NEWCASTLE UNIVERSITY

SCHOOL OF MATHEMATICS & STATISTICS

SEMESTER 2 Paper 1

MAS1343

Computational Probability and Statistics

Time allowed: 1 hour 30 minutes

Candidates should attempt all questions. Marks for each question are indicated. However you are advised that marks indicate the relative weight of individual questions, they do not correspond directly to marks on the University scale.

There are EIGHT questions on this paper.

Answers to questions should be entered directly on this question paper in the spaces provided. Rough work should be done on the blank sides of the pages, or in the blank pages at the end of the paper. The rough work will not be marked. This question paper must be handed in, attached inside an anonymised cover sheet, at the end of the examination.

Calculators may be used.

1. The following are the pulse rates (per minute) of 24 women, arranged in ascending order.

56	60	64	64	64	68	68	68	68	72	72	72
72	76	76	76	76	80	80	80	80	84	84	88

a) Find the sample mode for these data.

Answer:

b) Find the sample median for these data.

Answer:

c) Find the first and third quartiles for these data.

Answer:

[**8 marks**]

2. The sample variance of a set of n observations $x_1, x_2, \ldots, x_n$ is usually defined as:

$$\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

Show that this can be re-expressed as

$$\frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right).$$

Answer:

[**8 marks**]

3. Consider the following congruential generator:

$$r_i = (7r_{i-1} + 13) \bmod 10 \quad \text{for} \quad (i = 1, 2, 3, \ldots);$$

with $r_0 = 5$.

a) What is the value of the multiplier for this generator?

Answer:

b) Is the maximum period achieved for this generator? Give your reasons.

Answer:

c) Not including the seed, generate and write down the first 3 terms of the sequence of integers given by this generator.

Answer:

Question 3 continued on next page

d) How can we use congruential generators to generate values from a $U(0, 1)$ distribution.

Answer:

[**18 marks**]

4. Briefly describe what the following R functions do:

a) `length`

Answer:

b) `rev`

Answer:

c) `table`

Answer:

d) `dim`

Answer:

e) `data.frame`

Answer:

[11 marks]

5. The negative binomial distribution is a discrete probability distribution of the number of successes in a sequence of Bernoulli trials before a specified number r of successes occur. The pmf of this distribution is:

$$\Pr(Z = z) = \binom{z-1}{r-1} p^r (1-p)^{z-r} \quad \text{for } z = r, r+1, \ldots$$

and $0 < p < 1$.

The following three random numbers are simulated from the Uniform $U(0,1)$ distribution:

$$0.1257, \;\; 0.5612, \;\; 0.9243 \;\; .$$

Use these numbers to simulate three random numbers from the negative binomial distribution with $r = 2$ and $p = 0.5$. Show your reasoning.

Answer:

Question 5 continued on next page

[19 marks]

6. The following questions are related to the statistical computing package R:

a) What is the difference between $x == 5$ and $x = 5$ in R?

Answer:

b) In listing 1, what is the value of sq_total?

Listing 1: "A simple for loop"

```
total = 0
for(i in 2:5)
{
    total = total + 1
}
sq_total = total^2
```

Answer:

c) Write a piece of R code that would add one to the variable z, only if y>0.

Answer:

Question 6 continued on next page

d) In listing 2, what data types are the R variables `x`, `y` and `z`, i.e. are they Boolean, vectors, ... ?

Listing 2: "Variable types"

```
> x
[1] "s"
> y
[1] 1 2 3 4
> z
  a      b
1 1 Female
2 2   Male
3 3 Female
```

Answer:

[20 marks]

7. We have a biased coin, where the probability of obtaining heads is 0.75. Using the random numbers below, simulate throwing this biased coin four times. Show your reasoning.

$$U(0,1) \text{ Random numbers} : 0.92,\ 0.78,\ 0.55,\ 0.39$$

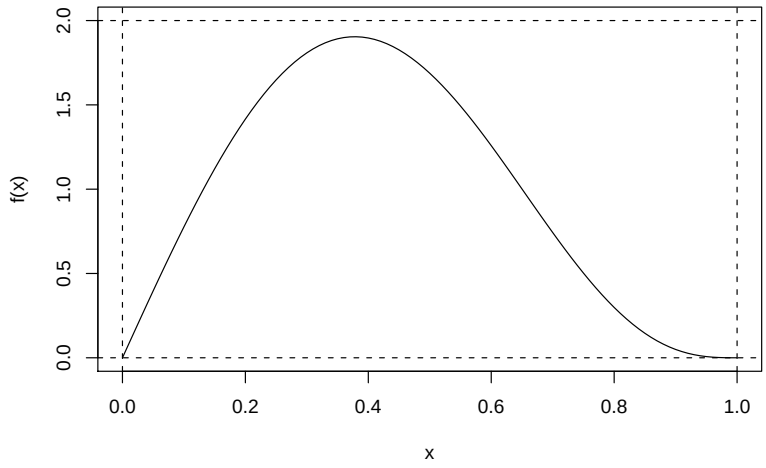
Answer:

[**6 marks**]

8. a) We wish to estimate the area under the pdf

$$f(x) = 8x\left(1 - x^2\right)^3 \quad \text{for} \quad 0 \leq x \leq 1$$

between $x = 0$ and $x = 1$ by using rejection sampling. The following is a plot of this pdf together with a possible rejection region.



Use the following twenty random numbers to estimate this area.

```
> round(runif(10, 0, 1),2)
 [1] 0.32 0.34 0.63 0.11 0.08 0.90 0.07 0.91 0.67 0.68

> round(runif(10, 0, 2),2)
 [1] 0.56 1.77 0.35 0.78 0.88 1.42 0.85 0.10 1.09 1.22
```

Answer:

Question 8 continued on next page

b) How close is your estimate using rejection sample to the correct value for this area?

Answer:

[**10 marks**]

THE END