CSE 544 Probability and Statistics for Data Scientists
Assignment - 2

Team Members

Akhil Arradi - 114353508

Sai Vikas Balabadhrapatruni - 114777850

Varshith Adavala - 114778433

Akhila Juturu - 114777498

①

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

$X \sim Nor(\underset{\mu}{0}, \underset{\sigma^2}{1})$ $\quad Y \sim Nor(1, 2)$

Given U & V as $U = X + \alpha Y \quad V = X - Y$

~~Using Linear Transformation~~ :

$$UV = (X+\alpha Y)(X-Y)$$
$$= X^2 - XY + \alpha XY - \alpha Y^2$$
$$= X^2 - \alpha Y^2 - XY(1-\alpha)$$

$$E[U] = E[X+\alpha Y]$$
$$= E[X] + E[\alpha Y]$$
$$= E[X] + \alpha E[Y]$$

$$E[V] = E[X-Y] \Rightarrow E[X] - E[Y]$$

$$E[UV] = E[X^2 - \alpha Y^2 - XY(1-\alpha)]$$
$$= E[X^2] - E[\alpha Y^2] - E[XY(1-\alpha)]$$
$$= E[X^2] - \alpha E[Y^2] - (1-\alpha)E[XY]$$

$$Cov(U,V) = E[UV] - E[U]E[V]$$
$$= (E[X^2] - \alpha E[Y^2] + E[XY](\alpha-1)] - [E[X] + \alpha E[Y]][E[X] - E[Y]]$$
$$= E[X^2] - \alpha E[Y^2] + E[XY](\alpha-1) - [(E[X])^2 - E[X]E[Y] + \alpha E[X]E[Y] - E[Y]^2 \alpha$$
$$= E[X^2] - \alpha E[Y^2] + (\alpha-1)\cancel{E[XY]} - (E[X])^2 - \cancel{(\alpha-1)E[XY]} + (E[Y])^2 \alpha$$
$$= E[X^2] - (E[X])^2 - \alpha [E[Y^2] - (E[Y])^2$$
$$= Var(X) - \alpha Var(Y)$$
$$= 1 - 2\alpha$$

(b) Given Covariance of Independent RVs is always zero

$$\mathrm{COV}(U,V)=0$$

$$\Rightarrow 1-2\alpha=0$$

$$\therefore \boxed{\alpha=\frac{1}{2}}$$

2. a) Given a non-negative random Variable X, $E[X]$ is

$$E[X] = \int_0^\infty x f(x)\,dx$$

$t>0$

Prove $E[X] \geq \int_t^\infty x f(x)\,dx$

Consider $E[X]$ as L.H.S & $\int_t^\infty x f(x)\,dx$ as R.H.S

$$E[X]=\int_0^\infty x f(x)\,dx$$

$$=\int_0^t x f(x)\,dx+\int_t^\infty x f(x)\,dx$$

$\because t>0$, $X\geq 0$ and $f(x)\geq 0$

$$\int_0^t x f(x)\,dx \geq 0$$

$$\therefore E[X]=\int_0^t x f(x)\,dx + \text{R.H.S}$$

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

2 (b) Using part (a) Prove $P_r(X>t) \leq \frac{E[X]}{t}$

We know that $E[X] \geq \int_t^\infty x f(x)\, dx$ — ①

Divide eq① by t

$$\frac{E[X]}{t} \geq \frac{\int_t^\infty x f(x)\, dx}{t}$$

$$\frac{E[X]}{t} \geq \sum_{x=t}^{\infty} \left(\frac{x}{t}\right) f(x)\, dx$$

$$\frac{E[X]}{t} \geq f(x=t)\, k_1 + f(x=t+1)\, k_2 + f(x=t+2)(k_3 \ldots\ldots$$

Where $k_1, k_2, k_3 \ldots k_\infty$ are constants ≥ 1

So removing this Constant does not change the inequality.

$$\Rightarrow \frac{E[X]}{t} \geq f(x=t) + f(x=t+1) + f(x=t+2) \ldots\ldots$$

$$\frac{E[X]}{t} \geq P_r(X \geq t)$$

$\because P_r(X=t) \geq 0$

$$\Rightarrow \frac{E[X]}{t} \geq P_r(X>t)$$

Hence proved

2(c) Using part (b), Prove that $Pr(|X-\mu| \geq t) \leq \frac{\sigma^2}{t^2}$

We know that from (b) $Pr(X>t) \leq \frac{E[X]}{t}$

Consider X & t as some functions of X & t.

$$\therefore \ Pr(f(x) > f(t)) \leq \frac{E[f(x)]}{f(t)}$$

Let $f(x) = (X-\mu)^2$
$f(t) = t^2$

$$\Rightarrow Pr((X-\mu)^2 \geq t^2) \leq \frac{E[(X-\mu)^2]}{t^2}$$

$$\Rightarrow Pr(|X-\mu| \geq t) \leq \frac{E((X-\mu)^2)}{t^2} \quad -① \quad \left[\begin{array}{l} \text{if } a^2 > b^2 \\ |a| \geq |b| \end{array}\right]$$

Consider $E((X-\mu)^2)$

$$\left[\text{W.K.T } Var(X) = E[X^2] - (E[X])^2\right]$$

$$\sigma^2 = E((X-\mu)^2) - E[(X-\mu)]^2$$

$$\Rightarrow E((X-\mu)^2) = \sigma^2 + (E[X-\mu])^2 \quad \left[\begin{array}{l} E[X-\mu] = E[X] - E[\mu] \\ = \mu - \mu \\ = 0 \end{array}\right]$$

$$-②$$

$$\Rightarrow E[(X-\mu)^2] = \sigma^2$$

~~From~~ Substitute ② in ①

$$\therefore Pr(|X-\mu| \geq t) \leq \frac{\sigma^2}{t^2}$$

Hence proved

3) a) $X_1, X_2, X_3, \ldots\ldots, X_K$ are independent RV's

$$f_{X_i}(x) = \lambda_i e^{-\lambda_i x}, \; x \geq 0 \quad ; \quad Z = \min(X_1, X_2, \ldots, X_K)$$

(i) CDF of Z

$$F_Z(z) = Pr(Z \leq z)$$
$$= 1 - Pr(Z > z)$$
$$= 1 - Pr(\min(X_1, X_2, X_3, \ldots\ldots, X_K) > z)$$

Since least value is $> z$, each X_i should be greater than z.

$$= 1 - Pr(X_1 > z, X_2 > z, \ldots\ldots X_K > z)$$
$$= 1 - \prod_{i=1}^{K} Pr(X_i > z)$$
$$= 1 - \prod_{i=1}^{K} \left[1 - Pr(X_i \leq z)\right]$$

CDF of X_i

$$F_{X_i}(x) = \int_0^{x_i} \lambda_i e^{-\lambda_i x}\, dx \Rightarrow \left.\frac{\lambda_i (e^{-\lambda_i x})}{-\lambda_i}\right|_0^{x} \Rightarrow \left. e^{-\lambda_i x}\right|_0^{x} \Rightarrow 1 - e^{-\lambda_i x}$$

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

then CDF of Z is

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

So the PDF of Z is

$$\frac{d}{dz}(F_Z(z)) \Rightarrow \sum_{i=1}^{K} \lambda_i \, e^{-\sum_{i=1}^{K} \lambda_i x}$$

(ii) Exponential distribution with $\hat{\lambda} = \sum_{i=1}^{K} \lambda_i$

$$\text{So, } E(Z) = \frac{1}{\hat{\lambda}} \Rightarrow \frac{1}{\hat{\lambda}} \Rightarrow \frac{1}{\sum_{i=1}^{K} \lambda_i}$$

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

Variance of X_i

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

$$E(X_i) = \int_0^\infty x \cdot Pr(X_i = x)\,dx \Rightarrow \int_0^\infty x \cdot \lambda_i e^{-\lambda_i x}\,dx$$

$$\Rightarrow \lambda_i \left[\frac{x \cdot e^{-\lambda_i x}}{-\lambda_i} \Big|_0^\infty + \frac{1}{\lambda_i} \int_0^\infty e^{-\lambda_i x}\,dx \right] \Rightarrow \lambda_i \left[0 + \frac{1}{\lambda} - \frac{e^{-\lambda_i x}}{\lambda_i} \Big|_0^\infty \right]$$

$$\Rightarrow \lambda_i \left(\frac{1}{\lambda_i^2} \right) = \frac{1}{\lambda_i}$$

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

$$\Rightarrow \frac{\lambda_i (2)}{\lambda_i} \int_0^\infty x \cdot e^{-\lambda_i x}\,dx \Rightarrow 2\left(\frac{1}{\lambda_i^2} \right) \text{ (from } E(X_i))$$

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

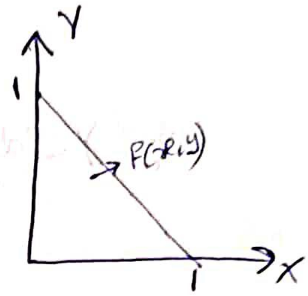
$$\therefore Var(Z) = \frac{1}{(\hat{\lambda})^2} = \frac{1}{\left(\sum_{i=1}^{k} \lambda_i \right)^2}$$

3)
(b) Given X and Y as jointly continuous random variable with PDF given by

$$f_{x,y}(x,y) = \begin{cases} cx+1 & x, y \geq 0,\ x+y<1 \\ 0 & \text{Otherwise} \end{cases}$$

i. Find the constant c, we write

From validity test



$$1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\,dx\,dy$$

$$= \int_0^1\int_0^{1-x} (cx+1)\,dy\,dx$$

$$= \int_0^1 (cx+1)(1-x)\,dx$$

$$\Rightarrow \int_0^1 cx - cx^2 + 1 - x = c\left|\frac{x^2}{2}\right|_0^1 - c\left|\frac{x^3}{3}\right|_0^1 + |x|_0^1 - \left|\frac{x^2}{2}\right|_0^1$$

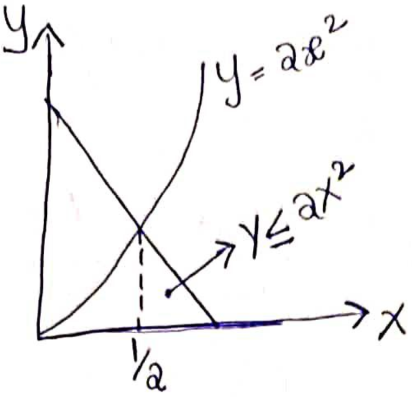
$$= \frac{c}{2} - \frac{c}{3} + 1 - \frac{1}{2} = \frac{1}{2} + \frac{1}{6}c$$

$$1 = \frac{1}{2} + \frac{1}{6}c$$

$$\Rightarrow \boxed{c = 3}$$

Thus, we conclude $c = 3$

iii) 3b $P(Y < 2X^2)$



$$P(Y < 2X^2) = \int_{-\infty}^{\infty}\int_{-\infty}^{2x^2} f_{XY}(x,y)\,dy\,dx$$

$$= \int_0^1 \int_0^{\min(2x^2,\,1-x)} (3x+1)\,dy\,dx$$

$$= \int_0^1 (3x+1)\min(2x^2, 1-x)\,dx$$

From above figure

$$= \int_0^{1/2} (3x+1)(2x^2)\,dx + \int_{1/2}^1 (3x+1)(1-x)\,dx$$

$$= \int_0^{1/2} [6x^3 + 2x^2]\,dx + \int_{1/2}^1 (3x + 1 - 3x^2 - x)\,dx$$

$$= 6\left[\frac{x^4}{4}\right]_0^{1/2} + 2\left[\frac{x^3}{3}\right]_0^{1/2} + \cancel{2}\left[\frac{x^2}{\cancel{2}}\right]_{1/2}^1 + \left[x\right]_{1/2}^1 - 3\left[\frac{x^3}{3}\right]_{1/2}^1$$

$$= 6\left[\frac{1}{4\times16}\right] + 2\left[\frac{1}{3(8)}\right] + \left[1 - \frac{1}{2} + 1 - \frac{1}{2} - 1 + \frac{1}{8}\right]$$

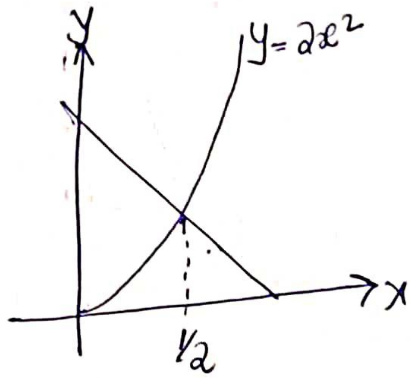
$$\Rightarrow \frac{6}{64} + \frac{2}{24} + \frac{3}{8} = \frac{18 + 16 + 72}{192}$$

$$= \frac{53}{96}$$

$$\therefore \quad P(Y < 2X^2) = \frac{53}{96}$$

3 b

(iv) Find $P(Y < 2X^2 \mid X = K)$



Let us divide it into 2 cases

$K < \frac{1}{2}$ and $K \geq \frac{1}{2}$

For $K < \frac{1}{2}$

$2K^2 < 1 - K$

$$\Rightarrow P_r(Y < 2X^2 \mid x = K) = \int_0^{2\cancel{x}^2} c(x+1)\, dy$$

W.K.T $c = 3$ and $x = K$

$$\Rightarrow \int_0^{2K^2} (3K+1)\, dy = (3K+1)\, y\Big|_0^{2K^2} = 2K^2(3K+1)$$

$$= \underline{\underline{6K^3 + 2K^2}}$$

For $K \geq 1/2$

As K value increases, the probability of $P_r(Y < 2X^2)$ also increases.

~~This~~ ~~is~~ ~~because~~

This is because

Y is between 0 & $1-K$

As K increases $1-K$ decreases

At $K = 1/2$

Y is some value between $[0, 1/2)$

For Y to be greater than or equal to $2K^2$, it should ~~be~~ atleast be $1/2$

As $2K^2 = 2\left(\frac{1}{4}\right) = 1/2$

But Y is always less than $1/2$, when $K \geq 1/2$

$\therefore$ Y is always less than $2K^2$, when $K \geq 1/2$

$$\therefore \Pr(Y < 2X^2 \mid X = K) = \begin{cases} 6K^3 + 2K^2 & K < 1/2 \\ 1 & K \geq 1/2 \end{cases}$$

3b (iii) Numerical Value from Uniform random Sampling using python is

Probability = 0.63572

$P(Y < 2X^2)$ from 3(b) ii is

$$P(Y<2X^2) = \frac{53}{96} = 0.55$$

The theoretical Value we got is the Continuous distribution function whereas in the Sampling we generate Values with a Variation called "Sampling Error". As programatically the Output is a Stimulated version of Continuous distribution

```
Probability of Y<2X^2 = 0.63503


...Program finished with exit code 0
Press ENTER to exit console.
```

3.iv. For $K < 0.5$, the values from the code and the formula are approximately equal with a difference of 0.05. This small difference might be because of randomness of the values picked in the code.

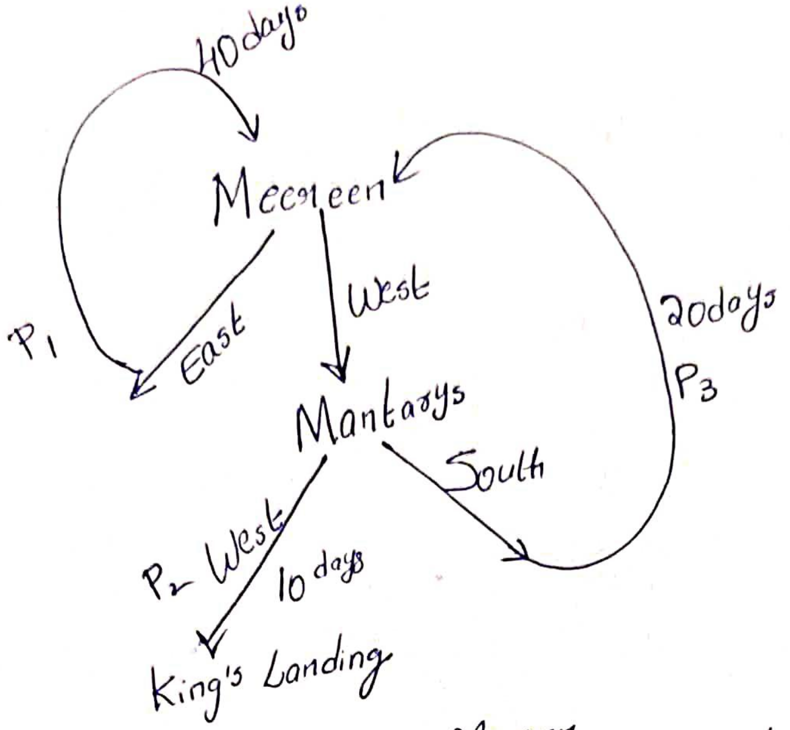
Whereas for $K > 5$, the values from the code and our observations exactly match to 1.

This is because, it is mathematically impossible for y to be $>= x^2$ for any value $x >= 0.5$. So the code and our observation exactly match to 1.

```
The probability that Y<2X^2 for the given distribution is 0.1682
Probability obtained Y<2X^2 given K=0.25 is :0.21875
The probability that Y<2X^2 for the given distribution is 1.0
Probability obtained Y<2X^2 given K=0.5 is :1
The probability that Y<2X^2 for the given distribution is 1.0
Probability obtained Y<2X^2 given K=0.75 is :1


...Program finished with exit code 0
Press ENTER to exit console.
```

4 ⓐ



Paths : $\begin{cases} P1 : \text{East from Meereen} \\ P2 : \text{West from Meeren and West from Mantarys} \\ P3 : \text{West from Meeren and East from Mantarys} \end{cases}$

W.K.T
Law of Total Expectation

$$E[X] = \sum_{\forall i} E(X|P_i)\,P_r(P_i)$$

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

If Path ① is choosen $E[X/P_1] = E[X+40]$ with Probability = $1/2$

" path ② " " $E[X|P_2] = E[10] = 10$ $\quad P_r(P_2) = \frac{1}{2}\cdot\frac{1}{2} = \frac{1}{4}$

" Path ③ is choosen $E[X|P_3] = E[X+20]$ $\quad P_r(P_3) = \frac{1}{2}\cdot\frac{1}{2} = 1/4$

$$\therefore E[X] = E[X/P_1]\,P_r(P_1) + E[X/P_2]\,P_r(P_2) + E(X|P_3)\,P_r(P_3)$$

$$= E[X+40]\left(\tfrac{1}{2}\right) + E(10)\left(\tfrac{1}{4}\right) + E[X+20]\left(\tfrac{1}{4}\right)$$

$$\overset{LOE}{=} \left(E[X] + E[40]\right)\tfrac{1}{2} + 10\left(\tfrac{1}{4}\right) + \left(E[X] + E[20]\right)\tfrac{1}{4}$$

$$E[X] = \frac{40}{2} + \frac{E[X]}{2} + \frac{10}{4} + \frac{20}{4} + \frac{E[X]}{4}$$

$$= 20 + \frac{E[X]}{2} + \frac{30}{4} + \frac{E[X]}{4}$$

$$E[X] - \frac{3}{4}E[X] = \frac{110}{4}$$

$$\frac{E[X]}{4} = \frac{110}{4}$$

$$\therefore \boxed{E[X] = 110}$$

(b) $Var[X]$

W.K.T $Var(X) = E(X^2) - (E(X))^2$

From ⓐ $E[X] = 110$ Using Law of Total Expectation

$$E[X^2] = E(X^2/P_1)Pr(P_1) + E(X^2/P_2)Pr(P_2) + E(X^2/P_3)Pr(P_3)$$

$$E[X^2] = E[(x+40)^2]\left(\frac{1}{2}\right) + E[(10)^2]\frac{1}{4} + E[(X+20)^2][1/4]$$

$$E[X^2] = E[x^2+1600+80x]\frac{1}{2} + E[100]\frac{1}{4} + E[x^2+400+40x]\frac{1}{4}$$

$$E[X^2] \overset{LOE}{=} (E[X^2]+E[1600]+E[80X])\frac{1}{2} + \frac{100}{4} + (E[X^2]+E[400]+E[40X])\frac{1}{4}$$

$$= \frac{E[X^2]}{2} + \frac{1600}{2} + \frac{80E[X]}{2} + \frac{100}{4} + \frac{E[X^2]}{4} + \frac{400}{4} + \frac{40E[X]}{4}$$

$$= \frac{E[X^2]}{2} + 800 + 40E[X] + 25 + \frac{E[X^2]}{4} + 100 + 10E[X]$$

$$E[X^2] = \frac{E[X^2]}{2} + 800 + 40E[X] + 925$$

$$E[X^2] = \frac{3}{4}E[X^2] + 50E[X] + 925$$

W.K.T $E[X] = 110$

$$E[X^2] - \frac{3}{4}E[X^2] = 50 \times 110 + 925$$

$$\frac{E[X^2]}{4} = 5500 + 925$$

$$\therefore E[X^2] = 6425 \times 4$$

$$\therefore E[X^2] = 25700$$

$$Var(X) = E[X^2] - (E[X])^2$$

$$= 25700 - (110)^2$$

$$= 25700 - 12100$$

$$= 13600$$

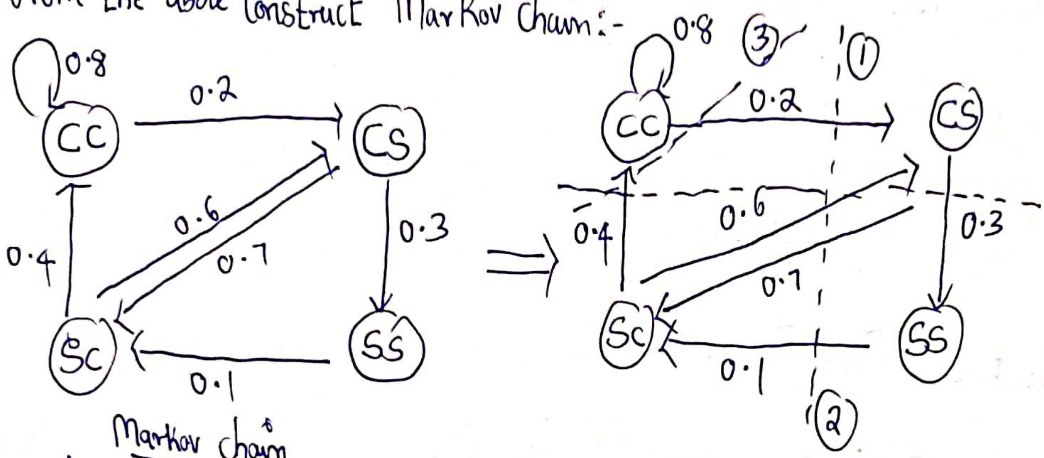
$$\boxed{\therefore Var(X) = 13600}$$

5.) Given,

$Pr(X_{i+1} \mid X_i, X_{i-1})$ as a Probability which does not follow Markovian Property.

$Pr[c|cc] = 0.8$; $Pr[c|cs] = 0.7$; $Pr[c|sc] = 0.4$; $Pr[c|ss] = 0.1$

(a). From the above Construct Markov chain:-



Markov chain

In-order to get the Steady-state Probabilities apply local balance to above Markov chain

Transition Matrix:-

	CC	CS	SC	SS
CC	0.8	0.2 ~~0.8~~	0	0
CS	0	0	0.7	0.3
SC	0.4	0.6	0	0
SS	0	0	0.1	0.9

Using local balancing $[\pi_{cc}, \pi_{cs}, \pi_{sc}, \pi_{ss}]$ are Steady State probabilities

From fig above:-

① $\pi_{cc}(0.2) + \pi_{sc}(0.6) = \pi_{cs}(0.7) + \pi_{ss}(0.1)$ —③

② $\pi_{sc}(0.4) + \pi_{sc}(0.6) = \pi_{cs}(0.7) + \pi_{cs}(0.3) \Rightarrow \boxed{\pi_{sc} = \pi_{cs}}$ —①

③ $\pi_{cc}(0.2) = \pi_{sc}(0.4) \Rightarrow \boxed{\pi_{cc} = 2\pi_{sc}}$ —②

Sub (1) & (2) in (3) Eq

Eq (3) ⇒ $\pi_{cc}(0.2) + \pi_{sc}(0.6) = \pi_{cs}(0.7) + \pi_{ss}(0.1)$

from Eq (1) ⇒ $\pi_{cc}(0.2) + \pi_{sc}(0.6) = \pi_{sc}(0.7) + \pi_{ss}(0.1)$

$\Rightarrow \pi_{cc}(0.2) - \pi_{ss}(0.1) = \pi_{sc}(0.1)$

$\Rightarrow 2\pi_{cc} - \pi_{ss} = \pi_{sc}$

from Eq (3) ⇒ $4\pi_{sc} - \pi_{sc} = \pi_{ss}$

$\Rightarrow \boxed{3\pi_{sc} = \pi_{ss}} \rightarrow (4)$

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

from (1) (2) (3) (4) ⇒ $2\pi_{sc} + \pi_{sc} + \pi_{sc} + 3\pi_{sc} = 1$

$7\pi_{sc} = 1 \Rightarrow$ $\boxed{\begin{aligned}\pi_{sc} &= 1/7\\ \pi_{ss} &= 3/7\\ \pi_{cs} &= 1/7\\ \pi_{cc} &= 2/7\end{aligned}}$ Sub in (1) (2) (3) (4)

$\therefore \pi_{cc}, \pi_{cs}, \pi_{sc}, \pi_{ss} = [2/7, 1/7, 1/7, 3/7]$

5.) b.) Pr (It will be Snowy 3 days from today)

From Property given in the problem

$Pr(X_{t+1} \mid X_t, X_{t-1})$ given the Steady State Probabilities Computed above

$\therefore$ In after Steady state Pr (snow 3 days from today) = $\pi_{ss} + \pi_{sc} = \frac{3}{7} + \frac{1}{7} = \frac{4}{7} = 0.57142$

```
Steady State: Power Iteration >> [0.2857, 0.1429, 0.1429 , 0.4286]
Probability that it will be snowy 3 days from today : 0.5714
```

⑥ Given the amounts which go into bottles of ketchups are Normally distributed with mean ~~3~~6oz & standard deviation 0.1 oz

Variance = σ^2

$$X \sim Nor(\overset{\mu}{36}, (\overset{\sigma}{0.1})^2)$$

The process is out of Control if $P(X < 35.8)$ or $P(X > 36.2)$

ⓐ Find probability that a bottle will be declared out of Control

$$P(X<35.8) + P(X>36.2)$$

Consider $P(X < 35.8)$ Given $X \sim Nor(36, (0.1)^2)$

Use standard Normal $Z \sim Nor(0, 1)$

$$\Phi(\alpha) = P_r(Z \le \alpha) = F_Z(\alpha)$$

Let ~~ax+~~ $ax+b = z$

$$Nor(a.36+b, a^2(0.1)^2) = Nor(0,1)$$

$$36a+b=0$$

$$a = -b/36$$

$$b = -360$$

$$a^2 = \frac{1}{(0.1)^2}$$

$$a = \frac{1}{0.1} = 10$$

$$10X - 360 = z$$

$$X = \frac{Z+360}{10}$$

$$X = \frac{Z}{10} + 36$$

$$F_X(35.8) = P_r(X \le 35.8)$$

$$\Rightarrow Pr\left(\frac{Z}{10}+36 \leq 35.8\right)$$

$$Pr(Z < -2)$$

$$\Rightarrow \Phi(-2) \underset{\text{From Table look up}}{=} 0.02275$$

Given the distribution is normal, it is symmetric

$$\therefore Pr(Z > \bar{X}+k\mu) = Pr(Z < X-k\mu)$$

Substitute $k=2$

$$Pr(Z > 36.2) = Pr(Z < 35.8)$$

$$Pr(Z > 36.2) + Pr(Z < 35.8)$$

$$\Rightarrow 2(Pr(Z<35.8))$$

$$= 2(0.02275) = \underline{\underline{0.0455}}$$

The probability that process is Out of Control is 0.0455

(b) Probability that the no. of bottles found out of control in an eight-hour day (16 inspections)

Probability that a Valid bottle picked = (1 - Probability that bottle is ~~knotted~~ out of control)

$$= 1 - 0.0455$$

$$= 0.9545$$

For 16 straight bottles to be Valid is

$$= (1-0.0455)^{16}$$

$$= (0.9545)^{16} = 0.4747$$

(c) There are 16 possibilities ~~Out~~ of in which ~~one~~ bottle One bottle can be out of Control among the 16 selected bottles

$$\text{Probability} = {}^{16}C_1 \left(\text{Probability that 15 are Valid}\right)\left(\text{Probability that 1 is out of control}\right)$$

$$= 16\,(0.9545)^{15}\,(0.0455)$$

$$= 16 \times 0.4973 \times 0.0455$$

$$= 0.362$$

(d) Given $Z \sim Nor(37, (0.4)^2)$

The process is out of Control if $P(X < 35.8)$ or $P(X > 36.2)$

Probability that a bottle will be declared out of Control

$$P(Z < 35.8) + P(Z > 36.2)$$

Consider $P(Z < 35.8)$

Use standard Normal $Z \sim Nor(0,1)$

$$\Phi(\alpha) = P_r(Z \le \alpha) = F_Z(\alpha)$$

$$\text{Let } aX + b = Z$$

$$Nor(a \cdot 37 + b,\; a^2 (0.4)^2) = Nor(0,1)$$

$$37a + b = 0$$

$a = \frac{-b}{37}$

$a^2 = \frac{1}{(0.4)^2}$

$a = \frac{1}{0.4} = 2.5$

$b = -37 \times \frac{1}{0.4}$

$b = -92.5$

$2.5X - 92.5 = Z$

$X = \frac{Z + 92.5}{2.5} = \frac{Z}{2.5} + 37 = 0.4Z + 37$

$F_X(35.8) = P_r(X \leq 35.8)$

$= P_r(0.4Z + 37 < 35.8)$

$= P_r(0.4Z < -1.2)$

$\Rightarrow P_r(Z < -3)$

$\Phi(-3) = 0.00135$

Consider $P(Z > 36.2)$ $\quad Z \sim (0,1)$

Use standard Normal

$F_X(36.2) = P_r(0.4Z + 37 > 36.2)$

$= P_r(0.4Z > 36.2 - 37)$

$= P_r(Z > -2) = 1 - P_r(Z \leq -2)$

$\Rightarrow 1 - \Phi(-2)$

$= \left[\Phi(-2) = 0.02275\right]$

$\Rightarrow 1 - 0.02275 \Rightarrow 0.97725$

$\therefore P_r(X < 35.8) + P_r(X > 36.2) = 0.00135 + 0.97725$

$= 0.00135 + 0.97725$

$\Rightarrow \underline{0.9786}$

7. Given that there are only n distinct types of Pokemons

(a) Let X is no. of days to capture atleast one Pokemon of all n distinct types.

Let us define $Y \sim Geometric(P)$

i.e $f_Y(y) = P(1-P)^{y-1}$ $\Rightarrow$ P_r(y flips to get first success)

The problem of selecting Pokemons can be modelled as

① Selecting 1^{st} Unique Pokemon:

Since we donot have any Pokemons already selected, any pokemon selected will be unique $\rightarrow$ Probability is 1.

Let $Y_1 \sim Geometric\left(P = \frac{n}{n}\right)$

② Selecting 2^{nd} unique Pokemon after 1^{st} has been selected.

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

This is because we can keep on selecting 1^{st} Pokemon with $1/n$ prob and success will be in selecting 2^{nd} pokemon with $P = \frac{n-1}{n}$

Continuing same logic for 3^{rd}, 4^{th}, ... n^{th} pokemon

$Y_3 \sim Geometric\left(\frac{n-2}{n}\right)$

$\vdots$

$Y_n \sim Geometric\left(1/n\right)$

We model this problem as a sum of Geometric random Variables which are independent.

X can be written as

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

$$E(X) = E(Y_1 + \cdots Y_n)$$

$$E(X) \overset{LOE}{=} E(Y_1) + E(Y_2) + E(Y_3) + \cdots E(Y_n) \quad -①$$

Calculating ~~RV~~ $E[X]$ of a RV which has geometric dist (P)

W.K.T Exepectation of Goemetric distribution is $\frac{1}{P}$

Formula $A \sim Ged(P)$

$$\left[E[A) = \sum_{K=1}^{\infty} (1-P)^{k-1} P = \frac{1}{P} \right]$$

For equ ①

$$\because E[X] = \frac{n}{n} + \frac{n}{n-1} + \frac{n}{n-2} + \cdots + \frac{n}{2} + \frac{n}{1}$$

$$E[x] = n\left(1 + \frac{1}{2} + \frac{1}{3} + \cdots \frac{1}{n-2} + \frac{1}{n-1} + \frac{1}{n}\right)$$

$$E[X] = n \sum_{i=1}^{n} \frac{1}{i}$$ where n is distinct pokemons.

$$\left[E[x] = n \sum_{i=1}^{n} \frac{1}{i} \right]$$

This is a not a closed form solution. This can be Considered as $n*$ sum of n terms in Harmonic mean.
But to Consider an approx. Solution

$$n \sum_{i=1}^{n} \frac{1}{i} \approx n \int_1^n \frac{1}{i}\, di \Rightarrow n\left[\ln i\right]_1^n$$

$$= n\ln(n) - 0$$

$$\simeq n\ln(n)$$

7 (b) $Var[X]$

(b) WKT $Var[X] = E[X^2] - (E[X])^2$

Variance of Geometric distribution is

$$\left[Var[X] = \frac{1-P}{P^2} \right] \rightarrow (1)$$

Since all Y_i's are independent with P_i probability

$$Var[X] = \sum_{i=1}^{n} Var[Y_i]$$

Y_i's are independent because they have modelled as problems having geometric distribution with probabilities.

~~Using LOE~~

$$Var[X] = Var(Y_1) + Var(Y_2) + \dots + Var(Y_n)$$

Using (1)

$$Var[X] = \sum_{i=1}^{n} Var[Y_i] = \sum_{i=1}^{n} \frac{1-P_i}{(P_i)^2}$$

$$= \sum_{i=1}^{n} \left(\frac{1}{(P_i)^2} - \frac{1}{P_i} \right) = \sum_{i=1}^{n} \left(\frac{1}{(P_i)^2} \right) - \sum_{i=1}^{n} \left(\frac{1}{P_i} \right)$$

$$= \left(\frac{n^2}{n^2} + \frac{n^2}{(n-1)^2} + \frac{n^2}{(n-2)^2} + \dots \frac{n^2}{1} \right) - \left(\frac{n}{n} + \frac{n}{n-1} + \frac{n}{n-2} + \dots \frac{n}{1} \right)$$

$$= \sum_{i=1}^{n} \left(\frac{n}{n-i-1} \right)^2 - \sum_{i=1}^{n} \left(\frac{n}{n-\cancel{i}+1} \right)$$

$$\Rightarrow n^2 \left(1 + \frac{1}{2^2} + \frac{1}{3^2} + \dots + \frac{1}{(n-2)^2} + \frac{1}{(n-1)^2} + \frac{1}{n^2} \right) -$$

$$n \left(1 + \frac{1}{2} + \frac{1}{3} + \dots \frac{1}{n-1} + \frac{1}{n} \right)$$

$$= n^2 \sum_{i=1}^{n} \frac{1}{i^2} - n \sum_{i=1}^{n} \frac{1}{i}$$

$$\left[Var(X) = n^2 \sum_{i=1}^{n} \frac{1}{i^2} - n \sum_{i=1}^{n} \frac{1}{i} \right]$$

This is not a closed form solution

~~This can be considered as sum of n terms in~~

But, to consider an approx solution

$$n^2 \sum_{i=1}^{n} \frac{1}{i^2} - n \sum_{i=1}^{n} \frac{1}{i} \Rightarrow n^2 \int_{1}^{n} \frac{1}{i^2}\, di - n \int_{1}^{n} \frac{1}{i}\, di$$

$$\Rightarrow n^2 \left[\frac{-1}{i}\right]_1^n - n \left[\ln(i)\right]_1^n$$

$$\Rightarrow n^2 \left[\frac{-1}{n^2} + 1\right] - n \left[\ln(n)\right]$$

$$\Rightarrow -1 + n^2 - n\left[\ln(n)\right]$$

$$\simeq n^2 - n\ln(n) - 1$$