In this document, we'll focus on using `R` to generate a) probabilities and b) realisations of random variables. `R` is capable of doing both of these very quickly, if you let it know what type of distribution you're interested in

Part 1: Discrete RV Probabilities

In this section, we will focus on probabilities for discrete and continuous distributions.
If we wanted to, there's nothing stopping us from directly programming probability mass functions into `R`. We have from the Week 3 videos, for instance, that if $X \sim Pois(\lambda)$, then

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

We could write the function

```
poisprobs<-function(x,lambda){
ans<-(exp(-lambda)*(lambda)^x)/factorial(x)
return(ans)}
```

1. Enter the above function into `R`, and use it to find $P(X = 4)$ when $\lambda = 3$.

2. Try writing a function in `R`, with inputs `x,n,p`, which returns the probability $P(X = x)$ when $X \sim Bin(n, p)$. **Hint:** the binomial coefficient $\binom{n}{k}$ can be found in `R` by inputting `choose(n,k)`.

Writing functions like this every time would be a bit of a pain, though, so someone else has done it for us!

3 Try inputting `dpois(4,3)`. What do you notice?

The "d" in `dpois` stands for "density". In the videos, I have made a distinction between the probability **mass** functions associated with discrete random variables, and the probability **density** functions associated with continuous random variables. `R`, in contrast, does not make this distinction.

4 Try inputting `dbinom(2,5,0.8)`. What does this value represent? Can you get the same value using the code you wrote for question 2?

One annoying task when dealing with discrete distributions is finding interval probabilities. For instance, if we have $X \sim Pois(8)$, how do we find the probability $P(X < 12)$? This isn't actually **difficult** to do, using the additive rule:

$$P(X < 12) = \sum_{i=0}^{11} P(X = i).$$

Finding each of the 12 probabilities involved is a bit of a slog, though.

Fortunately, `R` has us covered! We could calculate such a probability using a `for` loop, as follows:

```
ans<-0
for(i in 0:11){
ans<-ans+dpois(i,8)}
ans
```

We don't even need to do that much, though; we can just use the function `ppois`.

5 Enter the code above and use it to find $P(X < 12)$.

6 Use `ppois` to find the same value. Which inputs do you need to use with `ppois`, and in what order?

Here the "p" stands for "probability", because this function gives us probabilities. This is a bit confusing, perhaps, given `dpois` **also** gave us probabilities. When we come to look at continuous distributions, though, the density function itself won't give us probabilities, and we'll have to use the "probability function" in `R` instead. Note also thet what `R` calls the **probability function** is what we've defined in videos as being the **cumulative distribution function**.

7 Using just one line of code, find $P(X < 309)$ for $X \sim Bin(400, 0, 9)$.

Another useful set of functions here is `qpois`, `qbinom`, and so forth. Here, "q" stands for "quantile". We've talked about quantiles more than once before. In the context of random variables, a quantile of value i (where i is between 0 and 1) is the value the random variable will be equal to or less than $100i\%$ of the time.

This actually makes the probability function and the quantile distribution are inverses of each other.

8 Find the value of `qpois(ppois(11,8),8)`

9 Generate `x<-seq(0.23,0.24,by=0.001)`. Find the value of `qpois(x,4)`. Explain the nature of the resulting sequence - why the jump from 2 to 3?

Part 2: Continuous RV Probabilities

For continuous random variables, the density function no longer gives us probabilities as outputs. Instead, a function such as `dnorm` gives us the density function of a normal distribution.

We can draw a graph of a density function inside `R` quite quickly, using code such as

```
x<-seq(-3.5,3.5,by=0.01)
plot(x,dnorm(x),type=''l'')
```

9 Try drawing this same plot without the argument `type=''l''`. What difference does this make?

10 Use the `dnorm` help file to learn how to express the density function for X $N(\mu, \sigma^2)$, where $\mu \neq 0$ and $\sigma^2 \neq 1$. Draw the density function for one such plot.

Probabilities for continuous random variables have to be considered in terms of intervals. A function like rnorm makes that very simple.

11 For $X \sim N(0, 1)$, find $P(X < 1)$.

12 For $X \sim N(2, 6)$, find $P(0 < X < 3)$.

13 For $X \sim N(-1, 2)$, find $P(X > 0)$.

Note that, unlike with discrete random variables, there is no difference between being asked to find, say, $P(a \leq X \leq b)$, $P(a \leq X < b)$, $P(a < X \leq b)$, and $P(a < X < b)$ for a continuous random variable X. This is because, for example

$$P(X \leq x) = P(X < x) + P(X = x)$$

(using the additive rule), and the probability of a continuous random variable taking a specific value is always zero.

14 Generate `x<-seq(0.23,0.24,by=0.001)`. Find the value of `qexp(x,4)` (this is the quantile function for the exponential distribution). Explain the nature of the resulting sequence - why are there no jumps this time?

Part 3: Finding RV Realisations

We've now covered three of the four functions associated with binomial, Poisson, normal, and exponential distributions (along with many others) - the density function, the probability function, and the quantile function.

The remaining function is the **random generation** function, and we've actually seen this before more than once. This function allows us to generate some number n of realisations from a specified distribution.

15 Input `rexp(100,2)`.

16 Draw a histogram for the 100 values you have generated.

17 Repeat questions 15 and 16 for 1 000, 10 000, 100 000, and 1 000 000 values. What do you notice about the histograms?

18 Use the `dexp` function to draw the probability density function for $X \sim Exp(2)$.

We've used histograms before as rough estimators for probability density functions. They can also be used to generate rough probability estimates when we don't know the underlying distribution.

We'll look at an example of this in a situation where we **do** know the underlying distribution, in order to make what's going on clear. In general, though, we'd only use a histogram to estimate probabilities in a situation where we didn't have a probability distribution to use instead.

19 Generate 1000 values for the standard normal, and draw a histogram of this data, using rectangles of width 0.5, beginning at -4 and ending at 4.

20 Find how many values lie in the rectangles between -4 and -1. Use this to estimate the probability $P(Z < -1)$, where $Z \sim N(0, 1)$.

21 Find the actual value of $P(Z < -1)$ using `pnorm`, and compare it to your answer for part 20.