67577 Intro to Machine Learning Guy Lutsker 207029448

1.

$$\text{Calculate the projection of v} = (1,2,3,4)\text{on the vector w} = (0,-1,1,2):$$

$$first\ let\ us\ normalize\ w\ and\ define\ w' = \frac{w}{\|w\|} = \frac{w}{\sqrt{0\cdot 0 + (-1)^2 + 1\cdot 1 + 2\cdot 2}} = \frac{w}{\sqrt{6}}$$

$$and\ so\ w' = \begin{bmatrix} 0 \\ -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \end{bmatrix}$$

$$now, define\ v'\ the\ \ project\ of\ v\ onto\ w':$$

$$emmember\ formula\ from\ Linear\ algebra\ 2\ for\ the\ projection\ is\ v' = \langle w'|v\rangle \cdot w'$$

$$v' = \langle w'|v\rangle \cdot w' = \left(-\frac{2}{\sqrt{6}} + \frac{3}{\sqrt{6}} + \frac{8}{\sqrt{6}}\right) \cdot w' = \frac{9}{\sqrt{6}} \cdot w'$$

$$\rightarrow v' = \frac{9}{\sqrt{6}} \cdot \begin{bmatrix} 0 \\ -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \end{bmatrix} = \begin{bmatrix} 0 \\ -\frac{9}{6} \\ \frac{9}{6} \\ \frac{9}{3} \end{bmatrix} = \begin{bmatrix} 0 \\ -\frac{3}{2} \\ \frac{3}{2} \\ 3 \end{bmatrix}$$

$$note: could\ have\ also\ been\ achived\ by\ the\ formula\ in\ the\ first\ lecture:$$

$$v' = \|v\| \cos\theta \cdot \frac{w}{\|w\|} = \frac{\langle v, w\rangle}{\|w\|^2} w = \frac{3}{2} \begin{bmatrix} 0 \\ -1 \\ 1 \\ 2 \end{bmatrix}$$

2.

$$\text{Calculate the projection of v} = (1,2,3,4)\text{on the vector w} = (1,0,1,-1):$$

$$first\ let\ us\ normalize\ w\ and\ define\ w' = \frac{w}{\|w\|} = \frac{w}{\sqrt{1\cdot 1 + 0\cdot 0 + 1\cdot 1 + (-1)^2}} = \frac{w}{\sqrt{3}}$$

$$and\ so\ w' = \begin{bmatrix} \frac{1}{\sqrt{3}} \\ 0 \\ \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} \end{bmatrix}$$

$$now, define\ v'\ the\ \ project\ of\ v\ onto\ w':$$

$$v' = \langle w'|v\rangle \cdot w' = \left(\frac{1}{\sqrt{3}} + \frac{3}{\sqrt{3}} - \frac{4}{\sqrt{3}}\right) \cdot w' = 0$$

$$\rightarrow v \perp w$$

$$note: could\ have\ also\ been\ achived\ by\ the\ formula\ in\ the\ first\ lecture:$$

$$v' = \ \|v\| \cos\theta \cdot \frac{w}{\|w\|} = \frac{\langle v, w\rangle}{\|w\|^2} w = 0 \cdot w = 0$$

3.

Prove the angle between two non zero vectors v, w ∈ $\mathbb{R}^m$ is 90 iff $\langle \mathrm{v} | \mathrm{w} \rangle = 0$.

$$WLOG\ let\ 0 \neq v, w \in \mathbb{R}^m\ s.t\ the\ angle\ between\ them\ is + 90,$$
$$that\ means\ that\ if\ we\ project\ v\ onto\ w\ we\ will\ get\ the\ 0\ vector$$
$$and\ since\ neither\ v, nor\ w\ are\ the\ zero\ vector, we\ know\ that\ during\ the\ normalization$$
$$process\ w'(w\ normalized)\ will\ not\ be\ 0, and\ so\ the\ only\ way\ v'(v's\ projection\ onto\ w)$$
$$will\ be\ the\ zero\ vector \Leftrightarrow \langle \mathrm{v} | \mathrm{w} \rangle = 0.$$
$$that\ true\ since\ the\ formula\ for\ projection\ is:$$
$$0 = v' = \langle w' | v \rangle \cdot \overbrace{w'}^{not\ zero} \quad \overset{must\ happend}{\Rightarrow} \quad \langle w' | v \rangle = 0 \to \langle w | v \rangle = 0.\ \ as\ required$$
$$note: could\ have\ also\ been\ achived\ by\ the\ formula\ in\ the\ first\ lecture:$$
$$we\ know\ that\ \langle v, w \rangle = 0 \to \|v\|\|w\| \cos\theta = 0 \to \cos\theta = 0 \to \theta = \pm 90$$

4.

Prove that Orthonormal matrices are isometric transformations. That is let T : V → W be some linear transformation and A the corresponding matrix. Then if A is orthonormal then $\forall x \in V\ \ \|Ax\|_2 = \|x\|_2$

$$proof:$$
$$T\ is\ orthogonal, let\ v \in V\ lets\ look\ at:$$
$$in\ Linear\ Algebra\ 2\ we\ defined\ an\ orthogonal\ transformation\ T\ as\ a\ transformation\ that\ holds\ that$$
$$\forall v, w \in V\ \langle v, w \rangle = \langle T(v), T(w) \rangle$$
$$and\ so\ the\ proof\ is\ straight\ forward:$$
$$\|Av\|_2 = \sqrt{\langle T(v), T(v) \rangle} \quad \overset{Orthogonal}{=} \quad \sqrt{\langle v, v \rangle} = \|v\|_2$$

$$if\ the\ definition\ you\ seeked\ is\ that\ A\ is\ orthogonal\ if\ A^T A = I$$
$$then\ lets\ look\ at:$$
$$\langle Av, Av \rangle \quad \overset{Parsabel}{=} \quad (Av)^T Av = v^T A^T Av \quad \overset{orthogonal}{=} \quad v^T I v = v^T v \quad \overset{Parsabel}{=} \quad \langle v, v \rangle$$
$$\to \|Av\|_2 = \|v\|_2\ as\ required$$

5.

Assume A is invertible. Write a formula for the inverse of A using only the matrices U, D, V where UDV T is an SVD decomposition of A. Many learning algorithm implementations require calculating the inverse of a matrix. Explain why knowing the SVD decomposition of matrix is usefull in this context.

$$A^{-1} = (UDV^T)^{-1} = (V^T)^{-1} D^{-1} U^{-1} \quad \overset{\substack{V,U\ are\ orthogonal \\ and\ so\ V^T = V^{-1}}}{=} \quad VD^{-1}U^T$$

$$its\ usefull\ to\ know\ the\ SVD\ decomposition\ since\ it\ enables\ us\ to\ find\ the\ inverse\ of\ a\ function$$
$$farirly\ easily - just\ 3\ matrix\ multiplications, which\ can\ be\ easier\ than\ to\ apply$$
$$gauuses\ elimonation\ procces$$
$$since\ gaussian\ elimination\ is \sim O(n^3) and\ matrix\ is \sim O(n^2)$$
$$and\ so\ O(n^3) > 3 \cdot O(n^2)$$

6.

$$\text{Find an SVD of C} = \text{UDV}^T = \begin{bmatrix} 5 & 5 \\ -1 & 7 \end{bmatrix}$$

$$C^TC = \begin{bmatrix} 5 & -1 \\ 5 & 7 \end{bmatrix}\begin{bmatrix} 5 & 5 \\ -1 & 7 \end{bmatrix} = \begin{bmatrix} 26 & 18 \\ 18 & 74 \end{bmatrix}$$

$$also: C^TC = (UDV^T)^T(UDV^T) = VD^T \cdot \overbrace{U^TU}^{Orthogonal \to I} \cdot DV^T \overset{D=D^T}{=} VD^2V^T = \begin{bmatrix} 26 & 18 \\ 18 & 74 \end{bmatrix}$$

$$lets\ seek\ for\ D^2 \to diagnolize\ VD^2V^T = \begin{bmatrix} 26 & 18 \\ 18 & 74 \end{bmatrix}:$$

$$\chi_{VD^2U}(x) = \det\begin{bmatrix} 26 & 18 \\ 18 & 74 \end{bmatrix} - xI = \det\begin{bmatrix} 26-x & 18 \\ 18 & 74-x \end{bmatrix} = (26-x)(74-x) - 18^2$$

$$= x^2 - 100x + 1600 \to x \in \{80,20\} \to D^2 = \begin{bmatrix} 80 & 0 \\ 0 & 20 \end{bmatrix} \to \boldsymbol{D} = \begin{bmatrix} \boldsymbol{\sqrt{80}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\sqrt{20}} \end{bmatrix}$$

$$seek\ for\ V \to seek\ for\ eigenvectors:$$

$$for\ eigenvalue\ 80: \ker VD^2V^T - 80I = \ker\begin{bmatrix} -54 & 18 \\ 18 & -6 \end{bmatrix} \to \ker\begin{bmatrix} 0 & 0 \\ 18 & -6 \end{bmatrix} \to \ker\begin{bmatrix} 0 & 0 \\ 3 & -1 \end{bmatrix}$$

$$= span\left\{\begin{bmatrix} 1 \\ 3 \end{bmatrix}\right\} = span\left\{\begin{bmatrix} 1 \\ 3 \end{bmatrix} \cdot \frac{1}{\sqrt{10}}\right\}\ and\ \left\|\left\|\begin{bmatrix} \frac{1}{\sqrt{10}} \\ \frac{3}{\sqrt{10}} \end{bmatrix}\right\|\right\| = 1$$

$$for\ eigenvalue\ 20: \ker VD^2V^T - 20I = \ker\begin{bmatrix} 6 & 18 \\ 18 & 54 \end{bmatrix} \to \ker\begin{bmatrix} 6 & 18 \\ 0 & 0 \end{bmatrix} \to \ker\begin{bmatrix} 1 & 3 \\ 0 & 0 \end{bmatrix}$$

$$= span\left\{\begin{bmatrix} -3 \\ 1 \end{bmatrix}\right\} = span\left\{\begin{bmatrix} -3 \\ 1 \end{bmatrix} \cdot \frac{1}{\sqrt{10}}\right\}\ and\ \left\|\left\|\begin{bmatrix} -\frac{3}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} \end{bmatrix}\right\|\right\| = 1$$

$$and\ so\ \boldsymbol{V} = \begin{bmatrix} \boldsymbol{\frac{1}{\sqrt{10}}} & -\boldsymbol{\frac{3}{\sqrt{10}}} \\ \boldsymbol{\frac{3}{\sqrt{10}}} & \boldsymbol{\frac{1}{\sqrt{10}}} \end{bmatrix}\left(notice\ that\ V^TV = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \to V\ is\ orthogonal\right).$$

$$and\ V^{-1} = \begin{bmatrix} \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \\ -\frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \end{bmatrix} = V^T, since\ V\ is\ orthogonal$$

$$and\ from\ CV = UD\ we\ can\ deduce\ that: CVD^{-1} = U$$

$$U = CVD^{-1} = \begin{bmatrix} 5 & 5 \\ -1 & 7 \end{bmatrix}\begin{bmatrix} \frac{1}{\sqrt{10}} & -\frac{3}{\sqrt{10}} \\ \frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \end{bmatrix}D^{-1} = \begin{bmatrix} 5 & 5 \\ -1 & 7 \end{bmatrix}\begin{bmatrix} \frac{1}{\sqrt{10}} & -\frac{3}{\sqrt{10}} \\ \frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \end{bmatrix}\begin{bmatrix} \frac{1}{4\sqrt{5}} & 0 \\ 0 & \frac{1}{2\sqrt{5}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\frac{1}{\sqrt{2}}} & \boldsymbol{\frac{-1}{\sqrt{2}}} \\ \boldsymbol{\frac{1}{\sqrt{2}}} & \boldsymbol{\frac{1}{\sqrt{2}}} \end{bmatrix}$$

$which\ is\ also\ orthogonal.$

$$and\ in\ conclusion:$$

$$\text{C} = \begin{bmatrix} 5 & 5 \\ -1 & 7 \end{bmatrix} = \text{UDV}^T = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}\begin{bmatrix} \sqrt{80} & 0 \\ 0 & \sqrt{20} \end{bmatrix}\begin{bmatrix} \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \\ -\frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \end{bmatrix} \to ineed\ equals\ \begin{bmatrix} 5 & 5 \\ -1 & 7 \end{bmatrix}$$

$$as\ required\ :)$$

7.

. (Power Iteration) In this section we will implement an algorithm for SVD decomposition, we will use the relation between SVD of A to EVD of A^TA that we saw in recitation. For some $A \in M_{m\times n}(\mathbb{R})$, define $C_0 = A^TA$.

Let $\lambda_1, \lambda_2, ...\lambda_n$ be the eigenvalues of C_0, with the corresponding eigenvectors $v_1, v_2, ...v_n$, ordered such that $\lambda_1 \geqslant \lambda_2 \geqslant ... \geqslant \lambda_n$.

Assume $\lambda_1 > \lambda_2$, where λ_1 is the largest eigenvalue and λ_2 is the second-largest one.

Define $b_{k+1} = \frac{C_0 b_k}{\|C_0 b_k\|}$, and initialize b_0 randomly.

Show that: $\lim_{k\to\infty} b_k = v_1$

Hint: use EVD decomposition of C_0 and represent b_0 accordingly. You can assume that $b_0 = \sum_{i=0}^{n} a_i v_i$, where $a_1 \neq 0$. As b_0 is initialized randomly, the probability of $a_1 = 0$ is zero.

$$let\ b_0\ be\ a\ randomaly\ chosen\ vector\ in\ V\ s.t\ b_0 = \sum_{i=1}^{n} a_{0_i} v_i \quad (a_1 \neq 0)$$

$$first\ of\ all, notice\ that\ C_0 = A^TA \in \mathbb{R}^{nXn}\ is\ symmetric\ since: C_0^T = (A^TA)^T = A^TA = C_0$$

$$and\ so\ from\ the\ spectral\ theorem\ over\ \mathbb{R}\ C_0\ \ is\ self-adjoint \to \exists U, D \in \mathbb{R}^{nXn},$$
$$s.t\ U\ is\ orthogonal\ and\ D\ is\ diagonal\ s.t\ C_0 = UDU^{-1}.$$
$$also, we\ know\ from\ the\ question\ the\ eigen\ values\ of\ C_o\ and\ so\ we\ know\ that:$$
$$D_{ij} = \begin{cases} \lambda_i & i = j \\ 0 & i \neq j \end{cases}.$$

$$we\ can\ think\ of\ C_0\ as\ being\ diagonal (because\ we\ can\ use\ the\ eigenvectors\ as\ base$$
$$for\ the\ subspace\ we\ work\ in)\ and\ so\ from\ now\ on\ i\ will\ assume\ that\ C_0 = D$$

$$lets\ look\ at\ thed\ numerator\ of\ \ b_{k+1} = C_0^k b_0 = \sum_{i=1}^{n} C_0^k (a_{0_i} v_i) \overset{diagonal}{\triangleq}$$

$$\sum_{i=1}^{n} a_{k_i} \lambda_i^k v_i = a_1 \lambda_1^k \left(v_1 + \sum_{i=2}^{n} \frac{a_i}{a_1} \left(\frac{\lambda_i}{\lambda_1}\right)^k \cdot v_i \right)$$

$$meaning\ that\ for\ k \to \infty\ we\ get\ that:$$

$$\lim_{k\to\infty} b_k = \lim_{k\to\infty} b_{k+1} = \lim_{k\to\infty} \frac{a_1\lambda_1^k \left(v_1 + \sum_{i=2}^{n} \frac{a_i}{a_1}\left(\frac{\lambda_i}{\lambda_1}\right)^k \cdot v_i\right)}{\left\| a_1\lambda_1^k \left(v_1 + \sum_{i=2}^{n} \frac{a_i}{a_1}\left(\frac{\lambda_i}{\lambda_1}\right)^k \cdot v_i\right)\right\|} = \lim_{k\to\infty} \frac{a_1\lambda_1^k \left(v_1 + \sum_{i=2}^{n} \frac{a_i}{a_1}\left(\frac{\lambda_i}{\lambda_1}\right)^k \cdot v_i\right)}{|a_1\lambda_1^k| \left\| \left(v_1 + \sum_{i=2}^{n} \frac{a_i}{a_1}\left(\frac{\lambda_i}{\lambda_1}\right)^k \cdot v_i\right)\right\|}$$

$$\overset{\frac{a_1\lambda_1^k}{|a_1\lambda_1^k|} = \pm 1}{\triangleq}\ and\ \overset{\substack{\frac{\lambda_i}{\lambda_1} < 1 \\ because \\ \lambda_1 \geq \lambda\_i}}{\triangleq} \pm \frac{\left(v_1 + \sum_{i=2}^{n} \frac{a_i}{a_1} 0 \cdot v_i\right)}{\left\|\left(v_1 + \sum_{i=2}^{n} \frac{a_i}{a_1} 0 \cdot v_i\right)\right\|} = \pm \frac{v_1}{\|v_1\|} \overset{\|v_1\|=1}{\triangleq} \pm v_1$$

8.

Let $x \in \mathbb{R}^n$ be a fixed vector and $U \in \mathbb{R}^{n \times n}$ a fixed orthogonal matrix. Calculate the Jacobian of the function $f : \mathbb{R}^n \to \mathbb{R}^n$:

$$f(\sigma) = U diag(\sigma) U^T x$$

Here $diag(\sigma)$ is an $n \times n$ matrix where $diag(\sigma)_{ij} = \begin{cases} \sigma_i & i = j \\ 0 & i \neq j \end{cases}$

$$f(\sigma) = U \cdot \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \dots & \\ & & & \sigma_n \end{bmatrix} \cdot U^T \cdot x \overset{\substack{u_i\, is\, U's \\ i's\, column}}{\triangleq} [u_1\sigma_1, \dots, u_n\sigma_n]U^t x \overset{\substack{matrix \\ multiplication}}{\triangleq}$$

$$= \sum_{i=1}^{n} u_i \sigma_i u_i^T x \overset{\substack{\sigma_i\, is\, scalar\, and \\ so\, can\, change \\ position}}{\triangleq} \sum_{i=1}^{n} \sigma_i u_i u_i^T x$$

$$which\ means\ that\ the\ derivative\ of\ f_j\ according\ to\ \sigma_i = \frac{\partial f_i}{\partial \sigma_j} = \left[u_i u_i^T x\right]_j$$

$$because\ all\ of\ the\ \sigma_j (s.t\ j \neq i)\ will\ 'not\ survive'\ the\ derivative\ process$$

$$and\ f_j\ is\ the\ j's\ position\ in\ the\ output\ vector\ and\ so\ over\ all\ we\ get\ \left[u_i u_i^T x\right]_j$$

$$s.t\ Jacobian_{ij} = \left[u_i u_i^T x\right]_j$$

9.

Use the chain rule to calculate the gradient of h:

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

$$according\ to\ the\ chain\ rule\ , \nabla h = 2 \cdot \frac{1}{2} \cdot (f(\sigma) - y)^T \cdot \mathcal{J}_{f(\sigma)}$$

$$\to \nabla h = (f(\sigma)^T - y^T) \cdot \mathcal{J}_{f(\sigma)}$$

10.

Calculate the Jacobian of the softmax function (initial steps can be found in recitation file):

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

$$the\ function\ g\ operates\ thustly: g\left(\begin{bmatrix} z_1 \\ \cdots \\ z_K \end{bmatrix}\right) = \begin{bmatrix} S_1 \\ \cdots \\ S_K \end{bmatrix}\ s.t\ S_i = \frac{e^{z_i}}{\sum_{n=1}^{K} e^{z_n}}$$

$$we\ need\ to\ calculate\ the\ jacobian\ of\ g\ which\ is\ \left[\mathcal{J}_g\right]_{ij} = \frac{\partial S_i}{\partial z_j}$$

$$lets\ examine\ \frac{\partial S_i}{\partial z_j} = derive\left(\frac{e^{z_i}}{\sum_{n=1}^{K} e^{z_n}}\right) w.r.t\ z_j \overset{\substack{denote \\ \sum_{n=1}^{K} e^{z_n} = h}}{\triangleq} \frac{derive(e^{z_i}) w.r.t\ z_j}{h}$$

$$let\ us\ split\ the\ problem\ fot\ 2\ cases: i = j, i \neq j:$$

$$i = j:$$

$$\frac{\partial}{\partial z_j} \cdot \frac{e^{z_i}}{h} = \frac{e^{z_i} \cdot \sum_{n=1}^{K} e^{z_n} - e^{z_i} \cdot e^{z_j}}{(\sum_{n=1}^{K} e^{z_n})^2} = \frac{e^{z_i}}{\sum_{n=1}^{K} e^{z_n}} \cdot \frac{\sum_{n=1}^{K} e^{z_n}\ - e^{z_j}}{\sum_{n=1}^{K} e^{z_n}} = S_i - \left(1 - S_j\right)$$

$$i \neq j:$$

$$\frac{\partial}{\partial z_j} \cdot \frac{e^{z_i}}{h} = \frac{0 \cdot \sum_{n=1}^{K} e^{z_n} - e^{z_i} \cdot e^{z_j}}{(\sum_{n=1}^{K} e^{z_n})^2} = \frac{-e^{z_i} \cdot e^{z_j}}{(\sum_{n=1}^{K} e^{z_n})^2} = -\left(\frac{e^{z_i}}{\sum_{n=1}^{K} e^{z_n}} \cdot \frac{e^{z_j}}{\sum_{n=1}^{K} e^{z_n}}\right)$$

$$= -\left(S_i + S_j\right)$$

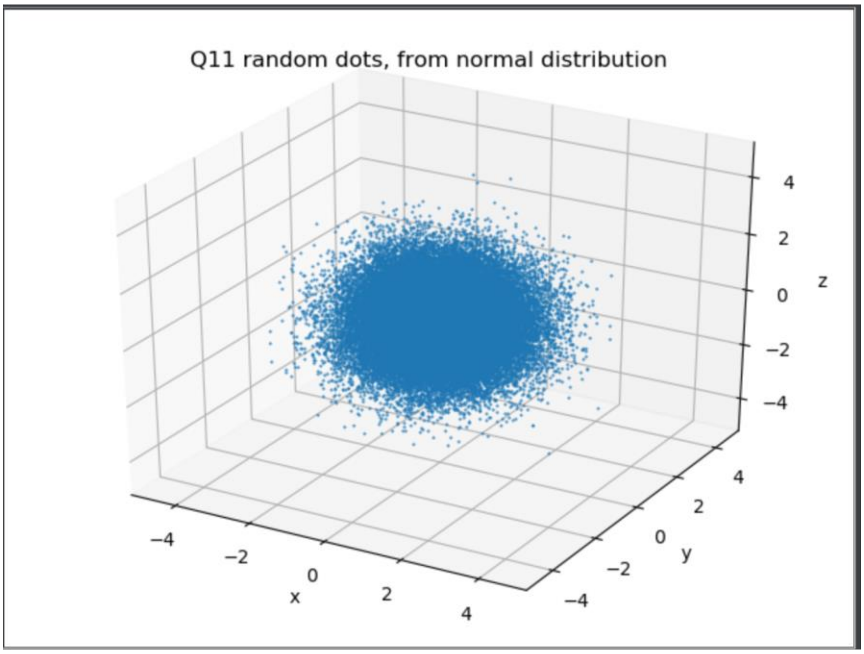
$$and\ so, the\ jacobian\ of\ g\ is:$$

$$\left[\mathcal{J}_g\right]_{ij} = \frac{\partial S_i}{\partial z_j} = \begin{cases} S_i - \left(1 - S_j\right) & i = j \\ -\left(S_i + S_j\right) & i \neq j \end{cases}$$

11.

$$Cov\ matrix = I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Plot:



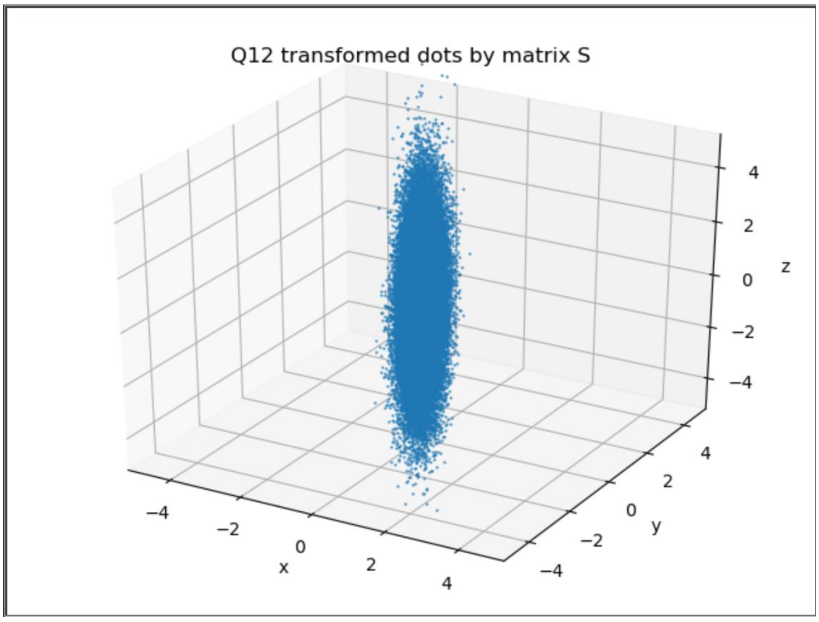
:

12. Transformed data with scaling matrix $S = \begin{bmatrix} 0.1 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 2 \end{bmatrix}$

Analytical Cov matrix we will define as $T = S \cdot I \cdot S^T = \begin{bmatrix} 0.01 & 0 & 0 \\ 0 & 0.25 & 0 \\ 0 & 0 & 4 \end{bmatrix}$

Numerical Cov matrix = $\begin{bmatrix} 0.0099 & 0.00006 & -0.0000003 \\ 0.00006 & 0.0247 & -0.008 \\ -0.0000003 & = 0.008 & 3.989 \end{bmatrix}$

Plot:



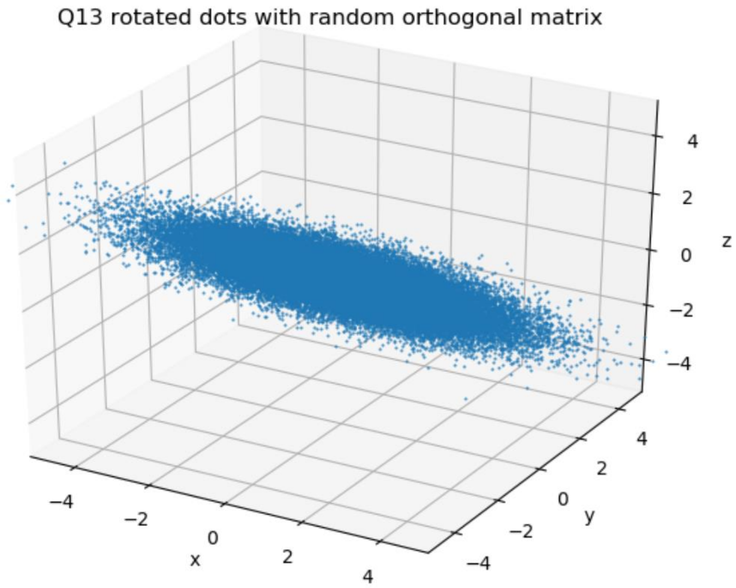
:

13. The random matrix I got: $O = \begin{bmatrix} -0.68337509 & 0.16398812 & -0.71141154 \\ 0.39172021 & -0.73994146 & -0.54684724 \\ -0.61607935 & -0.65237606 & 0.4414201 \end{bmatrix}$

Analytical Cov matrix we will define as $R = O \cdot T \cdot O^T =$
$\begin{bmatrix} 2.03581859 & 1.52312145 & -1.27866077 \\ 1.52312145 & 1.33458042 & -0.84729075 \\ -1.27866077 & -0.84729075 & 0.88960099 \end{bmatrix}$

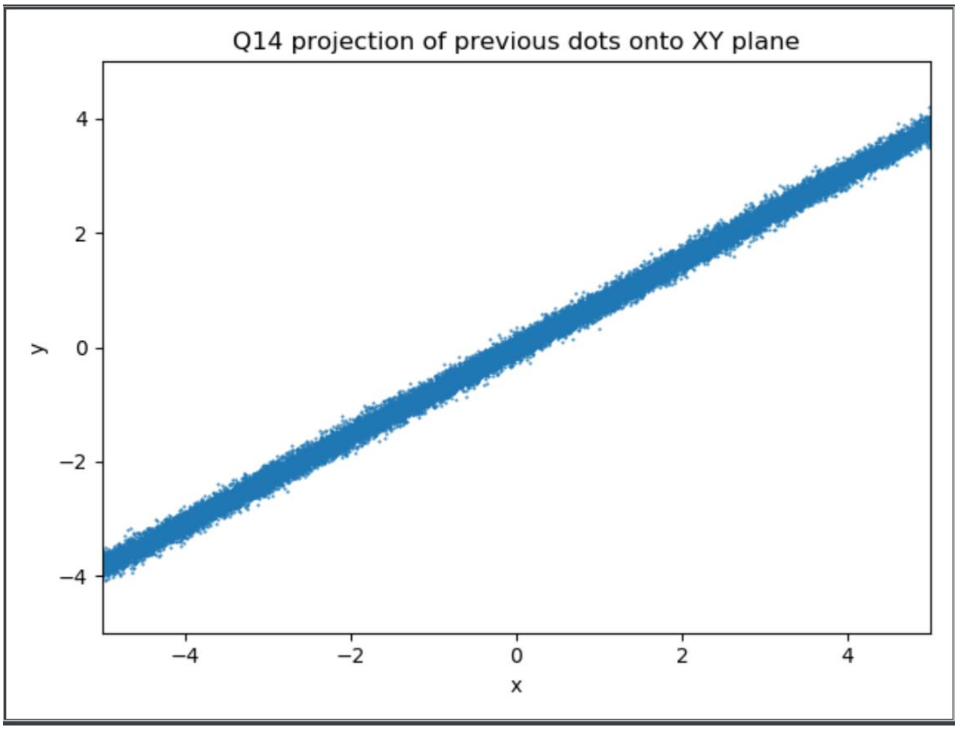
Numerical Cov matrix = $\begin{bmatrix} 2.03375934 & 1.52122005 & -1.27632887 \\ 1.52122005 & 1.33175511 & -0.84589351 \\ -1.27632887 & -0.84589351 & 0.88735952 \end{bmatrix}$

Plot:



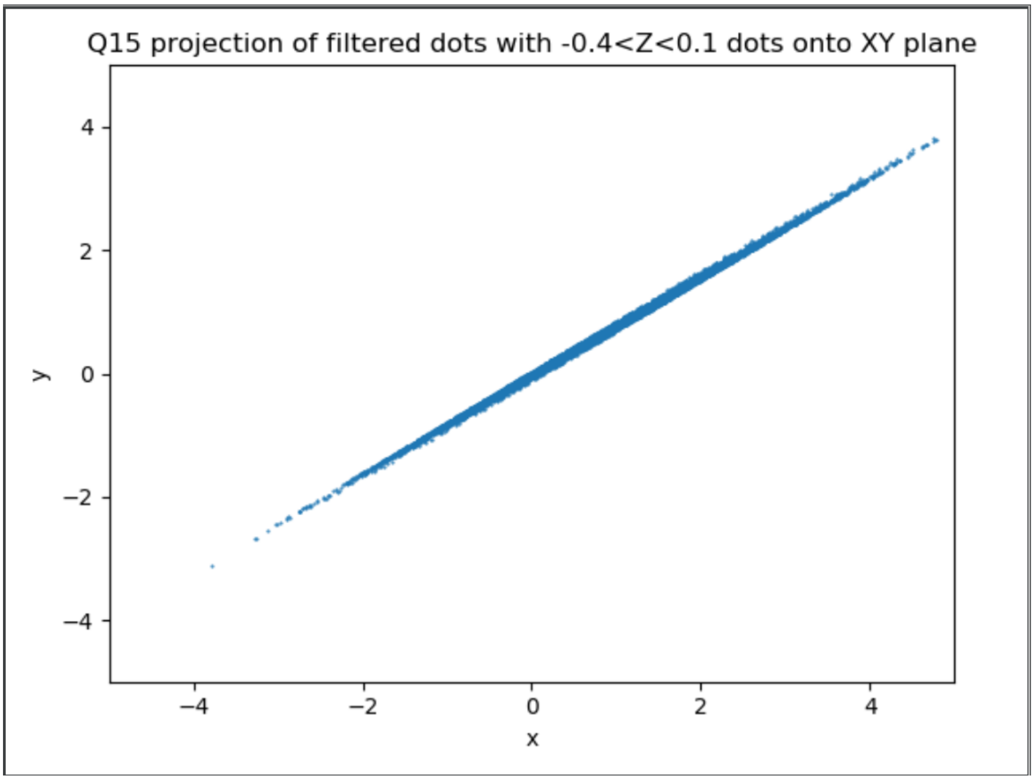
14.

Plot of the previous data set projected onto the XY plane:



15. Same plot as question 14 but only for Z values s.t $0.1 > z > -0.4$

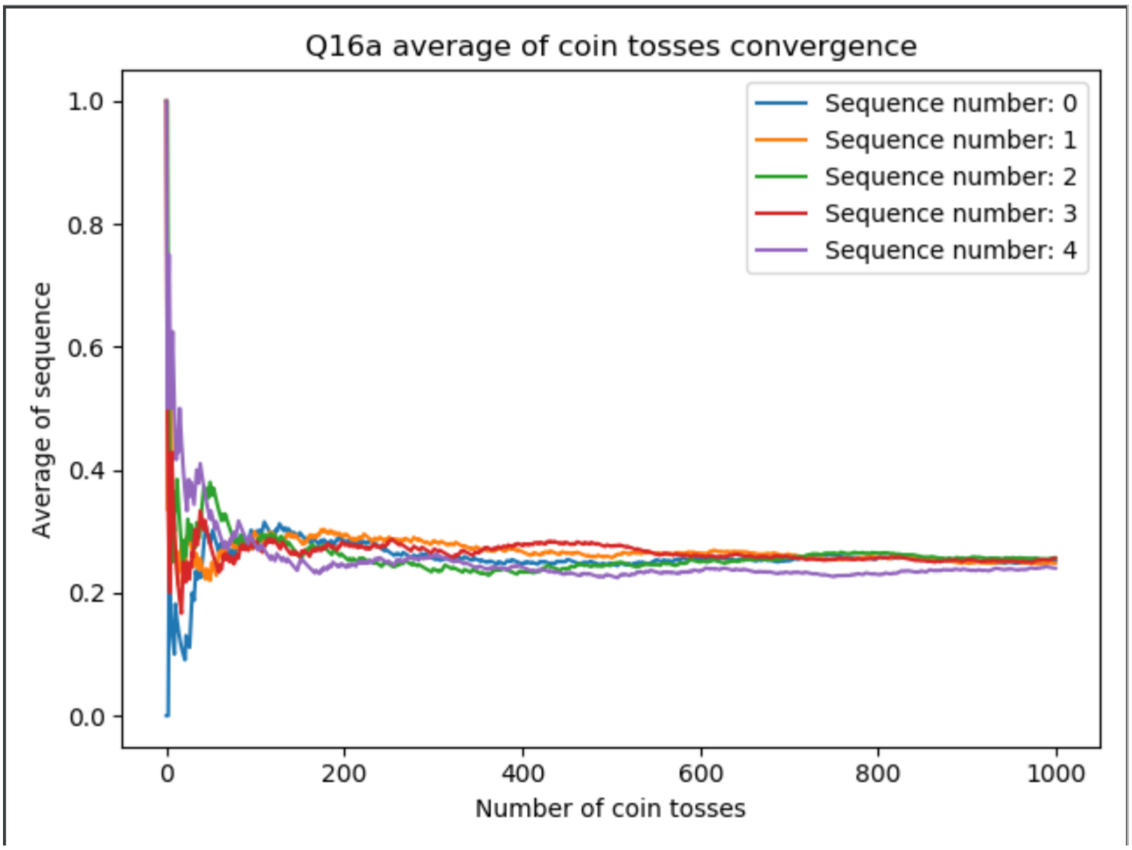
Plot:



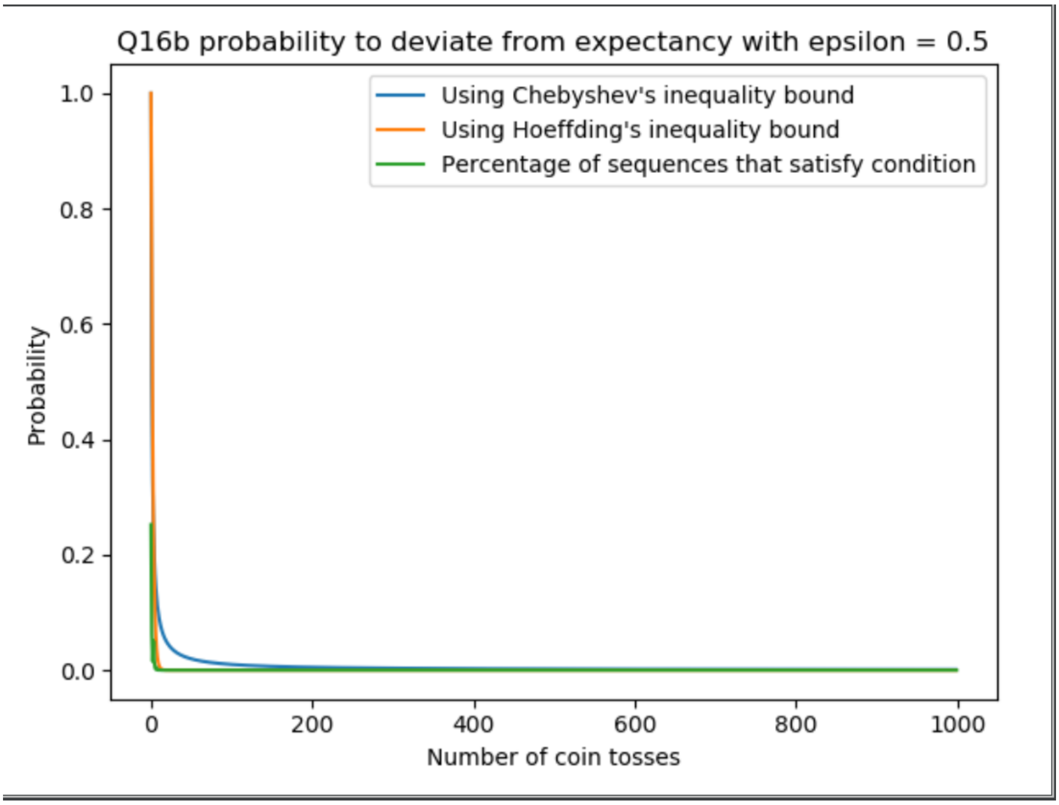
16. a.

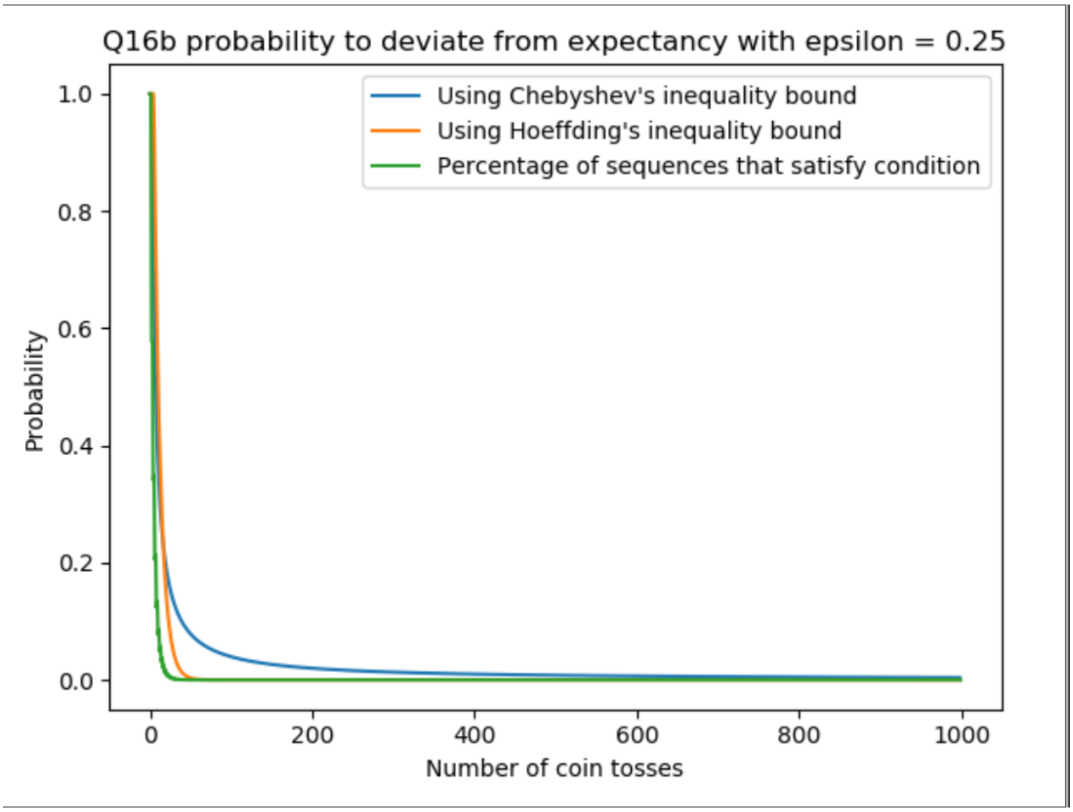
Plot of the first 5 sequences of 1000 tosses, s.t the plot shown the relationship between m(mean of tosses up to m) and the average of the sequences.

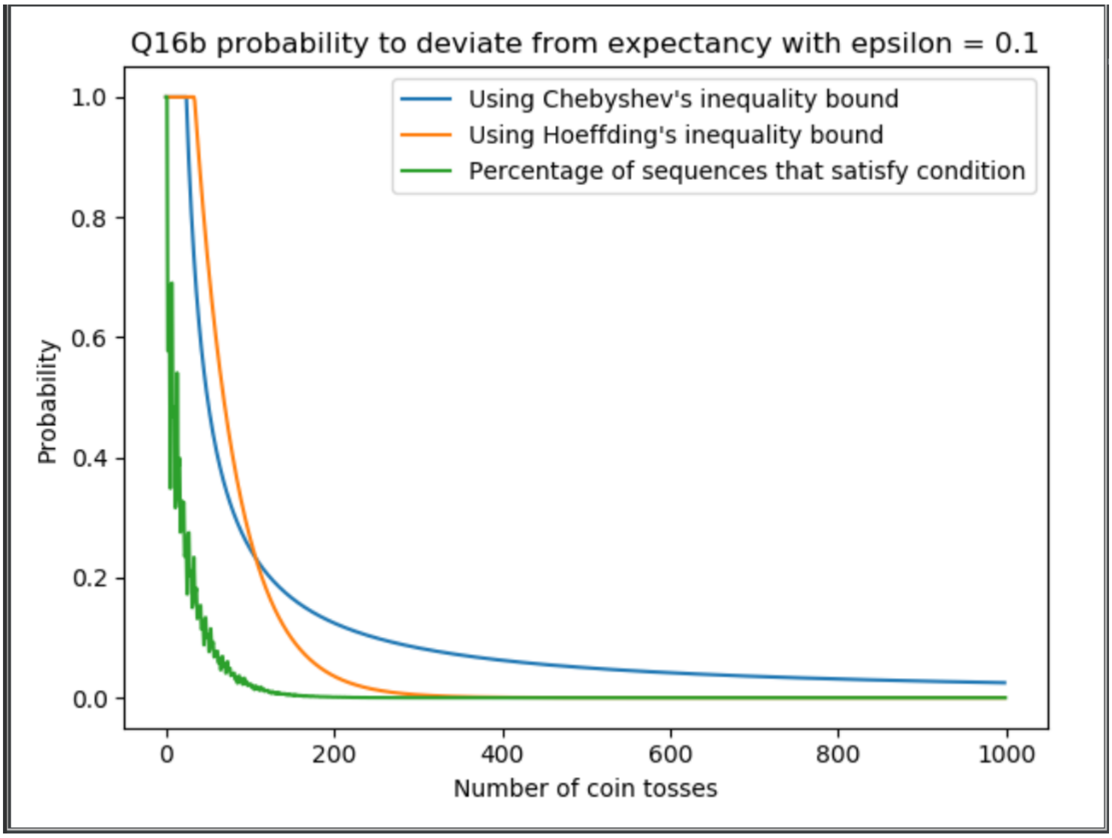
We would expect that as m grows the averages will converge, as the Weak Law of Large Numbers says, and indeed:

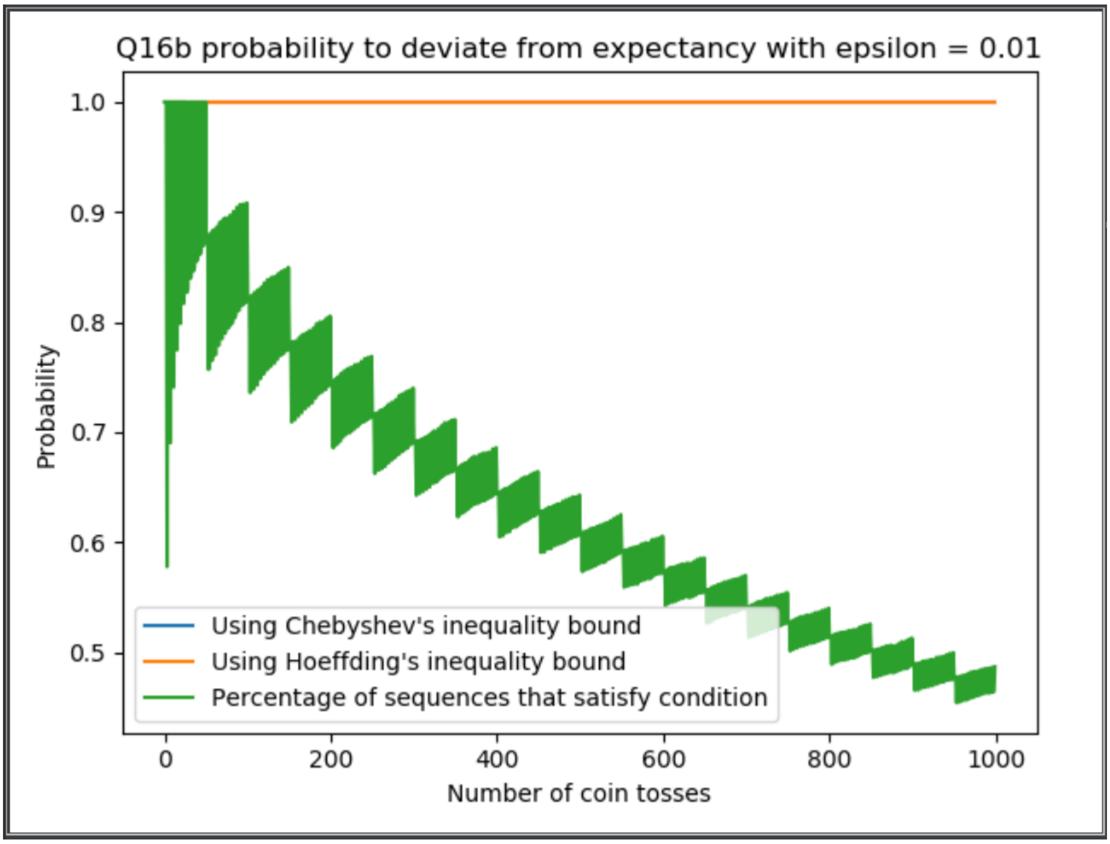


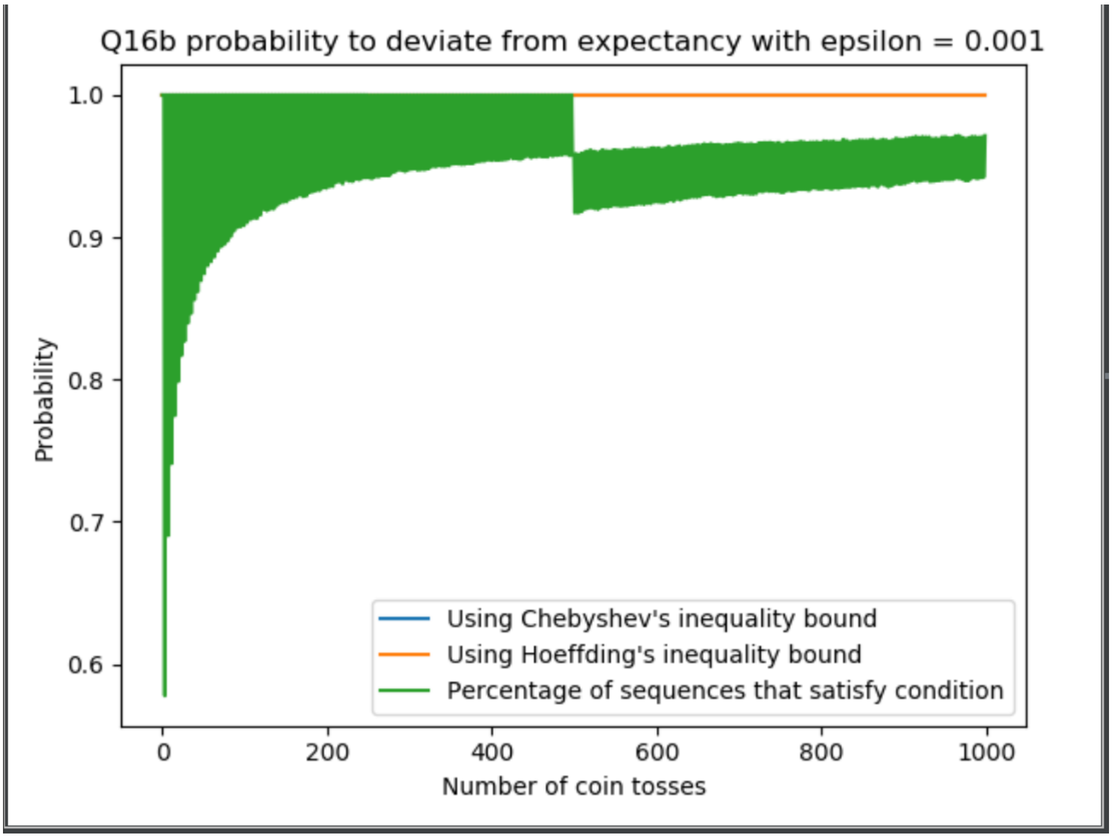
b. + c.: The plots I got in questions 16b and 16c:
I expect that as m grows the percentage of sequences that hold the claim will decrease.






Q16b probability to deviate from expectancy with epsilon = 0.1
Using Chebyshev's inequality bound
Using Hoeffding's inequality bound
Percentage of sequences that satisfy condition
Probability
1.0
0.8
0.6
0.4
0.2
0.0
0
200
400
600
800
1000
Number of coin tosses


Q16b probability to deviate from expectancy with epsilon = 0.01
Using Chebyshev's inequality bound
Using Hoeffding's inequality bound
Percentage of sequences that satisfy condition
Probability
1.0
0.9
0.8
0.7
0.6
0.5
0
200
400
600
800
1000
Number of coin tosses



(for clarity in the last 2 graphs the bound (orange& blue) the line is pretty much set only to 1)