INSTITUTE OF ACTUARIES OF INDIA

EXAMINATIONS

28thOctober 2014

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) Define the following types of categorical data and provide one example of each:

i) Dichotomous data (1)
ii) Nominal data (1)
iii) Ordinal data (1)
[3]

Q. 2) There is a box of red and black pens kept in it. 40% of the pens are red while the rest are black. A student is examining the contents of the box in random order, one by one, until she finds four red pens.

i) What is the probability that exactly 10 pens need to be examined before four red pens are found? (2)

ii) Find the expected number of pens that needs to be examined until two red pens are found? (2)
[4]

Q. 3) For each question addressed to a participant in a game show, the participant selects the number option 1, 2 or 3 at random. 200 random questions are addressed to him in the game show.

If T is the sum of all the number options chosen by the participant, calculate the approximate probability that T lies between 380 and 420 inclusive. **[5]**

Q. 4) The following frequency table contains the marks (obtained out of 100) of 72 graduating students in their final examination:

Marks	55	60	63	67	70	72	74	75	81	85	89	91	97
Frequency	1	3	7	8	5	4	11	7	9	8	5	2	2

i) Draw a boxplot for this data with appropriate labels. (*Graph paper is not required*) (3)

ii) Find the inter-quartile range for this data. (1)
[4]

Q. 5) Let X and Y be identically distributed and uncorrelated random variables such that the moment generating function of the random variable Z = X + Y is:

$$M_Z(t) = 0.09e^{-2t} + 0.24e^{-t} + 0.34 + 0.24e^{t} + 0.09e^{2t}, \quad -\infty < t < \infty$$

i) Compute E(Z) and Var(Z). (3)

ii) Using part (i), show that the marginal distribution of X (or Y) is:

Value	-1	0	1
Probability	0.30	0.40	0.30

(2)

iii) Hence construct a table showing the joint distribution of X and Y. (4)

iv) Are X and Y independent? Justify your answer. (3)
[12]

Q. 6) **i)** Suppose X denotes claim amount which follows an Exponential distribution with mean θ. Suppose q denotes the probability of a claim for an insured population of size w consisting of independent lives and let N be the number of claims. Assume that the random variable X is independent of N.

Let S denote the total reported claims. Show that the mean and variance of the random variable S are given as below:

- $E(S) = wq\theta$
- $Var(S) = wq(2 - q)\theta^2$ (2)

ii) A chemical factory has approached a life insurer for a group life cover quotation for its 1000 employees. After analysing the last 10 years' data for the factory, the actuary came up with the following summary statistics:

Age Group	Proportion	Death Count[1]	Average Benefit[2]
18 - 35	50%	25	5.0
36 - 50	30%	75	6.0
51 - 65	20%	100	7.5

[1] **Death Count**: Average number of deaths per 1000 lives over the last 10 years

[2] **Average Benefit**: Average claim amount (in Lakhs) paid on death of an employee over the last 10 years

From previous experience, the actuary knows that the claim sizes usually follow an Exponential distribution. You may assume that all lives, the number of claims and size of claims are independent across all age groups.

The actuary wants to set the single premium for the cover such that the probability that total claims will exceed premium received is 0.05. Applying results from part (i) for each age group in turn and using the normal approximation, determine the premium that the insurer will quote. (5)

[7]

Q. 7) Suppose $X_1, X_2, \ldots, X_{2n}$ are independent observations from the Bernoulli distribution with unknown parameter p $(0 < p < 1)$. An actuarial student wants to find a suitable estimator for p^2, say λ. One estimator that is of his special interest is:

$$\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} Z_i \text{ where } Z_i = X_{2i-1}X_{2i} \text{ for } i = 1, 2 \ldots n.$$

i) Argue why $Z_1, Z_2, \ldots, Z_n$ can be considered as independent observations from the Bernoulli distribution with parameter λ. (3)

ii) Show that $\hat{\lambda}$ is an unbiased estimator of λ. (1)

iii) Find the Cramér-Rao lower bound for the variance of unbiased estimators of λ. (5)

iv) Examine if the variance of $\hat{\lambda}$ attains the Cramér-Rao lower bound. (1)

[10]

Q. 8) A random sample of 25 values $x_1,\ x_2\ ,\dots,x_{25}$ was drawn from a normal population with mean θ and standard deviation $\theta(\theta > 0$ is unknown), and gave the following summary statistics:

$$\sum_{i=1}^{25} x_i = 69.66, \qquad \sum_{i=1}^{25} x_i^2 = 255.52.$$

Calculate a 95% confidence interval for θ. **[5]**

Q. 9) In preparation for the upcoming half-marathon a few months away, eight runners had begun their practice runs. Below is the data for the time (in minutes) taken by the runners to complete the half-marathon in the practice run and the time taken by them to complete it on the final day:

Runner ID	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Practice time taken (x)	154	162	175	185	142	136	140	145
Final time taken (y)	140	142	158	160	132	127	124	122

You are given:

$$\sum x = 1{,}239; \quad \sum y = 1{,}105$$

$$\sum x^2 = 194{,}095; \quad \sum y^2 = 154{,}141; \quad \sum\sum xy = 172{,}880$$

i) Determine the fitted linear regression equation of the final time taken by the runners on the practice time taken by them. (4)

ii) Assuming a full normal model, calculate an estimate of the error variance $\sigma^2$and obtain a 90% confidence interval for σ^2. (4)

iii) Calculate the proportion of variation explained by the model. Hence, comment on the fit of the model. (2)

[10]

Q. 10) A consumer testing service, wishing to test the accuracy of the thermostats of three different kinds of electric irons, set them at 55^0C and obtained the following actual temperature readings (in ^{0}C) by means of a thermocouple:

Iron	**Temperature**								***sum***	***sum of squares***
X	50	40	45	60	60	45	55	40	*395*	*19,975*
Y	70	75	80	85	82	81			*473*	*37,435*
Z	60	55	45	55	50	55	45		*365*	*19,225*

i) Calculate the estimate of the overall mean temperature. (1)

ii) Construct an appropriate ANOVA table and test whether the differences among the three means can be attributed to chance. State any assumptions you make in the process. (6)

iii) Compute an estimate of the underlying common variance in temperature readings. (1)

[8]

Q. 11) A quality control engineer is monitoring the number of defective items in each lot of shipment received by the factory. The proportion X of defective items in a lot can be assumed to be a random variable having the following probability density function:

$$f(x;\,\theta) = \theta(\theta+1)x^{\theta-1}(1-x); \quad 0 < x < 1 \quad \text{and} \quad \theta > 0 \text{ is unknown}$$

The engineer took readings of the proportion defective for each of 100 lots. The observed proportions are $x_1,\ x_2\,, \dots, x_{100}$.

Let $\hat{\theta}$ be the Maximum Likelihood Estimate (MLE) of θ.

i) Show that $\hat{\theta}$ satisfies the following quadratic equation:

$$\theta^2 + (2k+1)\,\theta + k = 0 \qquad \text{where } k^{-1} = \frac{1}{100}\sum_{i=1}^{100} \log_e x_i$$

(4)

ii) You are given:

$$\sum_{i=1}^{100} \log_e x_i = -58.334.$$

Compute the MLE of θ. (3)

The engineer tabulated the proportion of defective items observed into seven categories. The table below shows:

- the range of x-values for each category
- the frequency of observed proportions of defective items under each category
- the value of $\hat{\phi}_j$, the MLE of the probability that an observation falls in the j[th] category, calculated under the assumption that the underlying distribution is as stated above, for j = 1, 2 … 7. Of these, $\hat{\phi}_1$ and $\hat{\phi}_2$ are missing.

Category	**Observed Frequency**	$\hat{\phi}_j$
0.00 - 0.25	9	*
0.25 - 0.50	20	*
0.50 - 0.60	14	0.163
0.60 - 0.70	21	0.176
0.70 - 0.80	17	0.167
0.80 - 0.90	14	0.128
0.90 - 1.00	5	0.052

iii) Show that the maximum likelihood estimate of ϕ_1 is 0.051 and hence calculate the value of $\hat{\phi}_2$. (4)

iv) Perform an appropriate hypothesis test to investigate whether the assumed model fits the data well, and report your conclusions. (5)

[16]

Q. 12) The weight of an object is measured using an electronic scale that reports the true weight (100 milligrams) plus a random fluctuation that is normally distributed with zero mean and unknown variance σ^2.

A technician measures the weight of one object 10 times, and observes these values (x_k) independently: 100.6, 98.0, 101.2, 102.0, 104.8, 99.4, 97.2, 101.4, 95.4, 100.0

His supervisor picks a different electronic scale that reports the true weight (100 milligrams) plus a random fluctuation that is normally distributed with zero mean and unknown variance ψ^2.

He measures the weight of the same object 8 times, and observes these values (y_k) independently: 99.4, 102.3, 100.7, 98.8, 98.3, 101.6, 99.5, 99.4

The aim of the exercise is to check if using a different scale adds to the variability in the measurement.

i) Write down the distributions of $\left(\frac{X_k-100}{\sigma}\right)$and $\left(\frac{Y_k-100}{\psi}\right)$.

Hence, state the distributions of

$$V=\sum_{i=1}^{10}\left(\frac{X_k-100}{\sigma}\right)^2 \text{ and } W=\sum_{i=1}^{8}\left(\frac{Y_k-100}{\psi}\right)^2.$$ (2)

ii) Define $U=\dfrac{\sigma^2(V/10)}{\psi^2(W/8)}$.

State the distribution of $U\,\frac{\psi^2}{\sigma^2}$and explain why this can be regarded as the pivotal quantity for $\frac{\sigma^2}{\psi^2}$. (3)

iii) Hence, derive a 95% confidence interval for $\frac{\sigma^2}{\psi^2}$. (6)

iv) It is now required to test $H_0\colon \sigma^2=\psi^2$ against $H_1\colon \sigma^2\neq\psi^2$ at the 5% level. Compute the p-value for this hypothesis test and draw your conclusions. (3)

Use the following table of values for the Cumulative Distribution Function (CDF) of the relevant statistical distribution:

x	**2.50**	**2.75**	**3.00**	**3.25**	**3.50**	**3.75**	**4.00**	**4.25**	**4.50**	**4.75**	**5.00**
CDF	*0.89641*	*0.91747*	*0.93355*	*0.94596*	*0.95564*	*0.96327*	*0.96934*	*0.97422*	*0.97818*	*0.98141*	*0.98407*

v) Interpret the result obtained in part (iv) with reference to the confidence interval obtained in part (iii). (2)

[16]

****