22/1

# 2006

## STATISTICS

## Paper 1

*Time : 3 Hours ]* *[ Maximum Marks : 300*

---

### INSTRUCTIONS

*Candidates should attempt **all** the questions in Parts A, B & C. However, they have to choose only **three** questions in Part D. The number of marks carried by each question is indicated at the end of the question.*

*Answers must be written in English.*

*This paper has four parts :*

| | |
|---|---|
| **A** | 20 marks |
| **B** | 100 marks |
| **C** | 90 marks |
| **D** | 90 marks |

*Marks allotted to each question are indicated in each part.*

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

*Notations and symbols used are as usual.*

SEAL

## PART A

*4×5=20*

*Each question carries 5 marks.*

1. (a) If $\{A_n\}$, $n \geq 1$ is a sequence of events for which limit exists, then prove that

$$\lim_n P(A_n) = P(\lim_n A_n)$$

(b) If $X_1$ and $X_2$ are independent and identically distributed $B(1, p)$ random variables, then examine whether $(X_1 + 2X_2)$ is sufficient for $p$.

(c) Distinguish between simple and composite hypotheses. Illustrate.

(d) Explain the practical advantages of a sequential test procedure.

## PART B

*10×10=100*

*Each question carries 10 marks.*

**2.** (a) Stating the regularity conditions on $f(x, \theta)$, $\theta \in \Omega$, write down the Cramer – Rao lower bound for the variance of an unbiased estimator of $g(\theta)$.

(b) Let ~ $N(\theta, 1)$. Based on a random sample $X_1, \ldots, X_n$, construct an unbiased estimator of $\theta^2$. Also obtain the Cramer – Rao lower bound on variance. Is this bound attained in this case ?

**3.** (a) Let $f(x, y) = \frac{1}{4}[1 - x^3 y]$, $-1 < x, y < 1$ be the joint density function. Derive the marginal density functions of X and Y.

(b) Specify the probability density function (pdf) of a bivariate normal distribution. In the case of zero correlation show that the components are independently distributed.

**4.** Establish Holder's inequality and deduce Cauchy – Schwartz inequality as a particular case.

**5.** State Lindberg – Feller central limit theorem. Examine whether this result holds for the sequence $\{X_n\}$ of independent random variables with

$$P(X_n = -n) = \frac{1}{2\sqrt{n}} = P(X_n = n) \quad \text{and} \quad P(X_n = 0) = 1 - \frac{1}{\sqrt{n}}.$$

**6.** Given the characteristic function $\phi(t) = \exp\{-|t|\}$, show that the pdf exists and obtain the same.

**7.** Establish Lehmann – Scheffé theorem. State the significance of this result.

**8.** Outline the chi-square test for goodness of fit. Mention clearly the underlying assumptions.

**9.** Define the p-variate normal distribution. Derive its characteristic function.

**10.** Illustrate the application of the following :

(a) Hotellings' $T^2$ statistic

(b) Mahalanobis' $D^2$ statistic

**11.** Stating the necessary conditions on X and U, show that the least squares estimator of $\beta$ in the model $Y = X\beta + U$ is (a) unbiased and (b) consistent.

## PART C

*6×15=90*

*Each question carries 15 marks.*

**12.** (a) Define sufficient, minimal sufficient and complete minimal sufficient statistics.

(b) Define Fisher's information function and compute it for the one-parameter exponential family of density functions.

**13.** (a) Define a maximum likelihood estimator (MLE) and state its important properties.

(b) Obtain the MLE of $\theta$ based on random samples of $n$ observations from a uniform distribution over the interval $(\theta, \theta + 1)$.

**14.** Outline the following methods of estimation :

(a) Method of moments

(b) Method of minimum chi-square

**15.** Establish Markov inequality and hence deduce Chebychev's inequality.

**16.** (a) Derive the distribution of the sample multiple correlation when the population multiple correlation is zero in the trivariate case.

(b) Define orthogonal polynomials. How are these useful in regression analysis ?

**17.** Explain the following for a two-sample problem :

(a) Wilcoxon – Mann – Whitney test

(b) Kolmogorov – Smirnov test

[ Turn over

**PART D** *3×30=90*

*Answer any **three** of the following questions. Each question carries 30 marks.*

**18.** (a) State and prove Khinchine's weak law of large numbers (WLLN).

(b) Examine whether the WLLN holds for the sequence of random variables $\{X_n\}$ with $P(X_n = n) = \frac{1}{2n\sqrt{n}} = P(X_n = -n)$ and $P(X_n = 0) = 1 - \frac{1}{n\sqrt{n}}$.

(c) If $\{X_n\}$ is a sequence of independent random variables with

$$P(X_n = 0) = 1 - \frac{1}{n} = 1 - P(X_n = 1),$$

using Borel Cantelli lemma, prove that $X_n \to 0$ almost surely, as $n \to \infty$.

**19.** State and prove Neyman – Pearson fundamental lemma. Explain clearly the role of this result in tests of hypotheses.

**20.** Establish the Gauss – Markov theorem. What are the implications of this theorem for regression analysis ? Elaborate.

**21.** (a) State the advantages with non-parametric tests.

(b) Explain the assumptions and operating procedure of the following :

(i) Runs test

(ii) Sign test

(iii) Construction of confidence intervals for quantiles

**22.** (a) Explain the technique of Analysis of Variance. When and how is this technique useful ? Illustrate.

(b) Giving an example, state the objective of discriminant analysis. Define Fisher's linear discriminant function and explain its use.

SEAL