School of Electrical Engineering, KAIST
EE210 Probability and Introductory Random Processes
2022 Spring

Final

Thursday, June 16, 2022
1:00–3:00 pm

NAME: ______________________________

Student ID: ______________________________

- Don't forget to put your name and student ID.
- **Record all your solutions in this answer booklet. Only this answer booklet will be considered in the grading of your exam.**
- Be sure to show all relevant work and reasoning. A correct answer does not guarantee full credit, and a wrong answer does not guarantee loss of credit. You should clearly but concisely indicate your reasoning.

Problem	Your score	Max score
1		15
2		15
3		15
4		15
Total		60

Problem 1 **(15 Points)**

Consider two discrete random variables X_n and Y_n whose PMFs are defined as

$$p_{X_n}(x) = \begin{cases} 1 - 1/n, & \text{for } x = 0, \\ 1/n, & \text{for } x = 1, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$p_{Y_n}(x) = \begin{cases} 1 - 1/n, & \text{for } x = 0, \\ 1/n, & \text{for } x = n, \\ 0, & \text{otherwise.} \end{cases}$$

a) (2 points) Find the expected value and variance of X_n and Y_n.

Answer:

$\mathbb{E}[X_n] = 1/n, \qquad \text{var}(X_n) = (n-1)/n^2,$

$\mathbb{E}[Y_n] = 1, \qquad \text{var}(Y_n) = n - 1.$

Reasoning for Problem 1(a):

$$\begin{aligned} \mathbb{E}[X_n] &= 0 \cdot (1 - 1/n) + 1 \cdot 1/n = 1/n, \\ \text{var}(X_n) &= (0 - 1/n)^2 \cdot (1 - 1/n) + (1 - 1/n)^2 \cdot (1/n) = (n-1)/n^2, \\ \mathbb{E}[Y_n] &= 0 \cdot (1 - 1/n) + n \cdot 1/n = 1, \\ \text{var}(Y_n) &= (0 - 1)^2 \cdot (1 - 1/n) + (n-1)^2 \cdot (1/n) = n - 1. \end{aligned}$$

b) (5 points) What does the Chebyshev inequality tell us about the convergence of X_n and Y_n?

Answer:

X_n converges to 0 in probability;

We cannot conclude anything about the converge of Y_n through Chebyshev's inequality.

Reasoning for Problem 1(b): Using Chebyshev's inequality, we have

$$\lim_{n\to\infty} \mathbb{P}\left(|X_n - 1/n| \geq \epsilon\right) \leq \lim_{n\to\infty} \frac{n-1}{n^2\epsilon^2} = 0.$$

It follows that X_n converges to 0 in probability.

For Y_n, Chebyshev inequality suggests that

$$\lim_{n\to\infty} \mathbb{P}\left(|Y_n - 1| \geq \epsilon\right) \leq \lim_{n\to\infty} \frac{n-1}{\epsilon^2} = \infty.$$

Thus, we cannot conclude anything about the converge of Y_n through Chebyshev's inequality.

c) (5 points) Is Y_n convergent in probability? If so, to what value? (Hint: A sequence Z_n converges in probability to a number a if for any $\epsilon > 0$, $\lim_{n\to\infty} \mathbb{P}(|Z_n - a| \geq \epsilon) = 0$.)

Answer:

Yes, it converges to 0.

Reasoning for Problem 1(c):

For any $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(|Y_n| \geq \epsilon) \leq \lim_{n\to\infty} \frac{1}{n} = 0,$$

so Y_n converges to zero in probability.

d) (3 points) If a sequence of random variables converges in probability to a value a, does the corresponding sequence of expected values converge to a? Prove or give a counter example.

Answer: A counter example is Y_n. Y_n converges to 0 in probability yet its expectation value is 1 for all n.

Problem 2 (15 Points)

Consider a Bernoulli process $X_1, X_2, X_3, \dots$ with unknown probability of arrival q, i.e., $\mathbb{P}(X_i = 1) = q$ for all i. Define the k-th interarrival time T_k as

$$T_1 = Y_1, \quad T_k = Y_k - Y_{k-1}, \quad k = 2, 3, \dots$$

where Y_k is the time of the k-th arrival. In this problem, we estimate the arrival probability q from observed interarrival times $(t_1, t_2, t_3, \dots)$. Assume q is sampled from the random variable Q which is uniformly distributed over $[0, 1]$.

You may find the following integral useful: For any non-negative integer k and m,

$$\int_0^1 q^k (1-q)^m dq = \frac{k!m!}{(k+m+1)!}.$$

a) (5 points) Find the PMF of T_1.

Answer:

$$p_{T_1}(t) = \frac{1}{(t+1)t} \quad \text{for } t = 1, 2, \dots$$

Reasoning for Problem 2(a): By the total probability theorem, we have

$$p_{T_1}(t) = \int_0^1 p_{T_1|Q}(t, q) f_Q(q) dq = \int_0^1 q(1-q)^{t-1} dq = \frac{1!(t-1)!}{(t+1)!} = \frac{1}{(t+1)t} \quad \text{for } t = 1, 2, \dots$$

b) (5 points) Compute the least mean squares (LMS) estimate of Q conditioned on the first arrival time $T_1 = t_1$.

Answer:

$$\mathbb{E}[Q|T_1 = t_1] = \frac{2}{t_1 + 2}$$

Reasoning for Problem 2(b): The LSE estimate of q conditioned on $T_1 = t_1$ is equal to $\mathbb{E}[Q|T_1 = t_1]$, which can be calculated as

$$\begin{aligned}
\mathbb{E}[Q|T_1 = t_1] &= \int_0^1 f_{Q|T_1}(q|t_1) q dq \\
&= \int_0^1 \frac{p_{T_1|Q}(t_1|q) f_Q(q)}{p_{T_1}(t)} q dq \\
&= \int_0^1 t_1(t_1+1) q (1-q)^{t_1-1} q dq \\
&= t_1(t_1+1) \int_0^1 q^2 (1-q)^{t_1-1} dq = t_1(t_1+1) \frac{2(t_1-1)!}{(t_1+2)!} = \frac{2}{t_1+2}.
\end{aligned}$$

c) (5 points) Compute the maximum a posteriori (MAP) estimate of Q given the first k arrival times, $(T_1, T_2, \ldots, T_k) = (t_1, t_2, \ldots, t_k)$.

Answer:

$$\arg\max_q f_{Q|T_1,\ldots,T_k}(q|t_1,\ldots,t_k) = \frac{k}{\sum_{i=1}^k t_i}$$

Reasoning for Problem 2(c): The posterior probability of Q given $(T_1, T_2, \ldots, T_k) = (t_1, t_2, \ldots, t_k)$ is

$$\begin{aligned} f_{Q|T_1,\ldots,T_k}(q|t_1,\ldots,t_k) &= \frac{f_Q(q)\prod_{i=1}^k p_{T_i|Q}(T_i = t_i|Q = q)}{C} \\ &= \frac{q^k(1-q)^{(\sum_{i=1}^k t_i)-k}}{C} \end{aligned}$$

where $C = \int_0^1 f_Q(q)\prod_{i=1}^k p_{T_i|Q}(T_i = t_i|Q = q)dq$, which does not depend on q. The MAP estimate of Q is the value q that maximizes $f_{Q|T_1,\ldots,T_k}(q|t_1,\ldots,t_k)$. We can find the MAP estimate of Q by finding q that makes the first derivative of $f_{Q|T_1,\ldots,T_k}(q|t_1,\ldots,t_k)$ with respect to q equal to 0, i.e.,

$$kq^{k-1}(1-q)^{(\sum_{i=1}^k t_i)-k} - \left(\sum_{i=1}^k t_i - k\right) q^k(1-q)^{(\sum_{i=1}^k t_i)-k-1} = 0,$$

or equivalently,

$$k(1-q) - \left(\sum_{i=1}^k t_i - k\right) q = 0,$$

which yields the MAP estiamte

$$\hat{q}_{\text{MAP}} = \frac{k}{\sum_{i=1}^k t_i}.$$

Problem 3 (15 Points)

Alice, Bob and Charlie run laps around a track, with the duration of each lap (in hours) being exponentially distributed with parameter $\lambda_A = 21, \lambda_B = 23$ and $\lambda_C = 24$, respectively. Assume that all lap durations are independent.

a) (5 points) Write down the PMF of the total number L of completed laps (3 runners combined) over the first hour.

Answer:

$$p_L(l) = \begin{cases} \frac{68^l e^{-68}}{l!}, & l = 0, 1, 2, \ldots \\ 0, & \text{otherwise.} \end{cases}$$

Reasoning for Problem 3(a): Given the problem statement, we can treat Alice, Bob and Charlie's running as 3 independent Poisson processes, where the arrivals correspond to lap completions and the arrival rates indicate the number of laps completed per hour. Since the three processes are independent, we can merge them to create a new process that captures the lap completions of all three runners. This merged process will have arrival rate $\lambda_M = \lambda_A + \lambda_B + \lambda_C = 68$. The total number of completed laps, L, over the first hour is then described by a Poisson PMF with $\lambda_M = 68$ and $\tau = 1$.

b) (5 points) What is the probability that Alice finishes her first lap before any of the others?

Answer:

$$\frac{\lambda_A}{\lambda_A + \lambda_B + \lambda_C} = \frac{21}{68}.$$

Reasoning for Problem 3(b): The event that Alice is the first to finish a lap is the same as the event that the first arrival in the merged process came from Alice's process. This probability is $\frac{\lambda_A}{\lambda_A+\lambda_B+\lambda_C}$.

c) (5 points) Suppose that the runners have been running for a very long time when you arrive at the track. What is the distribution of the duration T of Alice's current lap? (This includes the duration of that lap both before and after the time of your arrival.)

You may use that Erlang distribution for the time Y_k of the k-th arrival in the Poisson process of rate λ is $f_{Y_k}(y) = \frac{\lambda^k y^{k-1} e^{-\lambda y}}{(k-1)!}$, $y \geq 0$.

Answer:

$$f_T(t) = \begin{cases} 21^2 t e^{-21t}, & t \geq 0 \\ 0, & \text{otherwise.} \end{cases}$$

Reasoning for Problem 3(c): This is an instance of the random incident paradox, so the duration of Alice's current lap consists of the sum of the duration from the time of your arrival until Alice's next lap completion and the duration from the time of your arrival back to the time of Alice's previous lap completion. This is the sum of 2 independent exponential random variables with parameter $\lambda_A = 21$ (i.e., a second-order Erlang random variable).

Problem 4 (15 Points)

In this problem, we want to have a Markov chain that models the spread of a virus. Assume a population of n individuals. At the beginning of each day (say 7am), each individual is either infected or susceptible (not yet infected but capable of being infected by contacts of infected people). Suppose that each pair of people (i, j), $i \neq j$, independently comes into contact with one another during the daytime (7am to 7pm) with probability p. Whenever a susceptible individual comes into contact with an infected individual, the susceptible individual is infected right away. Assume that during overnight (7pm to 7am next day), any individual who has been infected for at least 24 hours will recover with probability $0 < q < 1$ and return to being susceptible, independently of everything else (i.e., assume that a newly infected individual will spend at least one night without recovery).

a) (5 points) Suppose that there are m infected individuals at one daybreak (7am). The number of total population is n. What is the PMF of the number of new infected individuals N at the end of the daytime (7pm)?

Answer:

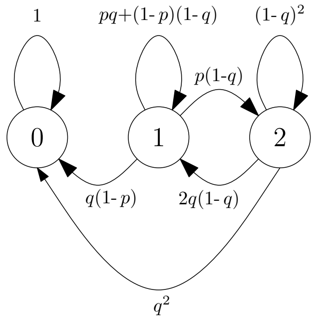
$$p_N(k) = \binom{n-m}{k} r^k (1-r)^{n-m-k}, \quad k = 0, 1, \ldots, n-m.$$

Reasoning for Problem 4(a): If m out of n individuals are infected, there are $n - m$ susceptible individuals. Each of these susceptible individuals will be independently infected during the daytime with probability $r = 1 - (1-p)^m$. Thus, the number of new infections N will be binomial random variable with parameters $n - m$ and r, so that

$$p_N(k) = \binom{n-m}{k} r^k (1-r)^{n-m-k}, \quad k = 0, 1, \ldots, n-m.$$

b) (5 points) Suppose that $n = 2$. Draw a Markov chain with states 0,1,2 (each of which indicates the number of infected individuals among the 2 at each daybreak) to model the spread of the virus.

Answer (without reasoning):



c) (5 points) Suppose that $n = 2$ and let's assume that the initial state is $X_0 = 1$ (the number of infected individuals at day 0 is equal to 1). Calculate the mean first passage time to the state 0, i.e., $t_1 = \mathbb{E}[\min\{n \geq 0 \text{ such that } X_n = 0\} | X_0 = 1]$ (the expected number of days to have 0 infected individual for the first time.)

Answer:

$$t_1 = \frac{1 - (1-q)^2 + p(1-q)}{(1 - pq - (1-p)(1-q))(1 - (1-q)^2) - 2pq(1-q)^2}.$$

Reasoning for Problem 4(c): We need to solve the following equations:

$$\begin{aligned} t_1 &= 1 + p_{11}t_1 + p_{12}t_2 \\ t_2 &= 1 + p_{21}t_1 + p_{22}t_2. \end{aligned}$$

We can find that

$$t_1 = \frac{1 - p_{22} + p_{12}}{(1 - p_{11})(1 - p_{22}) - p_{12}p_{21}}$$

By plugging in the transition probabilities,

$$t_1 = \frac{1 - (1-q)^2 + p(1-q)}{(1 - pq - (1-p)(1-q))(1 - (1-q)^2) - 2pq(1-q)^2}.$$