

# 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 EIGHT questions divided in two Sections.

Candidate has to attempt FIVE questions in all.

Question Nos. 1 and 5 are compulsory and out of the remaining, any THREE are to be attempted choosing at least ONE question 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. Attempt any five sub-parts:
  - (a) Let  $P(S_1, S_2) = \exp[-\lambda_1 - \lambda_2 - \mu + \lambda_1 S_1 + \lambda_2 S_2 + \mu S_1 S_2]$ ,  $(\lambda_1, \lambda_2, \mu) > 0$  be bivariate probability generating function (PGF) of  $(X, Y)$ . Find the marginal PGF of X and Y. Find also the PGF of  $X + Y$ . Examine whether X and Y are independent. 12
  - (b) Apply a suitable inversion theorem to obtain the probability density function of the random variable whose characteristic function is given by  $\phi(t) = 1 - |t| \leq 1$  12
  - (c) State and prove Chapman-Kolmogorov equation. 12
  - (d) State the method of minimum Chi-square for estimating a parameter and show that for large sample, the minimum Chi-square equations and likelihood equations are identical, providing identical estimates. 12
  - (e) Bring out the relation between uniformly most accurate (UMA) confidence sets of confidence coefficient  $1 - \alpha$  and uniformly most powerful non-randomized tests of level  $\alpha$ . 12
  - (f) Develop the Mann-Whitney-Wilcoxon test and obtain the mean and variance of the test statistic. How is the test carried out for large samples? 12
2. (a)(i) Define expectation of a random variable (r.v.). Show that if X is a non-negative r.v. with distribution function F then
 
$$E(X) = \int_0^\infty (1 - F(x)) dx,$$
 provided the expectation exists. 8
  - (a)(ii) Define convergence in probability (P), convergence in  $r^{th}$  mean ( $L_r$ ) and convergence in distribution (D). Show that convergence in e mean implies convergence in probability and convergence in probability implies convergence in distribution. 12
  - (b)(i) State a necessary and sufficient condition for a function  $\phi(t)$  on a real line to be the characteristic function of a random variable. Prove that it is necessary 12
  - (b)(ii) Explain the concept of strong and weak laws of large numbers. State and prove Khintchin's WLLN. 12
  - (c) Explain the concept of Central Limit Theorems (CLT). Let  $\{X_n\}$  be a sequence of uniformly distributed random variables over  $(-\beta n^\lambda, \beta n^\lambda)$ ,  $\beta > 0$ ,  $\lambda > 0$ . Test whether SLLN, WLLN and CLT hold for  $\{X_n\}$ . 16
3. (a)(i) If T be a sufficient statistic for an unknown parameter  $\theta$  and U be any unbiased estimator of  $\theta$ , then prove that  $\psi(T)$  of T such that  $\psi(t) = E(U|T=t)$  is also an unbiased and has smaller variance than that of U. 15

- (a)(ii) Let  $X_1, X_2, \dots, X_n$  be a random sample from a Poisson distribution with p.m.f.  

$$f_{\theta}(x) = \begin{cases} (e^{-\theta} \theta^x) / x! & x=0,1,2,\dots \\ 0 & \text{otherwise} \end{cases}$$
where the parameter  $\theta$  is unknown ( $\theta > 0$ ). Find the Minimum Variance Unbiased Estimator of  $(e^{-\theta} \theta^x) / k!$ . 12
- (b)(i) Define a consistent estimator for  $\theta = (\theta_1, \theta_2, \dots, \theta_k)$ . Show that  $T_n$  is a consistent estimator of  $\theta$  only if  $E(T_n) \rightarrow \theta$  and  $V(T_n) \rightarrow 0$  as  $n \rightarrow \infty$ . 15
- (b)(ii) Given a random sample of size  $n$  from a Cauchy population with location parameter  $\theta$ , show that the sample median is a consistent estimator of  $\theta$ . 10
- (c) A discrete random variable  $X$  has a distribution dependent on the specification of the values of a parameter  $\theta$  as indicated in the following table. Find the MLE of  $\theta$  when  $x=2$  is an observation of  $X$  8
- |                  |     |     |     |     |     |
|------------------|-----|-----|-----|-----|-----|
| $x$              | 0   | 1   | 2   | 3   | 4   |
| $P(x), \theta=0$ | 0.1 | 0.2 | 0.3 | 0   | 0.4 |
| $P(x), \theta=1$ | 0.2 | 0.3 | 0.2 | 0.1 | 0.2 |
| $P(x), \theta=2$ | 0.4 | 0.3 | 0.2 | 0.1 | 0   |
4. (a)(i) Given a random sample of size  $n$  from a Bernoulli  $(1, \theta)$  distribution, find the UMP size- $\alpha$  test for  $H_0: \theta = \theta_0$  against  $H_1: \theta > \theta_0$ . Derive the power function  $P(\theta)$  of the test. 10
- (a)(ii) Let  $\phi(x)$  be a test function such that  

$$\phi(x) = \begin{cases} 1 & \text{if } x > k \\ \gamma & \text{if } x = k \\ 0 & \text{if } x < k \end{cases}$$
where  $0 \leq \gamma \leq 1$  and if  $X$  has density  $f(x, \theta)$  which has monotone likelihood ratio in  $x$  where  $\theta$  is real parameter, then show that the power function of the test  $\phi(x)$  is increasing in  $\theta$ . 10
- (b)(i) Let  $(X_1, X_2, \dots, X_m)$  and  $(Y_1, Y_2, \dots, Y_n)$  be independent random samples from the continuous distribution functions  $F(x)$  and  $F(x-\theta)$ , respectively, where  $F$  is unknown. It is required to test  $H_0: \theta = 0$ . Write down the median test statistic  $T$  and its asymptotic distribution under  $H_0$ . 10
- (b)(ii) Suppose  $x_1, x_2, \dots, x_n$  are i.i.d.  $N(\mu, \sigma^2)$  observations where  $\mu$  is known. Obtain the uniformly most accurate (UMA) sets for  $\sigma^2$ . 10
- (c) Let  $X_1, X_2, \dots$  be i.i.d.  $N(\theta, 1)$  random variables and suppose that we want to test  $H_0: \theta = \theta_0$  against  $H_1: \theta = \theta_1 (> \theta_0)$ . Determine the size  $n$  of the fixed sample of the most powerful test of size- $\alpha$  and power  $1-\beta$ . Determine also average sample number under  $H_0 E_{\theta_0}(N)$  of the SPRT of the same problem with strength  $(\alpha, \beta)$ . 20

### SECTION 'B'

5. Attempt any five sub-parts:
- (a) In connection with the Gauss-Markov linear models, what do you mean by Design matrix and Normal equations? Discuss the least squares principle and use it to derive the normal equations. 12
- (b) Identify, giving reasons, which of the following models are linear? Wherever possible, suggest a transformation  $Z_i = g(Y_i)$  that will result in a linear model for  $Z_i$ : (i)  $Y_i = \beta_1 + \beta_2 x_i^2 + e_i$  (ii)  $Y_i = \beta_1 + \beta_2 x_i + e_i^2$  12
- (c) Define the Horvitz-Thompson estimator of the population total based on a sample of  $n$  distinct units from a finite population of  $N$  units with given positive inclusion probabilities of first and second orders. Show that the estimator is unbiased and derive the expression for its variance. 12
- (d) Discuss the advantages of stratified sampling in comparison to unstratified sampling. 12
- (e) What is missing plot technique? Describe, in detail, the analysis of a RBD with one missing observation. How will the analysis be modified if there are more than one 12

missing observations?

- (f) Establish the parametric relations of a BIBD with parameters  $v, b, r, n, \lambda$ . Show that in a symmetric BIBD, the number of treatments common between any two blocks is  $\lambda$ . **12**
6. (a) Consider the following linear model  

$$\begin{aligned} [y_1 &= \theta_1 + \theta_4 + \theta_5 + e_1; y_2 = \theta_2 - \theta_4 + e_2; y_3 = \theta_3 + 2\theta_5 + e_3; y_4 = \theta_1 + \theta_2 + 2\theta_5 + e_4; \\ y_5 &= \theta_1 + \theta_3 + \theta_4 - e_5] \end{aligned}$$
  
 Where  $e_i$ 's are i.i.d.  $N(0, \sigma^2)$  variables (i) Show that  $\sum_{i=1}^5 l_i \theta_i$  is estimable if and only if  $l_4 = l_1 - l_2$  and  $l_5 = l_1 + l_2 - 2l_3$ . (ii) Obtain the BLUE of  $\theta_1 + \theta_2 + \theta_3$  and  $\theta_3 - 2\theta_5$  **10**
- (b)(i) Define principal components and explain their significance in statistical practice. **10**
- (b)(ii) Let  $X \sim N_p(\mu, \Sigma)$  unknown and  $\mu = (\mu_1, \mu_2, \dots, \mu_p)'$ . Suggest a test for  $H_0: \mu_1 = \mu_2 = \dots = \mu_p$  against  $H_1: \text{all } \mu \text{ s are not equal}$ . **15**
- (c) Let  $X \sim N_p(\mu, \Sigma)$ . Let  $G(q \times p)$  be a matrix of rank  $q$ . show that  $GX \sim N_p(G\mu, G\Sigma G')$  **15**
7. (a)(i) Define first and second order inclusion probabilities. Evaluate them for SRSWR and SRSWOR. **8**
- (a)(ii) Define a ratio estimator for population mean on the basis of a sample of size  $n$  selected with SRSWOR. Derive the expressions for its bias and mean square error in large sample. **12**
- (b)(i) Describe a method for selection of  $n$  units from a finite population with (1) PPSWR and (2) PPSWOR. **10**
- (b)(ii) Define Hansen-Hurwitz Estimator (HHE) for the population mean and obtain its variance. **10**
- (c) Suppose a population consists of  $N$  first stage units (fsu) each of which contains  $M$  second-stage units (ssu). An SRSWOR of size  $n$  is taken from the  $N$  fsu and then from each selected fsu  $m$  ssu are selected at random without replacement. On the basis of the  $mn$  ssu suggest an unbiased estimator for the population mean. Find its variance. How will you determine  $a$  and  $m$  for a cost function  $c = a + nc_1 + nmc_2$ , where  $c$  = total cost,  $a$  = overhead cost,  $c_1$  = cost per fsu,  $c_2$  = cost per ssu, with minimum variance. **20**
8. (a) When is a two-way ANOVA model said to be mixed? How will you formulate the different hypotheses of this model? Describe how the hypotheses are tested. **20**
- (b)(i) Distinguish between elementary contrasts and orthogonal contrasts in the context of a set of  $v$  treatments. **6**
- (b)(ii) Construct a non-randomized layout of a  $(3^3, 3)$  partially confounded factorial experiment in four replicates, confounding second order interaction effects only in each replicate. What property does this design exhibit in respect of the loss of precision of the confounded effects? **16**
- (c) Why do we need to develop split plot design? Write down the model of the design and give the estimates of the parameters involved in the model. Write down the ANOVA table and give the expressions of the estimates of standard errors for the different types of comparisons. **18**