INSTITUTE OF ACTUARIES OF INDIA

EXAMINATIONS

20th May 2013

Subject CT3 – Probability & Mathematical Statistics

Time allowed: Three Hours (10.00 – 13.00)

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. ***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) Suppose two independent dice, one blue and one red, are rolled. Let A be the event that the face on the blue die is even, B the event that the face on the red die is even, and C the event that sum of the two faces is even. Show that the events A, B and C are pairwise independent and not mutually independent.

[4]

Q. 2) Let X be a random variable such that $|\mathbb{E}(X)| < \infty$ and

$$\mathbb{P}\left(X > \frac{1}{2} + x\right) = \mathbb{P}\left(X \leq \frac{1}{2} - x\right) \text{ for all x} \in \mathbb{R}$$

Show that the median of the random variable X is ½.

[2]

Q. 3) Suppose $\{Z_0, Z_1, Z_2 \ldots\}$ are independent random variables, each having a standard Normal distribution. In terms of members of this sequence, write down (with reasoning) random variables having:

i. A t-distribution with **5** degrees of freedom. (2)

ii. A F-distribution with **9** and **16** degrees of freedom (2)

[4]

Q. 4) LetX be a random variable having a gamma distribution with mean αλ and variance $\alpha\lambda^2$ and let the i^{th} cumulant of the distribution of X be denotedκ_i.Assuming the moment generating function of X, determine the values of κ_2, κ_3andκ_4.

[4]

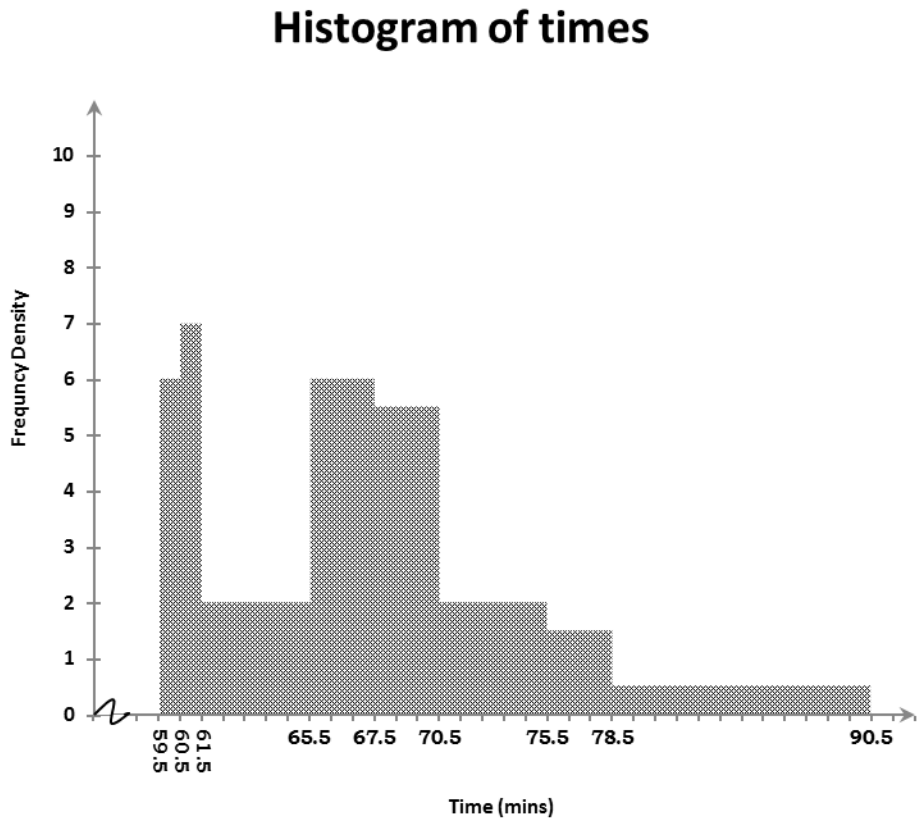
Q. 5) Let X and Y be the number of hours that a randomly selected person watches movies and sporting events, respectively, during a three-month period. The following information is known about X and Y:

- E[X] = 50
- E[Y] = 20
- Var[X] = 50
- Var[Y] = 30
- Cov[X, Y] = 10

100 people are randomly selected and observed for these three months. Let T be the total number of hours that these 100 people watch movies or sporting events during these three-month period. Compute an approximate value of$\mathbb{P}(T < 7050)$.

[3]

Q. 6) The histogram below shows the time (**t**) taken in minutes for 140 runners to complete a charity fund-raiser run:



i. Use the histogram to complete the following frequency table for the variable 't':

t	59.5 - 60.5	60.5 - 61.5	61.5 - 65.5	65.5 - 67.5	67.5 - 70.5	70.5 - 75.5	75.5 - 78.5	78.5 - 90.5
Frequency	12	14	16	24	33	20	*	*

(3)

ii. Use linear interpolation to approximate the values of the 3quartiles for t. (5)

iii. One measure of skewness is defined as:

$$skew = \frac{(Q_1 - 2Q_2 + Q_3)}{IQR}$$

Evaluate this measure and describe the skewness of the data.

(2)

[10]

Q. 7) A random variable X has the probability density function

$$f(x) = \begin{cases} c\ \ exp\left(-\dfrac{x^2}{2}\right), & x < 0 \\ \dfrac{c}{\sigma}\ \ exp\left(-\dfrac{x^2}{2\,\sigma^2}\right), & x \geq 0 \end{cases}$$

Here: σ > 0 is an unknown parameter and c is a given constant.

i. Find the value of **c**. (3)

ii. What is the well-known distribution when $\sigma = 1$?

The following random sample is drawn from the above population:

-0.42 2.80 -1.47 0.31 -0.95 -0.39 -0.51 0.20 1.15 2.18 (1)

iii. Compute the maximum likelihood estimate of σ. (6)

[10]

Q. 8) **i.** What are the three key properties that a pivotal quantity of the form $g(\underline{X}, \theta)$ should possess before it can be used to derive the confidence interval of θ? (3)

ii Assume that the salaries of chartered accountants have a N (μ,σ^2) distribution and both μ and σ^2 are unknown.

A random sample of 5 chartered accountants is taken and their salaries are: 639, 642, 652, 645 and 648 (in ₹ '000 p.a.).

a) Calculate a 95% confidence interval for the average salary of chartered accountants based on the sample. (5)

b) Calculate a 95% confidence interval of the form "$\sigma < L$" for the standard deviation of the salary of chartered accountants in the population. (3)

[11]

Q. 9) Generally it has been observed that the pace bowling attack of the Indian cricket team is *weaker* in terms of pace when compared to other countries' teams. The head coach of the National Cricket Academy of India decided to conduct a study to measure how much of this weakness can be attributed to dietary practices. As both male and female cricketers are trained at the Academy, the head coach decided to try out the new specialised diet on both genders so that he could also check if the impact would differ by gender too.

On the starting day of the experiment the head coach randomly selected 10 cricketers of both gender and asked them to bowl at full pace. He recorded the top speed achieved (X) while being accurate after bowling 10 consecutive deliveries. The players were then put on the specialised diet for next 10 weeks. At the end of the period, the same players were called upon and their top speed (Y) among 10 deliveries was recorded.

The data below gives the top bowling speeds (in km/hr.) achieved before and after being on the specialised diet for each of the 20 bowlers along with some summary statistics:

	Top Bowling Speed for Male Pacers before and after diet										Total
Bowler #	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	
Before: X	105	108	113	81	137	110	148	135	127	108	1,172
After: Y	127	145	156	129	162	124	148	147	151	132	1,421
Δ = Y - X	22	37	43	48	25	14	0	12	24	24	249
Δ * Δ	484	1,369	1,849	2,304	625	196	0	144	576	576	8,123

	Top Bowling Speed for Female Pacers before and after diet										Total
Bowler #	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	
Before: X	105	142	55	81	116	123	102	87	84	147	1,042
After: Y	118	134	92	113	134	142	134	129	116	137	1,249
Δ = Y - X	13	-8	37	32	18	19	32	42	32	-10	207
Δ * Δ	169	64	1,369	1,024	324	361	1,024	1,764	1,024	100	7,223

i. Construct a 95% confidence interval for the mean difference between the bowling speeds before and after taking the special diet for the male bowlers. (4)

ii. Construct a 95% confidence interval for the mean difference between the bowling speeds before and after taking the special diet for the female bowlers. (4)

iii. Comment briefly on the two confidence intervals obtained in parts (i) and (ii). (1)

iv. In order to perform a two-sample t-test to investigate whether the effects of the special diet differs between males and females, the head coach makes a critical assumption that the variances of the differences in the male and female samples do not differ at the 5% significance level. Verify if this holds. (3)

v. Hence, perform the above t-test at the 5% significance level. (4)

[16]

Q. 10) The insurance regulator has conducted a study to understand the relation between the number of branches a life insurance company operates with and the number of policies it sells. At the end of the month the regulator examined the records of 10 insurance companies. It obtained the total number of branches (x) and the number of policies (y) sold in the month. The collected data is given in the table below;

Company	**A1**	**A2**	**A3**	**A4**	**A5**	**A6**	**A7**	**A8**	**A9**	**A10**
Number of Branches (x)	5	9	2	3	3	1	1	6	5	4
Number of Policies (y)	73	120	34	46	35	24	26	93	45	66

A set of summarised statistics based on the above data is given below:

$$\sum x = 39; \sum x^2 = 207; \sum y = 562; \sum y^2 = 40{,}508; \sum xy = 2{,}853$$

i. Calculate the correlation coefficient between x and y. (4)

ii. Calculate the fitted linear regression equation of y on x. (2)

iii. Calculate the "total sum of squares" together with its partition into the "regression sum of squares" and the "residual sum of squares". (2)

iv. Using the values in part (iii), calculate the coefficient of determination R^2.
Comment briefly on its relationship with the correlation coefficient calculated in part (i). (2)

[10]

Q. 11) A popular fashion magazine conducted a survey on how much money women spend on parlours in a Tier II city of India. Six parlours were being compared with regard to the fees they charged from the female customers for one visit to the parlour. Independent random samples of 5 female customers from each parlour are examined and the fees charged (in ₹) are as follows:

Parlour	**A**	**B**	**C**	**D**	**E**	**F**
	298	306	354	294	248	294
	338	278	308	298	268	260
	254	262	326	272	280	312
	338	238	368	280	324	340
	266	234	298	336	298	284
Total	**1,494**	**1,318**	**1,654**	**1,480**	**1,418**	**1,490**

Consider the model:

$$Y_{ij} = \mu + \tau_i + e_{ij} \qquad i = 1, 2, \dots, 6 \qquad j = 1, 2, \dots, 5$$

$$e_{ij} \sim N(0, \sigma^2)$$

Here:

- y_{ij} is the fee charged by the i^{th} parlour to the j^{th} female customer.
- The e_{ij} are independent and identically distributed.
- σ^2 is the underlying unknown common variance.
- $\sum \tau_i = 0$.

A couple of summarised statistics based on the above data is given below:

$$\sum y_{ij} = 8{,}854; \quad \sum y_{ij}^2 = 2{,}647{,}876$$

i. Perform an analysis of variance (ANOVA) for these data and show that there are no significant differences, at the 5% level, between mean fees being charged by each parlour. (5)

ii. Calculate a 95% confidence interval for σ using $\frac{SS_R}{\sigma^2}$ as a pivotal quantity with a χ^2 distribution. (3)

[8]

Q. 12) Let Pbe a random variable having a beta distribution with parameters α (> 0) and β (> 0) defined over the region (0, 1).

i. Write down the probability density function of P. (1)

ii. Show that for any real number k (≥ 0):

$$\mathbb{E}[P^k] = \frac{\Gamma(\alpha+k)\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\alpha+\beta+k)}$$

Hence compute $\mathbb{E}[P]$ and $\mathbb{Var}[P]$. (6)

iii. Show that:

$$\mathbb{E}[P(1-P)] = \frac{\alpha\beta}{(\alpha+\beta)(\alpha+\beta+1)}$$ (4)

A pharmaceutical company is conducting a new drug trial and have chosen **n** patients admitted in the city hospital to study the efficacy of the drug. For i = 1, 2 … n, let:

- X_i denotes the random variable which takes the value of 1 if the trial is successful for the i^{th} patient and 0 otherwise;

- P_i denotes the probability that the drug trial will be successful for the i^{th} patient. As the patients are different, the in-charge of the trial is reluctant to assume that P_i are constant but rather follow a beta distribution with parameters α (> 0) and β (> 0).

Let $S = \sum_{i=1}^{n} X_i$ denote the total number of successful trials among the n patients.

iv. Using the theory of Conditional Expectations or otherwise, show that:

a) $\mathbb{E}[S] = \frac{n\alpha}{\alpha+\beta}$

b) $\mathbb{Var}[S] = \frac{n\alpha\beta}{(\alpha+\beta)^2}$

(7)

[18]

****