UNIVERSITY OF LONDON [C245 2002]

B.ENG. AND M.ENG. EXAMINATIONS 2002

For Internal Students of the Imperial College of Science, Technology and Medicine

This paper is also taken for the relevant examination for the Associateship of the City and Guilds of London Institute.

COMPUTING C245

STATISTICS

Date Thursday 25th April 2002 2.00 - 3.30 pm

Answer THREE questions

[*Before starting, please make sure that the paper is complete. There should be a total of FOUR questions. Ask the invigilator for a replacement if this copy is faulty.*]

1. (i) A slapdash lecturer has a probability of 0.1 of making a mistake on each of his slides, and these probabilities are independent from slide to slide.

   In a lecture in which he shows 5 slides:

   (a) What is the probability that none will contain an error?

   (b) What is the probability that they will all contain errors?

   (c) What is the probability that the first two will contain errors?

   (d) What is the probability that the first two, and none of the others, will contain errors?

   (e) What is the probability that exactly two will contain errors?

   (f) What is the probability that fewer than two will contain errors?

   (g) What is the probability that at least two will contain errors?

   (ii) Observation shows that there is a probability of 0.8 that a number dialled at random by an automatic telephone dialler will be answered. Furthermore, it is known that 10% of the telephone numbers in the directory used by the dialler have an answering machine attached. Whenever a number with an answering machine attached is dialled, the call is always answered (either by the machine or by a person).

   (a) What is the probability that the dialler will be answered when it calls a phone without an answering machine attached?

   (b) What is the probability that the first time the dialler is answered (by either a person or a machine) is at attempt 4?

   (c) What distribution would be an appropriate model for the number of attempts before the dialler is first answered?
   Give its name and the value of its parameter.

   (d) What is the expected number of attempts before the dialler is answered?

2. (i) For each of the following functions, state the values of c which make the functions legitimate probability density functions:

$$\text{(a)} \quad f(x) = \begin{cases} c, & a \leq x \leq b\,, \\ 0, & \text{otherwise}\,; \end{cases}$$

$$\text{(b)} \quad f(x) = \begin{cases} ce^{-\lambda x}, & x > 0\,, \\ 0, & \text{otherwise}\,; \end{cases}$$

$$\text{(c)} \quad f(x) = c\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right].$$

(ii) What are the mean values of the distributions given in part (i)?

(iii) What are the median values of the distributions given in part (i)?

(iv) A user of the internet normally uses search engine A, but is thinking of switching to search engine B.

(a) It is known that the times taken to locate particular items of information on the internet vary from search to search, and that the distribution of these times is right skewed. Draw a sketch indicating the shape of such a distribution.

(b) The log transforms of the search times are known to follow a normal distribution fairly closely. The mean log(time) for search engine A to locate items is known to be 1.5. The user has collected the information below on search engine B, which shows the log(time) values for 10 randomly chosen searches. Compute the mean and standard deviation of the log(times) in the sample.

2.6	2.2	1.5	1.4	1.4	1.2	1.8	1.1	1.0	2.9

(c) Using appropriate tables from the formula sheet, carry out a test of the hypothesis that the log(time) values using engine B are drawn from a distribution with a mean of 1.5, at the 5% level. In your answer, clearly state which distribution you use for the test statistic, and write down any formulae you use to compute the test statistic.

(d) What recommendation would you make to the user?

3. (i) Write brief descriptions of the following terms and, using discrete distributions, illustrate your explanations with simple examples:

   (a) Joint distribution.

   (b) Marginal distribution.

   (c) Conditional distribution.

   (d) Independence.

   (ii) A survey is conducted to compare the appeal of two different graphical user interfaces, GUI A and GUI B, for data mining. To do this, a random sample of data miners is asked a set of questions. One of the questions asks the users if the interface they used had taken much time to learn, with three possible responses: a short time, a moderate amount of time, or a long time. The results of the survey are given in the table below, with the cells of the table showing the number of users in the sample who took a short, moderate or long time under each system. Carry out an appropriate test at the 5% level, using any tables you need from the formula sheet, to explore whether the distribution of responses suggests that one of the systems takes less time to learn than the other. Indicate any test statistics you use, stating your conclusions clearly.

   Time taken to learn:

	Short	Moderate	Long
GUI A	12	15	20
GUI B	14	16	30

4. (i) Define the terms survivor function and hazard function. What forms are taken by the cumulative probability distribution, the survivor function and the hazard function of an exponential distribution?

(ii) A particular class of electronic component has a lifetime (measured in hours) which is known to be a random variable with probability density function $f(t) = \lambda e^{-\lambda t},\ t > 0.$

(a) If the mean lifetime of a random sample of 10 such components was found to be 20 hours, what would you estimate the variance of the sample lifetimes to be?

(b) Using the parameters for the model in part (a), what proportion of components would you expect to survive beyond 30 hours?

(c) If you know that a component is already 10 hours old, how much longer would you expect it to survive?

(iii) A system has three types of components, C1, C2 and C3, in the series/parallel arrangement below. Each of these types of components is known to have a lifetime which is a random variable following an exponential distribution. The mean lifetimes of components of types C1 and C2 are both 10 months and the mean lifetime of components of type C3 is 20 months. Calculate the probability that the system will fail before 20 months have passed.

```
                 |--------C2--------|
                 |                  |
-------C1--------|                  |----------------
                 |                  |
                 |--------C3--------|
```


END OF PAPER