CV201 HW1

Yakir Hadad ID: 313250276
Barak Bukre ID: 315453100

November 2019

1 Problem 1

Let $X = [X_1 X_2 X_3 X_4 X_5] \in S_1 \times S_2 \times S_3 \times S_4 \times S_5$ be a discrete random vector. Then

$$\begin{aligned} p(x_3, x_4, x_5) &= \sum\nolimits_{x_1, x_2} p(x) \\ &= \sum\nolimits_{x_1, x_2 \in \{-1,1\}^2} p(x) \\ &= \sum\nolimits_{x_1 \in \{-1,1\}} \sum\nolimits_{x_2 \in \{-1,1\}} p(x) \\ &= p(-1,-1,x3,x4,x5) + p(-1,1,x3,x4,x5) \\ &= p(1,-1,x3,x4,x5) + p(1,1,x3,x4,x5). \end{aligned} \tag{1}$$

2 Problem 2

Let $X \sim U\{1,6\}$ than $\forall x$, $p(x) = \dfrac{1}{6}$

$$\begin{aligned} E(X) &= \sum_{x=1}^{6} x p(x) \\ &= \frac{1}{6} \sum_{x=1}^{6} x \\ &= \frac{1}{6} \cdot \frac{6 \cdot 7}{2} \\ &= \frac{7}{2} = 3.5 \end{aligned} \tag{2}$$

3 Problem 3

$$\begin{aligned}
E(X^T) = & \\
&= \begin{bmatrix} \int_{\mathbb{R}^3} x_1 p(x)dx & \int_{\mathbb{R}^3} x_2 p(x)dx & \int_{\mathbb{R}^3} x_3 p(x)dx \end{bmatrix} \\
&= \begin{bmatrix} E(X_1) & E(X_2) & E(X_3) \end{bmatrix} \\
&= \begin{bmatrix} E(X_1) \\ E(X_2) \\ E(X_3) \end{bmatrix}^T \\
&= E(X)^T
\end{aligned} \tag{3}$$

Proof. $E(X^T) = E(X)^T$ $\square$

4 Problem 4

Let a = $\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$ and b = $\begin{bmatrix} b_1 & \dots & b_m \end{bmatrix}$

then suppose $c = a \cdot b = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \begin{bmatrix} b_1 & \dots & b_m \end{bmatrix} = \begin{bmatrix} a_1 b_1 & \dots & a_1 b_m \\ \vdots & \ddots & \vdots \\ a_n b_1 & \dots & a_n b_m \end{bmatrix}$

From the definition of matrix multiplication $c_{ij} = a_i b_j$

In our problem let $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ and $c = xx^T = \begin{bmatrix} x_1 x_1 & \dots & x_1 x_n \\ \vdots & \ddots & \vdots \\ x_n x_1 & \dots & x_n x_n \end{bmatrix}$

so $c_{ij} = x_i x_j$.

For $c^T \Rightarrow$

$$c_{ij}^T = x_j x_i = x_i x_j = c_{ij} \Rightarrow c^T = c \tag{4}$$

Proof. $xx^T = (xx^T)^T$ $\square$

5 Problem 5

Define $g(x) = xx^T \Rightarrow g_{ij}(x) = x_i x_j$ (Same as above if g(x) is c).
From definition 15 $E_{ij}(g(x)) = E(g_{ij}(x))$

$$\Rightarrow (R_x)_{ij} = E_{ij}(xx^T) = E_{ij}(g(x)) = E(g_{ij}(x)) = E(x_i x_j) \tag{5}$$

Proof. $R_x(i,j) = E(x_i x_j)$ $\square$

6 Problem 6

First notice that that for Y a random matrix, $E(Y^T) = E(Y)^T$

$$E(Y^T) = \begin{bmatrix} E(y_1y_1) & E(y_2y_1) & \dots & E(y_ny_1) \\ E(y_1y_2) & E(y_2y_2) & \dots & E(y_ny_2) \\ \vdots & \vdots & \ddots & \vdots \\ E(y_1y_n) & E(y_2y_n) & \dots & E(y_ny_n) \end{bmatrix} = E(Y)^T$$

From problem 4 $\Rightarrow XX^T = (XX^T)^T$.
Plus XX^T is a random matrix so it have the property from above.

$$R_X = E(XX^T) = E((XX^T)^T) = E(XX^T)^T = R_X^T \tag{6}$$

Proof. $R_X = R_X^T$ □

7 Problem 7

Let X be RV. First let define $Y = X - \mu$ and $g(Y) = YY^T$, more specific:

$$\begin{bmatrix} g_{11}(y) & g_{12}(y) & \dots & g_{1n}(y) \\ g_{21}(y) & g_{22}(y) & \dots & g_{2n}(y) \\ \vdots & \vdots & \ddots & \vdots \\ g_{n1}(y) & g_{n2}(y) & \dots & g_{nn}(y) \end{bmatrix}$$

$$=$$

$$\begin{bmatrix} (x_1-\mu_1)(x_1-\mu_1) & (x_1-\mu_1)(x_2-\mu_2) & \dots & (x_1-\mu_1)(x_n-\mu_n) \\ (x_2-\mu_2)(x_1-\mu_1) & (x_2-\mu_2)(x_2-\mu_2) & \dots & (x_2-\mu_2)(x_n-\mu_n) \\ \vdots & \vdots & \ddots & \vdots \\ (x_n-\mu_n)(x_1-\mu_1) & (x_n-\mu_n)(x_2-\mu_2) & \dots & (x_n-\mu_n)(x_n-\mu_n) \end{bmatrix}$$

From definition 14. Mean of matrix-valued function of continuous RV is the matrix of means of functions. So:

$$E(g(Y)) = \begin{bmatrix} E(g_{11}(y)) & E(g_{12}(y)) & \dots & E(g_{1n}(y)) \\ E(g_{21}(y)) & E(g_{22}(y)) & \dots & E(g_{2n}(y)) \\ \vdots & \vdots & \ddots & \vdots \\ E(g_{n1}(y)) & E(g_{n2}(y)) & \dots & E(g_{nn}(y)) \end{bmatrix}$$

$$= \begin{bmatrix} E((x_1-\mu_1)(x_1-\mu_1) & E((x_1-\mu_1)(x_2-\mu_2)) & \dots & E((x_1-\mu_1)(x_n-\mu_n)) \\ E((x_2-\mu_2)(x_1-\mu_1)) & E((x_2-\mu_2)(x_2-\mu_2)) & \dots & E((x_2-\mu_2)(x_n-\mu_n)) \\ \vdots & \vdots & \ddots & \vdots \\ E((x_n-\mu_n)(x_1-\mu_1)) & E((x_n-\mu_n)(x_2-\mu_2)) & \dots & E((x_n-\mu_n)(x_n-\mu_n)) \end{bmatrix} \tag{7}$$

Proof.

$$\Sigma_X = E((X-\mu)(X-\mu)^T) = E(g(Y))$$
$$(\Sigma_X)_{ij} = E_{ij}(g(Y)) = E((x_i-\mu_i)(x_j-\mu_j))$$

□

8 Problem 8

From problem 7 we can see that the diagonal of Σ_X is $\forall i, E((x_i - \mu)(x_i - \mu))$ which is exactly the definition of variance of X_i

Proof.

$$(\Sigma_X)_{ii} = VAR(X_i)$$

□

9 Problem 9

From problem 7

$\Sigma_X =$

$$\begin{bmatrix} E((x_1-\mu_1)(x_1-\mu_1) & E((x_1-\mu_1)(x_2-\mu_2)) & \dots & E((x_1-\mu_1)(x_n-\mu_n)) \\ E((x_2-\mu_2)(x_1-\mu_1)) & E((x_2-\mu_2)(x_2-\mu_2)) & \dots & E((x_2-\mu_2)(x_n-\mu_n)) \\ \vdots & \vdots & \ddots & \vdots \\ E((x_n-\mu_n)(x_1-\mu_1)) & E((x_n-\mu_n)(x_2-\mu_2)) & \dots & E((x_n-\mu_n)(x_n-\mu_n)) \end{bmatrix}$$

So to prove that $\Sigma_X = \Sigma_X^T$ we need to prove that $\forall i, j\ (\Sigma_X)_{ij} = (\Sigma_X)_{ji}$

$$\begin{aligned} (\Sigma_X)_{ij} &= E((x_i - \mu_i)(x_j - \mu_j)) \\ &= \iint_{\mathbb{R}} (x_i - \mu_i)(x_j - \mu_j) * p_{x_i x_j}(x_i, x_j) dx_i dx_j \\ &= \iint_{\mathbb{R}} (x_j - \mu_j)(x_i - \mu_i) * p_{x_i x_j}(x_i, x_j) dx_i dx_j \\ &= (\Sigma_X)_{ji} \end{aligned} \tag{8}$$

Proof.

$$(\Sigma_X)_{ij} = (\Sigma_X)_{ji}$$

□

10 Problem 10

$$\begin{aligned} \Sigma &= E((X-\mu)(X-\mu)^T) \\ &= E((X-\mu)(X^T-\mu^T)) \\ &= E(XX^T - X\mu^T - \mu X^T + \mu\mu^T) \\ &= E(XX^T) - E(X\mu^T) - E(\mu X^T) + E(\mu\mu^T) \\ &= E(XX^T) - E(X)\mu^T - \mu E(X^T) + \mu\mu^T \\ &= \{E(X) = \mu\} \\ &= E(XX^T) - \mu\mu^T - \mu\mu^T + \mu\mu^T \end{aligned} \tag{9}$$

Proof. $\Sigma = E(XX^T) - \mu\mu^T$ □

Using the previous proof

$$\begin{aligned}\Sigma &= E(XX^T) - \mu\mu^T \\ &= E(XX^T) - \mu E(X^T) \\ &= E(XX^T) - E(\mu X^T) \\ &= E(XX^T - \mu X^T)\end{aligned} \tag{10}$$

Proof. $\Sigma = E((X-\mu)X^T)$ □

Again using the first proof

$$\begin{aligned}\Sigma &= E(XX^T) - \mu\mu^T \\ &= E(XX^T) - E(X)\mu^T \\ &= E(XX^T) - E(X\mu^T) \\ &= E(XX^T - X\mu^T) \\ &= E(X(X^T - \mu^T))\end{aligned} \tag{11}$$

Proof. $\Sigma = E(X(X-\mu)^T)$ □

11 Problem 11

$$\begin{aligned}\Sigma_{YX} &= E((Y-E(Y))(X-E(X))^T) \\ &= E(YX^T - YE(X)^T - E(Y)X^T + E(Y)E(X)^T) \\ &= E(YX^T) - E(Y)E(X)^T - E(Y)E(X)^T + E(Y)E(X)^T \\ &= E(YX^T) - E(Y)E(X)^T\end{aligned} \tag{12}$$

Until now we didn't use any special things that we didn't use before. Using the property of the transpose of multiple matrices.

$$(AB)^T = B^T A^T \Rightarrow AB = (B^T A^T)^T$$

Continue from above

$$\begin{aligned}\Sigma_{YX} &= E(YX^T) - E(Y)E(X)^T \\ &= E(XY^T)^T - (E(X)E(Y)^T)^T \\ &= E(XY^T)^T - 2(E(X)E(Y)^T)^T + (E(X)E(Y)^T)^T \\ &= (E(XY^T - XE(Y)^T - E(X)Y^T + E(X)E(Y)^T)^T \\ &= E((X-E(X))(Y-E(Y))^T)^T \\ &= \Sigma_{XY}^T\end{aligned} \tag{13}$$

Proof.

$$\Sigma_{YX} = \Sigma_{XY}^T$$

□

12 Problem 12

A and b are constants, so $E(AX) = AE(X) = A\mu_x$ and $E(b) = b$

Proof.

$$\mu_Y = E(Y) = E(A\mu_X + b) = E(AX) + E(b) = A\mu_X + b$$

□

Using the proof above about μ_Y

$$\begin{aligned}
\Sigma_Y &= E((Y-\mu_Y)(Y-\mu_Y)^T) \\
&= E((AX+b-A\mu_X-b)(AX+b-A\mu_X-b)^T) \\
&= E((AX-A\mu_X)(AX-A\mu_X)^T) \\
&= E((A(X-\mu_X))(A(X-\mu_X))^T) \\
&= AE((X-\mu_X)(A(X-\mu_X))^T) \\
&= AE((X-\mu_X)(X-\mu_X)^T A^T) \\
&= AE((X-\mu_X)(X-\mu_X)^T)A^T \\
&= A\Sigma_X A^T
\end{aligned} \tag{14}$$

Proof.

$$\Sigma_Y = A\Sigma_X A^T$$

□

13 Problem 13

Notice that because Σ_X is diagonal matrix thats mean that for every 2 different variables that covariance is 0. and the variance for every X_i is σ

$$Cov(X_i, X_j) = 0, \forall i \neq j$$

$$Cov(X_i, X_i) = \sigma$$

And from probability if the covariance between 2 variables are 0 that's mean they independent and the expectation of the sum is $E(X+Y) = E(X) + E(Y)$ So now we can calculate the variance

$$Var(1^T X) = E((\sum_{i=1}^{n} X_i - E(\sum_{i=1}^{n} X_i))^2)$$

Using the independence behavior I mentioned before

$$E((\sum_{i=1}^{n} X_i - E(\sum_{i=1}^{n} X_i))^2) = E((\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} E(X_i))^2) = E((\sum_{i=1}^{n}(X_i - E(X_i)))^2)$$

Ill use the formula of square of sum in to keep calculate

$$\sum_{i=1}^{n} a_i = \sum_{i=1}^{n} a_i^2 + 2\sum_{i=1}^{n}\sum_{j=1}^{i-1} a_i a_j$$

$$\begin{aligned} E((\sum_{i=1}^{n}(X_i - E(X_i)))^2) &= E(\sum_{i=1}^{n}(X_i - E(X_i))^2 + 2\sum_{i=1}^{n}\sum_{j=1}^{i-1}(X_i - E(X_i))(X_j - E(X_j))) \\ &= \sum_{i=1}^{n} E((X_i - E(X_i))^2) + 2\sum_{i=1}^{n}\sum_{j=1}^{i-1} E((X_i - E(X_i))(X_j - E(X_j))) \\ &= \sum_{i=1}^{n} Var(X_i) + 2\sum_{i=1}^{n}\sum_{j=1}^{i-1} Cov(X_i, X_j) \\ &= \sum_{i=1}^{n} \sigma_i \end{aligned} \tag{15}$$

From problem 8 the variance of Random Variable i in RV X is at the covariance matrix $(\Sigma_X)_{ii}$, given that $\sigma_i = \sigma$ so :

Proof. $Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \sigma = n\sigma$ □

14 Problem 14

Let $X = [x_1, x_2 \dots x_n]^T \sim \mathcal{N}(\mu, \Sigma) \in \mathbb{R}^{n\times 1}$ and $A_i = [0 \dots 0, 1, 0 \dots 0] \in \mathbb{R}^{1\times n}$ vector of zeros and 1 in i^{th} place.
Let $G_i = A_i X \in \mathbb{R}$
From fact 9 affine transformation of Gaussian random vector is also random vector, so G_i is also random "vector" (1×1 dimensions) with expectation of $E(G_i) = A_i\mu$ and $cov(G_i) = A_i \Sigma A_i^T$. From matrix multiplication:

- $X_i = A_i X$
- $E(X_i) = \mu_i$
- $Var(X_i) = \Sigma_{ii}$

Proof.

$$G_i = X_i \sim \mathcal{N}(\mu, \Sigma)$$

□

Without limiting generality, everything we did can be done the same for Matrix A of zeros and ones that represents subgroup of X.

For example $A' = \begin{bmatrix} 0,1,0,0\dots,0\dots \\ 0,0,0,1\dots,0\dots \\ \vdots \\ 0,0,\dots,1_{2i},0\dots \end{bmatrix} \in \mathbb{R}^{\frac{n}{2}\times n}$ matrix with m rows, every row i is vector of 0's and 1 on the $2i$ cell. Again using fact 9 and matrix multiplication we will get that:

- $X' = A'X \in \mathbb{R}^k$
- $E(X') = [\mu_2, \mu_4, \dots]^T \in \mathbb{R}^{\frac{n}{2}}$
- $\Sigma_{X'} = \begin{bmatrix} \Sigma_{2,2} & 0 & \dots \\ 0 & \Sigma_{4,4} & \dots \\ \vdots & \vdots & \ddots \end{bmatrix} \in \mathbb{R}^{\frac{n}{2}\times\frac{n}{2}}$
- $X' \sim \mathcal{N}(E(X'), \Sigma_{X'})$

again this will work for every X' made of matrix A' of rows of 0's and 1 in one cell because of the property of fact 9 and matrix multiplication.

15 Problem 15

$$\mathbb{1}_A(w) = \begin{cases} 1 & w \in A \\ 0 & w \notin A \end{cases}$$

By definition of Indicators we get

Proof.

$$P(\mathbb{1} = 1) = P(w \in A) = P(A)$$

□

16 Problem 16

From definition 26 if X and Y are orthognal random vectors than:

$$E(XY^T) = 0_{n\times m}$$

I just want to make clear that $E(YX^T) = 0_{m\times n}$ too:

$$0_{n\times m} = E(XY^T) = E((XY^T)^T)^T = E(YX^T)^T$$
$$E(YX^T) = 0_{m\times n}$$

The correlation matrix of Z:

$$
\begin{aligned}
R_Z &= E(ZZ^T) = E([X^T\, Y^T]^T [X^T\, Y^T]) \\
&= \begin{bmatrix} E(XX^T) & E(XY^T) \\ E(YX^T) & E(YY^T) \end{bmatrix} = \begin{bmatrix} R_X & 0_{n\times m} \\ 0_{m\times n} & R_Y \end{bmatrix}
\end{aligned} \tag{16}
$$

Same as for the covariance matrix and using fact 27 that

$$\Sigma_{XY} = E(XY^T) - E(X)E(Y)^T = 0$$

:

$$
\begin{aligned}
\Sigma_Z &= E((Z - E(Z^T))(Z - E(Z^T))^T) \\
&= E(([X^T\, Y^T]^T - [E(X)^T\, E(Y)^T]^T)([X^T\, Y^T]^T - [E(X)^T\, E(Y)^T])^T \\
&= E([(X - E(X))^T\, (Y - E(Y))^T]^T\, [(X - E(X))^T\, (Y - E(Y))^T]) \\
&= \begin{bmatrix} E((X - E(X))(X - E(X))^T) & E((X - E(X))(Y - E(Y))^T) \\ E((Y - E(Y))(X - E(X))^T) & E((Y - E(Y))(Y - E(Y))^T) \end{bmatrix} \\
&= \begin{bmatrix} \Sigma_X & 0_{n\times m} \\ 0_{m\times n} & \Sigma_Y \end{bmatrix}
\end{aligned} \tag{17}
$$

17 Problem 17

Given X and Y RV and $X \perp\!\!\!\perp Y$, in other words $P(X, Y) = P(X)P(Y)$. If we calculate the Covariance of X and Y (Generally I calculate this for continues variables but the math will be the same for the discrete problems):
Uncorrelated variables if anad only if $E(XY^T) = E(X)E(Y)$

$$
\begin{aligned}
E(XY^T) &= \iint_{\mathbb{R}^n} xyp(x, y)dxdy \\
&= \iint_{\mathbb{R}^n} xyp(x)p(y)dxdy \\
&= \int_{\mathbb{R}^n} xp(x)dx \int_{\mathbb{R}^n} yp(y)dy \\
&= E(X)E(Y)
\end{aligned} \tag{18}
$$

Proof.

$$E(XY) = E(X)E(Y) \Rightarrow X \perp Y$$

$$X \perp\!\!\!\perp Y \Rightarrow X \perp Y$$

□

18 problem 18

$X_1 \backslash X_2$	0	1
0	0.5	0.1
1	0.3	0.1

Part (i)

$$\begin{aligned} p(X_1) &= p(X_1, X_2 = 0) + p(X_1, X_2 = 1) \\ p(X_2) &= p(X_1 = 0, X_2) + p(X_1 = 1, X_2) \end{aligned} \tag{19}$$

Part (ii)

$$\begin{aligned} E(X) &= E([X1X2]^T) = [E(X1)E(X2)]^T \\ &= \begin{bmatrix} 0\cdot(X_1 = 0) + 1\cdot(X_1 = 1) \\ 0\cdot(X_2 = 0) + 1\cdot(X_2 = 1) \end{bmatrix} \\ &= \begin{bmatrix} p(1,0) + p(1,1) \\ p(0,1) + p(1,1) \end{bmatrix} \\ &= \begin{bmatrix} 0.4 \\ 0.2 \end{bmatrix} \end{aligned} \tag{20}$$

Part (iii)

$$\begin{aligned} R_X &= E(XX^T) \\ &= \begin{bmatrix} E(X_1X_1) & E(X_1X_2) \\ E(X_2X_1) & E(X_2X_2) \end{bmatrix} \\ &= \begin{bmatrix} P(X_1 = 1) & P(X_1 = 1, X_2 = 1) \\ P(X_2 = 1, X_1 = 1) & P(X_2, X_2 = 1) \end{bmatrix} \\ &= \begin{bmatrix} 0.4 & 0.1 \\ 0.1 & 0.2 \end{bmatrix} \end{aligned} \tag{21}$$

Part (iv)

$$\begin{aligned} \Sigma_X &= E((X - E(X))(X - E(X))^T) \\ &= \begin{bmatrix} \Sigma_{X_1} & \Sigma_{X_1X_2} \\ \Sigma_{X_2X_1} & \Sigma_{X_2} \end{bmatrix} \end{aligned} \tag{22}$$

$$\begin{aligned} \Sigma_{X_1} &= E(X_1^2) - E(X_1)^2 = 1^2 \cdot 0.4 - 0.4^2 = 0.24 \\ \Sigma_{X_1X_2} &= E(X_1X_2) - E(X_1)E(X_2) = 1 \cdot 1 \cdot 0.1 - 0.4 \cdot 0.2 = -0.07 \\ \Sigma_{X_2X_1} &= \Sigma_{X_1X_2} = -0.07 \\ \Sigma_{X_2} &= E(X_2^2) - E(X_2)^2 = 1 \cdot 0.2 - 0.2^2 = 0.16 \end{aligned} \tag{23}$$

Proof.

$$\Sigma_X = \begin{bmatrix} 0.24 & -0.07 \\ -0.07 & 0.16 \end{bmatrix}$$

□

Part (v)
To see if $X1$ is independent to $X2$ we need to check that

$$\forall x_i, x_j \; P(X_1 = x_i, X_2 = x_j) = P(X_1 = x_i)P(X_2 = x_j)$$

Lets take for example $x_i = 1$ $x_j = 1$ than $P(1,1) = 0.1$ while $P(X_1 = 1) = 0.4$ and $P(X_2 = 1) = 0.2$ so $P(X_1 = 1)P(X_2 = 1) = 0.8$
$P(X_1 = 1, X_2 = 1) \neq P(X_1 = 1)P(X_2 = 1)$ which means that $X_1 \not\perp X_2$
Part (vi)
To check if X_1 and X_2 are correlated we need to check if

$$E(X_1X_2) \neq E(X_1)E(X_2)$$

We have the answer for this from part 3 :

$$0.1 \neq 0.4 \cdot 0.2$$

So X_1 and X_2 are correlated.