Tutorial 5 – MATH 4043 – Discrete Distributions Wrap-up

1. How many times should you toss a fair coin if you want a probability of 0.9 of getting heads at least once?

2. If $X \sim \mathrm{Bin}(n, p)$, what is the distribution of $Y = n - X$?

3. (Key Question) A pond has 100 fish. 60 are black and 40 are red. A fisherman catches 10 fish without replacement. What is the probability that he catches 3 black out of the 10?

4. Use both the exact Binomial calculation as well as the Poisson approximation to do the following calculations.

   (a) What is the probability that there are 15 heads in 30 fair coin tosses?

   (b) What is the probability that there will be fewer than 4 1's in 60 throws of a balanced die?

5. Assume that within any second, there is a probability of 0.25 that a Geiger counter records a radioactive signal. Assume that it is only possible to receive at most one signal within one second.

   Assume that time is discretised into 1 second intervals. Let X be the number of seconds taken for the 1st signal to arrive, thus $X \sim$ Geo(0.25)

   (a) What is the probability that the first signal is received after 3 seconds?

   (b) How many seconds do we need to wait in order to get about 10.5% probability of receiving the first signal?

   (c) What is the probability that the 2nd signal is received in the 4th second?

   (d) Use **R** to simulate the arrival of the signals over a 50 seconds interval (eg code "1" to denote a signal, "0" for no signal).

6. Use **R** and calculate by hand. The number of customers that arrive at a shop is distributed as Poisson with mean rate of 5 per hour.

   (a) What is the probability that at least 3 customers arrive at the shop within the hour?

   (b) What is the probability there is a most 2 customer arrivals within the hour?

   (c) What is the most likely number of customers that arrive at the shop within the hour?

   (d) Use **R** to simulate the number of customers arriving within each hour over a 10 hour period. **Hint:** Here is my example of 10 customers, simulate your own and interpret.

   ```
   > rpois(10,5)
    [1] 3 6 6 4 7 1 4 5 2 8
   ```

7. A shop has two entrances – front and back. The number of customers that arrive at the front entrance is distributed as Poisson with mean rate of 8 per hour. The number of customers that arrive at the back entrance is distributed as Poisson with mean rate of 6 per hour.

   (a) Use **R** to simulate the number of customers arriving via the front entrance within each hour over a 10 hour period.

   (b) Use **R** to simulate the number of customers arriving via the back entrance within each hour over a 10 hour period.

   (c) Merge your results from parts (a) and (b) to obtain the total number of customers arriving into the shop within each hour over the 10 hour period. N.B. You need to store your results from parts (a) and (b) first while doing those parts.

8. (Key Question) A winemaker is interested in the mean number of wine bottle (sales) at tasting events. He feels that the number of wine bottles sold has mean μ and that μ is either $8, 10, 12$. The winemaker assesses the prior probabilities of

$$P(\mu = 8) = 0.3,\ P(\mu = 10) = 0.60,\ P(\mu = 12) = 0.1.$$

   From past experience, he is willing to assume that the process has some variance given by the process. He randomly selects five tasting events and counts his number of sales he might sell, with the following results: $15, 2, 8, 10,$ and 7.

   (a) Find the winemaker's posterior distribution.

   (b) Compute the means and the variances of the prior and posterior distributions.

Tutorial 5 SOLUTIONS– MATH 4043 – Discrete Distributions Wrap-up

1. How many times should you toss a fair coin if you want a probability of 0.9 of getting heads at least once?

   Solution. Let $X \sim \text{Bin}(n, 0.5)$, where X counts the number of heads in n tosses. We want to find n such that
   $$P[X \geq 1] = 0.9.$$
   This is the same as $P[X = 0] = 1 - P[X \geq 1] = 0.1$. Thus solving
   $$P[X = 0] = C_0^n \ (0.5)^n = (0.5)^n = 0.1,$$
   we get
   $$\begin{aligned} \ln[(0.5)^n] &= \ln(0.1) \\ n \ln[(0.5)] &= \ln(0.1) \text{ (This is a property of logarithm)} \\ n &= \frac{\ln(0.1)}{\ln(0.5)} \\ n &= 3.32. \end{aligned}$$
   Need $n \geq 3.32$ so take next largest integer which is $n = 4$.

2. If $X \sim \text{Bin}(n, p)$, what is the distribution of $Y = n - X$?

   Solution. We need to find $p_Y(y) = P[Y = y] = P[n - X = y]$, since the form of the probability mass function of Y will give us information about how it is distributed.

   Now $P[n - X = y] = P[X = n - y]$, so we use the probability mass function of X evaluated at $x = n - y$. Thus
   $$\begin{aligned} p_Y(y) &= P[X = n - y] \\ &= \frac{n!}{(n-y)!(n-(n-y))!} p^{n-y} q^{n-(n-y)} \\ &= \frac{n!}{(n-y)!y!} p^{n-y} q^y = C_y^n q^y (1-q)^{n-y}. \end{aligned}$$
   This looks like the form of a Binomial pmf with parameters n and q, instead of n and p.

   So $Y \sim \text{Bin}(n, q)$ where $q = 1 - p$. The result makes sense since if X counts the number of heads in n tosses, then Y must count the number of tails in n tosses. So "success" for Y is obtaining tails instead of heads.

3. A pond has 100 fish. 60 are black and 40 are red. A fisherman catches 10 fish without replacement. What is the probability that he catches 3 black out of the 10?

   Solution. Here, $X \sim \text{Hyper}(100, 60, 10)$.

$$P[X = 3] = \frac{C_3^{60}\, C_7^{40}}{C_{10}^{100}} = 0.0369,$$

   using the formula in Lecture Notes Week 6.

4. Use both the exact Binomial calculation as well as the Poisson approximation to do the following calculations.

   Solution.

   (a) What is the probability that there are 15 heads in 30 fair coin tosses?
   Here let $X \sim \text{Bin}(30, 0.5)$ and $Y \sim \text{Pois}(15)$ since $\lambda = np$.

$$P[X = 15] = C_{15}^{30}(0.5)^{30} = 0.1445.$$

   and

$$P[Y = 15] = e^{-15}\frac{15^{15}}{15!} = 0.1024.$$

   Note that n is probably not that large enough for the approximation to work well in this case and the mean and the variance of the binomial are far from equal!

   (b) What is the probability that there will be fewer than 4 1's in 60 throws of a balanced die?
   Here let $X \sim \text{Bin}\left(60, \frac{1}{6}\right)$ and $Y \sim \text{Pois}(10)$.

$$\begin{aligned} P[X < 4] &= P[X = 0] + P[X = 1] + P[X = 2] + P[X = 3] \\ &= \left(\frac{5}{6}\right)^{60} + 60\left(\frac{5}{6}\right)^{59}\left(\frac{1}{6}\right) + \frac{60 \times 59}{2 \times 1} \times \left(\frac{5}{6}\right)^{58}\left(\frac{1}{6}\right)^{2} \\ &\quad + \frac{60 \times 59 \times 58}{3 \times 2 \times 1} \times \left(\frac{5}{6}\right)^{57}\left(\frac{1}{6}\right)^{3} \\ &= 0.0635 \end{aligned}$$

   and

$$\begin{aligned} P[Y < 4] &= P[Y = 0] + P[Y = 1] + P[Y = 2] + P[Y = 3] \\ &= \frac{e^{-10}10^0}{0!} + \frac{e^{-10}10^1}{1!} + \frac{e^{-10}10^2}{2!} + \frac{e^{-10}10^3}{3!} \\ &= e^{-10} + 10e^{-10} + 50e^{-10} + \frac{1000}{6}e^{-10} \\ &= 0.0103. \end{aligned}$$

Note that the approximations are not good in these cases compared to my lecture example.

5. Assume that within any second, there is a probability of 0.25 that a Geiger counter records a radioactive signal. Assume that it is only possible to receive at most one signal within one second.

   Solution. Assume that time is discretised into 1 second intervals. Let X be the number of seconds taken for the 1st signal to arrive, thus $X \sim \text{Geo}(0.25)$

   (a) What is the probability that the first signal is received after 3 seconds?

$$
\begin{aligned}
P[X > 3] &= P[X = 4] + P[X = 5] + \cdots \\
&= (0.25)(0.75)^3 + (0.25)(0.75)^4 + \cdots \\
&= (0.25)(0.75)^3[1 + 0.75 + 0.75^2 + \cdots] \\
&= (0.25)(0.75)^3 \frac{1}{1 - 0.75} \\
&= (0.75)^3 = \frac{27}{64}.
\end{aligned}
$$

   (b) How many seconds do we need to wait in order to get about 10.5% probability of receiving the first signal?

   This is to find n such at $P[X = n] \approx 0.105$. Solve

$$(0.25)(0.75)^{n-1} = 0.105.$$

   We can try different values of n, say $n = 2, 3, \cdots$ and so on until we get one that makes the LHS close to 0.105. When $n = 4$, we get $(0.25)(0.75)^{4-1} = 0.1055$, thus we need to wait 4 seconds.

   Alternatively, solve

$$
\begin{aligned}
(0.25)(0.75)^{n-1} &= 0.105 \\
\ln(0.25) + (n - 1)\ln(0.75) &= \ln(0.105) \text{ (property of logarithms)} \\
n &= 1 + \frac{\ln(0.105) - \ln(0.25)}{\ln(0.75)} \approx 4.01.
\end{aligned}
$$

   (c) What is the probability that the 2nd signal is received in the 4th second?

   Let X_1 be the number of seconds to the 1st signal, and X_2 be the number of seconds to the 2nd signal measured after the 1st signal. Then $X_1 \sim \text{Geo}(0.25)$ and

$X_2 \sim$ Geo(0.25) and we want to calculate $P[X_1 + X_2 = 4]$. Thus

$$
\begin{aligned}
&P[X_1 + X_2 = 4] \\
=&P[\{X_1 = 1\} \cap \{X_2 = 3\}] + P[\{X_1 = 2\} \cap \{X_2 = 2\}] \\
&+ P[\{X_1 = 3\} \cap \{X_2 = 1\}] \\
=&P[\{X_1 = 1\}]P[\{X_2 = 3\}] + P[\{X_1 = 2\}]P[\{X_2 = 2\}] \\
&+ P[\{X_1 = 3\}]P[\{X_2 = 1\}] \text{(because of independence)} \\
&= (0.25)(0.75)^2(0.25) + (0.75)(0.25)(0.75)(0.25) + (0.75)^2(0.25)(0.25) \\
&= 3 \times \frac{1}{16} \times \frac{9}{16} = \frac{27}{256}.
\end{aligned}
$$

(d) Use **R** to simulate the arrival of the signals over a 50 seconds interval (eg code "1" to denote a signal, "0" for no signal).

Within each second, the probability that there is a signal is 0.25, so within each second, the probability of getting a 1 is Bernoulli with $p = 0.25$, so this is Binomial Bin(1,0.25). Hence we need to simulate a sample of 50 Bernoulli Bern(0.25) or Binomial Bin(1,0.25) using the rbinom command in **R** . Note that your simulation results may be different.

```
> rbinom(50,1,0.25)
 [1] 1 0 1 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 1 1 1 0 0 0 0 0 0 1 0 0 0 0 0 1 1 0 1
[39] 0 0 0 1 1 0 1 0 0 0 0 1
```

My results shows a signal in the 1st second interval, then in the 3rd second interval, then in the 8th second interval and so on.

6. The number of customers that arrive at a shop is distributed as Poisson with mean rate of 5 per hour.

   Solution. Here $X \sim \text{Pois}(5)$.

   (a) What is the probability that at least 3 customers arrive at the shop within the hour?

   $$\begin{aligned} P[X \geq 3] &= 1 - P[X = 0] - P[X = 1] - P[X = 2] \\ &= 1 - P[X \leq 2] \\ &= 1 - \frac{e^{-5}5^0}{0!} - \frac{e^{-5}5^1}{1!} - \frac{e^{-5}5^2}{2!}. \end{aligned}$$

   Using **R** , we get

   ```
   > 1 - ppois(2,5)
   [1] 0.875348
   ```

   (b) What is the probability there is a most 2 customer arrivals within the hour? This is

   $$\begin{aligned} P[X \leq 2] &= P[X = 0] + P[X = 1] + P[X = 2] \\ &= \frac{e^{-5}5^0}{0!} + \frac{e^{-5}5^1}{1!} + \frac{e^{-5}5^2}{2!}. \end{aligned}$$

   Using **R** , we get

   ```
   >> ppois(2,5)
   [1] 0.124652
   ```

   This is also the complement of the event in part (a), so we can also take 1 minus the probability in part (a).

   (c) What is the most likely number of customers that arrive at the shop within the hour?

   Here we need to find x such that $P[X = x]$ is the maximum value. We use trial and error but we can try values around $X = 5$ since the mean is $\lambda = 5$ and the Poisson probability mass function is uni-modal, so quite likely its mode will be close to the mean. So it looks like we need to check out $p_X(4)$ and $p_X(5)$. We can use **R** .

   ```
   > dpois(4,5)
   [1] 0.1754674
   > dpois(5,5)
   [1] 0.1754674
   > dpois(6,5)
   [1] 0.1462228
   ```

It appears that both $X = 4$ and $X = 5$ are equally likely since they have the same probabilities and we can verify that they are equal.

$$P[X = 4] = \frac{e^{-5}5^4}{4!}$$

and

$$P[X = 5] = \frac{e^{-5}5^5}{5!} = \frac{e^{-5} \times 5 \times 5^4}{5 \times 4!} = \frac{e^{-5}5^4}{4!} = P[X = 4].$$

(d) Use **R** to simulate the number of customers arriving within each hour over a 10 hour period.

```
> rpois(10,5)
 [1] 3 6 6 4 7 1 4 5 2 8
```

My results show that I have 3 customers in the 1st hour, 6 in the 2nd hour, 6 in the 3rd hour and so on. Your simulation results may be different.

7. A shop has two entrances – front and back. The number of customers that arrive at the front entrance is distributed as Poisson with mean rate of 8 per hour. The number of customers that arrive at the back entrance is distributed as Poisson with mean rate of 6 per hour.

Solution. We can let $X_1 \sim \text{Pois}(8)$ be the number of customers arriving per hour via the front, and $X_2 \sim \text{Pois}(5)$ be the number of customers arriving per hour via the back. We store the respective simulation results in **R** as $x1$ and $x2$.

(a) Use **R** to simulate the number of customers arriving via the front entrance within each hour over a 10 hour period.

```
> x1 <- rpois(10,8)
> x1
 [1] 12 13  6  6  8  9  6  5 11  7
```

(b) Use **R** to simulate the number of customers arriving via the back entrance within each hour over a 10 hour period.

```
> x2 <- rpois(10,6)
> x2
 [1]  7  4  4  7 10  9  4  9  6  3
```

(c) Merge your results from parts (a) and (b) to obtain the total number of customers arriving into the shop within each hour over the 10 hour period. N.B. You need to store your results from parts (a) and (b) first while doing those parts.
To merge the results, we add $x1$ and $x2$.

```
> x1 + x2
 [1] 19 17 10 13 18 18 10 14 17 10
```