LAB100

Week 20: Probability calculations

Your aim for this workshop is to learn how to use RStudio to calculate probabilities relating to random variables.

- Calculate probabilities from the Binomial and Poisson densities
- Use the RStudio functions for the probability mass functions and cumulative mass functions
- Use `for` loops to calculate the probability of particular events

Remember that you need to change the working directory and open a new R script at the start of each workshop.

1 The Binomial Distribution

From the MATH104 lecture notes you know that the probability mass function (pmf) for a Binomial random variable X with size n and probability of success θ is:

$$\mathbb{P}(X = x) = \binom{n}{x} \theta^x (1-\theta)^{n-x} \quad \text{for} \quad x = 0, 1, \ldots, n.$$

When $n = 3$ and $\theta = 0.5$ we can calculate the pmf for each event in RStudio by:

```
choose(3,0) * 0.5^0 * (1 - 0.5)^(3 - 0)
choose(3,1) * 0.5^1 * (1 - 0.5)^(3 - 1)
choose(3,2) * 0.5^2 * (1 - 0.5)^(3 - 2)
choose(3,3) * 0.5^3 * (1 - 0.5)^(3 - 3)
```

The Binomial pmf is a standard function and can be evaluated using the `dbinom` command. It requires the following three arguments:

Argument	Definition
`x`	The value of the random variable
`size`	The number of Bernoulli trials
`prob`	The probability of success for each trial

Consequently, the pmf for the example above can be evaluated by:

```
binomialPMF <- dbinom( x=0:3, size=3, prob=0.5)
binomialPMF
barplot( binomialPMF, names.arg=0:3, main="Binomial(3, 0.5) PMF" )
```

Note that the name for this command is derived from calculating the `d`ensity of a `binom`ial distribution.

Rather than calculating the probability that the random variable is 'equal' to a number, we may wish to calculate the probability that the random variable is 'less than or equal to' a particular value. For example, when $X \sim \text{Binomial}(3, 0.5)$, what is the value for $\mathbb{P}(X \leq 1)$? Since the random variable is discrete, we can evaluate and add the probabilities $\mathbb{P}(X = 0)$ and $\mathbb{P}(X = 1)$.

```
dbinom( x=0, size=3, prob=0.5) + dbinom( x=1, size=3, prob=0.5)
sum( dbinom( x=0:1, size=3, prob=0.5) )
```

Alternatively, we can use the `pbinom` to calculate the cumulative `p`robability from a `binom`ial distribution. This command accepts the same `size` and `prob` arguments as the `dbinom` command, but the number in the probability statement is entered as the argument `q` (meaning `q`uantity). Therefore, to evaluate the above cumulative probability we can execute:

```
pbinom( q=1, size=3, prob=0.5)
```

The `pbinom` command only evaluates 'less than or equal to' events. Therefore, to calculate the probability $\mathbb{P}(X \geq 1)$ we use the law of total probability to determine the equalities:

$$\mathbb{P}(X \geq 1) = 1 - \mathbb{P}(X < 1) = 1 - \mathbb{P}(X \leq 0)$$

To evaluate this probability, we use the last statement as it contains a 'less than or equal to' event:

```
1 - pbinom( q=0, size=3, prob=0.75 )
```

2 The Poisson Distribution

The pmf of a Poisson Random variable X with rate parameter λ is:

$$\mathbb{P}(X = x) \quad = \quad \frac{\lambda^x \exp(-\lambda)}{x!} \quad \text{for} \quad x = 0, 1, 2, \ldots$$

When $\lambda = 4.6$, we can calculate the first three values of the pmf:

```
4.6^0 * exp(-4.6) / factorial( 0 )
4.6^1 * exp(-4.6) / factorial( 1 )
4.6^2 * exp(-4.6) / factorial( 2 )
```

Argument	Definition
x	The value of the random variable
lambda	The rate parameter

Similar to the Binomial distribution, the Poisson pmf is a standard function and can be evaluated using the `dpois` command. The name for this command is derived from calculating the density of a poisson distribution and it accepts the following arguments:

The first 10 probabilities from a Poisson pmf with $\lambda = 4.6$ can be evaluated by:

```
PoissonPMF <- dpois( x=0:9, lambda=4.6)
PoissonPMF
barplot( PoissonPMF, names.arg=0:9, main="Poisson(4.6) PMF")
```

Cumulative probabilities for a Poisson random variable can be determined using the `ppois` command. Can you derive the name of this command? The input arguments for this command are `q`, the quantity that we want to evaluate less than or equal to, and the rate parameter `labmda`.

Consider the random variable $X \sim \text{Poisson}(4.6)$, what is the value for $\mathbb{P}\left(X \leq 3\right)$? This can be derived by two methods: by calculating the sum of the pmf for when the random variable is equal to 0, 1, 2 and 3; or, by using the `ppois` command which evaluates the cumulative probability from a poisson distribution.

```
sum( dpois( x=0:3, lambda=4.6 ) )
ppois( q=3, lambda=4.6 )
```

Like the `pbinom` command, `ppois` evaluates 'less than or equal to' events. Therefore, to calculate the probability $\mathbb{P}\left(X \geq 6\right)$ we need to use the law of total probability to derive the alternative expression $1 - \mathbb{P}\left(X \leq 5\right)$.

```
1 - ppois( q=5, lambda=4.6 )
```

3 Simulating from a Distribution

In previous Workshops we have already seen how to simulate from a uniform random variable using `runif` and how to `sample` from a probability space. We can also simulate realisations of binomial and Poisson random variables.

The function `rbinom` can be used for binomial random variables, it takes three arguments.

Argument	Definition
n	The number of random values to return
size	The number of Bernoulli trials
prob	The probability of success for each trial

For instance, to simulate 1000 realisations of tossing an unbiased coin 10 times and counting the number of heads we can use

```
simulation <- rbinom(1000,10,0.5)
```

Similarly, the function `rpois` generates realisations from a Poisson random variable. For this we require the number of random values to return (`n`) and the rate parameter from the Poisson distribution (`lambda`). For example, assume that the number of floods of the River Lune in a year follows a Poisson distribution with rate parameter $\lambda = 1.5$. To simulate data for a 100 year period we would execute:

```
Floods <- rpois( n=100, lambda=1.5)
```

Note that we can still use `set.seed` if we want to replicate a particular sequence of random variables. However it is worth noting that equivalent simulations generated with the same seed but using a different technique will give different values. For instance, we can simulate from a Binomial distribution either by using `sample` with replacement or by using `rbinom`

```
set.seed(103)
bn.sim <- sum(sample(0:1,20,replace=TRUE,prob=c(0.3,0.7)))
set.seed(103)
bn.sim2 <- rbinom(1,20,0.7)
```

while these each simulate the same thing with the same seed they give different results. This is because, while they are both using the same sequence of pseudo-random numbers, they are using them in a different order.

4 Probability calculations

We can use the programming tools that have been introduced in previous workshops to calculate probabilities that would be difficult to compute just with a standard calculator.

For instance, suppose we have two independent random variables X and Y. Let $X \sim \text{Po}(\lambda)$ and $Y \sim \text{Bin}(n, p)$. Suppose we wish to calculate $\mathbb{P}(X + Y = 5)$. In principle we can find this probability by summing over all the individual combinations of X and Y that can lead to 5 i.e.

$$\mathbb{P}(X + Y = 5) = \sum_{x=0}^{5} \mathbb{P}(X = x \,\cap\, Y = 5 - x) = \sum_{x=0}^{5} \mathbb{P}(X = x)\mathbb{P}(Y = 5 - x)$$

we can use RStudio to calculate this probability. For instance let $\lambda = 4, n = 6$ and $p = 0.3$. Either we can use a `for` loop:

```
prob <- 0
for (x in 0:5) {
    prob <- prob + dpois(x,4)*dbinom(5-x,6,0.3)
}
```

alternative we can use `sum`

```
prob <- sum(dpois(0:5,4)*dbinom(5:0,6,0.3)
```

In each case we get the answer 0.1725.

Quiz 1: Simulate Poisson random variables

Set the random seed to 123 and then use `rpois` to simulate 100 realisation from a Poisson distribution with rate parameter $\lambda = 0.5$. What is the sum of the resulting variables?

(**A**) 51 (**B**) 53 (**C**) 49 (**D**) 42 (**E**) 47

Quiz 2: Binomial probability

Let $X \sim \text{Bin}(15, 0.4)$ find $\mathbb{P}(4 \leq X \leq 10)$

Quiz 3: Composite random variable I

Let $X \sim \text{Bin}(10, 0.2)$ and $Y \sim \text{Po}(3)$ be independent random variables. If $Z = XY$ calculate $\mathbb{P}(Z = 8)$.

Quiz 4: Composite random variable II

Let $X \sim \text{Bin}(20, 0.4)$ and $Y \sim \text{Bin}(19, 0.7)$ be independent random variables. If $Z = X + Y$, calculate the probability that Z is prime.

Quiz 5: Poisson probability

Suppose $X \sim \text{Po}(\lambda)$, find to 2 d.p. the value of λ such that $\mathbb{P}(X < 5) = 0.5$
