2021

STATISTICS

PAPER-I

Time Allowed — 3 Hours *Full Marks — 200*

If the questions attempted are in excess of the prescribed number, only the questions attempted first up to the prescribed number shall be valued and the remaining ones ignored.

*Answers may be given either in **English** or in **Bengali** but all answers must be in one and same language.*

Group-A

Answer *any four* questions.

1. (a) Give the classical definition of probability. Write down its limitations, if any.

(b) Define the following:

(i) Incompatible events

(ii) Independent events

(c) The probability that a person can hit a target is $^3/_5$ and the same probability for another person is $^4/_5$. If they fire together, show that the probability that the target will be hit by both of them is always more than $^2/_5$.

(d) State and prove Bayes Theorem. 6+6+10+8=30

2. (a) In a certain town, the proportions of males and females are equal. If 20% of males and 5% of females are unemployed, what is the probability that a randomly selected person is unemployed? If an unemployed person is selected at random, what is the probability that the person selected is a male?

(b) Prove that for any random variable X,

$$\mathrm{E}\,|X - C| \leq \mathrm{E}\,(X - C)^2$$

for all scalar C. When does equality hold?

(c) Suppose a coin is tossed until a head appears. If the coin is unbiased and X denotes the number of tosses required, find the median and mode of X.

(d) If $X \sim N(0, 1)$, compute $\mathrm{E}\,\{\phi(X)\}$, where $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, -\infty < x < \infty$. 7+8+8+7=30

3. (a) If $X \sim$ Uniform (0, 2), find the MGF of X and hence the variance of X.

(b) Show that for a positive random variable X, $\mathrm{E}\,(\mathrm{X}) \geq e^{\mathrm{E}(\log x)} \geq \left[\mathrm{E}\left(\frac{1}{\mathrm{X}}\right)\right]^{-1}$.

(c) State and prove Markov inequality. 9+15+6=30

4. (a) If $f(x,y)$ is a joint PDF, given by

$$f(x,y) = Ke^{-xy}y^2, x > 0, 0 < y < 3$$
$$= 0 \ ow.$$

(i) Find K.

(ii) Find the conditional PDF of X given $Y = y$.

(iii) Compute $E(e^X | Y = y)$.

(b) If $f(x,y) = \frac{1}{2}g(x,y,\rho) + \frac{1}{2}g(x,y,-\rho), -\infty < x < \infty, -\infty < y < \infty$ is a joint PDF, where $g(x,y,\rho)$ is the PDF of a $N_2(0,0,1,1,\rho)$ distribution, compute $E(X)$ and $E(XY)$. 5+6+8+11=30

5. (a) Suppose X_i are *iid* with CDF, $F(x) = \frac{x}{x+1}, 0 \le x < \infty$. If M_n is the maximum of $X_1, X_2, \ldots, X_n$, show that $\frac{M_n}{n} \xrightarrow{D} \frac{1}{Y}$, where $Y \sim \text{Exp}(1)$.

(b) Suppose $Y_1, \ldots, Y_n$ are *iid* uniform $(\theta, \theta + 1)$,

(i) show that

(a) $\bar{Y}$ is biased for θ,

(b) $MSE(\bar{Y}) \to \frac{1}{4}$ as $n \to \infty$.

(ii) Find a sufficient statistic for θ.

(c) Describe p value in the context of hypothesis testing. 12+12+6=30

6. (a) Suppose X_i are *iid* uniform $(0, \theta)$, $i = 1, \ldots, n$ and $Y_n = \max\{X_1, \ldots, X_n\}$.

(i) Compute the confidence coefficient for the set $[Y_n, \infty]$.

(ii) Find a pivotal quantity for θ in terms of Y_n.

(iii) If n is such that $b^n - a^n = 0{\cdot}96$, find the confidence coefficient of the confidence interval $\left[\frac{Y_n}{b}, \frac{Y_n}{a}\right]$.

(b) Suppose $X_i, i = 1,2,3,4$ are *iid* $\text{Exp}\left(\text{mean} = \frac{1}{\lambda}\right)$. Corresponding to the data (2,3,3,4), find the maximum likelihood estimate of λ when it is known that λ is either $\frac{1}{4}$ or $\frac{1}{3}$. (4+8+10)+8=30

Group-B

Answer *any two* questions.

7. (a) Consider a CRD with t treatments and replication number r_j, for the jth treatment, $j = 1, \ldots t$ with $\sum_{j=1}^{t} r_j = n$ (fixed). Show that the average variance of all estimated elementary treatment contrasts is minimised when $r_j = \frac{n}{t}$ for all $j = 1, \ldots t$.

(b) Consider a one-way ANOVA fixed effects model with K factor levels and n_K observations for the Kth factor level. If x_{ij} is the jth observation corresponding to the ith level, $j = 1, \ldots, n_i; i = 1,2, \ldots, K$, write down the expression of the F statistic F_1. If we define a new set of observations as $y_{ij} = Ax_{ij} + B$ and compute the F statistic F_2, establish a relation between F_1 and F_2. What is the implication of your finding?

(c) Describe the basic principles of design of experiments. 15+10+15=40

8. (a) Distinguish between sampling and non-sampling errors.

(b) From a population of N units, n units are drawn by SRSWR, of which only n_1, responded. Out of the remaining $n_2 = n - n_1$ non-responding units, information was later collected on u units, chosen using SRSWR. Show that $\hat{\mu} = \frac{n_1 \bar{y}_{n_1} + n_2 \bar{y}_u}{n}$ is an unbiased estimator of the population mean, where $\bar{y}_{n_1}(\bar{y}_u)$ is the sample mean based on responding initially (later) units. Also derive Var($\hat{\mu}$).

(c) Define a ratio estimator. Obtain its exact bias and approximate MSE. Also derive those for estimating the population mean. Define the regression estimator of population mean and compare it with that obtained for ratio method in terms of precision. 10+15+15=40

9. (a) In the context of 2^4 factorial experiment, describe Yate's procedure of forming treatment contrasts.

(b) Construct a non-randomized layout of $(2^5, 2^2)$ experiment, confounding ACD and BD, where A, B, C, D, E are the factors.

(c) Consider a $(2^4, 2^2)$ experiment with 4 factors A, B, C, D. One of the blocks is given by

a	b	cd	$abcd$

(i) Construct the other blocks.

(ii) Identify the confounded effects. 10+15+15=40