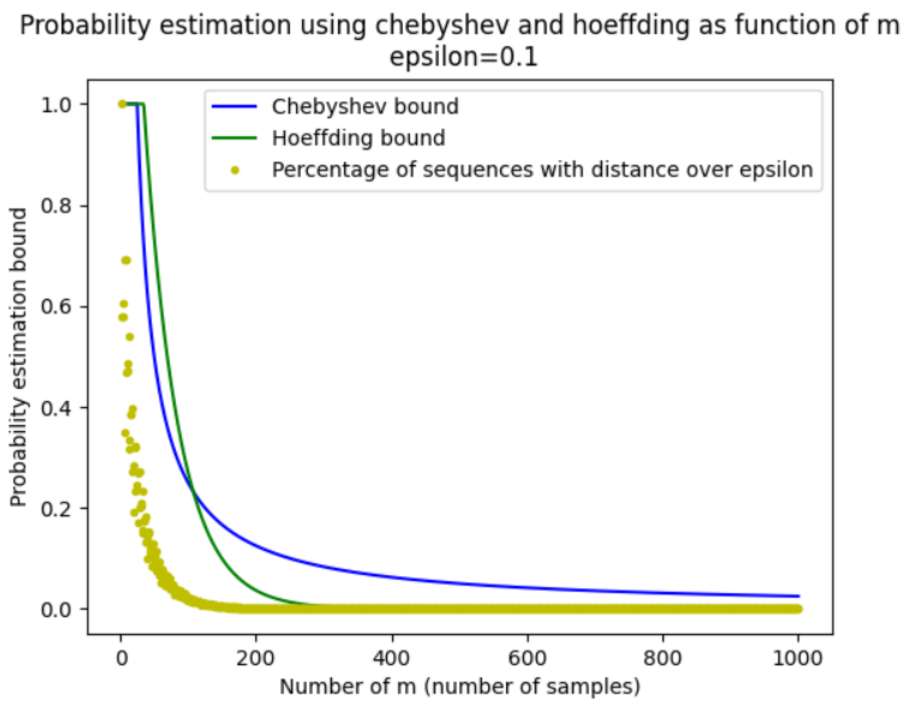


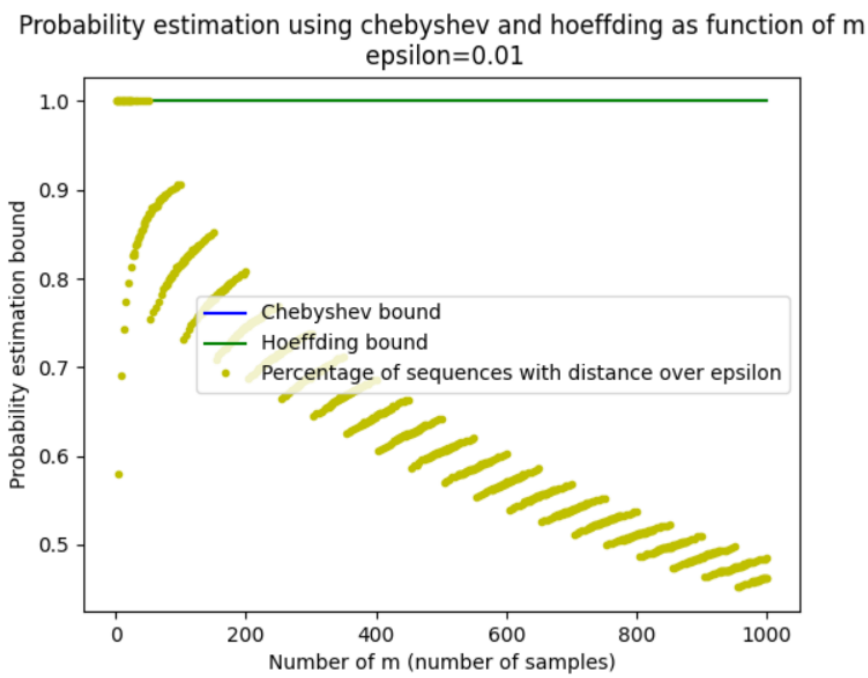
As seen in the graph, as m grows we expect to see convergence to 0.25 - the binomial probability for 1.

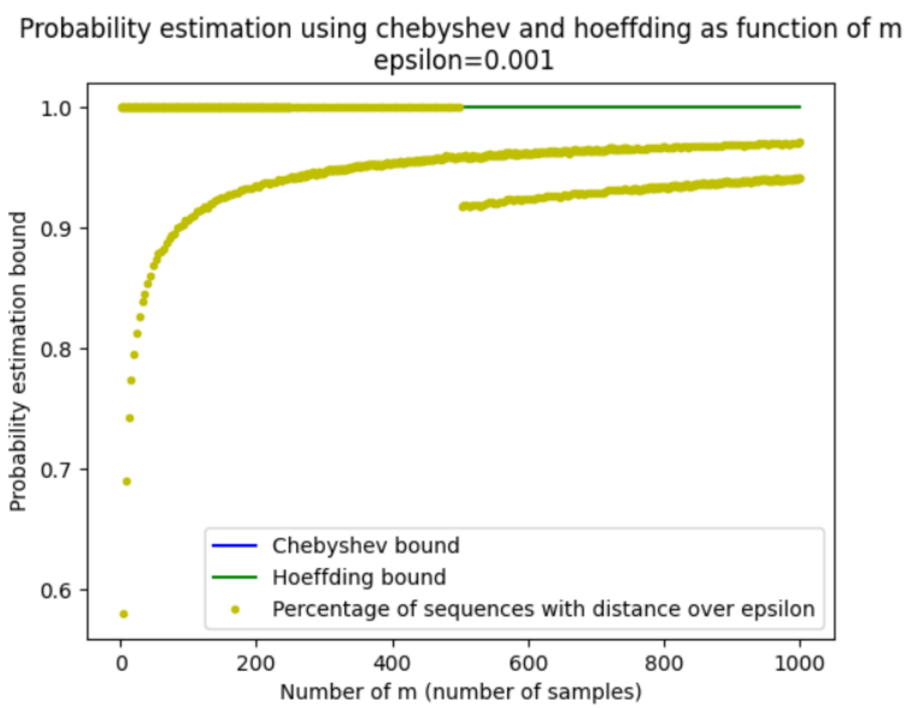
16. B+C




Probability estimation using chebyshev and hoeffding as function of m
epsilon=0.25
Chebyshev bound
Hoeffding bound
Percentage of sequences with distance over epsilon
Probability estimation bound
1.0
0.8
0.6
0.4
0.2
0.0
0
200
400
600
800
1000
Number of m (number of samples)


Probability estimation using chebyshev and hoeffding as function of m
epsilon=0.1
Chebyshev bound
Hoeffding bound
Percentage of sequences with distance over epsilon
Probability estimation bound
1.0
0.8
0.6
0.4
0.2
0.0
0
200
400
600
800
1000
Number of m (number of samples)





As epsilon becomes smaller, we get a higher precent of sequences differing from the expected value by more than epsilon. This makes sense - a bigger difference value holds for more sequences. Also, note that the upper bounds estimated by Chebyshev and Hoeffding hold well for any epsilon or m.