INSTITUTE OF ACTUARIES OF INDIA

EXAMINATIONS

8th November 2010

Subject CT3 – Probability & Mathematical Statistics

Time allowed: Three Hours (15.00 – 18.00 Hrs)

Total Marks: 100

INSTRUCTIONS TO THE CANDIDATES

1. ***Please read the instructions on the front page of answer booklet and instructions to examinees sent along with hall ticket carefully and follow without exception***
2. ***Mark allocations are shown in brackets.***
3. ***Attempt all questions, beginning your answer to each question on a separate sheet. However, answers to objective type questions could be written on the same sheet.***
4. ***In addition to this paper you will be provided with graph paper, if required.***

AT THE END OF THE EXAMINATION

Please return your answer book and this question paper to the supervisor separately.

Q. 1) Puneet, the owner of Hard Rock Café is interested in how much people spend at the cafe. He examines 20 randomly selected bill receipts and writes down the following data (in Rs.):

1500 2000 2500 2500 3500 2500 3500 3000 4000 4000 2000 2500 1500 1500 3500 2500 3500 4500 6000 6000

(a) Determine the mean and standard deviation of the sample. (3)

A mathematician came up with bounds on how much of the data must lie close to the mean. In particular, for any positive k, at least $1 - 1/k^3$ proportion of the data lies within the interval given by k times the standard deviation around the mean.

(b) For k = 2, determine the upper and lower bounds for the data. (1)

(c) Using your answer to part (b), comment on whether the mathematician's theorem is correct in light of the given data. (2)

[6]

Q. 2) It is desired to simulate an observation of the random variable X with probability density function

$$f(x) = \begin{cases} 1/k; & 0 \le x \le k \\ 0; & \text{otherwise} \end{cases}$$

A random number r is generated from the uniform distribution over [0, 1]. The following values are then calculated

$x_1 = rk$; $x_2 = k(1-r)$; $x_3 = r/k$

State, with reason, which of the above values are valid simulated observations of X? **[2]**

Q. 3) Suppose the probability of an individual being born on any particular day of the year is given by 1/365.

(a) What is the probability that 2 people meeting at random have the same birthday? (2)

(b) Suppose now that a group has 3 individuals. What is the probability that at least two of these individuals will share a birthday? What if the group has 4 individuals? (4)

(c) Show that a group must have 15 individuals such that the probability of finding at least 2 people with the same birthday is 25%. (3)

[9]

Q. 4) For the random variable X, you are given:

E[X] = θ, θ > 0

Var[X] = $\theta^2/25$

$\hat{\theta} = \frac{kX}{k+1}$; k > 0 and where $\hat{\theta}$ is the estimate of θ

$MSE_{\hat{\theta}}(\theta) = 2[bias_{\hat{\theta}}(\theta)]^2$

Find k. **[4]**

Q.5) Vivek's company owns a factory. It buys insurance to protect itself against major repair costs. Profit equals revenues, less the sum of insurance premiums, retained major repair costs, and all other expenses. Company will pay a dividend equal to the profit, if it is positive.

You are given:

(i) Revenue from the factory is 1.70.

(ii) The distribution of major repair costs (k) for the factory is

K	*Probability*
0	*0.4*
1	*0.3*
2	*0.2*
3	*0.1*

(iii) The insurance policy pays the major repair costs in excess of that factory's deductible of 1 (i.e. claims will be payable after deducting 1 provided claims are greater than 1, else nil). The insurance premium is 110% of the expected claims for the insurance company.

(iv) All other expenses are 20% of revenues.

Show that the expected dividend is equal to 0.368. **[4]**

Q.6) Let $X_1,\ X_2,\ X_3, \dots \dots \dots \dots, X_n$ be n mutually stochastically independent random variables, each with common *pdf* $f_X(x) = 3\,(1 - x^2),\ 0 < x < 1$. If *Y* is the minimum of these n variables, find the CDF and the *pdf* of *Y*. **[4]**

Q.7) The scores on the final exam in Varun's risk management class have a normal distribution with mean θ and standard deviation equal to 8. θ is a normally distributed random variable with mean equal to 75 and variance equal to 36.

Each year, Varun chooses a student at random and pays the student equal to the student's score.

However, if the student fails the exam (i.e. score ≤ 65), then there is no payment.

Calculate the conditional probability that the payment is less than 90, given that there is a payment. **[5]**

Q.8) A random variable X follows a "triangular" distribution specified by the density

$$f(x) = \begin{cases} 4x & : 0 < x < \frac{1}{2} \\ 4(1-x) & : \frac{1}{2} < x < 1 \\ 0 & : \text{otherwise} \end{cases}$$

(a) Show that the moment generating function of X is given by:

$$M_X(t) = \frac{4}{t^2}\left(e^{\frac{t}{2}} - 1\right)^2$$

(5)

(b) Let X_1 and X_2 be uniform independent random variables on the interval (0,1), i.e. each with density

$$f(x) = \begin{cases} 1 & : 0 < x < 1 \\ 0 & : \text{otherwise} \end{cases}$$

and moment generating function

$$M(t) = \frac{(e^t - 1)}{t}$$

(i) Use the definition and properties of moment generating functions to derive the moment generating function of $Y = \frac{1}{2}(X_1 + X_2)$. (2)

(ii) Hence, comment on the distribution of the mean of two random samples from the uniform distribution on the interval (0, 1). (1)

[8]

Q. 9) Suppose that the joint probability density function of the bivariate random variable (X,Y) is given by:

$f_{X;Y}\,(x; y) = 1 - \alpha\,(1 - 2x)(1 - 2y), \qquad 0 \le x \le 1;\ 0 \le y \le 1.$

(a) Work out *E(XY)* (3)

(b) Work out the marginal density functions $f_X(x)$ and $f_Y(y)$ and hence *E(X)* and *E(Y).* (3)

(c) Attempt to prove or disapprove the following statement: (4)

"In this example, the variables X and Y are independent **if and only if they are uncorrelated**." **[10]**

Q. 10) $\{S_n\}_{n=1}^{\infty}$ is a sequence of independent and identically distributed random variables, each with mean 5 and variance 25. S_n represents aggregate claims from a risk in year n. The insurer intends to calculate the annual risk premium, ∏, for this risk such that:

Pr [S_n > ∏] = 0.01

(i) Assuming S_n has an exponential distribution, show that ∏ = 23.03. (2)

(ii) Calculate the value of ∏ assuming S_n has a lognormal distribution. (4)

(iii) Assuming that S_n has an exponential distribution, calculate the value of: (4)

P[(S_1 ≤ 23.03) ∩ (S_1 + S_2 ≤ 46.06)] **[10]**

Q. 11) You are given the following random sample of 30 auto claims:

54	140	230	560	600	1,100	1,500	1,800	1,920	2,000
2,450	2,500	2,580	2,910	3,800	3,800	3,810	3,870	4,000	4,800
7,200	7,390	11,750	12,000	15,000	25,000	30,000	32,300	35,000	55,000

Test the hypothesis that auto claims follow a Continuous Distribution Function *F*(*x*) with the following percentiles:

x	*310*	*500*	*2,498*	*4,876*	*7,498*	*12,930*
F(x)	*0.16*	*0.27*	*0.44*	*0.65*	*0.83*	*0.95*

Group the data using the largest number of groups such that the expected number of claims in each group is at least 5.

Using the chi-square goodness-of-fit test, determine the results of the test at 5% level of significance. **[5]**

Q. 12) Let $x_1, x_2, x_3, \dots\dots\dots\dots, x_n$ be a sample from the random variable X with *pdf*

$$f(x) = \frac{1}{\theta}\, x^{\left(\frac{1-\theta}{\theta}\right)}, \quad 0 < x < 1, \quad \text{and} \quad 0 < \theta < \infty,$$

(a) Find the maximum likelihood estimator of θ. (4)

(b) Show that the above estimate is an unbiased estimator of θ. (4)

[8]

Q. 13) Sachin, owns a towing company which provides all towing services to members of the City Automobile Club.

For a particular towing, the information is as below

Towing Distance	**Towing Cost**	**Frequency**
0 - 9.99 km	80	50%
10 - 29.99 km	100	40%
30+ km	160	10%

The automobile owner must pay 10% of the cost and the remainder is paid by the City Automobile Club. The number of towings has a Poisson distribution with mean of 1000 per year. The number of towings and the costs of individual towings are all mutually independent.

Using the normal approximation for the distribution of aggregate towing costs, show that the probability that the City Automobile Club pays more than 90,000 in any given year is 10%. **[5]**

Q. 14) Rajeev compared protein intake among three groups of women:

- women eating a standard India diet (STD)
- women eating a lacto-vegetarian diet (LAC) and
- women eating a strict vegetarian diet (VEG)

The mean and standard deviation of protein intake as well as the group sizes are presented in the table below.

Group	Mean	Standard Deviation	Number of women in group
STD	75	9	10
LAC	57	13	10
VEG	47	17	6

(a) Perform an overall F test to determine whether there is a significant difference in mean protein intake between the three groups stating both the null and alternative hypotheses. (5)

(b) Obtain 95% confidence intervals for each of the group means. Which groups appear different from one another? (3)

(c) Repeat part (b) using pair-wise confidence intervals and hypothesis tests for whether there are differences in mean protein intake. (7)

[15]

Q.15) Anand obtains cash from an ATM (cash machine) for his girlfriend. He suspects that the rate at which she spends cash is affected by the amount of cash he withdrew at his previous visit to an ATM.

To investigate this, he deliberately varies the amounts he withdraws. For the next 10 withdrawals, he records, for each visit to an ATM, the amount x (in Rs.) withdrawn, and the number of hours, y, until his next visit to an ATM.

Withdrawal	1	2	3	4	5	6	7	8	9	10
x	40	10	100	110	120	150	20	90	80	130
y	56	62	195	240	170	270	48	196	214	286

(a) Calculate the equation of the regression line of y on x (4)

(b) Interpret, in context of the question, the gradient of the regression line (1)

[5]

****