Probability and Stochastic Processes (EL6303) October 27, 2017
NYU Tandon School of Engineering, Fall 2017
Instructor: *Dr. Elza Erkip*

Midterm Solutions

1. *(15 points)* Suppose the random variable N_t represents the number of e-mails you receive up to time t. We assume N_t has Poisson distribution with parameter μt, that is

$$P(N_t = k) = P(\mathrm{k\ emails\ up\ to\ time\ t}) = \mathrm{e}^{-\mu \mathrm{t}} \frac{(\mu \mathrm{t})^{\mathrm{k}}}{\mathrm{k}!}, \quad \mathrm{k} = 0, 1, 2, \ldots$$

We define T_1 as the arrival time of the first e-mail.

(a) Suppose you know $T_1 > t$. What does this imply about N_t?

(b) Find the cumulative distribution function (cdf) of T_1.

(c) Find the probability density function (pdf) of T_1.

<u>**Solution:**</u>

(a) $T_1 > t$ means that no emails has arrived up to time t, which means $N_t = 0$.

(b)
$$F_{T_1}(t) \triangleq P(T_1 \leq t) = P(N_t \neq 0) = 1 - P(N_t = 0) = 1 - e^{-\mu t}$$

(c)
$$f_{T_1}(t) = \frac{dF_{T_1}(t)}{dt} = \mu e^{-\mu t}.$$

2. *(22 points)* Consider the following communication system. Transmitter sends X, which takes values from the set $\{0, \ldots, M-1\}, M > 2$ and the receiver observes a noisy version, Y such that

$$Y = (X + Z)(\text{mod } M),$$

where mod M represents modulo M operation. We assume Z is independent of X and has Bernoulli(p) distribution with $P(Z = 0) = p$.

(a) Suppose $X \sim \text{Uniform}\{0, \ldots, M-1\}$.

i. Find the probability mass function (pmf) of Y.

ii. Find the conditional pmf of X given Y, that is $P(X = k|Y = j), k, j = 0, \ldots, M-1$.

(b) Suppose M is even and X is uniformly distributed over all *even* numbers between 0 and $M-1$, that is $P(X = i) = 2/M$ if i even, and zero otherwise .

i. Find the probability mass function (pmf) of Y.

ii. Find the conditional pmf of X given Y, that is $P(X = k|Y = j), k, j = 0, \ldots, M-1$.

(c) Considering your answers for the previous parts, which distribution of X allows you to better identify the transmitted value from the received one?

Solution:

(a) i.

$$P(Y = j) = P(X = j, Z = 0) + P(X = j-1, Z = 1) = \frac{1}{M}p + \frac{1}{M}(1-p) = \frac{1}{M},$$

for $j = 0, ..., M-1$. Note that the subtraction is in modulo M which gives $(0-1)(\text{mod}M) = M-1$.

ii.

$$P(X = k|Y = j) = \frac{P(X = k, Y = j)}{P(Y = j)}$$

$$= \begin{cases} \frac{P(X=k,Z=0)}{P(Y=j)} = \frac{P(X=k)P(Z=0)}{P(Y=j)} = \frac{p/M}{1/M} = p & \text{, if } k = j, \\ \frac{P(X=k,Z=1)}{P(Y=j)} = \frac{P(X=k)P(Z=1)}{P(Y=j)} = \frac{(1-p)/M}{1/M} = 1-p & \text{, if } k = j - 1\text{mod M}, \\ 0 & \text{otherwise.} \end{cases}$$

(b) i.

$$P(Y = j) = \begin{cases} P(X = j, Z = 0) = \frac{2p}{M}, & \text{if } j \text{ is even,} \\ P(X = j-1, Z = 1) = \frac{2(1-p)}{M}, & \text{if } j \text{ is odd.} \end{cases}$$

ii.

$$P(X = k|Y = j) = \frac{P(X = k, Y = j)}{P(Y = j)}.$$

If $Y = j$ is even, then we have:

$$P(X = k|Y = j) = \begin{cases} \frac{P(X=k,Z=0)}{P(Y=j)} = \frac{P(X=k)P(Z=0)}{P(Y=j)} = \frac{2p/M}{2p/M} = 1 & \text{, if } k = j, \\ 0 & \text{otherwise.} \end{cases}$$

On the other hand, if $Y = j$ is odd, we have:

$$P(X = k|Y = j) = \begin{cases} \frac{P(X=k,Z=1)}{P(Y=j)} = \frac{P(X=k)P(Z=1)}{P(Y=j)} = \frac{2(1-p)/M}{2(1-p)/M} = 1 & \text{, if } k = j - 1, \\ 0 & \text{otherwise.} \end{cases}$$

Hence, observing $Y = j$ will enable the receiver to identify the transmitted value of X correctly with probability 1.

(c) The distribution in part (b) since it identifies the transmitted value with probability 1.

3. *(22 points)* Three types of jobs arrive at a computer server:

- With probability p_1, a type 1 job arrives, requiring an exponentially distributed service time T_1 with parameter λ, i.e., $f_{T_1}(t_1) = \lambda e^{-\lambda t_1}$,
- With probability p_2, a type 2 job arrives, requiring a uniformly distributed service time T_2 in the interval $[0, T]$,
- With probability $p_3 = 1 - p_1 - p_2$, a type 3 job arrives, requiring a fixed service time $T_3 = K$, where $K < T$.

(a) Find $P(T_1 \leq K)$.

(b) Find $P(T_2 \leq K)$.

(c) Find the probability that service time of an arbitrary job, arriving according to the distribution (p_1, p_2, p_3), is less than or equal to K.

(d) If service time of a job is known to be less than or equal to K, what is the probability that it is a type 1 job?

Solution:

(a)

$$P(T_1 \leq K) = \int_0^K f_{T_1}(t_1)dt_1 = \int_0^K \lambda e^{-\lambda t_1} dt_1 = -e^{-\lambda t_1}]_{t_1=0}^{t_1=K} = 1 - e^{-\lambda K}.$$

(b)

$$P(T_2 \leq K) = \int_0^K \frac{1}{T} dt_1 = \frac{K}{T}.$$

(c) Using the law of total probability, we have:

$$P(S.T \leq K) = \sum_{i=1}^{3} P(S.T \leq K|\text{type } i) \cdot P(\text{type } i) = p_1(1 - e^{-\lambda K}) + p_2 \frac{K}{T} + p_3.$$

Where $S.T$ is service time.

(d) Using the Bayes' rule, we have:

$$P(\text{type } 1|S.T \leq K) = \frac{P(S.T \leq K|\text{type } 1)P(\text{type } 1)}{P(S.T \leq K)} = \frac{p_1(1 - e^{-\lambda K})}{p_1(1 - e^{-\lambda K}) + p_2 \frac{K}{S.T} + p_3}$$

4. *(20 points)*

(a) Suppose we have a communication link which introduces delay, represented by the *discrete* random variable D. We assume that D is uniformly distributed over $\{0, 1, 2, 3\}$. Find the mean of D.

(b) Now suppose we have access to two links, and you decide to send your file over both links at the same time. This way, you can observe a delay $X = \min(D_1, D_2)$, where D_1 and D_2 represent the random delays on links 1 and 2 respectively. We assume D_1 and D_2 are independent and uniformly distributed over $\{0, 1, 2, 3\}$.

i. Find the probability mass function (pmf) of X.

ii. Find the mean of X.

(c) Alternatively, you can split your file across the two links. This way, each link has a lower load, and less delay. To model this, we now assume that link delays are represented as D_3 and D_4, which are independent and uniformly distributed over $\{0, 1\}$. But in this case, you have to wait for each part of your file to arrive, so your delay is $Y = \max(D_3, D_4)$.

i. Find the pmf of Y.

ii. Find the mean of Y.

(d) Which method results in lower expected delay?

<u>Solution:</u>

(a) D is uniformly distributed over {0, 1, 2 ,3 } so we have

$$P(D = 0) = P(D = 1) = P(D = 2) = P(D = 3) = \frac{1}{4}$$

$$E[D] = \sum_{i=0}^{3} i \times P(D = i) = \frac{0 + 1 + 2 + 3}{4} = \frac{6}{4} = \frac{3}{2}.$$

(b) $X = min(D_1, D_2)$ where D_1 and D_2 are independent and uniformly distributed over {0,1,2,3}

i. We can see that X can only take values in {0,1,2,3}.

$$\begin{aligned} P(X=0) &= P(D_1 = 0 + D_2 = 0) \\ &= P(D_1 = 0) + P(D_2 = 0) - P(D_1 = 0, D_2 = 0) \\ &= \frac{1}{4} + \frac{1}{4} - \frac{1}{16} \\ &= \frac{7}{16} \end{aligned}$$

$$\begin{aligned} P(X=1) &= P((D_1 = 1, D_2 \geq 1) + (D_1 \geq 1, D_2 = 1)) \\ &= P(D_1 = 1, D_2 \geq 1) + P(D_1 \geq 1, D_2 = 1) - P(D_1 = 1, D_2 = 1) \\ &= \frac{1}{4} \times \frac{3}{4} + \frac{3}{4} \times \frac{1}{4} - \frac{1}{4} \times \frac{1}{4} \\ &= \frac{5}{16} \end{aligned}$$

$$\begin{aligned} P(X=2) &= P((D_1 = 2, D_2 \geq 2) + (D_1 \geq 2, D_2 = 2)) \\ &= P(D_1 = 2, D_2 \geq 2) + P(D_1 \geq 2, D_2 = 2) - P(D_1 = 2, D_2 = 2) \\ &= \frac{1}{4} \times \frac{2}{4} + \frac{2}{4} \times \frac{1}{4} - \frac{1}{4} \times \frac{1}{4} \\ &= \frac{3}{16} \end{aligned}$$

$$\begin{aligned} P(X=3) &= P(D_1 = 3, D_2 = 3) = \frac{1}{4} \times \frac{1}{4} \\ &= \frac{1}{16} \end{aligned}$$

ii. Calculating $E[X]$:

$$\begin{aligned} E[X] &= \sum_{i=0}^{3} i \times P(X = i) \\ &= 0 \times \frac{7}{16} + 1 \times \frac{5}{16} + 2 \times \frac{3}{16} + 3 \times \frac{1}{16} \\ &= \frac{7}{8} \end{aligned}$$

(c) $Y = \max(D_2, D_3)$ where D_2 and D_3 are independent and uniformly distributed over set {1,2}

i. Y can only take values 0 and 1.

$$\begin{aligned} P(Y=0) &= P(D_2=0, D_3=0)) \\ &= \frac{1}{2} \times \frac{1}{2} \\ &= \frac{1}{4} \end{aligned}$$

$$\begin{aligned} P(Y=1) &= 1 - P(Y=0) \\ &= \frac{3}{4} \end{aligned}$$

ii.

$$\begin{aligned} E[Y] &= \sum_{i=0}^{1} i \times P(Y=i) \\ &= 0 \times \frac{1}{4} + 1 \times \frac{3}{4} \\ &= \frac{3}{4} \end{aligned}$$

(d) We can see that $E[Y] = \frac{3}{4}$ is less than $E[X] = \frac{7}{8}$ and so the method in part (c) causes less expected delay.

5. *(22 points)* You do not have time to cook since you are studying for your midterm. You can order food from Elza's Diner (restaurant 1) or Erkip Burger (restaurant 2). The time it takes for the Diner food to arrive is represented by a random variable T_1, for the Burger food to arrive by a random variable T_2. We assume

$$T_1 \sim \text{Uniform}(0, 30), T_2 \sim \text{Uniform}(0, 20).$$

The longer it takes for the food to arrive, the colder it will get, and its value for you will be lower. We describe the *value* by a random variable

$$V_i = aL_i - bT_i,$$

where L_i is a random variable indicating how much you typically like the food from restaurant $i = 1, 2$, and a and b are constants. Note that the value can be negative. We assume $L_1 = 15$ and $L_2 \sim \text{Uniform}(10, 12)$. Also, L_2 and T_2 are independent.

(a) Suppose $a = 2, b = 1$. *Find and sketch* the probability density function of V_1.

(b) For $a = 2, b = 1$, find $P(V_2 > 10)$.

(c) For $a = 2, b = 1$,

i. Find $E(V_1)$.

ii. Find $E(V_2)$.

iii. Which restaurant would you order from if your goal is to maximize your mean value? Explain.

(d) If you didn't care about when you get the food, but only wanted to maximize how much you enjoy the food, how would you set a and b? Which restaurant would you choose? Explain.

<u>**Solution:**</u>

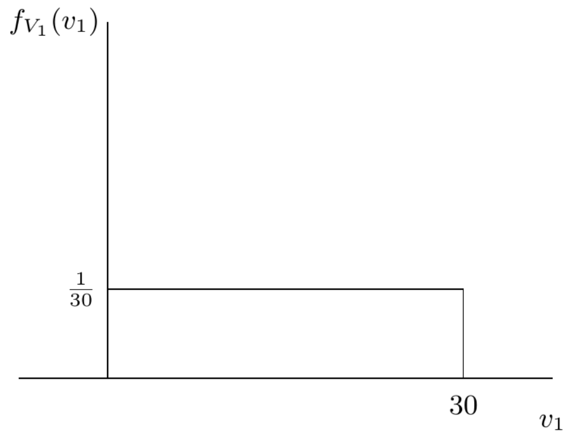
(a) $a = 2$ and $b = 1$, we have: $V_1 = 2 \times 15 - T_1 = 30 - T_1$

$$\begin{aligned}
P(30 - T_1 \leq v_1) &= P(30 - v_1 \leq T_1) \\
&= \frac{1}{30} \int_{30-v_1}^{\infty} U(t) - U(t-30) dt \\
&= \frac{1}{30} \int_{\max(30-v_1, 0)}^{30} dt \\
&= \begin{cases} 0, & \text{if } v_1 < 0 \\ \frac{v_1}{30}, & \text{if } 0 \leq v_1 \leq 30 \\ 1 & \text{if } 30 < v_1, \end{cases}
\end{aligned}$$

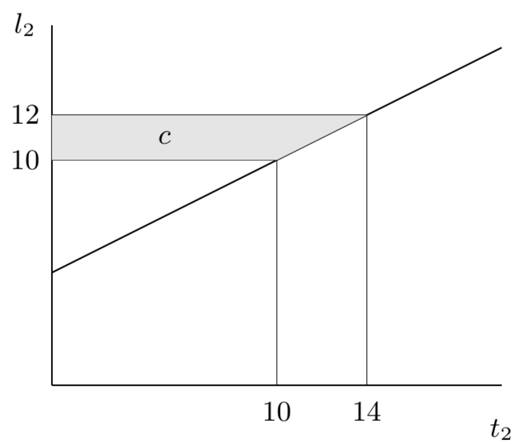
we know $f_{V_1}(v_1) = \frac{d}{dv_1} F_{V_1}(v_1)$. We have

$$f_{V_1}(v_1) = \begin{cases} 0, & \text{if } v_1 < 0 \\ \frac{1}{30}, & \text{if } 0 \geq v_1 \leq 30 \\ 0 & \text{if } 30 < v_1, \end{cases}$$

we can see that $V_1 \sim Uniform(0, 30)$



(b) $a = 2$ and $b = 1$, we have $V_2 = 2L_2 - T_2$



We want the probability of being in the shaded region which is equal to :

$$
\begin{aligned}
P(V_2 > 10) &= P(2L_2 - T_2 > 10) \\
&= \int\int_{t,l\in c} f_{L_2}(l) f_{T_2}(t) dl dt \\
&= \text{area of shaded region} \times \frac{1}{20} \times \frac{1}{2} \\
&= \frac{(10+14)\times(12-10)}{2} \times \frac{1}{20} \times \frac{1}{2} \\
&= \frac{3}{5}
\end{aligned}
$$

(c) $a = 2$ and $b = 1$, we have

i.

$$
\begin{aligned}
E[V_1] &= E[30 - T_1] \\
&= 30 - E[T_1] = 30 - 15 \\
&= 15
\end{aligned}
$$

ii.

$$
\begin{aligned}
E[V_2] &= E[2L_2 - T_2] \\
&= 2 \times E[L_2] - E[T_2] = 2 \times 11 - 10 \\
&= 12
\end{aligned}
$$

iii. Since $E[V_1] > E[V_2]$, you order from Restaurant 1, Elza's Diner.

(d) We don't care about the time it takes to get the food so $b = 0$ and a can be any positive number. We have,

$$
\begin{aligned}
E[V_1] &= 15 \times a \\
E[V_2] &= 11 \times a
\end{aligned}
$$

we see that $E[V_1]$ is bigger so we will order from Elza's Diner to have better quality.