IML - EX1

גיא קורנבליט, 308224948

(1) הראינו כי ההטלה של $\vec{v}$ על $\vec{w}$ מוגדרת ע"י $P_w(v) = \frac{\langle \vec{v},\vec{w}\rangle}{||\vec{w}||^2}\cdot\vec{w}$.

עבור $\vec{v}=(1,2,3,4)\,,\vec{w}=(0,-1,1,2)$, נחשב באמצעות המכפלה הסקלרית-

$$\|\vec{w}\| = 6,\ \langle \vec{v},\vec{w}\rangle = 0-2+3+8=9$$

ובסה"כ $P_w(v) = \frac{9}{36}\vec{w} = \frac{1}{4}\vec{w}$.

(2) באותו אופן, נחשב $\langle \vec{v},\vec{w}\rangle = 0$ ו-$\|\vec{w}\| = 3$ ובסה"כ $P_w(v)=\vec{0}$. כלומר $\vec{v}\perp\vec{u}$.

(3) **טענה:** יהיו $v,w\in\mathbb{R}^m \neq \vec{0}$, אזי $\langle \vec{v},\vec{w}\rangle = 0$ אם"ם הזווית שביניהם $\theta=\{\pm 90\}$.

הוכחה: הוכחנו טענה בתרגול לפיה $\langle \vec{v},\vec{w}\rangle = \|\vec{v}\|\,\|\vec{w}\|\,cos\theta$. מפני ש- $v,w\neq\vec{0}$ אזי בהכרח גם $\|\vec{v}\|\,,\|\vec{w}\|\neq 0$ (מתכונות הנורמה), ומכאן $\theta=\pm 90 \iff cos\theta = 0 \iff \langle \vec{v},\vec{w}\rangle = 0$.

■

(4) **טענה:** תהי $T\,:\,V\,\to\,W$ העתקה ליניארית, ו-A המ' המייצגת של T. אם A אורתוגונלית (עמודותיה בסיס אורתונורמלי) אזי T איזומטרית, כלומר לכל $x\in V$ $\|Ax\|_2 = \|x\|_2$.

הוכחה: נראה עבור המקרה הפרטי, עבור המכפלה הפנימית הסקלרית, ועבור המקרה הכללי.

(א) עבור המכפלה הסטנדרטית, מתקיים -

$$\begin{aligned}(||Ax||_2)^2 = \langle Ax,Ax\rangle = \underbrace{Ax\cdot Ax}_{dot\ product} = (Ax)^T Ax \\ = x^TA^TAx \underset{A^T=A^{-1}}{=} x^Tx = x\cdot x = \langle x,x\rangle = (\|x\|_2)^2\end{aligned}$$

$$\overset{sqrt\ mono.}{\Longrightarrow} \|Ax\|_2 = \|x\|_2$$

(ב) A מטריצה אורתוגונלית, בפרט ריבועית - לכן $dim(V)=dim(W)$ ומאיזומורפיות $V=W$ (כלומר A מייצגת אופרטור). כמו כן, וקטורי העמודה והשורות של A מהווים בסיסים או"נ ל-V ולכן A הינה למעשה מטריצת מעבר בסיס (תואם לאופרטור או"ג - סיבוב). יהי

$x \in V$, נסמן ב- $B = (b_1, ..., b_n)$ את הבסיס ש-A מעבירה את x. כלומר, מטענה שהוכחנו בליניארית, מתקיים

$$Ax = \langle b_1, x \rangle b_1 + ... + \langle b_n, x \rangle b_n$$

מכאן -

$$\begin{aligned}(\|Ax\|)^2 = \langle Ax, Ax \rangle &= \left\langle \sum_{i=1}^{n} \langle b_i, x \rangle b_i, \sum_{j=1}^{n} \langle b_j, x \rangle b_j \right\rangle \\ &\overset{linearity}{=} \sum_{i=1}^{n} \sum_{j=1}^{n} \langle b_i, x \rangle \cdot \langle b_j, x \rangle \cdot \langle b_i, b_j \rangle \\ &\overset{\delta_{ij}}{=} \sum_{i=1}^{n} \langle b_i, x \rangle \langle b_i, x \rangle = \left\langle \sum_{i=1}^{n} \langle b_i, x \rangle \cdot b_i | x \right\rangle = \langle x | x \rangle = (\|x\|)^2\end{aligned}$$

ממונוטוניות השורש נובע כי $\|Ax\| = \|x\|$. ■

(5) תהי A מטריצה הפיכה, $A = UDV^T$ הינו פירוק ה-SVD של A. נמצא נוסחה ל-A^{-1}.

תשובה: אם A הפיכה , אזי A בהכרח ריבועית, נסמן $A \in M_{n\times n}(\mathbb{R})$, ולכן נובע גם $U, V \in M_{n\times n}(\mathbb{R})$. בפרט, U, V מטריצות ריבועיות ואורתוגונליות, עמודתיהן בסיסים או"נ ולכן הגרעין שלהן טריוויאלי והן הפיכות. כמו כן, מטריצה או"ג צמודה לעצמה ולכן $U^T = U^{-1}$ ו-$V^T = V^{-1}$. מכאן, מתקיים

$$\begin{aligned}A^{-1} = \left(UDV^T\right)^{-1} &\underset{inverse\ operator}{=} \left(V^T\right)^{-1} D^{-1} U^{-1} \\ &= \left(V^{-1}\right)^T D^{-1} U^T = \left(V^T\right)^T D^{-1} U^T \\ &= V D^{-1} U^T\end{aligned}$$

נסיק כי על מנת לחשב את המטריצה ההפכית ל-A, מספיק לשחלף את U ו-V , ולחשב את המטריצה ההפכית של D באמצעות חלוקת הסקלרים על האלכסון הראשי.

(6) **נמצא פירוק** SVD **באמצעות** U, D, V **המקיימות** $C = UDV^T = \begin{pmatrix} 5 & 5 \\ -1 & 7 \end{pmatrix}$.

(א) בהתבסס על הקשר שראינו בתרגול בין פירוק SVS לפירוק EVD של C^TC, המטריצה הבאה לכסינה:

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

(ב) נסמן $A = C^TC = VDD^TV^T$. נחשב את פירוק ה-EVD של A (מציאת פולינום אפייני, חילוץ ע"ע, מציאת ו"ע והפיכתם לבסיס או"נ) ונקבל -

$$A = \begin{pmatrix} -\frac{3}{\sqrt{10}} & -\frac{1}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} & -\frac{3}{\sqrt{10}} \end{pmatrix} \begin{pmatrix} 20 & 0 \\ 0 & 80 \end{pmatrix} \begin{pmatrix} -\frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \\ -\frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \end{pmatrix}$$

$$\Longrightarrow V = \begin{pmatrix} -\frac{3}{\sqrt{10}} & -\frac{1}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} & -\frac{3}{\sqrt{10}} \end{pmatrix}, D = \begin{pmatrix} \sqrt{20} & 0 \\ 0 & \sqrt{80} \end{pmatrix}$$

(ג) כעת, נמצא את U. נוכל להכפיל מימין את פירוק ה-SVD של C ב-V, ומפני ש-V או"ג מתקיים $V^TV = I$ ולכן $CV = UD$. נכפיל מימין ב-D^{-1} ונקבל $CVD^{-1} = U$.

$$U = \begin{pmatrix} 5 & 5 \\ -1 & 7 \end{pmatrix} \begin{pmatrix} -\frac{3}{\sqrt{10}} & -\frac{1}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} & -\frac{3}{\sqrt{10}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{20}} & 0 \\ 0 & \frac{1}{\sqrt{80}} \end{pmatrix}$$
$$= \begin{pmatrix} -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}$$

(7) **אלגוריתם Power Iteration.**

תהי $A \in M_{m\times n}(\mathbb{R})$. נגדיר $C_0 = A^TA \in M_{n\times n}(\mathbb{R})$. יהיו $\lambda_1, .., \lambda_n$ ע"ע של C_0 כך ש-$\lambda_n \geq ... \geq \lambda_1$ (נניח $\lambda_1 > \lambda_2$), עם ו"ע מתאימים $v_1, ..., v_n$. נגדיר לכל $k \in [0, \infty)$ $b_{k+1} = \frac{C_0bk}{\|C_0bk\|}$, כך ש-$b_0$ מאותחל אקראית (כולם ב-$\mathbb{R}^n$). **טענה:** $\lim_{k\to\infty} b_k = v_1$.

הוכחה: יהי $B = (v_1, v_2, .., v_n)$ בסיס המורכב מקבוצת הו"ע של C_0, נניח כי B הינו בסיס אורתונורמלי של ו"ע (אחרת נוכל לנרמל אותו). נתבונן בוקטור המוגרל b_0 ונוכל להציגו כצירוף

ליניארי של איברי הבסיס, כלומר - $b_0 = \sum_{i=1}^{n} a_i v_i$, נניח כי $a_1 \neq 0$. נתבונן באיבר b_{k+1}:

$$b_{k+1} = \frac{C_0 b_k}{||C_0 b_k||} = \frac{C_0 \cdot \left(\frac{C_0 b_{k-1}}{||C_0 b_{k-1}||}\right)}{||C_0 \left(\frac{C_0 b_{k-1}}{||C_0 b_{k-1}||}\right)||} = \frac{\frac{C_0^2 b_{k-1}}{||C_0 b_{k-1}||}}{\left\|\frac{C_0^2 b_{k-1}}{||C_0 b_{k-1}||}\right\|} = \frac{\frac{C_0^2 b_{k-1}}{\cancel{||C_0 b_{k-1}||}}}{\frac{\left\|C_0^2 b_{k-1}\right\|}{\cancel{||C_0 b_{k-1}||}}}$$

$$= \frac{C_0^2 b_{k-1}}{||C_0^2 b_{k-1}||} = ... = \frac{C_0^{k+1} b_0}{\left\|C_0^{k+1} b_0\right\|}$$

כעת, מהלמה שהראינו בתרגול, C_0 לכסינה, כלומר קיימת מטריצה או"ג V (שעמודותיה וקטורי הבסיס B שהינם וקטורים עצמיים ל-C_0) ומטריצה אלכסונית D (על האלכסון מסודרים לפי גודל הע"ע של C_0), כך ש- $C_0 = VDV^T$. בפרט, עבור ו"ע $\vec{v}$ המתאים לע"ע λ מתקיים

$$C_0^k \vec{v} = C_0^{k-1} C_0 \vec{v} = C_0^{k-1} \lambda \vec{v} = ... = \lambda^k \vec{v}$$

מכאן -

$$C_0^k b_0 = C_0^k \left(a_1 v_1 + ... + a_n v_n\right) = \left(a_1 \lambda_1^k v_1 + ... + a_n \lambda_n^k v_n\right)$$

$$\underset{a_1 \neq 0}{=} a_1 \lambda_1^k \left(v_1 + \frac{a_2}{a_1} \left(\frac{\lambda_2}{\lambda_1}\right)^k v_2 + ... + \frac{a_n}{a_1} \left(\frac{\lambda_n}{\lambda_1}\right)^k v_n\right)$$

ומפני ש-$\lambda_n \geq ... \geq \lambda_2 > \lambda_1$, מתקיים גם $\left|\frac{\lambda_i}{\lambda_1}\right| < 1$ לכל $i \in [2, n]$, ולכן $\left(\frac{\lambda_i}{\lambda_1}\right)^k \overset{k\to\infty}{\longrightarrow} 0$.
בסה"כ נקבל

$$lim_{k\to\infty} b_k = lim_{k\to\infty} \frac{C_0^k b_0}{\left\|C_0^k b_0\right\|} = \frac{a_1 \lambda_1^k v_1}{\left\|a_1 \lambda_1^k v_1\right\|} = \vec{v_1}$$

נשים לב כי השוויון הוא עד כדי הסימן של $\vec{v}$. ■

(8) נתונים $x \in \mathbb{R}^n$ ו-$U \in M_{n\times n}(\mathbb{R})$ מטריצה אורתוגונלית. עבור $\sigma \in \mathbb{R}^n$, נתונה הפונקציה $f(\sigma) = U\text{diag}(\sigma) U^T x$ כאשר $\text{diag}(\sigma) \in \mathbb{R}^{n\times n}$ המקיימת

$$\text{diag}(\sigma)_{ij} = \begin{cases} \sigma_i & i = j \\ 0 & i \neq j \end{cases}$$

ראשית, נשים לב כי $f : \mathbb{R}^n \to \mathbb{R}^n$ מקבלת וקטור σ, שקובע ביחידות פירוק EVD של מטריצה כלשהי A. כלומר, בהינתן σ מתקיים $f(\sigma) = A_\sigma x$. יהי $B = (b_1, b_2.., b_n)$ בסיס או"נ של ו"ע של A_σ (וקטורי העמודה של U), אזי נוכל להציג את x כהטלה או"ג על המרחב בבסיס B, כלומר $x = \sum_{i=1}^{n} \langle x, b_i \rangle b_i$. מכאן,

נובע כי

$$f(\sigma) = A_\sigma x = \sum_{i=1}^{n} \langle x, b_i \rangle A_\sigma b_i = \sum_{i=1}^{n} \langle x, b_i \rangle \sigma_i \cdot b_i$$

$$= \langle x, b_1 \rangle \sigma_1 \cdot \begin{bmatrix} b_1^1 \\ \vdots \\ b_1^n \end{bmatrix} + \cdots + \langle x, b_n \rangle \sigma_n \cdot \begin{bmatrix} b_n^1 \\ \vdots \\ b_n^n \end{bmatrix}$$

$$= \begin{bmatrix} \langle x, b_1 \rangle \sigma_1 \cdot b_1^1 + \cdots + & \langle x, b_n \rangle \sigma_n b_n^1 \\ \vdots & \vdots \\ \langle x, b_1 \rangle \sigma_1 \cdot b_1^n + \cdots + & \langle x, b_n \rangle \sigma_n b_n^n \end{bmatrix}$$

נגדיר לכל $i \in [n]$ פונקציה $f_i : \mathbb{R}^n \to \mathbb{R}$ כך ש- $f_i(\sigma) = \sum_{k=1}^{n} \langle x, b_k \rangle \sigma_k \cdot b_k^i$. אזי $f(\sigma) = \begin{bmatrix} f_1(\sigma) \\ f_2(\sigma) \\ \vdots \\ f_n(\sigma) \end{bmatrix}$.

כעת, לכל $i, j \in [n]$ נתבונן בנגזרת החלקית

$$\frac{\partial f_i(\sigma)}{\partial \sigma_j} = \langle x, b_j \rangle b_j^i$$

ולכן מההגדרה מתקיים -

$$J_\sigma(f)=\begin{bmatrix}\frac{\partial f_1(\sigma)}{\partial\sigma_1} & \dots & \frac{\partial f_1(\sigma)}{\partial\sigma_n}\\ \vdots & & \vdots\\ \frac{\partial f_n(\sigma)}{\partial\sigma_1} & \dots & \frac{\partial f_n(\sigma)}{\partial\sigma_n}\end{bmatrix}=\begin{bmatrix}\langle x,b_1\rangle b_1^1 & \dots & \langle x,b_n\rangle b_n^1\\ \vdots & & \vdots\\ \langle x,b_1\rangle b_1^n & \dots & \langle x,b_n\rangle b_n^n\end{bmatrix}$$

$$=\begin{bmatrix}| & & |\\ \langle x,b_1\rangle b_1 & \dots & \langle x,b_n\rangle b_n\\ | & & |\end{bmatrix}$$

(9) בהינתן $y\in\mathbb{R}^n$, תהי $h:\mathbb{R}^n\to\mathbb{R}$ המוגדרת על ידי $h(\sigma)=\frac{1}{2}\|f(\sigma)-y\|^2$, נגדיר: $T: \mathbb{R}^n\to\mathbb{R}$ על ידי $T(\vec{v})=\frac{1}{2}\|\vec{v}\|^2$, ואת $g:\mathbb{R}^n\to\mathbb{R}^n$ באמצעות $g(\vec{v})=\vec{v}-\vec{y}$. מכאן, מתקיים $h=T\circ g\circ f$ ונובע מכלל השרשרת $J_\sigma(h)=J_{g\circ f(\sigma)}(T)\cdot J_{g(\sigma)}(g)\cdot J_\sigma(f)$. יהי $v\in\mathbb{R}^n$. $T(v)=\frac{1}{2}\sum_{i=1}^n v_i^2$ ולכן לכל $i\in[n]$, $\frac{\partial T(v)}{\partial v_i}=v_i$ כלומר $J_v(T)=v^T\in\mathbb{R}^{1\times n}$. לכל $i\in[n]$ נגדיר $g_i(v)=v_i-y_i$ ולכן $\frac{\partial g_i(v)}{\partial v_i}=\delta_{ij}$ ומכאן $J_v(g)=I_n$. בסה"כ נקבל -

$$\begin{aligned}J_\sigma(h)&=J_{g\circ f(\sigma)}(T)\cdot J_{g(\sigma)}(g)\cdot J_\sigma(f)\\ &=\begin{bmatrix}(A_\sigma x)_1-y_1 & \dots & (A_\sigma x)_1-y_1\end{bmatrix}I_nJ_\sigma(f)\\ &=\begin{bmatrix}\left(\sum_{j=1}^n\langle x,b_j\rangle\sigma_j\cdot b_j^1\right)-y_1\\ \vdots\\ \left(\sum_{j=1}^n\langle x,b_j\rangle\sigma_j\cdot b_j^1\right)-y_1\end{bmatrix}^T\cdot\begin{bmatrix}| & & |\\ \langle x,b_1\rangle b_1 & \dots & \langle x,b_n\rangle b_n\\ | & & |\end{bmatrix}\end{aligned}$$

נסמן $u^T=\begin{bmatrix}\left(\sum_{j=1}^n\langle x,b_j\rangle\sigma_j\cdot b_j^1\right)-y_1\\ \\ \left(\sum_{j=1}^n\langle x,b_j\rangle\sigma_j\cdot b_j^1\right)-y_1\end{bmatrix}^T:=\begin{bmatrix}u_1\\ \vdots\\ u_n\end{bmatrix}^T$ ולכן $\nabla h(\sigma)=J_\sigma(f)^Tu$.

(10) נחשב את היעקוביאן של פונקציית ה- Softmax. $g(\boldsymbol{z})_j=\frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}}$. לכל $i\in[K]$ נגדיר $f_i(z)=e^{z_i}$, ו-$h(\boldsymbol{z})=\sum_{k=1}^K e^{z_k}$. מתקיים $\frac{\partial g(z)_i}{\partial z_j}=\frac{\partial}{\partial z_j}\frac{f_i(z)}{h}$, כאשר $\frac{\partial f_i(\boldsymbol{z})}{\partial z_j}=\begin{cases}e^{z_j} & i=j\\ 0 & i\neq j\end{cases}$,

ו- $\frac{\partial h(\boldsymbol{z})}{\partial z_j} = e^{z_j}$. הראינו בכיתה כי עבור $i = j$,

$$\frac{\partial g\left(\boldsymbol{z}\right)_i}{\partial z_j} = g(\boldsymbol{z})_i \left(1 - g(\boldsymbol{z})_i\right)$$

ועבור $i \neq j$ מתקיים -

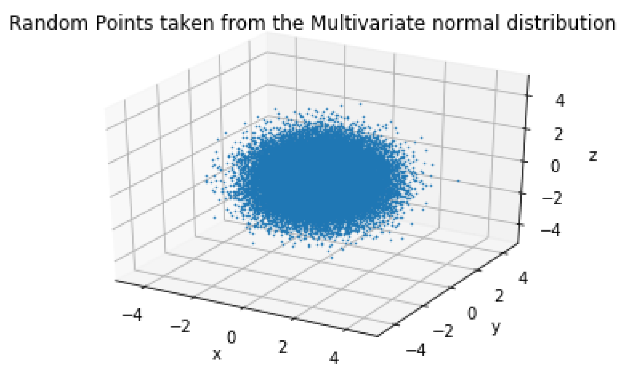
$$\frac{\partial}{\partial z_j}\frac{f_i(z)}{h} = \frac{0 \cdot \sum_{k=1}^{K} e^{z_k} - e^{z_j} \cdot e^{z_i}}{\left(\sum_{k=1}^{K} e^{z_k}\right)^2} = -\frac{e^{z_i}}{\sum_{k=1}^{K} e^{z_k}} \cdot \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}}$$

$$= -g\left(\boldsymbol{z}\right)_i g\left(\boldsymbol{z}\right)_j$$

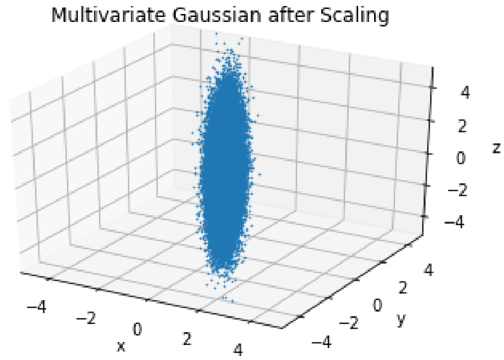
לפיכך -

$$J_{\boldsymbol{z}}\left(g\right) = \begin{bmatrix} g(\boldsymbol{z})_1 \left(1 - g(\boldsymbol{z})_1\right) & \dots & -g\left(\boldsymbol{z}\right)_1 g\left(\boldsymbol{z}\right)_n \\ \vdots & \ddots & \vdots \\ -g\left(\boldsymbol{z}\right)_n g\left(\boldsymbol{z}\right)_1 & \dots & g(\boldsymbol{z})_n \left(1 - g(\boldsymbol{z})_n\right) \end{bmatrix}$$

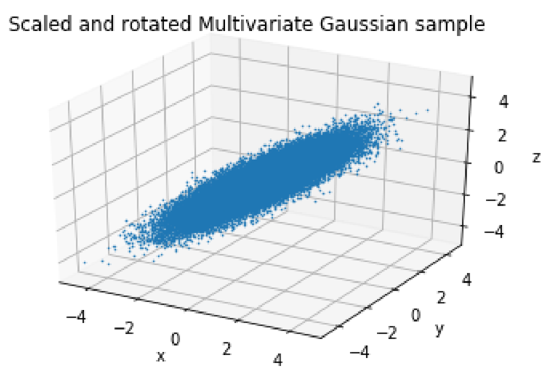
(11) ההתפלגות הרב-נורמלית:



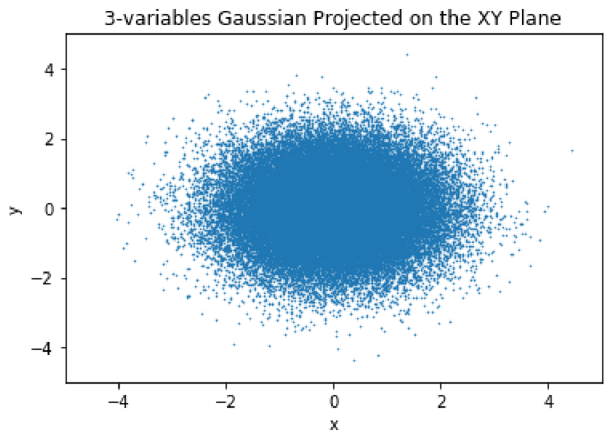
(12) מטריצת השונות המשותפת לאחר פעולת $scaling$ הינה בקירוב $\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 4 \end{bmatrix}$. כלומר, המשתנים הנורמליים שהוגרלו באופן בלתי מתואם (עבור מטריצת היחידה כמטריצת השונות המשותפת) נותרו בלתי מתואמים. בנוסף, נשים לב כי עבור המשתנה הראשון והשני השונות התאפסה, ולכן הטרנספורמציה הפכה אותם למשתנים קבועים (בקירוב).



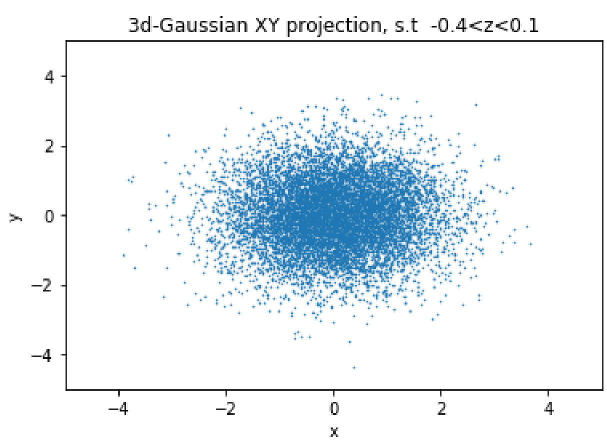
(13) מטריצת השונות המשותפת לאחר העתקה אורתוגונלית הינה $\begin{bmatrix} 0 & 1 & 0 \\ 1 & 4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$. כלומר, סיבוב או שיקוף הנתונים יצר קורלציה בין המשתנים.



(14) מההטלה של הנתונים על המישור, ניכר בבירור כי צורת ההתפלגות נשמרת. במילים אחרות, כפי שהראינו בכיתה, ההתפלגות השולית (לפי x, y) של התפלגות נורמלית הינה נורמלית, כאשר ההטלה מייצגת את ההתפלגות ביחס לשני משתנים אלו.

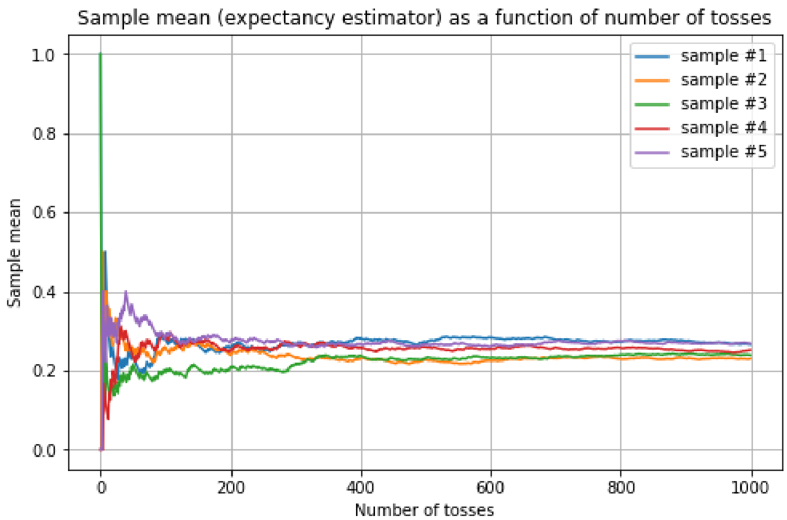


(15) בדוגמה זו, יצרנו התניה על המשתנה z, וכאשר מתבוננים בהטלה על מישור xy, כלומר על ההתפלגות השולית המותנית - ניכר כי גם במקרה זה ההתפלגות הינה נורמלית.

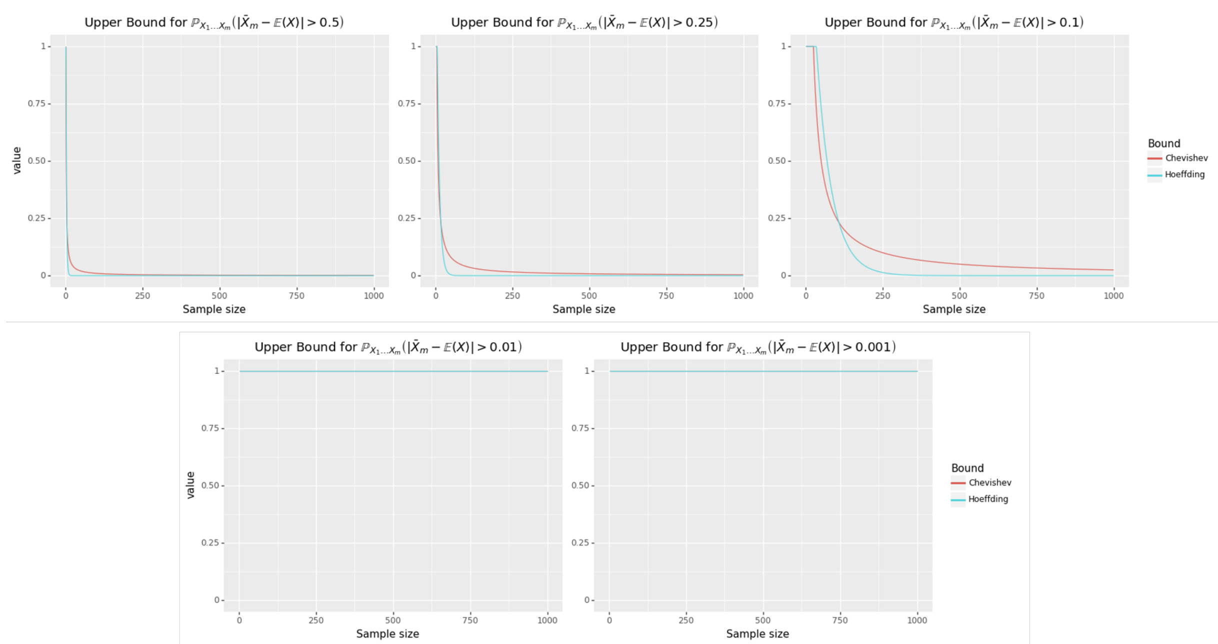


(16) בשאלה זו נדגמו $100k$ וקטורים מקריים של 1000 משתני ברנולי המייצגים הטלות מטבע. נגדיר לכל $m \in [1, 1000]$ אומד $\bar{X_m}$ לפרמטר של המשתנה.

(א) עבור 5 דגימות מהנתונים, ניכר כי ככל שגודל הדגימה עולה האומדן מתקרב לפרמטר של ההתפלגות (כפי שהוגדר - $.25$).



(ב) בהמשך לסעיף הקודם, הגרפים הנ"ל ממחישים כי ככל שגודל הדגימה עולה, כך האומדן מתקרב לפרמטר האמיתי. ההסתברות שהמרחק של האומד מהפרמטר יהיה גדול מ-ε הולך וקטן ככל שגודל הדגימה גדל. מנגד, עבור ε קטן מספיק, החסמים אינם מושפעים מגודל הדגימה, מפני שהגודל הנמדד אינו מספיק כדי לשפר את הקירוב מספיק (בשאיפה לאינסוף, החסם ירד).



(ג) כאשר בוחנים את אחוז הדגימות עבורן המרחק של האומדן מהפרמטר גדול מ-ε, ניכר כי המגמה היא מונוטונית יורדת. סביר להניח כי הקפיצות בהתפלגות אחוז הדגימות שנכשלו,

נובעות מהאקראיות במידע שהולכת ומצטמצמת (ובהתאם קו המגמה מתמתן) ככל שהאומד מתבסס על גודל דגימה גדול יותר. בנוסף, נשים לב כי לכל ε ההתפלגות נמצאת מתחת לחסמים שמצאנו.