

*Synchronous Online: Start 12pm, Wednesday 4th November*

# MAS2905: Probability and Bayesian Methods with R

Newcastle University  
`lee.fawcett@ncl.ac.uk`

Semester 1, 2020/21

# The story so far...

- You should now have completed **all of Chapter 2**
- We've looked at the basics of programming R:
  - Data types in R
  - Vectors and dataframes
  - Data summaries
  - Functions, for loops and if statements

# The story so far...

- You should now have completed **all of Chapter 2**
- We've looked at the basics of programming R:
  - Data types in R
  - Vectors and dataframes
  - Data summaries
  - Functions, for loops and if statements

# The story so far...

- You should now have completed **all of Chapter 2**
- We've looked at the basics of programming R:
  - Data types in R
  - Vectors and dataframes
  - Data summaries
  - Functions, for loops and if statements

# The story so far...

- **Test 1** will cover all of this, and a bit from today's lecture
  - Completed in Numbas
  - Goes live tomorrow, 16:00
  - **Deadline:** 16:00, Thursday 12 November
- We're now ready to move on – you should have already looked over Sections 3.1–3.2

# The story so far...

- **Test 1** will cover all of this, and a bit from today's lecture
  - Completed in Numbas
  - Goes live tomorrow, 16:00
  - **Deadline:** 16:00, Thursday 12 November
- We're now ready to move on – you should have already looked over Sections 3.1–3.2

# The story so far...

- **Test 1** will cover all of this, and a bit from today's lecture
  - Completed in Numbas
  - Goes live tomorrow, 16:00
  - **Deadline:** 16:00, Thursday 12 November
- We're now ready to move on – you should have already looked over Sections 3.1–3.2

## 3.1 The Binomial distribution

If  $X \sim \text{Bin}(n, p)$  then  $X$  has PMF (probability mass function) given by

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad \text{for } k = 0, 1, \dots, n.$$

There is no closed formula for the CDF (cumulative distribution function).

R commands:

- `dbinom` – calculate PMF
- `pbinom` – calculate CDF
- `rbinom` – generate random sample

## 3.1 The Binomial distribution

If  $X \sim \text{Bin}(n, p)$  then  $X$  has PMF (probability mass function) given by

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad \text{for } k = 0, 1, \dots, n.$$

There is no closed formula for the CDF (cumulative distribution function).

R commands:

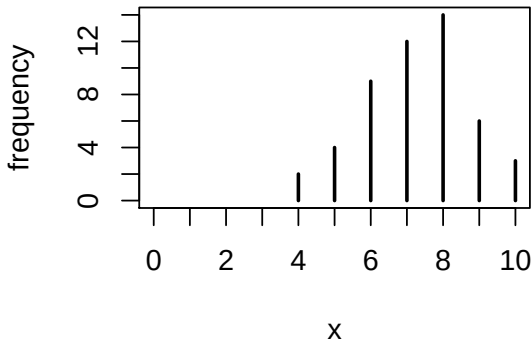
- `dbinom` – calculate PMF
- `pbinom` – calculate CDF
- `rbinom` – generate random sample

## 3.1 Commands for the Binomial distribution

```
1 dbinom(5, 10, 0.7)      #  $\Pr(X = 5), X \sim \text{Bin}(10, 0.7)$   
2 pbinom(4, 10, 0.7)     #  $\Pr(X \leq 4), X \sim \text{Bin}(10, 0.7)$   
3 rbinom(50, 10, 0.7)    # Sample a  $\text{Bin}(10, 0.7)$  distribution 50 times
```

## 3.1 Creating a bar plot

```
1 x = rbinom(50, 10, 0.7)
2 plot(table(x), xlim = c(0,10), xaxt = 'n', xlab = 'x',
3      ylab = 'frequency')
4 axis(1, at = seq(0, 10))
```



## 3.1 The Geometric distribution

If  $Y \sim \text{Geom}(p)$  then  $Y$  has PMF and CDF given by

$$p_Y(k) = (1 - p)^{k-1}p,$$

$$F_Y(k) = 1 - (1 - p)^k, \quad \text{for } k = 1, 2, \dots$$

Note that  $Y$  takes values in  $1, 2, 3, \dots$ :

- R uses a slightly different definition
- We use the definition that  $Y$  is the number of Bernoulli trials up to and including the first success
- R counts number of trials up to, but **not** including the first success, so in R geometric random variables take values  $0, 1, 2, \dots$

## 3.1 R commands for the Geometric distribution

Adjust the arguments to account for different definition:

```
1 dgeom(4, 0.2)           #  $\Pr(Y = 5), Y \sim \text{Geom}(0.2)$   
2 pgeom(2, 0.2)           #  $\Pr(Y \leq 3), Y \sim \text{Geom}(0.2)$   
3 1 + rgeom(100, 0.2)     # Sample a  $\text{Geom}(0.2)$  distribution 100 times
```

Here's a function to replace `dgeom` with our definition of the geometric distribution:

```
1 mydgeom = function(x, p) {  
2   dgeom(x-1, p)}  
}
```

## 3.1 The Poisson distribution

If  $Z \sim Po(\lambda)$  then  $Z$  has PMF given by

$$p_Z(k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad \text{for } k = 0, 1, 2, \dots$$

There is no closed formula for the CDF (cumulative distribution function).

R commands:

- `dpois` – calculate PMF
- `ppois` – calculate CDF
- `rpois` – generate random sample

## 3.1 R commands for the Poisson distribution

```
1 dpois(5, 3.5)      #  $\Pr(Z = 5), Z \sim Po(3.5)$   
2 ppois(2, 3.5)      #  $\Pr(Z \leq 2), Z \sim Po(3.5)$   
3 rpois(100, 3.5)    # Sample a  $Po(3.5)$  distribution 100 times
```

## 3.1 Summary

| Distribution | Binomial                 | Poisson                 | Geometric  |
|--------------|--------------------------|-------------------------|-----------------------------------------------------------------------------------------------|
| PMF          | <code>dbinom(...)</code> | <code>dpois(...)</code> | <code>dgeom(...)</code>                                                                       |
| CDF          | <code>pbinom(...)</code> | <code>ppois(...)</code> | <code>pgeom(...)</code>                                                                       |
| sample       | <code>rbinom(...)</code> | <code>rpois(...)</code> | <code>rgeom(...)</code>                                                                       |

## 3.2 Continuous random variables

R has functions for the uniform, exponential and normal distributions:

| Distribution | Uniform                 | Exponential            | Normal                  |
|--------------|-------------------------|------------------------|-------------------------|
| PDF          | <code>dunif(...)</code> | <code>dexp(...)</code> | <code>dnorm(...)</code> |
| CDF          | <code>punif(...)</code> | <code>pexp(...)</code> | <code>pnorm(...)</code> |
| quantile     | <code>qunif(...)</code> | <code>qexp(...)</code> | <code>qnorm(...)</code> |
| sample       | <code>runif(...)</code> | <code>rexp(...)</code> | <code>rnorm(...)</code> |

## 3.2 Continuous random variables – R examples

```
1 dunif(3, 2, 5)      #  $f_X(3), X \sim U(2, 5)$   
2 pexp(1.5, 5)      #  $F_Y(1.5), Y \sim \text{Exp}(5)$   
3 rnorm(10, 0, 2)    # Sample a  $N(0, 2^2)$  distribution 10 times
```

For the standard uniform ( $U(0, 1)$ ) and standard normal ( $N(0, 1)$ ) distributions you don't need to provide the parameters  $a = 0$ ,  $b = 1$  and  $\mu = 0$ ,  $\sigma = 1$  respectively. For example:

```
1 runif(20)          # Samples a  $U(0, 1)$  distribution 20 times  
2 pnorm(1.96)       #  $\Pr(Z < 1.96), Z \sim N(0, 1)$   
3 [1] 0.9750021
```

## 3.2 Quantiles

The quantile functions `qunif`, `qexp`, `qnorm` solve equations like

$$F_X(\alpha) = p$$

for  $\alpha$  given a probability  $p$ . For example:

```
1 qnorm(0.9750021)
2 [1] 1.96
```

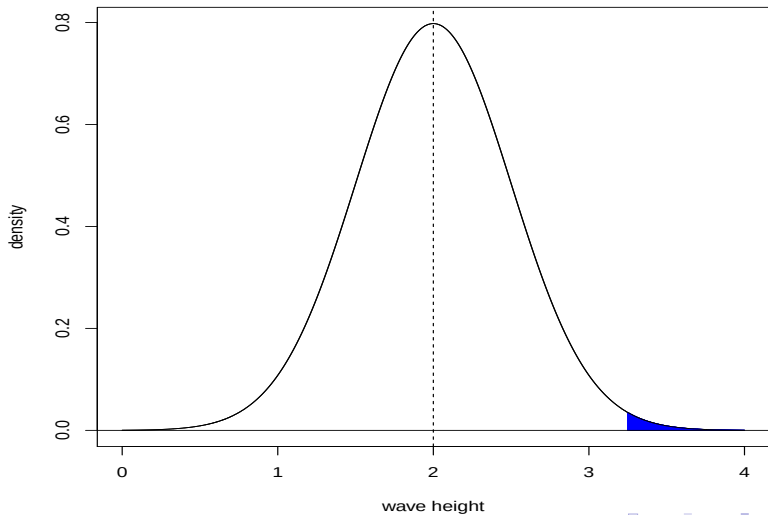
## Example 3.1

### Example

Suppose annual maximum wave heights observed off the coast at a flood-prone town are assumed Normally distributed, with mean 2 metres and standard deviation 0.5 metres.

- (a) Write down the R command to find the probability that the largest wave height next year will exceed 3.25 metres.
- (b) Write down the R command to estimate the height of a new sea wall such that we might expect the town to be flooded, on average, once per century. Why might our modelling assumption be invalid?

## Example 3.1: Solution (a)



## Example 3.1: Solution (a)

Could work out the old-fashioned way:

$$\begin{aligned}\Pr(X > 3.25) &= \Pr\left(Z > \frac{3.25 - 2}{0.5}\right) \\ &= \Pr(Z > 2.5) \\ &= 1 - \Pr(Z \leq 2.5) = 1 - 0.994 = 0.006.\end{aligned}$$

Or could just use R:

```
1 1-pnorm(3.25, 2, 0.5)
2 [1] 0.006209665
```

## Example 3.1: Solution (a)

Could work out the old-fashioned way:

$$\begin{aligned}\Pr(X > 3.25) &= \Pr\left(Z > \frac{3.25 - 2}{0.5}\right) \\ &= \Pr(Z > 2.5) \\ &= 1 - \Pr(Z \leq 2.5) = 1 - 0.994 = 0.006.\end{aligned}$$

Or could just use R:

```
1 1-pnorm(3.25, 2, 0.5)
2 [1] 0.006209665
```

## Example 3.1: Solution (a)

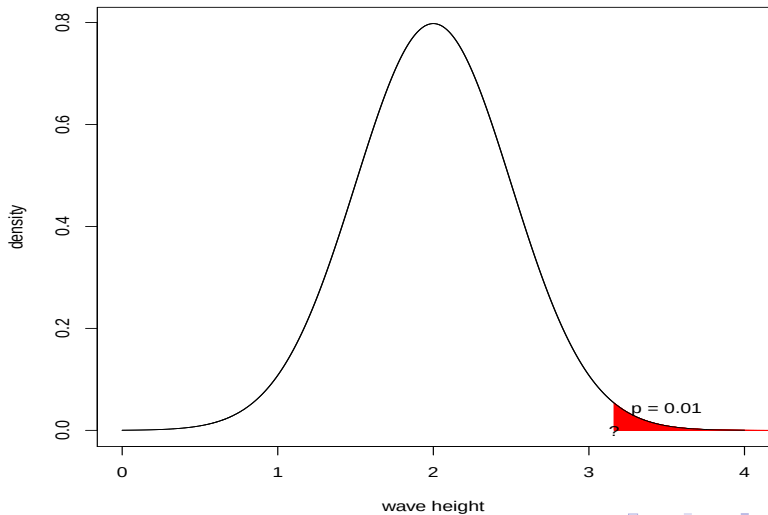
Could work out the old-fashioned way:

$$\begin{aligned}\Pr(X > 3.25) &= \Pr\left(Z > \frac{3.25 - 2}{0.5}\right) \\ &= \Pr(Z > 2.5) \\ &= 1 - \Pr(Z \leq 2.5) = 1 - 0.994 = 0.006.\end{aligned}$$

Or could just use R:

```
1 1-pnorm(3.25, 2, 0.5)
2 [1] 0.006209665
```

## Example 3.1: Solution (b)



## Example 3.1: Solution (b)

Similarly:

```
1 qnorm(0.99, 2, 0.5)
2 [1] 3.163174
```

# Quiz #1

What is the correct R code to find  $\Pr(X < 4)$ , when  $X \sim \text{Bin}(12, 0.3)$ ?



1. `rbinom(4, 12, 0.3)`
2. `pbinom(3, 12, 0.3)`
3. `pbinom(4, 12, 0.3)`
4. `dbinom(3, 0.3, 12)`

# Quiz #1

What is the correct R code to find  $\Pr(X < 4)$ , when  $X \sim \text{Bin}(12, 0.3)$ ?



1. `rbinom(4, 12, 0.3)`
2. `pbinom(3, 12, 0.3)`
3. `pbinom(4, 12, 0.3)`
4. `dbinom(3, 0.3, 12)`

## Quiz #2

Suppose the time (in hours) between collision 'near misses' at a known road safety hotspot in Newcastle is exponentially distributed with rate 0.01.

What is the correct R code to find the probability that the time between 'near misses' will exceed 5 days?



1. `rexp(5, 0.01)`
2. `1 - pexp(5, 0.01)`
3. `1 - pexp(120, 0.01)`
4. `pexp(120, 0.99)`

## Quiz #2

Suppose the time (in hours) between collision 'near misses' at a known road safety hotspot in Newcastle is exponentially distributed with rate 0.01.

What is the correct R code to find the probability that the time between 'near misses' will exceed 5 days?



1. `rexp(5, 0.01)`
2. `1 - pexp(5, 0.01)`
3. `1 - pexp(120, 0.01)`
4. `pexp(120, 0.99)`

## Example 3.2: Solution

Example performed live in RStudio during synchronous online lecture.

## Example 3.3: Solution

| Model                        | Question                                     | R code                            | Answer              |
|------------------------------|----------------------------------------------|-----------------------------------|---------------------|
| $X \sim \text{Po}(3)$        | Find $\Pr(X = 6)$ .                          | <code>dpois(6,3)</code>           | 0.0504              |
|                              | Find $\Pr(X < 5)$ .                          | <code>ppois(4,3)</code>           | 0.8153              |
|                              | Generate 3 observations.                     | <code>rpois(3,3)</code>           | 2, 9, 3             |
| $Y \sim \text{Bin}(10, 0.2)$ | Find $\Pr(Y \leq 4)$ .                       | <code>pbinom(4,10,0.2)</code>     | 0.9672              |
|                              | Find $\Pr(Y \geq 4)$ .                       | <code>1-pbinom(3,10,0.2)</code>   | 0.1209              |
|                              | Find $y$ such that $\Pr(Y \leq y) = 0.677$ . | <code>qbinom(0.677,10,0.2)</code> | 2                   |
| $Z \sim N(0, 1)$             | Find $\Pr(Z \leq 0.2)$ .                     | <code>pnorm(0.2)</code>           | 0.5793              |
|                              | Find $\Pr(Z > -2.5)$ .                       | <code>1-pnorm(-2.5)</code>        | 0.9938              |
|                              | Generate 3 observations.                     | <code>rnorm(3)</code>             | 0.993 -2.523 -2.468 |

## Example 3.3: Solution

| Model                        | Question                                     | R code                            | Answer              |
|------------------------------|----------------------------------------------|-----------------------------------|---------------------|
| $X \sim \text{Po}(3)$        | Find $\Pr(X = 6)$ .                          | <code>dpois(6,3)</code>           | 0.0504              |
|                              | Find $\Pr(X < 5)$ .                          | <code>ppois(4,3)</code>           | 0.8153              |
|                              | Generate 3 observations.                     | <code>rpois(3,3)</code>           | 2, 9, 3             |
| $Y \sim \text{Bin}(10, 0.2)$ | Find $\Pr(Y \leq 4)$ .                       | <code>pbinom(4,10,0.2)</code>     | 0.9672              |
|                              | Find $\Pr(Y \geq 4)$ .                       | <code>1-pbinom(3,10,0.2)</code>   | 0.1209              |
|                              | Find $y$ such that $\Pr(Y \leq y) = 0.677$ . | <code>qbinom(0.677,10,0.2)</code> | 2                   |
| $Z \sim N(0, 1)$             | Find $\Pr(Z \leq 0.2)$ .                     | <code>pnorm(0.2)</code>           | 0.5793              |
|                              | Find $\Pr(Z > -2.5)$ .                       | <code>1-pnorm(-2.5)</code>        | 0.9938              |
|                              | Generate 3 observations.                     | <code>rnorm(3)</code>             | 0.993 -2.523 -2.468 |

## Example 3.3: Solution

| Model                        | Question                                     | R code                            | Answer              |
|------------------------------|----------------------------------------------|-----------------------------------|---------------------|
| $X \sim \text{Po}(3)$        | Find $\Pr(X = 6)$ .                          | <code>dpois(6,3)</code>           | 0.0504              |
|                              | Find $\Pr(X < 5)$ .                          | <code>ppois(4,3)</code>           | 0.8153              |
|                              | Generate 3 observations.                     | <code>rpois(3,3)</code>           | 2, 9, 3             |
| $Y \sim \text{Bin}(10, 0.2)$ | Find $\Pr(Y \leq 4)$ .                       | <code>pbinom(4,10,0.2)</code>     | 0.9672              |
|                              | Find $\Pr(Y \geq 4)$ .                       | <code>1-pbinom(3,10,0.2)</code>   | 0.1209              |
|                              | Find $y$ such that $\Pr(Y \leq y) = 0.677$ . | <code>qbinom(0.677,10,0.2)</code> | 2                   |
| $Z \sim N(0, 1)$             | Find $\Pr(Z \leq 0.2)$ .                     | <code>pnorm(0.2)</code>           | 0.5793              |
|                              | Find $\Pr(Z > -2.5)$ .                       | <code>1-pnorm(-2.5)</code>        | 0.9938              |
|                              | Generate 3 observations.                     | <code>rnorm(3)</code>             | 0.993 -2.523 -2.468 |

## Example 3.3: Solution

| Model                        | Question                                     | R code                            | Answer              |
|------------------------------|----------------------------------------------|-----------------------------------|---------------------|
| $X \sim \text{Po}(3)$        | Find $\Pr(X = 6)$ .                          | <code>dpois(6,3)</code>           | 0.0504              |
|                              | Find $\Pr(X < 5)$ .                          | <code>ppois(4,3)</code>           | 0.8153              |
|                              | Generate 3 observations.                     | <code>rpois(3,3)</code>           | 2, 9, 3             |
| $Y \sim \text{Bin}(10, 0.2)$ | Find $\Pr(Y \leq 4)$ .                       | <code>pbinom(4,10,0.2)</code>     | 0.9672              |
|                              | Find $\Pr(Y \geq 4)$ .                       | <code>1-pbinom(3,10,0.2)</code>   | 0.1209              |
|                              | Find $y$ such that $\Pr(Y \leq y) = 0.677$ . | <code>qbinom(0.677,10,0.2)</code> | 2                   |
| $Z \sim N(0, 1)$             | Find $\Pr(Z \leq 0.2)$ .                     | <code>pnorm(0.2)</code>           | 0.5793              |
|                              | Find $\Pr(Z > -2.5)$ .                       | <code>1-pnorm(-2.5)</code>        | 0.9938              |
|                              | Generate 3 observations.                     | <code>rnorm(3)</code>             | 0.993 -2.523 -2.468 |

## Example 3.3: Solution

| Model                        | Question                                     | R code                            | Answer              |
|------------------------------|----------------------------------------------|-----------------------------------|---------------------|
| $X \sim \text{Po}(3)$        | Find $\Pr(X = 6)$ .                          | <code>dpois(6,3)</code>           | 0.0504              |
|                              | Find $\Pr(X < 5)$ .                          | <code>ppois(4,3)</code>           | 0.8153              |
|                              | Generate 3 observations.                     | <code>rpois(3,3)</code>           | 2, 9, 3             |
| $Y \sim \text{Bin}(10, 0.2)$ | Find $\Pr(Y \leq 4)$ .                       | <code>pbinom(4,10,0.2)</code>     | 0.9672              |
|                              | Find $\Pr(Y \geq 4)$ .                       | <code>1-pbinom(3,10,0.2)</code>   | 0.1209              |
|                              | Find $y$ such that $\Pr(Y \leq y) = 0.677$ . | <code>qbinom(0.677,10,0.2)</code> | 2                   |
| $Z \sim N(0, 1)$             | Find $\Pr(Z \leq 0.2)$ .                     | <code>pnorm(0.2)</code>           | 0.5793              |
|                              | Find $\Pr(Z > -2.5)$ .                       | <code>1-pnorm(-2.5)</code>        | 0.9938              |
|                              | Generate 3 observations.                     | <code>rnorm(3)</code>             | 0.993 -2.523 -2.468 |

## Example 3.3: Solution

| Model                        | Question                                     | R code                            | Answer              |
|------------------------------|----------------------------------------------|-----------------------------------|---------------------|
| $X \sim \text{Po}(3)$        | Find $\Pr(X = 6)$ .                          | <code>dpois(6,3)</code>           | 0.0504              |
|                              | Find $\Pr(X < 5)$ .                          | <code>ppois(4,3)</code>           | 0.8153              |
|                              | Generate 3 observations.                     | <code>rpois(3,3)</code>           | 2, 9, 3             |
| $Y \sim \text{Bin}(10, 0.2)$ | Find $\Pr(Y \leq 4)$ .                       | <code>pbinom(4,10,0.2)</code>     | 0.9672              |
|                              | Find $\Pr(Y \geq 4)$ .                       | <code>1-pbinom(3,10,0.2)</code>   | 0.1209              |
|                              | Find $y$ such that $\Pr(Y \leq y) = 0.677$ . | <code>qbinom(0.677,10,0.2)</code> | 2                   |
| $Z \sim N(0, 1)$             | Find $\Pr(Z \leq 0.2)$ .                     | <code>pnorm(0.2)</code>           | 0.5793              |
|                              | Find $\Pr(Z > -2.5)$ .                       | <code>1-pnorm(-2.5)</code>        | 0.9938              |
|                              | Generate 3 observations.                     | <code>rnorm(3)</code>             | 0.993 -2.523 -2.468 |

## Example 3.3: Solution

| Model                        | Question                                     | R code                            | Answer              |
|------------------------------|----------------------------------------------|-----------------------------------|---------------------|
| $X \sim \text{Po}(3)$        | Find $\Pr(X = 6)$ .                          | <code>dpois(6,3)</code>           | 0.0504              |
|                              | Find $\Pr(X < 5)$ .                          | <code>ppois(4,3)</code>           | 0.8153              |
|                              | Generate 3 observations.                     | <code>rpois(3,3)</code>           | 2, 9, 3             |
| $Y \sim \text{Bin}(10, 0.2)$ | Find $\Pr(Y \leq 4)$ .                       | <code>pbinom(4,10,0.2)</code>     | 0.9672              |
|                              | Find $\Pr(Y \geq 4)$ .                       | <code>1-pbinom(3,10,0.2)</code>   | 0.1209              |
|                              | Find $y$ such that $\Pr(Y \leq y) = 0.677$ . | <code>qbinom(0.677,10,0.2)</code> | 2                   |
| $Z \sim N(0, 1)$             | Find $\Pr(Z \leq 0.2)$ .                     | <code>pnorm(0.2)</code>           | 0.5793              |
|                              | Find $\Pr(Z > -2.5)$ .                       | <code>1-pnorm(-2.5)</code>        | 0.9938              |
|                              | Generate 3 observations.                     | <code>rnorm(3)</code>             | 0.993 -2.523 -2.468 |

## Example 3.3: Solution

| Model                        | Question                                     | R code                            | Answer              |
|------------------------------|----------------------------------------------|-----------------------------------|---------------------|
| $X \sim \text{Po}(3)$        | Find $\Pr(X = 6)$ .                          | <code>dpois(6,3)</code>           | 0.0504              |
|                              | Find $\Pr(X < 5)$ .                          | <code>ppois(4,3)</code>           | 0.8153              |
|                              | Generate 3 observations.                     | <code>rpois(3,3)</code>           | 2, 9, 3             |
| $Y \sim \text{Bin}(10, 0.2)$ | Find $\Pr(Y \leq 4)$ .                       | <code>pbinom(4,10,0.2)</code>     | 0.9672              |
|                              | Find $\Pr(Y \geq 4)$ .                       | <code>1-pbinom(3,10,0.2)</code>   | 0.1209              |
|                              | Find $y$ such that $\Pr(Y \leq y) = 0.677$ . | <code>qbinom(0.677,10,0.2)</code> | 2                   |
| $Z \sim N(0, 1)$             | Find $\Pr(Z \leq 0.2)$ .                     | <code>pnorm(0.2)</code>           | 0.5793              |
|                              | Find $\Pr(Z > -2.5)$ .                       | <code>1-pnorm(-2.5)</code>        | 0.9938              |
|                              | Generate 3 observations.                     | <code>rnorm(3)</code>             | 0.993 -2.523 -2.468 |

## Example 3.3: Solution

| Model                        | Question                                     | R code                            | Answer              |
|------------------------------|----------------------------------------------|-----------------------------------|---------------------|
| $X \sim \text{Po}(3)$        | Find $\Pr(X = 6)$ .                          | <code>dpois(6,3)</code>           | 0.0504              |
|                              | Find $\Pr(X < 5)$ .                          | <code>ppois(4,3)</code>           | 0.8153              |
|                              | Generate 3 observations.                     | <code>rpois(3,3)</code>           | 2, 9, 3             |
| $Y \sim \text{Bin}(10, 0.2)$ | Find $\Pr(Y \leq 4)$ .                       | <code>pbinom(4,10,0.2)</code>     | 0.9672              |
|                              | Find $\Pr(Y \geq 4)$ .                       | <code>1-pbinom(3,10,0.2)</code>   | 0.1209              |
|                              | Find $y$ such that $\Pr(Y \leq y) = 0.677$ . | <code>qbinom(0.677,10,0.2)</code> | 2                   |
| $Z \sim N(0, 1)$             | Find $\Pr(Z \leq 0.2)$ .                     | <code>pnorm(0.2)</code>           | 0.5793              |
|                              | Find $\Pr(Z > -2.5)$ .                       | <code>1-pnorm(-2.5)</code>        | 0.9938              |
|                              | Generate 3 observations.                     | <code>rnorm(3)</code>             | 0.993 -2.523 -2.468 |

## Example 3.3: Solution

| Model                        | Question                                     | R code                            | Answer              |
|------------------------------|----------------------------------------------|-----------------------------------|---------------------|
| $X \sim \text{Po}(3)$        | Find $\Pr(X = 6)$ .                          | <code>dpois(6,3)</code>           | 0.0504              |
|                              | Find $\Pr(X < 5)$ .                          | <code>ppois(4,3)</code>           | 0.8153              |
|                              | Generate 3 observations.                     | <code>rpois(3,3)</code>           | 2, 9, 3             |
| $Y \sim \text{Bin}(10, 0.2)$ | Find $\Pr(Y \leq 4)$ .                       | <code>pbinom(4,10,0.2)</code>     | 0.9672              |
|                              | Find $\Pr(Y \geq 4)$ .                       | <code>1-pbinom(3,10,0.2)</code>   | 0.1209              |
|                              | Find $y$ such that $\Pr(Y \leq y) = 0.677$ . | <code>qbinom(0.677,10,0.2)</code> | 2                   |
| $Z \sim N(0, 1)$             | Find $\Pr(Z \leq 0.2)$ .                     | <code>pnorm(0.2)</code>           | 0.5793              |
|                              | Find $\Pr(Z > -2.5)$ .                       | <code>1-pnorm(-2.5)</code>        | 0.9938              |
|                              | Generate 3 observations.                     | <code>rnorm(3)</code>             | 0.993 -2.523 -2.468 |

## Example 3.3: Solution

| Model                        | Question                                     | R code                            | Answer              |
|------------------------------|----------------------------------------------|-----------------------------------|---------------------|
| $X \sim \text{Po}(3)$        | Find $\Pr(X = 6)$ .                          | <code>dpois(6,3)</code>           | 0.0504              |
|                              | Find $\Pr(X < 5)$ .                          | <code>ppois(4,3)</code>           | 0.8153              |
|                              | Generate 3 observations.                     | <code>rpois(3,3)</code>           | 2, 9, 3             |
| $Y \sim \text{Bin}(10, 0.2)$ | Find $\Pr(Y \leq 4)$ .                       | <code>pbinom(4,10,0.2)</code>     | 0.9672              |
|                              | Find $\Pr(Y \geq 4)$ .                       | <code>1-pbinom(3,10,0.2)</code>   | 0.1209              |
|                              | Find $y$ such that $\Pr(Y \leq y) = 0.677$ . | <code>qbinom(0.677,10,0.2)</code> | 2                   |
| $Z \sim N(0, 1)$             | Find $\Pr(Z \leq 0.2)$ .                     | <code>pnorm(0.2)</code>           | 0.5793              |
|                              | Find $\Pr(Z > -2.5)$ .                       | <code>1-pnorm(-2.5)</code>        | 0.9938              |
|                              | Generate 3 observations.                     | <code>rnorm(3)</code>             | 0.993 -2.523 -2.468 |

## Example 3.3: Solution

| Model                        | Question                                     | R code                            | Answer              |
|------------------------------|----------------------------------------------|-----------------------------------|---------------------|
| $X \sim \text{Po}(3)$        | Find $\Pr(X = 6)$ .                          | <code>dpois(6,3)</code>           | 0.0504              |
|                              | Find $\Pr(X < 5)$ .                          | <code>ppois(4,3)</code>           | 0.8153              |
|                              | Generate 3 observations.                     | <code>rpois(3,3)</code>           | 2, 9, 3             |
| $Y \sim \text{Bin}(10, 0.2)$ | Find $\Pr(Y \leq 4)$ .                       | <code>pbinom(4,10,0.2)</code>     | 0.9672              |
|                              | Find $\Pr(Y \geq 4)$ .                       | <code>1-pbinom(3,10,0.2)</code>   | 0.1209              |
|                              | Find $y$ such that $\Pr(Y \leq y) = 0.677$ . | <code>qbinom(0.677,10,0.2)</code> | 2                   |
| $Z \sim N(0, 1)$             | Find $\Pr(Z \leq 0.2)$ .                     | <code>pnorm(0.2)</code>           | 0.5793              |
|                              | Find $\Pr(Z > -2.5)$ .                       | <code>1-pnorm(-2.5)</code>        | 0.9938              |
|                              | Generate 3 observations.                     | <code>rnorm(3)</code>             | 0.993 -2.523 -2.468 |

## Example 3.3: Solution

| Model                        | Question                                     | R code                            | Answer              |
|------------------------------|----------------------------------------------|-----------------------------------|---------------------|
| $X \sim \text{Po}(3)$        | Find $\Pr(X = 6)$ .                          | <code>dpois(6,3)</code>           | 0.0504              |
|                              | Find $\Pr(X < 5)$ .                          | <code>ppois(4,3)</code>           | 0.8153              |
|                              | Generate 3 observations.                     | <code>rpois(3,3)</code>           | 2, 9, 3             |
| $Y \sim \text{Bin}(10, 0.2)$ | Find $\Pr(Y \leq 4)$ .                       | <code>pbinom(4,10,0.2)</code>     | 0.9672              |
|                              | Find $\Pr(Y \geq 4)$ .                       | <code>1-pbinom(3,10,0.2)</code>   | 0.1209              |
|                              | Find $y$ such that $\Pr(Y \leq y) = 0.677$ . | <code>qbinom(0.677,10,0.2)</code> | 2                   |
| $Z \sim N(0, 1)$             | Find $\Pr(Z \leq 0.2)$ .                     | <code>pnorm(0.2)</code>           | 0.5793              |
|                              | Find $\Pr(Z > -2.5)$ .                       | <code>1-pnorm(-2.5)</code>        | 0.9938              |
|                              | Generate 3 observations.                     | <code>rnorm(3)</code>             | 0.993 -2.523 -2.468 |

## Example 3.3: Solution

| Model                        | Question                                     | R code                            | Answer              |
|------------------------------|----------------------------------------------|-----------------------------------|---------------------|
| $X \sim \text{Po}(3)$        | Find $\Pr(X = 6)$ .                          | <code>dpois(6,3)</code>           | 0.0504              |
|                              | Find $\Pr(X < 5)$ .                          | <code>ppois(4,3)</code>           | 0.8153              |
|                              | Generate 3 observations.                     | <code>rpois(3,3)</code>           | 2, 9, 3             |
| $Y \sim \text{Bin}(10, 0.2)$ | Find $\Pr(Y \leq 4)$ .                       | <code>pbinom(4,10,0.2)</code>     | 0.9672              |
|                              | Find $\Pr(Y \geq 4)$ .                       | <code>1-pbinom(3,10,0.2)</code>   | 0.1209              |
|                              | Find $y$ such that $\Pr(Y \leq y) = 0.677$ . | <code>qbinom(0.677,10,0.2)</code> | 2                   |
| $Z \sim N(0, 1)$             | Find $\Pr(Z \leq 0.2)$ .                     | <code>pnorm(0.2)</code>           | 0.5793              |
|                              | Find $\Pr(Z > -2.5)$ .                       | <code>1-pnorm(-2.5)</code>        | 0.9938              |
|                              | Generate 3 observations.                     | <code>rnorm(3)</code>             | 0.993 -2.523 -2.468 |

## Example 3.3: Solution

| Model                        | Question                                     | R code                            | Answer              |
|------------------------------|----------------------------------------------|-----------------------------------|---------------------|
| $X \sim \text{Po}(3)$        | Find $\Pr(X = 6)$ .                          | <code>dpois(6,3)</code>           | 0.0504              |
|                              | Find $\Pr(X < 5)$ .                          | <code>ppois(4,3)</code>           | 0.8153              |
|                              | Generate 3 observations.                     | <code>rpois(3,3)</code>           | 2, 9, 3             |
| $Y \sim \text{Bin}(10, 0.2)$ | Find $\Pr(Y \leq 4)$ .                       | <code>pbinom(4,10,0.2)</code>     | 0.9672              |
|                              | Find $\Pr(Y \geq 4)$ .                       | <code>1-pbinom(3,10,0.2)</code>   | 0.1209              |
|                              | Find $y$ such that $\Pr(Y \leq y) = 0.677$ . | <code>qbinom(0.677,10,0.2)</code> | 2                   |
| $Z \sim N(0, 1)$             | Find $\Pr(Z \leq 0.2)$ .                     | <code>pnorm(0.2)</code>           | 0.5793              |
|                              | Find $\Pr(Z > -2.5)$ .                       | <code>1-pnorm(-2.5)</code>        | 0.9938              |
|                              | Generate 3 observations.                     | <code>rnorm(3)</code>             | 0.993 -2.523 -2.468 |

## Example 3.3: Solution

| Model                        | Question                                     | R code                            | Answer              |
|------------------------------|----------------------------------------------|-----------------------------------|---------------------|
| $X \sim \text{Po}(3)$        | Find $\Pr(X = 6)$ .                          | <code>dpois(6,3)</code>           | 0.0504              |
|                              | Find $\Pr(X < 5)$ .                          | <code>ppois(4,3)</code>           | 0.8153              |
|                              | Generate 3 observations.                     | <code>rpois(3,3)</code>           | 2, 9, 3             |
| $Y \sim \text{Bin}(10, 0.2)$ | Find $\Pr(Y \leq 4)$ .                       | <code>pbinom(4,10,0.2)</code>     | 0.9672              |
|                              | Find $\Pr(Y \geq 4)$ .                       | <code>1-pbinom(3,10,0.2)</code>   | 0.1209              |
|                              | Find $y$ such that $\Pr(Y \leq y) = 0.677$ . | <code>qbinom(0.677,10,0.2)</code> | 2                   |
| $Z \sim N(0, 1)$             | Find $\Pr(Z \leq 0.2)$ .                     | <code>pnorm(0.2)</code>           | 0.5793              |
|                              | Find $\Pr(Z > -2.5)$ .                       | <code>1-pnorm(-2.5)</code>        | 0.9938              |
|                              | Generate 3 observations.                     | <code>rnorm(3)</code>             | 0.993 -2.523 -2.468 |

## Example 3.3: Solution

| Model                        | Question                                     | R code                            | Answer              |
|------------------------------|----------------------------------------------|-----------------------------------|---------------------|
| $X \sim \text{Po}(3)$        | Find $\Pr(X = 6)$ .                          | <code>dpois(6,3)</code>           | 0.0504              |
|                              | Find $\Pr(X < 5)$ .                          | <code>ppois(4,3)</code>           | 0.8153              |
|                              | Generate 3 observations.                     | <code>rpois(3,3)</code>           | 2, 9, 3             |
| $Y \sim \text{Bin}(10, 0.2)$ | Find $\Pr(Y \leq 4)$ .                       | <code>pbinom(4,10,0.2)</code>     | 0.9672              |
|                              | Find $\Pr(Y \geq 4)$ .                       | <code>1-pbinom(3,10,0.2)</code>   | 0.1209              |
|                              | Find $y$ such that $\Pr(Y \leq y) = 0.677$ . | <code>qbinom(0.677,10,0.2)</code> | 2                   |
| $Z \sim N(0, 1)$             | Find $\Pr(Z \leq 0.2)$ .                     | <code>pnorm(0.2)</code>           | 0.5793              |
|                              | Find $\Pr(Z > -2.5)$ .                       | <code>1-pnorm(-2.5)</code>        | 0.9938              |
|                              | Generate 3 observations.                     | <code>rnorm(3)</code>             | 0.993 -2.523 -2.468 |

## 3.3 A more advanced model

- Number of arrivals per day at an IT help-desk is modelled using a Poisson distribution
- Mean of the Poisson distribution might vary from day-to-day
- Suppose the number of arrivals  $X \sim Po(\Lambda)$
- $\Lambda$  is itself a random variable, with  $\Lambda \sim Exp(c)$  for a constant  $c = 0.05$

What can we say about the distribution of  $X$ ?

- 1 What are the expectation and variance of  $X$ ?
- 2 What is  $\Pr(X > 30)$ ?

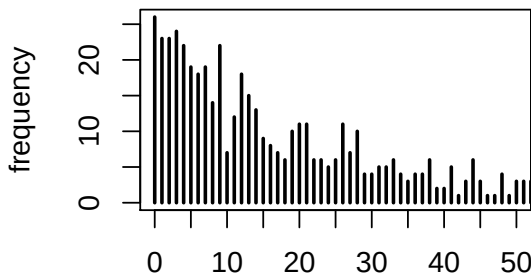
## 3.3 Arrival model – R code

```
1 arrivals = function(n, c = 0.05) {  
2   x = vector(mode = 'numeric', length = n)  
3   for (i in 1:n) {  
4     lambda = rexp(1, rate = c)  
5     x[i] = rpois(1, lambda)  
6   }  
7   return(x)  
8 }
```

## 3.3 Arrival model – R code

Bar plot:

```
1 x = arrivals(500)
2 plot(table(x), xlim = c(0, 50), xaxt = 'n', xlab = 'x'
3     , ylab = 'frequency')
4 axis(1, at = seq(0, 50, 5))
```



## 3.3 Using simulated samples

- The **sample mean** is an approximation to the distribution mean
- The same applies for the sample variance
- To calculate  $\Pr(X > 30)$  we count the proportion of times this occurs in the sample

We expect the approximation to improve **as we increase the sample size**.

## 3.3 Using simulated samples

```
1 mean(x)
2 [1] 19.618
3 var(x)
4 [1] 382.1764
5 sum(x>30)/500
6 [1] 0.202
```

## 3.4 Mixtures of normal distributions

Suppose

- $\mu_1, \mu_2, \sigma_1 > 0, \sigma_2 > 0$  are fixed constants,
- $w_1, w_2$  are positive constants with  $w_1 + w_2 = 1$ .

Consider the following function:

$$f(x) = w_1 f_1(x) + w_2 f_2(x)$$

where  $f_1$  and  $f_2$  are the density functions for  $Z_1 \sim N(\mu_1, \sigma_1^2)$  and  $Z_2 \sim N(\mu_2, \sigma_2^2)$  respectively.

Check that  $f(x)$  represents a valid probability density function.

## 3.4 Mixtures of normal distributions

First note that  $f(x) \geq 0$  everywhere. Also

$$\begin{aligned}\int_{-\infty}^{\infty} f(x) dx &= w_1 \int_{-\infty}^{\infty} f_1(x) dx + w_2 \int_{-\infty}^{\infty} f_2(x) dx \\ &= w_1 + w_2 = 1.\end{aligned}$$

## 3.4 Mixtures of normal distributions

First note that  $f(x) \geq 0$  everywhere. Also

$$\begin{aligned}\int_{-\infty}^{\infty} f(x) dx &= w_1 \int_{-\infty}^{\infty} f_1(x) dx + w_2 \int_{-\infty}^{\infty} f_2(x) dx \\ &= w_1 + w_2 = 1.\end{aligned}$$

## 3.4 Mixtures of normal distributions

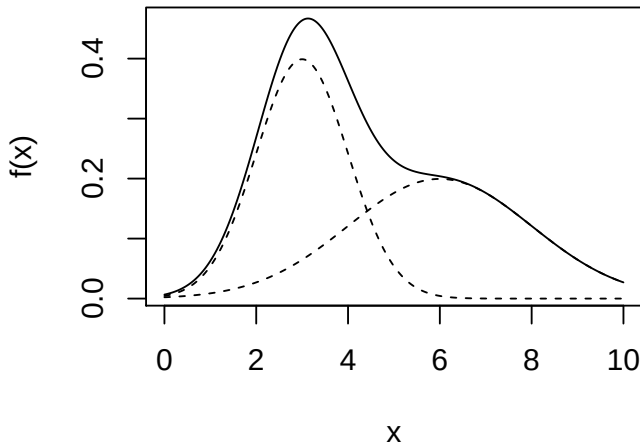
If a random variable  $X$  has PDF corresponding to  $f(x)$  we say it is a **mixture** of normal distributions  $N(\mu_1, \sigma_1^2)$  and  $N(\mu_2, \sigma_2^2)$  with weights  $w_1, w_2$ .

Note that  $X$  is **not** the sum of two normal random variables i.e.

$$X \neq w_1 Z_1 + w_2 Z_2 \quad \text{where } Z_i \sim N(\mu_i, \sigma_i^2), \quad i = 1, 2.$$

## 3.4 Example

For example: suppose  $\mu_1 = 3, \sigma_1 = 1$  and  $\mu_2 = 6, \sigma_2 = 2$  with  $w_1 = w_2 = 1/2$ .



## 3.4 Sampling from a mixture distribution

- 1 Sample a random variable  $J \in \{1, 2\}$  such that  $\Pr(J = 1) = w_1$  and  $\Pr(J = 2) = w_2$ .
- 2 The random variable  $J$  tells you which **component** of the mixture to sample  $X$  from.
- 3 If  $J = 1$  then sample  $X$  from  $N(\mu_1, \sigma_1^2)$ , but if  $J = 2$  then sample  $X$  from  $N(\mu_2, \sigma_2^2)$ .

## 3.4 Sampling from a mixture distribution

```
1  normal.mixture = function(n,mu1,sig1,w1,mu2,sig2,w2) {  
2    p = c(w1, w2)  
3    x = vector(mode = 'numeric', length = n)  
4    for (i in 1:n) {  
5      j = sample(c(1, 2), 1, prob = p)  
6      if (j==1) {  
7        x[i] = rnorm(1, mu1, sig1)  
8      }  
9      else {  
10       x[i] = rnorm(1, mu2, sig2)  
11     }  
12   }  
13   return(x)  
14 }
```

## Example 3.4

### Example

Investigations into infants suffering from *severe idiopathic respiratory distress syndrome* have shown that whether the infant survives may be related to their weight at birth.

Suppose that you are interested in developing a distribution for the birth weight  $X$  of such infants.

You have a normal  $N(2.3, 0.52^2)$  distribution for the birth weight (in kg) of infants who survive and a normal  $N(1.7, 0.66^2)$  distribution for infants who die.

## Example 3.4

### Example

- (a) If you believe that the proportion of infants that survive is 0.6, write down a suitable Normal mixture distribution for  $X$ .
- (b) Produce some R code to sample from the normal mixture distribution you propose in part (a).
- (c) Use your code in part (b) to simulate a sample of size 1000 from your normal mixture distribution, and produce a histogram of your sample.

## Example 3.4: Solution

Using our notation from earlier, define

$$J = \begin{cases} 1 & \text{If the infant survives} \\ 2 & \text{If the infant dies} \end{cases}$$

Then

$$\Pr(J = 1) = 0.6 \quad (= w_1) \quad \text{and}$$

$$\Pr(J = 2) = 0.4 \quad (= w_2).$$

## Example 3.4: Solution

Using our notation from earlier, define

$$J = \begin{cases} 1 & \text{If the infant survives} \\ 2 & \text{If the infant dies} \end{cases}$$

Then

$$\begin{aligned} \Pr(J = 1) &= 0.6 \quad (= w_1) \quad \text{and} \\ \Pr(J = 2) &= 0.4 \quad (= w_2). \end{aligned}$$

## Example 3.4: Solution

We also know that:

- If  $J = 1$ , we need to sample from  $N(2.3, 0.52^2)$
- If  $J = 2$ , we need to sample from  $N(1.7, 0.66^2)$

So we have

$$0.6 \cdot N(2.3, 0.52^2) + 0.4 \cdot N(1.7, 0.66^2).$$

## Example 3.4: Solution

We also know that:

- If  $J = 1$ , we need to sample from  $N(2.3, 0.52^2)$
- If  $J = 2$ , we need to sample from  $N(1.7, 0.66^2)$

So we have

$$0.6 \cdot N(2.3, 0.52^2) + 0.4 \cdot N(1.7, 0.66^2).$$

## Example 3.4: Solution

Therefore we can use our `normal.mixture` function from earlier, with  $\mu_1 = 2.3$ ,  $\sigma_1 = 0.52$ ,  $\mu_2 = 1.7$  and  $\sigma_2 = 0.66$ , giving:

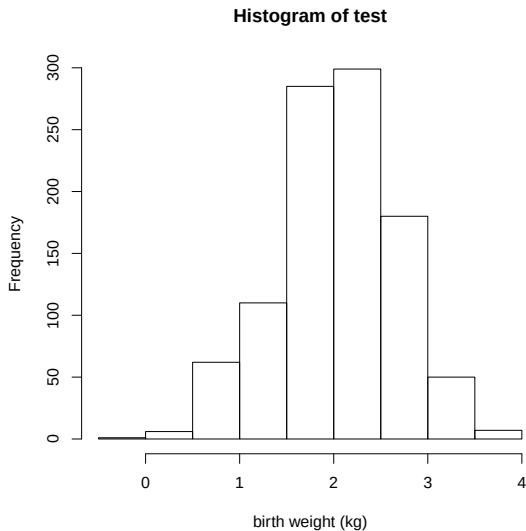
```
1 test = normal.mixture(1000,2.3,0.52,0.6, 1.7,0.66,0.4)
2 hist(test)
```

## Example 3.4: Solution

Therefore we can use our `normal.mixture` function from earlier, with  $\mu_1 = 2.3$ ,  $\sigma_1 = 0.52$ ,  $\mu_2 = 1.7$  and  $\sigma_2 = 0.66$ , giving:

```
1 test = normal.mixture(1000,2.3,0.52,0.6, 1.7,0.66,0.4)
2 hist(test)
```

## Example 3.4: Solution



- That's us done for today, thanks for coming along!
- Non-Synchronous activities to be completed this week, in the following order:
  1. **Test 1** goes live tomorrow at 16:00 – we've now covered everything that might be on the test!
  2. Before next Wednesday's synchronous online lecture, you will need to watch the short video **Home Study 4** – this will be posted in the week 4 materials in Canvas next Monday, along with the accompanying course notes for Chapter 4

- That's us done for today, thanks for coming along!
- Non-Synchronous activities to be completed this week, in the following order:
  1. **Test 1** goes live tomorrow at 16:00 – we've now covered everything that might be on the test!
  2. Before next Wednesday's synchronous online lecture, you will need to watch the short video **Home Study 4** – this will be posted in the week 4 materials in Canvas next Monday, along with the accompanying course notes for Chapter 4