CSM – 68/18

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 **three** of the remaining questions, selecting at least **one** from each Section.*

SECTION – A

1. Attempt any **three** of the following sub-parts : 20×3 = 60

   (a) Let $\{X_n\}$ be a sequence of random variables. If $X_n \xrightarrow{P} X$, where X is a random variable, then show that $g'(X_n) \xrightarrow{P} g(X)$ where g is a continuous function.

(b) Find the probability density function of a distribution function of a random variable whose characteristic function is defined by $e^{-|x|}$.

(c) State and prove Gauss-Markov Theorem and explain its applications in linear estimation.

(d) What is the relation between Hotelling's T^2 statistic and Mahalanobis D^2 statistic. Also, show that Hotelling T^2 statistic is invariant under any linear transformation.

2. (a) A problem in statistics is given to 3 students A, B and C whose chances of solving it are $\frac{1}{2}$, $3/4$ and $\frac{1}{4}$ respectively. What is the probability that the problem will be solved if all of them try independently ?

(b) A random variable X has the p. d. f. given by $f(x) = 6x(1 - x), 0 \le x \le 1$. Find the mean, mode and standard deviation.

(c) State and prove Kolmogorov's Strong Law of Large Numbers. 20×3 = 60

3. (a) Show that the normal distribution is a limiting form of binomial distribution.

(b) If X and Y are independent Gamma variates with parameters μ and υ respectively, then show that $\frac{X}{X+Y}$ is a $\beta_1(\mu, \upsilon)$ variate.

(c) The equations of two regression lines obtained in a correlation analysis are as follows :

$3X + 12Y = 19, 3Y + 9X = 46$

Obtain :

(i) The value of correlation coefficient.

(ii) The mean values of X and Y. $20 \times 3 = 60$

4. (a) In a study of a random sample of 120 students, the following results are obtained :

$\bar{X}_1 = 68$ $\bar{X}_2 = 70$ $\bar{X}_3 = 74$

$s_1^2 = 100$ $s_2^2 = 25$ $s_3^2 = 81$

$r_{12} = 0.60$ $r_{13} = 0.70$ $r_{23} = 0.65$

$\left[s_i^2 = v(X_i)\right]$; where X_1, X_2 and X_3 denote percentage of marks obtained by a student in I test, II test and the final examination respectively.

(i) Obtain the least square regression equation of X_3 on X_1 and X_2.

(ii) Estimate the percentage marks of a student in the final examination, if he gets 60% and 67% in tests – I and II respectively.

(b) Let $X \sim N(\mu, \Sigma)$, then derive the distribution of the quadratic form $X' A X$ where A is a positive definite symmetric matrix.

(c) Explain linear discriminant analysis and its applications in social sciences. $20 \times 3 = 60$

SECTION – B

5. Answer any **five** sub-questions : $12 \times 5 = 60$

(a) State the Cramer-Rao lower bound for the variance of an unbiased estimator. Compute the bound for estimation of $g(\lambda) = P[X = 0]$, where X has Poisson (λ) distribution. Does there exist an MVB estimator of $g(\lambda)$? If not, state how you would determine the UMVUE.

(b) Define a Uniformly Most Powerful test. Derive it for testing $H_0 : \sigma^2 \le \sigma_0^2$ against $H_1 : \sigma^2 \ge \sigma_0^2$ for the variance of a normal distribution with mean 0. Derive an expression for the power function.

(c) Stating the hypothesis, derive the stopping bounds on the sample sum for an SPRT(B, A) applied to the parameter p of a binomial distribution.

(d) Compute the Kolmogorov-Smirnov statistic for testing the hypothesis that the following sample has come from a distribution with pdf $f(x) = 2x, 0 < x < 1$.

0.22, 0.11, 0.84, 0.64, 0.42

(e) Find an expression for the efficiency of LSD over RBD with rows as blocks.

(f) Explain the techniques of stratified and systematic sampling. Show that for estimating the mean, when there is a linear trend in the population, stratified sampling is n times more precise than systematic sampling.

6. (a) Construct the level-α UMPU test of $H_0 : \sigma^2 = \sigma_0^2$ against $H_1 : \sigma^2 \neq \sigma_0^2$ given a random sample the $N(0, \sigma^2)$ model. Obtain the cut-off value and an expression for the power function.

(b) Consider the SPRT for testing the mean of $N(\mu, \sigma_0^2)$, σ_0^2 known. Derive stopping bounds on the sample sum at the n^{th} stage. Also, obtain an expression for $h(\mu) \neq 0$ satisfying $\int\{f_1(x)/f_0(x)\}^{h(\mu)} f_\mu(x)dx = 1$. Explain how you would construct the OC and ASN curves.

(c) Describe the Likelihood Ratio test procedure and state its asymptotic properties.

25+25+10 = 60

7. (a) Explain the regression method of estimating the population mean $\bar{Y}$ when the population mean $\bar{X}$ of the auxiliary variable is unknown. Derive the variance of the estimator and conditions under which it is smaller than the estimator under SRSWOR ignoring the auxiliary variable.

(b) Under the PPS scheme, derive a sufficient condition for WOR estimator of the population total Y_{HT} to be more efficient than the WR estimator. What is your conclusion for the equal probability sampling ?

(c) Describe Warner's randomised response model giving an instance of its application.

25+25+10 = 60

8. (a) Describe a one-way random effects model and give a method of estimating the components of variance under the model.

(b) Explain the need of confounding in factorial experiments. Setup the ANOVA table for testing the main effects of a 2^4 factorial experiment carried out in a single replicate of 4 incomplete blocks confounding the interactions ABC and BCD.

(c) Describe a BIBD and give an outline of its intra-block analysis. 20+20+20 = 60