MAS1343: Computational Probability and Statistics (with R)

Outline solutions to revision questions

2. R-based questions

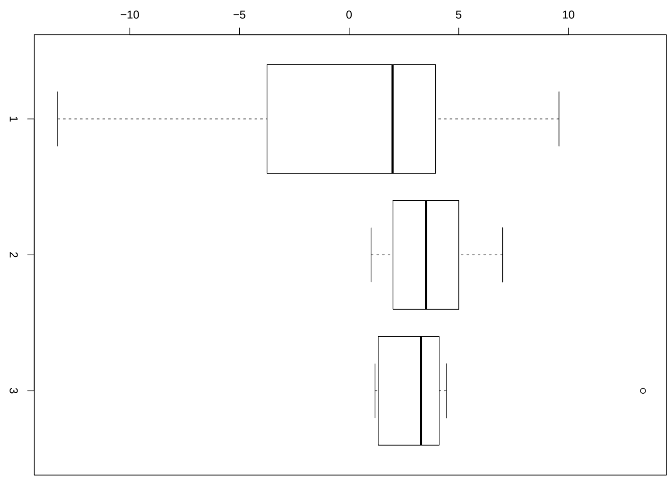
1. 11
2. 17
3. 2 3 4 5 6
4. 2 0 -2
5. 1 2 10 10
6. 5
7. `TRUE`
8. `TRUE`
9. -2 -1 7 8 9 10
10. dataframe
11. `x`: vector of doubles; `y`: logical; `z`: vector of doubles
12. `==` tests for equality; `!=` tests for inequality
13. (a) vector of doubles; (b) 10; (c) vector of logicals; (d) 10; (e) vector of doubles; (f) 5
14. (a) `y[,1:2]`; (b) `y[1:2,]`; (c) `apply(y,1,median)`; (d) `apply(y,2,mean)`; (e) `y$c3[y$c1>0]`

3. Summary statistics

	Mean	Median	IQR	Range	Variance
Dataset 1	0.605	1.98	8.19	22.86	37.280
Dataset 2	4	3.5	3.50	6	4.154
Dataset 3	4.101	3.27	3.23	12.22	18.558

4. Graphs

1. Note: these boxplots have been produced in R – the quartiles may be slightly different to those you calculate by hand!



2. Categorical data – so bar chart or pie chart would do here.
3. This would be inappropriate because although we usually give our age as an integer value, age is actually continuous. Therefore, a histogram would be more appropriate.
4. See Table 4.2 in the lecture notes.

5. R programming

1. 4
2. 6
3. 15
4. 4
5. 4
6. 6
7. 4
8. 2
9. 2
10. 4
11. `TRUE`
12. 1
13. 2
14. 4
15. 1
16. –5
17. 16

6. Random number generation

1. (a) 2; (b) 0; (c) 1; (d) 7
2. (a) 17, 4, 6, 14, 19; (b) (i) b and m have no common factors other than 1, (ii) the only prime that divides m is 3, and $m - 1 = 3$ is a multiple of 3, and (iii) $m - 1$ is not a multiple of 4 – thus, the maximum period *is* achieved; (c) Period is 27; (d) Divide by m to give (0.6296, 0.1481, 0.2222, 0.5185, 0.7037); (e) Suitability: maximum period *is* acheived - however, in practice, we would rather use a *much* larger value for m to avoid early repeatability.
3. See solutions to practical 5 on BB.

7. Discrete random numbers

1. We have:

Outcome	Head	Tail
Prob.	0.5	0.5
Cum. Prob.	0.5	1.0

   This gives: (Tail, Tail, Head, Head).

2. We get:

x	1	2	3	4	5	6
$\Pr(X \leq x)$	0.10	0.25	0.40	0.55	0.70	1.00

This gives: (6, 6, 1, 2, 6, 3, 6, 5, 6, 6, 4, 2). Probably would have been suspicious as there are far more sixes than we'd expect!

3. We use

$$X = 1 + \left[\frac{\ln(1-U)}{\ln(1-p)}\right].$$

This gives:(2,1,4,1,1).

4. If $X \sim Po(3)$, then we have:

x	0	1	2	3	4	5	6	7	8	9	10
$\Pr(X = x)$	0.05	0.149	0.224	0.224	0.168	0.101	0.050	0.022	0.008	0.003	0.001
$\Pr(X \leq x)$	0.05	0.199	0.423	0.647	0.815	0.916	0.966	0.988	0.996	0.999	1.000

This gives: (2, 3, 1, 1, 2, 5, 2, 1).

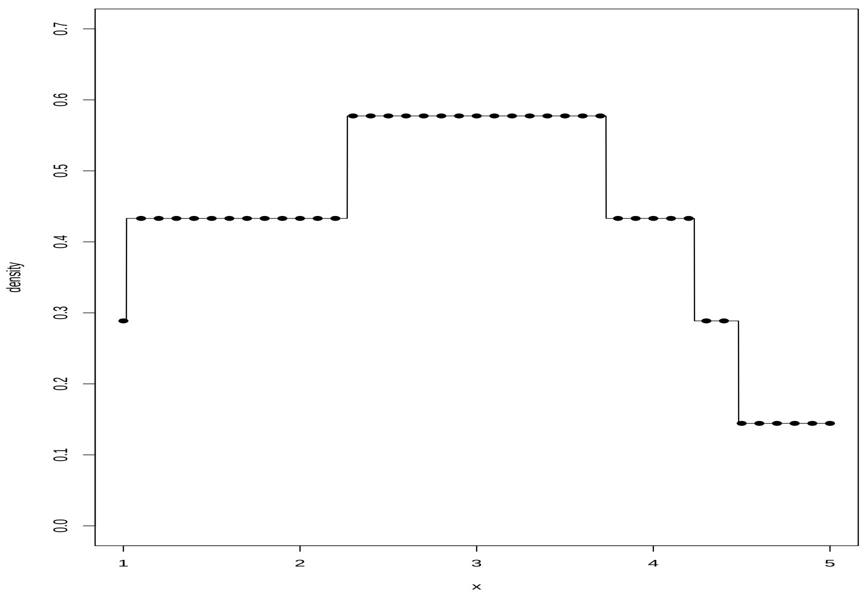
5. See solutions on BlackBoard.

8. Monte Carlo methods

1. Cannot use regions (1) or (2), as the x-axis needs to cover the range $(-2, 2)$. Both regions (3) and (4) will work, as the maximum value for $f(x) = 2 + x$ in the range $-2 \leq x \leq 2$ is $f(x) = 4$, and both regions (3) and (4) will cover this.

9. Kernel density methods

1. See lecture notes page 63.
2. Kernel 1 is *not* valid, as integrating the function between -1 and 1 does not give an area of 1. All of the other kernels are valid, as they have an area of 1 between their limits.
3. See lecture notes page 63 – (9.1) and (9.2)
4. Plot should look something like:



Plotting points can be calculated by consdering a range of x-values to plot over – say $1 \longrightarrow 5$. Then, for example, we would have for $x_i = 1.0$:

$$
\begin{aligned}
\hat{f}(1.0) &= \frac{1}{4}\left[K(1.0 - 2.00) + K(1.0 - 2.50) + K(1.0 - 2.75) + K(1.0 - 4.00)\right] \\
&= \frac{1}{4}\left[K(-1) + K(-1.5) + K(-1.75) + K(-3)\right] \\
&= \frac{1}{4}\left[\frac{1}{\sqrt{3}} + \frac{1}{\sqrt{3}} + 0 + 0\right] \\
&= 0.2886751
\end{aligned}
$$

Similarly, for other points, we get:

x_i	$\hat{f}(x_i)$
1.0	0.2886751
1.1	0.4330127
1.2	0.4330127
1.3	0.4330127
1.4	0.4330127
1.5	0.4330127
$\vdots$	$\vdots$
4.9	0.1443376
5.0	0.1443376