EECS 126 Probability and Random Processes University of California, Berkeley: Spring 2019
Kannan Ramchandran April 9, 2019

Midterm 2

Last Name	First Name	SID

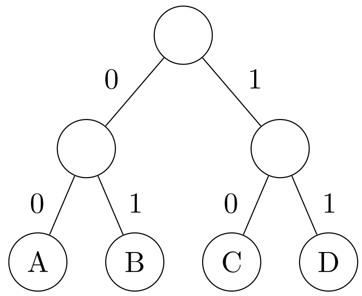
***Rules*.**

- No form of collaboration between the students is allowed. If you are caught cheating, you may fail the course and face disciplinary consequences.
- You have 10 minutes to read the exam and 100 minutes to complete it.
- The exam is not open book; we are giving you a cheat sheet. No calculators or phones allowed.
- Unless otherwise stated, all your answers need to be justified. Show all your work to get partial credit.
- Maximum you can score is 114 but 100 points is considered perfect.

Problem	points earned	out of
Problem 1		44
Problem 2		20
Problem 3		32
Problem 4		18
Total		114

Problem 1: Answer these questions briefly but clearly.

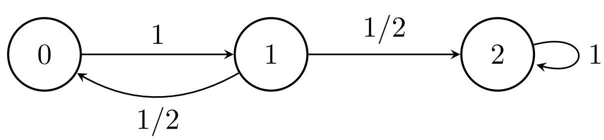
(a) [4] Which of the following frequencies for A,B,C and D can generate the following Huffman tree? (Select all that apply.)



- ◯ $p_A = 0.4\,,\; p_B = 0.3\,,\; p_C = 0.2\,,\; p_D = 0.1$
- ◯ $p_A = 0.35\,,\; p_B = 0.25\,,\; p_C = 0.2\,,\; p_D = 0.2$
- ◯ $p_A = 0.25\,,\; p_B = 0.25\,,\; p_C = 0.25\,,\; p_D = 0.25$
- ◯ $p_A = 0.2\,,\; p_B = 0.35\,,\; p_C = 0.2\,,\; p_D = 0.25$

2,3

(b) [**2+2+2**] Consider the the Markov Chain $(X_n)_{n\in\mathbb{N}}$ whose transitions are given by



1. X_n converges almost surely. True False

   If true, it converges a.s. to (N/A if false): __________

2. X_n converges in probability. True False

   If true, it converges i.p. to (N/A if false): __________

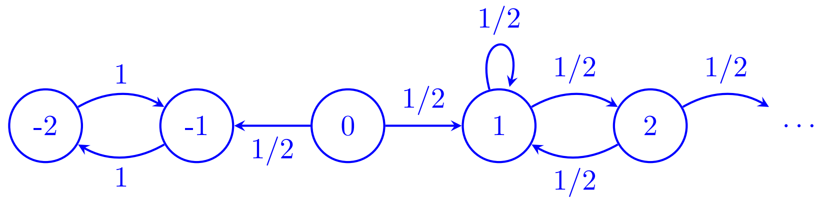
3. X_n converges in distribution. True False

   If true, it converges in distribution to (N/A if false): __________

All True. It converges to 2. The possible values of the sequence (X_n) look like

$(0, 1, 0, 1, 0, 1 \ldots$ (some finte number of 0,1s and eventually) $0, 1, 2, 2, 2, \ldots 2$ forever) except for the one sequence of infinitely repeating $0, 1$ oscillations and this has probability 0. So with probability 1(almost surely), X_n converges to 2. Since it converges almost surely, it certainly converges in probability and in distribution to 2.

(c) [6] Construct a reversible Markov Chain which has one positive recurrent class, one transient class, and one null recurrent class.



Since there is one PR class there is a unique stationary distribution, which is just the stationary distribution of the PR class viewed in isolation ($(1/2, 1/2)$ in the above example). The long term averages for the transient state (0) and the null recurrent states $(1, 2\dots)$ are 0. We just need to construct the PR class to satisfy detailed balance. It is easily verified that this is one such case.

(d) [6] Random Walk on a Random Graph: A particle performs a random walk on a graph with 3 vertices, labeled 1, 2 and 3, starting from state 1. However, at each time-step, before the particle makes a move, the edges of the graph are re-sampled according to an independent $G(3, 1/2)$ distribution. Once the edges have been sampled, the particle chooses a neighbor of its current state uniformly at random from the vertices connected to it. What is the expected number of time-steps before the particle hits state 3?

(*Note: $G(n, p)$ refers to the Erdos-Renyi random graph on n vertices where each edge exists independently with probability p, as introduced in lecture.*)

Let T_i be the expected hitting time to state 3 starting from state i. By symmetry, $T_1 = T_2$ and $T_3 = 0$.

With probability $1/4$ the particle stays in state 1 in the next step. With probability $(1/4 + 1/8)$ the particle goes to state 2. With probability $(1/4 + 1/8)$ it goes to state 3. So we have the single equation:

$$T_1 = 1 + (1/4)T_1 + (3/8)T_1$$

Solving this gives $T_1 = 8/3$.

Alternative Solution:
The probability of hitting 3 on the next step is $3/8$ from state 1 or state 2. Therefore, the process is equivalent to flipping a coin with a $3/8$ probability of coming up heads. The hitting time to state 3 therefore is just the expectation of a $Geom(3/8)$ random variable which is $8/3$.

(e) [3+3] Interarrival Times: A factory's production line outputs items according to a Poisson Process with rate λ.

1. If each item is defective with probability 1/3, what is the distribution of the time between the arrivals of two successive defective items? What is its expectation?

   $\text{Exp}(\lambda/3)$, $3/\lambda$

2. If every third item is defective, what is the distribution of the time between the arrivals two successive defective items? What is its expectation?

   $\text{Erlang}(3, \lambda)$, $3/\lambda$

(f) [2+2+2] Convergence in Probability: Let $X_1, ..., X_n$ be independent continuous random variables Uniform in [0,1]. Let $Y_n = \max(\{X_1, ..., X_n\})$, and $Z_n = \max(\{X_1, ..., X_n\} \backslash Y_n)$. In other words Y_n is the largest element of the n variables and Z_n is the second largest element.

1. What is $\mathbb{P}(Y_n < 1 - \epsilon)$, for some $\epsilon > 0$?

   $\mathbb{P}(Y_n < 1 - \epsilon) = (\mathbb{P}(X_1 < 1 - \epsilon))^n = (1 - \epsilon)^n$

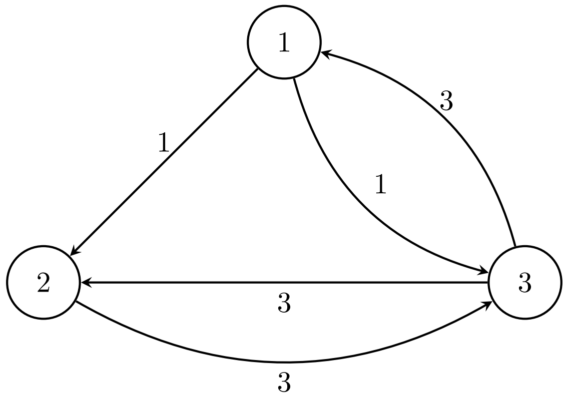
2. What is $\mathbb{P}(Z_n < 1 - \epsilon)$, for some $\epsilon > 0$?

   There are two cases: corresponding to whether the maximum of $X_1, ..., X_n$ is less or greater than $1 - \epsilon$. Hence, $\mathbb{P}(Z_n < 1 - \epsilon) = (1 - \epsilon)^n + n(1 - \epsilon)^{n-1}\epsilon$

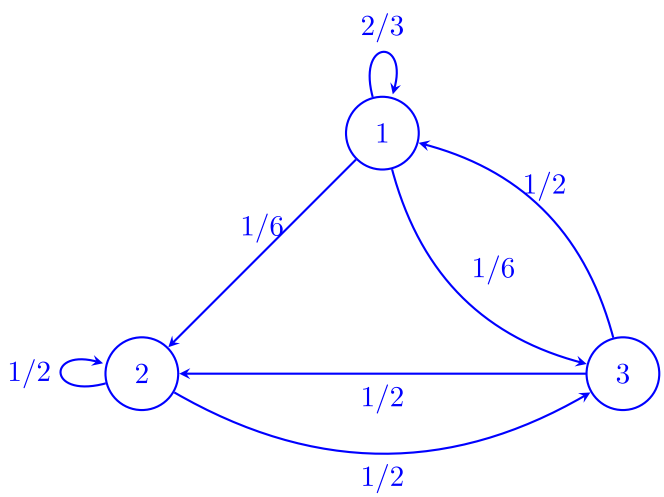
3. Show that $Y_n - Z_n$ converges in probability to 0.
   Hint: You may use the fact that if $A_n \xrightarrow{\mathbb{P}} a$ and $B_n \xrightarrow{\mathbb{P}} b$ for constants a and b, then $A_n + B_n \xrightarrow{\mathbb{P}} a + b$

   Let $\epsilon > 0$. Then $\mathbb{P}(Y_n < 1 - \epsilon) = (1 - \epsilon)^n \to 0$ as $n \to \infty$, so $Y_n \xrightarrow{\mathbb{P}} 1$. Then $\mathbb{P}(Z_n < 1 - \epsilon) = (1 - \epsilon)^n + n(1 - \epsilon)^{n-1}\epsilon \to 0$ as $n \to \infty$. The left term approaches 0 as before, and also $(1 - \epsilon)^{n-1}$ decreases exponentially faster than n grows, so that the right term also approaches 0. Thus, $Z_n \xrightarrow{\mathbb{P}} 1$. Using the hint yields the answer directly.

(g) [4] Simulated CTMC: Consider the Continuous-time Markov Chain (CTMC) shown below.



Construct a Discrete-time Markov Chain (DTMC) that has the same stationary distribution as the above chain.



(h) [6] Cascaded BEC channel: It is desired to transmit reliably over a composite channel comprising a cascade of two back-to-back BEC(p) channels (Binary Erasure Channels with erasure probability equal to p). What is the capacity (i.e. maximum rate at which you can transmit reliably) of this composite channel? (Recall that the capacity of a single BEC(p) channel is $(1-p)$ bits per channel use.)

The two cascaded BECs effectively form a BEC with erasure probability $p+p-p^2 = 2p-p^2$. Thus the new rate is $1-2p+p^2 = (1-p)^2$

Problem 2 [20]: Frisbee Attempts

The probability that a frisbee player will catch the frisbee depends on the results of her last two attempts, and is given by

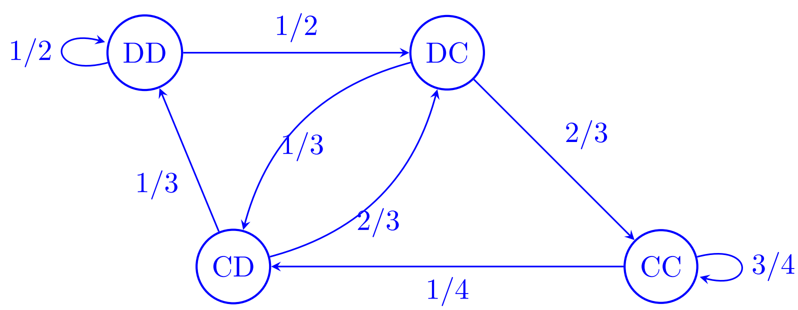
$$\mathbb{P}(\text{catch}) = \begin{cases} 1/2, & \text{if she dropped during both of her last two attempts} \\ 2/3, & \text{if she catches in exactly one of the two the last two attempts} \\ 3/4, & \text{if she catches in both of her last two attempts.} \end{cases}$$

(a) **[3]** Let $X_i = 1$ if you just caught the frisbee, and $X_i = 0$ otherwise; show with the help of an example that X_i is not a Markov chain.

It's not as it depends on last two catches. Based on this we can construct an example showing the Markov Property doesnt hold.

(b) **[5]** Let the states be denoted by ordered tuples (X_i, X_{i+1}); argue that this is a Markov chain. Draw the transition diagram.

Let the states be {DD, DC, CD, CC}, where DC means the second last attempt was a drop and the last attempt was a catch, etc. The underlying Markov chain is



(c) **[5]** Find the stationary distribution.

Solving for stationary distribution yields $\pi\,[\text{DD}, \text{DC}, \text{CD}, \text{CC}] = [1/8,\ 3/16,\ 3/16,\ 1/2]$.

(d) **[4]** Calculate the long-term fraction of times she catches the frisbee.

The limiting fraction of catches $= \frac{1}{2}[2\pi(CC) + \pi(CH) + \pi(HC)] = \frac{11}{16}$.

(e) **[3]** Is this chain reversible? Justify your answer.

No. Check states DD and DC, for example.

Problem 3 [32]: please do this thanks ray

Students ask logistical questions on Piazza according to a Poisson process of rate λ. Ray checks Piazza according to an independent Poisson process of rate μ. Every time Ray checks Piazza, he answers all logistical questions instantaneously.

(a) **[3]** Let's say Ray checked Piazza for the first time at time t, and there were n unanswered logistical questions. What is the expected time when the first question showed up?

This is the first order statistic of n uniform random variables in [0,t]. Hence, the expectation is $t/(n+1)$.

(b) **[4]** Ray just finished answering logistical questions, what is the expected number of logistical questions he will have to answer next time he checks Piazza?

We can represent the number of questions Ray will have to answer as the r.v. $X \sim Geom(\frac{\mu}{\mu+\lambda}) - 1$ (same as hw9 q1b). Thus, the expected number of questions he will have to answer is $((\mu+\lambda)/\mu) - 1 = \lambda/\mu$. .

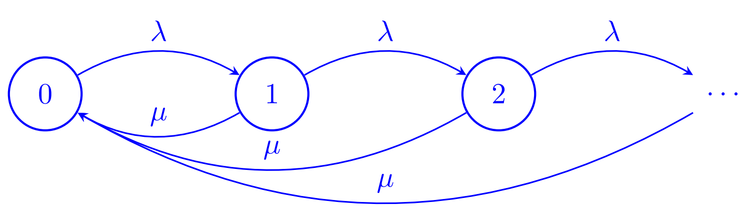
Alternative Solution:

Let T be the time until Ray checks piazza again. Then $\mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X|T)) = \mathbb{E}(\lambda T) = \lambda\mathbb{E}(T) = \lambda/\mu$

(c) **[5]** What is the expected number of times Ray will check Piazza before he is greeted with a nightmare of $\geq n$ unanswered logistical questions (including the time he's greeted with the nightmare)?

We can model this as a $Geom(p)$ random variable with $p = \left(\frac{\lambda}{\lambda+\mu}\right)^n$. Thus, the expected number of times until this event occurs is $\left(\frac{\lambda+\mu}{\lambda}\right)^n$.

(d) **[3]** We can model the number of outstanding logistical requests as a CTMC with the natural numbers as the state space (0, 1, 2, ...). Draw this CTMC.



(e) **[6]** Find the long term fraction of time for which there are no outstanding logistical questions.

We can solve this using the balance equations of the states starting from 1. We know that $(\lambda+\mu)\pi_1 = \lambda\pi_0$, and $(\lambda+\mu)\pi_2 = \lambda\pi_1$ and so on. By induction, we see that $\pi_i = \left(\frac{\lambda}{\lambda+\mu}\right)^i \pi_0$ for any state i. We just need to normalize this to 1 to formulate a valid stationary distribution. Since this is a geometric series, we can set $\pi_0 = \frac{\mu}{\lambda+\mu}$ to have the series converge to 1.

(f) **[3]** Is this CTMC positive recurrent, null recurrent or transient? (If it depends on λ, μ, specify how so).

It is positive recurrent for all λ, μ as a stationary distribution exists regardless of the values of λ, μ.

(g) **[8]** A student just posted the third logistical question since Ray's last check. What is the expected time until the next moment when there are exactly two outstanding logistical questions?

This is just the expected time until Ray's next check ($\frac{1}{\mu}$) + the expected time for the next question to be posted after that ($\frac{1}{\lambda}$) + the expected hitting time to 2 from 1($\beta_2(1)$).
To calculate $\beta_2(1)$, we have

$$\begin{aligned}
\beta_2(1) &= \frac{1}{\lambda+\mu} + \frac{\mu}{\lambda+\mu}\beta_2(0) \\
&= \frac{1}{\lambda+\mu} + \frac{\mu}{\lambda+\mu}(1/\lambda + \beta_2(1)) \\
\frac{\lambda}{\lambda+\mu}\beta_2(1) &= \frac{1+\frac{\mu}{\lambda}}{\lambda+\mu} \\
\beta_2(1) &= \frac{\lambda+\mu}{\lambda^2}
\end{aligned}$$

So the final answer is

$$\frac{1}{\mu} + \frac{1}{\lambda} + \frac{\lambda+\mu}{\lambda^2} = \frac{(\lambda+\mu)^2}{\lambda^2\mu}$$

Problem 4 [18]: Plants vs. Zombies

Efe decides the current version of Plants vs. Zombies is too easy for him, so he undertakes the task of writing a new version of the game. He's currently writing the randomizer for zombies entering the screen, and needs your help to analyze how many zombies will appear on each level. Suppose that at the beginning of each level Efe creates $Y \sim$ Poisson(μ) zombie generators, and each zombie generator independently creates zombies according to a Poisson Process with parameter λ. Each level lasts for T seconds, where T is fixed. Let N be the number of zombies generated in a given level.

(a) **[6]** Find $\mathbb{E}[N]$.

$\mathbb{E}[N] = \mathbb{E}[\mathbb{E}[N|Y]]$
Conditioned on Y, we can merge the individual rate λ processes to get a Poisson process with rate $Y\lambda$.
Thus, conditioned on Y, $N \sim$ Pois($Y\lambda T$).
Thus $\mathbb{E}[N|Y] = YT\lambda$
From that, we have the final result as $T\lambda\mu$

(b) **[6]** Find var(N).

We use the Law of Total Variance.

$$\text{var}(N) = \mathbb{E}[\text{var}(N|Y)] + \text{var}(\mathbb{E}[N|Y])$$

Note that $N|Y \sim$ Poisson(λTY)

$$\begin{aligned}\text{var}(N) &= \mathbb{E}[\lambda TY] + \text{var}(\lambda TY)\\ &= \lambda T\mathbb{E}[Y] + \lambda^2 T^2 \text{var}(Y)\\ &= \lambda\mu T + \lambda^2 T^2 \mu\\ &= \lambda\mu T(1 + \lambda T)\end{aligned}$$

(c) **[6]** Now suppose that Efe runs his simulation for 100 levels. Assume you are told that $\lambda < 2$ zombies/sec and $\mu < 5$. Using the CLT and $\sum_{i=1}^{100} N_i/n$ as an estimator, construct a 95% confidence interval for $\mathbb{E}[N]$.
Hint: For $Z \sim \mathcal{N}(0,1), \mathbb{P}(-1.96 \le Z \le 1.96) = 0.95$.

Let N_i be the number of zombies created on a level. We want to use the CLT to find a confidence interval for $\mathbb{E}[N]$.
The CLT states that $\frac{A_n - \mathbb{E}[N]}{\sigma_N^2/n} \sim \mathcal{N}(0,1)$ where the sample average $A_n = \frac{\sum_{i=1}^{n} N_i}{n}$. We have already computed σ_N^2 in part (b).
Using the hint, a 95% confidence interval for $\mathbb{E}[N]$ is $(A_n - 1.96\sqrt{\frac{\sigma_N^2}{n}}, A_n + 1.96\sqrt{\frac{\sigma_N^2}{n}})$
Plugging in $n = 100$ gives us

$$(A_n - 0.196\sqrt{\lambda\mu T(1+\lambda T)}, A_n + 0.196\sqrt{\lambda\mu T(1+\lambda T)})$$

Now we just need to use our upper bounds on λ and μ to get the result -

$$(A_n - 0.196\sqrt{20T^2 + 10T}, A_n + 0.196\sqrt{20T^2 + 10T})$$