

MAS1343 - 2014/15 Exam Solutions

Question 1

(a) $\bar{x} = 30.75, s = 46.63$ [1,1]

Quantiles: 13.50 (will accept 14.50), 21.50, 28.75 (will accept 28.25), IQR = 15.25 (will accept 13.75) [2,1,2,1]

(b) Median - as the mean will be distorted by the extremely large value of 226 [1]

(c) Boxplot [6 marks]:

– IQR = $37.5 - 21.75 = 15.75$.

– Upper whisker: $37.5 + 1.5 \times 15.75 = 61.12$. So upper whisker goes to 42, with 70 being an outlier.

– Lower whisker: 6. No outliers

Marks: 1 mark for the “box”, 1 mark for axis labels, 2 marks for the outlier, 1 mark for the lower whisker, 1 mark for upper.

Lose 2 marks if the graph is “scrappy”, i.e. they didn’t use a ruler.

Question 2

x	0	1	2	...
$\Pr[X = x]$	0.1296	0.3456	0.3456	...
$\Pr[X \leq x]$	0.1296	0.4752	0.8208	...

Table 0.1: $\text{Bin}(4, 0.4)$ CDF

From the table, the random numbers are 0, 1, 2. [2 marks per number]

Question 3

(a) Let $X \sim \text{Geom}(p)$, so that X takes the random values $1, 2, 3, \dots$ with probabilities $p, (1-p)p, (1-p)^2p, \dots$. Hence,

$$\Pr[X = k] = p(1-p)^{k-1}.$$

Hence,

$$\Pr[X = 1] + \Pr[X = 2] + \dots + \Pr[X = k-1] \leq U < \Pr[X = 1] + \dots + \Pr[X = k],$$

$$\sum_{j=1}^{k-1} (1-p)^{j-1} p \leq U < \sum_{j=1}^k (1-p)^{j-1} p. \quad (0.1)$$

This is GP, so

$$1 - (1-p)^{k-1} \leq U < 1 - (1-p)^k,$$

i.e.

$$k-1 \leq \frac{\log(1-U)}{\log(1-p)} < k,$$

with the inequality sign reversing since $\log(1-p) < 0$. Thus, we observe $X = k$ if

1. $X > \frac{\log(1-U)}{\log(1-p)}$
2. $X - 1 \leq \frac{\log(1-U)}{\log(1-p)}$, i.e. $X \leq 1 + \frac{\log(1+U)}{\log(1-p)}$.

Both (1) and (2) are satisfied by $X = 1 + \left\lfloor \frac{\log(1-U)}{\log(1-p)} \right\rfloor$, where $\lfloor \cdot \rfloor$ denotes “the integer part”. **Bookwork**

[7 marks]

(b) 2 [1 mark]

(c) `rgeom(1,0.8)` [1 mark]

Question 4

(a) (i) 234, 108 [2 marks]

(ii) This would increase the period of the generator [1 mark]

(b) Anything sensible - [2 marks per property]. For example

- A sequence of random numbers has a high probability of containing no identical consecutive elements.
- A sequence of numbers which is indistinguishable from ‘true random’ numbers.
- It should be impossible to calculate, or guess, from any given sub-sequence, any previous or future values in the sequence.
- It should be impossible, for all practical purposes, for an attacker to calculate, or guess the values used in the random number algorithm.

Question 5

Six values fall under the curve. So estimate is

$$\frac{6}{8} \times \text{Box Area} = \frac{36}{8} = 4.5$$

[10 marks: 3 for using box area in calculation, 3 for using ‘8’ in denominator, 4 for getting points correct]

Question 6

(a) [1 mark for each]

(i)

$$\int_{-\infty}^{\infty} K(t)dt = 1$$

(ii) $K(-t) = K(t), \forall t$

(b) Choose three from *Epanechnikov*, *Uniform*, *Gaussian* and *Triangular*. [2 marks for each]
Make sure it’s apparent from the sketches that the area is 1, and make sure the sketches cover the range $(-1, 1)$.

Question 7

2 marks per question

- (a) `length(dd$a)` or `length(dd[,1])` [1 mark]
- (b) `sample(dd$b,1)` or `sample(dd[,2],1)` [2 marks]
- (c) `dd[dd$a < dd$b | dd$a !=5,]` [3 marks] - 1 mark for inner expression
- (d) Calculates the standard deviation of `a` conditional on the value of `c` [2 marks]
- (e) `c` is a string. [2 marks]

Question 8

- (a) 8 [3 marks]
- (b) FALSE [5 marks - give partial credit for understanding]
- (c) 7 and 7 [7 marks - give partial credit for understanding]

Question 9

- (a) It calculates the median [3 marks]
- (b) Does not account for an even sample size [3 marks]
- (c) Three of: [2 marks for each]
 - Indent the code
 - Use sensible function name, i.e. change `g` to `MyMedian`
 - `n` shouldn't be a global variable
 - Use the inbuilt `median(x)` function

Question 10

Hard question

- (a) Variance [6 marks (4 if they say standard deviation)]
- (b) Possible reasons are: [2 marks per reason]
 - One pass in calculating the mean and variance
 - Reduces numerical instability
 - Calculating the variance and mean as data arrives.