CSE 544, Spring 2020, Probability and Statistics for Data Science

Assignment 2: Random variables and Markov chain **Solution**

1. Transformation of Normal random variable **(Total 5 points)**

a) Given:

$X \sim Nor(\mu, \sigma^2); Y = aX + b$

$F_Y(y) = P(Y \leq y) = P(aX + b \leq y)$

$\Rightarrow F_Y(y) = F_X\left(\frac{y-b}{a}\right)$

Differentiating both sides w.r.t y, we get

$$f_Y(y) = \frac{1}{a} f_X\left(\frac{y-b}{a}\right) \qquad -(1)$$

$$f_Y(y) = \frac{1}{a\sqrt{2\pi\sigma^2}} e^{-\frac{\left(\left(\frac{y-b}{a}\right)-\mu\right)^2}{2\sigma^2}}$$

$$f_Y(y) = \frac{1}{a\sqrt{2\pi\sigma^2}} e^{-\frac{\left(y-(a\mu+b)\right)^2}{2(a\sigma)^2}}$$

$$f_Y(y) = \frac{1}{\sqrt{2\pi(a\sigma)^2}} e^{-\frac{\left(y-(a\mu+b)\right)^2}{2(a\sigma)^2}}$$

$$f_Y(y) = \frac{1}{\sqrt{2\pi(\sigma')^2}} e^{-\frac{(y-\mu')^2}{2(\sigma')^2}}$$

$\therefore Y \sim Nor(\mu', \sigma'^2)$ where $\mu' = (a\mu + b)$ and $\sigma' = a\sigma$

b) Given $X \sim Nor(0,1), X \sim Nor(0,1), X \perp Y \; and \; Z = X + Y$

$$f_Z(z) = \int_{x=-\infty}^{\infty} f_{XY}(x, z-x)dx$$

$$f_Z(z) = \int_{x=-\infty}^{\infty} f_X(x) f_Y(z-x)dx$$

$$f_Z(z) = \frac{1}{2\pi}\int_{x=-\infty}^{\infty} e^{\left(-\frac{x^2}{2}\right)} e^{\left(-\frac{(z-x)^2}{2}\right)} dx$$

$$f_Z(z) = \frac{1}{2\pi}\int_{x=-\infty}^{\infty} e^{\left(-\frac{2x^2+z^2-2xz}{2}\right)} dx = \frac{1}{2\pi}\int_{x=-\infty}^{\infty} e^{\left(-\left(x^2+\frac{z^2}{2}-xz\right)\right)} dx$$

$$f_Z(z) = \frac{1}{2\pi}\int_{x=-\infty}^{\infty} e^{\left(-\left(x^2+\frac{z^2}{2}-xz\right)\right)} dx = \frac{1}{2\pi}\int_{x=-\infty}^{\infty} e^{\left(-\left(x-\frac{z}{2}\right)^2-\frac{z^2}{4}\right)} dx$$

$$f_Z(z) = \frac{1}{2\pi} e^{-\frac{z^2}{2.2}} \int_{x=-\infty}^{\infty} e^{-\left(x-\frac{z}{2}\right)^2} dx$$

$$f_Z(z) = \frac{1}{2\pi} e^{-\frac{z^2}{2.2}} \int_{x=-\infty}^{\infty} e^{-\frac{\left(x-\frac{z}{2}\right)^2}{2.\frac{1}{2}}} dx$$

Let $\left(x-\frac{z}{2}\right) = w$, then

$$f_Z(z) = \frac{1}{\sqrt{2}*\sqrt{2\pi}} e^{-\frac{z^2}{2.2}} \left[\frac{\sqrt{2}}{\sqrt{2\pi}}\int_{x=-\infty}^{\infty} e^{--\frac{w^2}{2.\frac{1}{2}}} dw\right]$$

Term inside the bracket is normal distribution with variance $\frac{1}{2}$, therefore it integrates to 1.

$$f_Z(z) = \frac{1}{\sqrt{2\pi*2}} e^{-\frac{z^2}{2.2}}$$

Therefore $Z \sim N(0,2)$.

2. Introduction to Covariance (Total 5 points)

a) Since the coin is fair, each RV is from Ber(0.5)

Let X_1, X_2, X_3 be the three flips.

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

$Cov(X,Y) = Cov(X_1 + X_2, X_3 + X_2)$

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

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

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

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

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

$\because Flips\ are\ indepentent$

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

b) $Y = X^2$

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

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

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

$\therefore Cov(X,Y) = 0$

c) No. 2b) is a counterexample.
Since covariance only captures linear relationships, dependent random variables with no linear relationship will have zero covariance.

3. Inequalities **(Total 10 points)**

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:

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

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

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

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

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

4. Functions of Random Variables. **(Total 10 points)**

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

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

(i)

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

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

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

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

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

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

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

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

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

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

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

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

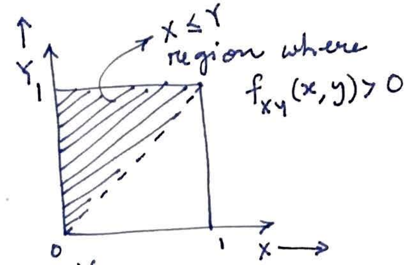
$$\therefore E[Z] = \frac{1}{(\lambda_1 + \lambda_2 + \cdots + \lambda_k)}$$

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

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

4.b)

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



$Z = XY$

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

$= P(XY \le z) = 1 - P(XY > z)$

For finding the lower limit of integration over Y

$0 \le X \le Y \le 1$

For any value of $\underline{\underline{Z \ge z}}$

$Z \geq z$

$\Rightarrow XY \geq z$ —— ①

$Y \geq X$ [Given]

$\Rightarrow Y^2 \geq XY$ —— ②

$\Rightarrow Y^2 \geq XY \geq z$ [From ① and ②]

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

$\therefore$ For any any value of z

$$F_Z(z) = 1 - \int_{y=\sqrt{z}}^{1} \int_{x=\frac{z}{y}}^{y} f_{XY}(x,y)\, dx\, dy$$

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

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

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

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

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

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

5. Daenerys returns to King's Landing, almost. (Total 10 points)

a)

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

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

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

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

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

$$E[X] = 55$$

b) Similar to (1)

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

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

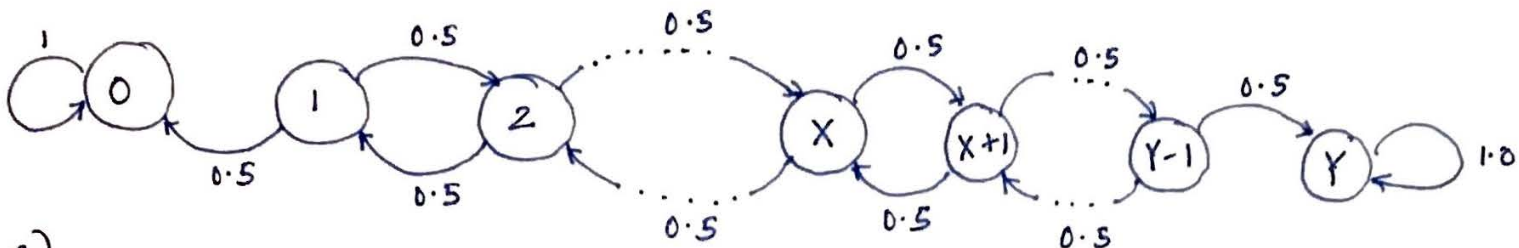
$$E[X^2] = \frac{1}{2}E[400 + 40X + X^2] + \frac{1}{2}(\frac{1}{2}.25 + \frac{1}{2}E[X^2 + 100 + 20X])$$

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

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

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

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



a)

Let each state represents the current amount you have

Let $p_i \equiv$ Probability of reaching state (Y) (target state) given the current state is "i".

$\because$ 0 and Y are the absorbing / stopping state

$\therefore$ $p_0 = 0$ and $p_Y = 1$ —— ①

Now $p_i = 0.5\, p_{i-1} + 0.5\, p_{i+1}$

$\therefore$ $p_i = \dfrac{(p_{i+1} + p_{i-1})}{2}$

$\therefore$ Sequence $\{p_i\}_{i=0}^{Y}$ is an AP.

Let the common difference of AP = "d."

$p_i = p_0 + i d$ —— ②

From ①, ② we get $\boxed{d = \frac{1}{Y}}$

$\boxed{\therefore\ p_X = \frac{X}{Y} + 0 = \frac{X}{Y}}$

b) Similar to above approach,

Let $q_i \equiv$ probability of reaching state 0 from state "i".

$\therefore$ $q_i = 0.5\, q_{i+1} + 0.5\, q_{i-1}$ —— ①

$\because$ 0 and Y are the absorbing states.

$q_0 = 1$ and $q_Y = 0$ —— ②

Similar to the structure of $\{p_i\}$, $\{q_i\}$ forms an AP

On solving $q_i = q_0 + di$ and ②

We get $\boxed{\text{"}d\text{"} = -\frac{1}{Y}}$ and $\boxed{q_X = 1 - \frac{X}{Y}}$

c) Since the stopping criteria of the random walk is reaching the value of 0 or Y. (only two possible absorptio states)

∴

For state 0, the value of the stack 0

For state Y, —— " —————— " —— Y

If the current state (starting amount) is "X", then the probability of reaching state "Y" $= P_X$

Probability of reaching state 0 from "X" $= (1 - P_X)$

∴ Expected value of stack at the end $= 0\,(1 - P_X) + Y\,P_X$

$$= Y \cdot \frac{X}{Y}$$

$$= \underline{\underline{X}}$$

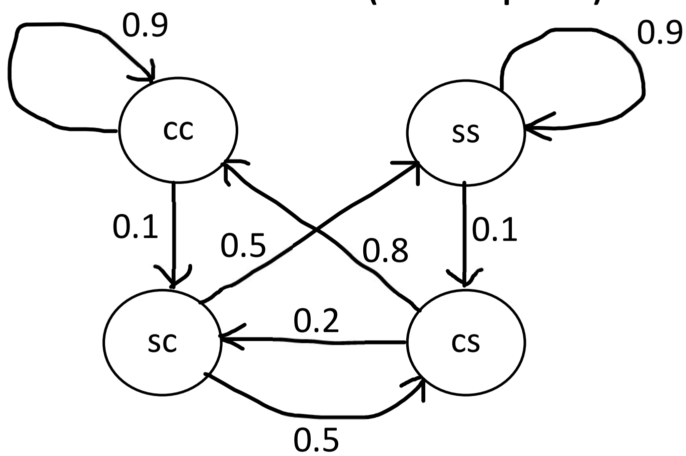
6. Stay Away from stocks!! (Total 15 points)

d)

	6.a	**6.b**	**6.c**
TEST CASE 1 >>	0.6675	0.3325	100.125
TEST CASE 2 >>	0.9997	0.0003	199.94
TEST CASE 3 >>	1.0	0.0	250

7. Dependence on past two states **(Total 15 points)**

$$
\begin{array}{c|cccc}
 & cc & sc & cs & ss \\
\begin{array}{c} cc \\ sc \\ cs \\ ss \end{array} &
\multicolumn{4}{c}{\begin{bmatrix} 0.9 & 0.1 & 0.0 & 0.0 \\ 0 & 0.0 & 0.5 & 0.5 \\ 0.8 & 0.2 & 0.0 & 0.0 \\ 0 & 0.0 & 0.1 & 0.9 \end{bmatrix}}
\end{array}
$$



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.

$$
\begin{array}{c|cccc}
 & cc & sc & cs & ss \\
\begin{array}{c} cc \\ sc \\ cs \\ ss \end{array} &
\multicolumn{4}{c}{\begin{bmatrix} .53 & .06 & .06 & .33 \\ .53 & .06 & .06 & .33 \\ .53 & .06 & .06 & .33 \\ .53 & .06 & .06 & .33 \end{bmatrix}}
\end{array}
$$