UNIVERSITY OF SURREY©

B. Sc. Undergraduate Programmes in Mathematical Studies
M. Math. Undergraduate Programmes in Mathematical Studies

Level HE1 Examination

Module MS132 PROBABILITY AND STATISTICS

Time allowed – 2 hours Spring Semester 2007

Attempt FOUR questions. If any candidate attempts more than FOUR questions only the best FOUR solutions will be taken into account.
Cambridge Statistical Tables and a formula sheet will be provided.

Question 1

(a) Suppose that in a survey concerning the reading habits of students it is found that:
60% read magazine A
50% read magazine B
50% read magazine C
30% read magazines A and B
20% read magazines B and C
30% read magazines A and C
10% read all three magazines.

(i) What percentage read magazine A or magazine B? [3]

(ii) What percentage read exactly two magazines? [4]

(iii) What percentage do not read any of the magazines? [4]

(b) (i) A representative committee for the Student Union of size 10 has to be selected from 25 female students and 15 male students. Write down an expression for the probability that exactly half of the committee members are female and evaluate this probability. [Give your answer correct to four decimal places] [4]

(ii) Now suppose that a representative committee for the Student Union of size 10 is to be selected from 2500 female students and 1500 male students.

i. Write down an expression for the probability that exactly half the committee members are females and evaluate this probability.
[Give your answer correct to four decimal places] [4]

ii. Write down a binomial random variable that can be used to approximate the probability that exactly r of the 10 committee members are females. Use the this random variable to estimate the probability that exactly half the committee members are female.
[Give your answer correct to four decimal places] [6]

Question 2

(a) In a factory, a certain brand of chocolates is packed into boxes on four different production lines A_1, A_2, A_3 and A_4. Records show that a small percentage of boxes are not packed properly for sale: 1% from A_1, 4% from A_2, 2% from A_3 and 2% from A_4. The percentage of total output that have come from the production lines are 35% from A_1, 20% from A_2, 24% from A_3 and 21% from A_4.

(i) What is the probability, that a box chosen at random from the whole output came from production lines A_1 or A_4? [1]

(ii) What is the probability, that a box chosen at random from the whole output is not properly packed? [4]

(iii) If we find a box, which is packed properly, what is the probability it came from production line A_3? [4]

(b) A random variable X has distribution given by probability function

$$p(x) = p^x(1-p), x = 0, 1, 2, ..,$$

$$0 < p < 1$$

(i) Show that the probability generating function $G_x(s)$ for this distribution is $G_x(s) = \frac{1-p}{1-sp}$. [6]

(ii) Using $G_x(s)$ from part (i) evaluate the expectation $E[X]$. [5]

(iii) Using $G_x(s)$ from part (i) evaluate the variance $Var[X]$. [5]

Question 3

(a) In an experiment to test the effectiveness of two slimming diets, 25 persons followed diet A and 30 persons followed diet B for one month. The losses in weight were recorded in kilograms. Let x_i denote the weight loss of the i^{th} person on diet A. Let y_i denote the weight loss of the i^{th} person on diet B. The results are given:

Sample	Diet A	Diet B
Size	25	30
$\sum x_i, \sum y_i$	27.5	57.1
$\sum x_i^2, \sum y_i^2$	37.1	115.6
Mean	1.1	1.9
Variance	0.274	0.243
S.D.	0.523	0.493

(i) Test the null hypothesis H_0 : variances of observations from diets A,B are the same against alternative H_1 : variances of observations from diets A,B are different. [6]

(ii) Carry out a statistical test to determine if there is any evidance to indicate that diet B is more effective than diet A. [6]

(b) A bag contains a very large collection of black marbles and white marbles. 8192 random samples of 6 marbles are drawn from the bag. The frequencies of the number of black marbles in these samples are tabulated below:

Number of black marbles	1	2	3	4	5	6	Total
Frequency	45	255	1115	2505	2863	1409	8192

Test the hypothesis that the data fit a binomial distribution with the ratio of the numbers of black to white marbles in the bag being 3:1. [13]

Question 4

(a) A continuous random variable X has normal distribution with mean $E[X] = a$ and variance $Var[X] = b^2$. Find the probability that $|X - a| \leq 3b$. [5]

(b) A continuous random variable has probability density function, $f(x)$, given by:

$$f(x) = \begin{cases} \frac{x}{c} & \text{if } 0 \leq x \leq 1; \\ 2 - \frac{x}{c} & \text{if } 1 \leq x \leq 2; \\ 0 & \text{otherwise.} \end{cases}$$

(i) Find constant c. [3]

(ii) Find mean $E[x]$ and variance $Var[X]$. [3]

(iii) Find probability $P(0.25 < X < 0.75)$. [4]

(iv) Now suppose that the above function $f(x)$ is the probability density function for the distribution of pizza deliveries during lunch time, i.e. between noon and 2 pm, where X is the number of hours after 12 pm. All orders taken after 1.25 are classified as 'after-peak'. The manager wants to divide the rest of lunch time into 'pre-peak' period and 'peak' period which would have equal proportions of the orders. Find the interval corresponding to 'peak' time. [10]

Question 5

A fair dice is thrown three times. The result of first throw is scored as $X_1 = 1$ if the dice shows 5 or 6 and $X_1 = 0$ otherwise; X_2 and X_3 are scored likewise for the second and third throws. Let $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_3$.

Show that $P(Y_1 = 0, Y_2 = -1) = \frac{4}{27}$. Calculate the remaining probabilities in the bivariate distribution of the pair (Y_1, Y_2) and display the joint probabilities in an appropriate table. [6]

(a) Find the marginal probability distributions of Y_1 and Y_2. [4]

(b) Calculate the means and variances of Y_1 and Y_2. [5]

(c) Calculate the covariance of Y_1 and Y_2. [4]

(d) Find the conditional distribution of Y_1 given $Y_2 = 0$. [4]

(e) Find the conditional mean of Y_1 given $Y_2 = 0$. [2]

Question 6

(a) (i) $x_1, \ldots, x_n$ is a random sample of size n from an Exponential random variable with mean $\frac{1}{\theta}$ and probability density function:

$$f(x) = \theta \exp(-\theta x), x > 0$$

$$\theta > 1$$

Find $\hat{\theta}$, the maximum likelihood estimator of θ. [8]

(ii)) If the random variable X has the Exponential probability density function given in part (i), show that

$$P(a < X \leq b) = \exp(-a\theta) - \exp(-b\theta)$$

[4]

(b) In an experiment batteries are subjected to a magnetic field and their lifetimes recorded. The distribution of lifetimes Y (in months) is thought to be Exponential. One hundred batteries are put on test and the results are given in the following table:

Lifetimes	0-1	1-2	2-3	3-4	4-5	5-6	6-7	7-8	8+	Total
Frequency	36	30	15	5	5	5	3	1	0	100

Test the goodness-of-fit of the Exponential model, using a suitable estimate for the parameter θ. [13]

INTERNAL EXAMINER: R. Retkute