

Problem |

a) Since the coin is fair, each RV is from $\text{Ber}(0.5)$

Let X_1, X_2, X_3 be the three flips.

$$X = X_1 + X_2 \text{ and } Y = X_2 + X_3, E[X_i] = \mu_i$$

$$\text{Cov}(X, Y) = \text{Cov}(X_1 + X_2, X_3 + X_2)$$

$$\Rightarrow \text{Cov}(X, Y) = E[(X_1 + X_2 - E[X_1 + X_2])(X_3 + X_2 - E[X_3 + X_2])]$$

$$\Rightarrow \text{Cov}(X, Y) = E[((X_1 - \mu_1) + (X_2 - \mu_2))((X_2 - \mu_2) + (X_3 - \mu_3))] - \text{By LOE } E[X_1 + X_2] = \mu_1 + \mu_2$$

$$\Rightarrow \text{Cov}(X, Y) = E[((X_1 - \mu_1) + (X_2 - \mu_2))((X_2 - \mu_2) + (X_3 - \mu_3))] - \text{By LOE } E[X_1 + X_2] = \mu_1 + \mu_2$$

$$\Rightarrow = E[(X_1 - \mu_1)(X_2 - \mu_2)] + E[(X_1 - \mu_1)(X_3 - \mu_3)] + E[(X_2 - \mu_2)(X_3 - \mu_3)] + E[(X_2 - \mu_2)^2]$$

$$\Rightarrow \text{Cov}(X, Y) = \text{Cov}(X_1, X_2) + \text{Cov}(X_1, X_3) + \text{Cov}(X_2, X_3) + \text{Var}(X_2)$$

\because Flips are independent

$$\Rightarrow \text{Cov}(X, Y) = \text{Var}(X_2) = \frac{1}{4} \quad \because X_i \perp X_j \Rightarrow \text{Cov}(X_i, X_j) = 0 \text{ and } X \sim \text{Ber}(p) \Rightarrow \text{Var}(X) = p(1-p)$$

b) $Y = X^2$

$$\text{Cov}(X, Y) = E[XY] - E[X]E[Y]$$

$$E[X]E[Y] = 0 \quad \because E[X] = \frac{1}{5}[0 - 5 - 2 + 2 + 5]$$

$$E[XY] = E[X^3] = 0$$

$$\therefore \text{Cov}(X, Y) = 0$$

c) No. 2b) is a counterexample.

Since covariance only captures linear relationships, dependent random variables with no linear relationship will have zero covariance.

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

$$\begin{aligned} E[X] &= \int_0^\infty xf(x)dx && \because x \geq 0 \\ \Rightarrow E[X] &= \int_0^t xf(x)dx + \int_t^\infty xf(x)dx \\ \int_0^t xf(x)dx &\geq 0 && \because t > 0, X \geq 0 \text{ and } f(x) \geq 0 \\ \Rightarrow E[X] &\geq \int_t^\infty xf(x)dx \end{aligned}$$

— (1)

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

Solⁿ:

$$\begin{aligned} E[X] &\geq \int_t^\infty xf(x)dx \\ \Rightarrow E[X] &\geq \int_t^\infty tf(x)dx \\ \Rightarrow E[X] &\geq t \int_t^\infty f(x)dx = tP(X > t) \\ \Rightarrow P(X > t) &\leq \frac{E[X]}{t} \end{aligned}$$

— (2)

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

Solⁿ:

$$E[X] \geq tP(X > t) \quad \text{— From (2)}$$

$$\begin{aligned} \Rightarrow P(X > t) &\leq \frac{E[X]}{t} \\ \Rightarrow P((X - \mu)^2 > t^2) &\leq \frac{E[(X - \mu)^2]}{t^2} \\ \Rightarrow P((X - \mu)^2 > t^2) &\leq \frac{\sigma^2}{t^2} \\ \Rightarrow P(|X - \mu| \geq t) &\leq \frac{\sigma^2}{t^2} \end{aligned}$$

Problem 3.

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

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

(i)

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

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

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

$$F_Z(z) = 1 - P(X_1 > z, X_2 > z, \dots, X_k > z)$$

$$F_Z(z) = 1 - P(X_1 > z, X_2 > z, \dots, X_k > z)$$

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

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

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

$$f_Z(z) = \frac{\partial F_Z(z)}{\partial z}$$

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

$$\therefore Z \sim \text{exp}(\lambda_1 + \lambda_2 + \dots + \lambda_k)$$

(ii) $\because Z \sim \text{exp}(\lambda_1 + \lambda_2 + \dots + \lambda_k)$

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

(iii) $\because Z \sim \text{exp}(\lambda_1 + \lambda_2 + \dots + \lambda_k)$

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

3b

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

$$Z = XY$$

$$\begin{aligned} F_Z(z) &= P(Z \leq z) \\ &= P(XY \leq z) = 1 - P(XY > z) \end{aligned}$$

For finding the lower limit of integration over Y

$$0 \leq X \leq Y \leq 1$$

For any value of Z

$$\Rightarrow XY \geq z \quad \text{---} \quad (1)$$

$$Y \geq X \quad [\text{Given}]$$

$$\Rightarrow Y^2 \geq XY \quad \text{---} \quad (2)$$

$$\Rightarrow Y^2 \geq XY \geq z \quad [\text{From } (1) \text{ and } (2)]$$

$$\Rightarrow Y \geq \sqrt{z}$$

∴ For any any value of Z

$$F_Z(z) = 1 - \int \int f_{xy}(x, y) dx dy$$

$y = \sqrt{z} \quad x = \frac{z}{y}$

$$F_Z(z) = 1 - 2 \int_{\sqrt{z}}^1 \int_0^{z/y} dx dy$$

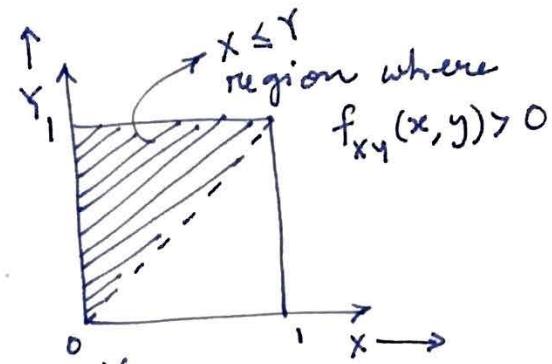
$$F_Z(z) = 1 - 2 \int_{\sqrt{z}}^1 \left(y - \frac{z}{y} \right) dy$$

$$F_Z(z) = 1 - 2 \left(\frac{1 - (\sqrt{z})^2}{2} - z \left(0 - \frac{\ln(z)}{2} \right) \right)$$

$$\boxed{\therefore F_Z(z) = z - z \ln(z)}$$

$$\therefore f_Z(z) = \frac{d F_Z(z)}{dz}$$

$$\boxed{\therefore f_Z(z) = -\ln(z) \quad \forall z \in (0, 1)}$$



Problem 4.

a)

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

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

$$E[X] = \frac{1}{2}E[X+20] + \frac{1}{2}\left(\frac{1}{2} \cdot 5 + \frac{1}{2}E[X+10]\right)$$

$$E[X] = \frac{1}{2}E[X+20] + \frac{1}{2}\left(\frac{1}{2} \cdot 5 + \frac{1}{2}E[X+10]\right)$$

$$E[X] = \frac{55}{4} + \frac{3E[X]}{4}$$

$$E[X] = 55$$

b) Similar to (1)

$$E[X^2] = \frac{1}{2}E[X^2|East] + \frac{1}{2}(E[X^2|West, West]P(West) + E[X^2|West, South]P(South))$$

$$E[X^2] = \frac{1}{2}E[(X+20)^2] + \frac{1}{2}\left(\frac{1}{2} \cdot 5^2 + \frac{1}{2}E[(X+10)^2]\right)$$

$$E[X^2] = \frac{1}{2}E[400 + 40X + X^2] + \frac{1}{2}\left(\frac{1}{2} \cdot 25 + \frac{1}{2}E[X^2 + 100 + 20X]\right)$$

$$E[X^2] = \frac{925}{4} + 25E[X] + \frac{3}{4}E[X^2]$$

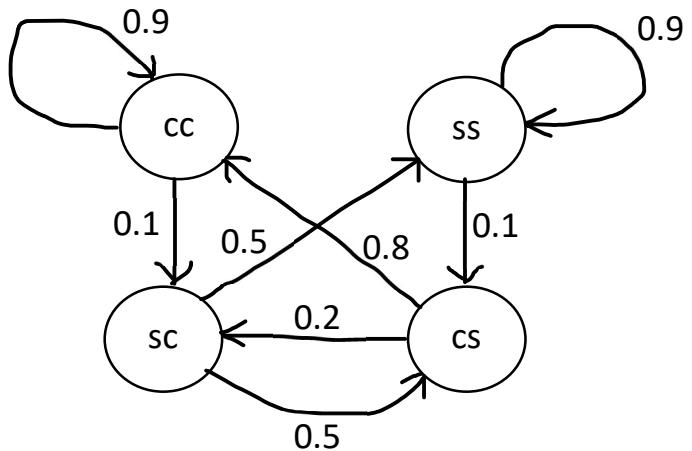
$$E[X^2] = 100E[X] + 925 = 625$$

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

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

Problem 5

$$\begin{array}{l}
 \text{cc} \quad \text{sc} \quad \text{cs} \quad \text{ss} \\
 \text{cc} \left[\begin{array}{cccc} 0.9 & 0.1 & 0.0 & 0.0 \\ 0 & 0.0 & 0.5 & 0.5 \\ 0.8 & 0.2 & 0.0 & 0.0 \\ 0 & 0.0 & 0.1 & 0.9 \end{array} \right]
 \end{array}$$



Given:

$$Pr[c|cc] = 0.9$$

$$Pr[c|cs] = 0.8$$

$$Pr[c|sc] = 0.5$$

$$Pr[c|ss] = 0.1$$

$$Pr[s|cc] = 0.1$$

$$Pr[s|cs] = 0.2$$

$$Pr[s|sc] = 0.5$$

$$Pr[s|ss] = 0.9$$

(a) The Markov chain and transition matrix will be as shown on the figure.

Solving the following stationary equations:

$$\pi_{cc} = 0.9\pi_{cc} + 0.8\pi_{cs}$$

$$\pi_{cs} = 0.5\pi_{sc} + 0.1\pi_{ss}$$

$$\pi_{sc} = 0.1\pi_{cc} + 0.2\pi_{cs}$$

$$\pi_{ss} = 0.5\pi_{sc} + 0.9\pi_{ss}$$

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

$$\therefore \pi_{sc} = \frac{1}{15}; \pi_{cc} = \frac{8}{15}; \pi_{cs} = \frac{1}{15}; \pi_{ss} = \frac{5}{15}$$

$$(b) P = \pi_{sc} + \pi_{ss} = \frac{1}{15} + \frac{5}{15} = \frac{6}{15}$$

(c) Transition matrix raised to the power 100.

$$\begin{array}{l}
 \text{cc} \quad \text{sc} \quad \text{cs} \quad \text{ss} \\
 \text{cc} \left[\begin{array}{cccc} .53 & .06 & .06 & .33 \\ .53 & .06 & .06 & .33 \\ .53 & .06 & .06 & .33 \\ .53 & .06 & .06 & .33 \end{array} \right]
 \end{array}$$

Q6.

- (a) Since $X = (X_1, \dots, X_k)$ is multinomial, every linear combination $t_1 X_1 + t_2 X_2 + \dots + t_k X_k$ is Normal. In particular, for $t_j = 1$, and $t_i = 0 \neq i \neq j$ we have $t \cdot X_j = X_j$ a Normal distribution
- (b) Note that $P(X+Y) = P(X+S, X) = P(S=-1) = 1/2$

Since a linear combination of X, Y has non-zero probability mass at -1 , the linear combination is not Normal. Therefore (X, Y) can't be Normal.

- (c) Since Z, W are i.i.d $N(0, 1)$,
 $t_1 Z \sim N(0, t_1^2)$ and $t_2 W \sim N(0, t_2^2)$ and
are independent. Therefore $t_1 Z + t_2 W \sim N(0, t_1^2 + t_2^2)$. Since this holds for any t_1, t_2 , (Z, W) is a Multivariate Normal.

Consider $t_1, t_2 \in \mathbb{R}$

$$\text{We have } t_1(Z+2W) + t_2(3Z+5W)$$

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

This is Normal $(0, (t_1 + 3t_2)^2 + (2t_1 + 5t_2)^2)$.

$\therefore (Z+2W, 3Z+5W)$ is a multivariate Normal.

(d) Since $X = (X_1, \dots, X_n)$ and $Y = (Y_1, \dots, Y_m)$ are Multivariate Normals,

$a_1 X_1 + a_2 X_2 + \dots + a_n X_n$ and

$b_1 Y_1 + b_2 Y_2 + \dots + b_m Y_m$ are Normals.

Since X & Y are independent,

$\sum_{i=1}^m a_i X_i$ and $\sum_{j=1}^n b_j Y_j$ are independent.

Therefore $\sum_{i=1}^m a_i X_i + \sum_{j=1}^n b_j Y_j$ is Normal

for any a_i 's & b_j 's

This shows that $(X_1, \dots, X_n, Y_1, \dots, Y_m)$ is a Multivariate Normal.

(e). From Fact 2, we have

$$\text{Cor}(X+Y, X-Y)$$

$$= \text{Var}(X) - \text{Cov}(X, Y) + \text{Cov}(Y, X) - \text{Var}(Y)$$

$$= 0.$$

From Fact 1, we have $X+Y$ and $X-Y$ to be independent.

We also have $X+Y \sim N(0, 2)$ and

$$X-Y \sim N(0, 2)$$

Therefore $t_1(X+Y) + t_2(X-Y)$ is also Normal.

Problem 7

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

$$X = t_1 + t_2 + t_3 + \dots + t_d$$

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

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

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

(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:

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

