Probability and Statistics 1 2010–11

Progress Test

07 January 2011

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

1. A continuous random variable X takes values in the range $(-1, 1)$ and satisfies $\mathbb{E}(X) = \mu_1$, $\mathbb{E}(X^2) = \mu_2$. The density function of X is of the form

$$f(x) = \begin{cases} ax^2 + bx + c & \text{for } -1 < x < 1 \\ 0 & \text{for } x < -1 \text{ or } x > 1 \end{cases},$$

where a, b and c are constants.

(i) Write down the distribution function $F(x)$ of X in terms of a, b and c. **[2 marks]**

(ii) Find expressions for a, b and c in terms of μ_1 and μ_2. **[4 marks]**

(iii) If $\mathbb{E}(X) = \frac{1}{2}$ and $\mathrm{Var}\,(X) = \frac{1}{4}$, calculate $\mathbb{P}(X > 0)$. **[3 marks]**

2. All 25 computers in a computer lab have their software updated. As a result each computer, independently, is left with a number of faults which has a Poisson distribution with mean 0.1. 24 students are due to take an exam in the lab.

(i) Calculate the probability that a given computer has at least one fault. **[1 mark]**

(ii) Let X denote the number of computers with at least one fault.

a) Write down the distribution of X: give the name of the distribution and the values of any parameters.

b) Calculate the probability that there are not enough fault-free computers to accommodate all the students. **[3 marks]**

(iii) If a computer has exactly one fault, there is a 50% chance that it will still be able to run Excel and, independently, a 50% chance that it can run Word. The computer exam requires both Word and Excel. Given that exactly 23 of the computers have no faults, what is the probability that at least one of the remaining two computers can run Excel and at least one can run Word? **[4 marks]**

3. All short-listed applicants for a job are invited to attend an interview, a psychometric test or both. Three-fifths of the applicants attend an interview and three-quarters of them attend a test. Jobs are offered to 10% of those who attend only the interview, to 20% of those who attend only the test and to 40% of those who attend both.

(i) What proportion of short-listed applicants attend both an interview and a test? **[1 mark]**

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

(iii) Calculate the proportion of short-listed applicants who are offered a job. **[2 marks]**

(iv) If an applicant is offered a job, what is the probability that he or she attended a psychometric test? **[2 marks]**

4. Two fair standard dice, one green and one yellow, are rolled simultaneously. Events A, B and C are defined as follows:

$$\begin{aligned} A &= \{\text{the green die shows an even number}\} \\ B &= \{\text{the yellow die shows an even number}\} \\ C &= \{\text{the sum of the scores on the two dice is odd}\} \end{aligned}$$

(i) Show that the events A and B are independent, events A and C are independent and events B and C are independent. **[2 marks]**

(ii) a) Explain what is meant by *physical independence*.

b) Which of the pairs of events are physically independent? **[2 marks]**

(iii) a) Calculate $\mathbb{P}(A \cap B \cap C)$.

b) Is it possible to find an event D such that $D \subset A \cup B \cup C$ and $\mathbb{P}(D) = \mathbb{P}(A) \times \mathbb{P}(B) \times \mathbb{P}(C)$? Give a reason for your answer. **[3 marks]**

5. Owners of two adjacent coconut plantations are comparing the annual yields of their trees, which they have recorded in tabular form as shown below:

Grower A		**Grower B**	
No. of coconuts	**No. of trees**	**No. of coconuts**	**Percentage of trees**
40–50	12	21–35	5
51–60	30	36–50	17
61–70	64	51–60	16
71–80	29	61–80	46
81–90	9	81–100	16

(i) Use a cumulative frequency diagram to compare the yields of Grower A's trees with those of Grower B and comment on your findings. **[4 marks]**

(ii) Calculate the mean and median of each data set, correct to the nearest whole number. **[4 marks]**

6. (i) A student notes down the numbers of the buses which stop at the bus stop outside her window over a 15-minute period. Giving a reason for your answers, state

a) which of the following types of display could meaningfully be used to illustrate this data set:

histogram, cumulative frequency diagram, bar chart, box-and-whisker plot, pie chart

b) which of mean, median and mode could be used as a suitable measure of location. **[3 marks]**

(ii) The student also notes down the length of time (in seconds) for which each bus is standing at the bus stop:

28 175 10 245 75 3 50 60 20 42

a) Illustrate the data by means of a stem-and-leaf diagram.

b) Calculate the five-point summary of the data set and determine whether any outliers are present.

c) Construct a box-and-whisker plot of the data set and comment on its symmetry or skewness. **[8 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 2010–11

Solutions to Progress Test

1. (i) Integrating,

$$\begin{aligned} F(x) &= \int_{-1}^{x} (ay^2 + by + c)\,dy \\ &= \frac{a}{3}(x^3+1) + \frac{b}{2}(x^2-1) + c(x+1) \end{aligned}$$

(ii) We have

$$\begin{aligned} 1 &= \int_{-1}^{1} f(x)\,dx = \frac{2}{3}a + 2c \\ \mu_1 &= \int_{-1}^{1} xf(x)\,dx = \frac{2}{3}b \\ \mu_2 &= \int_{-1}^{1} x^2 f(x)\,dx = \frac{2}{5}a + \frac{2}{3}c \end{aligned}$$

Therefore $b = \frac{3}{2}\mu_1$, $a = \frac{15}{8}(3\mu_2 - 1)$ and $c = \frac{3}{8}(3 - 5\mu_2)$.

(iii) $\mu_1 = \frac{1}{2}$, $\mu_2 = \mathrm{Var}\,(X) + [\mathbb{E}(X)]^2 = \frac{1}{2}$. Therefore $b = \frac{3}{4}$, $a = \frac{15}{16}$ and $c = \frac{3}{16}$. Therefore $\mathbb{P}(X > 0) = 1 - F(0) = 1 - \left[\frac{a}{3} - \frac{b}{2} + c\right] = 1 - \frac{5}{16} + \frac{3}{8} - \frac{3}{16} = \frac{7}{8}$.

2. (i) Let Y denote the number of faults on a single computer. Then $Y \sim \mathrm{Pois}(0.1)$, so $\mathbb{P}(Y \geq 1) = 1 - \mathbb{P}(Y = 0) = 1 - e^{-0.1} = 0.0952$.

(ii) a) $X \sim \mathrm{Bin}(25, 0.0952)$.

b) If $X \geq 2$ there will not be enough working computers to accommodate all the students. $\mathbb{P}(X \geq 2) = 1 - \mathbb{P}(X \leq 1) = 1 - [0.0821 + 0.2158] = 0.7021$.

(iii) Let p denote the conditional probability that a computer has exactly one fault given that it has at least one fault.

If both faulty computers have more than one fault, neither Excel nor Word can run at all.

If one has more than one fault and the other does not, the probability that both Excel and Word can run is 0.25.

If both faulty computers have exactly one fault, the probability that one of them can run Word is $1 - (\frac{1}{2})^2 = \frac{3}{4}$. The same is true for Excel. Therefore the probability that both Excel and Word can run is $(\frac{3}{4})^2 = \frac{9}{16}$.

So the probability we are looking for is $\frac{9}{16}p^2 + \frac{1}{4} \times 2p(1-p) = \frac{1}{2}p + \frac{1}{16}p^2$.

Now $p = \frac{\mathbb{P}(Y=1)}{\mathbb{P}(Y>0)} = 0.9508$, so the answer we are looking for is 0.5319.

3. (i) Let I denote the event that the applicant is invited for an interview, T the test. Then $1 = \mathbb{P}(I \cup T) = \mathbb{P}(I) + \mathbb{P}(T) - \mathbb{P}(I \cap T) = 1.35 - \mathbb{P}(I \cap T)$. Therefore $\mathbb{P}(I \cap T) = 0.35$.

(ii) Diagram.

(iii) Let J denote the event that the applicant is offered a job. Then

$$\begin{aligned} \mathbb{P}(J) &= \mathbb{P}(J \,|\, I \cap T)\,\mathbb{P}(I \cap T) + \mathbb{P}(J \,|\, I \cap T^c)\,\mathbb{P}(I \cap T^c) + \mathbb{P}(J \,|\, I^c \cap T)\,\mathbb{P}(I^c \cap T) \\ &= 0.4 \times 0.35 + 0.1 \times 0.25 + 0.2 \times 0.4 = 0.14 + 0.025 + 0.08 = 0.245. \end{aligned}$$

(iv) $\mathbb{P}(T \,|\, J) = \frac{\mathbb{P}(J \,|\, I \cap T)\,\mathbb{P}(I \cap T) + \mathbb{P}(J \,|\, I^c \cap T)\,\mathbb{P}(I^c \cap T)}{\mathbb{P}(J)} = \frac{0.14 + 0.08}{0.245} = \frac{44}{49} = 0.8980.$

4. (i) Each of A, B and C has probability $\frac{1}{2}$.

There are 36 possible outcomes. Of these, the number in which both dice show an even number is $3 \times 3 = 9$. $\mathbb{P}(A \cap C)$ is the probability that the green die shows an even number and the yellow die shows an odd number, which is also $\frac{1}{4}$, and similarly for $B \cap C$.

(ii) Two events are physically independent if they are determined by separate experiments.

A and B are physically independent.

(iii) a) Clearly $\mathbb{P}(A \cap B \cap C) = 0$.

b) No. $\mathbb{P}(A) \times \mathbb{P}(B) \times \mathbb{P}(C) = \frac{1}{8}$. Since there are 36 elementary events in Ω, each with equal probability, there is no event with probability $\frac{1}{8}$.

5. (i) Diagram. Comments.

(ii) Grower A has 144 trees. By linear interpolation, Median $= 60.5 + \frac{72-42}{64} \times (70.5 - 60.5) = 65.2$

The sample mean is $(12 \times 45 + 30 \times 55.5 + 64 \times 65.5 + 29 \times 75.5 + 9 \times 85.5)/144 = 65.0$.

Grower B has only reported percentages. In this case, Median $= 60.5 + \frac{50-38}{46} \times (80.5 - 60.5) = 65.7$

And the sample mean is $(5 \times 28 + 17 \times 43 + 16 \times 55.5 + 46 \times 70.5 + 16 \times 90.5)/100 = 64.5$.

6. (i) a) We are dealing here with nominal data. That means that histogram, cumulative frequency diagram and box plot are inappropriate. Either a pie chart or a bar chart would be suitable.

b) Only the mode is acceptable.

(ii) a) Stem-and-leaf diagram:

```
200 | 45
100 | 75
100 |
  0 | 50 60 75
  0 | 03 10 20 28 42          Key: 100 | 25 = 125
```

b) Minimum is 3, lower quartile 20, median 46, upper quartile 75, maximum 245.

With an IQR of 55, any value above 240 is a definite outlier, and a value between 157.5 and 240 is a possible outlier. Thus 175 is a possible outlier, 245 a definite outlier.

c) Box plot.

Comments: The stem-and-leaf diagram appears to be right-skewed. The difference between the upper quartile and the median is 29, whilst the difference between the median and the lower quartile is 26, again suggesting right skew. The sample mean is 70.8, greater than the median.