ORIE 5530, Fall 2020: Individual exercise

****

Rules of the game :

- This is an individual assignment. You must do it by yourself. Do not discuss questions or answers with others until the exercise is due.
- There are three questions and a maximum total of 75 points.
- Write clearly and legibly. Show your derivation clearly. This will allow for partial credit if the final answer is wrong but the approach is correct.
- There is empty space for your answers. You could use additional pages. Make sure you include all your solution in your submitted version.
- You can use class notes (posted or your own handwritten) and homework solutions.

****

Name: Hongyi Nie

Net ID: hn327

1. (20 points) A dreaded new disease causes extreme fear of probability-related courses. A blood test is 90% effective in detecting this disease when it is in fact present. However, the blood test also yields a false positive result for 10% of healthy people tested. (That is, if a healthy person is tested, then with probability 0.1 the test will conclude that the person has the disease.) Suppose that 1% (0.01) of the students at Cornell actually have the disease.

   a. What is the probability that a student that is selected at random and tested for the disease has a positive outcome (i.e., the result of the test is positive)?

   b. What is the probability that a randomly tested student has the disease, given that his/her test result is positive?

   c. Suppose that 1000 students are selected at random and are tested. What are the mean and the variance of the number of people that test positive?

   d. In the setting of part (c), provide one bound and one approximation for the probability that the number of students that are tested positive exceeds 125. For the latter you do not have to reach a final number. It suffices to write an explicit mathematical expression.

a. Let A = positive test result Y = have desease
   B = negative test result N = not have desease.

$p(A) = p(A|Y) \times p(Y) + p(A|N) \cdot p(N)$.
$0.01 \times 0.9 + 0.99 \times 0.1 = 0.108 = 10.8\%$

b) $p(Y|A) = \dfrac{p(A|Y) \cdot p(Y)}{p(A)}$
$= 0.9 \times 0.01 / 0.108 = 0.0833 = 8.33\%$

c). Since $A \sim$ Bernoulli. $(p = 0.108)$,
Let Z = the event k student have A among 1000 students
$\sim$ Binominal $(p = 0.108,\ n = 1000)$

$E(Z) = 1000 \times p(A) = 1000 \times 0.108 = 108.$

$Var(Z) = 1000\, p(A) \cdot (1 - p(A)) = 1000 \times 0.108 \times (1 - 0.108) = 96.336$

d). $X = 125$, $\mu = 108$, $\sigma = \sqrt{96.336} = 9.815$

$Z = \dfrac{X - \mu}{\sigma} = \dfrac{125 - 108}{9.815} = 1.7320$

at 95% confidence level:

$p(Z \geq z) = 1 - p(Z \leq z) = 1 - 0.9582 = 0.0418 = 4.18\%$

CI: $\bar{X} \pm z \times \dfrac{\sigma}{\sqrt{n}} = (108 - 1.96 \times \dfrac{9.815}{\sqrt{1000}},\ 108 + 1.96 \times \dfrac{9.815}{\sqrt{1000}})$
$= (107.39,\ 108.61)$

$\because$ the upperbound is 108.61 - probability of test positive excess 125 = 4.18%

2. (25 points) Let $\{X_n : n \geq 0\}$ be a Markov chain taking values in $\{1, 2, 3\}$ with transition matrix

$$\begin{array}{cc} & \text{output } X_1 \\ \text{input } X_0 \quad P = & \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1 & 0 & 0 \\ 1/3 & 1/3 & 1/3 \end{pmatrix} \end{array}.$$

Compute the following

(a) $\mathbb{E}(X_1|X_0 = 3)$, $\mathbb{E}(X_1^2|X_0 = 3)$ and $Var(X_1|X_0 = 3)$ (the expectation and variance of X_1 given that $X_0 = 3$)

For parts (b),(c) and (d) assume that the initial distribution is $\alpha = (1/2, 0, 1/2)$ (that is, we initialize the chain so that $\mathbb{P}\{X_0 = 1\} = \mathbb{P}\{X_0 = 3\} = 1/2$ and $\mathbb{P}\{X_0 = 2\} = 0$).

(b) $Var(X_1)$ and $\mathbb{E}[X_1]$.

(c) $\mathbb{P}(X_1 = 3, X_2 = 1)$

(d) $\mathbb{P}(X_1 = 3|X_2 = 1)$ $\quad p(X_9=2 | X_1=3, X_4=1, X_7=1)$

(e) $\mathbb{P}(X_9 = 2|X_1 = 3, X_4 = 1, X_7 = 1)$

a) $E(X_1 | X_0=3) = \sum_{i=0}^{3} i\, p(X_1 = i | X_0 = 3)$

$= 1 \times \frac{1}{3} + 2 \times \frac{1}{3} + 3 \times \frac{1}{3} = 2$

$E(X_1^2 | X_0=3) = \sum_{i=0}^{3} i^2 p(X_1 = i | X_0 = 3)$

$= 1^2 \times \frac{1}{3} + 2^2 \times \frac{1}{3} + 3^2 \times \frac{1}{3} = \frac{14}{3}$

$Var(X_1 | X_0=3) = E(X_1^2 | X_0=3) - E(X_1|X_0=3)^2 = \frac{14}{3} - 2^2 = \frac{2}{3}.$

b). $E[X_1] = \sum_{i=0}^{3} i\, p(X_1 = i) = \sum_{i=0}^{3} i [p(X_1 = i | X_0=1) \cdot p(X_0=1)$
$+ p(X_1=i | X_0=3) \cdot p(X_0=3)]$

$= 1 \times [p(X_1=1 | X_0=1) \cdot p(X_0=1) + p(X_1=1|X_0=3) \cdot p(X_0=3)]$
$+ 2 \times [p(X_1=2 | X_0=1) p(X_0=1) + p(X_1=2 | X_0=3) \cdot p(X_0=3)]$
$+ 3 \times [p(X_1=3 | X_0=1) p(X_0=1) + p(X_1=3 | X_0=3) p(X_0=3)]$
$= 1 \times [0 + \frac{1}{3} \times \frac{1}{2}] + 2 \times [\frac{1}{2} \times \frac{1}{2} + \frac{1}{3} \times \frac{1}{2}] + 3 [\frac{1}{2} \times \frac{1}{2} + \frac{1}{3} \times \frac{1}{2}]$
$= 2.25 = \frac{9}{4}$

$$E[X_1^2] = \sum_{i=0}^{3} i^2 p(X_1=i) = \sum_{i=0}^{3} i^2 [p(X_1=i \mid x_0=1)\cdot p(X_0=1)$$
$$+ p(X_1=i \mid X_0=3)\cdot p(X_0=3)]$$

$$= 1^2 \times [p(X_1=1 \mid X_0=1)\cdot p(X_0=1) + p(X_1=1 \mid X_0=3)\cdot p(X_0=3)]$$
$$+ 2^2 \times [p(X_1=2 \mid X_0=1) p(X_0=1) + p(X_1=2 \mid X_0=3)\cdot p(X_0=3)]$$
$$+ 3^2 \times [p(X_1=3 \mid X_0=1) p(X_0=1) + p(X_1=3 \mid X_0=3) p(X_0=3)]$$
$$= 1^2 \times [0 + \tfrac{1}{3}\times\tfrac{1}{2}] + 2^2 \times [\tfrac{1}{2}\times\tfrac{1}{2} + \tfrac{1}{3}\times\tfrac{1}{2}] + 3^2 [\tfrac{1}{2}\times\tfrac{1}{2} + \tfrac{1}{3}\times\tfrac{1}{2}]$$
$$= \frac{67}{12}$$

$$Var(X_1) = E(X^2) - E(X)^2 = \frac{67}{12} - \left(\frac{9}{4}\right)^2 = \frac{25}{48}$$

c. $p(X_1=3, X_2=1) = p(X_2=1 \mid X_1=3)\cdot p(X_1=3)$

$$= \frac{1}{3} \times p(X_1=3) = \frac{1}{3} \times \frac{5}{12} = \frac{5}{36}$$

since $p(X_1=1) = p(X_1=3 \mid X_0=1) p(X_0=1) + p(X_1=3 \mid X_0=3) p(X_0=3)$

$$= \frac{1}{2}\times\frac{1}{2} + \frac{1}{3}\times\frac{1}{2} = \frac{5}{12}$$

d. $$p^2 = \begin{pmatrix} 2/3 & 1/6 & 1/6 \\ 0 & 1/2 & 1/2 \\ 4/9 & 5/18 & 5/18 \end{pmatrix}$$

$$p(X_1=3 \mid X_2=1) = \frac{p(X_1=3, X_2=1)}{p(X_2=1)}$$

$$p(X_2=1) = p(X_2=1 \mid X_1=1)\cdot p(X_1=1)$$
$$+ p(X_2=1 \mid X_1=3)\cdot p(X_1=3)$$
$$= \frac{2}{3}\times\frac{1}{2} + \frac{4}{9}\times\frac{1}{2} = \frac{5}{9}$$

e. $p(X_9=2 \mid X_1=3, X_4=1, X_7=1) = p(X_9=2 \mid X_7=1)$ by Markov Chain property.

$$= \frac{1}{6}.$$

3. (30 points) Waiting for a cab: There is a taxi stop at the corner of Sherman and Foster in Evanston, IL. Taxis arrive every geometric number of minutes with parameter $p^T = 1/2$ (T for Taxi). If a taxi arrives and there are already two taxis waiting at the stop, this taxi leaves. Customers arrive to the stop every geometric number of minutes with parameter $p^c = 1/2$ (c for customer). If when a customer arrives, there is a cab (or more) waiting, the customer immediately takes the cab and they both leave. If when a customer arrives there are not cabs, the customer waits unless there are already two people waiting at the station, in which case the customer leaves. Thus, there can be at most two people waiting at any given time and there can be at most two taxis waiting at any given time. However, if there are cabs waiting there cannot be customer waiting and vice versa (because as soon as there is a customer and a cab, they both leave).

(a) Describe the above as a discrete time Markov chain over the states $\{-2, -1, 0, 1, 2\}$ (where -2 stands for two cabs waiting and 2 for two customers waiting). Explicitly write the transition probabilities (in a matrix or diagram) in terms of p^c and p^t.

(b) Say at time 0 there are two customer waiting in the station. What is the expected time until the station empties for the first time?

(c) Write explicitly the equations you would have to solve to obtain the long-run fraction of time $\pi = \{\pi_{-2}, \pi_{-1}, \pi_0, \pi_1, \pi_2\}$ that the chain spends in each state? **You do not have to solve these equations**. What is, in terms of π, the long run fraction of time that there are no cabs in the station?

(d) Again, in terms, of π, what is the average number of customers, per minute, that find the station full (with two people waiting)?

a). $p^T = 1/2 = 1 - p^T$; $p^c = 1/2 = 1 - p^c$

$$P = \begin{array}{c|ccccc} & -2 & -1 & 0 & 1 & 2 \\ \hline -2 & 1-p^cp^t & p^cp^t & 0 & 0 & 0 \\ -1 & p^cp^t & 2p^cp^t & p^cp^t & 0 & 0 \\ 0 & 0 & p^cp^t & 2p^cp^t & p^cp^t & 0 \\ 1 & 0 & 0 & p^cp^t & 2p^cp^t & p^cp^t \\ 2 & 0 & 0 & 0 & p^cp^t & 1-p^cp^t \end{array}$$

b). E(Station Empty at 1st time) = E(from 2 customer to 1 customer) + E(from 1 customer to 0 customer)

E(from 2 customer to 1 customer) $= \frac{1}{p} = \frac{1}{p^cp^t} = \frac{1}{1/4} = 4$

$\sim$ Geometric $(p = p^cp^t)$

E(from 1 customer to 0 customer) $= x$

$x = \frac{1}{4} \times 1 + \frac{1}{2} \times (1 + x) + \frac{1}{4} \times (1 + \underbrace{E(\text{from } 2 \text{ to } 1)}_{=4} + x)$

$x = \frac{1}{4} + \frac{1}{2} + \frac{1}{2}x + \frac{5}{4} + \frac{1}{4}x = 2 + \frac{3}{4}x$

$\Rightarrow \frac{1}{4}x = 2 \qquad x = 8$

Thus E(station empty 1st time) $= 4 + 8 = 12$

c). $\pi_{-2} = \frac{3}{4}\pi_{-2} + \frac{1}{4}\pi_{-1}$

$\pi_{-1} = \frac{1}{4}\pi_{-2} + \frac{1}{2}\pi_{-1} + \frac{1}{4}\pi_0$

$\pi_0 = \frac{1}{4}\pi_{-1} + \frac{1}{2}\pi_0 + \frac{1}{4}\pi_1$

$\pi_1 = \frac{1}{4}\pi_0 + \frac{1}{2}\pi_1 + \frac{1}{4}\pi_2$

$\pi_2 = \frac{3}{4}\pi_2 + \frac{1}{4}\pi_1$

d). $(1 - p^+) = \frac{1}{2}$

$p^c = \frac{1}{2}$

Avg # of customers to find station full

$= (1 - p^+)p^c\pi_2$

$= \frac{1}{2} \times \frac{1}{2}\pi_2 = \frac{1}{4}\pi_2$