INSTITUTE OF ACTUARIES OF INDIA

EXAMINATIONS

11th November 2013

Subject CT3 – Probability & Mathematical Statistics

Time allowed: Three Hours (10.30 – 13.30 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. *Please check if you have received complete Question Paper and no page is missing. If so, kindly get new set of Question Paper from the Invigilator.*

AT THE END OF THE EXAMINATION

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

Q. 1) The data below shows the percentages of customers who bought newspaper *A* from a magazine stall in city *K* for Monday to Friday in a randomly selected week:

62% 55% 63% 58% 62%

i) Find the mean and the median of the above data. (2)

ii) Let a% and b% be the respective percentages of customers who bought newspaper *A* from the stall for Saturday and Sunday in that week.

a) Write down (with reasoning) the least possible value of the median of the whole week's data. (1)

b) It is known that the mean and the median of the whole week's data is same as that obtained in part (**i**). Write down (with reasoning) one pair of possible values of 'a' and 'b' (3)

iii) The stall-keeper claims that since the mean and the median found in (**i**) exceed 50%, newspaper *A* has the largest market share among the newspapers in city *K*. Discuss. (1)

[7]

Q. 2) An urn contains 10 balls: 4 red and 6 blue. A second urn contains 16 red balls and an unknown number of blue balls. A single ball is drawn from each urn. The probability that both balls are of the same colour is 0.44.

Calculate the number of blue balls in the second urn.

[3]

Q. 3) Suppose the lifetime, T, of an electronic device is a random variable having a probability density function:

$$f_T(t) = 0.5\, e^{-0.5\, t}\,; \qquad t > 0$$

The device is given an efficiency value V = 5 if it fails before time t = 3. Otherwise, it is given a value V = 2 T.

i) Show that the probability density function of V is given by:

$$f_V(v) = \begin{cases} 0 & v < 5,\ \ 5 < v < 6 \\ 1 - e^{-1.5} & v = 5 \\ 0.25e^{-0.25v} & v \geq 6 \end{cases}$$

(4)

ii) Compute the expected efficiency value of the electronic device. (3)

[7]

Q. 4) Consider a random sample, $X_1, X_2 \ldots X_n$ from a population with mean μ and variance σ^2. Denote $\bar{X}$ as sample mean and S^2 as sample variance.

i) Show that:

$$S^2 = \frac{1}{2n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(X_i - X_j\right)^2$$

(4)

ii) Show that $\mathbb{E}\left[\left(X_i - X_j\right)^2\right] = 2\sigma^2$ for $\forall$ i ≠ j, and hence prove that S^2 is an unbiased estimator of σ^2 (3)

[7]

Q. 5) There are two independent random variables X and Y with probability density functions g and h respectively, where for any x > 0, we have:

$$g(x) = 3^{18}\, x^{17}\, \frac{e^{-3x}}{17!}\ ; \qquad h(x) = 3^{6}\, x^{5}\, \frac{e^{-3x}}{5!}$$

Find the probability density function of S = X + Y using the convolution formula:

$$f_S(s) = \int_{-\infty}^{\infty} g(x)\ h(s-x)\, dx$$

[5]

Q. 6) Let N be the number of claims on a motor insurance policy in one year. Suppose the claim amounts $X_1, X_2 \ldots$ are independent and identically distributed random variables, independent of N. Let S be the total amount claimed in one year for that insurance policy.

i) Write down expressions for E(S) and Var (S) in terms of the mean and variance of N and X_1. (1)

An insurer has sold 100 such motor insurance policies. For all policies, if there is a claim, it has mean μ (> 0) and variance σ^2. For the i^{th} policy (i = 1, 2 … 100), the actuary has two options to model the number of reported claims for the policy:

Option 1: A Bernoulli distribution with parameter β_i $(0 < \beta_i < 1)$

Option 2: A Poisson distribution with parameter β_i $(0 < \beta_i < 1)$

Let T and T**’** be the total amount claimed on the whole portfolio in one year for Options 1 and 2 respectively.

ii) Determine the mean and variance of T. (3)

iii) Determine the mean and variance of T’, and compare your answers with those obtained in part **(ii).** (4)

[8]

Q. 7) Consider a random variable U that has a uniform distribution on [0, 1] and let F be the cumulative distribution function of the standard normal distribution.

Define a random variable X as below:

$$X = \begin{cases} -F^{-1}\left(U + \frac{1}{2}\right) & \text{if } 0 \le U < \frac{1}{2} \\ -F^{-1}\left(U - \frac{1}{2}\right) & \text{if } \frac{1}{2} < U \le 1 \end{cases}$$

i) Show that X has a standard normal distribution. (7)

ii) Hence, simulate two observations of X using the random numbers 0.619 and 0.483 selected from the U (0,1) distribution (3)

[10]

Q. 8) Country A owns 10 nuclear plants which she claims are built for peaceful purposes. Country B suspects that some of these nuclear plants are used for producing nuclear weapons, and is considering an invasion of A.

To help her make the decision, country B sends a secret agent into A to investigate the true functions of two randomly selected nuclear plants there. If any of the two investigated nuclear plants is found to produce nuclear weapons, country B will invade country A; otherwise she will abandon the idea of invasion, trusting that A’s nuclear plants are all used for peaceful purposes. The secret agent is absolutely professional, and never makes any mistake in his investigation.

Let θ ε {0, 1 … 10} be the number of weapon-producing nuclear plants in country A. Let X be the number of nuclear plants found to be producing nuclear weapons by the secret agent. Thus X is a random variable with possible values 0, 1 or 2.

i) Formulate the decision making process for country B in terms of a hypothesis test based on the unknown parameter θ. (1)

ii) What is the type of error made by B if she does not invade A when some of the nuclear plants in A are indeed producing weapons? (1)

iii) Describe the critical region adopted by country B in terms of observation X. (1)

iv) Find the probability of a Type I error. (1)

[4]

Q. 9) It has been suggested that the outcome of successive tosses of a specially minted gold coin may not be independent and the following model has been proposed:

- The probability of getting a head on first toss is 0.5.
- The outcome of subsequent tosses depends on what happened in the previous toss of the coin.

Let θ be the probability that the outcome of successive tosses are of the same type.

In a study involving '*n'* independent three-toss experiments, the frequencies *a, b, c, d, e, f, g* and *h* of the eight possible outcome sequences and the associate probabilities are as follows:

HHH	HHT	HTH	THH
a	b	c	d
$\frac{1}{2}\theta^2$	$\frac{1}{2}\theta(1-\theta)$	$\frac{1}{2}(1-\theta)^2$	$\frac{1}{2}\theta(1-\theta)$

TTH	THT	HTT	TTT
e	f	g	h
$\frac{1}{2}\theta(1-\theta)$	$\frac{1}{2}(1-\theta)^2$	$\frac{1}{2}\theta(1-\theta)$	$\frac{1}{2}\theta^2$

Let S be a random variable denoting the total number of outcomes of type HH or TT observed in the above '*n*' three-toss experiments. Let s denote the observed value of S in the above experiment.

i) Show that $s = 2a + b + d + e + g + 2h$. (2)

ii) Show that the likelihood function for θ is proportional to $\theta^s(1-\theta)^{2n-s}$, and hence derive the maximum likelihood estimator (MLE) of θ. (4)

iii) Derive the asymptotic variance of the MLE of θ, and hence determine its approximate asymptotic distribution. [*Hint: Use the result* $E(S) = 2n\theta$] (3)

In a particular study involving 1000 three-toss experiments, the observed frequencies were:

a	b	c	d	e	f	g	h
135	131	125	125	123	115	125	121

iv) Calculate a large-sample 95% confidence interval for θ based on the above data. (4)

v) It has been suggested that the above model is incorrect in its assumption that the probability of a head on the first toss is 0.5. Use the above data to investigate this criticism by carrying out a suitable hypothesis test. Clearly state any assumptions you make in the process. (5)

[18]

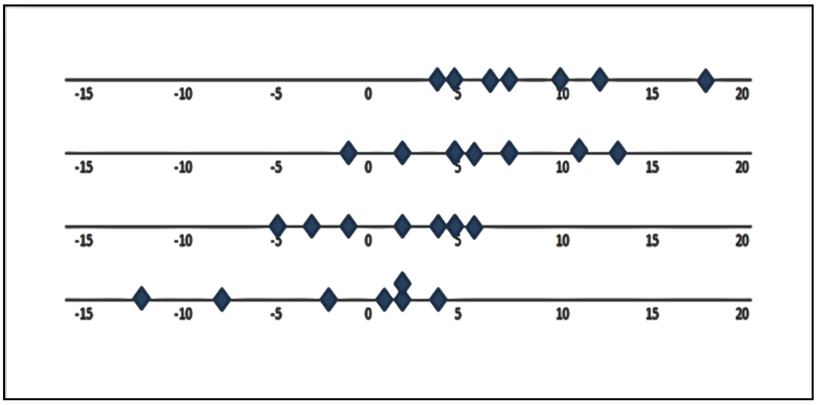
Q. 10) The government's urban planning committee is concerned about the increasing traffic within the major cities and the time required for commutation within the city. In order to understand the situation, it commissions an external agency to conduct a survey looking into the difference between the times taken to commute in the peak hours (i.e. office hours) compared with that in the non-peak hours (i.e. non-office hours).

The study was carried out in 4 major cities with a randomly selected sample of 7 commuters in each city. The table in the next page shows the difference between the time taken (in minutes) for each individual to commute to office in peak hours and non-peak hours:

Cities →	A	B	C	D	$\sum y_{\cdot j}$	$\sum y_{\cdot j}^2$
	18	2	2	-8	14	396
	7	11	-3	4	19	195
	8	13	-5	2	18	262
	4	-1	5	-12	-4	186
	10	5	4	1	20	142
	5	8	6	-2	17	129
	12	6	-1	2	19	185
Total	**64**	**44**	**8**	**-13**	**103**	**1,495**

The agency wanted to infer on the mean time differences among the 4 cities and thus decides to use an analysis of variance (ANOVA) approach.

i) The diagram given below compares the time differences in commutation in the four cities (in order). Suggest brief comments one can make on the basis of the plot. (2)



ii) State the assumptions required to carry out the ANOVA. (2)

iii) Carry out the ANOVA test and draw your conclusion at the 5% significance level. (4)

iv) Carry out an analysis of the mean differences using a least significant difference approach at the 5% significance level. (6)

[14]

Q. 11) Dell and IBM are well known in computer industry. If the computer industry is doing well then we may expect the stocks of these two companies as well to increase in value. If the industry goes down then we would expect both may go down as well. The table below gives data on the share prices (in US $) of Dell (X) and IBM (Y) at the end of each month for a calendar year:

X	27.9	40.7	37.8	31.6	37.5	31.6	29.2	24.5	30.9	25.6	37.9	30.0
Y	97.4	105.0	145.5	126.2	114.2	106.7	76.7	65.6	68.9	82.2	95.6	78.5

$\sum x = 385.2$; $\sum x^2 = 12{,}666.58$; $\sum y = 1{,}162.5$; $\sum y^2 = 119{,}026.9$; $\sum xy = 38{,}191.41$

i) Calculate the least squares fit regression line in which IBM share price is modelled as the response and the Dell share price as the explanatory variable. (4)

ii) Determine a 95% confidence interval for the slope coefficient of the model. State any assumptions made. (5)

iii) Use the fitted model to construct 95% confidence intervals for the mean IBM share price when the Dell Share price is US $ 40. (4)

[13]

Q. 12) The quality control division of a factory suspect defects in the production of a model of mobile phone battery which has resulted in reduction of battery life.

In a post-production inspection of the last batch, it was found that the average time spent in conversation was 290 minutes in a sample of 60 batteries. Before this, the experience has been such that the life of the battery followed a normal distribution with a mean of 300 minutes and a standard deviation of 30 minutes.

Test if the suspicion, that there has been a reduction in the battery life compared to the past experience, is true at the 1% level of significance.

[4]

****