עבודה 1 – מבוא למודלים גרפיים ולמידה עמוקה

וקטורים אקראיים

1. נגדיר סדרה אינסופית $\langle A_i \rangle_{i=1}^{\infty}$ כך שלכל $i > n$: $A_i = \emptyset$. ברור שזו סדרה של קבוצות זרות כי הן ריקות. אזי:

$$\sum_{i=1}^{n} P(A_i) = \sum_{i=1}^{n} P(A_i) + 0 = \sum_{i=1}^{n} P(A_i) + \sum_{i=n+1}^{\infty} 0 = \sum_{i=1}^{n} P(A_i) + \sum_{i=n+1}^{\infty} P(A_i) = \sum_{i=1}^{\infty} P(A_i) = P\left(\bigcup_{i=1}^{\infty} A_i\right)$$

$$= P\left(\left(\bigcup_{i=1}^{n} A_i\right) \cup \left(\bigcup_{i=n+1}^{\infty} A_i\right)\right) = P\left(\left(\bigcup_{i=1}^{n} A_i\right) \cup \left(\bigcup_{i=n+1}^{\infty} \emptyset\right)\right) = P\left(\left(\bigcup_{i=1}^{n} A_i\right) \cup \emptyset\right)$$

$$= P\left(\bigcup_{i=1}^{n} A_i\right)$$

כנדרש.

2. יהי מאורע A. אזי:

$$1 = P(\Omega) = P(A \cup A^C) = P(A) + P(A^C) \rightarrow P(A^C) = 1 - P(A)$$

כנדרש.

3. *(1)* $X^{-1}(B) = \Omega$ כיוון שלכל קלט של X הפלט שלו ב-B. אזי $P\left(X^{-1}(B)\right) = 1$

(2) $X^{-1}(B) = \emptyset$ כיוון שאין ב-B אף איבר שיכול להיות פלט של X. אזי $P\left(X^{-1}(B)\right) = 0$.

4. זו לא פונקציית CDF כיוון שהיא לא רציפה מימין ב-$x = 42$.

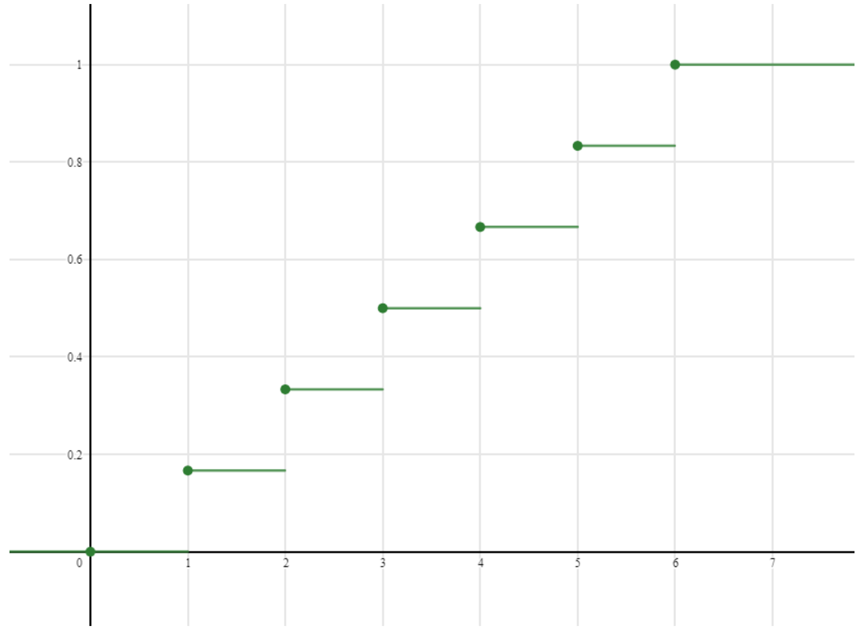
5. *(1)* X רציף לכן ההסתברות שיקבל כל ערך ספציפי היא 0.

(2) $B = \{x\}$ עבור $x \in R^n$ כלשהו.

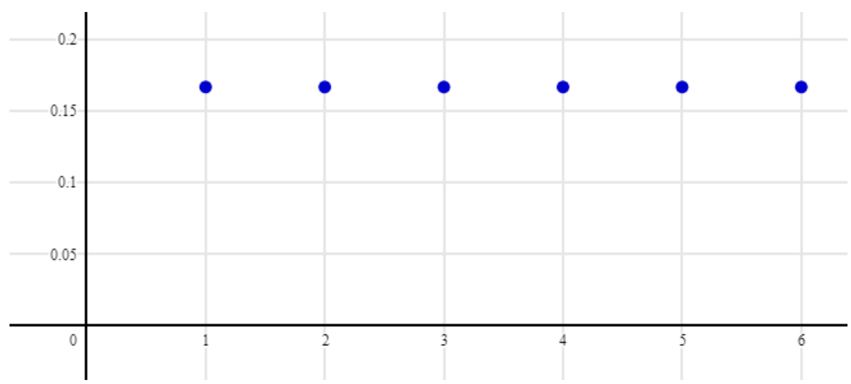
(3) אם כל B_i בת מנייה אז האיחוד שלהן בת מנייה ואז מתקבלת הסתברות 0 כיוון ש-X רציף.

6.

$$f(x) = \begin{cases} 0 & : x < 1 \\ \frac{1}{6} & : 1 \le x < 2 \\ \frac{2}{6} & : 2 \le x < 3 \\ \frac{3}{6} & : 3 \le x < 4 \\ \frac{4}{6} & : 4 \le x < 5 \\ \frac{5}{6} & : 5 \le x < 6 \\ 1 & : x \ge 6 \end{cases}$$



$$p(x) = \begin{cases} \frac{1}{6} & : x = 1 \\ \frac{1}{6} & : x = 2 \\ \frac{1}{6} & : x = 3 \\ \frac{1}{6} & : x = 4 \\ \frac{1}{6} & : x = 5 \\ \frac{1}{6} & : x = 6 \end{cases}$$



7.

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

8.

$$E(X) = \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \frac{1}{6} \cdot 3 + \frac{1}{6} \cdot 4 + \frac{1}{6} \cdot 5 + \frac{1}{6} \cdot 6 = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = \frac{1}{6} \cdot 21 = 3.5$$

9. $(xx^T)^T = (x^T)^T x^T = xx^T$ כנדרש.

10. נסמן כל תא ב-X ע"י $X_{i,j}$. אזי:

$$E(X^T)_{i,j} = E(X_{j,i}) = (E(X)^T)_{i,j}$$

11. מהגדרת כפל מטריצות, $(XX^T)_{i,j} = X_iX_j$. אזי $(R_X)_{i,j} = E(XX^T)_{i,j} = E(X_iX_j)$.

12. ההוכחה זהה לשאלה 9.

13. $(\sum_X)_{i,j} = E\left((X-\mu)(X-\mu)^T\right)_{i,j} = E\left((X-\mu)_i(X-\mu)_j\right) = E\left(\left(X_i - E(X_i)\right)\left(X_j - E(X_j)\right)\right)$

14. $(\sum_X)_{i,i} = E\left(\left(X_i - E(X_i)\right)^2\right) = Var(X)$ לפי ההגדרה.

15. $(\sum_X)^T = E\left((X-\mu)(X-\mu)^T\right) = R_{X-\mu}$ וכבר ראינו ש-$R_{X-\mu}$ סימטרי.

16. (1) מתוך זהות ידוע על *covariance*:

$$(\sum)_{i,j} = E\left(\left(X_i - E(X_i)\right)\left(X_j - E(X_j)\right)\right) = E(X_iX_j) - \mu_i\mu_j = (E(XX^T) - \mu\mu^T)_{i,j}$$

(2) $E\left((X-\mu)X^T\right)_{i,j} = E\left((X-\mu)_iX_j\right) = E\left(X_iX_j - \mu_iX_j\right) = E(X_iX_j) - \mu_iE(X_j) = E(X_iX_j) - \mu_i\mu_j$

אזי זו מטריצת השונויות המשותפות כלומר $\sum$.

(3) לפי התוצאה הקודמת: $E(X(X-\mu)^T) = (E(X(X-\mu)^T)^T)^T = E\left((X-\mu)X^T\right)^T = \sum^T = \sum$

17. *(a)* $R_{YX} = E(YX^T) = (E(YX^T)^T)^T = E(XY^T)^T = (R_{XY})^T$

(b)

$$\sum_{YX} = E\left((Y-\mu_Y)(X-\mu_X)^T\right) = \left(E\left((Y-\mu_Y)(X-\mu_X)^T\right)^T\right)^T = E\left((X-\mu_X)(Y-\mu_Y)^T\right)^T = (\sum_{XY})^T$$

(c)

$$(\Sigma_{XY})_{i,j} = E\left((X-\mu_X)(Y-\mu_Y)^T\right)_{i,j} = E\left((X-\mu_X)_i(Y-\mu_Y)_j\right) = E\left(\left(X_i - E(X_i)\right)\left(Y_j - E(Y_j)\right)\right) = E(X_iY_j) - E(X_i)E(Y_j) = (R_{XY} - \mu_X\mu_Y^T)_{i,j}$$

18. $\mu_Y = E(AX+b) = E(A)E(X) + E(b) = A\mu_X + b$

$$\Sigma_Y = \Sigma_{AX+b} = E\left((AX+b-\mu_Y)(AX+b-\mu_Y)^T\right) =$$

$$E\left((AX+b-A\mu_X-b)(AX+b-A\mu_X-b)^T\right) = E\left((AX-A\mu_X)(AX-A\mu_X)^T\right) =$$

$$E(A(X-\mu_X)(X-\mu_X)^TA^T) = E(A)E\left((X-\mu_X)(X-\mu_X)^T\right)E(A^T) = A\,\Sigma_X\,A^T$$

19. נסמן: $Y = 1^TX$. אזי לפי הסעיף הקודם: $\Sigma_Y = 1^T\,\Sigma_X\,1 = \sigma^2 1^T \cdot 1 = \sigma^2 n$ וזו השונות של Y.

20. נניח ש-A,B ב"ת בהינתן C. אזי:

$$P(A|B,C) = \frac{P(A\cap B\cap C)}{P(B\cap C)} = \frac{P(A\cap B\cap C)P(C)}{P(B\cap C)P(C)} = P(A\cap B|C)\cdot\frac{P(C)}{P(B\cap C)} = P(A|C)P(B|C)\cdot P(B|C)^{-1} = P(A|C)$$

כעט נניח ש-$P(A|B,C) = P(A|C)$. אזי:

$$P(A\cap B|C) = \frac{P(A\cap B\cap C)}{P(C)} = \frac{P(A\cap B\cap C)P(B\cap C)}{P(C)P(B\cap C)} = P(A|B,C)P(B|C) = P(A|C)P(B|C)$$

והתנאי השלישי גם שקול בגלל הסימטריות של אי תלות כלפי A,B.

21. $E(1_A) = 1\cdot P(A) + 0\cdot P(A^C) = P(A)$

22. כפי שלמדנו: $R_Z = \begin{bmatrix} R_X & R_{XY} \\ R_{YX} & R_Y \end{bmatrix}$. אבל בהגדרה $R_{XY}, R_{YX} = 0$ כאשר X,Y אורתוגונליים לכן:

$$R_Z = \begin{bmatrix} R_X & 0 \\ 0 & R_Y \end{bmatrix}$$

באופן דומה אנו יודעים ש-$\Sigma_Z = \begin{bmatrix} \Sigma_X & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_Y \end{bmatrix}$. אבל בהגדרה $\Sigma_{XY}, \Sigma_{YX} = 0$ כאשר X,Y בלתי מתואמים לכן $\Sigma_Z = \begin{bmatrix} \Sigma_X & 0 \\ 0 & \Sigma_Y \end{bmatrix}$.

23. יהיו X, Y וקטורים מקריים בלתי תלויים. ראשית נראה שכל X_i, Y_j ב"ת:

$$\begin{aligned} p(x_i, y_j) &= \sum_{\substack{x_1,\dots,x_{i-1},x_{i+1},\dots,x_n \\ y_1,\dots,y_{j-1},y_{j+1},\dots,y_m}} p(x,y) = \sum_{\substack{x_1,\dots,x_{i-1},x_{i+1},\dots,x_n \\ y_1,\dots,y_{j-1},y_{j+1},\dots,y_m}} p(x)p(y) \\ &= \sum_{x_1,\dots,x_{i-1},x_{i+1},\dots,x_n} \sum_{y_1,\dots,y_{j-1},y_{j+1},\dots,y_m} p(x)p(y) = \sum_{x_1,\dots,x_{i-1},x_{i+1},\dots,x_n} p(x) \sum_{y_1,\dots,y_{j-1},y_{j+1},\dots,y_m} p(y) \\ &= \sum_{x_1,\dots,x_{i-1},x_{i+1},\dots,x_n} p(x)p(y_j) = p(x_i)p(y_j) \end{aligned}$$

אזי לכל זוג X_i, Y_j הם ב"ת. לכן:

$$E(XY^T)_{i,j} = E(X_iY_j) = E(X_i)E(Y_j) = \left(E(X)E(Y^T)\right)_{i,j}$$

כנדרש.

24. *(a)* $p_{X_1}(0) = p(0,0) + p(0,1) = 0.6$, $p_{X_1}(1) = p(1,0) + p(1,1) = 0.4$

$p_{X_2}(0) = p(0,0) + p(1,0) = 0.8$, $p_{X_2}(1) = p(0,1) + p(1,1) = 0.2$

(b) $E(X_1) = p_{X_1}(1) = 0.4$, $E(X_2) = p_{X_2}(1) = 0.2$. אזי $E(X) = \begin{bmatrix} 0.4 \\ 0.2 \end{bmatrix}$.

(c) $E(X_1^2) = E(X_1) = 0.4$, $E(X_2^2) = E(X_2) = 0.2$, $E(X_1X_2) = p(1,1) = 0.1$. לכן:

$$R_{XY} = \begin{bmatrix} 0.4 & 0.1 \\ 0.1 & 0.2 \end{bmatrix}$$

(d)
$$\Sigma_Y = \begin{bmatrix} 0.4 - 0.4^2 & 0.1 - 0.4 \cdot 0.2 \\ 0.1 - 0.4 \cdot 0.2 & 0.2 - 0.2^2 \end{bmatrix} = \begin{bmatrix} 0.24 & 0.02 \\ 0.02 & 0.16 \end{bmatrix}$$

(e,f) המשתנים מתואמים כיוון שהשונות המשותפת שלהם היא לא 0 (היא 0.02). כיוון שהם מתואמים, הם תלויים.

25. *(a)* p_{X_1} ו-p_{X_2} שניהם מקבלים ערך של 0.5 עבור 0 ו-1.

(b) $E(X_1) = E(X_2) = 0.5 \cdot 1 + 0.5 \cdot 0 = 0.5$. אזי $E(X) = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$

(c) $E(X_1^2) = E(X_2^2) = E(X_1) = E(X_2) = 0.5$. $E(X_1X_2) = p(1,1) = 0.25$. אזי:

$$\Sigma_X = \begin{bmatrix} 0.5 - 0.5^2 & 0.25 - 0.5^2 \\ 0.25 - 0.5^2 & 0.5 - 0.5^2 \end{bmatrix} = \begin{bmatrix} 0.25 & 0 \\ 0 & 0.25 \end{bmatrix}$$

(d) כן כיוון ש-$p(x_1, x_2) = 0.25 = 0.5^2 = p(x_1)p(x_2)$.

(e) כן כיוון שבמטריצת הקורלציה ניתן לראות שהשונות המשותפת של X_1, X_2 היא 0.

(f) $p_Y(1,1) = p(1,0) + p(0,1) = 0.5$, $p_Y(2,0) = p(1,1) = 0.25$, $p_Y(0,0) = p(0,0) = 0.25$

(g) $p_{Y_1}(0) = p_Y(0,0) = 0.25$, $p_{Y_1}(1) = p_Y(1,1) = 0.5$, $p_{Y_1}(2) = p_Y(2,0) = 0.25$.

$p_{Y_2}(0) = p_Y(0,0) + p_Y(2,0) = 0.25 + 0.25 = 0.5$, $p_{Y_2}(1) = p_Y(1,1) = 0.5$

(h) $E(Y_1) = E(X_1) + E(X_2) = 0.5 + 0.5 = 1$, $E(Y_2) = p_{Y_2}(1) = 0.5$. אזי: $E(Y) = \begin{bmatrix} 1 \\ 0.5 \end{bmatrix}$.

(i) $E(Y_1^2) = 1 \cdot 0.5 + 4 \cdot 0.25 = 1.5$, $E(Y_2^2) = E(Y_2) = 0.5$, $E(Y_1 Y_2) = p_Y(1,1) = 0.5$. אזי:

$$\Sigma_Y = \begin{bmatrix} 1.5 - 1^2 & 0.5 - 1 \cdot 0.5 \\ 0.5 - 1 \cdot 0.5 & 0.5 - 0.5^2 \end{bmatrix} = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.25 \end{bmatrix}$$

(j) $p_Y(0,0) = 0.25 \neq 0.25 \cdot 0.5 = p_{Y_1}(0) p_{Y_2}(0)$ אזי הם תלויים.

(k) ניתן לראות במטריצה שהשונות המשותפת של Y_1, Y_2 היא 0 לכן הם בלתי מתואמים.

26. נניח ש-X, Y ב"ת בהינתן Z. אזי:

$$p(x|y,z) = \frac{p(x,y,z)}{p(y,z)} = \frac{p(x,y,z)p(z)}{p(y,z)p(z)} = p(x,y|z) \cdot \frac{p(z)}{p(y,z)} = p(x|z)p(y|z) \cdot p(y|z)^{-1} = p(x|z)$$

כעט נניח ש-$p(x|y,z) = p(x|z)$. אזי:

$$p(x,y|z) = \frac{p(x,y,z)}{p(z)} = \frac{p(x,y,z)p(y,z)}{p(z)p(y,z)} = p(x|y,z)p(y|z) = p(x|z)p(y|z)$$

והתנאי השלישי גם שקול בגלל הסימטריות של אי תלות כלפי x, y.

27. *(a)* נבחין: $\exp(x + xz + yz) = \exp(x + xz)\exp(yz)$.

נגדיר: $f(x,z) = \exp(x + xz)$, $g(y,z) = \exp(yz)$ בתחום B, מחוץ לו f, g הן 0. אזי f, g אי-שליליות ומקיימות $p(x,y,z) = f(x,z)g(y,z)$ לכן X, Y ב"ת בהינתן Z.

(b) נשים לב שלפי עובדה 18, X ו-Y ב"ת בהינתן Z א"ממ אפשר לפרק את הפונקציה $p(x,y,z)$ לשתי פונקציות f,g כך ש- $p(x,y,z) = f(x,z)g(y,z)$, כלומר לפרק את הפונקצייה p לשתי פונקציות כאשר כל אחת תלויה בשתי משתנים. כדי להפריד את אחד המשתנים מהפונקציה נצתרך למצוא את פונקציית ההתפלגות השולית שלו כלומר לגזור לפי המשתנה, בגלל שהנגזרת של exp(x)’ = exp(x)·x’ לא נוכל להפריד את אחד המשתנים מהפונקציה. ולכן לא נוכל למצוא f,g כאלה. (כי נקבל שלמשל הנגזרת לפי x היא exp(xyz)·yz)

שרשראות מרקוב

1. *(a)*

x_1	x_3	$p(x_1,x_3)$
0	0	$(1-\theta)^2$
0	1	$\theta(1-\theta)$
1	0	$\theta(1-\theta)$
1	1	θ^2

(b) $Y_1 = X_1$ אזי יש להם את אותה התפלגות.

(c) Y_2 מתפלג ברנולי אזי $p(y_2) = \binom{2}{y_2}\theta^{y_2}(1-\theta)^{2-y_2}$

(d) Y_3 גם מתפלג ברנולי אזי $p(y_3) = \binom{3}{y_3}\theta^{y_3}(1-\theta)^{3-y_3}$

<u>הערה:</u> בכל הסעיפים הבאים נניח שהמאורעות שמותנים בהם לא ריקים. למשל בהסתברות $p(y_1|x_1)$ אז $x_1 \in \{0,1\}$.

(e) $Y_1 = X_1$ לכן $p(y_1|x_1) = \begin{cases}1 & if\ y_1 = x_1\\ 0 & if\ y_1 \neq x_1\end{cases}$

(f) $Y_2 = X_1 + X_2$ לכן: $p(y_2|x_1,x_2) = \begin{cases}1 & if\ y_2 = x_1 + x_2\\ 0 & if\ y_2 \neq x_1 + x_2\end{cases}$

(g) נתון לנו y_2 וברור שלא צריך יותר מזה כדי לדעת את ההתפלגות של y_2: $p(y_2'|y_1,y_2) = \begin{cases}1 & if\ y_2' = y_2\\ 0 & if\ y_2' \neq y_2\end{cases}$

(h) $Y_3 = X_1 + X_2 + X_3$. כיוון שנתונים X_1, X_2 מתקבל ש-Y_3 מתפלג באופן דומה ל-X_3:

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

(i) $X_1 = Y_1$ לכן זה שנתון Y_1 לא משנה דבר וההתפלגות היא כמו זו מהסעיף הקודם.

(j) $Y_2 = X_1 + X_2$ לכן שוב לא נוספו שום נתונים חדשים והתשובה היא כמו סעיף *h*.

(k) $Y_3 = Y_2 + X_3$ לכן: $p(y_3|x_1,x_2,x_3,y_2) = \begin{cases}1 & if\ y_3 = y_2 + x_3\\ 0 & if\ y_3 \neq y_2 + x_3\end{cases}$

(l) נתון Y_3 לכן מתקבל: $p(y_3'|y_1,y_2,y_3) = \begin{cases}1 & if\ y_3' = y_3\\ 0 & if\ y_3' \neq y_3\end{cases}$

(m) כל המשתנים ב"ת לכן $p(x_i|x_{1:(i-1)}) = p(x_i) = p(x_i|x_{i-1})$. כלומר זו שרשרת מרקוב פשוט בגלל שהמשתנים הקודמים לא משנים כלל עבור משתנה כלשהו.

(n) נגדיר את $y_0 = 0$. נתבונן ב-$p(y_i|y_{1:(i-1)})$:

$$p(y_i|y_{1:(i-1)}) = \frac{p(y_{1:i})}{p(y_{1:(i-1)})} = \frac{p\left(\bigcap_{j=1}^{i} X_j = y_j - y_{j-1}\right)}{p\left(\bigcap_{j=1}^{i-1} X_j = y_j - y_{j-1}\right)} = \frac{p\left((X_i = y_i - y_{i-1}) \cap \left(\bigcap_{j=1}^{i-1} X_j = y_j - y_{j-1}\right)\right)}{p\left(\bigcap_{j=1}^{i-1} X_j = y_j - y_{j-1}\right)}$$

כיוון שה-X_i-ים ב"ת:

$$\frac{p\left((X_i = y_i - y_{i-1}) \cap \left(\bigcap_{j=1}^{i-1} X_j = y_j - y_{j-1}\right)\right)}{p\left(\bigcap_{j=1}^{i-1} X_j = y_j - y_{j-1}\right)} = \frac{p(X_i = y_i - y_{i-1})p\left(\bigcap_{j=1}^{i-1} X_j = y_j - y_{j-1}\right)}{p\left(\bigcap_{j=1}^{i-1} X_j = y_j - y_{j-1}\right)}$$
$$= p(X_i = y_i - y_{i-1}) = \frac{p(X_i = y_i - y_{i-1})p(y_{i-1})}{p(y_{i-1})} = \frac{p(X_i = y_i - y_{i-1} \cap y_{i-1})}{p(y_{i-1})}$$
$$= \frac{p(y_i, y_{i-1})}{p(y_{i-1})} = p(y_i | y_{i-1})$$

כנדרש.

2. יהיו $P_1, P_2 \epsilon M_{nxn}^{\mathbb{R}}$ מטריצות מעבר מרקוביות, ולכן סטוכסטיות בפרט,
נשים לב כי לכל וקטור שורה במטריצות אלה: $\sum_{i=1}^{n} P_{1_{ij}} = 1 = \sum_{i=1}^{n} P_{2_{ij}}$ נסתכל על וקטור שורה של המכפלה:
$$\sum_{i=1}^{n} (P_1 P_2)_{ji} = \sum_{i=1}^{n} (\sum_{k=1}^{n} (P_1)_{jk} (P_2)_{ki}) = \sum_{k=1}^{n} (P_1)_{jk} (\sum_{i=1}^{n} (P_2)_{ki}) = \sum_{k=1}^{n} (P_1)_{jk} \cdot 1 = 1$$
הסבר: ניתן להחליף את סדר הסכימה בסגלל שהסכום $\sum_{k=1}^{n} P_{ik}$ לא תלוי ב-i

3. השיוויון הראשון- נכון כיוון שאנו מפצלים את המאורע $X_{n+1} = j$ למאורעות זרים שמהווים חלוקה של כל המאורעות האפשריים.
השיוויון השני- נכון לפי נוסחת בייס.
השיוויון השלישי- נכון כיוון שהמשתנים הם מרקוביים.
השיוויון הרביעי- נכון כיוון שכפל הוא קומוטטיבי.
השיוויון החמישי- נכון לפי ההגדרה של הפונקציות p, p_n.
השיוויון השישי- נכון לפי הנחת האינדוקציה.
השיוויון האחרון- נכון לפי הגדרת כפל מטריצות.

4. נסמן ב-π_0 את ההסתברות ההתחלתית, ונקבל מהטענה הקודמת ש:
$$\pi_n = p(x_n) = p_n(i,j) = P^n{}_{ij} = \pi_0 P^n$$

5. *(a)* נסמן ב-π^j את האיבר ה-j בוקטור π. ניזכר שלפי הגדרת כפל מטריצות ריבועיות בתא ה-(i,j) של המטריצה $A \cdot B$ יש את ה-dot product של השורה ה-i של A והעמודה ה-j של B. נבחין שבעמודה ה-j של P יש את הוקטור $\pi^j \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$.
לכל וקטור v, $v^T \pi^j \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = \pi^j \sum_{i=1}^{N} v^i$. אזי: $(P^2)_{i,j} = \pi^j \sum_{i=1}^{N} \pi^i$ אבל הסכום $\sum_{i=1}^{N} \pi^i$ הוא 1 מהגדרת וקטור סטוכסטי לכן $(P^2)_{i,j} = \pi^j = (P)_{i,j}$, כנדרש.
(b) הוכחנו עבור $n = 2$. נניח נכונות עבור $n > 2$ כלשהו. אזי: $P^{n+1} = P^n P = PP = P$, כנדרש.
(c) $(\tilde{\pi} P)_j = \sum_{i=1}^{N} \tilde{\pi}^i (P)_{i,j} = \sum_{i=1}^{N} \tilde{\pi}^i \pi^j = \pi^j \sum_{i=1}^{N} \tilde{\pi}^i = \pi^j \cdot 1 = \pi^j$, כנדרש.
(d) לפי השאלה הקודמת: $\pi_n = \pi_0 P^n$. אבל ראינו ש-$P^n = P$ לכן $\pi_n = \pi_0 P$. ראינו ש-$\pi_0 P = \pi$ לכן $\pi_n = \pi$.

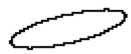
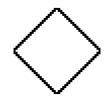
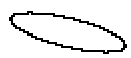
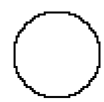
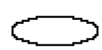
מטריצות SPD, מכפלה פנימית, נורמות ומרחקים

1. תהי A מטריצה SPD. אזי אין לה את הע"ע 0. לכן $Ker(A) = \{0\}$ לכן היא הפיכה.

2. תהי Q מטריצה SPD ושני וקטורים x, y.
קומוטטיביות:
$$< x, y >_Q = x^T Qy = (x^T Qy)^T = y^T Q^T x^{T^T} = y^T Qx =< y, x >_Q$$
משמר חיבור וקטורים:|
$$< x + y, z >_Q = (x + y)^T Qz = (x^T + y^T)Qz = x^T Qz + y^T Qz =< x, z >_Q +< y, z >_Q$$
משמר כפל בסקלר:
$$< cx, y >_Q = (cx)^T Qy = cx^T Qy = c < x, y >_Q$$
מכפלה של כל וקטור עם עצמו אי-שלילית, ורק עבור וקטור ה-0 מתקבל 0: זו ההגדרה של מטריצה PD.

3. $\|x\|_Q = \sqrt{< x, x >_Q} = \sqrt{x^T Qx} = \sqrt{x^T LL^T x} = \sqrt{(L^T x)^T (L^T x)} = \|L^T x\|_I = \|L^T x\|_{l_2}$3
כנדרש.

תרגיל בתכנות 1

<table>
<tr><td>לפי מרחק $Q2$:
</td><td>לפי מרחק l_1:
</td></tr>
<tr><td>לפי מרחק $Q3$:
</td><td>לפי מרחק l_2:
</td></tr>
<tr><td colspan="2">לפי מרחק $Q1$:
</td></tr>
</table>

וקטורים אקראיים גאוסיאנים

1. נבחין:

$$p(x)=\frac{1}{\sqrt{(2\pi)^n}}\exp\left(-\frac{1}{2}\|x\|\right)=\frac{1}{\sqrt{(2\pi)^n}}\exp\left(-\frac{1}{2}\sum_{i=1}^{n}x_i^2\right)=\frac{1}{\sqrt{(2\pi)^n}}\prod_{i=1}^{n}\exp\left(-\frac{1}{2}x_i^2\right)$$
$$=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}x_i^2\right)=\prod_{i=1}^{n}N(x_i;0,1)$$

עבור תת קבוצה J של אינדקסים מתוך $1,\dots,n$ נסמן ב-$_Jx$ את איברי הוקטור X עם אינדקסים שלא ב-J. נוכיח באינדוקציה על $|J|=k$ את התכונה הבאה: $p({}_Jx)=\prod_{\substack{i=1\\ i\notin J}}^{n}N(x_i;0,1)$.

$k=0$: כבר הוכחנו.

נניח נכונות עבור $|J|=k$ ונוכיח עבור קבוצה עם $|J|=k+1$ (כאשר $k+1<n$). נבחר $j\in J$ כלשהו. נסמן $J'=J\backslash\{j\}$. נבחין ש-$|J'|=k$. אזי:

$$p({}_Jx)=\int_{-\infty}^{\infty}p({}_{J'}x)x_j=\int_{-\infty}^{\infty}\prod_{\substack{i=1\\ i\notin J'}}^{n}N(x_i;0,1)\,x_j=\prod_{\substack{i=1\\ i\notin J}}^{n}N(x_i;0,1)\int_{-\infty}^{\infty}N(x_j;0,1)dx_j=\prod_{\substack{i=1\\ i\notin J}}^{n}N(x_i;0,1)$$

כנדרש. בהינתן x_s כלשהו, נגדיר את $J=\{1,\dots,n\}\backslash\{s\}$. אזי:

$$p(x_s)=p({}_Jx)=\prod_{\substack{i=1\\ i\notin J}}^{n}N(x_i;0,1)=N(x_s;0,1)$$

אזי כל x_s מתפלג זהה עם התפלגות נורמלית סטנדרטית. כעט נוכיח אי-תלות: עבור קבוצה כלשהי S כל אינדקסים, נגדיר את $J=\{1,\dots,n\}\backslash S$. אזי:

$$p(x_i\ s.t.\ i\in S)=p({}_Jx)=\prod_{\substack{i=1\\ i\notin J}}^{n}N(x_i;0,1)=\prod_{i\in S}N(x_i;0,1)=\prod_{i\in S}p(x_i)$$

לכן בנוסף להתפלגות זהה, המשתנים גם בלתי תלויים לפי ההגדרה, מש"ל.

2. ממשפט 4 ומכך ש-X מתפלג גאוסיאנית נקבל ש-Y מתפלג גאוסיאנית, נחשב את התוכלת והשונות של Y:

$$\mathbb{E}(Y)=\mathbb{E}(aX+b)=a\mathbb{E}(X)+b=a\mu+b$$
$$\mathbb{V}(Y)=\mathbb{V}(aX+b)=a^2\mathbb{V}(X)=a^2\sigma^2$$

ולכן נקבל ש- $Y\sim\mathcal{N}(a\mu+b,a^2\sigma^2)$

3. (a) נגדיר את A_i להיות וקטור שורה באורך n עם 1 בתא ה-i ו-0 בשאר. כמו ברמז: $X_i = A_i X$ לכל i. אזי כיוון שלמדנו שטרנספורמציה לינארית של וקטור גאוסיאני היא גם וקטור גאוסיאני:

$X_i = A_i X \sim N\left(A_i\mu, A_i \sum \ A_i^T\right) = N\left(\mu_i, (\sum\ )_{i,i}\right)$, כנדרש.

(b) עבור $S = i_1, \dots, i_k$ סדרת אינדקסים שונים מתוך $\{1, \dots, n\}$, נסמן $V_S = [V_{i_1} \quad \cdots \quad V_{i_k}]^T$ עבור וקטור כלשהו V ונגדיר את A_S להיות מטריצה $k \times n$ כך שלכל $1 \le j \le k$: $(A_S)_{j,i_j} = 1$ ושאר המטריצה אפסים. אזי $(A_S V)_j = V_{i_j} \cdot 1 + 0 \dots 0 = V_{i_j}$ כלומר $A_S V = V_S$ ובפרט $A_S X = X_S$. זו טרנספורמציה לינארית אזי כפי שלמדנו: $X_S = A_S X \sim N(A_S\mu, A_S \sum \ A_S^T)$.

כפי שהסברנו, $A_S\mu = \mu_S$. נתבונן במטריצה $A_S \sum \ A_S^T$. מתקיים:

$(A_S \sum \ A_S^T)_{j,t} = \sum_{p=1}^n (A_S)_{j,p}(\sum \ A_S^T)_{p,t} = (\sum \ A_S^T)_{i_j,t} = \sum_{q=1}^n (\sum\ )_{i_j,q}(A_S^T)_{q,t} = \sum_{q=1}^n (\sum\ )_{i_j,q}(A_S)_{t,q} = (\sum\ )_{i_j,i_t}$

אזי המטריצה $A_S \sum \ A_S^T$ היא פשוט המטריצה $\sum$ שבה משמיטים שורות ועמודות המיוחסות לאיברים שלא ב-X_S, כך שאכן ב- $\mu, \sum$ יש את כל המידע הנדרש ע"מ לדעת את ההתפלגות של X_S, כנדרש.

4. נשתמש במשפט 10, תוך כדי ההנחה ש-*X,Y* ב"ת ונקבל:

$$Z = \begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \begin{bmatrix} \sum_X & \sum_{XY} \\ \sum_{XY}^T & \sum_Y \end{bmatrix}\right) = \mathcal{N}\left(\begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \begin{bmatrix} \sum_X & 0 \\ 0 & \sum_Y \end{bmatrix}\right)$$

5. (a) המטריצה $A = [I_{n\times n} \quad I_{n\times n}]$ מקיימת את הנדרש:

$$A\begin{bmatrix} X \\ Y \end{bmatrix} = [I_{n\times n} \quad I_{n\times n}]\begin{bmatrix} X \\ Y \end{bmatrix} = IX + IY = X + Y = Z$$

(b) לפי הוכחה קודמת, הוקטור $\begin{bmatrix} X \\ Y \end{bmatrix}$ הוא גאוסיאני. A ראינו כבר שטרנספורמציה לינארית על וקטור גאוסיאני היא גם גאוסיאנית אזי

$$Z \sim N\left(A\begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, A\begin{bmatrix} \sum_X & 0 \\ 0 & \sum_Y \end{bmatrix}A^T\right) = N\left(\mu_X + \mu_Y, A\begin{bmatrix} \sum_X \\ \sum_Y \end{bmatrix}\right) = N(\mu_X + \mu_Y, \sum_X + \sum_Y)$$