Probability and Statistics 1 2007–08

Progress Test

16 January 2008, 2.30 – 4.00 pm

Answer all 6 questions.
The total number of marks available is 50.

1. A lecturer marks an exam on a continuous scale from 0 to 1. The students' marks may be assumed to be independent random variables with density function of the form $f_X(x) = cx^2(1-x)$, $(0 < x < 1)$.

(i) Sketch the function f_X. **[1 mark]**

(ii) Judging by your sketch, are there likely to be more students achieving marks between 0.8 and 1 than between 0 and 0.2? Explain. **[1 mark]**

(iii) Calculate the value of the constant c. **[2 marks]**

(iv) Find the mean and variance of the distribution. **[3 marks]**

(v) Calculate $\mathbb{P}(0 < X < 0.2)$ and $\mathbb{P}(0.8 < X < 1)$. **[2 marks]**

2. During the daytime Harry receives text messages at an average rate of one every 20 minutes.

(i) Would a Poisson process be a reasonable model for the arrival of text messages? Give a reason for your answer: no marks will be awarded for a simple 'Yes' or 'No'. **[2 marks]**

(ii) Assuming a Poisson process model is appropriate, calculate the probability that Harry receives at least 3 texts during a 50-minute lecture. **[2 marks]**

(iii) Three-fifths of the texts Harry receives are from his friends, a quarter from his family. (Assume the two groups are mutually exclusive.) Given that Harry gets 4 texts, what is the probability that 3 are from friends and 1 is from a member of the family? **[3 marks]**

3. The user of a malfunctioning computer thinks there is a 60% chance that the problem is caused by faulty software, a 25% chance of a hardware problem, or a 15% chance that the operating system is at fault.
A test program designed to detect software conflicts has a 95% probability of giving a positive result if the software is causing the problem, but also a 10% chance of a positive result if the cause is hardware or operating system.

(i) Draw a tree diagram to illustrate the various possible outcomes if the test is applied. **[2 marks]**

(ii) Calculate
(a) the probability that the first test gives a positive result;
(b) the conditional probability that the computer has a software conflict given that the result of the test is positive. **[4 marks]**

(iii) If the test gives a positive result, the user applies a diagnostic routine to investigate the operating system: this is certain to give a positive result if the malfunction is caused by the operating system, but also has a 20% chance of a positive result if the software or hardware is at fault. Given that the test program gives a positive result and the diagnostic routine a negative result, find the probability that the computer is suffering from (a) an operating system fault, (b) a hardware fault. **[4 marks]**

4. A study suggests that 30% of people have a talent in some academic field (event A), 40% in one or more sports (event S) and 50% have a creative talent (event C). Academic talent and creative talent are independent of one another, according to the study, but an individual who has a talent for sport has only a one in five chance of being academically talented, a one in five chance of having a creative talent and only a one in 20 chance of having both.

(i) Express in set-theoretic notation the event that an individual has a talent for sport but is not academically or creatively talented. **[1 mark]**

(ii) Express in words the event $(A \cap S)^c \cup C$. **[1 mark]**

(iii) Write down in set-theoretic notation the complement of the event $(A \cap S)^c \cup C$. **[1 mark]**

(iv) Calculate the probability of the event $(A \cap S)^c \cup C$. **[3 marks]**

(v) What is the probability that an individual has neither academic nor sporting nor creative talent? **[2 marks]**

5. A school teacher has developed a new set of revision materials to help children learn mathematics. She has given the new materials to half her class and given the previous set of materials to the other half. The children's scores on the end-of-term exam are as follows:

Materials	**Scores**	**Sample Mean**
New	61 63 68 72 73 76 80 80 81 81 81 82 87 88 96	77.93
Old	59 61 64 67 67 68 70 71 72 74 74 77 77 78 78	70.47

(i) Investigate whether either of the data sets contains any outliers. **[3 marks]**

(ii) Use a box-and-whisker plot to illustrate the performances of the two classes in such a way as to aid comparison. **[2 marks]**

(iii) Comment on the location, spread and skewness of the two data sets. **[3 marks]**

6. A golfer visits a practice area and hits 24 balls towards the hole from a distance of 50 metres. The distances of the balls from the hole, measured in metres and arranged in ascending order, are as follows:

0.4	0.6	0.9	1.3	1.4	1.6	1.6	1.9	2.0	2.1	2.4	2.6
2.7	2.9	3.3	3.6	3.9	4.7	5.1	5.3	6.6	6.9	7.8	9.4

(i) Use a histogram to display the data. Explain your choice of interval widths. **[5 marks]**

(ii) Identify the modal interval. **[1 mark]**

(iii) Use your grouped data to estimate the sample mean. Comment on the difference between your answer and the sample mean calculated from the raw data, which is 3.375. **[2 marks]**

Formulae

$$\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) + \mathbb{P}(A \cap B \cap C).$$

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B \mid A)\mathbb{P}(A)}{\mathbb{P}(B)}$$

Distribution	Notation	Mean	Variance	$p(x)$	Domain
Binomial	$\mathrm{Bin}(k, \theta)$	$k\theta$	$k\theta(1-\theta)$	$\binom{k}{x} \theta^x (1-\theta)^{k-x}$	$x = 0, 1, \ldots, k$
Poisson	$\mathrm{Pois}(\lambda)$	λ	λ	$e^{-\lambda} \frac{\lambda^x}{x!}$	$x = 0, 1, \ldots$
Geometric	$\mathrm{Geom}(\theta)$	θ^{-1}	$\theta^{-2}(1-\theta)$	$\theta (1-\theta)^{x-1}$	$x = 1, 2, \ldots$

Probability and Statistics 1 2007–08

Solutions to Progress Test

1. (i) Sketch. $\boxed{1}$

(ii) 0.8 to 1 is likelier, as the density is very flat at the bottom end. $\boxed{1}$

(iii) $1 = \int_0^1 cx^2(1-x)\,dx = c\left[\frac{x^3}{3} - \frac{x^4}{4}\right]_0^1 = \frac{c}{12}$, so that $c = 12$. $\boxed{2}$

(iv) $\mathbb{E}(X) = 12\int_0^1 x^3(1-x)\,dx = 0.6$. $\boxed{1}$
$\mathbb{E}(X^2) = 12\int_0^1 x^4(1-x)\,dx = 0.4$. $\boxed{1}$
Hence $\mathrm{Var}\,(X) = 0.4 - 0.36 = 0.04$. $\boxed{1}$

(v) $F(x) = \int_0^x f(y)\,dy = 4x^3 - 3x^4$. We require $F(0.2) - F(0) = 0.032 - 0.0048 = 0.0272$ and $F(1) - F(0.8) = 1 - [2.048 - 1.2288] = 0.1808$. $\boxed{2}$

2. (i) Any reason will do, as long as it is sensible. $\boxed{2}$

(ii) Let X be the number of texts received during the lecture, so that $X \sim \mathrm{Pois}(2.5)$. We are looking for $\mathbb{P}(X \geq 3) = 1 - p_X(0) - p_X(1) - p_X(2) = 1 - e^{-2.5}(1 + 2.5 + 3.125) = 0.4562$. $\boxed{2}$

(iii) The probability that the first three texts are from friends and the 4th is from a family member is $0.6^3 \times 0.25 = 0.054$. But there are 4 possible orders in which the 4 texts can be received, so we multiply by 4 to get 0.216. $\boxed{3}$

3. (i) Tree diagram. $\boxed{2}$

(ii) (a) $\mathbb{P}(\text{test positive}) = .6 \times .95 + .4 \times .1 = .61$. $\boxed{2}$
(b) $0.6 \times 0.95/0.61 = 0.9344$. $\boxed{2}$

(iii) $\mathbb{P}(\text{first test pos} \cap \text{second test neg}) = .6 \times .95 \times .8 + .25 \times .1 \times .8 = .476$. $\boxed{1.5}$
Then we have (a) 0, clearly. $\boxed{1}$ (b) $0.02/0.476 = \frac{5}{119} = 0.0420$. $\boxed{1.5}$

4. (i) $S \cap A^c \cap C^c$ or equivalent. $\boxed{1}$

(ii) Either the individual has a creative talent, or else he does not have both artistic and sporting talents. $\boxed{1}$

(iii) The complement is $A \cap S \cap C^c$. $\boxed{1}$

(iv) $\mathbb{P}(A \mid S) = 0.2$ indicates that $\mathbb{P}(A \cap S) = 0.2\mathbb{P}(S) = 0.08$. In addition, $\mathbb{P}(A \cap C \mid S) = 0.05$, showing that $\mathbb{P}(A \cap C \cap S) = 0.05\mathbb{P}(S) = 0.02$. Hence $\mathbb{P}(A \cap S \cap C^c) = \mathbb{P}(A \cap S) - \mathbb{P}(A \cap S \cap C) = 0.06$, and therefore $\mathbb{P}((A \cap S)^c \cup C) = 0.94$. $\boxed{3}$

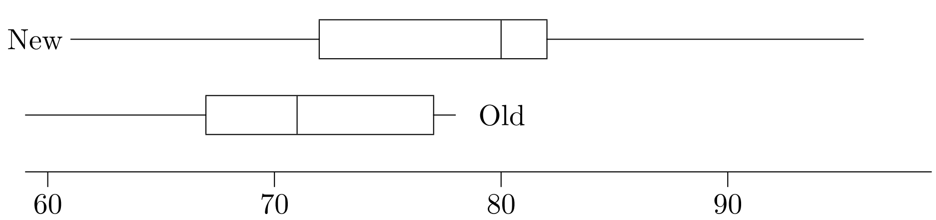
(v) We can use the formula provided:

$$\begin{aligned}\mathbb{P}(A \cup S \cup C) &= \mathbb{P}(A) + \mathbb{P}(S) + \mathbb{P}(C) - \mathbb{P}(A \cap S) - \mathbb{P}(A \cap C) - \mathbb{P}(S \cap C) + \mathbb{P}(A \cap S \cap C) \\ &= 0.3 + 0.4 + 0.5 - 0.08 - 0.3 \times 0.5 - 0.08 + 0.02 = 0.91\end{aligned}$$

so that the required probability is 0.09. $\boxed{2}$

5. (i) Quartiles of New data set are 72, 80 and 82; of Old are 67, 71 and 77. No outliers. $\boxed{3}$

(ii)

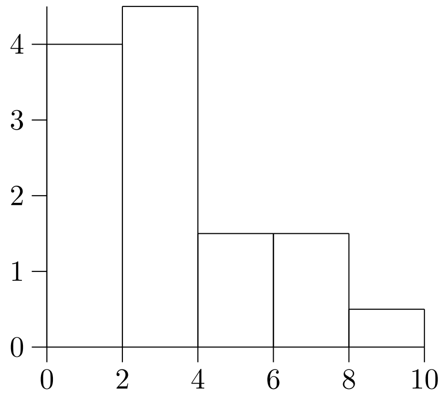


$\boxed{2}$

(iii) Location: the new materials seem to do noticeably better than the old ones. $\boxed{1}$
Spread: the IQRs are the same, though the range of results achieved with the new materials is wider. $\boxed{1}$
Skewness: comparing UQ–Median with Median–LQ suggests that the New data set is left skewed, the Old slightly right skewed. Comparing mean with median suggests that both are right skewed, the New rather more so than the Old. Nothing very conclusive in either case. $\boxed{1}$

6. (i) We would look for interval widths which give about 5 intervals, so an interval width of 2 is reasonable. This means that some intervals have low frequency, but that's life. $\boxed{1}$

Interval	frequency	width	frequency density
0–1.95	8	2	4
1.95–3.95	9	2	4.5
3.95–5.95	3	2	1.5
5.95–7.95	3	2	1.5
7.95–9.95	1	2	0.5



$\boxed{4}$

(ii) The modal interval is 1.95–3.95. $\boxed{1}$

(iii) Estimate is $(8 \times 0.95 + 9 \times 2.95 + 3 \times 4.95 + 3 \times 6.95 + 1 \times 8.95)/24 = 3.283$. $\boxed{1}$
This is marginally below the actual sample mean, indicating a slight tendency for individual observations to fall in the upper half of the interval to which they have been assigned. $\boxed{1}$