INSTITUTE OF ACTUARIES OF INDIA

EXAMINATIONS

10th May 2011

Subject CT3 – Probability & Mathematical Statistics

Time allowed: Three Hours (15.00 – 18.00)

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) If $Z \sim N(0,1)$, then for any a > 0, show that

$$P[Z > a] \leq \frac{1}{2} e^{-\frac{1}{2}a^2}$$

[5]

Q. 2) The gross incomes of a random sample of 10 school teachers were recorded (to the nearest 100 and in units of 1,000) and gave the following summary statistics: $\sum x = 321.6, \sum x^2 = 10628.31$. Income tax is charged at a single rate of 30% on that part of a person's income above 10,000. Calculate the sample mean and standard deviation of the 10 net incomes.

[3]

Q. 3) A uniform random number X divides [0, 1] into two segments. Let R be the ratio of the smaller versus the larger segment. Compute the density of R.

[5]

Q. 4) X has a Poisson distribution with a mean of 2. Y has a geometric distribution on the integers 0, 1, 2 ... and also with mean 2. You are given that X and Y are independent. Show that

$$P[X = Y] = \frac{1}{3} e^{-2/3}$$

[5]

Q. 5) You are the consulting actuary to a group of venture capitalists financing a search for pirate gold.

It is a risky undertaking: With probability 0.80, no treasure will be found and thus the outcome will be zero.

The rewards are high: With probability 0.20, treasure will be found. The outcome, if treasure is found, is uniformly distributed to [1000, 5000].

You use the inverse transformation method to simulate the outcome, where large random numbers from the uniform distribution on [0, 1] corresponds to large outcomes. Your random numbers for the first five trials are 0.95, 0.65, 0.75, 0.55 and 0.85.

Calculate the average of the outcomes of these first five trials.

[3]

Q. 6) In order to simulate an observation of a normal random variable it is suggested that

$$S = \sum_{i=0}^{n} X_i$$

is used, where $X_1, X_2, \ldots X_n$ is a random sample from a continuous uniform distribution on the interval (-0.5, 0.5).

(a) Determine the approximate distribution of S. (2)

(b) Determine the value of *n* which should be used if S is required to represent a standard normal random variable. (1)

(c) Explain why S has the same coefficient of skewness as a standard normal random variable. (1)

[4]

Q. 7) The random variable X takes values on the interval [0, 2].
You are given:

- $P(X = 1) = 0.25$
- $F(x \mid X < 1) = x^2$
- $F(x \mid X > 1) = x - 1$
- $E(X) = 1$

Find $F(1)$.

[5]

Q. 8) If X ~ Poisson(θ) and θ is large then the approximation

$$\frac{X-\theta}{\sqrt{\theta}} \sim approx.\, N(0,1)$$

may be used to derive an approximate confidence interval for θ.

You have observed a single value of 200 from this distribution. Using the above approximation and a continuity correction, derive an approximate 95% symmetrical confidence interval for θ.

[7]

Q. 9) A random sample of size n is drawn from the distribution with probability density function

$$f(x) = \frac{\theta}{(\theta+x)^2}; \qquad 0 < x < \infty,\ \theta > 0$$

Derive the asymptotic variance of the maximum likelihood estimator $\hat{\theta}$ of θ.

[5]

Q. 10) Let X be a random variable distributed in the interval (0, 1) under the density function

$$f(x \mid \theta) = \frac{\Gamma(3\theta)}{\Gamma(\theta)\Gamma(2\theta)} x^{\theta-1}(1-x)^{2\theta-1}$$

Here, θ (> 0) is some unknown parameter. The gamma function Γ(.) defined by

$$\Gamma(a) = \int_0^\infty u^{a-1} e^{-u}\, du,$$

satisfies Γ(a) = (a - 1)! if a is a positive integer.

(a) Derive the mode for the variable X under different permissible values of θ. (3)

Suppose we wish to test by means of the most powerful test, the following hypothesis based on a single observation of X:

$$H_0: \theta = 1 \quad against \quad H_1: \theta = 2$$

(b) Show that the critical region of the test has the following form for some constant k:

$$\{x \in (0,1): x(1-x)^2 > k\}$$

Suppose that X is observed to be $\frac{1}{7}$. (3)

(c) Using the shape of the critical region as obtained in (b), obtain the most powerful test for testing H_0 against H_1. (3)

(d) Hence, show that the p-value given by the above test is $\frac{27}{49}$. (1)

[10]

Q. 11) Suppose X is a continuous random variable with probability density function expressed by

$$f(x) = \begin{cases} \frac{1}{2}e^{-x}, & for\ x \geq 0 \\ \frac{1}{2}e^{+x}, & for\ x < 0 \end{cases}$$

Compute the moment generating function of X and deduce the mean and variance of X.

[5]

Q. 12) Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution a density function:

$$f(x;\ \theta) = \begin{cases} 2\theta^{-2}x, & 0 < x \leq \theta \\ 0, & otherwise \end{cases}$$

Here θ > 0 is an unknown parameter.

(a) Find the maximum likelihood estimator $\hat{\theta}$ of θ. (3)

(b) Find the mean square error of $\hat{\theta}$. (5)

[8]

Q. 13) We want to poll a population to estimate the fraction 'p' of eligible voters who will vote for the current ruling party in the country's next general election. We want to find the smallest number 'n' we need to poll to estimate 'p' with a margin of error of ±3% with 95% confidence. Find n.

[3]

Q. 14) Let X have one or the other of the discrete distributions given in the following table:

Value of X	=	1	2	3	5	7
$H_0 :\ f(x)$	=	0.2	0.3	0.1	0.3	0.1
$H_1 : f(x)$	=	0.3	0.1	0.3	0.2	0.1

(a) Determine the most powerful test for H_0 against H_1 at level α = 0.3 (4)

(b) Compare the power of this test with that of the test which rejects H_0 if X = 5 (1)

[5]

Q. 15) The lengths of 900 sentences randomly sampled from the book "*The Works, Wealth and Happiness of Mankind*" written by H. G. Wells are counted and denoted by $X_1, X_2, \ldots, X_{900}$. Here a length of a sentence is defined to be the number of words in that sentence.

A summary of the data is given by the contingency table below:

ln(X) ε	**(-∞, 1]**	**(1, 1.5]**	**(1.5, 2]**	**(2, 2.5]**	**(2.5, 3]**
Freq	1	6	40	121	290
ln(X) ε	**(3, 3.5]**	**(3.5, 4]**	**(4, 4.5]**	**(4.5, 5]**	**(5, ∞)**
Freq	250	156	34	2	0

Under a proposed model "M", $\ln(X_1), \ln(X_2), \ldots, \ln(X_{900})$ are independent and identically distributed according to $N(\mu, \sigma^2)$.

Assuming the above model is correct, define:

$$f_1(\mu,\sigma) = P[\ln(X_1) \le 1]$$
$$f_j(\mu,\sigma) = P\left[\frac{j}{2} < \ln(X_1) \le \frac{j+1}{2}\right] \quad for\ j = 2, 3, \ldots, 9$$
$$f_{10}(\mu,\sigma) = P[\ln(X_1) > 5]$$

(a) Write down the expression for $f_j(\mu,\sigma)$ $for\ j = 1, 2, \ldots, 10$ in terms of standard normal cumulative distribution function $\Phi(*)$. (2)

Suppose you know that $\mu = 3$ and $\sigma = 0.6$ under model "M". The following table gives the values of expected frequencies E(j) for each of the cells as defined above:

j	**1**	**2**	**3**	**4**	**5**
E(j)	0.387	5.202	37.125	140.229	*
j	**6**	**7**	**8**	**9**	**10**
E(j)	*	*	*	*	*

(b) Complete the above table. (2)

(c) Perform a chi-square goodness-of-fit test to check if the data is indeed a sample from the model "M". Use p-value to draw your conclusions. (4)

Now you are told that you don't know what the values of μ and σ are. You look at the actual data and obtain the following estimates $\hat{\mu} = 3.0177$ and $\hat{\sigma} = 0.5916$. You decide to redo the chi-square test using these estimates. Your revised test statistic value is 11.355.

(d) Has your conclusion changed from (c) and why? (2)

[10]

Q. 16) A student after reading up linear regression on a statistics textbook decides to apply the theory in practice. He computes the average rainfall (y) recorded in his city on each calendar day (x) during the month of September over the last 10 years.

He wants to fit a linear regression line for y on x and aims to use it to predict rainfall for the month of September in the next year.

You are given the following:

- $S_{xy} = \sum(x_i - \bar{x})(y_i - \bar{y}) = 599.62$
- $S_{yy} = \sum(y_i - \bar{y})^2 = 1{,}020.60$

(a) Calculate the least square estimate of the slope parameter $\hat{\beta}$ (3)

(b) State the estimate of error variance: $\hat{\sigma}^2$ and calculate it. (2)

(c) Find SS_{RES} , i.e., variability between the responses and their fitted values and so "unexplained" by the model. (1)

(d) Define and estimate the coefficient of determination R^2. (1)

(e) Test the significance of the regression coefficient under the hypothesis that there is no linear relationship between y and x. Comment on your result with the result of R^2 value. (3)

[10]

Q. 17) A psychologist was interested in whether different TV shows lead to more positive outlook in life. People were split in four groups and then taken to a room to view a program. The four groups saw: The Tom & Jerry Show, Futurama, The BBC News, No Program. After program a blood sample was taken and serotonin levels (y) measured (more serotonin means more happy).

Shows	serotonin levels (y)								Σy	Σy²
Tom & Jerry Show	11	7	8	14	11	10	5		66	676
Futurama	4	8	6	11	9	8			46	382
The BBC News	4	3	2	2	3	6			20	78
No Program	7	7	5	4	3	4	4	4	38	196

(a) Calculate the estimates for the:

i. Overall serotonin level (1)

ii. Common underlying variance in serotonin level. (3)

(b) Perform an analysis of variance to see if real difference exists between the mean serotonin levels. State your confidence level. (3)

[7]

****