Probability and Statistics 1 2009–10

Progress Test

11 January 2010

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

1. A random variable X has density function of the form

$$f(x) = \begin{cases} cx^2(1-x) & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases} .$$

(i) Calculate the value of the constant c. **[1 mark]**

(ii) Evaluate the probability that X takes a value greater than $\frac{2}{3}$. **[2 marks]**

(iii) Calculate the expectation and variance of X. **[3 marks]**

2. A bus stop is used by the number 15 bus and the number 47 bus. Number 15 buses arrive according to a Poisson process with rate parameter $\lambda_{15} = 6$ per hour, number 47 buses according to an independent Poisson process with rate parameter $\lambda_{47} = 4$ per hour.

(i) Calculate the probability that, over a 15-minute period, at least two number 47 buses arrive at the bus stop. **[2 marks]**

(ii) Calculate the probability that exactly one bus arrives at the bus stop over the course of 12 minutes. **[2 marks]**

(iii) Given that exactly one bus arrives at the bus stop in a 12-minute period, what is the probability, p, that it is a number 15 bus? **[1 mark]**

(iv) Suppose that each bus is either a number 15 bus with probability p or a number 47 bus otherwise. Calculate the probability that, of the first four buses to arrive, no more than one of them is a number 15. **[2 marks]**

(v) Show that the probability that no buses of any kind stop at the bus stop between time 0 and time t is e^{-10t}. **[2 marks]**

(vi) Let T be the time when the first bus stops at the bus stop. Explain why $T > t$ is the event that no buses stop at the bus stop between time 0 and time t, and hence find the distribution function and density function of T. **[3 marks]**

3. If there is no engineering work on the railway line, I go to work by train, and there is a 10% chance that I am delayed by half an hour or more, but the delay is never as much as an hour. When there *is* engineering work, I toss a coin to decide whether to go by bus or by train. On the train there is a one in four chance that I experience delays of an hour or more on my journey to work and another one in four chance that I am delayed by between half an hour and an hour. If I take the bus instead, I am certain to experience a delay of at least half an hour, but there is only a 15% chance that I will be delayed by an hour or more. The probability that there are engineering works taking place on a given day is 0.2.

(i) Draw a tree diagram to illustrate the situation, marking in the probabilities. **[2 marks]**

(ii) If there is engineering work, what is the probability that I am delayed by more than an hour? **[1 mark]**

(iii) If I am delayed by between half an hour and an hour, what is the probability that I took the bus? **[3 marks]**

4. A die is used to generate a number in the range 1 to 6. Events A, B and C are defined by:

$A =$"the number showing on the die is even"
$B =$"the number showing on the die is divisible by 3"
$C =$"the number showing on the die is not 6"

(i) List the outcomes which are contained in the event $A^c \cup C^c$. **[1 mark]**

(ii) Use set-theoretic notation to write down, in terms of A, B and C, an event which contains only the outcomes 1 and 5. **[2 marks]**

(iii) There are two dice available for this experiment. One of them is fair, producing each number with probability $\frac{1}{6}$, but the other is biased, producing a six with probability $\frac{1}{2}$ or any other number with probability $\frac{1}{10}$ each. Consider the following four statements. In each case, determine whether the statement is true for both dice, whether it is false for both dice, or whether it is true for one and false for the other. Give reasons for your answers: no marks will be awarded for answers without reasons.

a) $\mathbb{P}(B \cup C) = \mathbb{P}(B) + \mathbb{P}(C)$; **[1 mark]**

b) A and B are independent events; **[2 marks]**

c) $\mathbb{P}(A \mid C) > \mathbb{P}(C \mid A)$; **[3 marks]**

d) $\mathbb{P}(B) < \mathbb{P}(A^c \cup C^c)$. **[1 mark]**

5. A Statistics exam is taken by 10 students on a Philosophy degree and 8 students on a Zoology degree. The marks they achieve are as follows:

Philosophy students	25	32	38	44	44	48	50	58	62	92
Zoology students	40	45	53	55	58	62	70	78		

(i) Calculate the median and quartiles of each data set. **[3 marks]**

(ii) Identify any outliers which may be present. **[2 marks]**

(iii) Use a box-and-whisker diagram to illustrate the two data sets. **[2 marks]**

6. The percentage return on an investment fund in 15 successive years was as follows:

0.9 22.9 8.7 10.0 11.7 21.9 4.5 12.4 12.9 8.2 5.2 8.1 15.3 5.1 3.5

(i) Use a stem-and-leaf diagram to illustrate the data. **[2 marks]**

(ii) Comment on the symmetry or skewness of the sample, using at least two different methods to assess the skewness. **[2 marks]**

(iii) a) Draw a cumulative frequency diagram to illustrate the data set. **[3 marks]**

b) On the same diagram draw the cumulative distribution function of a continuous uniform distribution on the range $(0, 25)$. Does your diagram lead you to believe that the continuous uniform distribution is a good fit to the data? (Give reasons for your answer.) **[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 2009–10

Solutions to Progress Test

1. (i) $1 = \int_0^1 cx^2(1-x)\,dx = \frac{c}{12}$, so that $c = 12$.

(ii) $\mathbb{P}\left(X > \frac{2}{3}\right) = \int_{2/3}^1 12x^2(1-x)\,dx = \left[4x^3 - 3x^4\right]_{2/3}^1 = 1 - \left(4 \times \frac{8}{27} - 3 \times \frac{16}{81}\right) = \frac{11}{27}$.

(iii) $\mathbb{E}(X) = c\int_0^1 x^3(1-x)\,dx = \frac{12}{20} = 0.6$.
$\mathbb{E}(X^2) = c\int_0^1 x^4(1-x)\,dx = \frac{12}{30} = 0.4$.
Hence $\mathrm{Var}\,(X) = 0.4 - 0.6^2 = 0.04$.

2. (i) Let X be the number of 47 buses which arrive in 15 minutes. Then $X \sim P(1)$, and we need $\mathbb{P}(X \geq 2) = 1 - \mathbb{P}(X \leq 1) = 1 - e^{-1}(1+1) = 0.2642$.

(ii) The probability of no 15s and one 47 is $0.8e^{-0.8} \times e^{-1.2}$, whereas the probability of one 15 an no 47s is $e^{-0.8} \times 1.2e^{-1.2}$. So the probability of just one arrival is $2e^{-2} = 0.2707$.

(iii) Given that there is only one bus, the probability that it is a number 15 is 0.6.

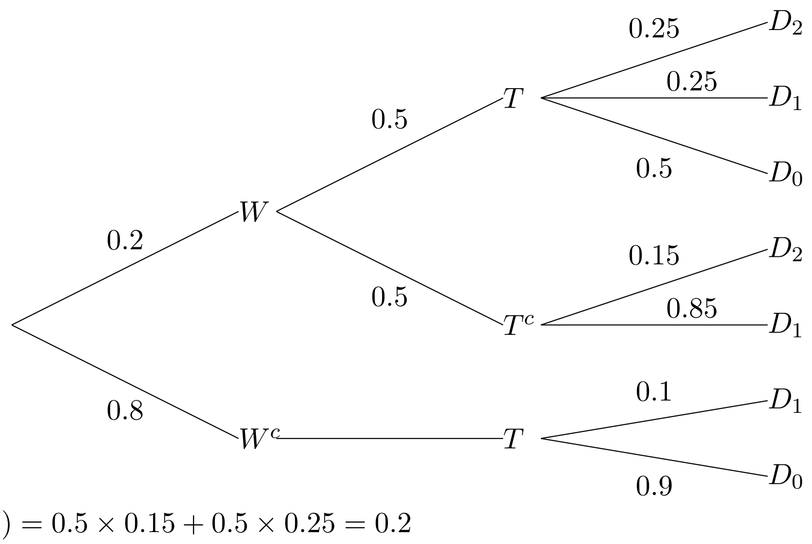
(iv) Let Y be the number of 15s in the first four buses. Then $Y \sim \mathrm{Bin}(4, 0.6)$, so that $\mathbb{P}(Y \leq 1) = 0.4^4 + 4 \times 0.4^3 \times 0.6 = 0.1792$.

(v) Let X_{15} and X_{47} be the number of 15 buses and 47 buses stopping between time 0 and time t. Then $X_{15} \sim P(6t)$ and $X_{47} \sim P(4t)$, so $\mathbb{P}(X_{15} = 0) = e^{-6t}$, $\mathbb{P}(X_{47} = 0) = e^{-4t}$, and by independence the required probability is the product of these two.

(vi) If $T > t$, that means that the first bus arrives after time t, so that no buses have arrived before time t. Conversely, $T < t$ would mean that at least one bus has arrived before time t.
We have $F_T(t) = \mathbb{P}(T \leq t) = 1 - \mathbb{P}(T > t) = 1 - e^{-10t}$.
Therefore $f_T(t) = 10e^{-10t}$ for $t > 0$.

3. Let W be the event that there is engineering work, T the event that I take the train, D_0 the event of no delay (or less than half an hour's delay), D_1 the event of a delay of between half an hour and an hour, D_2 the event of a delay of an hour or more.

(i) The tree diagram should include a key defining any events which are referred to by letters and should look like this:



(ii) $\mathbb{P}(D_2 \mid W) = 0.5 \times 0.15 + 0.5 \times 0.25 = 0.2$

(iii) $\mathbb{P}(T^c \mid D_1) = \frac{\mathbb{P}(T^c \cap D_1)}{\mathbb{P}(D_1)}$. Now $\mathbb{P}(T^c \cap D_1) = 0.2 \times 0.5 \times 0.85 = 0.085$, and $\mathbb{P}(D_1) = 0.085 + 0.2 \times 0.5 \times 0.25 + 0.8 \times 0.1 = 0.085 + 0.025 + 0.08 = 0.19$. So the required conditional probability is $\frac{0.085}{0.19} = \frac{17}{38} = 0.4474$.

4. (i) $A^c \cup C^c = (A \cap C)^c = \{1, 3, 5, 6\}$.

(ii) $(A \cup B)^c$ will do, or $A^c \cap B^c$ or $A^c \cap B^c \cap C$.

(iii) a) False for both, since B and C are not mutually exclusive.

b) For the fair die, $\mathbb{P}(A \cap B) = \frac{1}{6} = \mathbb{P}(A) \times \mathbb{P}(B)$, so they are independent. For the biased die, however, $\mathbb{P}(A \cap B) = \mathbb{P}(6) = \frac{1}{2}$, whilst $\mathbb{P}(A) = 0.7$ and $\mathbb{P}(B) = 0.6$.

c) The LHS is $\mathbb{P}(A \cap C)/\mathbb{P}(C)$, whereas the RHS is $\mathbb{P}(A \cap C)/\mathbb{P}(A)$.
So we are really asking whether $\mathbb{P}(A) > \mathbb{P}(C)$.
For the fair die $\mathbb{P}(A) = \frac{1}{2}$ and $\mathbb{P}(C) = \frac{5}{6}$, so the answer is no. But for the biased die, $\mathbb{P}(A) = 0.7$, $\mathbb{P}(C) = 0.5$, so the answer is yes.

d) B is a subset of $A^c \cap C^c$, so this will always be true.

5. (i) There are 10 Philosophy students, so the quartiles are 38 and 58, with median 46.
There are 8 Zoology students, so the quartiles are 49 and 66, with median 56.5.

(ii) IQR for P is 20; therefore possible outliers for Philosophy are values above $58 + 1.5 \times 20 = 88$ or below $38 - 1.5 \times 20 = 8$. Thus 92 is a possible outlier.
IQR for Z is 17; therefore possible outliers for Zoology are values above $66 + 1.5 \times 17 = 91.5$ or below $49 - 1.5 \times 17 = 23.5$. Thus there are no possible outliers.

(iii) The student should draw a single diagram against a linear horizontal scale, clearly labelling the boxes relating to the two data sets.

6. (i) There are a number of possible ways to do the diagram. Students may want to round the observations up to the nearest integer, or may not bother to do that. The interval width could be 5 or 10. Here are a couple of examples.

```
20 | 19 29              Key:              20 | 2 3                 Key:
10 | 53                 stem - 1          10 | 0 2 2 3 5           stem - 1
10 | 00 17 24 29        leaf - 0.1         0 | 1 4 5 5 5 8 8 9     leaf - 1
00 | 51 52 81 82 87
00 | 09 35 45
```

(ii) Looking at the stem-and-leaf diagram, the tail is to the right, suggesting positive skew.
The quartiles are 5.1 and 12.9, with median 8.7, again suggesting positive skew.
The sample mean is 10.1, larger than the median, so positive skew is suggested.

(iii) a) See separate sheet for the cumulative frequency diagram.

b) The fact that the line is at some points quite far from the line of the cumulative distribution function of the continuous uniform indicates that the uniform distribution is not a good fit to the data.