Problem 1

Calculate the expectation of U and V.

$E(U) = E(X+\alpha Y)$

$= E(X) + \alpha E(Y)$

$= 0 + \alpha \times 1$

$= \alpha$

$E(V) = E(X-Y)$

$= E(X) - E(Y)$

$= 0 - 1$

$= -1$

Compute $E(X^2)$ and

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

$E(X^2) = Var(X) + E(X)^2$

$= 1 + 0$

$= 1$

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

$E(Y^2) = Var(Y) + E(Y)^2$

$= 2 + 1$

$= 3$

Compute the value of $E(UV)$

$E(UV) = E((X+\alpha Y) \times (X-Y))$

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

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

$= 1 + (\alpha-1)E(X)E(Y) - 3\alpha$

$= 1 - 3\alpha$

a)

Covariance:

Compute the covariance of U and V

$$\begin{aligned}Cov(U,V) &= E(UV) - E(U)\times E(V)\\ &=(1-3\alpha) + \alpha\\ &=1 - 2\alpha\end{aligned}$$

b)

Condition for independence

If U and V are independent events, then it satisfies the following condition.

$E(UV) = E(U) \times E(V)$

Substitute the respective value in the above equation and simplify to compute the value of α

$(1-3\alpha) = \alpha \times (-1)$

$1 - 3\alpha = -\alpha$

$1 - 2\alpha = 0$

$\alpha = 1/2$

Problem 2

Let X be a non-negative RV with mean μ and variance σ^2, and let t > 0 be some real number.

a) Prove the following: $E[X] \geq \int_t^\infty xf(x)dx$

$$E[X] = \int_0^\infty xf(x)dx \qquad \because x \geq 0$$

$$\Rightarrow \; E[X] = \int_0^t xf(x)dx + \int_t^\infty xf(x)dx$$

$$\int_0^t xf(x)dx \geq 0 \qquad \because t > 0, X \geq 0 \; and \; f(x) \geq 0$$

$$\Rightarrow \; E[X] \geq \int_t^\infty xf(x)dx \qquad -(1)$$

b) With the help of part(a), prove the following inequality: $\Pr(X > t) \leq \frac{E[X]}{t}$

Sol^n:

$$E[X] \geq \int_t^\infty xf(x)dx$$

$$\Rightarrow \; E[X] \geq \int_t^\infty tf(x)dx$$

$$\Rightarrow \; E[X] \geq t\int_t^\infty f(x)dx = tP(X > t) \qquad -(2)$$

$$\Rightarrow \; P(X > t) \leq \frac{E[X]}{t}$$

c) Using the inequality proved in part (b), prove the following: $\Pr(|X - \mu| \geq t) \leq \frac{\sigma^2}{t^2}$

Sol^n:

$$E[X] \geq tP(X > t) \qquad - From \; (2)$$

$$\Rightarrow \; P(X > t) \leq \frac{E[X]}{t}$$

$$\Rightarrow \; P((X - \mu)^2 > t^2) \leq \frac{E[(X - \mu)^2]}{t^2}$$

$$\Rightarrow \; P((X - \mu)^2 > t^2) \leq \frac{\sigma^2}{t^2}$$

$$\Rightarrow \; P(|X - \mu| \geq t) \leq \frac{\sigma^2}{t^2}$$

Problem 3

a) $X_1, X_2, \dots, X_k$ are k independent exponential random variables with

$$f_{X_i}(x) = \lambda_i e^{-\lambda_i x},\ x \geq 0 \qquad \forall\ i \in \{1, 2, \dots, k\} \text{ and}$$

(i)

Let $Z = \min(X_1, X_2, \dots, X_k)$

$F_Z(z) = P(Z \leq z)$

$F_Z(z) = P(\min(X_1, X_2, \dots, X_k) \leq z)$

$F_Z(z) = 1 - P(\mathrm{X}_1 > z, \mathrm{X}_2 > z, \dots, \mathrm{X}_\mathrm{k} > z)$

$F_Z(z) = 1 - P(\mathrm{X}_1 > z, \mathrm{X}_2 > z, \dots, \mathrm{X}_\mathrm{k} > z)$

$F_Z(z) = 1 - \prod_{i=1}^{k} P(X_i > z)$

$F_Z(z) = 1 - \prod_{i=1}^{k} \exp(-\lambda_i z)$

$F_Z(z) = 1 - \exp(-z(\lambda_1 + \lambda_2 + \cdots + \lambda_k))$

$f_Z(z) = \dfrac{\partial F_z(z)}{\partial z}$

$f_Z(z) = (\lambda_1 + \lambda_2 + \cdots + \lambda_k) \exp(-z(\lambda_1 + \lambda_2 + \cdots + \lambda_k))$

$\therefore Z \sim \exp(\lambda_1 + \lambda_2 + \cdots + \lambda_k)$

(ii) $\because Z \sim \exp(\lambda_1 + \lambda_2 + \cdots + \lambda_k)$

$$\therefore E[Z] = \frac{1}{(\lambda_1 + \lambda_2 + \cdots + \lambda_k)}$$

(iii) $\because Z \sim \exp(\lambda_1 + \lambda_2 + \cdots + \lambda_k)$

$$\therefore Var(Z) = \frac{1}{(\lambda_1 + \lambda_2 + \cdots + \lambda_k)^2}$$

Q3.b) $\because \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\, dx\, dy = 1$

$$1 = \int_{0}^{1}\int_{0}^{1-x} cx+1 \; dy \; dx$$

$$1 = \int_{0}^{1} (cx+1)(1-x)\, dx$$

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

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

ii. Pr $(Y < 2X^2)$

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

$$= \int_{0}^{1}\int_{0}^{\min(2x^2,1-x)} (3x+1)\, dy\, dx.$$

$$= \int_{0}^{1} (3x+1) \min(2x^2, 1-x)\, dx$$

$$= \int_{0}^{1/2} 2x^2 (3x+1)\, dx + \int_{1/2}^{1} (3x+1)(1-x)\, dx$$

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

Q3 b) iii.

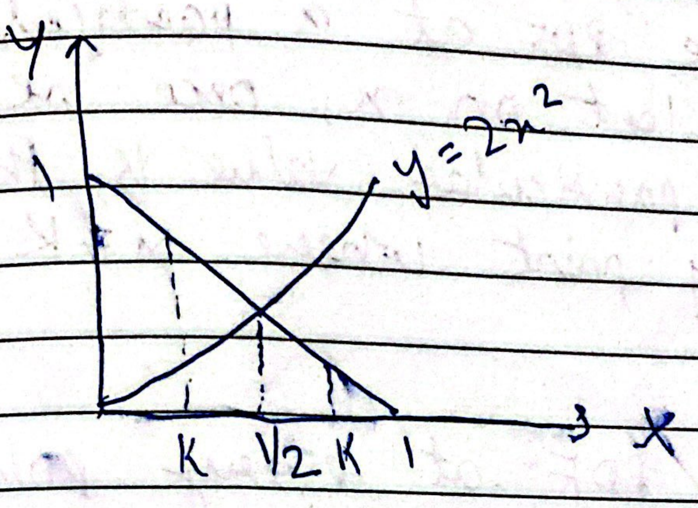
The two values do not match, since we are not accounting for the PDF while randomly sampling points to calculate $P(Y < 3x^2)$ numerically.

The random sampling assumes a uniform PDF, while in reality our PDF is dependent on x.

Scanned with CamScanner

Q3 b) (iv)

$P_r(Y < 2X^2 | X = K)$



For $K <= 1/2$;

$\min(2K^2, 1-K) = 2K^2$. (From the graph).

~~Hence for~~ $\cancel{Y \in [0, 2K^2], P(Y < 2K^2) =}$

Hence, $Y <= 2K^2$ for $Y \in [0, 2K^2]$.

and $Y > 2K^2$ for $Y \in (2K^2, 1-K)$.

$P(Y < 2X^2 | X = K) = 2K^2 / 1-K$.

For $K > 1/2$;

It is evident from the graph that $\min(2K^2, 1-K) = 1-K$. Hence, for every point in the non-zero region of the P.D.F,

$Y < 2K^2$.

$P(Y < 2X^2 | X = K) = \frac{1-K}{1-K} = 1$.

Q3b.

v. Yes, the values match.

Since the PDF at a particular point is dependent on x, once we fix x to a particular value K, the PDF for every point where $x = K$ is the same.

Since the PDF at every point is the same, random sampling is the correct way to sample points and calculate probability numerically.

Applying the formula from Q3b iv)

$P(Y < 2X^2 \mid X = 0.25) = 1/6$.

$P_n(Y < 2X^2 \mid X = 0.5) = 1$.

$P_n(Y < 2X^2 \mid X = 0.75) = 1$.

Q4.

a)

$$E(X) = E[X|East]\,P(East) + E[X|West]\,P(West)$$

$$E(X) = \frac{1}{2}E(X|East) + \frac{1}{2}E(X|West, West)\,P(West) + E(X|West, South)\,P(South) \quad \text{--- (1)}$$

$$E(X) = \frac{1}{2}E(X+40) + \frac{1}{2}\left(\frac{1}{2}\cdot 10 + \frac{1}{2}E(X+20)\right)$$

$$E(X) = \frac{E(X)}{2} + 20 + \frac{10}{4} + \frac{E(X)+20}{4}$$

$$\frac{E(X)}{4} = 27.5$$

$$\boxed{E(X) = 110}$$

b) $E(x^2) = \frac{1}{2} E(x^2 | East) + \frac{1}{2} (E(x^2 | West, West) P(West)$
$+ E(x^2 | West, South) P(South))$

$$E(x^2) = \frac{1}{2} E((x+40)^2) + \frac{1}{2}\left(\frac{1}{2} \cdot 10^2 + \frac{1}{2} E((x+20)^2)\right)$$

$$E(x^2) = \frac{1}{2}\left(E(x^2 + 80x + 1600)\right) + \frac{1}{2}\left(\frac{E(x^2 + 40x + 400)}{2} + 50\right)$$

$$\frac{E(x^2)}{4} = 40E(x) + 800 + 25 + 10E(x) + 100$$

$$\frac{E(x^2)}{4} = 50E(x) + 925$$

$$E(x^2) = 200E(x) + 3700 = 25,700$$

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

$$Var(x) = 13,600$$

Scanned with CamScanner

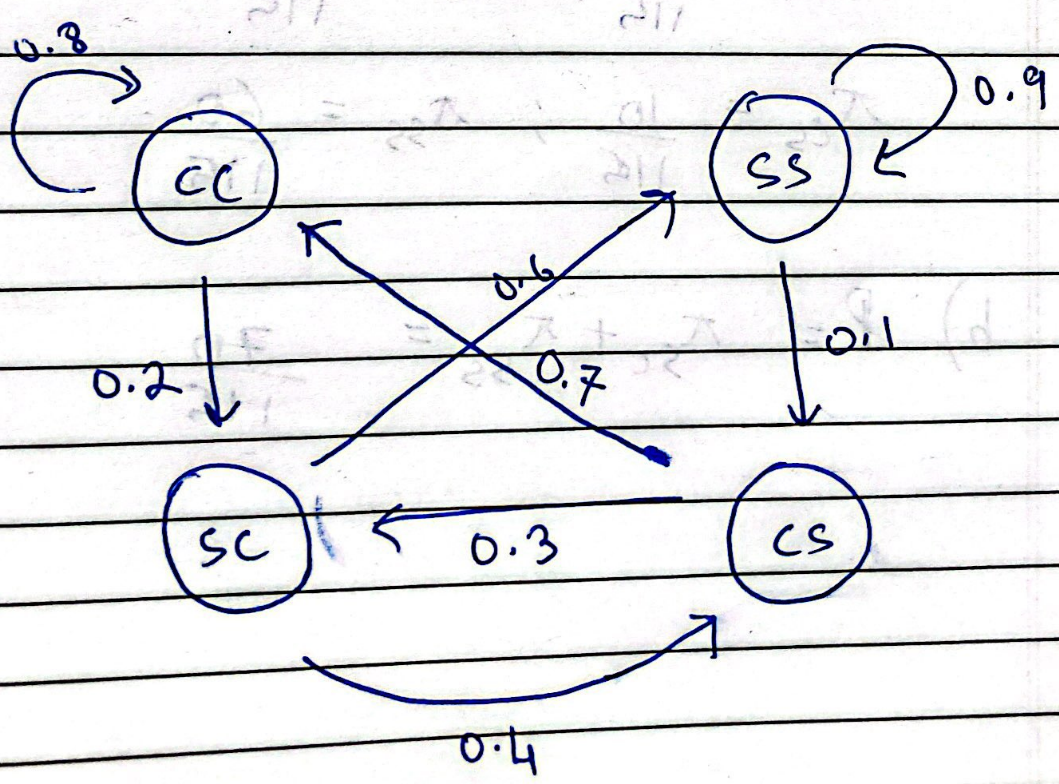
Q5.

$$\begin{array}{c|cccc} & CC & CS & SC & SS \\ \hline CC & 0.8 & 0 & 0.2 & 0 \\ CS & 0.7 & 0 & 0.3 & 0 \\ SC & 0 & 0.4 & 0 & 0.6 \\ SS & 0 & 0.1 & 0 & 0.9 \end{array}$$

Given:

$Pr(c|cc) = 0.8$, $Pr(c|cs) = 0.7$, $Pr(c|sc) = 0.4$, $Pr(c|ss) = 0.1$

$Pr(s|cc) = 0.2$, $Pr(s|cs) = 0.3$, $Pr(s|sc) = 0.6$, $Pr(s|ss) = 0.9$



a) Solving the following equations;

$$\pi_{cc} = 0.8\pi_{cc} + 0.7\pi_{cs}$$

$$\pi_{cs} = 0.4\pi_{sc} + 0.1\pi_{ss}$$

$$\pi_{sc} = 0.2\pi_{cc} + 0.3\pi_{cs}$$

$$\pi_{ss} = 0.9\pi_{cs} + 0.6\pi_{sc}$$

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

$\therefore \pi_{sc} = \frac{10}{115}$; $\pi_{cc} = \frac{35}{115}$;

$\pi_{cs} = \frac{10}{115}$; $\pi_{ss} = \frac{60}{115}$.

b) $P = \pi_{sc} + \pi_{ss} = \frac{70}{115}$.

Scanned with CamScanner

c) Transition matrix raised to power 100.

$$\begin{array}{c|cccc} & CC & SC & CS & SS \\ \hline CC & 0.30 & 0.08 & 0.08 & 0.52 \\ SC & 0.30 & 0.08 & 0.08 & 0.52 \\ CS & 0.30 & 0.08 & 0.08 & 0.52 \\ SS & 0.30 & 0.08 & 0.08 & 0.52 \end{array}$$

Problem 6

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

a. We are given $X \sim N(36, 0.1)$. We compute the probability:

$$
\begin{aligned}
P(X < 35.8) + P(X > 36.2) &= P(z < 35.8-36/0.1) + P(z > 36.2-36/0.1) \\
&= P(z < -2) + P(z > 2) \\
&= 0.0228 + (1 - 0.9772) \\
&= 0.0456.
\end{aligned}
$$

b. This is binomial with $n = 16$, $p = 0.0456$.

$P(X = 0) = ({}^{16}C_0)\ (0.0456)^0\ (1 - 0.0456)^{16} = 0.4739.$

c. This is binomial with $n = 16$, $p = 0.0456$.

$P(X = 1) = ({}^{16}C_1)\ (0.0456)^1\ (1 - 0.0456)^{15} = 0.3623.$

d. Now $X \sim N(37, 0.4)$. We compute the probability:

$$
\begin{aligned}
P(X < 35.8) + P(X > 36.2) &= P(z < 35.8-37/0.4) + P(z > 36.2-37/0.4) \\
&= P(z < -3) + P(z > -2) \\
&= 0.0013 + (1 - 0.0028) \\
&= 0.97
\end{aligned}
$$

Problem 7

We model this problem as a sum of geometric random variables which are independent.

$$X = t_1 + t_2 + t_3 + ... + t_d$$

Each t_i represents the number of days to get the i^{th} new pokemon.

(a)

So, $E[t_1]$ will be 1, obviously, because any pokemon we get on the first day will be a new pokemon. And if p_i is the probability of success (that is we get a new pokemon on any day), if we already have found $i-1$ distinct pokemons,

$$p_{i+1} = \frac{d-i}{d},$$

because for i^{th} day, any of the $d-i$ pokemons can be the new pokemon.
Each of these t_i are geometric random variables with probability of success p_i, hence

$$E[t_i] = \frac{1}{p_i}$$

$$
\begin{aligned}
E[X] &= \sum_{i=1}^{d} E[t_i] \\
E[X] &= 1 + \frac{d}{d-1} + \frac{d}{d-2} + ... + \frac{d}{1} \\
E[X] &= d(1 + \frac{1}{2} + \frac{1}{3} + ... + \frac{1}{d-2} + \frac{1}{d-1} + \frac{1}{d}) \\
E[X] &= dH_d
\end{aligned}
$$

(b)

$$Var(t_i) = \frac{1-p_i}{p_i^2} = \frac{d(i-1)}{(d+1-i)^2}$$

All of the t_i are independent of each other. They depend on i which represents the number of pokemons that have already been found, but the number of days to get a new pokemon is independent. So we have:

$$
\begin{aligned}
Var(X) &= \sum_{i=1}^{d} Var[X_i] \\
&= d\sum_{i=1}^{d} \frac{(i-1)}{(d+1-i)^2} \\
&= d \cdot [\frac{1}{(d-1)^2} + \frac{2}{(d-2)^2} + \cdots + \frac{d-1}{1}]
\end{aligned}
$$