Stat 302 Assignment 2 (solution)

Q1. (a) Suppose X is the number of UBC students infected by H1N1 and follows a Poisson random variable with $\lambda = 5$ per week, then

$$\begin{aligned} P(X > 10) &= 1 - \sum_{i=1}^{10} P(X = i) \\ &= 0.0137. \end{aligned}$$

(b) Suppose random variable Y is the number of weeks with more than 10 infected students. Clearly Y follows a binomial distribution with $n = 4$ and $p = P(X > 10) = 0.0137$. Then

$$\begin{aligned} P(Y \geq 2) &= 1 - P(Y = 0) - P(Y = 1) \\ & \quad (\text{or } P(Y = 2) + P(Y = 3) + P(Y = 4)) \\ &= 0.0011. \end{aligned}$$

Q2. (a)

$$\begin{aligned} \frac{f(k; n, p)}{f(k-1; n, p)} &= \frac{\binom{n}{k} p^k (1-p)^{n-k}}{\binom{n}{k-1} p^{k-1} (1-p)^{n-k+1}} \\ &= \frac{n-k+1}{k} \times \frac{p}{1-p} \\ &= 1 + \frac{(n+1)p - k}{k(1-p)}. \end{aligned}$$

(b)

$$\begin{aligned} P(X \geq k) &= \sum_{i=k}^{n} \binom{n}{i} p^i (1-p)^{n-i} \\ &= \sum_{i=k}^{n} \binom{n}{n-i} p^i (1-p)^{n-i} \\ &= \sum_{j=n-k}^{0} \binom{n}{j} p^{n-j} (1-p)^j, \;\; j = n - i \\ &= P(Y \leq n - k). \end{aligned}$$

Q3. Denote X as the number of alpha particles emitted per minute by a radioactive substance, and we know X follows a Poisson process with rate $\lambda = 4$ per minute.

(a) $P(X = 3) = \frac{4^3 e^{-4}}{3!} = 0.1953$.

(b) $P(X \leq 3) = \sum_{i=0}^{3} \frac{4^i e^{-4}}{i!} = 0.4335$.

(c) Let Y_i denote the interval length between the $(i-1)$th and the ith event, which is described by a Poisson process. Then Y_i's are i.i.d. random variables following exponential distribution with p.d.f. $f_Y(y) = \lambda e^{-\lambda y},\ y > 0$. The expected time is

$$
\begin{aligned}
E(\sum_{i=1}^{100} Y_i) &= \sum_{i=1}^{100} E(Y_i) \\
&= 100E(Y_i) \\
&= 100 \times \frac{1}{4} \\
&= 25\text{minutes}.
\end{aligned}
$$

Q4. (a) Denote the time of crashes within one year as X, which follows Poisson distribution with rate of $\lambda = 1.5$ per **year**. Then the probability that three or more crashes will occur next year is

$$
\begin{aligned}
P(X \geq 3) &= 1 - P(X \leq 2) \\
&= 1 - \sum_{i=0}^{2} \frac{1.5^i e^{-1.5}}{i!} \\
&= 0.1912.
\end{aligned}
$$

(b) Suppose Z_0, Z_1 and Z_2 as the occurrence time of three consecutive crashes, and Y_1 and Y_2 are the in-between times, that is, $Y_1 = Z_1 - Z_0$, $Y_2 = Z_2 - Z_1$. Assuming a Poisson process with rate $\lambda_1 = 1.5/12$ per **month**, we know that Y_1 and Y_2 are i.i.d. random variables following exponential distribution with rate λ_1.

The question **IS** asking $P(Y_2 < 3)$. Then

$$
\begin{aligned}
P(Y_2 < 3) &= 1 - e^{-3\lambda_1} \\
&= 0.3127.
\end{aligned}
$$

If you were unfortunately misled to work with $P(Y_1 + Y_2 < 3)$, here are two possible solutions.

- Smart solution provided by students ♪: We could consider 3 months as one unit, and let X represent the number of plane crashes in that period. Obviously X follows a Poisson distribution with a rate of $\lambda_2 = 1.5/4$. Then the probability that next two plane crashes occur within three months can be represented by:

$$\begin{aligned} P(X \geq 2) &= 1 - \sum_{i=0}^{1} \frac{\lambda_2^i e^{-\lambda_2}}{i!} \\ &= 0.055. \end{aligned}$$

- Another complicated solution: Since Y_1 and Y_2 are i.i.d. random variables, $W = Y_1 + Y_2$ follows Gamma distribution Gamma$(2, 1/\lambda_0)$. Therefore, we obtain $P(W < 3) = 0.055$.

Q5. We first compute the c.d.f. $F_X(k)$ and then obtain the p.m.f. as

$$\begin{aligned} p_X(k) &= P(X = k) = P(X \leq k) - P(X \leq k-1) \\ &= F_X(k) - F_X(k-1), \quad k = 1, 2, \ldots, 10 \end{aligned}$$

We have

$$\begin{aligned} F_X(k) &= P(X \leq k) \\ &= P(\max(X_1, X_2, X_3) \leq k) \\ &= P(X_1 \leq k, X_2 \leq k, X_3 \leq k) \\ &= P(X_1 \leq k)P(X_2 \leq k)P(X_3 \leq k), \end{aligned}$$

where the last equality follows from the independence of the events $\{X_1 \leq k\}$, $\{X_2 \leq k\}$, $\{X_3 \leq k\}$. Next we determine $P(X_1 \leq k)$.

Because your test scores are (integer) values between 1 and 10 with equal probability 1/10, i.e.

$$P(X_1 = k) = \frac{1}{10}, \quad k = 1, 2, \ldots, 10$$

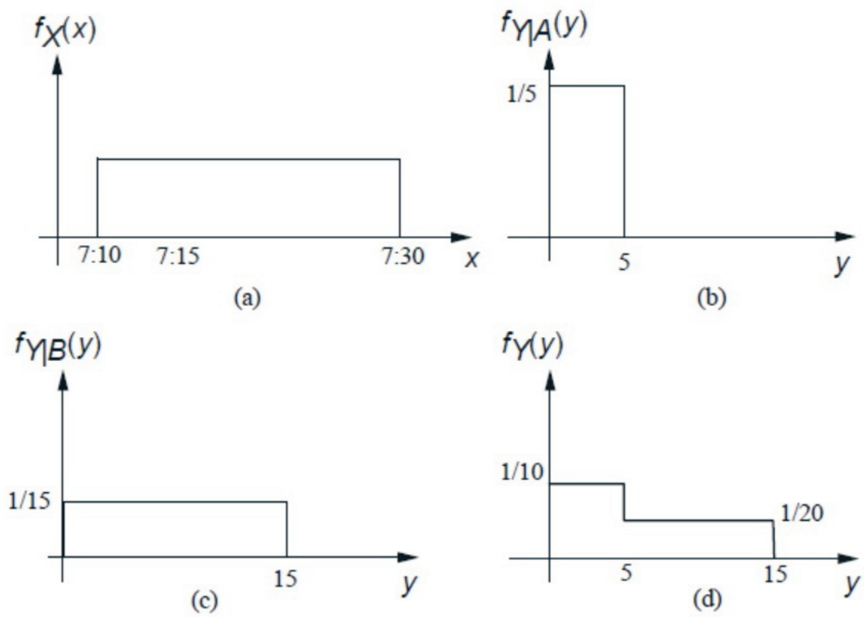
Therefore,

$$P(X_1 \le k) = P(X_1 = 1) + P(X_1 = 2) + \cdots + P(X_1 = k) = \frac{k}{10}.$$

Similarly, $P(X_2 \le k) = P(X_3 \le k) = \frac{k}{10}$. So $F_X(k) = (\frac{k}{10})^3$. Thus the p.m.f. is given by

$$p_X(k) = F_X(k) - F_X(k-1) = \left(\frac{k}{10}\right)^3 - \left(\frac{k-1}{10}\right)^3, \qquad k = 1, 2, \ldots, 10.$$

Q6. (a) Solution I.



The time of your arrival, denoted by X, is a uniform random variable on the interval from 7:10 to 7:30, i.e. $X \sim U(0, 20)$. The p.d.f. of X is

$$f_X(x) \quad = \quad \begin{cases} \frac{1}{20} & \text{for } 0 \le x \le 20, \\ 0 & \text{otherwise} \end{cases}$$

see Fig. 1(a).

Let A and B be the events

$A = \{\text{arrive at station between 7:10 and 7:15}\} = \{0 \le X \le 5\} = \{\text{you board the 7:15 train}\}$,

$B = \{\text{arrive at station between 7:15 and 7:30}\} = \{5 \le X \le 20\} = \{\text{you board the 7:30 train}\}$.

Then

$$P(A) = P(0 \le X \le 5) = \int_0^5 \frac{1}{20} dx = \left[\frac{1}{20} x\right]_0^5 = \frac{5}{20} = \frac{1}{4}$$

and

$$P(B) = P(5 \le X \le 20) = \int_5^{20} \frac{1}{20} dx = \left[\frac{1}{20} x\right]_5^{20} = \frac{15}{20} = \frac{3}{4}$$

Let Y be the waiting time.

Conditioning on the event A, your waiting time Y is also uniform and takes values between 0 and 5 minutes, i.e. $Y|A \sim U(0, 5)$. Therefore the p.d.f. of $Y|A$ is

$$f_{Y|A}(y) \quad = \quad \begin{cases} \frac{1}{5} & \text{for } 0 \le y \le 5, \\ 0 & \text{otherwise} \end{cases}$$

See Fig. 1(b).

Similarly, conditioned on B, Y is uniform and takes values between 0 and 15 minutes, i.e. $Y|B \sim U(0, 15)$. Therefore the p.d.f. of $Y|B$ is

$$f_{Y|B}(y) \quad = \quad \begin{cases} \frac{1}{15} & \text{for } 0 \le y \le 15, \\ 0 & \text{otherwise} \end{cases}$$

See Fig. 1(c).

The p.d.f of Y is obtained by

$$f_Y(y) = P(A) f_{Y|A}(y) + P(B) f_{Y|B}(y).$$

In particular, for $0 \le y \le 5$,

$$f_Y(y) = \frac{1}{4} \times \frac{1}{5} + \frac{3}{4} \times \frac{1}{15} = \frac{1}{10};$$

for $5 < y \le 15$, $f_{Y|A}(y) = 0$, hence,

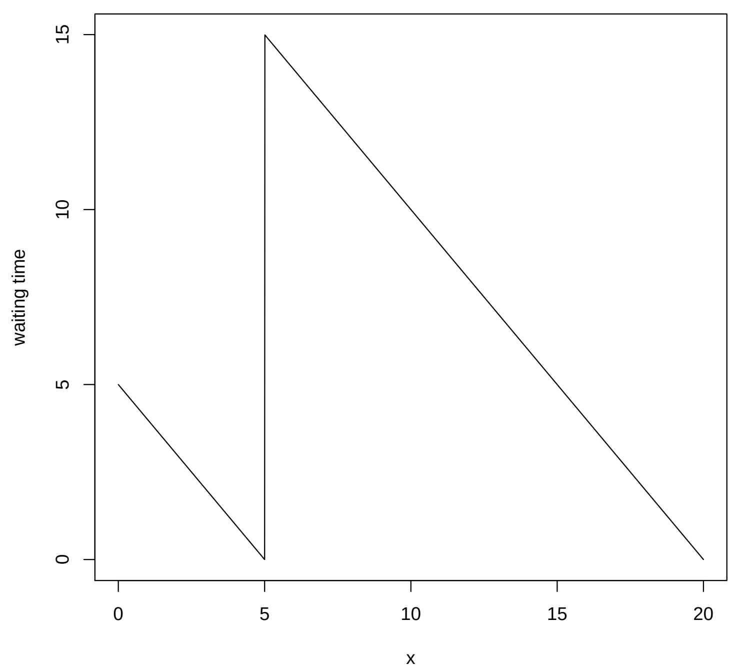
$$f_Y(y) = \frac{1}{4} \times 0 + \frac{3}{4} \times \frac{1}{15} = \frac{1}{20}.$$

We can also express the p.d.f. of Y as follows:

$$f_Y(y) \;=\; \begin{cases} \frac{1}{10} & \text{for } 0 \le y \le 5, \\ \frac{1}{20} & \text{for } 5 < y \le 15, \\ 0 & \text{otherwise} \end{cases}$$

The p.d.f of Y is shown in Fig. 1(d).

Solution II. According to the question, the following graph displays the relationship between arriving time X and waiting time Y.



We also know that

$$\{Y \le y\} \;=\; \begin{cases} \emptyset & \text{for } y < 0, \\ \{5 - y \le X \le 5\} \cup \{20 - y \le X \le 20\} & \text{for } 0 \le y \le 5, \\ \{0 \le X \le 5\} \cup \{20 - y \le X \le 20\} & \text{for } 5 < y \le 15, \\ \{0 \le X \le 20\} & \text{for } 15 < y. \end{cases}$$

Therefore, we can obtain the c.d.f. of Y

$$
\begin{aligned}
F(y) &= \begin{cases} 0 & \text{for } y < 0, \\ P(5-y \le X \le 5) + P(20-y \le X \le 20) & \text{for } 0 \le y \le 5, \\ P(0 \le X \le 5) + P(20-y \le X \le 20) & \text{for } 5 < y \le 15, \\ 1 & \text{for } 15 < y. \end{cases} \\
&= \begin{cases} 0 & \text{for } y < 0, \\ \int_{5-y}^{5} \frac{1}{20} dy + \int_{20-y}^{20} \frac{1}{20} dy = \frac{y}{10} & \text{for } 0 \le y \le 5, \\ \int_{0}^{5} \frac{1}{20} dy + \int_{20-y}^{20} \frac{1}{20} dy = \frac{5+y}{20} & \text{for } 5 < y \le 15, \\ 1 & \text{for } 15 < y. \end{cases}
\end{aligned}
$$

Then we may have the p.d.f. of Y.

(b) The waiting time Y should between 0 to 15 minutes, the expectation of Y is given by

$$
\begin{aligned}
E(Y) &= \int_0^{15} y \cdot f_Y(y) dy = \int_0^5 y \frac{1}{10} dy + \int_5^{15} y \frac{1}{20} dy \\
&= \left[\frac{1}{20} y^2 \right]_0^5 + \left[\frac{1}{40} y^2 \right]_5^{15} = \frac{5^2}{20} + \frac{15^2 - 5^2}{40} = 6.25,
\end{aligned}
$$

(c) Denote by m the median waiting time, by the definition of median, the c.d.f. of Y at m should be $1/2$. i.e. $F_Y(m) = 1/2$. From (a) we know,

$$
\begin{aligned}
F_Y(m) &= P(Y \le m) = \int_0^m f_Y(y) dy \\
&= \begin{cases} 0 & \text{for } m < 0 \\ \int_0^m \frac{1}{10} dy & \text{for } 0 \le m \le 5, \\ \int_0^5 \frac{1}{10} dy + \int_5^m \frac{1}{20} dy & \text{for } 5 < m \le 15, \\ 1 & \text{for } m > 15 \end{cases}
\end{aligned}
$$

It is easy to see that when $m = 5$, $F_Y(m) = \int_0^5 \frac{1}{10} dy = 1/2$. So the median waiting time is 5 minutes.

(d) The median is a "better" summary because Y is a skewed distribution.

Q7. The c.d.f. of X is given by

$$\begin{aligned}
F_X(t) &= P(X \leq t) = \int_0^t f_X(x)dx \\
&= \begin{cases} 0 & \text{for } t \leq 0 \\ \int_0^t \frac{1}{8}xdx & \text{for } 0 < t < 4, \\ 1 & \text{for } t \geq 4 \end{cases} \\
&= \begin{cases} 0 & \text{for } t \leq 0 \\ t^2/16 & \text{for } 0 < t < 4, \\ 1 & \text{for } t \geq 4 \end{cases}
\end{aligned}$$

(a) $P(X \leq t) = 1/4$ implies

$$\int_0^t \frac{1}{8}xdx = \left[\frac{1}{16}x^2\right]_0^t = \frac{t^2}{16} = 1/4.$$

Thus, $t = 2$.

(b) $P(X \geq t) = 1/2$ implies $P(X < t) = 1/2$.

From $P(X < t) = P(X \leq t) = \frac{t^2}{16} = 1/2$, we obtain $t = \sqrt{8} = 2.8284$

(c) Possible value of Y are $\{0, 1, 2, 3, 4\}$. Y is discrete random variable. The p.m.f. of Y is given by

$$\begin{aligned}
P(Y = 0) &= P\{(X \leq 0.5) \cup (X = 1.5) \cup (X = 2.5) \cup (X = 3.5)\} \\
&= P(X \leq 0.5) = \frac{0.5^2}{16} = \frac{1}{64} \\
P(Y = 1) &= P(0.5 < X < 1.5) = P(X < 1.5) - P(X < 0.5) = \frac{1.5^2 - 0.5^2}{16} = \frac{1}{8} \\
P(Y = 2) &= P(1.5 < X < 2.5) = P(X < 2.5) - P(X < 1.5) = \frac{2.5^2 - 1.5^2}{16} = \frac{2}{8}
\end{aligned}$$

$$P(Y=3) = P(2.5 < X < 3.5) = P(X < 3.5) - P(X < 2.5) = \frac{3.5^2 - 2.5^2}{16} = \frac{3}{8}$$
$$P(Y=4) = P(X \geq 3.5) = 1 - \frac{3.5^2}{16} = \frac{15}{64}$$

Q8. Let X_1 denote the lifetime of component 1, X_2 denote the lifetime of component 2, and X_3 denote the lifetime of component 3.

The lifetime (in day) of each component approximately follows an exponential distribution with a mean lifetime of 100 days, i.e. $X_i \sim \exp(\lambda)$ and $E(X_i) = 100$ for $i = 1, 2, 3$. Note that the expectation of exponential random variable is $1/\lambda$, therefore, in this question, the parameter $\lambda = 1/100 = 0.01$ per day. Hence, the p.d.f. of X_i is $f(x) = 0.01e^{-0.01x}, x \geq 0$ and the c.d.f. of X_i is $F(x) = P(X \leq x) = 1 - e^{-0.01x}, x \geq 0$.

(a)The lifetime of the first component is $X_1 \sim \exp(0.01)$.

$$\begin{aligned} P(50 \leq X_1 \leq 100) &= F(100) - F(50) \\ &= [1 - e^{-0.01\times 100}] - [1 - e^{-0.01\times 50}] \\ &= (1 - e^{-1.0}) - (1 - e^{-0.5}) = 0.63212 - 0.39347 = 0.23865 \end{aligned}$$

(b) The probability that component i ($i = 1, 2, 3$) has lifetimes between 50 and 150 days is

$$P(50 \leq X_i \leq 150) = F(150) - F(50) = (1 - e^{-1.5}) - (1 - e^{-0.5}) = 0.77687 - 0.39347 = 0.3834$$

Let Y denote the number of components (of the 3) with lifetime between 50 and 150 days, then $Y \sim Binomial(3, 0.3834)$. Therefore,

$$P(Y=2) = \binom{3}{2} 0.3834^2 (1 - 0.3834)^1 = 0.2719$$

(c) Denote by T the lifetime of the entire system. Since the components are arranged in series, we have $T = \min\{X_1, X_2, X_3\}$. The c.d.f. of T is

$$\begin{aligned} F_T(t) = P(T \leq t) &= 1 - P(T > t) = 1 - P(\min\{X_1, X_2, X_3\} > t) \\ &= 1 - P(X_1 > t, X_2 > t, X_3 > t) \\ &= 1 - P(X_1 > t)P(X_2 > t)P(X_3 > t) \end{aligned}$$

$$\begin{aligned}
&= 1-[1-P(X_1 \le t)][1-P(X_2 \le t)][1-P(X_3 \le t)] \\
&= 1-[1-(1-e^{-0.01t})]^3 \\
&= 1-[e^{-0.01t}]^3 = 1-e^{-0.03t}, \qquad t \ge 0.
\end{aligned}$$

where the forth equality follows from the independence of the events $\{X_1 > t\}$, $\{X_2 > t\}$, $\{X_3 > t\}$. Notice that $T \sim \exp(0.03)$. Thus, the mean of T is $1/0.03 = 33.33$.

Denote by m the median lifetime, we need to solve $F_T(m) = 1/2$, i.e. $1 - e^{-0.03m} = 1/2$. $e^{-0.03m} = 0.5 \Rightarrow m = \frac{\log(0.5)}{-0.03} = 23.105$. Thus, the median lifetime is 23.105 days.

(d) If the 3 components are arranged in parallel, the lifetime of entire system is $T = \max\{X_1, X_2, X_3\}$. Similar as Q5, the c.d.f. of T is given by

$$\begin{aligned}
F_T(t) &= P(T \le t) \\
&= P(\max(X_1, X_2, X_3) \le t) \\
&= P(X_1 \le t, X_2 \le t, X_3 \le t) \\
&= P(X_1 \le t)P(X_2 \le t)P(X_3 \le t) \\
&= (1-e^{-0.01t})^3, \qquad t \ge 0.
\end{aligned}$$

where the last equality follows from the independence of the events $\{X_1 \le k\}$, $\{X_2 \le k\}$, $\{X_3 \le k\}$. Next we determine $P(X_1 \le k)$.

The median lifetime can be solved from $F_T(m) = 1/2 = 0.5$, i.e.

$$(1-e^{-0.01m})^3 = 0.5$$

$$e^{-0.01m} = 1-\sqrt[3]{0.5} \Rightarrow m = \frac{\log(1-\sqrt[3]{0.5})}{-0.01} = 157.8426$$

If the 3 components are arranged in parallel, the median lifetime of entire system is 157.84 days, which is much longer than the median lifetime when the system is arranged in series.

Q9. Suppose

$$X \sim N(\mu, \sigma^2) = N(3, 4)$$

and Z denotes the standard normal distribution.

(a)

$$\begin{aligned} P(X > 4) &= P\left(\frac{X-3}{2} > \frac{1}{2}\right) \\ &= P\left(Z > \frac{1}{2}\right) \\ &= 1 - \Phi(.5) = 0.3085. \end{aligned}$$

(b)

$$\begin{aligned} P(2 < X < 4) &= P\left(-\frac{1}{2} < \frac{X-3}{2} < \frac{1}{2}\right) \\ &= P\left(-\frac{1}{2} < Z < \frac{1}{2}\right) \\ &= \Phi(.5) - \Phi(-.5) = 0.3829. \end{aligned}$$

(c)

$$\begin{aligned} P(|X-3| < 2) &= P\left(\frac{|X-3|}{2} < 1\right) \\ &= P\left(-1 < Z < 1\right) \\ &= \Phi(1) - \Phi(-1) = 0.6827. \end{aligned}$$

(d) $P(|X-3| < 2)$ is larger than $P(|X-2| < 2)$, because X is symmetric distributed and centered around 3.

(e) Since $P(|X-3| < c) = 0.90$, we know $P(X - 3 < c) = 0.95$.

$$\begin{aligned} P(X - 3 < c) &= 0.95 \\ P\left(\frac{X-3}{2} < \frac{c}{2}\right) &= 0.95 \\ P\left(Z < \frac{c}{2}\right) &= 0.95 \\ \frac{c}{2} &= 1.645 \\ c &= 3.29. \end{aligned}$$

Q10. Suppose $X \sim N(\mu, \sigma^2)$ where μ is unknown and $\sigma = 0.05\mu$.

(a) We need to set a value for μ such that $P(X < 25) = 0.1$.

$$\begin{aligned}
P(X < 25) &= 0.1 \\
P\left(\frac{X-\mu}{.05\mu} < \frac{25-\mu}{.05\mu}\right) &= 0.1 \\
P\left(Z < \frac{25-\mu}{.05\mu}\right) &= 0.1 \\
\frac{25-\mu}{.05\mu} &= -1.28 \\
\mu &= 26.7094.
\end{aligned}$$

(b) We just need to replace 25 with 50 and redo the above procedure to find the solution $\mu = 53.4188$.