D–GT–M–TTA

STATISTICS

Paper—I

Time Allowed : Three Hours	Maximum Marks : 200

INSTRUCTIONS

Candidates should attempt Question Nos. 1 and 5 which are compulsory, and any THREE of the remaining questions, selecting at least ONE question from each Section.

All questions carry equal marks.

Marks allotted to parts of a question are indicated against each.

Answers must be written in ENGLISH only.

Assume suitable data, if considered necessary, and indicate the same clearly.

(Notations and symbols are as usual)

Important Note

All parts/sub-parts of a question being attempted are to be answered contiguously on the answer-book. That is, where a question is being attempted, all its constituent parts/sub-parts must be answered before moving on to the next question.

Pages left blank, if any, in the answer-book(s) must be clearly struck out. Answers that follow pages left blank may not be given credit.

Section—A

1. Answer the following : 8×5=40

(a) A bag contains 10 coins—one worth 10 rupees and each of other nine worth 5 rupees. A person draws one coin at a time without replacement until he draws the coin worth 10 rupees. What is the expected total amount he has drawn (including the last coin worth 10 rupees)?

(b) What is the distribution of a random variable with characteristic function

$$\frac{1}{2}e^{-\frac{1}{2}t^2} + \frac{1}{2}\,?$$

(c) There are two coins one with probability $\frac{1}{3}$ of heads and the other with probability $\frac{2}{3}$ of heads. One of the coins is picked at random and let S_n be the number of heads obtained when the chosen coin is tossed n times. Which of the following is/are correct?

Given any $\varepsilon > 0$

(i) $\lim\limits_{n \to \infty} P\left(\left| \frac{S_n}{n} - \frac{1}{2} \right| < \varepsilon \right) = 1$

(ii) $\lim\limits_{n \to \infty} P\left(\left| \frac{S_n}{n} - \frac{1}{3} \right| < \varepsilon \right) = \frac{1}{2}$

Explain your answer.

(d) If X has a Poisson distribution and the prior distribution of its parameter λ is a gamma distribution with parameters α and β, then show that the posterior distribution of λ given $X = x$ is a gamma distribution with the parameters $\alpha + x$ and $\beta / (\beta + 1)$. Also obtain the mean of the posterior distribution of λ.

(e) (i) State and prove Wald's fundamental identity on sequential analysis.

(ii) Prove that the SPRT terminates with probability 1.

2. (a) Let X be a random variable with normal distribution $N(0, 1)$. Find the distribution of $Y = e^X$. 10

(b) State and prove Lindeberg-Levy central limit theorem. 10

(c) A fair coin is tossed independently n times. Let S_n be the number of heads obtained. Use Chebyshev's inequality to find a lower bound of the probability that $\frac{S_n}{n}$ differs from $\frac{1}{2}$ by less than $0 \cdot 1$, when (i) $n = 100$ and (ii) $n = 100000$. 10

(d) A monkey jumps to get a fruit from a tree. Suppose the probability that the monkey fails to get the fruit at the nth attempt is $\frac{1}{n}$. Using Borel-Cantelli lemma, show that the monkey will eventually get the fruit with probability 1. 10

3. *(a)* Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from the uniform population given by

$$f(x;\ \theta) = \begin{cases} 1, & \text{for } \theta - \frac{1}{2} < x < \theta + \frac{1}{2} \\ 0, & \text{elsewhere} \end{cases}$$

Find the maximum likelihood estimator for θ. Comment on this estimator. 10

(b) State factorization theorem of sufficient statistic.

Let $X_i\ (i = 1, 2, \ldots, n)$ be a random sample from a distribution with p.d.f.

$$f(x;\ \theta) = \begin{cases} \theta x^{\theta - 1}, & \text{if } 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

where $0 < \theta < \infty$. Find a sufficient statistic for θ. 10

(c) If $X_1, X_2, \ldots, X_n$ constitute a random sample from the population given by

$$f(x;\ \theta) = \begin{cases} e^{-(x - \theta)}, & x > 0 \\ 0, & \text{elsewhere} \end{cases}$$

show that the smallest sample value is a consistent estimator of the parameter θ. 10

(d) Find the critical region of the likelihood ratio test for testing the null hypothesis $H_0 : \mu = \mu_0$ against the composite alternative $H_1 : \mu \neq \mu_0$, on the basis of a random sample of size n from a normal population with known variance σ^2. 10

4. *(a)* Describe the method of constructing similar regions in testing composite hypothesis. Show that every test based on a similar region will have Neyman-Pearson structure with respect to a sufficient statistic T, if the family of distributions $f_{\theta}(T)$ of T for $\theta \in \Theta_0$ is boundedly complete. 10

(b) State and prove inversion theorem of characteristic function. 12

(c) On a genetic inbreeding experiment with single pair of allelomorphs A and a, the progeny of a mating $Aa \times Aa$ results in 50% heterozygotes Aa, where the proportion of heterozygotes being reduced by one-half at each generation. Find the transition matrix of various genotypes and the proportion of heterozygotes at rth generation. 8

(d) Write notes on the following : 5×2=10

(i) Kolmogorov-Smirnov two-sample test

(ii) Wilcoxon-Mann-Whitney test

Section—B

5. Answer the following : 8×5=40

(a) Describe the technique of analysis of variance for testing equality of a number of regression equations.

(b) Define 'estimable parametric function'. State the conditions of estimability. Show that such a function has a unique linear estimator which is best unbiased.

(c) Define Hotelling's T^2 and Mahalanobis's D^2 statistics, and obtain the relationship between them. Explain the uses of these statistics.

(d) Show that Horvitz-Thompson estimate is an unbiased estimate of the population mean. Obtain the variance of the estimate of the mean under Horvitz-Thompson estimate.

(e) Prove that the number of treatments common between any two blocks in a symmetrical BIBD is λ.

6. (a) State the Cramer-Rao lower bound. Give two examples, one in which the lower bound is attained and another where it is not attained. 10

(b) What is principal component analysis? Discuss its uses with examples. Show that the principal components are all uncorrelated. 10

(c) Describe the AOV for a two-way classification (random effects model) with one observation per cell. 10

(d) Describe Fisher's method of discrimination between two groups on the basis of multiple measurements. Explain how this procedure can be represented as a problem in regression. 10

7. (a) Show that, for estimation of population mean

$$V_{\text{opt}} \le V_{\text{prop}} \le V_{\text{SR}}$$ 10

(b) A population consists of $nk + p$ units, where n, p, k are positive integers $(p < k)$. A systematic sample is selected with k as the class interval. Show that the sample mean is a biased estimate of the population mean. Obtain an unbiased estimate of the population mean. 10

(c) Compare ratio and regression methods of estimation. Show that the ratio method gives a biased estimate of the population total. Derive an approximate expression for the bias. 10

(d) Why is it said that a split-plot design confounds main effects? Give an outline of analysis of a split-plot design. 10

8. (a) The yield of one plot of a $p \times p$ LSD is missing. Estimate the missing observation and give an outline of the analysis of the design. 10

(b) For two-stage sampling, where the first-stage units are of equal size, obtain an estimator of population mean. Also obtain the expression for the variance of the estimator. 10

(c) Discuss, in detail, how total and partial confounding can be done in 3^3-factorial experiment. Give the partition of d.f., contents of the blocks and calculation of sum of squares in each case. 10

(d) Write explanatory notes on the following : 5×2=10

(i) Lattice designs

(ii) Non-sampling errors

★ ★ ★