EECS 126 Probability and Random Processes University of California, Berkeley: Spring 2018
Kannan Ramchandran March 21, 2018

Midterm 2

Last Name	First Name	SID

- You have 10 minutes to read the exam and 105 minutes to complete this exam.
- The maximum you can score is 125, but 100 points is considered perfect.
- The exam is not open book, but you are allowed to consult the cheat sheet that we provide. No calculators or phones.
- No form of collaboration between the students is allowed. If you are caught cheating, you may fail the course and face disciplinary consequences.
- Show all work to get any partial credit.
- Take into account the points that may be earned for each problem when splitting your time between the problems.

Problem	points earned	out of
Problem 1		48
Problem 2		25
Problem 3		25
Problem 4		27
Total		100 (+25)

Problem 1: Answer these questions briefly but clearly. [48]

(a) [6] Let $X_1, X_2, \ldots$ be a sequence of i.i.d. Uniform($[0,1]$) random variables and define $Y_n = \max\{X_1, \ldots, X_n\}$. Prove that $Y_n \xrightarrow{\mathbb{P}} 1$.

Fix $\epsilon > 0$. Then, $\mathbb{P}(Y_n < 1-\epsilon) = \mathbb{P}(X_1 < 1-\epsilon)^n = (1-\epsilon)^n \to 0$ as $n \to \infty$, so $Y_n \to 1$ in probability.

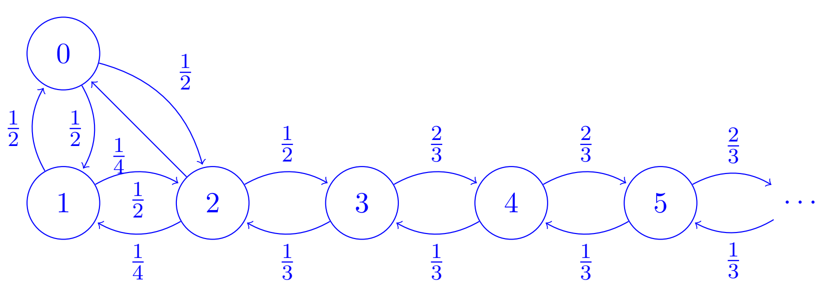
(b) [6] Consider a Poisson process $(N(t),\ t \geq 0)$ with rate λ. Does $n^{-1}N(n)$ converge almost surely as $n \to \infty$, and if so, to what? Explain carefully.

Yes, we can write $N(n) = \sum_{i=1}^{n}[N(i) - N(i-1)]$ and $N(1) - N(0), \ldots, N(n) - N(n-1)$ are i.i.d. Poisson(λ) random variables (because of the independent stationary increments property of the Poisson property), so by the usual SLLN we get $n^{-1}N(n) \to \mathbb{E}[N(1)] = \lambda$ almost surely as $n \to \infty$.

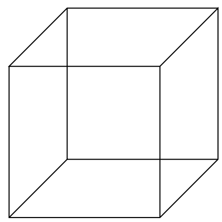
(c) [6] Let $S_N = X_1 + \cdots + X_N$, where $N \sim$ Geometric(p) and $X_1, X_2, \ldots$ are i.i.d. Exponential(λ) random variables. Using Poisson splitting find the distribution of S_N.

If we take a Poisson process of rate λ and split it with probability p, then $X_1, X_2, X_3, \ldots$ represent the interarrival times of the original Poisson process, and N represents the number of arrivals needed before an arrival is sent to the split process. Thus, S_N is the time of the first arrival in the split process, and since the split process is a Poisson process of rate $p\lambda$, it follows that $S_N \sim$ Exponential($p\lambda$).

(d) [6] Construct an irreducible, transient, aperiodic Markov chain with no self-loops.



(e) [6] Consider a 3-dimensional hypercube: the vertices are strings in $\{0,1\}^3$ where two vertices are connected by an edge if and only if they differ by exactly one bit. Start a random walk on the hypercube with starting vertex $X_0 = 110$. At each discrete time step, choose an edge leaving the current vertex uniformly at random and take this edge to the next vertex. What is the probability that the random walk hits 000 before 111?



By symmetry, the probability of absorption in 000 before 111 is the same for vertices 110, 101, and 011; call this q. Similarly, the probability of absorption is the same for vertices 001, 010, and 100; call this p. Then, first step analysis yields the equations

$$
\begin{aligned}
p &= \frac{1}{3} + \frac{2}{3} \cdot q, \\
q &= \frac{1}{3} \cdot 0 + \frac{2}{3} \cdot p.
\end{aligned}
$$

Solving this system gives $q = 2/5$.

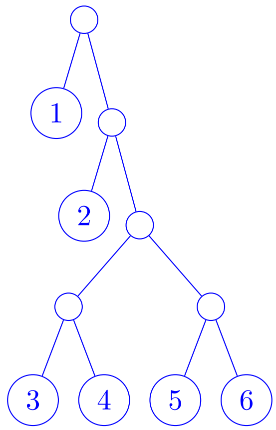
(f) [**6**] There are 6 people visiting a hospital and we know that exactly one of them is sick. The probability p_i that the i-th person is sick is given by

$$(p_1, p_2, p_3, p_4, p_5, p_6) = (0.5, 0.25, 0.1, 0.05, 0.05, 0.05).$$

The sickness can be diagnosed by a blood test.

For the first sample, you mix the blood of some of the people and test the mixture, noting that if none of the people in the sample are sick, the test will come up negative. You proceed, mixing and testing, stopping when the sick person has been determined. Give a strategy for mixing the blood of the people, for which the expected number of tests required is at most 2.

Let X be the index of the person who is sick. For each subset $S \subseteq \{1, \ldots, 6\}$, mixing the blood of the people in S and getting the blood tested is equivalent to asking the question "is $X \in S$?", and the optimal sequence of questions to ask is given by the Huffman encoding of the distribution of X. Here is a Huffman encoding tree for the distribution of X (it is not unique):

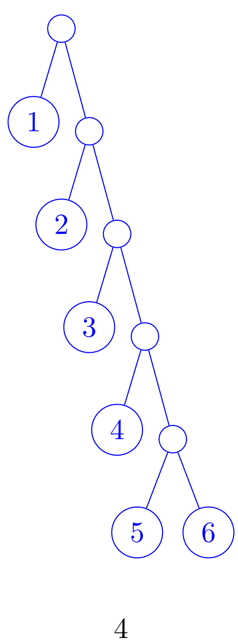


The tree corresponds to the following strategy.

- Ask if $X \in \{1\}$.
  - If the answer is yes, then person 1 is sick.
  - If the answer is no, ask if $X \in \{2\}$.
    - If the answer is yes, then person 2 is sick.
    - If the answer is no, ask if $X \in \{3, 4\}$.
      - If the answer is yes, then ask if $X \in \{3\}$.
        - If the answer is yes, then person 3 is sick.
        - If the answer is no, then person 4 is sick.
      - If the answer is no, then ask if $X \in \{5\}$.
        - If the answer is yes, then person 5 is sick.
        - If the answer is no, then person 6 is sick.

The expected number of questions for this strategy is $1 \cdot 0.5 + 2 \cdot 0.25 + 4 \cdot 0.25 = 2$.

Here is a curiosity. Another optimal Huffman encoding is given by the following tree:

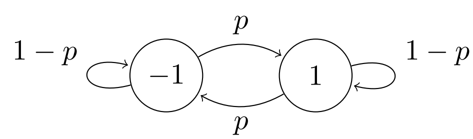


This corresponds to sequentially testing the first person, then the second person, etc. However, once we have tested the first five people, we do not need to carry out the last test—by process of elimination, if none of the first five people were sick, then the last person must be sick. Therefore, the expected number of tests is only $0.5+2\cdot 0.25+3\cdot 0.1+4\cdot 0.05+5\cdot 0.05 = 1.75$.

(g) [6] Consider a Markov chain on the state space $\{-1,1\}$ such that $P(-1,-1) = P(1,1) = 1-p$ and $P(-1,1) = P(1,-1) = p$ where $p \in (0,1)$. Our goal is to estimate the unknown parameter p. For $i = 1,\dots,n$, suppose that we observe $Y_i := X_i X_{i+1}$, and we will use the estimator

$$\hat{p} = \frac{\sum_{i=1}^n (1-Y_i)}{2n}.$$

Using the CLT, give a 95% approximate confidence interval for p using the estimator $\hat{p}$.



First, observe that $\mathbb{E}[\hat{p}] = p$.

Since Y_1 has variance $1-(1-2p)^2 = 4p(1-p) \le 1$, then

$$\operatorname{var}\hat{p} \le \frac{1}{4n}.$$

Using this upper bound on the variance, since $\approx 95\%$ of the area under the Gaussian curve is within ± 2 standard deviations, then the confidence interval for p is

$$\Big(\hat{p} - \frac{1}{\sqrt{n}}, \hat{p} + \frac{1}{\sqrt{n}}\Big).$$

(h) [6] Recall that for a sequence $X_1,\dots,X_n$ of i.i.d. Bernoulli(p) random variables, the typical set, with parameter $\epsilon > 0$, can be written as

$$A_\epsilon^{(n)} = \left\{(x_1,\dots,x_n) \in \{0,1\}^n : \left|\sum_{i=1}^n x_i - np\right| \le n\epsilon\right\}.$$

For $n = 10,\ p = 0.6$ and $\epsilon = 0.1$, calculate $\left|A_{0.1}^{(10)}\right|$.

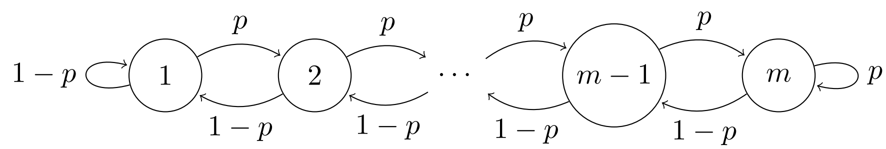
$$A_{0.1}^{(10)} = \left\{(x_1,\dots,x_{10}) \in \{0,1\}^{10} : 5 \le \sum_{i=1}^{10} x_i \le 7\right\}.$$

Therefore

$$\left|A_{0.1}^{(10)}\right| = \binom{10}{5} + \binom{10}{6} + \binom{10}{7}.$$

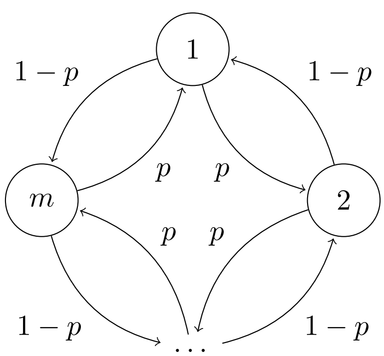
Problem 2: Reversible or Not? [25]

(a) [6] Consider a random walk on $\{1,\ldots,m\}$ given by $P(i,i+1)=p$, $P(i+1,i)=1-p$ for all $i=1,\ldots,m-1$, and $P(1,1)=1-p$, $P(m,m)=p$. Assume that $p\notin\{0,1\}$. Which of the following option(s) is/are true about the chain? (i) It is irreducible. (ii) It is aperiodic. (iii) It is reversible.



It is irreducible and aperiodic (since it has a self-loop). It is reversible because its graph is a tree (a line, in fact).

(b) [6] Consider a modification of the chain in (b) where the random walk is on a circle: now, $P(i,i+1 \bmod m)=p$ and $P(i,i-1 \bmod m)=1-p$ for all $i=1,\ldots,m$. Assume $p\notin\{0,1\}$. Which of the following option(s) is/are true about the chain? (i) It is irreducible. (ii) It is aperiodic.



By the assumptions on p, the chain is irreducible; it is aperiodic if m is odd, and it is periodic (with period 2) if m is even.

(c) [6] Does the chain have a unique stationary distribution, and if so, what is it? Under what conditions does it converge to the stationary distribution regardless of the initial distribution?

Yes, a unique stationary distribution exists because the chain is irreducible and by symmetry it is uniform on the state space. The chain converges to the stationary distribution if it is aperiodic (m odd).

(d) **[7]** For what values of p is the chain reversible? (Assume $p \notin \{0, 1\}$.)

Solution 1:

We already know from part (c) that the stationary probabilities are given by

$$\pi(i) = \frac{1}{m}, \text{ for } i = 1, \dots, m,$$

so the detailed balance equation can be written as

$$\begin{aligned} P(m, 1) &= P(1, m), \\ P(i, i+1) &= P(i+1, i), \text{ for } i = 1, \dots, m-1, \end{aligned}$$

which is true if and only if

$$\begin{aligned} p &= 1 - p \\ p &= \frac{1}{2}. \end{aligned}$$

Solution 2:

The detailed balance equations can be written as

$$\begin{aligned} \pi(m)p &= \pi(1)(1-p), \\ \pi(i)p &= \pi(i+1)(1-p), \text{ for } i = 1, \dots, m-1, \end{aligned}$$

which are equivalent to

$$\begin{aligned} \pi(m) &= \frac{1-p}{p}\pi(1), \\ \pi(i) &= \left(\frac{p}{1-p}\right)^{i-1}\pi(1), \text{ for } i = 1, \dots, m-1, \end{aligned}$$

and from this formulation we see that they have a solution if and only if

$$\begin{aligned} \left(\frac{p}{1-p}\right)^{m-1} &= \frac{1-p}{p} \\ \left(\frac{p}{1-p}\right)^{m} &= 1 \\ \frac{p}{1-p} &= 1 \\ p &= \frac{1}{2}. \end{aligned}$$

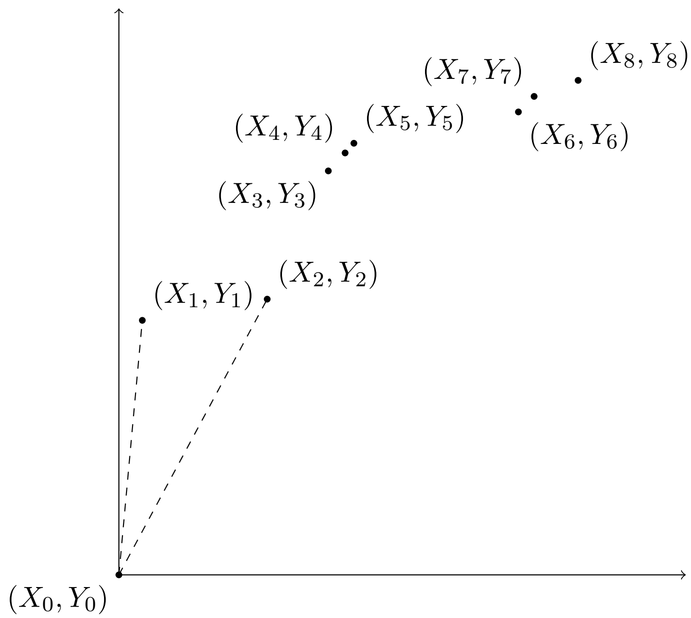
Problem 3: A New Dimension for the Poisson Process [25]

Hint: For all of the following parts, you will find it useful to think about the question in terms of the Poisson processes (recall the PPP: Poisson process perspective).

(a) [12] Let $(U_n)_{n\in\mathbb{N}}$ be i.i.d. Exponential(μ) and $(V_n)_{n\in\mathbb{N}}$ be i.i.d. Exponential(λ) independent of each other, and for each n let $X_n := \sum_{i=1}^n U_i$, $Y_n := \sum_{i=1}^n V_i$. Think of

$$(X_0, Y_0), (X_1, Y_1), (X_2, Y_2), \ldots$$

as a sequence of points in the plane $\mathbb{R}^2$. What is the probability that the number of points which land in the box $[0, 2]^2$ is at least 2? (You may leave your answer in terms of summations, no need to simplify. Do not leave your answer as an integral.)



Think of $(X_n)_{n\in\mathbb{N}}$ and $(Y_n)_{n\in\mathbb{N}}$ as independent Poisson processes. The number of points with X-coordinate ≤ 2 is Poisson with mean 2μ and similarly the number of points with Y-coordinate ≤ 2 is Poisson with mean 2λ, so the number of points which land in the box is the minimum of these two independent Poisson random variables. If we let A and B denote these independent Poisson random variables, then

$$\begin{aligned}\mathbb{P}(\min\{A, B\} \ge 2) &= \mathbb{P}(A \ge 2)\mathbb{P}(B \ge 2) \\ &= \big(1 - \exp(-2\mu) - 2\mu\exp(-2\mu)\big)\big(1 - \exp(-2\lambda) - 2\lambda\exp(-2\lambda)\big).\end{aligned}$$

(b) [4] What is the probability that the line connecting (X_0, Y_0) to (X_1, Y_1) has slope > 1?

This is the same as the probability that $Y_1 > X_1$, which happens with probability $\mu/(\mu + \lambda)$.

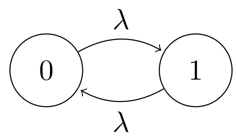
(c) **[9]** What is the probability that the line connecting (X_0, Y_0) to (X_2, Y_2) has slope > 1?

Hint: Think about the merged process.

This is asking for the probability that the second arrival from the Y process occurs after the second arrival from the X process. In the merged process, each arrival comes from the X process independently and with probability $\mu/(\mu+\lambda)$, so the probability that we see two arrivals from the X process before the second arrival from the Y process can be found by enumerating all of the possibilities: XX, YXX, XYX. Let $p := \mu/(\mu + \lambda)$. The probability we desire is $p^2 + 2(1 - p)p^2$.

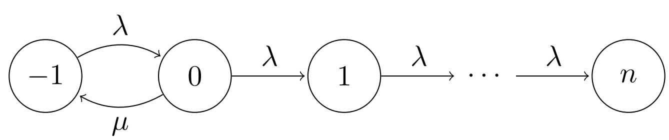
Problem 4: Time Passing Continuously, Markov! [27]

(a) [15] Consider a CTMC on state space $\{0, 1\}$ with $Q(0, 1) = Q(1, 0) = \lambda$. Given that at time 3 there have been a total of 6 times that the CTMC has switched states, what is the probability that the CTMC switched exactly 2 times by time 1?



The switches of the CTMC occur at the arrivals of a Poisson process so the 6 unordered arrival times are uniformly distributed on $[0, 3]$, so the probability is $\binom{6}{2}(2/3)^4(1/3)^2$.

(b) [10] Consider a CTMC on $\{-1, 0, 1, \ldots, n\}$ with $Q(0, -1) = \mu$, and for $i = -1, 0, 1, \ldots, n-1$, $Q(i, i + 1) = \lambda$. What is the expected time, starting from state 0, to hit state n?



At state 0, with probability $\lambda/(\mu+\lambda)$, the chain moves right, and if the chain moves right, then the additional expected time it takes to reach state n is $(n - 1)/\lambda$. If the chain moves left, then it takes on expectation $1/\lambda$ time to return to 0, so we get

$$\begin{aligned}
\mathbb{E}_0[T_n] &= \frac{1}{\mu + \lambda} + \frac{\lambda}{\mu + \lambda}\frac{n - 1}{\lambda} + \frac{\mu}{\mu + \lambda}\Big(\frac{1}{\lambda} + \mathbb{E}_0[T_n]\Big), \\
\frac{\lambda}{\mu + \lambda}\,\mathbb{E}_0[T_n] &= \frac{n + \mu/\lambda}{\mu + \lambda}, \\
\mathbb{E}_0[T_n] &= \frac{n\lambda + \mu}{\lambda^2}.
\end{aligned}$$

(c) [2] Was this exam too easy, too hard, or just right? There is no wrong answer.