Tutorial 9 – MATH 4043

This tutorial contains **a** continuous assessment item.

1. This question deals with the discretisation of an exponentially distributed random variable. Let X be distributed as exponential with $\lambda = 5$. Define another random variable Y with probability mass function $p_Y(y) = P[Y = y]$ such that

$$P[Y = y] = P[y - 1 < X \leq y] \text{ for } y = 1, 2, \cdots$$

   (a) Obtain an expression for $p_Y(y)$ in terms of y and simplify.

   (b) Identify the distribution of Y from its probability mass function.

2. Let X be distributed as exponential $\text{Exp}(\lambda)$.

   (a) Show that $P[X > 7|X > 2] = P[X > 5]$.

   (b) Let $a > 0$ and $b > 0$. Show that

$$P[X > a + b|X > a] = P[X > b].$$

   This is known as the memoryless property of the exponential random variable.

3. Let Y be a binomially distributed random variable where $Y \sim \text{Bin}(100, 0.45)$. Let X be a normally distributed random variable such that $\mu = E[X] = E[Y]$ and $\sigma^2 = \text{Var}[X] = \text{Var}[Y]$. You may use **R** to assist you in your calculations.

   (a) Calculate $P[Y = 45]$ and $P[44.5 \leq X \leq 45.5]$. Note that because X is a continuous random variable, we cannot calculate $P[X = 45]$ since it will be zero.

   (b) Calculate $P[40 \leq Y \leq 45]$ and $P[39.5 \leq X \leq 45.5]$.

   (c) What do you observe in your calculations?

   This is known as the Normal approximation to the binomial. It works reasonably well when n is large and p is neither too small or too large.

4. There is another inequality that is useful for a non-negative valued random variable X. This is called Markov's inequality and it is

$$P[X \geq a] \leq \frac{E[X]}{a}.$$

Let X be exponential with $\lambda = 2$.

(a) Calculate $P[X \geq 2]$.

(b) Compare with the Markov inequality.

(c) Calculate $P[X < 1]$.

(d) Compare with the Markov inequality.

(e) Challenge problem: Prove Markov's inequality. You can assume that X is any continuous random variable.

5. . Let X be distributed as Poisson with $\lambda = 16$.

(a) Calculate $P[6 \leq X \leq 26]$.

(b) Compare your result with an appropriate use of Chebyshev's Inequality.

6. **Continuous Assessment Item**. Suppose the waiting times (in hours) for a customer arriving in a queue shop (timed immediately after the previous customer or after time 0 in the case of the very 1st customer) is distributed as exponential Exp(6). Use **R** to simulate the arrival times of the customers within a 2 hour period. Suppose you ended up with k customers arriving in the 2 hour period, use **R** to calculate $P[S_k \leq 2]$, where S_k is the sum of the k waiting times.

Solutions.

1. This question deals with the discretisation of an exponentially distributed random variable. Let X be distributed as exponential with $\lambda = 5$. Define another random variable Y with probability mass function $p_Y(y) = P[Y = y]$ such that

$$P[Y = y] = P[y - 1 < X \leq y] \text{ for } y = 1, 2, \cdots$$

   (a) Obtain an expression for $p_Y(y)$ in terms of y and simplify.

   (b) Identify the distribution of Y from its probability mass function.

Solution.

   (a) From Question 1, for $x > 0$, $F_X(x) = 1 - e^{-\lambda x}$.

$$\begin{aligned} P[Y = y] &= P[y - 1 < X \leq y] \\ &= F_X(y) - F_X(y - 1) \\ &= 1 - e^{-\lambda y} - (1 - e^{-\lambda(y-1)}) \\ &= e^{-\lambda(y-1)} - e^{-\lambda y} \\ &= (e^{-\lambda})^{y-1}(1 - e^{-\lambda}). \end{aligned}$$

   Thus $p_Y(y) = (e^{-\lambda})^{y-1}(1 - e^{-\lambda})$ for $y = 1, 2, \cdots$.

   (b) The pmf of Y take the form $p_Y(y) = (1-p)^{y-1}p$, where $p = 1-e^{-\lambda}$ and $1-p = e^{-\lambda}$. This is the form of a Geometric Geo(p) random variable. Thus $Y \sim$ Geo(p) and we have discretise X into a geometric random variable.

2. Let X be distributed as exponential Exp(λ).

   (a) Show that $P[X > 7|X > 2] = P[X > 5]$.

   (b) Let $a > 0$ and $b > 0$. Show that

$$P[X > a + b|X > a] = P[X > b].$$

   This is known as the memoryless property of the exponential random variable.

Solution. We use $F_X(x) = 1 - e^{-\lambda x}$ for $x > 0$ from Question 1. Thus for any $x > 0$, $P[X > x] = 1 - F_X(x) = 1 - (1 - e^{-\lambda x}) = e^{-\lambda x}$.

(a)

$$\begin{aligned} P[X>7|X>2] &= \frac{P[\{X>7\}\cap\{X>2\}]}{P[\{X>2\}]} \\ &= \frac{P[\{X>7\}]}{P[\{X>2\}]} \\ &\quad (\text{since } \{X>7\}\subset\{X>2\}) \\ &= \frac{e^{-7\lambda}}{e^{-2\lambda}} \\ &= e^{-5\lambda} = P[X>5]. \end{aligned}$$

(b)

$$\begin{aligned} P[X>a+b|X>a] &= \frac{P[\{X>a+b\}\cap\{X>a\}]}{P[\{X>a\}]} \\ &= \frac{P[\{X>a+b\}]}{P[\{X>a\}]} \\ &\quad (\text{since } \{X>a+b\}\subset\{X>a\}) \\ &= \frac{e^{-\lambda(a+b)}}{e^{-\lambda a}} \\ &= e^{-\lambda b} = P[X>b]. \end{aligned}$$

3. Let Y be a binomially distributed random variable where $Y \sim \text{Bin}(100, 0.45)$. Let X be a normally distributed random variable such that $\mu = E[X] = E[Y]$ and $\sigma^2 = \text{Var}[X] = \text{Var}[Y]$. You may use **R** to assist you in your calculations.

   (a) Calculate $P[Y = 45]$ and $P[44.5 \le X \le 45.5]$. Note that because X is a continuous random variable, we cannot calculate $P[X = 45]$ since it will be zero.

   (b) Calculate $P[40 \le Y \le 45]$ and $P[39.5 \le X \le 45.5]$.

   (c) What do you observe in your calculations?

   This is known as the Normal approximation to the binomial. It works reasonably well when n is large and p is neither too small or too large.

   Solution. $Y \sim \text{Bin}(100, 0.45)$ and $X \sim N(45, \sqrt{24.75})$. Note that I will be using **R** to assist in the calculations.

   (a) $P[Y = 45] = C^{100}_{45}(0.45)^{45}(0.55)^{35} = 0.0799875$.
   For $P[44.5 \le X \le 45.5]$, we have

```
> pnorm(45.5,45,sqrt(24.75)) - pnorm(44.5,45,sqrt(24.75))
[1] 0.08005562
```

(b) $P[40 \le Y \le 45] = P[Y \le 45] - P[Y \le 39]$. Using **R** , we get

```
> pbinom(45,100,0.45) - pbinom(39,100,0.45)
[1] 0.4070621
```

For $P[39.5 \le X \le 45.5]$, we get

```
> pnorm(45.5,45,sqrt(24.75)) - pnorm(39.5,45,sqrt(24.75))
[1] 0.4055653
```

(c) Both pairs of answers in both parts are very close to each other, so the normal approximation to the binomial works.

4. There is another inequality that is useful for a non-negative valued random variable X. This is called Markov's inequality and it is

$$P[X \ge a] \le \frac{E[X]}{a}.$$

Let X be exponential with $\lambda = 2$.

(a) Calculate $P[X \ge 2]$.

(b) Compare with the Markov inequality.

(c) Calculate $P[X < 1]$.

(d) Compare with the Markov inequality.

(e) Challenge problem: Prove Markov's inequality. You can assume that X is any continuous random variable.

Solution. Using Question 1, $F_X(x) = 1 - e^{-2x}$ for $x > 0$ and hence $P[X \ge x] = e^{-2x}$. Also $E[X] = \frac{1}{2}$ using the results of Question 1. Note that a must be positive.

(a) $P[X \ge 2] = e^{-2\times 2} = e^{-4} = 0.01831564$.

(b) The Markov inequality yields $P[X \ge 2] \le \frac{0.5}{2} = \frac{1}{4}$. The actual answer in part (a) is less that $\frac{1}{4}$.

(c) $P[X < 1] = F_X(1) = 1 - e^{-2} = 0.8646647$.

(d) $P[X < 1] = 1 - P[X \ge 1] > 1 - \frac{E[X]}{1} = 1 - \frac{1}{2} = \frac{1}{2}$ using the Markov inequality in the opposite direction. The actual answer in part (c) is greater than $\frac{1}{2}$.

(e) We will prove it for X any positive random variable with pdf $f_X(x)$. All points x in the region $\{X \ge a\}$ must satisfy $x \ge a$. The proof follows the style of the proof of the Tchebyshev inequality.

$$\begin{aligned}
E[X] &= \int_0^\infty x f_X(x)\,dx \\
&\geq \int_a^\infty x f_X(x)\,dx \quad \text{(since integrating over a smaller region)} \\
&\geq \int_a^\infty a f_X(x)\,dx \quad \text{(since } x > a \text{ in this smaller region)} \\
&= a \int_a^\infty f_X(x)\,dx \\
&= aP[X \geq a].
\end{aligned}$$

Thus $E[X] \geq aP[X \geq a]$ and hence $P[X \geq a] \leq \frac{E[X]}{a}$.

5. Let X be distributed as Poisson with $\lambda = 16$.
   (a) Calculate $P[6 \leq X \leq 26]$.
   (b) Compare your result with an appropriate use of Chebyshev's Inequality.

   Solutions.

   (a) Calculate $P[6 \leq X \leq 26]$. By using **R**, we have that

   $$P(6 \leq X \leq 26) = P(X \leq 26) - P(X \leq 6) =$$

   ```
   > ppois(26,16) - ppois(6,16)
   [1] 0.988535
   ```

   (b) Compare your result with an appropriate use of Chebyshev's Inequality. By using Chebyshev's inequality,

   $$P(|X - E(X)| \geq a) \leq Var(X)/a^2$$

   we have $E(X) = 16$, $Var(X) = 16$, and $a = 10$,

   $$P(|X - 16| \geq 10) = P(X \geq 26 \textit{ or } X \leq 6) \leq 16/100 = 0.16.$$

   From part (a), we can calculate $P(X \geq 26 \textit{ or } X \leq 6)$ since

   $$P(X \geq 26 \textit{ or } X \leq 6) = 1 - P(6 \leq X \leq 26) = 0.011465.$$

   This is clearly less than 16%.

6. **Continuous Assessment Item**. Suppose the waiting times (in hours) for a customer arriving in a queue shop (timed immediately after the previous customer or after time 0 in the case of the very 1st customer) is distributed as exponential Exp(6). Use **R** to simulate the arrival times of the customers within a 2 hour period. Suppose you ended up with k customers arriving in the 2 hour period, use **R** to calculate $P[S_k \leq 2]$, where S_k is the sum of the k waiting times.