CSM – 68/17
Statistics
Paper – I

Time : 3 hours

Full Marks : 300

The figures in the right-hand margin indicate marks.

*Candidates should attempt Q. No. 1 from Section – A and Q. No. 5 from Section – B which are compulsory and any **three** of the remaining questions selecting at least **one** from each Section.*

SECTION – A

1. Answer any **three** of the following :

(a) (i) The diameter of an electric cable, say X, is assumed to be a continuous random variable with p. d. f. $f(x) = 6(x)(1 - x), 0 \le x \le 1$. Show that the above is a p. d. f. and determine a number b such that $P(X < b) = P(X > b)$. 10

(ii) X, Y have joint p. d. f. :

$f(x, y) = x e^{-x(y+1)}$ $(x \geq 0, y \geq 0)$

Find the marginal and conditional p. d. f's. 10

(b) (i) Find the characteristic function of the Gamma distribution. 10

(ii) Ten coins are thrown simultaneously. Find the probability of getting at least seven heads. 10

(c) (i) If the regression equations are : 10

$8X - 10Y + 66 = 0$

$40X - 18Y = 214$

Find the mean values of X and Y and the correlation coefficient between X and Y.

(ii) In a trivariate distribution : 10

$r_{12} = 0.7$, $r_{23} = r_{31} = 0.5$

Find $r_{23.1}$ and $R_{1.23}$

(d) (i) If $X \sim N(\mu, \Sigma)$, then show that $Y = CX \sim N(C\mu, C\Sigma C')$ for C non-singular. 10

(ii) What is the relation between Mahalanobis D^2 and Hotelling's T^2

statistics. Explain any two applications of Hotelling's T^2 statistics. 10

2. (a) If two dice are thrown, what is the probability that the sum is : 10+10 = 20

(i) Greater than 8

(ii) Neither 7 nor 11

(b) Let (X, Y) have joint p.d.f. :

$$f(x, y) = \begin{cases} \frac{1+xy}{4}, & |x|<1, \quad |y|<1 \\ 0 & \text{elsewhere} \end{cases}$$

Show that X, Y are not independent. 20

(c) $\{X_K\}$, K = 1, 2, ... is a sequence of independent random variables each taking values –1, 0, 1. Given that $P(X_u = 1) = \frac{1}{K} =$

$P(X_K = -1)$, $P(X_K = 0) = 1 - \frac{2}{K}$.

Examine, if the law of large numbers holds for this sequence. 20

3. (a) Derive the moment generating function of Poisson distribution $P(\lambda)$. Hence derive its mean and variance. 20

(b) If $X_1, X_2, \cdots\cdots, X_n$ are independent random variables, X_i having an exponential distribution with parameter θ_i, $i = 1, 2, \cdots, n$; then $Z = \min(X_1, X_2, \cdots\cdots X_n)$ has exponential distribution with parameter $\sum_{i=1}^{n} \theta_i$. 20

(c) State the multiple linear regression model with the assumptions. Explain a procedure to estimate the parameters of the model. Define the coefficient of determination R^2 for this model. 20

4. (a) Explain the principle of least squares and describe its applicatons in fitting a curve of the form $Y = a\, e^{(bX + cX^2)}$. 20

(b) Let $X \sim N(0, I_p)$. If $X' A X$ is a quadratic form of rank r in X then show that, $X' A X$ is distributed as χ^2, if A is an idempotent matrix. 20

(c) Derive the Bayesian classification rule to classify an observation into one of the two mutlivariate normal populations with equal covariance matrices. 20

SECTION – B

5. Answer any **three** of the following :

(a) (i) Let $X_1, X_2, \ldots\ldots X_n$ be a random sample from Bernoulli distributoin :

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

Obtain the sufficient statistic for θ. 10

(ii) Let $X_1, X_2, \ldots\ldots X_n$ be a random sample from the uniform distribution with p. d. f :

$$f(x, \theta) = \frac{1}{\theta}, \quad 0 < x < \infty, \theta > 0$$
$$0 \quad \text{elsewhere}$$

Obtain the maximum likelihood estimator for θ. 10

(b) (i) If $x \geq 1$ is the critical region for testing $H_0 : \theta = 2$ against the alternative $\theta = 1$, on the basis of the single observation from the population, $f(x, \theta) = \theta \exp(-\theta x), \theta \leq x < \infty$. Obtain the values of type I and type II errors. 10

(ii) Explain Wald's SPRT and describe the test procedure for binomial distribution. 10

(c) (i) Explain the role of auxiliary variables in the ratio and regression methods of estimation. Show that regression estimator is more efficient than ratio estimator for estimating the population mean. 10

(ii) Obtain the sampling variance of the mean based on systematic sample and compare the variance with that based on simple random sampling and stratified random sampling. 10

(d) (i) Explain the three basic principles and their importance in the design of experiments. 10

(ii) Discuss, in detail, the analysis of BIBD using intra-block information only. 10

6. (a) State and prove Rao-Blackwell theorem. 20

(b) Show that the maximum likelihood estimators are consistent and sufficient, if it exists. 20

(c) Let $X \sim N(\mu, \sigma^2)$. Construct the likelihood ratio test to test $H_0 : f = \mu_0$ against $H_1 : \mu \neq f_0$ when σ^2 is unknown. 20

7. (a) Explain the following non-parametric tests : 10+10 = 20

    (i) Kolmogorov-Smirnov Test (two sample)

    (ii) Run test

(b) (i) State and prove Wald's fundamental identity.

    (ii) Write a short note on OC and ASN functions. 10+10 = 20

(c) Describe the advantages of stratified random sampling with illustrations. Compare the efficiencies of the Neyman and proportional allocations with that of an unstratified random sample of the same size. 20

8. (a) Write short notes on the following : 10+10 = 20

    (i) Non-sampling error

(ii) Hansen-Hurwitz and Horvitz-Thompson estimator

(b) Describe the layout of a 2^3 experiment where all the interactions are partially confounded. 20

(c) Give the layout and analysis of a LSD with one missing value. Compare the efficiency of LSD over CRD and RBD. 20