

# MAS1343: Computational Probability and Statistics (with R)

**Revision Notes** 

Dr. Lee Fawcett, School of Mathematics & Statistics

## Overview

As you will have noticed, this course has a large coursework component.

- End of semester examination worth 60% of your overall mark. In order to pass this module you must get at least 35% in the exam.
- Assessment of course-work worth 40%:
  - A single CBA;
  - 6 practicals (one of which consisted of pencil-and-paper exercises)

#### End of term exam

The majority of people finish the end of semester exam with plenty of time left. If you happen to be one of these people, then remember good exam technique:

- Double check all your working
- Is your writing legible? It better had be...
- Check all graphs for labels!

#### Chapter 2: R

#### **Example questions**

1. What is the value of  $\times$ ?

> x = 5> x = 2 \* x + 1> x

2. What is the value of  $\times$ ?

> x = 5; y= 6 > x = x + 2\*y > x + 5 > x

3. What is the value of  $\times$ ?

> x = seq(2, 6)

4. What is the value of blob?

> blob = seq(2, -2, by=-2)

5. What is the value of z?

> x = c(1,10,100,1000); > y = c(1, 5, 10, 100) > z = x/y

6. What is the value of y?

> x = seq(1, 5)> y = length(x + 1)

7. What is the value of z?

> x = TRUE; y = FALSE> z = x & !y

8. What is the value of z?

> x = TRUE; y = FALSE; z = 5 > x = (y / !y) & x > z = x

9. What is the value of z?

> x = seq(-2, 10) > z = x[(x >= 7) | x <0]

10. Suppose movies is a data frame. What type of variable is m1?

> m1 = movies[movies\$ Budget > 10000,]

11. What type of **R** data structure are x, y, z in the following code?

> x
[1] -2 -1 0 1 2 3 4 5 6 7 8 9 10
> y
[1] FALSE
> z
[1] -2 -1 7 8 9 10

- 12. What is the difference between == and ! = ?
- 13. Consider the following vectors:
  - > x = 1:10
    > y = x > 5
    > z = x[x>5]
    - a) What type is x? For example, character, boolean, ..
    - b) What is the length of x?
    - c) What type is y?
    - d) What is the the length of y?
    - e) What type is z?
    - f) What is the the length of z?
- 14. Suppose that y is a data frame with 3 columns and 100 rows. The column names are c1, c2 and c3. All values in y are doubles. Write down the **R** code to:
  - a) Select the first two columns.
  - b) Select the first two rows.
  - c) Calculate the median values for each row
  - d) Calculate the mean values for each column
  - e) Select the values column c3 where the values in c1 are greater than 0.

## **Chapter 3: Summary statistics**

#### Example questions

Data set 1	-3.76	1.1	-5.01	9.57	1.98	-4.92	2.92	
	4.43	3.45	-1.83	9.07	2.34	-3.73	-13.29	6.75
Data set 2	3	3	4	7	2	7	7	5
	5	1	2	2	5	3		
Data set 3	1.45	3.27	1.18	13.4	3.78	1.2	4.43	

For the following practice data sets, calculate the mean, median, IQR, range, and  $s^2$ :

Table 1: Practice data sets.

## **Chapter 4: Graphs**

#### **Example questions**

- 1. For the data sets in Table 1, construct box-plots.
- 2. The hair colour of students in MAS1343 was collected and tabulated. What graphic would you recommend to display this type of data?
- 3. The age of students doing a mathematics degree was collected. Someone recommends that the data is displayed using a bar-chart. Briefly explain why this would be inappropriate.
- 4. Standard rules for determining histogram bin-width are given in table 3. Summary statistics for the internet movie database are given for movie budget, movie length and movie rating. Complete table 2.

				Number of bins		
Variable	Range	S	IQR	Sturges	Scott's	Free-Dia
Budget	$2 \times 10^8$	23039711	$8 \times 10^6$			
Length	249.0	17.3	17.0			
Rating	8.1	1.5	2.0			

Table 2: Bin sizes for the movie data. In all cases n = 5000.

Rule	Formula
Sturges'	$k_{ST} = \lceil \log_2 n + 1 \rceil$
Scott's	$h_{SC} = 3.49 \times s \times n^{-1/3}$
Freedman-Diaconis	$h_{FR} = 2 \times IQR(x) \times n^{-1/3}$

Table 3: Standard bin width rules in R.

## Chapter 5: R programming

#### **Example questions**

1. What is the value of y?

```
> fun1 = function(x) {
+            return(2*x)
+            }
> y = fun1(2)
```

2. What is the value of y?

```
> fun1 = function(x) {
+  z = 3
+  return(x*z)
+ }
> y = fun1(2)
```

3. What is the value of y?

```
> fun1 = function(x) {
+  z = 3
+  return(x*z)
+ }
> k = 5
> y = fun1(k)
```

4. What is the value of *y*?

```
> fun1 = function(x=3) {
+            return(2*x)
+ }
> y = fun1(2)
```

5. What is the value of y?

```
> fun1 = function(x=3) {
+            return(2*x)
+            }
> (y = fun1(2))
[1] 4
```

6. What is the value of y?

```
> fun1 = function(x=3) {
+ return(2*x)
+ }
> y = fun1()
```

7. What is the value of y?

```
> fun1 = function(x=3) {
+     x = 2
+     return(2*x)
+ }
> y = fun1()
```

8. What is the value of y?

```
> x = 1
> fun1 = function(x) {
+ x = 2
+ return(x)
+ }
> y= fun1(3)
```

9. What is the value of y?

```
> x = 1
> fun1 = function(x=3) {
+ x = 2
+ return(x)
+ }
> y= fun1()
```

10. What is the value of y?

> x = 1
> fun1 = function(x=3) {
+ return(2\*x)
+ }
> x = fun1(x)
> y = fun1(x)

11. What is the value of y?

```
> x = 5
> if(x==6) {
+ y = FALSE
+ } else {
+ y = TRUE
+ }
```

12. What is the value of y?

```
> x = 5
> if(x > 6){
+ y = 0
+ }else if(x >= 5) {
+ y = 1
+ } else {
```

+ y = 2 + }

13. What is the value of z?

```
> x = 5
> y = "male"
> if(x > 6 & y == "Female"){
+ z = 0
+ }else if(x < 5 & y == "Male") {
+ z = 1
+ } else {
+ z = 2
+ }
```

14. What is the value of  $\times$ ?

```
> total = 0
> for(blob in 1:4) {
+  total = total + 1
+ }
> x = total
```

15. What is the value of  $\times$ ?

```
> total = 0
> for(blob in 1:10) {
+    if(blob > 9) {
+       total = total + 1
+    }
+ }
> x = total
```

16. In the following piece of **R** code, what is the value of y?

```
> x = 0
> for(i in -3:6) {
+ if(i <= 1) {
+ x = x + i
+ }
+ }
> y = x
```

17. In the following piece of **R** code, what is the value of y?

```
> x = 1
> for(i in 4:6) {
+ x = x*i
+ }
> y = x
```

## **Chapter 6: Random number generation**

#### **Example questions**

- 1. Find the values of
  - a) 8946 mod 4
  - b) 1067 mod 11
  - c) 124 mod 3
  - d) 999 mod 8
- 2. Consider the following congruential generator:

$$r_i = (ar_{i+1} + b) \mod m \text{ for } i = 1, 2, \dots, m,$$
 (1)

where the seed is 0, the multiplier is 4, the additive constant is 17 and the modulo is 27.

- a) Generate the first 5 terms of the sequence produced by this generator.
- b) Is the maximum possible period for this choice of modulo achieved?
- c) What is the period for this generator?
- d) Using the values of  $r_1, \ldots, r_5$ , generate values from a U(0, 1) distribution.
- e) Comment on the suitability of this particular congruential generator for generating random numbers.
- 3. See question 2 from practical 5.

## Chapter 7: Discrete random numbers

#### **Example questions**

1. Using the following four random numbers, simulate tossing a fair coin.

```
0.907, 0.792, 0.050, 0.125,
```

2. Suppose the outcome when a weighted six-sided dice is rolled is a discrete random variable *X* having the following probability mass function:

x	1	2	3	4	5	6
$\Pr[X = x]$	0.10	0.15	0.15	0.15	0.15	0.30

Suppose that the following is a list of random numbers u simulated from a U(0, 1) distribution:

0.907, 0.792, 0.050, 0.125, 0.772, 0.343, 0.722, 0.572, 0.938, 0.860, 0.419, 0.215.

For each *u*, simulate an observation on the random variable *X*. If somebody had told you this was a fair dice, would these outcomes have made you suspicious?

3. Let the discrete random variable *X* have a Geom(0.5) distribution. The following five numbers are pseudo-random numbers from a U(0, 1) distribution.

For each pseudo-random number *u*, simulate an observation on the random variable *X*.

4. Using the following set of eight random numbers from a U(0,1) distribution, simulate eight observations on a Poisson(3) random variable *X*.

0.417, 0.486, 0.162, 0.108, 0.274, 0.882, 0.364, 0.119

5. See also practical 5, question 3.

## **Chapter 8: Monte-Carlo methods**

#### **Example questions**

1. Suppose we wished to calculate

$$\int_{-2}^{2} x + 2dx$$

using Monte-Carlo integration. Determine if the following rectangular regions are suitable:

- a) Region 1:  $0 \le x \le 1$  and  $0 \le y \le 4$
- b) Region 2:  $-5 \le x \le 5$  and  $0 \le y \le 4$
- c) Region 3:  $-2 \le x \le 2$  and  $0 \le y \le 6$
- d) Region 4:  $-2 \le x \le 2$  and  $0 \le y \le 4$

Which rejection region would you choose?

## **Chapter 9: Kernel Density methods**

#### **Example questions**

- 1. State the definition of a kernel.
- 2. Sketch each of the following functions and comment on whether they are valid kernels.
  - a) Kernel 1

$$\begin{cases} \frac{3}{2}(1-t^2) & -1 < t < 1 \\ 0 & \text{otherwise.} \end{cases}$$

 $\begin{cases} \frac{1}{4} & -2 < t < 2\\ 0 & \text{otherwise.} \end{cases}$ 

- b) Kernel 2:
- c) Kernel 3

$$\begin{cases} \frac{1}{2} & -2 < t < -1 \\ \frac{1}{2} & 1 < t < 2 \\ 0 & \text{otherwise.} \end{cases}$$

d) Kernel 4:

 $\begin{cases} \frac{3}{2}t^2 & -1 < t < 1 \\ 0 & \text{otherwise.} \end{cases}$ 

- 3. *K* is a non-negative real-valued integrable. For *K* to be considered a suitable kernel, what two requirements should it fulfil?
- 4. Using the Uniform Kernel

$$K(t) = \begin{cases} \frac{1}{\sqrt{3}} & -\sqrt{3} < t < \sqrt{3} \\ 0 & \text{otherwise.} \end{cases}$$

Sketch a density plot for the following 4 points