UNIVERSITY COLLEGE LONDON

EXAMINATION FOR INTERNAL STUDENTS

MODULE CODE	:	**MATH7501**
ASSESSMENT PATTERN	:	**MATH7501A**
MODULE NAME	:	**Probability and Statistics**
DATE	:	**30-Apr-08**
TIME	:	**14:30**
TIME ALLOWED	:	**2 Hours 0 Minutes**

2007/08-MATH7501A-001-EXAM-177

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. Let $P(\cdot)$ be a probability function defined on some sample space, and let A, B and C be events.

   (a) Use the axioms of probability to show that $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

   (b) What is meant by saying that A and B are (i) disjoint (ii) independent?

   (c) What is meant by saying that A, B and C are (i) mutually disjoint (ii) mutually independent?

   (d) If $P(A \cup B) = 0.6$ and $P(A) = 0.2$, find $P(B)$ if

      (i) A and B are disjoint;

      (ii) A and B are independent;

      (iii) $P(B|A) = 1$.

   (e) Show that if A, B and C are mutually independent, then A and $B \cup C$ are independent.

2. (a) Suppose that X is a discrete random variable taking non-negative integer values. Let $\Pi(s)$ denote the probability generating function (pgf) of X. Show that $\Pi'(1) = \mathrm{E}[X]$ and that $\Pi''(1) = \mathrm{E}[X(X-1)]$, where $\Pi'(\cdot)$ and $\Pi''(\cdot)$ denote the first and second derivatives of the pgf.

   (b) A field contains two types of plants. Type A is attractive to insects, whereas type B is not. The number of insects to be found on a plant of type A can be modelled as a Poisson random variable with mean μ. No insects will ever be found on a plant of type B. The proportion of type B plants in the field is p. Let X be the number of insects found on a randomly chosen plant.

      (i) Find the probability mass function of X.

      (ii) Show that the probability generating function of X is given by $\Pi(s) = p + (1-p)e^{\mu(s-1)}$. Use this result to find the mean and variance of X.

3. The lifetime of a certain type of washing machine is exponentially distributed, with a mean of μ years. The manufacturer offers a guarantee on every machine sold: if the machine breaks down irreparably within the first year of use (so that its lifetime is less than one year), it will be replaced free of charge.

   (a) What proportion of machines need to be replaced free of charge?

   (b) Any replacement machine that fails within the first year of its own lifetime will also be replaced free of charge. The lifetimes of successive replacements are mutually independent. For every machine sold, how many replacements does the company expect to have to provide?

   (c) The cost of manufacturing and supplying a washing machine is £200, and the machines sell for £500 each. If the mean lifetime is 5 years, how much profit does the manufacturer expect to make per machine sold?

   (d) It is suggested that the manufacturer may be able to increase profitability by making the machines more reliable, so that fewer replacements are required. However, this will increase production costs. Specifically, the cost (in pounds) of manufacturing and supplying a machine with a mean lifetime of μ years is $C(\mu) = 180 + 4\mu$. The manufacturer wishes to keep the selling price fixed at £500 per machine. What value of μ is maximises the expected profit?

   (e) Briefly, explain why the exponential distribution might, as a first approximation, provide a reasonable model for the distribution of washing machine lifetimes in practice. Does this model have any unrealistic features?

4. (a) Daily summer temperatures in a certain part of the USA are normally distributed with mean 28°C and standard deviation 3°C.

      (i) What is the probability that the temperature on a particular day will exceed 32.2°C?

      (ii) What is the probability that the temperature on a particular day will be between 25°C and 32.2°C?

      (iii) In the USA, a heat wave is sometimes defined as a period of three or more consecutive days with temperatures in excess of 32.2°C. Assuming that temperatures on successive days are independent, calculate the probability of experiencing a heat wave during a given three-day period.

      (iv) Another possible definition of a heat wave is any three-day period when the average temperature over the three days exceeds some threshold τ°C. What should this threshold be in order that the probability of a heat wave under this definition is the same as that found in part (iii)?

   (b) Suppose X is a normally distributed random variable with mean μ and variance σ^2. Let $\Phi(\cdot)$ denote the distribution function of the standard normal distribution. Show that $U = \Phi\left((X - \mu)/\sigma\right)$ has a uniform distribution, and give the parameters of this distribution.

5. In an experiment to compare the effectiveness of two drugs, each drug was administered to a different group of sick patients and the time (in days) for each patient to recover was recorded. The recovery times were as follows:

Drug A	79	84	108	114	120	103	122	120		
Drug B	91	103	90	113	108	87	100	80	99	54

(a) Calculate the sample mean and standard deviation for each set of measurements.

(b) Test, at the 5% level, the hypothesis that the two sets of observations come from distributions with the same variance. State your conclusions clearly.

(c) Assuming that the underlying variances are the same in each group, construct a 95% confidence interval for the underlying difference in mean recovery times for the two drugs. Comment on the result.

6. (a) Let X be a continuous random variable with a uniform distribution on the range $(0, \theta)$. Find the mean and variance of X, along with its distribution function $F(x) = P(X \leqslant x)$.

(b) Suppose $X_1, \ldots, X_n$ are independent identically distributed random variables, each with distribution function $F_X(x)$. Let M be the largest of the $\{X_i\}$. Express the event "$M \leqslant x$" as an intersection of n independent events, each involving exactly one of the $\{X_i\}$. Hence deduce that the distribution function of M is $F_M(x) = [F_X(x)]^n$.

(c) Use the results from parts (a) and (b) to write down an expression for the distribution function of the largest among n independent $U(0, \theta)$ random variables. Find the corresponding density function.

(d) A sample $X_1, \ldots, X_n$ of independent observations from $U(0, \theta)$ is available, and it is required to estimate θ. Two estimators are proposed: $\hat{\theta}_1 = 2\overline{X}$, where $\overline{X}$ is the sample mean, and $\hat{\theta}_2 = (n+1)M/n$, where M is the largest of the $\{X_i\}$.

(i) Show that $\hat{\theta}_1$ and $\hat{\theta}_2$ are both unbiased estimators of θ.

(ii) Which, if either, of the estimators is better? Justify your answer using appropriate calculations.