Assignment - 2

CSE - 544

Group

i) Irfan Ahmed (113166464)

ii) Sharfuddin Mohammed (112070993)

iii) Anurag Yepuri (113070893)

iv) Fatemeh Aslan Beigi (113870646)

1)

$$Cov(X,Y) = E[(X-E[X])(Y-E[Y])] = E[XY] - E[X]E[Y]$$

a)

$$E[X] = \frac{1}{4}\times 0 + \frac{1}{2}\times 1 + \frac{1}{4}\times 2 = 1$$

$$E[Y] = \frac{1}{4}\times 0 + \frac{1}{2}\times 1 + \frac{1}{4}\times 2 = 1$$

$$E(XY) = \frac{3}{8}\times 0 + \frac{2}{8}\times 1 + \frac{2}{8}\times 2 + \frac{1}{8}\times 4 = \frac{10}{8}$$

$$cov[X,Y] = \frac{10}{8} - 1 = \frac{2}{8} = \frac{1}{4} = 0.25$$

b) $Y = X^2$

$$E[X] = \frac{1}{5}(-5 + -2 + 0 + 2 + 5) = 0$$

$$E(Y) = \frac{2}{5}\times 25 + \frac{2}{5}\times 4 + \frac{1}{5}\times 0 = \frac{2}{5}(25+4) = \frac{58}{5} = 11.6$$

$E[XY] = 0 \longrightarrow$ cause X is symmetric & fair

$$cov[XY] = 0$$

c) No, As we have seen in part b $Y = X^2$ & these two RVs are not independent but their Covariance is zero cause X is a fair dice & its values are symmetric.

Problem 2

a) Given a non-negative random variable X, $E[X]$ is

$$E[X] = \int_0^\infty x\, p(X=x)\, dx = \int_0^t x\, p(X=x)\, dx + \int_t^\infty x\, p(X=x)\, dx$$

But in the interval $[0,t]$, $x \geq 0$ and $p(X=x) \geq 0 \;\forall x$.

Thus we have, $x\, p(X=x) \geq 0 \Rightarrow \int_0^t x\, p(X=x)\, dx \geq 0$.

Thus, $E[X] \geq \int_t^\infty x\, p(X=x)\, dx$.

$\Rightarrow E[X] \geq \int_t^\infty x\, p(X=x)\, dx$ where $p(X=x) = f(x)$

b) Consider ~~$P(X \geq t)$~~ $E[X]$.

We just showed that $E[X] \geq \int_t^\infty x\, p(X=x)\, dx$ for all $t \geq 0$

But in the interval t to ∞, $x \geq t$

Thus, $x\, p(X=x) \geq t\, p(X=x)$ in $[t, \infty)$

Therefore, $E[X] \geq \int_t^\infty x\, p(X=x)\, dx \geq \int_t^\infty t\, p(X=x)\, dx$

$\Rightarrow E[X] \geq \int_t^\infty t\, p(X=x)\, dx = t \int_t^\infty p(X=x)\, dx$

Thus, $E[X] \geq t \int_t^\infty p(X=x)\, dx$

But $\int_t^\infty p(X=x)$ is $P(X \geq t)$

Thus, $E[X] \geq t\, P(X \geq t) \Rightarrow P(X \geq t) \leq \frac{E[X]}{t}$

But ~~Therefore~~ we have, $P(X \geq t) \geq P(X > t)$

Thus, $P(X > t) \leq \frac{E[X]}{t}$.

Scanned with CamScanner

c) Consider the inequality $(x-\mu)^2 \geq t^2$

$\Rightarrow (x-\mu)^2 - t^2 \geq 0 \quad \Rightarrow \quad (\cancel{(x-\mu)-t}) \geq 0$

$\Rightarrow \big((x-\mu)-t\big)\big((x-\mu)+t\big) \geq 0$

Thus, $(x-\mu)-t \geq 0$ and $(x-\mu)+t \geq 0$ or — ①

$(x-\mu)-t \leq 0$ and $(x-\mu)+t \leq 0$ — ②

From eqn. ①, we have $(x-\mu) \geq t$ and $(x-\mu) \geq -t \Rightarrow (x-\mu) \geq t$

From eqn ②, we have $(x-\mu) \leq t$ and $(x-\mu) \leq -t \Rightarrow (x-\mu) \leq -t$.

We get eqns ③ and ④ because t is positive.

From eqns ③ and ④, $(x-\mu) \geq t$ and $(x-\mu) \leq -t$

Combining these equations we have, $|x-\mu| \geq t$

Thus, the values of x which satisfy $|x-\mu| \geq t$ are exactly the same as those that satisfy $(x-\mu)^2 \geq t^2$.

Consider, the Random variable $Y = (x-\mu)^2$, Here the Random variable Y is non-negative as $(x-\mu)^2 \geq 0$

From b) we have, $P(Y \geq t^2) \leq \dfrac{E[Y]}{t^2}$

But, $E[Y] = E[(x-\mu)^2] = \sigma^2$

Thus, $P((x-\mu)^2 \geq t^2) \leq \dfrac{\sigma^2}{t^2} \Rightarrow P(|x-\mu| \geq t) \leq \dfrac{\sigma^2}{t^2}$

3.) a.) $X_1, X_2, \ldots, X_K$ are independent RV's $\quad f_{X_i}(x) = \lambda_i e^{-\lambda_i x}, \; x \geq 0$

$Z = \min(X_1, X_2, \ldots, X_K)$

cdf of Z $\quad F_Z(z) = Pr(Z \leq z) = 1 - Pr(Z > z) = 1 - Pr(\min(X_1, X_2, \ldots, X_K) > z)$

since minimum is $> z$, each X_i should be greater than z.

$$= 1 - Pr(X_1 > z, X_2 > z, \ldots X_K > z) \overset{ind}{=} 1 - \prod_{i=1}^{K} Pr(X_i > z)$$

$$= 1 - \prod_{i=1}^{K} [1 - Pr(X_i \leq z)] \quad -①$$

cdf of X_i

$$F_{X_i}(x) = \int_0^{x_i} \lambda_i e^{-\lambda_i x} dx = \lambda_i \frac{(e^{-\lambda_i x})}{-\lambda_i}\Big|_0^x = e^{-\lambda_i x}\Big|_x^0 = 1 - e^{-\lambda_i x}$$

$$1 - Pr(X_i \leq x) = e^{-\lambda_i x}$$

$$\therefore ① \Rightarrow 1 - \prod_{i=1}^{K} e^{-\lambda_i x} = 1 - e^{-\sum_{i=1}^{K} \lambda_i x}$$

pdf of $Z = \frac{d}{dz}(F_Z(z)) = \underline{\sum_{i=1}^{K} \lambda_i \; e^{-\sum_{i=1}^{K} \lambda_i x}}$ $\rightarrow$ Exponential distribution with $\hat{\lambda} = \sum_{i=1}^{K} \lambda_i$

Expectation of X_i

$$E[X_i] = \int_{0}^{\infty} x \, Pr(X_i = x) dx = \int_0^\infty x \cdot \lambda_i e^{-\lambda_i x} dx = \lambda_i \left[\frac{x e^{-\lambda_i x}}{-\lambda_i}\Big|_0^\infty + \frac{1}{\lambda_i} \int_0^\infty e^{-\lambda_i x} dx \right]$$

$(x \geq 0)$

$$E[X_i] = \lambda_i \left[0 + \frac{1}{\lambda} \frac{-e^{-\lambda_i x}}{\lambda_i}\Big|_0^\infty \right] = \lambda_i \left(\frac{1}{\lambda_i^2}\right) = \frac{1}{\lambda_i}$$

$$\lim_{x \to \infty} x \cdot e^{-\lambda_i x} = \lim_{x \to \infty} \frac{x}{e^{\lambda_i x}} \text{ (Using L-Hospital's rule)} = \lim_{x \to \infty} \frac{1}{\lambda_i e^{\lambda_i x}} = 0$$

$$\text{(or) } \lim_{x \to \infty} \frac{x}{e^{+\lambda_i x}} = \lim_{x \to \infty} \frac{x}{\left(1 + \lambda_i x + \frac{(\lambda_i x)^2}{2!} + \ldots\right)} = \lim_{x \to \infty} \frac{1}{\left(\frac{1}{x} + \lambda_i + \ldots\right)} = \frac{1}{0 + \lambda_i + \infty + \ldots} = 0$$

$$\therefore E(Z) = \frac{1}{\hat{\lambda}} = \frac{1}{\sum_{i=1}^{K} \lambda_i}$$

Scanned with CamScanner

iii) $Var(Z) = E(Z^2) - E^2(Z)$

Variance of x_i

$Var(x_i) = E(x_i^2) - E^2(x_i)$

$$E(x_i^2) = \int_0^\infty x^2 \lambda_i e^{-\lambda_i x} dx = \lambda_i \left(\frac{x^2 e^{-\lambda_i x}}{-\lambda_i} \Big|_0^\infty + 2\int_0^\infty \frac{x \cdot e^{-\lambda_i x}}{+\lambda_i} dx \right)$$

$$= \frac{\lambda_i (2)}{\lambda_i} \int_0^\infty x \cdot e^{-\lambda_i x} dx = 2\left(\frac{1}{\lambda_i^2}\right)$$

↳ From E(X)

$$\therefore Var(x_i) = \frac{2}{\lambda_i^2} - \frac{1}{\lambda_i^2} = 1/\lambda_i^2$$

$$\therefore Var(Z) = \frac{1}{(\hat{\lambda})^2} = \frac{1}{\left(\sum_{i=1}^{K} \lambda_i\right)^2}$$

Problem 3

b) Given Random variables X and Y with joint pdf

$$f_{XY}(x,y) = \begin{cases} 2, & 0 \le x \le y \le 1 \\ 0, & \text{otherwise} \end{cases}$$

Now, to find pdf of $Z = XY$, we consider the cdf of ~~ex~~ $Z = XY$.

So, we have $cdf(z) = P(Z \le z) = P(XY \le z)$

The combined pdf of X, Y takes non-zero values only when $0 \le x \le y \le 1$. So, we will consider values of X, Y which satisfy this inequality.

We have,

$$0 \le x \le y \le 1 \quad - (1)$$

We need to find $Pr(Z \le z)$ such that $XY \le z$

Thus, we have,

$$Y \le \frac{z}{X} \quad - (2)$$

Now, from equations (1) & (2), $y \le \min\left(1, \frac{z}{X}\right)$

Consider, the case when $\frac{z}{X} \le 1 \Rightarrow x \ge z \quad - (3)$

$$\Rightarrow y \le \frac{z}{x} \le 1$$

We are considering values of x where $x \le y$

$$\Rightarrow x \le \frac{z}{x} \Rightarrow x^2 - z \le 0 \Rightarrow x \le \sqrt{z} \ (\text{Since } x, z \ge 0) - (4)$$

From equations (3) and (4), we have limits on X and Y, when $\frac{z}{X} \le 1$. Thus, the cdf ~~in the~~ under this condition is:

$Pr(Z \le z \mid z/x \le 1, x \le y) = Pr(z \le x \le \sqrt{z},\ x \le y \le z/x)$

$$= \int_{z}^{\sqrt{z}} \int_{x}^{z/x} 2 \, dy\, dx$$

$$= 2\int_{z}^{\sqrt{z}} \left(\frac{z}{x} - x\right) dx = 2\left(z \log x \Big|_{z}^{\sqrt{z}} - \frac{x^2}{2}\Big|_{z}^{\sqrt{z}}\right)$$

$$= 2z\left(\frac{\log z}{2} - \log z\right) - 2\left(\frac{z}{2} - \frac{z^2}{2}\right)$$

$$= -z \log z - \frac{z}{2}\times 2 + \frac{z^2}{2}\times 2 = -z\log z - z + z^2 \quad \text{— (4)}$$

Now, consider, the case when, $\frac{z}{x} \ge 1 \Rightarrow x \le z$

Then, we have, $y \le 1$

So, $Pr(Z \le z \mid z/x > 1, x \le y) = \int_{0}^{z} \int_{x}^{1} 2\, dy\, dx$

$$= 2\int_{0}^{z} (1-x)\, dx$$

$$= 2.\left(1\Big|_{0}^{z} - \frac{x^2}{2}\Big|_{0}^{z}\right)$$

$$= 2\left(z - \frac{z^2}{2}\right) = 2z - z^2 \quad \text{— (5)}$$

Thus, we have, $Pr(Z \le z \mid x \le y) = Pr(Z \le z \mid \frac{z}{X} \le 1, x \le y) +$
$Pr(Z \le z \mid \frac{z}{X} > 1, x \le y)$

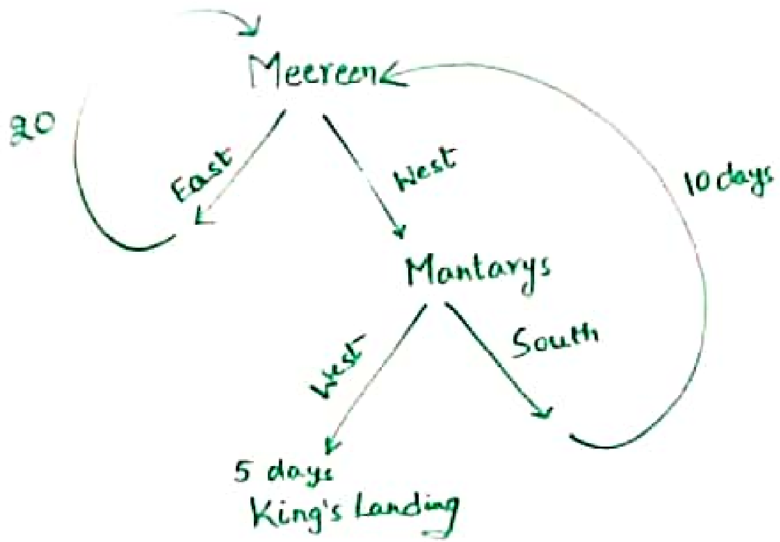
= From eqns (4) & (5),

$Pr(Z \le z \mid x \le y) = 2z - z^2 + (-z\log z - z + z^2)$

$Pr(Z \le z \mid x \le y) = -z \log z + z$ — (6)

Thus, $Pr(Z = z \mid x \le y)' = -\log z$ (Differentiating (6))

4) a)



Paths: { P1: East from Meereen.
P2: West from Meereen and West from Mantarys
P3: West from Meereen and East from Mantarys }

$E(X) = \sum_{\forall i} E(X|P_i) \Pr(P_i)$ [Law of Total Expectation]

X = no. of days to reach King's Landing

If P_1 is chosen: $X' = 20 + X$ with prob $1/2$

If P_2 is chosen $\Rightarrow X = 5$ with prob $1/2 \cdot 1/2 = 1/4$

If P_3 is chosen $\Rightarrow X' = 10 + X$ with prob $1/4$

$\therefore E(X) = E(X|P_1) \Pr(P_1) + E(X|P_2) \Pr(P_2) + E(X|P_3) \Pr(P_3)$

$E(X) = E(20+X) \cdot 1/2 + E(5) \cdot 1/4 + E(10+X) \, 1/4$

$E(X) \overset{\text{LoE}}{=} (20 + E(X)) \, 1/2 + 5 \times 1/4 + (10 + E(X)) \, 1/4$

$E(X) = 10 + \frac{E(X)}{2} + 5/4 + 10/4 + \frac{E(X)}{4}$

$\frac{E(X)}{4} = 10 + 5/4 + 10/4 \Rightarrow \boxed{E(X) = 55}$ //

Scanned with CamScanner

4) b) $Var(X) = E(X^2) - (E(X))^2$

$$E(X^2) = E(X^2|P_1)Pr(P_1) + E(X^2|P_2)Pr(P_2) + E(X^2/P_3)Pr(P_3) \text{ [Law of Total Expectation]}$$

$$= E((20+X)^2)\,\tfrac{1}{2} + E(25)\,\tfrac{1}{4} + E((10+X)^2)\,\tfrac{1}{4}$$

$$E(X^2) = E(400 + 40X + X^2)\,\tfrac{1}{2} + 25 \times \tfrac{1}{4} + E(100 + 20X + X^2)\,\tfrac{1}{4}$$

$$E(X^2) = (400 + 40E(X) + E(X^2))\,\tfrac{1}{2} + 25/4 + (100 + 20E(X) + E(X^2))\,\tfrac{1}{4}$$

$$= (400 + 40 \times 55 + E(X^2))\,\tfrac{1}{2} + 25/4 + (100 + 20 \times 55 + E(X^2))\,\tfrac{1}{4}$$

$$E(X^2) = 2600 \times \frac{1}{2} + \frac{E(X^2)}{2} + \frac{25}{4} + 1200 \times \frac{1}{4} + \frac{E(X^2)}{4}$$

$$\frac{E(X^2)}{4} = \frac{2600}{2} + 25/4 + 1200/4 \Rightarrow E(X^2) = 6425$$

$$Var(X) = E(X^2) - (E(X))^2 = 6425 - 55^2 = 3400$$

$$\boxed{Var(X) = 3400}$$

5) Given

$P_m(x_{t+1} | x_t, x_{t-1})$ is the Markovian property

Let $\Omega = \{CC, CS, SC, SS\}$

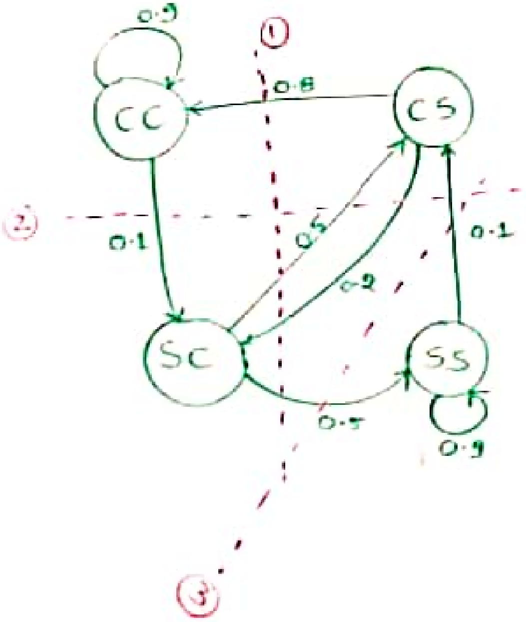
$P(c|cc) = 0.9$

$P(c|cs) = 0.8$

$P(c|sc) = 0.5$

$P(c|ss) = 0.1$

a)



Transition Probabilities

	CC	CS	SC	SS
CC	0.9	0	0.1	0
CS	0.8	0	0.2	0
SC	0	0.5	0	0.5
SS	0	0.1	0	0.9

Using Local Balance [$\pi_{cc}, \pi_{cs}, \pi_{sc}, \pi_{ss}$ are steady state probabilities]

1 $\Rightarrow$ $\pi_{cs} \times 0.8 + 0.2 \times \pi_{cs} = 0.5\pi_{sc} + 0.5 \times \pi_{sc}$

$$\boxed{\pi_{cs} = \pi_{sc}}$$

2 $\Rightarrow$ $\pi_{cc} \times 0.1 + 0.2 \times \pi_{cs} = 0.5\pi_{sc} + 0.1 \times \pi_{ss}$

3 $\Rightarrow$ $0.5\pi_{sc} = 0.1\pi_{ss}$

$$\boxed{\pi_{sc} = \frac{1}{5}\pi_{ss}}$$

Substituting 1 & 3 results in 2

$\pi_{cc} \times 0.1 + 0.2\pi_{sc} = 0.5\pi_{sc} + 0.5\pi_{sc}$

$0.1\pi_{cc} = 0.8\pi_{sc} \Rightarrow \boxed{\pi_{sc} = \frac{1}{8}\pi_{cc}}$

Also we have

Sum of all steady state prob = 1

$\pi_{cc} + \pi_{sc} + \pi_{cs} + \pi_{ss} = 1$

From 1, 2, 3

$\frac{1}{8}\pi_{cc} = \pi_{cs} = \pi_{sc} = \frac{\pi_{ss}}{5} = k$

$8k + k + k + 5k = 1$

$k = 1/15$

$\pi_{cc}, \pi_{cs}, \pi_{sc}, \pi_{ss} = [8/15, 1/15, 1/15, 1/3]$

5.) b.)

Pr(it will be snowy 3 days from today)

From Markovian property stated in problem:

$$Pr(X_{t+1} | X_t, X_{t-1}, \ldots, X_1) = Pr(X_{t+1} | X_t, X_{t-1})$$

$\therefore$ Pr(it will snow in 3 days from today) $= \pi_{SS} + \pi_{SC}$

[From steady state]

$$= 1/3 + 1/15$$

$$= 6/15 = 2/5$$

$$\boxed{Pr(\cdot) = 0.4}$$

```
Steady state: [0.53 0.07 0.07 0.33]
Probability it'll snow 3 days from today 0.4
```

6) $X = (X_1, \dots, X_K)$ is Multivariate Normal

a)

$\Rightarrow X = t_1 X_1 + t_2 X_2 + \dots + t_K X_K$ for any $t_i \in \mathbb{R}$

Considering all values of $t_i = 0$ except for $t_j \rightarrow$ This is a possible linear combination.

This would mean $X = t_j X_j$

Given if X is normal then X_j is normal since its only an affine transformation. else take $t_j = 1 \Rightarrow X = X_j$

X is normal, X_j is normal. Can be proved for any j.

b) $X = \mathcal{N}(0,1)$, $S = \{-1, 1\}$, $Y = SX$ is normal.

$Z = t_1 X_1 + t_2 Y$. let $t_1 = t_2 = 1$.

$Z = X + Y = X + SX \Rightarrow Z = \begin{cases} 2X & \text{with prob } 1/2 \\ 0 & \text{with prob } 1/2 \end{cases}$

For $z \neq 0$, $Pr(Z = z) = \frac{1}{2} Pr(Z = 2X) = \frac{1}{2} Pr(X = z/2)$

At $z = 0$, $Pr(Z = 0) = \frac{1}{2} + \frac{1}{2} Pr(2X = 0) = \frac{1}{2} + \frac{1}{2} Pr(X = 0)$ — ①

But in the neighbourhood of 0, $Pr(Z = 0^- \simeq -\delta) = \frac{1}{2} Pr(X = 0^-) = \frac{1}{2} Pr(X = 0)$

$Pr(Z = 0^+ \simeq +\delta) = \frac{1}{2} Pr(X = 0^+) = \frac{1}{2} Pr(X = 0)$ — ②

From eqns ① and ②, $Pr(Z = z)$ is not continuous at $z = 0$, thus Z is not normal c.) Z, W are $\mathcal{N}(0,1)$ i.i.d

$$X = t_1 Z + t_2 W$$

$t_1 Z$ is $\hat{Z}$ with $\mathcal{N}(0, t_1^2)$

$t_2 W$ is $\hat{W}$ with $\mathcal{N}(0, t_2^2)$

$$X = \hat{Z} + \hat{W}$$

$$F_X(x) = \Pr(X \le x) = \Pr(\hat{Z} + \hat{W} \le x) = \Pr(\hat{Z} \le x - \hat{W})$$

cdf
$$F_X(x) = \int_{-\infty}^{\infty} \int_{-\infty}^{x-\hat{w}} f_{\hat{Z},\hat{W}}(\hat{z}, \hat{w})\, dz\, dw$$

pdf
$$f_X(x) = \int_{-\infty}^{\infty} f_{\hat{Z},\hat{W}}(x - \hat{w}, \hat{w})\, dw \overset{\text{i.i.d}}{=} \int_{-\infty}^{\infty} f_Z(x-w) \cdot f_W(w)\, dw$$

$$= \int_{-\infty}^{\infty} \frac{1}{2\pi} e^{-w^2/2} \cdot e^{-(x-w)^2/2}\, dw$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-w^2/2 - x^2/2 + w^2/2 + xw}\, dw = \frac{1}{2\pi} \cdot e^{-x^2/4} \int_{-\infty}^{\infty} e^{-(w - \frac{x}{2})^2} dw$$

$\downarrow$

$\sqrt{\pi}$ (Integral of gaussian from $-\infty$ to ∞) $= 1$

$$= \frac{1}{2\pi} e^{-x^2/4} \cdot \sqrt{\pi} = \frac{1}{\sqrt{4\pi}} e^{-x^2/4}$$

$X \sim \mathcal{N}(0, 2)$

---

let $X = t_1(Z + 2W) + t_2(3Z + 5W)$

$$= (t_1 + 3t_2) Z + (2t_1 + 5t_2) W$$

X is a linear combination of Z & W which are i.i.d's

As shown above X in this case will also be normal.

d.) $X = (X_1, \dots, X_n)$, $Y = (Y_1, \dots, Y_m)$ are multivariate normals.

X is independent of Y.

$W = (X_1, \dots, X_n, Y_1, \dots Y_m)$.

$$W = \underbrace{t_1 X_1 + t_2 X_2 + \dots + t_n X_n}_{\mathcal{N}(\mu_X, \sigma_X^2)} + \underbrace{c_1 Y_1 + \dots + c_m Y_m}_{\mathcal{N}(\mu_Y, \sigma_Y^2)} \quad \forall\, t_i, c_i \in \mathbb{R}$$

$W = X + Y$

Similar to (c) part, if two variables are normally distributed & independent linear combination of them is also Normal.

$$W \sim \mathcal{N}(\underbrace{\mu_X + \mu_Y}_{\text{LOE}}, \underbrace{\sigma_X^2 + \sigma_Y^2}_{\text{sum of variance } (\perp)}).$$

e.) We need to show every part of $(X+Y)$ is uncorrelated with $(X-Y)$.

$\operatorname{cov}(X+Y, X-Y) = \operatorname{cov}(X,X) + -\operatorname{cov}(X,Y) + \operatorname{cov}(X,Y) - \operatorname{cov}(Y,Y)$ (From fact 2).

$= \operatorname{var}(X) \cancel{+ 2\operatorname{cov}(X,Y)} - \operatorname{var}(Y)$

X, Y are i.i.d standard normals.

$\Rightarrow \operatorname{Var}(X) = \operatorname{Var}(Y)$

$\therefore \operatorname{cov}(X+Y, X-Y) = 0 \Rightarrow X+Y, X-Y$ are uncorrelated.

Using Fact 1,

$X = (X_1, X_2)$, X is Multivariate Normal, $X_1 \perp X_2$

$X_1 = X+Y$, $X_2 = X-Y$; ~~$X+Y$ & $X-Y$ are independent ($\because \operatorname{cov}(\ ) = 0$)~~

Since X, Y are iid, linear combi^n of X, Y is also normal [X+Y is normal, X−Y too]

$\therefore$ X_1 is normal, X_2 is normal, $\operatorname{cov}(X_1, X_2) = 0 \Rightarrow X_1$ is $\perp$ X_2

From fact 1, $X = (X_1, X_2)$ is Multivariate Normal.

e) $(X+Y, X-Y)$

$Cov(X+Y, X-Y) = Cov(X,X) - Cov(X,Y) + Cov(X,Y) - Cov(Y,Y)$

$= Cov(X,X) - Cov(Y,Y) = 1-1 = 0$

$X,Y \longrightarrow$ are normal & independent $\longrightarrow t_1 X + t_2 Y$ is normal

for any t_1, t_2

$(X+Y, X-Y) \xrightarrow{is}$ multivarite normal $\xrightarrow{cause}$ $t_1(X+Y) + t_2(X-Y)$

$= X(t_1+t_2) + Y(t_1 - t_2) = X t_1' + Y t_2' \longrightarrow$ is normal $\xrightarrow{so}$

$(X+Y, X-Y)$ is a multivarite normal & $Cov(X+Y, X-Y)$ is 0

$\xrightarrow{so}$ $(X+Y)$ and $(X-Y)$ are independent.

7.) X is no. of days needed to capture atleast one Pokemon of all 'n' distinct types.

Let us define $Y \sim \text{Geometric}(p) \rightarrow$ number of trials to get first success.

i.e. $f_Y(y) = p(1-p)^{y-1}$

The problem of selecting pokemons as per constraints can be modelled as:

① Selecting 1^{st} unique pokemon: Since we do not have any pokemons already selected, any pokemon selected is unique. $\rightarrow$ Probability is 1.

let $Y_1 \sim \text{Geometric}\left(p = \frac{n}{n}\right)$

② Selecting 2^{nd} unique pokemon after 1^{st} has been selected:

↳ This is a geometric distribution where $p = \frac{n-1}{n}$. This is so, because we can keep on selecting 1^{st} pokemon with $1/n$ prob and success would be in selecting 2^{nd} pokemon with $p = \frac{n-1}{n}$.

$$Y_2 \sim \text{Geometric}\left(p = \frac{n-1}{n}\right)$$

Continuing the same logic for $3^{rd}, 4^{th}, \ldots$ n^{th} pokemons.

$$Y_3 \sim \text{Geometric}\left(\frac{n-2}{n}\right)$$
$$\vdots$$
$$Y_{n-1} \sim \text{Geo}\left(2/n\right)$$
$$Y_n \sim \text{Geo}\left(1/n\right)$$

Our original problem, X can be written as.

$$X = Y_1 + Y_2 + Y_3 + \ldots + Y_n$$

$$E(X) \overset{LOE}{=} E(Y_1) + E(Y_2) + E(Y_3) + \ldots E(Y_n)$$

Calculating expectation of a RV which has geometric dist (p)

$$A \sim \text{Geo}(p)$$
$$f_A(a;p) = p(1-p)^{a-1}$$

$$E(A) = \sum_{a=0}^{\infty} a \cdot p(1-p)^{a-1}$$ → This results to $1/a$

$$\therefore \quad E(X) = \frac{N}{N} + \frac{N}{N-1} + \frac{N}{N-2} + \ldots + \frac{N}{2} + \frac{N}{1}$$

$$E(X) = N \sum_{i=1}^{N} \frac{1}{i}$$ where N is distinct pokemons.

ii) $Var(X) = E(X^2) - E(X)$

~~$E(X^2) =$~~ Since all Y_i's are independent with p_i probability.

$$Var(X) \overset{\perp}{=} \sum_{i=1}^{n} Var(Y_i)$$

Y_i's are $\perp$ because they have modelled as problems having geometric dist with probabilities. Y_i is itself an event now.

$$Var(X) = Var(Y_1) + Var(Y_2) + \ldots + Var(Y_N)$$

$$Var(\text{Geometric}(p)) = \frac{1}{p^2} - \frac{1}{p}$$

$$Var(X) = 0 + \frac{N^2}{(N-1)^2} - \frac{N}{N-1} + \frac{N^2}{(N-2)^2} - \frac{N}{(N-2)} + \dots + N^2 - N$$

$$= \frac{N^2}{N^2} - \frac{N}{N} + \frac{N^2}{(N-1)^2} - \frac{N}{N-1} + \frac{N^2}{(N-2)^2} - \frac{N}{N-2} + \dots$$

$$Var(X) = N^2 \sum_{i=1}^{N} \frac{1}{i^2} - N \sum_{i=1}^{N} \frac{1}{i}$$