Probability and Statistics 1 2008–09

Progress Test

January 2009

Answer all 5 questions.
The total number of marks available is 50.

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

$$f(x) = \begin{cases} 0 & \text{if } x < 1 \\ a + \frac{b}{x^4} & \text{if } x > 1 \end{cases}$$

where a and b are constants.

(i) Calculate the values of a and b. **[2 marks]**

(ii) Evaluate the probability that X lies between 1 and 1.5. **[2 marks]**

(iii) Evaluate the expectation and variance of X. **[2 marks]**

2. A small taxi company employs only two drivers, Andrzej and Boris. In dry weather each has a 1 in 1000 chance per journey of having an accident, but in wet weather the probability of an accident is 1 in 500 if Andrzej is driving, 1 in 200 if Boris is driving. When a taxi is ordered, the driver is equally likely to be Andrzej or Boris when the weather is dry, but Andrzej is twice as likely as Boris to be selected in wet weather. The weather is wet 30% of the time.

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

(ii) Calculate the probability that an accident takes place during a journey given that the weather is wet. **[1 mark]**

(iii) Given that a journey involved an accident, what is the probability that Andrzej was driving? **[3 marks]**

(iv) The company finds that accidents occur according to a Poisson process with mean 9 per year.

(a) Calculate the probability that there are no accidents in the first three months of the year.

(b) Write down the distribution of Y, the number of months in a single year during which no accidents take place.

(c) Evaluate the probability that the first accident of 2009 occurs during March.

(d) If there are 4 accidents in the first 6 months of the year, what is the probability that Andrzej was involved in no more than one of them? **[6 marks]**

3. A motorists' organisation suggests that about 15% of the motorists in London have not bought the compulsory motor insurance (event I), about 10% have not paid their car tax (event T), and about 5% are driving a vehicle which is not listed in the official Register of Motor Vehicles (event R). Vehicles which are not listed in the Register cannot be taxed or insured.

(i) Explain in words what it means to say that a motorist belongs to $(I \cap T)^c \cup R$. **[1 mark]**

(ii) Describe, in set theoretic notation, the event that a motorist is driving a vehicle which *is* listed in the Register but is either untaxed or uninsured (or both). **[1 mark]**

(iii) Show, by means of a Venn diagram, the relationship between the events I, T and R. **[2 marks]**

(iv) Let p denote the conditional probability that a motorist is uninsured given that he or she has not paid the car tax.

(a) Find an expression involving p for the probability that a given motorist is driving around legally (event $(I \cup T \cup R)^c$).

(b) Show that at least 80% of motorists are driving legally, assuming the motoring organisation's figures are correct. **[5 marks]**

4. A gambling game is based around a pack of 20 cards, numbered from 1 to 4 in each of five colours (red, blue, green, yellow and purple). In the first stage of the game, a single card is drawn from the pack and placed on the table.

(i) Write down the sample space, Ω, of this experiment. **[1 mark]**

(ii) Give an example of a pair of mutually exclusive events in this sample space, each having non-zero probability. **[1 mark]**

(iii) Give an example of a pair of independent events in this sample space, each with probability strictly between 0 and 1. **[2 marks]**

(iv) A second card is now selected from the pack and placed on the table next to the first. Let X denote the number on the first card, Y the number on the second card. Calculate the probability function of $X + Y$. **[3 marks]**

5. The running times (in minutes) of the 27 films directed by Steven Spielberg between 1964 and 2007 are as follows (films arranged in chronological order):

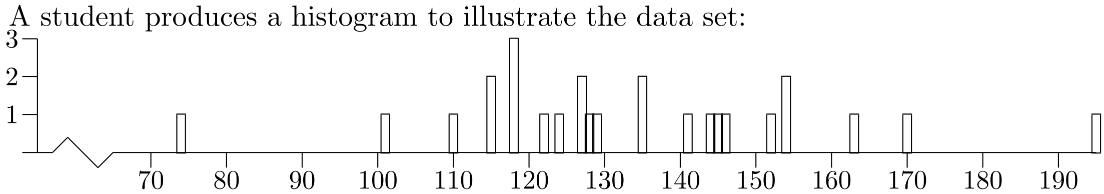
135 26 74 110 124 135 118 115 115 101 118 154 154 127
122 144 127 195 129 152 170 146 145 141 128 118 163

[Source: Wikipedia]

(i) Illustrate this data set by means of a stem-and-leaf diagram. **[2 marks]**

(ii) Identify the median and quartiles and investigate whether the data set contains any outliers. **[4 marks]**

(iii) A student produces a histogram to illustrate the data set:



Comment on the histogram and indicate how it could be improved. **[3 marks]**

(iv) During the same period Steven Spielberg was involved (as author, producer or executive producer) in the production of a further 18 films. The running times (in minutes) of these 18 films may be summarised as follows:

Minimum = 69, L Quartile = 91, Median = 113.5, U Quartile = 129, Maximum = 144.

(a) Draw a box-and-whisker diagram to illustrate the two data sets.
(b) Compare the two data sets in terms of location, spread and skewness. **[7 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	$\mathrm{Bin}(k, \theta)$	$k\theta$	$k\theta(1-\theta)$	$\binom{k}{x}\theta^x (1-\theta)^{k-x}$	$x = 0, 1, \dots, k$
Poisson	$\mathrm{Pois}(\lambda)$	λ	λ	$e^{-\lambda}\,\frac{\lambda^x}{x!}$	$x = 0, 1, \dots$
Geometric	$\mathrm{Geom}(\theta)$	θ^{-1}	$\theta^{-2}(1-\theta)$	$\theta\,(1-\theta)^{x-1}$	$x = 1, 2, \dots$

Probability and Statistics 1 2008–09

Solutions to Progress Test

1. (i) $1 = \int_1^\infty (a + bx^{-4})\,dx$. The first term will be infinite unless $a = 0$. The second term comes to $\frac{1}{3}b$. Hence $a = 0$, $b = 3$.

(ii) $\int_1^{1.5} 3x^{-4}\,dx = [-x^{-3}]_1^{1.5} = 1 - \frac{8}{27} = \frac{19}{27} = 0.7037$.

(iii) We have

$$\mathbb{E}[X] = \int_1^\infty 3x^{-3}\,dx = 1.5, \quad \mathbb{E}[X^2] = \int_1^\infty 3x^{-2}\,dx = 3,$$

so that $\mathrm{Var}\,(X) = 0.75$.

2. (i) Tree diagram.

(ii) $\frac{2}{3} \times \frac{1}{500} + \frac{1}{3} \times \frac{1}{200} = 0.003$.

(iii) Let C denote the event that an accident occurs, A (B) the event that Andrzej (Boris) is the driver, D the event that the weather is dry. Then

$$\begin{aligned}\mathbb{P}(A \cap C) &= \mathbb{P}(D)\,\mathbb{P}(A \mid D)\,\mathbb{P}(C \mid A \cap D) + \mathbb{P}(D^c)\,\mathbb{P}(A \mid D^c)\,\mathbb{P}(C \mid A \cap D^c) \\ &= 0.7 \times \frac{1}{2} \times \frac{1}{1000} + 0.3 \times \frac{2}{3} \times \frac{2}{1000} = \frac{0.75}{1000}\end{aligned}$$

and similarly

$$\mathbb{P}(B \cap C) = 0.7 \times \frac{1}{2} \times \frac{1}{1000} + 0.3 \times \frac{1}{3} \times \frac{5}{1000} = \frac{0.85}{1000}.$$

Therefore $\mathbb{P}(C) = \frac{1.6}{1000}$ and $\mathbb{P}(A \mid C) = \frac{0.75}{1.6} = \frac{15}{32} = 0.4688$.

(iv) (a) $e^{-2.25} = 0.1054$.
(b) Y is Binomial, with parameters $k = 12$, $\theta = e^{-0.75} = 0.4724$.
(c) $0.4724^2(1 - 0.4724) = 0.1177$.
(d) Let Z be the number of accidents involving Andrzej. Given that there were 4 accidents altogether, $Z \sim \mathrm{Bin}\left(4, \frac{15}{32}\right)$, so that $\mathbb{P}(Z \le 1) = \mathbb{P}(Z = 0) + \mathbb{P}(Z = 1) = 0.0780 + 0.2811 = 0.3608$.

3. (i) Either the car is unregistered or it *is* insured or it *is* taxed.

(ii) $R^c \cap (I \cup T)$.

(iii) Venn diagram (R is contained within $I \cap T$).

(iv) (a) Since $R \subset I \cap T$, we have $I \cup T \cup R = I \cup T$.
$\mathbb{P}(I \mid T) = p \Rightarrow \mathbb{P}(I \cap T) = p\mathbb{P}(T) = 0.1p \Rightarrow \mathbb{P}(I \cup T) = 0.25 - 0.1p$.
Hence $\mathbb{P}((I \cup T \cup R)^c) = 1 - \mathbb{P}(I \cup T) = 0.75 + 0.1p$.
(b) $0.1p = \mathbb{P}(I \cap T) \ge \mathbb{P}(R) = 0.05$. Hence $p \ge 0.5$ and therefore the proportion of motorists driving legally is at least 0.8.

4. (i) $\Omega = \{R1, 2, \ldots, R4, B1, \ldots, P4\}$.

(ii) Any two disjoint events will do.

(iii) Plenty to choose from. For example, $A = \{\text{colour is red}\}$, $B = \{\text{number is 1}\}$. The answer should include a proof that the two events are independent.

(iv) The probability that both cards show the number j is $\frac{5}{20} \times \frac{4}{19} = \frac{2}{38}$ for each j. The probability that they show the numbers i and j (in some order) is $2 \times \frac{5}{20} \times \frac{5}{19} = \frac{5}{38}$. for each $i \ne j$.
We therefore have

Total	2	3	4	5	6	7	8
Probability	$\frac{2}{38}$	$\frac{5}{38}$	$\frac{7}{38}$	$\frac{10}{38}$	$\frac{7}{38}$	$\frac{5}{38}$	$\frac{2}{38}$

5. (i) Intervals of width 10 or 20 are OK, though 10 is a bit better.

```
Stem  Leaf
190 | 5
180 |
170 | 0
160 | 3
150 | 2 4 4
140 | 1 4 5 6
130 | 5 5
120 | 2 4 7 7 8 9
110 | 0 5 5 8 8 8
100 | 1
 90 |
 80 |
 70 | 4
 60 |
 50 |
 40 |
 30 |
 20 | 6
```

(ii) Median = 14th value = 128 mins; LQ = 7th value = 118 mins; UQ = 146 mins.
Definite outliers are $>$ 230 mins or $<$ 34 mins (26 is a definite outlier);
Possible outliers are $>$ 188 mins or $<$ 76 mins (195 and 74 are both possible outliers).

(iii) Various possible comments. :
The definite outlier has been omitted. If omitted, some mention should be made of this fact.
The intervals are much too narrow: a better choice would be about 20.
The shape of the data set is not obvious: there are too many intervals.
Broken axes are not really desirable.

(iv) (a) Box-and-whisker diagram.
(b) Location: The films directed by Spielberg (call this data set D) have all 3 quartiles higher than the other set (set P).
Spread: the IQR of set D is 28, whereas that of set P is 38; from this it appears that set P has greater spread. On the other hand the range of set D is much greater than that of set P.
Skewness: For set D, med−LQ=11, UQ−med=17, indicating positive skew.
For set P, med−LQ=22.5, UQ−med=15.5, indicating negative skew.