MTH 201: Probability and Statistics

Quiz 2

22/05/2023

Sanjit K. Kaul

No books, notes, or devices are allowed. Just a pen and eraser. Any exchange of information related to the quiz with a human or machine will be deemed as cheating. Institute rules will apply. Explain your answers. Show your steps. Approximate calculations are fine as long as the approximations are reasonable. You have 45 minutes.

Question 1. 30 **marks** Random Variable X has the PDF

$$f_X(x) = \begin{cases} c & x \in (0, c), \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Answer the following questions.

9 (a) Calculate c that ensures that $f_X(x)$ is a valid PDF. State the properties that a PDF must satisfy, which you used to calculate c.

7 (b) An experiment whose outcome is governed by $f_X(x)$ is known to have resulted in a number smaller than 4. State the event that captures the observation. Derive the conditional PDF of X, given the event.

7 (c) Calculate the conditional variance of X.

7 (d) Derive the conditional probability that X took a value greater than 2.

Question 2. 30 **marks** A very large number of runners from around the world are known to participate in an informal yearly race conducted by Gossip Runners. Of all the runners that take part in the race, 80% reach the finish line within the time allotted for the race and 15% reach the finish line but take more than the allotted time. The rest don't finish the race. The time taken to reach the finish line for those who finish within the allotted time is distributed as a uniform$(1, 3)$ random variable. For those who reach the finish line but not in the allotted time, the time taken to reach the finish line is uniform$(3, 5)$. Calculate the expected value of time taken to finish by those who reach the finish line.

You need to solve only one of the two questions below. If you attempt both, your total score in this quiz will include larger of the two scores and disregard the other.

Question 3. 40 **marks** We have two timers A and B. Timer A times out after one or more integer number of ticks where the number of ticks is distributed as a geometric(q) random variable. That is, the timer times out in any tick, starting with tick 1 and independently of any other tick, with probability q. Timer B times out after a number of ticks that is distributed as a geometric(p) random variable. The number of ticks after which the timers time out are independent of each other. An experiment involves starting the timers simultaneously and noting down the number of ticks each timer took to time out. Derive the conditional PMF of the number of ticks after which timer A times out, given that B took a larger number of ticks than A to time out.

Question 4. 40 **marks** We have two timers A and B. Timer A times out after exponential(q) minutes. Timer B times out after an exponential(p) minutes. The number of minutes after which the timers time out are independent of each other. An experiment involves starting the timers simultaneously and noting down the number of minutes each timer took to time out. Derive the conditional PDF of the number of minutes after which timer A times out, given that B took a larger number of minutes than A to time out. [Hint: Recall that RV $X \sim$ exponential(λ) has a PDF $f_X(x) = \lambda e^{-\lambda x}, x \geq 0,$ and $f_X(x) = 0,$ otherwise.]

(8)

(a) $f_X(x) \geq 0, \quad x \in \mathbb{R}$. 1

Also $\int_{x \in \mathbb{R}} f_X(x) = 1$. 1

Therefore, we require

$$\int_0^c c \, dx = 1$$

$$\Rightarrow c^2 = 1$$

$$c = \pm 1.$$ 1

Since $f_X(x) \geq 0$, $c = +1$. 1

We have

$$f_X(x) = \begin{cases} 1 & (0,1) \\ 0 & \text{otherwise.} \end{cases}$$

5

(b) The event is $\{X < 4\}$

$$f_{X|X<4}(x) = \begin{cases} \dfrac{f_X(x)}{P[X<4]} & x \in (0,4) \\ 0 & \text{otherwise} \end{cases}$$

This is key.

$$= \begin{cases} \dfrac{f_X(x)}{1} & x \in (0,4) \\ 0 & \text{otherwise} \end{cases}$$

?

$$= \begin{cases} f_X(x) & x \in (0,1) \\ 0 & \text{otherwise} \end{cases}$$

Essentially, $f_{X|X<4}(x) = f_X(x)$,

$x \in \mathbb{R}$.

This isn't surprising, given that X must take values in $(0,1)$.

(c) $\mathrm{Var}[X \mid X<4] = \mathrm{Var}[X]$

Because $f_{X|X<4}(x) = f_X(x)$.

$$= E[X^2] - (E[X])^2$$

$$E[X] = \int_0^1 x \, dx = \left(\frac{x^2}{2}\right)\Big|_0^1 = \frac{1}{2}.$$

3.5

$$E[X^2] = \int_0^1 x^2\,dx = \left(\frac{x^3}{3}\right)_0^1 = \frac{1}{3}$$

$$\mathrm{Var}[X] = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}.$$ 3.5

(d) $$P[X>2 \mid X<4] = \frac{P[X>2,\, X<4]}{P[X<4]}$$

7

$$= \frac{P[2<X<4]}{P[X<4]} = \frac{0}{1} = 0.$$

Q2 Let X be the time for which any runner takes part in the race.

Let A be the event that a runner finishes in time

Let B be the event that a runner finishes but exceeds the allotted time.

$f_{X|A}(x)$ is uniform$(1,3)$

That is

$$f_{X|A}(x) = \begin{cases} \frac{1}{2} & x \in (1,3) \\ 0 & \text{otherwise} \end{cases}$$

3

Also,

$$f_{X|B}(x) = \begin{cases} \frac{1}{2} & x \in (3,5) \\ 0 & \text{otherwise} \end{cases}$$

3

We are interested in

$$E[X \mid A \cup B].$$

That is, we are interested in the average time for runners who finished.

$$P[\text{Runner exceeds allotted time} \mid \text{Runner finished}]$$

$$= \frac{P[\text{Runner exceeds allotted time, Runner finished}]}{P[\text{Runner finished}]}$$

$$= \frac{P[B]}{P[A \cup B]} = \frac{P[B]}{P[A]+P[B]} = \frac{0.15}{0.8+0.15}$$

$$= \frac{0.15}{0.95} = \frac{15}{95}$$

3

$$P\left[\text{Runner runs for a less than or equal to the allotted time} \,\middle|\, \text{Runner finished}\right]$$

$$= \frac{P[A]}{P[A \cup B]} = \frac{P[A]}{P[A] + P[B]} = \frac{0.8}{0.95} = \frac{80}{95}$$

5

Let $C = A \cup B$.

$$f_{X|C}(x) = f_{X|A}(x)\, P[A|C] + f_{X|B}\, P[B|C]$$

And

10

OR

The expectation of interest is

$$E[X|C] = E[X|A]\, P[A|C] + E[X|B]\, P[B|C]$$

$$= 2\,P[A|C] + 4\,P[B|C]$$

$$= 2\left(\frac{80}{95}\right) + 4\left(\frac{15}{95}\right)$$

$$= \frac{160+60}{95} = \frac{220}{95} \simeq 2.32.$$

Note that 2.32 is less than 3, which is smallest time those that finish but exceed the allotted time take. This makes sense as a significant fraction $\frac{80}{95}$ finish within the allotted time and take in $(1,3)$ hours to finish.

The $\frac{15}{95}$ fraction makes the average of 2 over who finish in allotted time increase to 2.32.

MTH 201: Probability and Statistics

Quiz 2

22/05/2023

Sanjit K. Kaul

No books, notes, or devices are allowed. Just a pen and eraser. Any exchange of information related to the quiz with a human or machine will be deemed as cheating. Institute rules will apply. Explain your answers. Show your steps. Approximate calculations are fine as long as the approximations are reasonable. You have 45 minutes.

Question 1. 30 **marks** Random Variable X has the PDF

$$f_X(x) = \begin{cases} c & x \in (0, c), \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Answer the following questions.

9 (a) Calculate c that ensures that $f_X(x)$ is a valid PDF. State the properties that a PDF must satisfy, which you used to calculate c.

7 (b) An experiment whose outcome is governed by $f_X(x)$ is known to have resulted in a number smaller than 4. State the event that captures the observation. Derive the conditional PDF of X, given the event.

7 (c) Calculate the conditional variance of X.

7 (d) Derive the conditional probability that X took a value greater than 2.

Question 2. 30 **marks** A very large number of runners from around the world are known to participate in an informal yearly race conducted by Gossip Runners. Of all the runners that take part in the race, 80% reach the finish line within the time allotted for the race and 15% reach the finish line but take more than the allotted time. The rest don't finish the race. The time taken to reach the finish line for those who finish within the allotted time is distributed as a uniform$(1, 3)$ random variable. For those who reach the finish line but not in the allotted time, the time taken to reach the finish line is uniform$(3, 5)$. Calculate the expected value of time taken to finish by those who reach the finish line.

You need to solve only one of the two questions below. If you attempt both, your total score in this quiz will include larger of the two scores and disregard the other.

Question 3. 40 **marks** We have two timers A and B. Timer A times out after one or more integer number of ticks where the number of ticks is distributed as a geometric(q) random variable. That is, the timer times out in any tick, starting with tick 1 and independently of any other tick, with probability q. Timer B times out after a number of ticks that is distributed as a geometric(p) random variable. The number of ticks after which the timers time out are independent of each other. An experiment involves starting the timers simultaneously and noting down the number of ticks each timer took to time out. Derive the conditional PMF of the number of ticks after which timer A times out, given that B took a larger number of ticks than A to time out.

Question 4. 40 **marks** We have two timers A and B. Timer A times out after exponential(q) minutes. Timer B times out after an exponential(p) minutes. The number of minutes after which the timers time out are independent of each other. An experiment involves starting the timers simultaneously and noting down the number of minutes each timer took to time out. Derive the conditional PDF of the number of minutes after which timer A times out, given that B took a larger number of minutes than A to time out. [Hint: Recall that RV $X \sim$ exponential(λ) has a PDF $f_X(x) = \lambda e^{-\lambda x}, x \geq 0,$ and $f_X(x) = 0,$ otherwise.]

Q3 Let N_A be the number of clicks to time out for A.

Let N_B be the same for B.

$$P[N_A = n] = \begin{cases} (1-q)^{n-1} q & n = 1, 2, \ldots \\ 0 & \text{otherwise} \end{cases}$$

$$P[N_B = n] = \begin{cases} (1-p)^{n-1} p & n = 1, 2, \ldots \\ 0 & \text{otherwise.} \end{cases}$$

N_A and N_B are independent RV(s).

The event of interest is $N_B > N_A$.

We want the PMF

$$P_{N_A | N_B > N_A}(n), \quad n \in \mathbb{R}.$$

$$P_{N_A | N_B > N_A}(n)$$

$$= P[N_A = n \mid N_B > N_A]$$

$$= \frac{P[N_A = n, N_B > N_A]}{P[N_B > N_A]}$$

Stating the probability of interest. Expanding it suitably.

5

$$= \frac{\left[P[N_A = n, N_B = n+1] + P[N_A = n, N_B = n+2] + \cdots \right]}{P[N_B > N_A]}$$

$$= \frac{\sum_{j=1}^{\infty} P[N_A = n, N_B = n+j]}{P[N_B > N_A]}$$

$$P[N_B > N_A] = \sum_{n=1}^{\infty} \sum_{j=1}^{\infty} P[N_A = n, N_B = n+j]$$

Calculating $P[N_B > N_A]$ including simplification 15

Therefore; for $n = 1, 2, \ldots$

$$P_{N_A | N_B > N_A}(n) = \frac{\displaystyle\sum_{j=1}^{\infty} P[N_A = n, N_B = n+j]}{\displaystyle\sum_{n=1}^{\infty} \sum_{j=1}^{\infty} P[N_A = n, N_B = n+j]}$$

Simplifying the numerator 15 using independence etc.

Using the independence of N_A and N_B, we can write

For $n = 1, 2, \dots$

$$p_{N_A | N_B > N_A}(n)$$

$$= \frac{\sum_{j=1}^{\infty} P[N_A = n]\, P[N_B = n+j]}{\sum_{n=1}^{\infty} \sum_{j=1}^{\infty} P[N_A = n]\, P[N_B = n+j]}$$

$$= \frac{P[N_A = n]\, P[N_B > n]}{\sum_{n=1}^{\infty} P[N_A = n]\, P[N_B > n]}$$

$$= \frac{(1-q)^{n-1} q\, (1-p)^n}{\sum_{n=1}^{\infty} (1-q)^{n-1} q\, (1-p)^n}$$

$$= \frac{[(1-q)(1-p)]^{n-1}}{\sum_{n=1}^{\infty} (1-q)^{n-1}(1-p)^{n-1}}$$

Final answer

$$\boxed{[1-(1-q)(1-p)][(1-q)(1-p)]^{n-1}}$$

Note that the PMF of A, given that A was the faster timer, that is $N_A < N_B$, is

$$\text{Geom}\left(1-(1-q)(1-p)\right)$$

Q4 Let N_A be the no. of minutes A takes to expire.

Let N_B be the minutes B takes to expire.

$$N_A \sim \text{exponential}(\alpha)$$

$$N_B \sim \text{exponential}(\beta)$$

N_A and N_B are independent RV(s).

We want the PDF

$$f_{N_A \mid N_A < N_B}(x) \quad \forall x.$$

Consider the joint conditional PDF

$$f_{N_A, N_B \mid N_A < N_B}(x, y)$$

5

$$= \begin{cases} \dfrac{f_{N_A,N_B}(x,y)}{P[N_A < N_B]} & \{(x,y) : x \geq 0, y \geq 0, x < y\} \\ 0 & \text{otherwise} \end{cases}$$

Let's calculate $P[N_A < N_B]$.

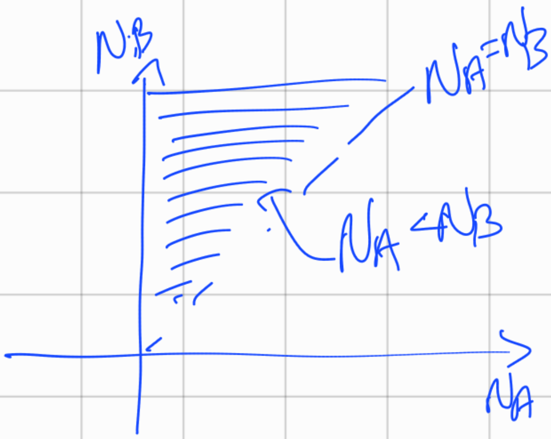
We need the joint PDF

$$f_{N_A,N_B}(x,y) = f_{N_A}(x)\, f_{N_A}(y) \quad \forall x,y.$$

since the RVs are independent.

$$f_{N_A,N_B}(x,y) = \begin{cases} qe^{-qx}\, pe^{-py} & x \geq 0, y \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

5



$P[N_A < N_B]$ Calculating this 10

$$= \int_0^\infty \int_0^y qe^{-qx}\, pe^{-py}\, dx\, dy$$

$$= \int_0^\infty pe^{-py} \left(\int_0^y qe^{-qx}\, dx \right) dy$$

$$= \int_0^\infty p e^{-py} \left(1 - e^{-qy}\right) dy$$

$$= \int_0^\infty p e^{-py} dy - \int_0^\infty p e^{-py} e^{-qy} dy$$

$$= 1 - \frac{p}{p+q} \int_0^\infty \underbrace{(p+q) e^{-(p+q)y}}_{\text{PDF of } \exp(p+q)} dy$$

$$= 1 - \frac{p}{p+q}$$

$$= \frac{q}{p+q}.$$

Calculating this (10)

$$\therefore f_{N_A, N_B | N_A < N_B}(x, y)$$

$$= \begin{cases} \dfrac{pq e^{-qx} e^{-py}}{(q/(p+q))} & \{(x,y) : x \geq 0, y \geq 0, x < y\} \\ 0 & \text{otherwise.} \end{cases}$$

Note that $f_{N_A | N_A < N_B}(x)$

5 $= \int_x^\infty f_{N_A, N_B | N_A < N_B}(x, y)\, dy,$ $x \geq 0.$

$$= \int_x^\infty (p+q) p\, e^{-qx} e^{-py}\, dy$$

$$= (p+q) e^{-qx} \int_x^\infty p\, e^{-py}\, dy$$

$$= (p+q) e^{-qx} \left(e^{-px}\right)$$

$$= (p+q) e^{-(p+q)x}, \quad x \geq 0$$

$$\therefore \quad f_{N_A | N_A < N_B}(x) = \begin{cases} (p+q) e^{-(p+q)x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Final answer 5

Note that the conditional PDF of N_A is exponential $(p+q)$.

We have a faster exponential that times out on an average in $\frac{1}{p+q}$ minutes.

Compare this with $\frac{1}{p}$ for N_B and $\frac{1}{q}$ for N_A.