Final Part II

Thursday, July 2, 2020
1:00–3:15 pm

- Be sure to **show all relevant work and reasoning** in your answer sheet. 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.
- Please be clear in writing—we can't grade what we can't decipher!
- Don't forget to upload your answer sheet for
  - Problem 1 by 1:30pm
  - Problem 2 by 2:05pm
  - Problem 3 by 2:40pm
  - Problem 4 by 3:15pm

  through KLMS. The system will be automatically closed at each due time. If the system does not work, you should email it to `ee210b_20spring@kaist.ac.kr` by the due time. Late submissions will not be accepted/graded.

Problem 1 (15 Points)

Let's consider a signal transmission (communication) problem over a noisy channel. Our goal is to transmit a signal X through a noisy channel. The noisy channel adds a Gaussian noise $Z \sim N(0, \sigma^2)$ to the transmitted signal, so that the observed signal at the receiver is $Y = X + Z$. Let's assume that X is independent of Z and X can take value a with probability $p_a > 0$ and value b with probability $p_b > 0$ where $p_a + p_b = 1$ and $b > a$. Conditioned on $X = a$, the observed signal is $Y \sim N(a, \sigma^2)$ and conditioned on $X = b$, the observed signal is $Y \sim N(b, \sigma^2)$.

Given $Y = y$, we want to make a guess of the value for $X \in \{a, b\}$. Let's denote our guess as $G \in \{a, b\}$. Our goal is to minimize $\mathbf{P}(G \neq X)$, the probability that our guess is incorrect.

Let's consider the following strategy for the guessing: Compare two conditional probabilities $\mathbf{P}(X = a|Y = y)$ and $\mathbf{P}(X = b|Y = y)$ and say that our guess is $G = a$ if $\mathbf{P}(X = a|Y = y) \geq \mathbf{P}(X = b|Y = y)$ and that $G = b$ if $\mathbf{P}(X = a|Y = y) < \mathbf{P}(X = b|Y = y)$.

a) (5 points) Consider the case where $p_a = p_b = 1/2$. Specify the range of y such that our guess is $G = a$. (Hint: You may need to find the threshold r such that if $y \leq r$ your guess $G = a$.)

Solution: The range of y such that $\mathbf{P}(X = a|Y = y) \geq \mathbf{P}(X = b|Y = y)$ is

$$\frac{\mathbf{P}(X = a)\mathbf{P}(Y = y|X = a)}{\mathbf{P}(Y = y)} \geq \frac{\mathbf{P}(X = b)\mathbf{P}(Y = y|X = b)}{\mathbf{P}(Y = y)}$$

by Bayes rule.(**1 point**) Since $p_a = p_b = 1/2$, above inequality can be simplified as

$$\mathbf{P}(Y = y|X = a) \geq \mathbf{P}(Y = y|X = b). \qquad (\textbf{1 point})$$

Since $\mathbf{P}(Y = y|X = a) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-a)^2}{2\sigma^2}}$ and $\mathbf{P}(Y = y|X = b) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-b)^2}{2\sigma^2}}$,(**1 point**) the range of y where $\mathbf{P}(X = a|Y = y) \geq \mathbf{P}(X = b|Y = y)$ becomes

$$y \leq \frac{(b^2 - a^2)}{2(b - a)} = \frac{b + a}{2}. \qquad (\textbf{2 points})$$

b) (5 points) For the case $p_a = p_b = 1/2$, calculate the probability $\mathbf{P}(G \neq X)$, the probability that our guess is incorrect. Write down this probability in terms of the CDF for the standard normal. (Remind that the CDF for the standard normal is $\Phi(r) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{r} e^{-t^2/2} dt$.)

Solution: When we let $r := \frac{b+a}{2}$,

$$\begin{aligned}
\mathbf{P}(G \neq X) &= \frac{1}{2}\mathbf{P}(G \neq X|X = a) + \frac{1}{2}\mathbf{P}(G \neq X|X = b) && (\mathbf{1\ point})\\
&= \frac{1}{2}\mathbf{P}(Y \geq r|X = a) + \frac{1}{2}\mathbf{P}(Y \leq r|X = b) && (\mathbf{1\ point})\\
&= \frac{1}{2}\mathbf{P}\left(\frac{Y-a}{\sigma} \geq \frac{r-a}{\sigma}|X = a\right) + \frac{1}{2}\mathbf{P}\left(\frac{Y-b}{\sigma} \leq \frac{r-b}{\sigma}|X = b\right)\\
& && (\mathbf{1\ point})\\
&= \frac{1}{2}\left(1 - \mathbf{P}\left(\frac{Y-a}{\sigma} \leq \frac{b-a}{2\sigma}|X = a\right)\right)\\
&\quad + \frac{1}{2}\mathbf{P}\left(\frac{Y-b}{\sigma} \leq \frac{a-b}{2\sigma}|X = b\right)\\
&= \frac{1}{2}\left(1 - \Phi\left(\frac{b-a}{2\sigma}\right)\right) + \frac{1}{2}\left(\Phi\left(\frac{a-b}{2\sigma}\right)\right) && (\mathbf{2\ point})\\
&= \frac{1}{2}\left(1 - \Phi\left(\frac{b-a}{2\sigma}\right)\right) + \frac{1}{2}\left(1 - \Phi\left(\frac{b-a}{2\sigma}\right)\right)\\
&= 1 - \Phi\left(\frac{b-a}{2\sigma}\right)
\end{aligned}$$

c) (5 points) Consider a general $p_a, p_b > 0$, not necessarily $p_a = p_b = 1/2$. Specify the range of y such that our guess is $G = a$ in terms of p_a and p_b. Again, you may need to find the threshold r such that if $y \leq r$ your guess $G = a$. Describe how r should change as p_a increases (or p_b increases) while satisfying $p_a + p_b = 1$.

Solution: The range of y such that $\mathbf{P}(X = a|Y = y) \geq \mathbf{P}(X = b|Y = y)$ is

$$\frac{\mathbf{P}(X = a)\mathbf{P}(Y = y|X = a)}{\mathbf{P}(Y = y)} \geq \frac{\mathbf{P}(X = b)\mathbf{P}(Y = y|X = b)}{\mathbf{P}(Y = y)}, \qquad (\mathbf{1\ point})$$

which is equivalent to

$$p_a\mathbf{P}(Y = y|X = a) \geq p_b\mathbf{P}(Y = y|X = b). \qquad (\mathbf{1\ point})$$

Since $\mathbf{P}(Y = y|X = a) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(y-a)^2}{2\sigma^2}}$ and $\mathbf{P}(Y = y|X = b) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(y-b)^2}{2\sigma^2}}$,(**1 point**) the range of y where $\mathbf{P}(X = a|Y = y) \geq \mathbf{P}(X = b|Y = y)$ becomes

$$y \leq \frac{(b^2 - a^2) + 2\sigma^2 \ln\frac{p_a}{p_b}}{2(b-a)} = \frac{b+a}{2} + \frac{\sigma^2 \ln\frac{p_a}{p_b}}{b-a}. \qquad (\mathbf{1\ point})$$

As p_a increases, the threshold $r := \frac{(b^2-a^2)+2\sigma^2\ln\frac{p_a}{p_b}}{2(b-a)} = \frac{b+a}{2} + \frac{\sigma^2\ln\frac{p_a}{p_b}}{b-a}$ should move to the right and as p_b increases, the threshold r should move to the left.(**1 point**)

Problem 2 (15 Points)

Consider a Bernoulli process $X_1, X_2, X_3, \ldots$ 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

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

where Y_k is the time of the k-th arrival. In this problem, we estimate the arrival probability q from observed interarrival times $(t_1, t_2, t_3, \ldots)$.

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

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

Assume q is sampled from the random variable Q which is uniformly distributed over $[0, 1]$.

a) (5 points) Find the PMF of T_1, $p_{T_1}(t)$.

Solution: By the total probability theorem, we have

$$\begin{aligned} p_{T_1}(t) &= \int_0^1 p_{T_1|Q}(t, q) f_Q(q) dq = \int_0^1 q(1-q)^{t-1} dq \qquad (\textbf{2 point}) \\ &= \frac{1!(t-1)!}{(t+1)!} = \frac{1}{(t+1)t} \quad \text{for } t = 1, 2, \ldots \qquad (\textbf{3 point}) \end{aligned}$$

b) (5 points) Compute the least mean squares (LMS) estimate of Q conditioned on the first arrival time $T_1 = t_1$, i.e., find $\mathbb{E}[Q|T_1 = t_1]$.

Solution: 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

$$\begin{aligned} \mathbb{E}[Q|T_1 = t_1] &= \int_0^1 p_{Q|T_1}(q|t_1) q dq \\ &= \int_0^1 \frac{p_{T_1|Q}(t_1|q) f_Q(q)}{p_{T_1}(t)} q dq \qquad (\textbf{2 point}) \\ &= \int_0^1 t_1(t_1+1) q (1-q)^{t_1 - 1} q dq \\ &= t_1(t_1+1) \int_0^1 q^2 (1-q)^{t_1-1} dq = t_1(t_1+1) \frac{2(t_1-1)!}{(t_1+2)!} = \frac{2}{t_1+2}. \ (\textbf{3 point}) \end{aligned}$$

c) (5 points) Compute the maximum a posteriori (MAP) estimate of Q given the first k arrival times, $(T_1, T_2, \ldots, T_k) = (t_1, t_2, \ldots, t_k)$, i.e., find $\arg\max_q f_{Q|T_1,\ldots,T_k}(q|t_1, \ldots, t_k)$.

Solution: The posterior probability of Q given $(T_1, T_2, \ldots, T_k) = (t_1, t_2, \ldots, t_k)$ is

$$\begin{aligned} f_{Q|T_1,\ldots,T_k}(q|t_1, \ldots, t_k) &= \frac{f_Q(q) \prod_{i=1}^k p_{T_i|Q}(T_i = t_i | Q = q)}{C} \qquad (\textbf{2 point}) \\ &= \frac{q^k (1-q)^{(\sum_{i=1}^k t_i) - k}}{C} \qquad (\textbf{1 point}) \end{aligned}$$

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

$$\hat{q}_{\text{MAP}} = \frac{k}{\sum_{i=1}^k t_i}. \qquad \textbf{(2 point)}$$

Problem 3 (15 Points)

The voters in a given town arrive at the place of voting according to a Poisson process of rate $\lambda = 100$ voters per hour. The voters independently vote for candidate A and candidate B each with probability $1/2$. Assume that the voting starts at time 0 and continues indefinitely.

a) (5 points) Conditioned on that 1000 voters arrived during the first 10 hours of voting, find the probability that candidate A receives n of those votes.

Solution: We can consider the splitting of the original Poisson process with $\lambda = 100$ into two Poisson processes, each of which indicates the votes for candidate A and B, respectively, by splitting arrivals into two streams using independent coin flips of a fair coin. The first poisson process with $\lambda_A = 50$ indicates the arrival of voters who vote for candidate A and the other Poisson process with $\lambda_B = 50$ indicates arrival of voters who vote candidate B. The two Poisson processes are independent.

Since the arrivals are splitted into two processes with probability 1/2 and 1/2, conditioned on that 1000 voters arrived during the first 10 hours, the probability that candidate A receives n votes follows the binomial distribution with $n = 1000$ and $p = 1/2$. (**3 point**)
So that the probability is equal to

$$\binom{1000}{n}\left(\frac{1}{2}\right)^n. \qquad \textbf{(2 point)}$$

This can also be found by calculating $\frac{\mathbf{P}(N_A(10)=n, N_B(10)=1000-n)}{\mathbf{P}(N_A(10)+N_B(10)=1000)}$ where $N_A(\tau), N_B(\tau)$ is the number of arrivals for the Poisson process for A and B, respectively, for time duration of τ:

$$\frac{\mathbf{P}(N_A(10)=n, N_B(10)=1000-n)}{\mathbf{P}(N_A(10)+N_B(10)=1000)} = \frac{\frac{(50\cdot 10)^n e^{-50\cdot 10}}{n!}\,\frac{(50\cdot 10)^{1000-n} e^{-50\cdot 10}}{(1000-n)!}}{\frac{(100\cdot 10)^{1000} e^{-100\cdot 10}}{1000!}} = \binom{1000}{n}\left(\frac{1}{2}\right)^n.$$

For student who multiplied the conditional term : (**-2 point**)

b) (5 points) Let $T_{1,A}$ be the arrival of the first voter who votes for candidate A. Find the pdf of $T_{1,A}$, $f_{T_{1,A}}(t)$.

Solution: $T_{1,A}$ is the first arrival time of the Poisson process with rate $\lambda_A = 50$. (**3 point**) Its pdf is the exponential(λ_A), i.e.,

$$f_{T_{1,A}}(t) = 50e^{-50t}, \text{ for } t \geq 0. \qquad \textbf{(2 point)}$$

c) (5 points) Define V_B as the number of voters for candidate B who arrive before the first voter for A. Find the pmf of V_B.

Solution: The pmf of the number of voters for candidate B who arrive before the first voter for A follows the geometric distribution with $p = 1/2$, thus

$$p_{V_B}(k) = (1/2)^k \cdot 1/2 = (1/2)^{k+1}. \qquad \textbf{(5 point)}$$

We can also find this by calculating the pmf for V_B conditioned on $T_{1,A} = t_1$ and by using the total probability theory. Note that

$$p_{V_B|T_{1,A}}(k|t_1) = \frac{(50t_1)e^{-50t_1}}{k!}$$

By using the pdf of $T_{1,A}$ which is $f_{T_{1,A}} = 50e^{-50t}$ for $t \geq 0$, we get

$$\begin{aligned}
p_{V_B}(k) &= \int_0^\infty p_{V_B|T_{1,A}}(k|t_1) f_{T_{1,A}}(t_1) dt_1 (\mathbf{2\ point}) \\
&= \int_0^\infty \frac{(50t_1)^k e^{-50t_1}}{k!} 50e^{-50t_1} dt_1 (\mathbf{1\ point}) \\
&= \frac{(50)^{k+1}}{k!} \int_0^\infty {t_1}^k e^{-100t_1} dt_1 \\
&= \frac{(50)^{k+1}}{k!} \frac{k!}{100^{k+1}} = (1/2)^{k+1}. (\mathbf{2\ point})
\end{aligned}$$

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 each daybreak (7 am), 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 an infected individual comes into contact with a susceptible individual, the susceptible individual is infected right away. Assume that during overnight (7pm to 7am next day), any individual who has been infected will recover with probability $0 < q < 1$ and return to being susceptible, independently of everything else.

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

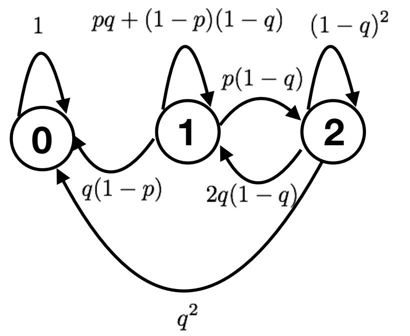
Solution: If m out of n individuals are infected, there are $n - m$ susceptible individuals. Each of these susceptible individuals will be independently infected during the daytime with probability $r = 1 - (1-p)^m$. Thus, the number of new infections N will be binomial random variable with parameters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

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

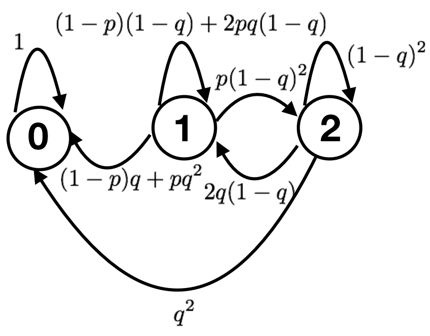
Partial point: $r = 1 - (1-p)^m$ (3 points),

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

Solution:



Assuming that a new infected individual cannot be recovered on the same day



Assuming that a new infected individual can be recovered on the same day

Partial point: it depends on the figure of the Markov chain you draw and the number of correct transition probabilities.

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.)

Solution: 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

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

Partial point: it depends on how you set up the equation to solve the problem. Even though the transition probabilities calculated in problem *b*) was wrong, I give you full points if the equation for t_1 is appropriate.

Every description of partial points is explained in KLMS.