GMDL212, HW1

Yuval Margalit

Part I
Random Vectors

Problem 1:

יהיו $A_1, ..., A_n$ סדרת מאורעות זרים בזוגות נראה כי $P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$
נשלים את הסדרה לסדרה אינסופית $A_1, ..., A_n, A_{n+1}, ...$ כך ש $A_{n+1} = A_{n+2} = ... = \emptyset$.
אז $P(\bigcup_{i=1}^{n} A_i) = P(\bigcup_{i=1}^{\infty} A_i) \overset{P(\emptyset)=0}{=} \sum_{i=1}^{\infty} P(A_i) \overset{\text{property 3}}{=} \sum_{i=1}^{n} P(A_i)$

Problem 2:

יהי A מאורע נראה כי $P(A^c) = 1 - P(A)$
$\Omega = A \cup A^c \Rightarrow 1 = P(\Omega) = P(A \cup A^c) \overset{fact1}{=} P(A) + P(A^c) \Rightarrow P(A^c) = 1 - P(A)$

Problem 3:

יהי X וקטור רנדומי אשר הטווח שלו הוא $\mathbb{R}$

1. נניח $B = \mathbb{R}$ **אז** $X^{-1}(B) = \{\omega \in \Omega | X(\omega) \in B\} = \{\omega \in \Omega | X(\omega) \in \mathbb{R}\} \overset{F(x)\subset\mathbb{R}}{=} \Omega$
ולכן $P(X^{-1}(B)) = P(\Omega) = 1$

2. נניח $B = \emptyset$ **אז** $X^{-1}(B) = \{\omega \in \Omega | X(\omega) \in B\} = \{\omega \in \Omega | X(\omega) \in \emptyset\} = \emptyset$
ולכן $P(X^{-1}(B)) = P(\emptyset) = 0$

Problem 4:

אראה כי F אינה רציפה מימין ב42 ולכן לא יכול להיות CDF
$\lim_{x \to 42^+} F(x) = 1$ אבל $F(42) = 0$ ולכן F אינה רציפה מימין

Problem 5:

יהי X וקטור רדומי רציף n מימדי

1. $P(X=x)=\lim_{x'\to x^-}P(x'<X\leq x)=F(x)-\lim_{x'\to x}F(x')$
F רציפה ולכן $F(x)-\lim_{x'\to x^-}F(x')=F(x)-F(x)=0$

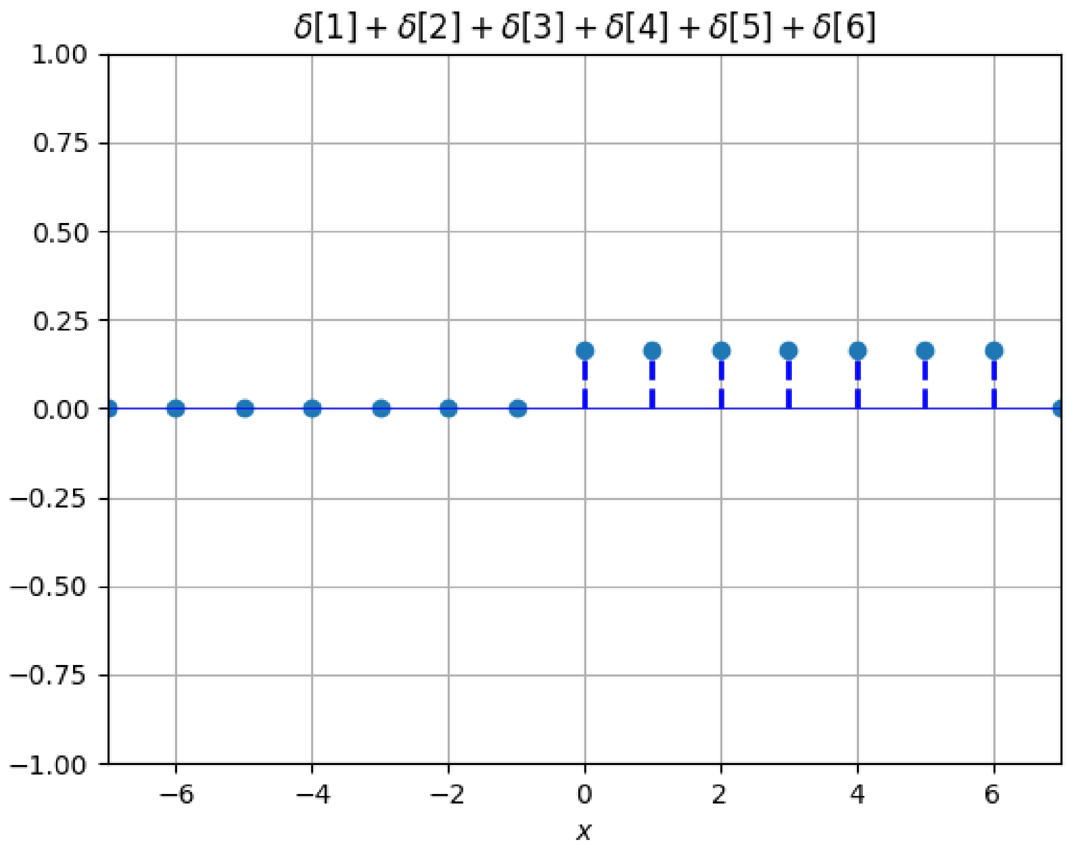
2. בעבור $B=\{\underbrace{(1,..,1)^T}_{n-times}\}$ יתקיים כי $P(X^{-1}(B))=F(X=(1,..,1)^T)\stackrel{1}{=}0$

3. נבחר את הקבוצות הבאות $\forall i\geq i, B_i=\{\underbrace{(i,..,i)^T}_{n-times}\}$

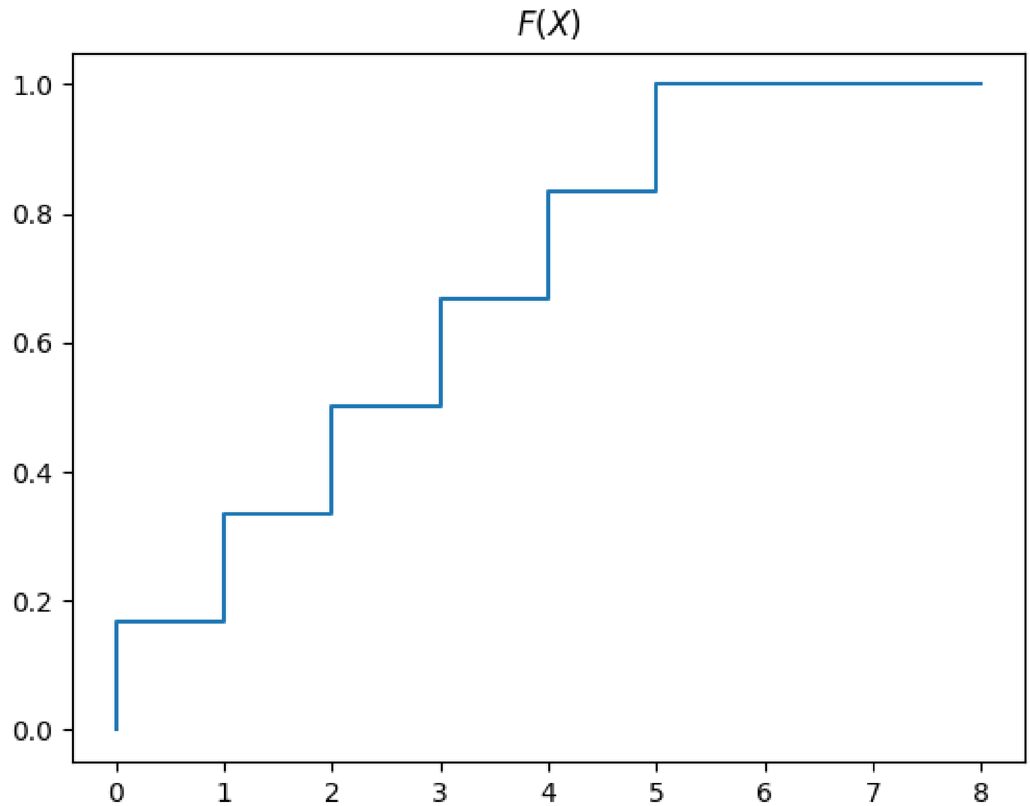
נשים לי כי אכן $\forall i,j\geq 1, i\neq j\to B_i\cap B_j=\emptyset$, נתבונן ב $X^{-1}(\bigcup_{i=1}^{\infty}B_i)$, נשים לב שכיוון ו$X$ הוא פונקציה מתקיים $\forall i,j\geq 1, i\neq j\to$ $X^{-1}(B_i)\cap X^{-1}(B_j)=\emptyset$

ולכן $X^{-1}(\bigcup_{i=1}^{\infty}B_i)$מהווה איחוד של מאורעות זרים. אז מתכונה 3 של פונקציות הסתברות $P(X^{-1}(\bigcup_{i=1}^{\infty}B_i))=\sum_{i=0}^{\infty}P(X^{-1}(\bigcup_{i=1}^{\infty}B_i))=$ $\sum_{i=0}^{\infty}0=0$

problem 6:



$\delta[1] + \delta[2] + \delta[3] + \delta[4] + \delta[5] + \delta[6]$
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−6
−4
−2
0
2
4
6
x



problem 7:

$p(x_3, x_4, x_5) = \sum_{x_1, x_2} p(x)$

problam 8:

$$E(X) = \sum_{i=1}^{6} P(X = i)i = \frac{1}{6} \sum_{i=1}^{6} i = \frac{21}{6} = 3.5$$

problem 9:

$$(xx^T)^T = (x^T)^T x^T = xx^T$$

problem 10:

נניח $E(X) = \mu$ ונגדיר $g(x) = x^T$

$E(X^T) = E(g(X)) = \sum_x g(x)p(x) = \sum_x x^T p(x) = (\sum_x xp(x))^T = \mu^T$

באופן דומה הדבר נכון גם למקרה הרציף

problem 11:

נגדיר $g(x) = xx^T$ אז $E(XX^T) = E(g(X))$ ולכן מ definition 23/2

$E(XX^T)_{i,j} = E(g(X))_{i,j} = E((g(X))_{ij}) = E(X_iX_j)$

problem 12:

$$(R_x)^T = E(XX^T)^T \overset{problem10}{=} E((XX^T)^T) \overset{problem9}{=} E(XX^T) = R_x$$

problem 13:

נגדיר $g(x) = (x - E(x))(x - E(x))^T$ אז $\Sigma_X = E(g(X))$ ולכן מ definition 23/2

$(\Sigma_X)_{i,j} = E(g_{i,j}(X)) = E((X - E(X))_i((X - E(X))^T)_j) = E((X_i - \mu_i)(X_j - \mu_j)) = cov(X_i, X_j)$

problem 14:

$$(\Sigma_X)_{i,i} \overset{problem13}{=} E((X_i - \mu_i)(X_i - \mu_i)) = E((X_i - \mu_i)^2) = V(X_i)$$

problem 15:

$$(\Sigma_X)^T = (E((X - \mu)(X - \mu)^T))^T \overset{problem9}{=} E(((X - \mu)^T)^T(X - \mu)^T) = E((X - \mu)(X - \mu)^T)$$

problem 16:

1.
$$\begin{aligned}\Sigma_X = E((X - \mu)(X - \mu)^T) = E(XX^T - X\mu^T - \mu X^T + \mu\mu^T) &\overset{mean-is-linear}{=} \\ =E(XX^T) - E(X\mu^T) - E(\mu X^T) + E(\mu\mu^T) &\overset{problem-10}{=} \\ =E(XX^T) - \mu\mu^T - \mu\mu^T + \mu\mu^T = E(XX^T) - \mu\mu^T\end{aligned}$$

2.
$$E((X - \mu)X^T) = E(XX^T - \mu X^T) \overset{mean-is-linear}{=} E(XX^T) - E(\mu X^T) \overset{problem-10}{=} E(XX^T) - \mu\mu^T = \Sigma_X$$

3.
$$E(X(X - \mu)^T) = E(XX^T - X\mu^T) \overset{mean-is-linear}{=} E(XX^T) - E(X\mu^T) \underset{E(c)=c}{\overset{problem-10}{=}} E(XX^T) - \mu\mu^T = \Sigma_X$$

problem 17:

1.

$$R_{XY} = E(XY^T), R_{YX} = E(YX^T), R_{XY}^T = (E(XY^T))^T \overset{problem-10}{=} E((Y^T)^T X^T) = E(YX^T) = R_{YX}$$

2.

$$\begin{aligned}\Sigma_{XY}^T = &(E((X-E(X))(Y-E(Y))^T))^T \overset{problem-10}{=} E(((X-E(X))(Y-E(Y))^T)^T) \\ =&E(((Y-E(Y))^T)^T(X-E(X))^T) = E((Y-E(Y))(X-E(X))^T) = \Sigma_{YX}\end{aligned}$$

3.

$$\begin{aligned}\Sigma_{XY}^T = &(E((X-E(X))(Y-E(Y))^T))^T = E(XY^T - X((E(Y))^T - E(X)Y^T + E(X)(E(Y))^T) \overset{mean-is-linear}{=} \\ =&E(XY^T) - E(X(E(Y))^T) - E(E(X)Y^T) + E(E(X)(E(Y))^T) \underset{E(c)=c}{\overset{E(cX)=cE(X)}{=}} \\ =&E(XY^T) - E(X)(E(Y))^T - E(X)E(Y^T) + E(X)(E(Y))^T \overset{problem-10}{=} E(XY^T) - \mu_X\mu_Y^T - \mu_X\mu_Y^T + \mu_X\mu_Y^T \overset{definition-27}{=} \\ =&R_{XY} - \mu_X\mu_Y^T\end{aligned}$$

problem 18:

$$\mu_Y = E(Y) = E(AX+b) \overset{mean-is-linear}{=} E(AX) + E(B) = AE(X) + b = A\mu_X + b$$

$$\begin{aligned}\Sigma_Y = &E((Y-E(Y))(Y-E(Y))^T) = E((AX+b-A\mu_X-b)((AX+b-A\mu_X-b)^T) = \\ =&E((AX-A\mu_X)((AX-A\mu_X)^T) = E(A(X-\mu_X)(X-\mu_X)^T A^T) = \\ =&AE((X-\mu_X)(X-\mu_X)^T)A^T = A\Sigma_X A^T\end{aligned}$$

problem 19:

יהי X וקטור רנדומי ממימד N עם תוחלת μ וקובריאנס $\Sigma_X = \sigma^2 I$ כאשר I מטריצת היחידה מסדר n ו $\sigma > 0$
נסמן $Y = 1^T X = \sum_{i=1}^{n} X_i$ אז מהשאלה הקודמת $\Sigma_Y = \mathbf{1}^{\mathrm{T}}\Sigma_X\mathbf{1} = (\sigma^2\mathbf{1})^T\mathbf{1} = \sum_{i=1}^{n}\sigma^2 = n\sigma^2$

problem 20:

יהיו A, B, C מאורעות.
$\Leftarrow$: נניח $A \perp\!\!\!\perp B|C$ אז
$P(A|B,C) = \frac{P(A\cap B\cap C)}{P(B\cap C)} = \frac{P(A,B|C)P(C)}{P(B|C)P(C)} = \frac{P(A|C)P(B|C)P(C)}{P(B|C)P(C)} = P(A|C)$
$\Rightarrow$: נניח $P(A|B,C) = P(A|C)$ אז $\frac{P(A\cap C)}{P(C)} = \frac{P(A\cap B\cap C)}{P(B\cap C)}$
$P(A,B|C) = \frac{P(A\cap B\cap C)}{P(C)} = \frac{P(A\cap B\cap C)P(B\cap C)}{P(B\cap C)P(C)} = \frac{P(A\cap C)P(B\cap C)}{P(C)P(C)} = P(A|C)P(B|C)$

problem 21:

$E(\mathbb{1}_A) = \sum_{i=0}^{1} P(\mathbb{1}_A = i) * i = P(\mathbb{1}_A = i) = P(A)$

problem 22:

יהיו $X \in R^n, Y \in R^m$ וקטורים מקריים אורטוגונליים כלומר $E(XY^T) = 0_{m\times n}$

$$R_Z = E(ZZ^T) \overset{problem-11}{=} \begin{bmatrix} E(XX^T) & E(XY^T) \\ E(YX^T) & E(YY^T) \end{bmatrix} = \begin{bmatrix} R_X & 0 \\ 0 & R_Y \end{bmatrix}$$

יהיו $X \in R^n, Y \in R^m$ וקטורים מקריים חסרי קורלציה כלומר $E(XY^T) = \mu_X \mu_Y^T$
אז

$$\Sigma_z \overset{fact\,9}{=} \begin{bmatrix} \Sigma_X & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_y \end{bmatrix} \overset{Definition\,27}{=} \begin{bmatrix} \Sigma_X & E(XY^T) - \mu_X \mu_Y^T \\ E(YX^T) - \mu_Y \mu_X^T & \Sigma_y \end{bmatrix} = \begin{bmatrix} \Sigma_X & 0 \\ 0 & \Sigma_y \end{bmatrix}$$

problem 23:

יהיו X, Y וקטורים רנדומיים ונניח $X \perp\!\!\!\perp Y$ מהגדרה $p(x,y) = p(x)p(y)$

$$E(XY^T) = \sum_{x,y} xy^T p(x,y) = \sum_x \sum_y xy^T p(x)p(y) \overset{sum-is-linear}{=} \sum_x xp(x)(\sum_y y^T p(y)) \overset{problem-10}{=} \sum_x xp(x)E(Y^T) = E(X)(E(Y))^T$$

ההוכחה זהה למקרה הרציף כיוון ואינטגרל הינו לינארי

problem 24:

1. $$p(X_1) = \sum_{x_2} p(x_1, x_2) = p(x_1, 0) + p(x_1, 1) = \begin{cases} 0.6 & x_1 = 0 \\ 0.4 & x_1 = 1 \end{cases}$$
   $$p(X_1) = \sum_{x_1} p(x_1, x_2) = p(0, x_2) + p(1, x_2) = \begin{cases} 0.8 & x_2 = 0 \\ 0.2 & x_2 = 1 \end{cases}$$
2. $$E(X) \overset{Definition20}{=} \begin{bmatrix} E(X_1) \\ E(X_2) \end{bmatrix} = \begin{bmatrix} 0.4 \\ 0.2 \end{bmatrix}$$
3. $$R_X = E(XX^T) \overset{probelm-11}{=} \begin{bmatrix} E(X_1^2) & E(X_1X_2) \\ E(X_2X_1) & E(X_2^2) \end{bmatrix} = \begin{bmatrix} 0.4 & 0.08 \\ 0.08 & 0.2 \end{bmatrix}$$
4. $$\Sigma_X \overset{fact-8}{=} E(XX^T) - \mu_X \mu_X^T = \begin{bmatrix} 0.4 & 0.08 \\ 0.08 & 0.2 \end{bmatrix} - \begin{bmatrix} 0.4 \\ 0.2 \end{bmatrix} \begin{bmatrix} 0.4 & 0.2 \end{bmatrix} = \begin{bmatrix} 0.4 & 0.08 \\ 0.08 & 0.2 \end{bmatrix} - \begin{bmatrix} 0.16 & 0.08 \\ 0.08 & 0.04 \end{bmatrix} = \begin{bmatrix} 0.24 & 0 \\ 0 & 0.16 \end{bmatrix}$$
5. X_1, X_2 אינם בלתי תלויים כיוון ש $p(X_1 = 1, X_2 = 1) = 0.1 \neq 0.08 = 0.4 * 0.2 = p(X_1 = 1)p(X_2 = 1)$,
   כלומר פונקציית ההתפלגות המשותפת אינה מכפלה של ההתפלגויות השוליות
6. X_1, X_2 חסרי קורלציה כיוון ו $E(X_1X_2) = 0.08 = 0.4 * 0.2 = E(X_1)E(X_2)$

probelm 25:

1. $$p(X_1)=\sum_{x_2}p(x_1,x_2)=p(x_1,0)+p(x_1,1)=\begin{cases}0.5 & x_1=0\\ 0.5 & x_1=1\end{cases}$$
$$p(X_1)=\sum_{x_1}p(x_1,x_2)=p(0,x_2)+p(1,x_2)=\begin{cases}0.5 & x_2=0\\ 0.5 & x_2=1\end{cases}$$

2. $$E(X)\overset{Definition20}{=}\begin{bmatrix}E(X_1)\\ E(X_2)\end{bmatrix}=\begin{bmatrix}0.5\\ 0.5\end{bmatrix}$$

3. $$E(XX^T)\overset{probelm-11}{=}\begin{bmatrix}E(X_1^2) & E(X_1X_2)\\ E(X_2X_1) & E(X_2^2)\end{bmatrix}=\begin{bmatrix}0.5 & 0.25\\ 0.25 & 0.5\end{bmatrix}$$
$$\Sigma_X\overset{fact-8}{=}E(XX^T)-\mu_X\mu_X^T=\begin{bmatrix}0.5 & 0.25\\ 0.25 & 0.5\end{bmatrix}-\begin{bmatrix}0.5\\ 0.5\end{bmatrix}\begin{bmatrix}0.5 & 0.5\end{bmatrix}=\begin{bmatrix}0.5 & 0.25\\ 0.25 & 0.5\end{bmatrix}-\begin{bmatrix}0.25 & 0.25\\ 0.25 & 0.25\end{bmatrix}=\begin{bmatrix}0.25 & 0\\ 0 & 0.25\end{bmatrix}$$

4. X_1,X_2 בלתי תלויים כיוון ו $\forall 0\le i,j\le 1, p(i,j)=0.25=0.5*0.5=p(i)*p(j)$

5. כיוון ו X_1,X_2 בלתי תלויים הם גם חסרי קורלציה (המשפט משאלה 23 קיים גם למשתנים מקריים סקלריים (מקרה פרטי))

6. נשים לב כי $\forall y\in\Omega, Y_1(y)\in\{0,1,2\}, Y_2(y)\in\{0,1\}$

נתבונן בטבלה הבאה:

$X_1\backslash X_2$	0	1
0	$Y_1=0,Y_2=0$	$Y_1=1,Y_2=1$
1	$Y_1=1,Y_2=1$	$Y_1=2,Y_2=0$

מהטבלה ניתן לראות כי קיימים רק 3 צירפוים אפשריים של Y_1 ו - Y_2

$$p_{Y_1,Y_2}(0,0)=p_X(0,0)=0.25$$

$$p_{Y_1,Y_2}(1,1)=p_X(1,0)+p_X(0,1)=0.25+0.25=0.5$$

$$p_{Y_1,Y_2}(2,0)=p_X(1,0)=0.25$$

7. $$p_{Y_1}(y_1)=\sum_{y_2}p_{Y_1,Y_2}(y_1,y_2)=p_{Y_1,Y_2}(y_1,0)+p_{Y_1,Y_2}(y_1,1)=\begin{cases}0.25 & y_1=0\\ 0.5 & y_1=1\\ 0.25 & y_1=2\end{cases}$$
$$p_{Y_2}(y_2)=\sum_{y_1}p_{Y_1,Y_2}(y_1,y_2)=p_{Y_1,Y_2}(0,y_2)+p_{Y_1,Y_2}(1,y_2)+p_{Y_1,Y_2}(2,y_2)=\begin{cases}0.5 & y_2=0\\ 0.5 & y_2=1\end{cases}$$

8. $$E(Y)\overset{Definition20}{=}\begin{bmatrix}E(Y_1)\\ E(Y_2)\end{bmatrix}=\begin{bmatrix}1\\ 0.5\end{bmatrix}$$

9. $E(YY^T) \overset{probelm-11}{=} \begin{bmatrix} E(Y_1^2) & E(Y_1Y_2) \\ E(Y_1Y_2) & E(Y_2^2) \end{bmatrix} = \begin{bmatrix} 1.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}$
$$\Sigma_X \overset{fact-8}{=} E(YY^T) - \mu_X\mu_X^T = \begin{bmatrix} 1.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix} - \begin{bmatrix} 1 \\ 0.5 \end{bmatrix}\begin{bmatrix} 1 & 0.5 \end{bmatrix} = \begin{bmatrix} 1.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix} - \begin{bmatrix} 1 & 0.5 \\ 0.5 & 0.25 \end{bmatrix} = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.25 \end{bmatrix}$$
10. Y_1, Y_2 אינם בלתי תלויים כיוון ש $p(Y_1 = 1, Y_2 = 1) = 0.5 \neq 0.25 = 0.5 * 0.5 = p(Y_1 = 1)p(Y_2 = 1)$, כלומר פונקציית ההתפלגות המשותפת אינה מכפלה של ההתפלגויות השוליות
11. Y_1, Y_2 חסרי קורלציה כיוון ו $E(Y_1Y_2) = 0.5 = 1 * 0.5 = E(Y_1)E(Y_2)$

problem 26:

יהיו X, Y, Z וקטורים מקריים ויהיו x,y,zערכים אפשריים שלהם בהתאמה
$\Rightarrow$ נניח $X \perp\!\!\!\perp Y|Z$ אז
$$P(x|y,z) = \frac{P(x,y,z)}{P(y,z)} = \frac{P(x,y|z)P(z)}{P(y|z)P(z)} = \frac{P(x|z)P(y|z)P(z)}{P(y|z)P(z)} = P(x|z)$$
$:\Rightarrow$ נניח $P(x|y,z) = P(x|z)$ אז $\frac{P(x,z)}{P(z)} = \frac{P(x,y,z)}{P(y,z)}$
$$P(x,y|z) = \frac{P(x,y,z)}{P(z)} = \frac{P(x,y,z)P(y,z)}{P(y,z)P(z)} = \frac{P(x,z)Py,z)}{P(z)P(z)} = P(x|z)P(y|z)$$

problem 27:

1. נשים לב כי $\forall x, y, z \in \mathbb{R}$exp(x+xz+yz)=exp(x+xz)*exp(yz)ולכן ניתן להגדיר

$$f(x,z) = \begin{cases} \exp(x+xz) & (x,z) \in B \\ 0 & otherwise \end{cases}, g(y,z) = \begin{cases} \exp(yz) & (y,z) \in B \\ 0 & otherwise \end{cases}$$

ויתקבל כי $p(x,y,z) \propto f(x,z)g(y,z)$ ולכן מעובדה 18 $X \perp\!\!\!\perp Y|Z$

2. נשים לב כי $\exp(xyz) = (\exp(xy))^z$. ולכן יהיה קייפ פירוק כמו בעובדה 18 רק אם קיימת פקטוריזציה ל $\exp(xy)$ לשתי פונקציות במשתנה יחיד.
לא קיים פירוק כזה ולכן לא קיים פירוק מתאים ל $\exp(xyz)$ ולכן מעובדה 18 $X \not\perp\!\!\!\perp Y|Z$

Part II
Markov Chains

problem 1:

(d)

y_3	$p(y_3)$
0	$(1-\theta)^3$
1	$\binom{3}{1}\theta(1-\theta)^2 = 3\theta(1-\theta)^2$
2	$\binom{3}{2}\theta^2(1-\theta)^3 = 3\theta^2(1-\theta)$
3	θ^3

(i)

$$p(y_3|x_{11},x_2,y_1) \overset{Y_1 \equiv X_1}{=} p(y_3|x_1,x_2) \overset{h}{=} \begin{cases} x_1+x_2 & with\, probability\, \theta \\ x_1+x_2+1 & with\, probability\, \theta \end{cases}$$

(j)

$$p(y_3|x_1,x_2,y_2) = p(y_3|y_2) = \begin{cases} y_2 & with\, probability\, \theta \\ y_2+1 & with\, probability\, \theta \end{cases}$$

(k)

נשים לב כי $Y_3 = X_3 + X_2 + X_1 = X_3 + Y_2$ ולכן

$$p(y_3|x_1,x_2,x_3,y_2) = p(y_3|x_3,y_2) = 1^{1_{y_3=x_3+y_2}}$$

(l)

$$p(y_3|y_1,y_2,y_3) = p(y_3|y_3) = 1^{1_{y_3=y_3}}$$

problem 2:

יהיו $P_1, P_2 \in M_{n\times n}$מטריצות מעבר של שרשראות מרקוב נראה כי $P_3 = P_1P_2$ הינה מטריצת מעבר של שרשרת מרקוב כלשהי.

- יהיו $1 \le i,j \le n$ נראה כי $0 \le (P_3)_{i,j} \le 1$.

$$(P_3)_{ij} = \sum_{k=1}^{n} (P_1)_{ik}(P_2)_{kj}$$

כיוון ו P_2 מטריצת מעבר אז $\forall 0 \leq l, m \leq n, (P_2)_{lm} \leq 1$ ולכן

$$(P_3)_{ij} = \sum_{k=1}^{n} (P_1)_{ik}(P_2)_{kj} \leq \sum_{k=1}^{n} (P_1)_{ik} = 1$$

בנוסף כיוון וכל הכניסות בP_1, P_2 אי שליליות בהכרח $0 \leq \sum_{k=1}^{n} (P_1)_{ik}(P_2)_{kj} = (P_3)_{ij}$.

-

$$\sum_{j=0}^{n} (P_3)_{ij} = \sum_{j=0}^{n}\sum_{k=1}^{n} (P_1)_{ik}(P_2)_{kj} = \sum_{k=0}^{n}\sum_{j=1}^{n} (P_1)_{ik}(P_2)_{kj} =$$
$$= \sum_{k=0}^{n} (P_1)_{ik} \sum_{j=1}^{n} (P_2)_{kj} \overset{eq.\,21}{=} \sum_{k=0}^{n} (P_1)_{ik} \overset{eq.\,21}{=} 1$$

problem 3:

$P(X_{n+1} = j | X_0 = i) = \sum_{k=1}^{n} P(X_{n+1} = j, X_n = k | X_0 = i)$ השיוויון נובע מזה שצד שמאל הינו המרג'ינל של צד ימין, ואנו מסצעים סכימה על כל ערכי X_n.

$= \sum_{k=1}^{n} P(X_{n+1} = j | X_n = k, X_0 = i) P(X_n = k | X_0 = i)$ נובע מכלל בייס

$\sum_{k=1}^{n} P(X_{n+1} = j | X_n = k) P(X_n = k | X_0 = i)$ נובע ממרקוביות $P(x_n | x_1, ..., x_{n-1}) = P(x_n | x_{n-1})$

$\sum_{k=1}^{n} P(Xn = k | X0 = i) P(Xn + 1 = j | Xn = k)$ קומטטביות כפל של סקלרים

$\sum p_n(i,k)p(k,j)$ לפי הגדרה להסתברות לעבור בk צעדים בשרשרת מרקוב

$(the\, i-th\, row\, of\, P^n)\, times\, (the\, j-th\, column\, of\, P)$ מהנחת האינדוקציה ומקרה הבסיס

$= the\, (i,j)\, entry\, of\, P^n$ מהגדרת כפל מטריצות

problem 4:

$$\pi_0 P^n = \begin{bmatrix} \sum_{k=1}^{N} (\pi_0)_k P_{1k} & ... & \sum_{k=1}^{N} (\pi_0)_k P_{Nk} \end{bmatrix} \overset{problem\, 3}{=} \begin{bmatrix} \sum_{k=1}^{N} P(X_0 = k) P(X_n = i | X_0 = k) & ... & \sum_{k=1}^{N} P(X_0 = k) P(X_n = N | X_0 = k) \end{bmatrix} =$$
$$\overset{bayes}{=} \begin{bmatrix} \sum_{k=1}^{N} P(X_n = i, X_0 = k) & ... & \sum_{k=1}^{N} P(X_n = N, X_0 = k) \end{bmatrix} \overset{marginal}{=} \begin{bmatrix} \sum_{k=1}^{N} P(X_n = i) & ... & \sum_{k=1}^{N} P(X_n = N) \end{bmatrix}$$

problem 5:

תהי P מטריצת מעברים מגודל $N\times N$ אשר כל שורתיה זהות ויהי $P=\begin{bmatrix}\pi\\ \vdots\\ \pi\end{bmatrix}$ עבור וקטור התפלגות כלשהו π מגודל $1\times N$

(a)

$$P_{ij}^2=\sum_{k=1}^{N}P_{ik}P_{kj}=\sum_{k=1}^{N}P_{ik}\pi_j=\pi_j\sum_{k=1}^{N}P_{ik}\overset{\pi\, is\, distrebution}{=}\pi_j=P_{ij}$$

(b)

אוכיח באינדוקציה על n.
בסיס: הוכח בסעיף הקודם, נניח שהטענה נכונה עבור n נראה עבור $n+1$
$P^{n+1}=P^nP$ אז מהנחת האינדוקציה $P^nP=PP=P^2=P$

(c)

יהי $\widetilde{\pi}$וקטור התפלגות מגודל $1\times N$

$$\widetilde{\pi}P=\begin{bmatrix}\sum_{k=1}^{N}\widetilde{\pi}_kP_{1k} & \dots & \sum_{k=1}^{N}\widetilde{\pi}_kP_{Nk}\end{bmatrix}=\begin{bmatrix}\sum_{k=1}^{N}\pi_1\widetilde{\pi}_k & \dots & \sum_{k=1}^{N}\pi_N\widetilde{\pi}_k\end{bmatrix}=\begin{bmatrix}\pi_1\sum_{k=1}^{N}\widetilde{\pi}_k & \dots & \pi_N\sum_{k=1}^{N}\widetilde{\pi}_k\end{bmatrix}\overset{\sum_{i=1}^{N}\widetilde{\pi}_i=1}{=}\begin{bmatrix}\pi_1 & \dots & \pi_N\end{bmatrix}=\pi$$

(d)

יהיו $n\geq 1$ ו π_0 וקטור התפלגות כלשהו מגודל $1\times N$
$\pi_n\overset{fact\,1}{=}\pi_0P^{n-1}\overset{(b)}{=}\pi_0P\overset{(c)}{=}\pi$

Part III
SPD Matrices, Inner Products, Norms, and Metrics

problem 1:

תהי A מטריצת SPD
מעובדה 1 אנו יודעים כי כל ערכיה העצמיים של A חיוביים ולכן שונים מאפס, אזי A הפיכה (מטריצה הפיכה אמ"מ 0 לא ע"ע שלה)
ולכל λ ע"ע של A מתקיים λ^{-1} הינו ע"ע של A^{-1}. הוכחה: יהי x וע של A ויהי λ העע המתאים לו אז:

$$A^{-1}Ax=x$$
$$A^{-1}Ax=A^{-1}\lambda x=\lambda A^{-1}x$$

λ עע של A ולכן שונה מאפס אז ניתן לחלק בו. אז$A^{-1}x=\lambda^{-1}x$ולכן λ^{-1} הינו ע"ע של A^{-1}
אז גם כל הע"ע של A^{-1}הינם חיוביים ולכן מעובדה 1 A^{-1} הינה SPD

problem 2:

נזכר כי כפל מטריצות הינו אסוציטיבי ודיסטרביוטיבי
תהי $Q \in M_{n\times n}(\mathbb{R})$ SPD נראה כי $\langle,\rangle_Q : (x,y) \mapsto xTQy$ מכפלה פנימית.

- $\langle x,y\rangle_Q = \langle y,x\rangle_Q$
  יהיו $x,y \in \mathbb{R}^n$
  $$\langle x,y\rangle_Q = x^TQy = \left[\sum_{i=1}^n x_iQ_{i1}, \sum_{i=1}^n x_iQ_{i2}, ..., \sum_{i=1}^n x_iQ_{in}\right] y = \sum_{j=1}^n y_j \sum_{i=1}^n x_iQ_{ij} = \sum_{i=1}^n\sum_{j=1}^n x_iy_jQ_{ij}$$
  $$\langle y,x\rangle_Q = y^TQx = \left[\sum_{i=1}^n y_iQ_{i1}, \sum_{i=1}^n y_iQ_{i2}, ..., \sum_{i=1}^n y_iQ_{in}\right] x = \sum_{j=1}^n x_j \sum_{i=1}^n y_iQ_{ij} = \sum_{i=1}^n\sum_{j=1}^n x_iy_jQ_{ij}$$
  אז $\langle x,y\rangle_Q = \langle y,x\rangle_Q$
- $\langle \alpha x+\beta y,z\rangle_Q = \alpha\langle x,z\rangle_Q + \beta\langle y,z\rangle_Q$
  יהיו $x,y,z \in \mathbb{R}^n$
  $$\langle \alpha x+\beta y,z\rangle_Q = (\alpha x+\beta y)^TQz \overset{distributive}{=} \alpha x^TQz + \beta y^TQz \overset{associative}{=} \alpha\langle x,z\rangle_Q + \beta\langle y,z\rangle_Q$$
- מהגדרת PD מתקיימת התכונה $\forall x, \langle x,x\rangle_Q \geq 0$
- בשאלה 1 הוכחנו כי SPD הינה הפיכה ולכן $dim(Ker(Q)) = 0$ כלומר $x = 0_v \iff Qx = 0$

problem 3:

תהי $Q \in M_{n\times n}(\mathbb{F})$ SPD ויהי $LL^T = Q$ פירוק הצ'יקולסקי שלה.
$$||x||_Q = (x^TQx)^{\frac{1}{2}} = (x^TLL^Tx)^{\frac{1}{2}} = ((L^Tx)^T(L^Tx))^{\frac{1}{2}} = (\sum_{i=1}^n (L^Tx)_i^2)^{\frac{1}{2}} = ||L^Tx||_{l_2}$$

Computer Exercise 1:

In [1]:

```python
import numpy as np
import pylab
pylab.ion()
from pylab import plt
```

In [2]:

```python
x2,x1 = np.mgrid[-3:3:0.01,-3:3:0.01]
```

In [3]:

```python
def QnormSquared(Q):
    return x1**2*Q[0,0]+x2**2*Q[1,1]+2*x1*x2*Q[0,1]
```

In [4]:

```python
def axis_xy(g=None):
    # mimics Matlab's "axis xy" command;
    # making the y axis pointing upward when using imshow
    if g is None:
        g = plt.gca()
    bottom, top = g.get_ylim()
    if top<bottom:
        g.set_ylim(top,bottom)
```

In [5]:

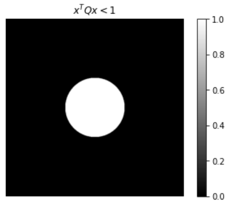
```python
def visualize(Q):
    plt.figure()
    plt.imshow((QnormSquared(Q)<1).astype(np.int64),cmap='gray')
    axis_xy()
    plt.colorbar()
    plt.axis('off')
    plt.title(r'$x^TQx<1$')
    plt.subplots_adjust(wspace=0.4)
```

In [6]:

```python
Q=np.eye(2)
print ('Q=\n',Q)
visualize(Q)
```

```
Q=
 [[1. 0.]
 [0. 1.]]
```

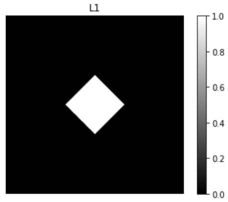
$x^TQx < 1$



In [7]:

```python
plt.figure()
plt.imshow((abs(x1)+abs(x2)<1).astype(np.int64),cmap='gray')
axis_xy()
plt.colorbar()
plt.axis('off')
plt.title(r'L1')
plt.subplots_adjust(wspace=0.4)
```

L1

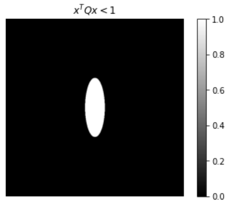


In [8]:

```python
Q=np.eye(2)*np.diag([9,1])
print ('Q=\n',Q)
visualize(Q)
```

```
Q=
 [[9. 0.]
 [0. 1.]]
```

$x^TQx < 1$

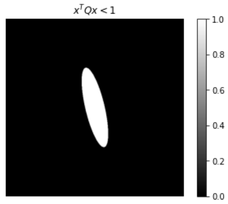


In [9]:

```python
Q=np.array([[9,2],[2,1]])
print ('Q=\n',Q)
visualize(Q)
```

```
Q=
 [[9 2]
 [2 1]]
```

$x^TQx < 1$

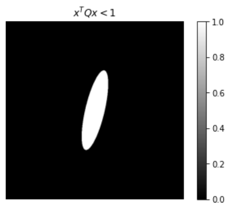


In [10]:

```python
Q=np.array([[9,-2],[-2,1]])
print ('Q=\n',Q)
visualize(Q)
```

```
Q=
 [[ 9 -2]
 [-2  1]]
```

$x^TQx < 1$



Part IV
Gaussian Random Vectors

probelm 1:

יהי $X = \begin{bmatrix} X_1 & ... & X_n \end{bmatrix}^T \sim N(0_{n\times 1}, I_{n\times n})$.

מהגדרה $p(x) = \frac{1}{(2\pi)^{\frac{n}{2}}}\exp(-\frac{1}{2}||x||_{l2}^2) = \frac{1}{(2\pi)^{\frac{n}{2}}}\exp(-\frac{1}{2}\sum\limits_{i=1}^{n} x_i^2) = \prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}}\exp(\frac{-x_i^2}{2})$

יהי $1 \leq i \leq n$

אז $p_{X_i}(x) = \int\limits_{x_1,...,x_{i-1},x_{i+1},..,x_n}\prod_{j=1}^{n}\frac{1}{\sqrt{2\pi}}\exp(\frac{-x_j^2}{2}) = \frac{1}{\sqrt{2\pi}}\exp(\frac{-x_i^2}{2})$, אז $X_i \sim N(0,1)$ וגם $p(x) = \prod_{i=1}^{n} p_{X_i}(x_i)$

באופן דומה ניתן להציג כל התפלגות שולית של מספר משתנים כמכפלה של ההתפלגויות השוליות של כל משתנה בודד וגם כל המשתנים הבודדים מתפלגים זהה אז

הם i.i.d

problem 2:

יהי X מ"מ כך ש $X \sim N(\mu, \sigma^2)$ ויהיו $a, b \in \mathbb{R}$.

$F_{aX+b}(x) = P(aX+b \leq x) = P(X \leq \frac{x-b}{a}) = \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\frac{x-b}{a}}\exp(-\frac{1}{2}(\frac{(t-\mu)^2}{\sigma^2}))d$

נציב $t = \frac{u-b}{a}$ אז $dt = \frac{du}{a}$ ונקבל כי

$$\frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\frac{x-b}{a}}\exp(-\frac{1}{2}(\frac{(t-\mu)^2}{\sigma^2}))dt = \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\text{x}}\frac{\exp(-\frac{1}{2}(\frac{(\frac{u-b}{a}-\mu)^2}{\sigma^2}))}{a}du = \frac{1}{\sqrt{2\pi\sigma^2a^2}}\int_{-\infty}^{\text{x}}\exp(-\frac{1}{2}(\frac{(\frac{u-b}{a}-\mu)^2}{\sigma^2}))du =$$

$$F_{aX+b}(x) = \frac{1}{\sqrt{2\pi\sigma^2a^2}}\int_{-\infty}^{\text{x}}\exp(-\frac{1}{2}(\frac{(u-(b+a\mu))^2}{\sigma^2a^2}))du$$

אז $p_{aX+b}(x) = F'_{aX+b}(x) = \frac{1}{\sqrt{2\pi\sigma^2a^2}}\exp(-\frac{1}{2}(\frac{(x-(b+a\mu))^2}{\sigma^2a^2}))$

ולכן $(aX+b) \sim N(a\mu+b, a^2\sigma^2)$

problem 3:

יהי $X = \begin{bmatrix} X_1 & ... & X_n \end{bmatrix}^T \sim N(\mu, \Sigma)$

(a)

יהי $1 \leq i \leq n$ אז $X_i = \begin{bmatrix} 0_{1\times i-1} & 1 & 0_{1\times(n-1)} \end{bmatrix} X$ ולכן מעובדה 7 הינו גאוסיאני.

בנוסף מעובדה 7 מתקבל כי:

$E(X_i) = \mu_i, \sigma_{X_i}^2 = \begin{bmatrix} 0_{1\times i-1} & 1 & 0_{1\times(n-i)} \end{bmatrix}\Sigma\begin{bmatrix} 0_{1\times i-1} & 1 & 0_{1\times(n-i)} \end{bmatrix}^T = \begin{bmatrix} \Sigma_{i1} & ... & \Sigma_{in} \end{bmatrix}\begin{bmatrix} 0_{1xi} & 1 & 0_{1\times(n-i)} \end{bmatrix}^T = \Sigma_{ii}$

(b)

יהיו $1 \le k < n$ ויהיו $1 \le i_1 < ... < i_k \le n$.

נסמן $A = \begin{bmatrix} 0_{1x(i_1-1)} & 1 & 0_{1x(n-i_1)} \\ \vdots & \dots & \vdots \\ 0_{1x(i_k-1)} & 1 & 0_{1x(n-i_k)} \end{bmatrix}$ אז מעובדה 7 $Y = AX$ הינו גאוסייאני. נשים לב כי מהגדרת A $Y = \begin{bmatrix} X_{i_1} \\ \vdots \\ X_{i_k} \end{bmatrix}$

בנוסף מעובדה 7 נקבל כי $E(Y) = A\mu_x = \begin{bmatrix} \mu_{i_1} \\ \vdots \\ \mu_{i_k} \end{bmatrix}$ וגם כי $(\Sigma_Y)_{kj} = \Sigma_{i_k i_j}$

problem 4:

יהיו $X \sim N(\mu_X, \Sigma_X)$ וקטור רנדומי ממימד n, $Y \sim N(\mu_Y, \Sigma_Y)$ וקטור רנדומי ממימד m. נניח כי $X \perp\!\!\!\perp Y$.

מהגדרת תוחלת על ו"ר ומעובדה 13 בפרק על וקטורים רנדומליים אנו מקבלים כי עבור $Z = \begin{bmatrix} X \\ Y \end{bmatrix}, E(Z) = \begin{bmatrix} E(X) \\ E(Y) \end{bmatrix}, \Sigma_Z = \begin{bmatrix} \Sigma_X & 0_{n\times m} \\ 0_{m\times n} & \Sigma_Y \end{bmatrix}$.

נראה כי $Z \sim N(\begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \begin{bmatrix} \Sigma_X & 0_{n\times m} \\ 0_{m\times n} & \Sigma_Y \end{bmatrix})$.

$p(z) = p(x,y) \overset{X \perp\!\!\!\perp Y}{=} p(x)p(y)$

אז

$$p(z) = (\frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma_X|^{\frac{1}{2}}}\exp(-\frac{1}{2}||x-\mu_x||_{\Sigma_X^{-1}}))(\frac{1}{(2\pi)^{\frac{m}{2}}|\Sigma_Y|^{\frac{1}{2}}}\exp(-\frac{1}{2}||y-\mu_x||_{\Sigma_Y^{-1}})) = \frac{1}{(2\pi)^{\frac{m+n}{2}}|\Sigma_X|^{\frac{1}{2}}|\Sigma_Y|^{\frac{1}{2}}}\exp(-\frac{1}{2}(||x-\mu_x||_{\Sigma_X^{-1}}$$

$$||z-\mu_Z||_{\Sigma_Z^{-1}} = (\begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} \mu_X \\ \mu \end{bmatrix})^T \begin{bmatrix} \Sigma_X & 0_{n\times m} \\ 0_{m\times n} & \Sigma_Y \end{bmatrix}^{-1} (\begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} \mu_X \\ \mu \end{bmatrix}) = \begin{bmatrix} x-\mu_X \\ y-\mu_Y \end{bmatrix}^T \begin{bmatrix} \Sigma_X^{-1} & 0_{n\times m} \\ 0_{m\times n} & \Sigma_Y^{-1} \end{bmatrix} \begin{bmatrix} x-\mu_X \\ y-\mu_Y \end{bmatrix} =$$

$$= \begin{bmatrix} (x-\mu_X)^T\Sigma_X^{-1} \\ (y-\mu_Y)^T\Sigma_Y^{-1} \end{bmatrix}^T \begin{bmatrix} x-\mu_X \\ y-\mu_Y \end{bmatrix} = (x-\mu_X)^T\Sigma_X^{-1}(x-\mu_X) + (y-\mu_Y)^T\Sigma_Y^{-1}(y-\mu_Y) = ||x-\mu_x||_{\Sigma_X^{-1}} + ||y-\mu_x||_{\Sigma_Y^{-1}}$$

נשים לב כי כיוון ומדובר במטריצתץ בלוקים אלכסונית מתקיים כי : $det(\begin{bmatrix} \Sigma_X & 0_{n\times m} \\ 0_{m\times n} & \Sigma_Y \end{bmatrix}) = det(\Sigma_X) \times det(\Sigma_Y)$

אז קיבלנו כי $p(z) = p(x,y)\frac{1}{(2\pi)^{\frac{m+n}{2}}|\Sigma_X|^{\frac{1}{2}}\Sigma_Y|^{\frac{1}{2}}}\exp(-\frac{1}{2}(||x-\mu_x||_{\Sigma_X^{-1}} + ||y-\mu_x||_{\Sigma_Y^{-1}}) = \frac{1}{(2\pi)^{\frac{m+n}{2}}|\Sigma_Z|^{\frac{1}{2}}}\exp(-\frac{1}{2}(||z-\mu_Z||_{\Sigma_Z^{-1}})$

ולכן $Z \sim N(\begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \begin{bmatrix} \Sigma_X & 0_{n\times m} \\ 0_{m\times n} & \Sigma_Y \end{bmatrix})$

problem 5:

יהיו $X \sim N(\mu_X, \Sigma_X), Y \sim N(\mu_Y, \Sigma_Y)$ ו"ר ממימד n כל אחד. כך ש $X \perp\!\!\!\perp Y$. נסמן $Z = X + Y$

(a)

צ.ל $A \in M_{n\times 2n}(\mathbb{R})))))$ כך ש $Z = A\begin{bmatrix} X \\ Y \end{bmatrix}$

$$\begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} X_1 \\ \vdots \\ X_n \\ Y_1 \\ \vdots \\ Y_n \end{bmatrix}$$

נבחר
$$A = \begin{bmatrix} 1 & 0_{1\times(n-1)} & 1 & 0_{1\times(n-2)} & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0_{1\times(i-1)} & 1 & 0_{1\times(n-1)} & 1 & 0_{1\times(n-i)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0_{1\times(n-2)} & 1 & 1_{1\times(n-1)} & 0 \end{bmatrix}$$

$$A \begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} 1 & 0_{1\times(n-1)} & 1 & 0_{1\times(n-2)} & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0_{1\times(i-1)} & 1 & 0_{1\times(n-1)} & 1 & 0_{1\times(n-i)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0_{1\times(n-2)} & 1 & 1_{1\times(n-1)} & 0 \end{bmatrix} \begin{bmatrix} X_1 \\ \vdots \\ X_n \\ Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} X_1 + Y_1 \\ \vdots \\ X_n + Y_n \end{bmatrix} = X + Y$$

(b)

נשתמש ב A שמצאנו בסעיף א אז מעובדות 13 ו7 אנו יודעים כי $Z \sim N(A \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, A \begin{bmatrix} \Sigma_X & 0_{n\times n} \\ 0_{n\times n} & \Sigma_Y \end{bmatrix} A^T)$

באופן זהה לסעיף א נקבל כי $E(Z) = A \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix} = \mu_X + \mu_Y$

$$A \begin{bmatrix} \Sigma_X & 0_{n\times n} \\ 0_{n\times n} & \Sigma_Y \end{bmatrix} A^T = \begin{bmatrix} \Sigma_X & \Sigma_Y \end{bmatrix} A^T = \Sigma_X + \Sigma_Y$$