UNIVERSITY COLLEGE LONDON

University of London

EXAMINATION FOR INTERNAL STUDENTS

For the following qualifications :-

B.Sc. *M.Sci.*

Mathematics M252: Probability and Statistics

COURSE CODE	:	**MATHM252**
UNIT VALUE	:	**0.50**
DATE	:	**02-MAY-02**
TIME	:	**10.00**
TIME ALLOWED	:	**2 hours**

TURN OVER

All questions may be attempted but only marks obtained on the best **four** *solutions will count.*
The use of an electronic calculator **is** *permitted in this examination.*

New Cambridge Statistical Tables are provided.

1. (a) Define each of the following terms, as used in the context of probability theory:
   (i) Experiment.
   (ii) Sample space.
   (iii) Event.
   (iv) Random variable.
   (v) Bernoulli trial.

   (b) Define the conditional probability of an event A given an event B, and state any conditions necessary for the definition to be valid.

   (c) Show that if A, B and C are 3 events such that all probabilities are well-defined,

$$\frac{P(A|C)}{P(B|C)} = \frac{P(A \cap C)}{P(B \cap C)}.$$

   (d) The following table summarises the results of an experiment whereby different doses of a flu vaccine were given to a sample of 1000 individuals. The individuals were monitored over a winter period to see how many of them caught flu:

	Dose level		
Disease status	**Low**	**Medium**	**High**
Flu	24	9	13
No flu	289	100	565

   Use this table to provide a numerical illustration of the result derived in part (c), by defining appropriate events and evaluating the corresponding probabilities.

2. Machine components are packaged in boxes of 20. It is known that a particular box of components contains 3 that are defective. To isolate the defective ones, components are taken from the box one at a time, without replacement, and inspected.

   (a) Find the probability that

      (i) Of the first 5 components to be inspected, exactly 2 are defective.

      (ii) The 7th component to be inspected is the 2nd defective one found.

      (iii) At most 10 components need to be inspected to find all of the defectives.

   (b) Let Y be the number of components inspected, up to and including the last of the defectives. Find the probability mass function of Y and show that, for appropriate values of k, it can be written in the form

$$P(Y = k) = c(k-1)(k-2) \ ,$$

   for some constant c. Deduce that the mode of Y is 20.

3. (a) Let X be an exponentially distributed random variable having density

$$f(x) = \lambda e^{-\lambda x} \qquad (x > 0, \lambda > 0) \ .$$

   Find the mean and standard deviation of X.

   (b) Suppose X is an exponentially distributed random variable as above. Find $P(a < X \leqslant b)$, where a and b are real numbers with $0 < a < b$.

   (c) Now suppose that Y is a discrete random variable obtained by rounding non-integer values of X up to the next highest integer i.e. $Y = k$ if $(k-1) < X \leqslant k$. Find the probability mass function of Y, and deduce that Y has a geometric distribution. Give the parameter of this distribution.

   (d) By considering simple situations in which the exponential and geometric distributions arise, explain briefly why the result in part (c) is 'obvious'.

4. In a certain university, exam papers are set with the intention that students' final marks, averaged over a large number of papers, will be normally distributed with a mean of 55% and a standard deviation of 10%.

   (a) Assuming that the final marks really *do* have the intended distribution, what proportion of students will obtain final marks:

   (i) over 70%?

   (ii) below 35%?

   (iii) in the range [60%, 70%)?

   (b) In a particular cohort of 25 students, no student obtains a final mark in excess of 70%. Would you regard this as evidence that the exam papers were too difficult? Your answer should be supported by an appropriate probability calculation, stating clearly any assumptions that you make.

   (c) Is it reasonable to expect that the students' final marks will be normally distributed?

5. In an experiment to determine the effectiveness of a new medical treatment designed to prolong survival after surgery, operations were carried out on 16 mice. 7 of the mice (the 'test group') were then randomly selected to receive the new treatment, and the other 9 (the 'control group') were given no treatment. The survival times, in days, for the test group were as follows:

$$94,\ 197,\ 16,\ 38,\ 99,\ 141,\ 23\ .$$

   The survival times for the controls were

$$52,\ 104,\ 146,\ 10,\ 50,\ 31,\ 40,\ 27,\ 46\ .$$

   (Data taken from B. Efron and R. Tibshirani: *An Introduction to the Bootstrap*, 1993).

   (a) Assume that the two groups of measurements can be regarded as independent samples from populations with means μ_1 and μ_2, and variances σ_1^2 and σ_2^2, respectively. Calculate estimates of μ_1, μ_2, σ_1^2 and σ_2^2.

   (b) Test, at the 95% level, the hypothesis that the $\sigma_1^2 = \sigma_2^2$. State your conclusions clearly.

   (c) Assuming that $\sigma_1^2 = \sigma_2^2$, calculate a 95% confidence interval for $\mu_1 - \mu_2$.

   (d) What do these analyses tell you about the effectiveness of the new treatment? State any additional assumptions you have made during the analyses, and comment on their validity.

6. (a) n pairs of observations $(x_1, y_1), \ldots, (x_n, y_n)$ have been generated from the linear regression model

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \qquad (i = 1, \ldots, n) \ ,$$

where the $\{\varepsilon_i\}$ are independent random variables having common mean zero and variance σ^2.

Suppose the value of β_1 is known. Find the least-squares estimate of β_0. What is this estimate when $\beta_1 = 0$?

(b) Suppose that $X_1, \ldots, X_n$ are independent identically distributed random variables. It is known that they are drawn from some family of probability distributions indexed by a parameter θ, but the value of θ is unknown. Suppose that, on the basis of the n observations available, there are several possible estimators of θ. Explain clearly the issues you would consider when deciding which of these estimators to use. Take care to define any technical terms that you use in your explanation.