

STATISTICS

(Paper I)

Time Allowed : Three Hours
Maximum Marks : 300

QUESTION PAPER SPECIFIC INSTRUCTIONS

(Please read each of the following instructions carefully before attempting questions)

There are TWELVE questions divided in three Sections.

Candidate has to attempt FIVE questions in all, selecting at least ONE and not more than TWO questions from each Section.

The number of marks carried by a question/part is indicated against it.

Wherever any assumptions are made for answering a question, they must be clearly indicated.

Word limit in questions, wherever specified, should be adhered to.

SECTION 'A'

1. (a) What is a random experiment? Consider the random experiment “a die is rolled n times”. Define the associated sample space and state the number of elementary events in it. What is the event that “number 1 shows at least once”? How many elementary events are there in this event? 25
- (b) A fair die is rolled twice. Let A be the event that the first throw shows a number ≤ 2 and B the event that the second throw shows at least 4. Describe the event $A \cap B$ and find $P(A \cap B)$. 15
- (c) A point is picked at random from the unit square $\Omega = \{(x, y); 0 \leq x \leq 1, 0 \leq y \leq 1\}$. Find the probability that the point (i) lies in the circle with centre $(1/2, 1/2)$ and radius $1/2$, (ii) lies at most a unit distance from the origin $(0, 0)$. 20
2. (a) When is a random variable said to be of the discrete type? What is the pmf of such a random variable? Define the corresponding distribution function. Give an example of such a random variable. 20
- (b) Let (X, Y) be jointly distributed with pdf

$$f(x, y) = \begin{cases} 2 & \text{if } 0 < x < y < 1 \\ 0 & \text{otherwise} \end{cases}$$
 Find all the marginal and conditional density functions. 20
- (c) An urn contains 3 red and 2 white balls. A random sample of two balls is drawn with replacement. Let $X=0$ if the first ball drawn is white and $X=1$ if the first ball drawn is red. Let $Y=0$ if the second ball drawn is white and $Y=1$ if the second ball drawn is red. Find the conditional expectation $E[X/Y]$. 20
3. (a) State and prove the Borel-Cantelli lemma. 28
- (b) Verify whether the following sequences of independent random variables satisfy the WLLN: (i) $P[X_k = +k] = P[X_k = -k] = 1/2\sqrt{k}$, $P[X_k = 0] = 1 - 1/\sqrt{k}$ (ii) $P[X_k = k^\lambda] = P[X_k = -k^\lambda] = 1/2$, $\lambda > 0$ 16
- (c) Do the following functions define distribution functions of random variables? (i)

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } 0 \leq x < 1/2 \\ 1 & \text{if } x \geq 1/2 \end{cases}$$
 (ii)

$$F(x) = \begin{cases} 0 & \text{if } x < 1 \\ 1/2 & \text{if } x = 1 \\ -1/x & \text{if } x > 1 \end{cases}$$
 [?] Justify your answer in each case. 16
4. (a) Define the characteristic function of a random variable. Determine the characteristic function of a Poisson variable. Are the following functions characteristic functions? (i) $\exp[-t^4]$ (ii) $\exp[itc]$ (iii) $\phi(t) = 1$ for all t 20
- (b) State the condition when the moment sequence of a random variable defines a unique distribution. A random variable X has moments about origin given by $\mu_n' = c/(c+n)$, $n=1, 2, \dots$, 20

where $c > 0$. Show that the pdf is given by $f(x) = cx^{c-1}$, $0 < x < 1$.

- (c) Give an example where a negative binomial distribution is appropriate. Give its pmf and determine the mean and variance of the distribution. 20

SECTION 'B'

5. (a) State and prove the Cramer-Rao inequality for the variance of an unbiased estimator. State the regularity conditions needed. 30
- (b) State the properties of maximum likelihood estimators. 15
- (c) Let T be an MVB estimator of $g(\Theta)$, while S is an unbiased estimator with efficiency e_θ . If ρ_θ is the correlation between T and S , then show that $\rho_\theta = \sqrt{e_\theta}$. 15
6. (a) On the basis of a random sample of size n from a distribution with pdf

$$f_\theta(x) = \begin{cases} \theta \exp(-\theta x), & x > 0, \theta > 0 \\ 0 & \text{otherwise} \end{cases}$$
 derive the MP critical region for testing $H: \Theta = \Theta_0$ against $K: \Theta = \Theta_1 (> \Theta_0)$. Will this region be UMP against $K: \Theta > \Theta_0$? Will this region be also MP against $K: \Theta = \Theta_2 (< \Theta_0)$? 30
- (b) Let $(X_1, X_2, \dots, X_n)$ be a random sample from a distribution with pmf $f_\theta(x) = \theta^x (1-\theta)^{1-x}$, if $x = 0, 1$ ($0 < \theta < 1$). Find the UMP critical region of size α , for testing $H: \Theta = \Theta_0$ against $K: \Theta > \Theta_0$. 20
- (c) When is a family of densities said to have monotone likelihood ratio? Show that the one-parameter exponential family of densities belongs to this class of MLR densities. 10
7. (a) Derive the asymptotic distribution of the LR test criterion for testing a simple null hypothesis in the one-parameter family of densities. 30
- (b) For the SPRT with stopping bounds A and B ($A > B$) and strength (α, β) show that

$$A \leq (1-\beta)/\alpha \text{ and } B \geq \beta/(1-\alpha) \text{ for } 0 < \alpha < 1, 0 < \beta < 1.$$
 30
8. (a) What is the one-sample problem concerning the location of a probability distribution? State the classical normal theory test for this problem. Describe a non-parametric version of this test and consider both one sided and two-sided alternatives. 35
- (b) If $X_{(r)}$ and $X_{(s)}$, $r < s$, are order statistics for a random sample of size n from a continuous distribution, then show that $(X_{(r)}, X_{(s)})$ is a confidence interval for the population quantile of order p , with confidence coefficient

$$I_p(r, n-r+1) - I_p(s, n-s+1),$$
 where $I_p(v_1, v_2)$ is the incomplete beta function. 25

SECTION 'C'

9. (a) Describe the model and the analysis of variance of a complete two-way classified data with equal (> 1) number of observations per cell. State clearly what you will do if the interaction effects (i) are not significant (ii) are significant. 30
- (b) In the case of two-way classified data with one observation per cell derive the expectations of the MS components of the ANOVA table. 30
10. (a) In the multiple linear regression model with k independent variables, β_k is the estimated regression coefficient for the k^{th} independent variable β_k . Show that the statistic for testing the hypothesis $\beta_k = 0$ is provided by

$$t = \hat{\beta}_k / \sqrt{\text{Var}(\hat{\beta}_k)},$$
 where $\text{Var}(\hat{\beta}_k)$ is the estimated variance of $\hat{\beta}_k$. What is the critical region for the alternatives $\beta_k \neq 0$? 30
- (b) In the multiple linear regression model with usual normality assumptions obtain the $100(1-\alpha)\%$ confidence interval for a regression coefficient β_i which is symmetric with respect to $\hat{\beta}_i$. 15

- (c) Indicate, with the help of the ANOVA table, how you will test for the linearity of regression. 15
11. (a) If $X \sim N(\Theta, \Sigma)$ then show that the marginal distribution of each $X_i \sim N(\theta_i, \sigma_{ii})$, where $X = (X_1, X_2, \dots, X_p)'$, $\Theta = (\theta_1, \dots, \theta_p)'$ and $\Sigma = (\sigma_{ij})$. Is the converse true? 25
- (b) If $X = (X_1, X_2)' \sim N(\Theta, \Sigma)$ and X_1, X_2 are r - and s -vectors respectively with $r+s=p$, then show that $X_1 - \Sigma_{12} \Sigma_{22}^{-1} X_2$ and X_2 are independently normally distributed. The partition of Σ is

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{pmatrix}$$
 20
- (c) Let the covariance matrix of a four-dimensional normal random vector $X = (X_1, X_2, X_3, X_4)'$ be given by

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{pmatrix}$$
Find the partial correlation coefficient between $(i-1)^{\text{th}}$ and $(i+1)^{\text{th}}$ components of X when the i^{th} component is held fixed. 15
12. (a) Define Hotelling's T^2 statistic and Mahalanobis' D^2 statistic and state the relationship between the two. Describe the various uses of these statistics in test of hypotheses. 30
- (b) What are orthogonal polynomials? State how these are used in fitting regression curves and obtain the LS estimates of the coefficients in the regression equation. Indicate how the technique of ANOVA is used in deciding the appropriate degree of the polynomial to be fitted. 30