Total No. of Questions : 5]

SEAT No. :

P572

[Total No. of Pages : 3

[5841]-101

M.Sc. (Statistics)

ST - 11 : BASICS OF REAL ANALYSIS AND CALCULUS

(2019 Pattern) (Semester - I)

Time : 3 Hours] *[Max. Marks : 70*

Instructions to the candidates:

1) ***All questions are compulsory.***
2) ***Figures to the right indicate full marks.***
3) ***Use of statistical tables and scientific calculator is allowed.***
4) ***Symbols and abbreviations have their usual meaning.***

Q1) Answer the following questions : **[5 × 2 = 10]**

a) Define Matric space with an illustration.

b) Check whether the following set is countable set. Justify your answer.

A = {2, 4, 6, ...}

c) Define Cauchy sequence with an illustration.

d) Show that if $A \subset B$ and $B \subset C$ the $(A \cup B) \subset C$.

e) Define Monotone sequence with an illustration.

Q2) Answer the following questions (Any 3) : **[3 × 5 = 15]**

a) Prove that $\sqrt{7}$ is an irrational number.

b) Examine whether $d(x-y) = (x-y)^2$ for x and y in R is a metric or not.

c) State and prove Heine - Borel theorem.

d) State and prove the Archimedean principle of real numbers.

P.T.O.

Q3) Answer the following question (Any 3) : **[3 × 5 = 15]**

a) Examine whether the following series $\sum a_n$ converges or diverges.

i) $\sum_{n=1}^{\infty}(-1)^n$,

ii) $\sum_{n=0}^{\infty}\frac{1}{n!}$

b) State and prove cauchy criteria for the convergence of series of real numbers.

c) Prove or disprove the set of rational numbers is dense in R.

d) Prove that if p > 0 then $\sqrt[n]{p}=1$.

Q4) Answer the following questions (Any 3): **[3 × 5 = 15]**

a) Define : An open set, closed set in metric space. Show that finite union of closed set is closed.

b) Show that a converge sequence in a metric space is bounded. Is the converse is true? Justify your answer.

c) Examine whether the following sequence is converge. Obtain the limit in case of it converges.

i) $\{i^n\}$

ii) $\sqrt[n]{n}$

iii) $\sqrt{n+1}-\sqrt{n}$

d) State and prove finite intersection property of compact set.

Q5) Answer the following questions (Any 1) : **[15]**

I. a) Define the following terms with suitable example. **[8]**

i) Common refinement of two partitions.

ii) Reimann sum associated with a partition.

iii) Improper integral of first kind.

iv) Local maximum of a function.

b) State and Prove Bolzano-Weierstrass theorem in R. **[7]**

II. a) Consider the function $f(x) = x^2$ on [0,1]. Let Pn be the partition $\left\{0, \frac{1}{n}, \frac{2}{n}, \ldots\ldots\ldots\ldots \frac{n-1}{n}, 1\right\}$ of [0,1], n $\geq$1. Compute L(Pn,f) and U(Pn,f) for each n. Deduce that f is Riemann integrable and hence find the integral. **[9]**

b) State & Prove fundamental theorem of calculus. **[6]**

❄ ❄ ❄

Total No. of Questions : 5]

SEAT No. :

P573

[Total No. of Pages : 3

[5841]-102

M.Sc. (Statistics)

ST - 12 : LINEAR ALGEBRA AND NUMERICAL METHODS

(2019 Pattern) (Semester - I)

Time : 3 Hours] ***[Max. Marks : 70***

Instructions to the candidates:

1) ***All questions are compulsory.***
2) ***Figures to the right indicate full marks.***
3) ***Use of statistical tables and scientific calculator is allowed.***
4) ***Symbols and abbreviations have their usual meaning.***

Q1) Attempt each of the following : **[5 × 2 = 10]**

a) Define subspace of matrix

b) Define Inner product of two vectors

c) Define Eigen value of a matrix

d) Define Row space of matrix

e) Define Null space of matrix

Q2) Attempt any Three of the following : **[3 × 5 = 15]**

a) Solve the following system of linear equations by using Gauss Jordan method.

$$6X - Y + Z = 13$$

$$X + Y + Z = 9$$

$$10X + Y - Z = 19$$

b) Define characteristic root and characteristic vector and prove that λ^k is an eigen value of A^k, where k positive integer.

c) Verify $A = PDP^T$ Spectral Decomposition of following matrix.

$$A = \begin{pmatrix} -3 & 5 \\ 4 & -2 \end{pmatrix}$$

P.T.O.

d) Find inverse of square matrix (partition method) following matrix.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 5 & 6 & 1 \end{bmatrix}$$

Q3) Attempt any Three of the following : **[3 × 5 = 15]**

a) Define Moore Penrose generalize inverse of matrix. Show that Moore Penrose generalize inverse is unique.

b) Consider a Quadratic Form

$$3x_1^2 + 2x_2^2+3x_3^2+2x_1x_3$$

i) Write the above Quadratic from in matrix form

ii) Find Eigen value of the matrix of Quadratic Form. Also examine nature of the Quadratic form.

c) Prove that for a real symmetric matrix eigen values are real.

d) Reduce the given matrix to upper traingular form and find its determinant.

$$P = \begin{bmatrix} 3 & 8 & 7 \\ 1 & 2 & 4 \\ -1 & 3 & 2 \end{bmatrix}$$

Q4) Attempt any Three of the following : **[3 × 5 = 15]**

a) State and prove a necessary and sufficient condition for a quadratic form to be non-negativedefinite.

b) Find the characteristics root and characteristics vector of given matrix.

$$B = \begin{bmatrix} 1 & 1 & 3 \\ 1 & 5 & 1 \\ 3 & 1 & 1 \end{bmatrix}$$

Q5) Attempt any One of the following : **[1 × 15 = 15]**

a) i) Prove : The characteristics root of a real symmetric matrix are real. **[8]**

ii) Find inverse and MPG inverse of following matrix.

$$A = \begin{bmatrix} 1 & 0 & 2 \\ 0 & -1 & 1 \\ 1 & 2 & 0 \end{bmatrix}$$

Verify the AGA = A holds for the given matrix. **[7]**

b) i) Explain Helmert matrix and Find the Helmert matrix of order 4 (i.e. n = 4) **[8]**

ii) Describe Gram-Smidt orthogonalization process. **[7]**

❄ ❄ ❄

Total No. of Questions : 5]

SEAT No. :

P574

[Total No. of Pages : 4

[5841]-103
M.Sc. (Statistics)
ST-13 : PROBABILITY DISTRIBUTIONS
(2019 Pattern) (Semester - I)

Time : 3 Hours] ***[Max. Marks : 70***

Instructions to the candidates:

1) ***All the questions are compulsory.***
2) ***Figures to the right indicate full marks.***
3) ***Use of scientific calculators and statistical tables is allowed.***
4) ***Symbols and abbreviations have their usual meaning.***

Q1) a) Define the following terms : **[2 × 3 = 6]**

i) Symmetric probability distribution around 'a'

ii) Continuity theorem of probability

iii) Bivariate poisson distribution

b) Attempt the following : **[2 × 2 = 4]**

i) Let x be an random variable defined on $(\Omega, \mathbb{F}, \mathbb{P})$ Let $g(.)$ be any Borel-measurable function on $\mathbb{R}$. Then show that g(x) is also a random variable.

ii) A fair coin is tossed three times. Let x be the number of heads in the three tosses and let y be the absolute difference between no. of heads and number of tails.

Find the joint p.m.f. of (x.y).

Q2) Attempt any 3 questions out of 4 **[3 × 5 = 15]**

a) Define the cumulative distribution function, F(x) of a random variable X based on probability space $(\Omega, \mathbb{F}, \mathbb{P})$. Show that F($x$) is right continuous function of x.

b) Let X and Y be any two random variables defined on probability space $(\Omega, \mathbb{F}, \mathbb{P})$ then prove that :

$E(x) = E_y(E(x|y))$

c) Let (x,y) be jointly distributed random variables with p.d.f.

$$f_{x_1,x_2}(x_1,x_2) = 2e^{-(x_1+x_2)} \quad : \quad 0 < x_1 < x_2 < \infty$$
$$= 0 \quad : \quad \text{Elsewhere}$$

Then derive the p.d.f of y = x_1+x_2

d) A random sample of size 4 is drawn from a population with U (0,5) distribution. Find the comutative distribution function of 3rd order statistic $X_{(3)}$. Find the probability $P(X_{(3)}>2)$

Q3) Attempt any 3 questions out of 4 : **[3 × 5 = 15]**

a) Let $\mathbb{F}_1$ and $\mathbb{F}_2$ are any two fields defined over the sample space Ω then prove that $(\mathbb{F}_1 \cap \mathbb{F}_2)$ is also a field.

b) Let x be a discrete random variable with p.g.f.

$P_x(s) = \frac{8}{5}(2+3S^2)$. Find the p.m.f. of x.

c) State the p.d.f. of non-central chi-square distribution derive its mean and variance.

d) A fair coin is tossed twice. Let x(w) be the no. of heads in outcome w. Let $\mathbb{F}$ be the σ-field given as :

$\mathbb{F} = \{\phi, \{HH\}, \{HT, TH, TT\}, \Omega\}$

Examine whether x(w) is a random variable.

Q4) Attempt any 3 questions out of 4 : **[3 × 5 = 15]**

a) Let X be any non-negative random variable with p.d.f.,

$$f(x) = \begin{cases} e^{-x} & ; x > 0 \\ 0 & , elsewhere \end{cases}$$

Show that the limiting distribution of $(X_{(n)} - \log n)$ is $e^{-e^{-x}}$ where $X_{(n)}$ is largest order statistics based on random sample of size n taken from $f(x)$.

b) Let X be a random variable with p.d.f. **[4]**

$$f(x) = \frac{\beta}{x^{\beta+1}} \ ; x \geq 1, \beta > 0$$

Check whether the raw moments μ_r, r = 1, 2, 3 .... exist for $\beta = 3$.

c) Describe sign test for testing the hypothesis about p^{th} population quantile. Obtain the probability distribution of the test statistic under the null hypothesis.

d) Consider the bivariate negative binomial distribution with p.m.f.

$$p\,(X = x, Y = y) = \frac{(x+y+k-1)!}{x!\,y!\,(k-1)!} p_1^x\, p_2^y (1-p_1-p_2)^k$$

Where x, y = 0, 1, 2, ...; and $k \geq 1$ is an integer p_1, $p_2 \in (0,1)$. Obtain the marginal p.m.f. of $u = x + y$.

Q5) Attempt any one out of 2 questions. **[1 × 15 = 15]**

a) i) Let x be a random variable with c.d.f.,

$$F(x) = \begin{cases} 0 & , \quad x < 0 \\ \frac{1}{2} & , \quad x = 0 \\ \frac{1}{2} + \frac{x}{2} & , \quad 0 < x < 1 \\ 1 & , \quad x \geq 1 \end{cases}$$

Check whether F(x) is discrete or continuous? **[10]**

ii) Define Random variable. If x be a random variable defined on measurable space $(\Omega, \mathbb{F})$ then prove that $y = ax + b$ is also a random variable where a & b are any two real numbers. **[5]**

b) i) Suppose a set function P $(A_k) = \int_{A_k} \bar{e}^{x}\, dx$

where $A_k = \left\{ x \mid 2 - \frac{1}{k} \le x \le 3 \right\}$

verify whether $\lim_{k\to\infty} \mathrm{P}(\mathrm{A}_k) = \mathrm{P}\left[\lim_{k\to\infty} \mathrm{A}_k\right]$ **[10]**

ii) Let (X, Y) be bivariate random variable with joint p.d.f.

$$f(x, y) = \begin{cases} 2 & \text{if } 0 < x < y < 1 \\ 0 & ,\text{otherwise} \end{cases}$$

Find conditional p.d.f. of x given y. Also find E (X | Y = y). **[5]**

❄ ❄ ❄

Total No. of Questions : 5]

SEAT No. :

P575

[Total No. of Pages : 3

[5841]-104
M.Sc.
STATISTICS
ST - 14 Sampling Theory (4 Credits)
(2019 Pattern) (Semester - I)

Time : 3 Hours] *[Max. Marks : 70*

Instructions to the candidates:

1) ***All the questions are compulsory.***
2) ***Figures to the right indicate full marks.***
3) ***Use of statistical tables and scientific calculator is allowed.***
4) ***Symbols and abbreviations have their usual meaning.***

Q1) **Attempt the following questions.** **[2 Each]**

a) Explain the method of systematic sampling when population size is multiple of sample size.

b) Define the stratified Random Sampling and give one real life situation where stratified random sampling is used.

c) Give the two real life situations where the two-stage sampling is applicable.

d) State the assumptions used in ratio and regression method of estimation.

e) Define the mean square error and derive the relationship between variance and mean squared error.

Q2) **Attempt any three of the following.** **[5 Each]**

a) Determine the size of sample to be drawn from the population using SRRSWOR method for prespecified relative error in estimation of population mean using sample mean.

b) Let (Xi, Yi) be the values with respect the two variables X and Y associated with the units having labels I, i=1,2, ...n. Let $\overline{x_n}, \overline{y_n}$, be the sample means and $S_{XY} = \frac{1}{N-1}\sum_{i=1}^{N}\left(Xi - \overline{X_n}\right)\left(Yi - \overline{Y_n}\right)$, then under simple random sampling without replacement $Cov\left(\overline{x_n}, \overline{y_n}\right) = \frac{N-n}{nN}S_{XY}$.

P.T.O.

c) Prove that in case of inverse sampling estimator $\frac{(m-1)}{n-1}$ is an unbiased estimator for the population proportion P, where m is predetermined number of units possessing given attribute in the sample of size n.

d) Explain any one of the method of selecting a sample of size n by SRSWR with Probability Proportional to Size (PPS) and state its limitations.

Q3) Attempt any three of the following. **[5 Each]**

a) What is the difference between the cluster sampling and stratified random sampling? Show that weighed average of stratum sample mean with weight as size of the stratum is an unbiased estimator of the population mean.

b) For stratified sampling, explain Optimal and Proportional allocation. Also show that the optimum allocation is better than proportional allocation

c) Explain balanced systematic sampling with the illustration.

d) Show that systematic sampling is more precise than simple random sampling without replacement, if the variance within systematic sample is larger than the population variance as a whole.

Q4) Attempt any three of the following. **[5 Each]**

a) Show that ratio estimator of population mean is more precise than the regression estimator of the population mean.

b) Show that for the Probability Proportional to Size without Replacement (PPSWOR) design, Horvitz Thomson estimator is unbiased for the population mean and obtain its variance.

c) Find variance of the estimator of the population mean in case of two stage sampling having equal number of first stage units.

d) State an unbiased estimator of population means in case of post stratification. Hence find variance of the estimator.

Q5) Attempt the following questions.

a) i) Derive the Yates corrected estimator of the population mean. **[8]**

ii) Distinguish between sampling and non-sampling errors. Discuss in detail the mathematical model for measurement of observational errors. **[7]**

b) i) Define regression estimator of population mean in double sampling. Show that it is biased estimator. Also find the mean squared error of this estimator. **[8]**

ii) With two strata, sampler would like to have $n1 = n2$ for the administrative convenience, instead of using the values given by Neyman allocation. If V and V_{opt} denote the variances given by $n1 = n2$ and Neyman allocations, respectively, show that the $\frac{V - V_{opt}}{V_{opt}} = \left(\frac{r-1}{r+1}\right)^2$, where $r = \frac{n1}{n2}$ given by Neyman's allocation. Assume N1 and N2 large. **[7]**

❍❍❍❍

Total No. of Questions : 5]

SEAT No. :

P576

[Total No. of Pages : 2

[5841]-201
M.Sc.
STATISTICS
ST - 21 : Probability Theory
(2019 Pattern) (Semester - II)

Time : 3 Hours] *[Max. Marks : 70*

Instructions to the candidates:

1) ***All questions are compulsory.***
2) ***Figures to the right indicate full marks.***
3) ***Use of statistical tables and scientific calculator is allowed.***
4) ***Symbols and abbreviations have their usual meaning.***

Q1) **Attempt each of the the following:**

a) Define each of the following **[1 each]**

i) Sigma field

ii) Increasing sequence of sets

b) Explain **[1 each]**

i) Lebesgue stielje's measure.

ii) Vector valued random variable.

c) State sigma field generated by class $A = \{A, B\}$ defined on sample space. **[2]**

d) Explain arbitrary random variable with suitable example. **[2]**

e) Define convex function. Check the convexity of the function X^2. **[2]**

Q2) **Attempt any three of the followings:** **[3 × 5 = 15]**

a) $\{An\}_{n \geq 1}$ is a sequence of real sets:

If $An = \left[a, b + \frac{1}{n}\right]$, Find $\lim\limits_{n \to \infty}$ An

b) State and prove decomposition theorem of distribution function.

c) State and prove Holder's inequality.

d) State and prove continuity theorem of probability measure.

P.T.O.

Q3) Attempt any three of the following: **[3 × 5 = 15]**

a) Define convergence of sequence of random variable. Explain with suitable example.

b) Compare convergence in distribution and almost sure convergence.

c) State and prove convergence theorem for expectations.

d) If $X_n \xrightarrow{p} \mathbf{X}$ then show that $aX_n + b \xrightarrow{p} aX + b$.

Q4) Attempt any three of the following: **[3 × 5 = 15]**

a) Compare convergence in distribution and convergence in rth mean.

b) If $\{X_n\}_{n\geq 1}$ is sequence of random variables them prove that

$$X_n \xrightarrow{p} O \text{ if and only if } E\left(\frac{|X_n|}{1+1X_n}\right) \xrightarrow{p} O \, as \, n \to \infty$$

c) State and prove Slutkey's theorem.

d) If $|x_n| y, y$ is integrable and $X_n \xrightarrow{a-s} X$ then prove that $E(X_n) \to E(X)$.

Q5) Attempt any one of the following: **[3 × 3 = 9]**

A) a) Prove the following

i) $P\left(AB^C \cup BA^C\right) = P(A) + P(B) - 2P(AB)$

ii) If $A \subset B$ then $P(A) < P(B)$

iii) Monotone increasing sequence of sets is convergent.

b) State and prove Jensen's inequality. **[6]**

Hence or otherwise comment on

$E(\ln(X))$ and $\ln(E(X))$

B) a) Define field induced by random variable X. Explain it with suitable example. **[5]**

b) State Lindberg Feller form. Give the suitable example to explain it. **[5]**

c) Show that Borel function of a random variable X is random variable. **[5]**

❍❍❍❍

Total No. of Questions : 5]

SEAT No. :

P577

[Total No. of Pages : 3

[5841]-202
M.A./M.Sc.
STATISTICS
ST22 : Regression Analysis
(2019 Pattern) (Semester - II) (4 Credits)

Time : 3 Hours] *[Max. Marks : 70*

Instructions to the candidates:

1) ***All questions are compulsory.***
2) ***Figures to the right side indicate full marks.***
3) ***Use of statistical tables and scientific calculator is allowed.***
4) ***Symbols and abbreviations have their usual meaning.***

Q1) **Attempt all questions.** **[5 × 2 = 10]**

a) Explain the no intercept term model.

b) State and prove any two properties of direct regression estimators.

c) Show that solution of normal equation actually minimizes the residual sum of squares.

d) Describe Box and Cox transformation in regression model.

e) Define link function for the generalized linear model.

Q2) **Attempt any 3 questions out of 4 questions.** **[3 × 5 = 15]**

a) Define Polynomial regression model. Estimate the parameters of orthogonal polynomial regression model and give its variance.

b) State and prove Gauss-Markoff theorem.

c) Discuss the method of maximum likelihood estimation for estimating parameters of simple linear regression model.

d) Obtain 100 $(1 - \alpha)\%$ confidence interval on regressor parameter β_1 of the simple linear regression.

P.T.O.

Q3) **Attempt any 3 questions out of 4 questions.** **[3 × 5 = 15]**

a) Explain the concept of multicollinearity in multiple regression model. Also, explain sources of multicollinearity.

b) State all the assumptions made for random errors in any particulars model. Explain the graphical diagnostic techniques used to verify these assumptions.

c) Discuss the method of principle component regression for dealing with the problem of multicollinearity.

d) For the generalized linear model, explain the following link function.

i) Binomial function

ii) Inverse binomial

iii) Inverse gamma

Q4) **Attempt any 3 questions out of 4 questions.** **[3 × 5 = 15]**

a) For nonlinear regression model $y = \theta_1 e^{\theta_2 x} + \varepsilon$, obtain least square normal equation. Now, linearize the model and obtain least square estimates of parameters and hence compare the two methods.

b) Discuss Wald test in logistic regression model.

c) Consider the problem of regression of y on two predictor variables X_1 and X_2. It is observed that instead of using multiple linear regression model, a second order response surface can be fitted to get more precise results. Discuss the use of orthogonal polynomials to simplify the analysis of this model.

d) Define the following terms:

i) Mallows Cp-statistic

ii) PRESS residuals

iii) Leverage points

iv) Hat matrix

v) R^2 and adjusted R^2

Q5) **Attempt any 1 question out of 2 questions.** **[1 × 15 = 15]**

a) i) Explain the logistic regression model with single explanatory variable. For a sample of size *n*. Hence, obtain the maximum likelihood equation. **[8]**

ii) Describe the generalized linear model. Define link function and obtain it for poisson distribution. **[7]**

b) i) Describe the problem of autocorrelation. **[2]**

ii) For the simple linear regression model, with first order autoregressive errors, Discuss the Durbin-Watson test to detect the presence of autocorrelation in errors. **[7]**

iii) For the nonlinear regression model $y = \theta_1 e^{\theta_2 + \theta_3 x} + \varepsilon$, obtain least squares normal equation. Now, linearize the model and obtain least square estimates of parameters. Compare the two methods. **[6]**

❍❍❍❍

Total No. of Questions : 5]

SEAT No. :

P578

[Total No. of Pages : 3

[5841]-203
M.Sc.
STATISTICS
ST 23 : Statistical Inference - I
(2019 Pattern) (Semester - II) (4 Credits)

Time : 3 Hours] *[Max. Marks : 70*

Instructions to the candidates:

1) ***All questions are compulsory.***
2) ***Figures to the right indicate full marks.***
3) ***Use of statistical tables and scientific calculator is allowed.***
4) ***Symbols and abbreviations have their usual meaning.***

Q1) **Attempt all questions:** **[5 × 2 = 10]**

a) Distinguish between estimator and estimate.

b) Define the term estimability. Prove or disprove: If X_1, X_2 is a random sample from *Bernoulli* $(p), p \in (0,1)$ then $p(1-p)$ is estimable.

c) State Rao-Blackwell Lehmann-Scheffe' theorem.

d) State Chapman-Robin bounds.

e) Define the terms: Prior distribution and posterior distribution.

Q2) **Attempt any THREE of the following.** **[3 × 5 = 15]**

a) State the factorization criterion for a sufficient statistic. Using it show that the order statistic is a sufficient statistic for any parametric family of continuous distributions.

b) Let $X_1, X_2, \ldots, X_n$ be *i.i.d.* random variables with *pdf* $f(x;\theta), \theta \in \Theta \subseteq \mathbb{R}$. Let $T = T(X_1, X_2, \ldots, X_n)$ be a statistic. Show that $I_T(\theta) \le I_{(X_1, X_2, \ldots, X_n)}(\theta)$ with the equality holds if and only if T is sufficient for θ.

P.T.O.

c) Define sufficiency. Let $X_1, X_2, ..., X_n$ be *i.i.d.* geometric random variables with parameter θ. Find the sufficient statistic for θ.

d) Define complete statistic. Let $X_1, X_2, ..., X_n$ be *i.i.d.* $U(\theta, \theta+1)$. Show that $\left(X_{(1)}, X_{(n)}\right)$ is sufficient for θ but not complete for θ.

Q3) **Attempt any Three of the following.** **[3 × 5 = 15]**

a) State and prove Rao Blackwell theorem.

b) Let $X_1, X_2, ..., X_n$ be independent uniformly distributed over $(-i\theta, i\theta), \theta > 0, i = 1, 2, ..., n$. Obtain minimal sufficient statistic for θ.

c) Prove that for distribution belonging to one parameter exponential family, UMP test can be obtained for one sided alternative.

d) Let $X_1, X_2, ..., X_n$ be a random sample from Gamma$\left(1, \frac{1}{\lambda}\right)$. To estimate λ, let the priori *pdf* on λ be $\pi(\lambda) = e^{-\lambda}$ if $\lambda > 0$ and the loss function be squared error. Find the Bayes estimator for λ.

Q4) **Attempt any Three of the following:** **[3 × 5 = 15]**

a) Let *X* have the following possible distributions :

X	0	2	4	6
$P_1(x)$	0.2	0.3	0.15	0.35
$P_2(x)$	0.4	0.1	0.05	0.45

For testing $H_0 : X \sim P_1(x)$ against $H_1 : X \sim P_2(x)$ we use the tests

$$\Phi_1(x) = \begin{cases} 1, & if\ x = 4 \\ 0.3, & if\ x = 2 \\ 0, & otherwise \end{cases} \quad \text{and} \quad \Phi_2(x) = \begin{cases} 1, & if\ x = 0, 4 \\ 0.1, & if\ x = 2 \\ 0, & otherwise \end{cases}$$

Obtain the sizes of Φ_1 and Φ_2 and the powers of the tests Φ_1 and Φ_2. Can you compare these tests?

b) Define ancillary statistic with illustration. Prove that, any ancillary statistic is independent of complete sufficient statistic.

c) What is pivotal quantity? Explain its use in constructing of confidence intervals.

d) Define monotone likelihood ratio property of a distribution. Show that the binomial distribution satisfies MLR property.

Q5) Attempt any one of the following: **[1 × 15 = 15]**

a) i) Write note on shortest expected length confidence interval (SELCI). **[8]**

ii) Define uniformly most accurate confidence interval. Let $X_1, X_2, ..., X_n$ be a random sample from $N(\theta, \sigma^2), \sigma^2$ is known. Obtain level α uniformly most accurate lower bound on θ. **[7]**

b) i) State and prove Neyman-Pearson Lemma. Let $X_1, X_2, ..., X_n$ be a random sample from $f(x;\theta) = \frac{\theta}{x^2}$, where $0 < \theta \le x < \infty$. Find an MP test of its size for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1 (\theta_0 < \theta_1)$. **[9]**

ii) Let

$$f(x, y, \lambda, p) = \binom{x}{y} p^y (1-p)^{(x-y)} \frac{e^{-\lambda}\lambda^x}{x!}, \quad y = 0,1,2,...,x; x = 0,1,2,...,$$

$\lambda > 0$ and $0 < p < 1$ be joint *pmf* of (X, Y). Show that it belongs to two parameter exponential family and obtain minimal sufficient statistic $(T_1, T_2)'$ for $(p, \lambda)'$. **[6]**

❍❍❍❍

Total No. of Questions : 5]

SEAT No. :

P579

[Total No. of Pages : 4

[5841]-204

M.Sc. (Statistics)

ST 24 : Multivariate Analysis

(2019 Pattern) (Semester - II)

Time : 3 Hours] *[Max. Marks : 70*

Instructions to the candidates:

1) *All questions are compulsory.*
2) *Figures to the right side indicate full marks.*
3) *Use of statistical tables and scientific calculator is allowed.*
4) *Symbols and abbreviations have their usual meaning.*

Q1) **Attempt all questions.** **[5 × 2 = 10]**

a) Let $\underline{X}$ have variance - convariance matrix $\sum = \begin{bmatrix} 25 & -2 & 4 \\ -2 & 4 & 1 \\ 4 & 1 & 9 \end{bmatrix}$ Find the correlation between X_1 and $\frac{X_2}{2}+\frac{X_3}{2}$.

b) Distinguish between t-test and Hotelling T^2 statistic.

c) Why Wishart distribution is called generalization of χ_n^2 distribution. Justify your answer.

d) Define the following

i) Factor loadings

ii) Factor scores

e) If $\underline{X} \sim N_2(\underline{0},\Sigma)$ where $\sum = \begin{bmatrix} 1 & 0.7 \\ & 1 \end{bmatrix}$ then find $P(X_1 < X_2)$.

P.T.O.

Q2) Attempt any three of the following. **[3 × 5 = 15]**

a) Let the random vector $\underline{X} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} \sim N_4\left(\underline{\mu}, \Sigma\right)$ with $\underline{\mu} = \begin{bmatrix} 3 \\ 2 \\ -2 \\ 0 \end{bmatrix}$ and

$$\Sigma = \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}. \text{ Let } A = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 1 & -2 & 0 \\ 1 & 1 & 1 & -3 \end{bmatrix}, \text{ find}$$

i) $E\left(A\underline{X}\right)$

ii) $V\left(A\underline{X}\right)$

b) Write a note on cluster analysis. Describe a method of forming clusters from given observations by using a distance function.

c) Explain the term 'Discriminant Analysis' with an illustration.

d) Let $\underline{X} \sim N_p\left(\underline{\mu}, \Sigma\right)$ and $\underline{X}$ is partition as $\begin{bmatrix} \underline{X}^{(1)} \\ \underline{X}^{(2)} \end{bmatrix}$, then obtain the marginal distributions of $\underline{X}^{(1)}$ and $\underline{X}^{(2)}$.

Q3) Attempt any three of the following. **[3 × 5 = 15]**

a) Let $\underline{X} \sim N_4\left(\underline{\mu}, \Sigma\right)$, where $\underline{\mu} = (1,2,1,2)'$ and $\Sigma = 4I$. Derive the joint distribution $Y_1 = 2X_1 + X_2 + 2X_3$ and $Y_2 = X_2 + 2X_3 + X_4$

b) Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from $N_p\left(\underline{\mu}, \Sigma\right)$ where Σ is unknown covariance matrix. Derive test statistic to test $H_0 : \mu_1 = \mu_2 = \ldots = \mu_p$.

c) Suppose the radom vector X of three components X_1, X_2 & X_3 having variance-convariance matrix $\sum = \begin{bmatrix} 13 & -4 & 2 \\ -4 & 13 & -2 \\ 2 & -2 & 10 \end{bmatrix}$. Find the first two population principal components. How much population variance explained by first two PC out of total variation.

d) Consider the following function

$$\Phi_{(x,y)}(t_1,t_2) = exp\left[i(3t_1 + t_2) - \left(\frac{3}{2}t_1^2 + t_1t_2 + t_2^2\right)\right]; t_1, t_2 \in R$$

Identify the corresponding distribution with parameters.

Q4) Attempt any Three of the following. **[3 × 5 = 15]**

a) If $\underline{X} \sim N_p(\underline{\mu}, \Sigma)$ and $\underline{X}$ is partition as $\begin{bmatrix} \underline{X}^{(1)} \\ \underline{X}^{(2)} \end{bmatrix}$. Find the conditional distribution of $\underline{X}^{(1)}$ given $\underline{X}^{(2)} = \underline{x}^{(2)}$.

b) Define non-hierarchical cluster analysis. What is a feature of non hierarchical cluster analysis. Explain k-means clustering method in brief.

c) Show that for a sample from a multivariate normal population, sample mean $\overline{X}$ and the corrected sum of squares and sum of product (CSSSP) matrix S are independent. State the major results that you use in the proof.

d) Let $\underline{X}_1, \underline{X}_2, ..., \underline{X}_n$ be random sample of size n from p variate normal distribution with mean vector $\underline{\mu}$ and variance covariance matrix $\sum$. Derive the maximum likelihood estimators of the parameters $\underline{\mu}$ and $\sum$.

Q5) Attempt any one of the following. **[1 × 15 = 15]**

A) a) Let $\underline{X} \sim N_3\left(\underline{\mu}, \Sigma\right)$ with $\underline{\mu} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ and $\Sigma = \begin{bmatrix} 2 & 1 & 3 \\ & 2 & 3 \\ & & 9 \end{bmatrix}$, then find the distribution of : **[8]**

i) $\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ ii) $\begin{bmatrix} X_1 - X_3 \\ X_2 + X_3 \end{bmatrix}$

iii) $\begin{bmatrix} & X_1 & \\ X_1 & + & X_2 \\ X_1 & + & X_3 \end{bmatrix}$ iv) $\begin{bmatrix} X_1 & - & X_2 \\ X_1 & + & 2X_2 \\ X_1 & - & 3X_3 \end{bmatrix}$

b) Derive the expression to obtain first canonical correlation and first pair of canonical variables. **[7]**

B) a) If $D \sim W_p\left(n, \Sigma\right)$, then show that,

i) $\dfrac{|D|}{|\Sigma|}$ is distributed as the product of p independent χ^2 variates with $n\left(n-1\right)....n-\left(p-1\right)d.f.$ respectively.

ii) $\dfrac{\sigma pp}{dpp} \sim \chi^2_{n-(p-1)}$, where d^{pp} and σ^{pp} are the last elements of D^{-1} and $\sum^{-1}$ respectively. **[8]**

b) Analyse the following data by using one Way MANOVA and draw the conclusion : **[7]**

Treatment Observations

i) $\begin{bmatrix} 9 \\ 3 \end{bmatrix} \begin{bmatrix} 6 \\ 2 \end{bmatrix} \begin{bmatrix} 9 \\ 7 \end{bmatrix}$ ii) $\begin{bmatrix} 0 \\ 4 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \end{bmatrix}$

iii) $\begin{bmatrix} 3 \\ 8 \end{bmatrix} \begin{bmatrix} 1 \\ 9 \end{bmatrix} \begin{bmatrix} 2 \\ 7 \end{bmatrix}$

❍❍❍❍

Total No. of Questions : 5]

SEAT No. :

P580

[Total No. of Pages : 2

[5841]-301

M.Sc.

STATISTICS

Applied Stochastic Processes (4 Credits)

(2019 Pattern) (Semester - III) (ST-31)

Time : 3 Hours] ***[Max. Marks : 70***

Instructions to the candidates:

1) ***All questions are compulsory:***
2) ***Figures to the right indicate full marks.***
3) ***Use of statistical tables and scientific calculator is allowed.***
4) ***Symbols and abbreviations have their usual meaning.***

Q1) Attempt all questions: **[5×2=10]**

a) Define Stochastic Process. With suitable example.

b) Define The terms: Index Set, State Space with example Markov sequence.

c) Define Poisson Process and write postulates of Poisson Process.

d) Define Weiner process and its Properties.

e) Give the example of each : Non Markov chain and Non homogenous Markov chain.

Q2) Attempt any Three question out of Four questions: **[3×5=15]**

a) Show that stationary distribution is unique.

b) Construct a random walk with absorbing barrier 0 and 4. Find either the states are persistent or transient and check the ergodicity.

c) Describe the Galton - Watson branching chain.

d) Consider the markov Chain on {0, 1, 2} having t.p.m. is the Markov chain is irreducible.

$$P = \begin{bmatrix} 1 & 0 & 0 \\ 1/2 & 1/2 & 0 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}$$

Q3) Attempt any Three question out of Four questions: **[3×5=15]**

a) Write short note on Ergodic Markov chain.

b) Describe the Gambler ruin problem. Find the probability that player A gets ultimately ruined. What is the expected duration of the game.

c) Renewal process $\{N(t), t \geq 0\}$ with inter-arrival time distribution F. Note that, the event $\{N(t) \geq n\}$ then prove renewal equation.

d) N(t) is Poisson Process, derive its mean and variance.

Q4) Attempt any Three question out of Four questions: **[3×5=15]**

a) Explain Random walk markov chain. Show that random walk is persistent if P = ½, otherwise it is transient.

b) Derive Yule Furry Process.

c) State and Prove Chapmans-Kolmogorov equation.

d) Derive the relation between Poisson Process and Binomial Distribution.

Q5) Attempt any one question out of Two questions: **[1×15=15]**

a) i) Prove that $P_n(s) = P_{n-1}[P(s)]$
$= P[(P_{n-1})(s)]$ also prove mean of Branching Process. **[9]**

ii) Prove Kolmogorov backward and forward equations of birth and death process. **[6]**

b) i) Find the expected duration of game in the gambler ruin problem. **[7]**

ii) Let $\{X_n\ n = 0, 1, 2, \text{----------}\}$ be Markov chain with state space S = {0, 1, 2} and one step TPM is . **[8]**

$$P = \begin{bmatrix} 3/4 & 1/4 & 0 \\ 1/4 & 1/2 & 1/4 \\ 0 & 3/4 & 1/4 \end{bmatrix}$$

The Initial distribution $P(X_0 = 0) = P(X_0 = 1) = 1/4$, $P(X_0 = 2) = 2/4$

Find

1) $P[X_1 = 1/ X_0 = 2]$
2) $P[X_2 = 2\ X_1 = 1/X_0 = 2]$
3) $P[X_2 = 2\ X_1 = 1\ X_0 = 2]$
4) $P[X_2 = 1]$

✠ ✠ ✠

Total No. of Questions :5]

SEAT No. :

P581

[Total No. of Pages :6

[5841] - 302
M.Sc.
STATISTICS
ST-32 (A) : Bayesian Inference
(2019 Pattern) (Semester - III) (4 Credits)

Time : 3 Hours] ***[Max. Marks :70***

Instructions to the candidates:

1) ***All questions are compulsory.***
2) ***Figures to the right indicate full marks.***
3) ***Use of statistical tables and scientific calculator is allowed.***
4) ***Symbols and abbreviations have their usual meaning.***

Q1) Attempt all questions. **[5×2=10]**

a) Explain the term subjective prior.

b) Represent the squared error loss function, absolute error loss function and zero-one (all-or-nothing) loss function graphically.

c) Define improper prior with an illustration.

d) Write the procedure to obtain Bayes estimator.

e) What is credible interval? Is it unique?

Q2) Attempt any THREE of the following: **[3×5=15]**

a) Let X be a continuous random variable with pdf $f(x|\theta)=\frac{3\theta^3}{x^4}$, $x>\theta$. Let the prior pdf of θ be given by $\pi(\theta)=e^{-(\theta-1)}$, $\theta>1$. Compute the posterior distribution of θ.

b) Define squared error loss function and obtain the Bayes estimator under squared error loss function.

P.T.O.

c) Explain the term maximum aposterior (MAP) estimator. Let X be a random variable having geometric distribution with parameter p. The prior distribution of p is given by $\pi(p)=\begin{cases} 2p & \text{if } 0 \le p \le 1 \\ 0 & \text{otherwise} \end{cases}$. Find the MAP estimate of p.

d) Let X and Y be jointly normal and $X \sim N(0,1)$, $Y \sim N(1,4)$ and $\rho\ (X,Y) = 12$. Find a 95% credible interval for X, given $Y = 2$ is observed.

Q3) Attempt any THREE of the following: **[3×5=15]**

a) Write note on EM-Algorithm.

b) Define the Jeffrey's prior. Let $X \sim$ *Binomial* (n,p), find the Jeffrey's prior of p.

c) Suppose $\underline{X} = (X_1, X_2,...,X_{10})$ is a random sample of size 10 from $N(\mu,4)$ with $\bar{X} = 2.1$. Suppose we assume a normal prior $N(100,4)$ for μ. Find the Bayesian estimator of μ under quadratic error loss.

d) Explain Bayes rule. A man is known to speak the truth two out of three times. He throws a die and reports that the number obtained is four. Find the probability that the number obtained is actually a four.

Q4) Attempt any THREE of the following: **[3×5=15]**

a) Define the highest posterior density credible interval (HPDCI) for a real valuedparameter θ. Assuming that the posterior distribution of θ to be symmetricand unimodal, obtain HPDCI for θ.

b) If X follows *Poisson* distribution with parameter λ and the prior distribution of λ is *Gamma* with parameters α and β. Find posterior distribution of λ based on $Y = \sum_{i=1}^{n} X_i$ which is sufficient statistics for λ, where $X_1,X_2,...,X_n$ is random sample of size n from *Poisson* distribution.

c) Write short note on informative prior and non-informative prior.

d) i) Let $X \sim N(\theta, 4)$ and assume that we are using squared error loss. Let $\hat{\theta} = X$. Find the expected loss (risk) for this estimator.

ii) Let $X \sim B(n, \theta)$ and prior distribution of θ is $\beta_1(a,b)$. Find the posterior distribution of θ .Is given prior conjugate or not?

Q5) Attempt any ONE of the following: **[1×15=15]**

a) i) Suppose that the signal $X \sim N(0, \sigma_x^2)$ is transmitted over a communication channel. Assume that the received signal is given by $Y = X + W$, where $W \sim N(0, \sigma_w^2)$ is independent of X. Find the maximum likelihood (ML) and maximum aposterior (MAP) estimate of X given $Y = y$. **[8]**

ii) Write note on MH algorithm. **[7]**

b) i) Explain the term Bayes factor and Bayesian information criterion (BIC) with illustration. **[9]**

ii) Explain the terms: Decision function, Loss function and Bayes risk. **[6]**

Total No. of Questions :5]

P581

[5841] - 302

M.Sc.

STATISTICS

ST-32 (B) : Statistical Quality Control

(2019 Pattern) (Semester - III) (4 Credits)

Time : 3 Hours] *[Max. Marks :70*

Instructions to the candidates:

1) ***All questions are compulsory.***
2) ***Figures to the right indicate full marks.***
3) ***Use of statistical tables and scientific calculator is allowed.***
4) ***Symbols and abbreviations have their usual meaning.***

Q1) Attempt all questions. **[5×2=10]**

a) State the relationship between capability index C_p and performance index C_{pk}.

b) Find sample size required for p-chart, so that there will be atleast 95% samples have one or more defectives. Give that process fraction defective is 0.05.

c) Explain the working of single sampling plan with curtailed inspection.

d) Describe Nelson control chart for low defect counts.

e) State four western Electric rule for the process control.

Q2) Attempt any 3 questions out of 4 questions. **[3×5=15]**

a) The following is $\bar{X}$ and S chart based on $n = 4$ and 3-sigma control limits:

$\bar{X}$ -chart	S-chart
UCL = 710	UCL = 18.08
CL = 700	CL = 7.979
LCL = 690	LCL = 0

i) Estimate parameters μ and σ .

ii) If the specifications are at 705 ± 15 and process output is normally distributed, estimate the fraction non-confirming.

iii) For the $\overline{X}$ control chart, find the probability of type-I error, assuming σ is constant. (Take $C_4 = 0.9213$ and $n = 4$).

b) Write a short note on DMAIC procedure.

c) Explain the concept of OC curve with reference to the control chart. Also, obtain the OC for the $\overline{X}$ -chart.

d) Obtain the Average Run Length (*ARL*) for:

i) Shewart control chart

ii) CUSUM chart.

Q3) Attempt any 3 questions out of 4 questions. **[3×5=15]**

a) Discuss EWMA control chart for monitoring process mean and variance. Explain how it is better than Shewhart $\overline{X} - \mathrm{R}$ control chart.

b) Explain the construction and working of Confirming Run Length (CRL) chart for process fraction defective.

c) Write a note on synthetic control chart and find the expression for ARL (0).

d) What is the importance of OC curve in the selection sampling plans? Describe the impact of the sample size and acceptance number of the OC curve. What is the disadvantage of having an acceptance number as zero?

Q4) Attempt any 3 questions out of 4 questions. **[3×5=15]**

a) Explain *V*-mask method of implementing CUSUM chart for mean.

b) Obtain the relation between number of defectives and capability index C_p.

c) Explain the constructions of control chart for residuals after fitting first order autocorrelated model.

d) Describe of MIL STD sampling plan.

Q5) Attempt any 1 question out of 2 questions. **[1×15=15]**

a) i) Define the capability index C_p and performance index C_{PK}. Obtain $(1-\alpha)\%$ confidence interval for the both indices. **[8]**

ii) Explain the equivalence between control chart and testing of hypothesis. **[7]**

b) Using CUSUM chart, check whether the following process is under control or not, if the target value is 175, $k = 2$, $h = 4.77$, $\hat{\sigma} = 20$? the observations are as follows. 160, 186, 190, 250, 158, 195, 135, 285, 215, 150. **[15]**

Total No. of Questions : 5]

SEAT No. :

P582

[Total No. of Pages : 3

[5841]-303

M.Sc.

STATISTICS

ST 33 : Design and Analysis of Experiments (4 credits)

(2019 Pattern) (Semester - III)

Time : 3 Hours] *[Max. Marks : 70*

Instructions to the candidates:

1) *All questions are compulsory:*
2) *Figures to the right indicate full marks.*
3) *Use of statistical tables and scientific calculator is allowed.*
4) *Symbols and abbreviations have their usual meaning.*

Q1) Attempt all questions: **[5×2=10]**

a) Sate and prove Fisher's inequality.

b) Define the terms:

i) Ridge system

ii) Orthogonal array

c) state test statistic for testing the hypothesis of equality of all treatment effects in two-way classification model.

d) Write the possible models those can be used to analyze 3^3 factorial design.

e) Define the term quality according to prof. G. Taguchi.

Q2) Attempt any Three question out of Four questions: **[3×5=15]**

a) For two-way classification model with interaction give the model and assumption. Also, obtain estimates of parameters in the model.

b) Show that for symmetric BIBD any two blocks have treatments common.

c) Write the linear model for one way classification model with single covariate.Describe the test procedure for testing the hypothesis of no covariate in this model.

d) Explain Tuckey's mehtod for comparing pairs of treatment means.

P.T.O.

Q3) Attempt any Three question out of Four questions: **[3×5=15]**

a) What is resolution of two level fractional factorial designs? Discuss on resolution III with illustration.

b) What are 3-level factorial designs? Represent the runs of 3^2 and 3^3 factorial design graphically. Write ANOVA for 3^3 factorial design.

c) Confound the 2^5 factorial design in four blocks each of size 8, where the effects to be confounded are ADE and BCE.

d) Explain the concept of minimum aberration of design.

Q4) Attempt any Three question out of Four questions: **[3×5=15]**

a) What is response surface methodology? Discuss about any two statistical designs that can be used to fit the response surface models.

b) State the transformation of factorial variables required to obtain the canonical form for the second order response surface model. Discuss how this canonical analysis can be used to get information about the nature of fitted response surface.

c) What is robust design? Discuss the role of inner and outer arrays in designing robust experiments.

d) Describe the concept behind signal to noise (S/N) ratio. Hence discuss its use in analyzing the matrix experiments.

Q5) Attempt any one question out of Two questions: **[1×15=15]**

a) i) For BIBD with parameters v, r, b, k and λ, prove that $bk = rv$ and $\lambda(v - 1) = r(k - 1)$

ii) Obtain half fraction of 2^3 factorial design with highest possible design.

iii) Consider the following response surface model fitted for predicting response y for given predictor variables X_1 and X_2.

$$\hat{y} = 70 - 16x_1 + 11x_2 - 9x_1^2 - 6x_2^2 - 2x_1x_2$$

Obtain the stationary point and response at stationary point.

[5+5+5]

b) i) An engineer is interested in the effects of cutting speed (A), tool geometry (B), cutting angle (C) on the life (in hours) of a machine. Two levels of each factor are chosen and the replicates are given in the following table.

Treatment Combination	Replicate		
	I	II	III
'1'	22	31	25
a	32	43	29
b	35	34	50
ab	55	47	46
c	44	45	38
ac	40	37	36
bc	60	50	54
abc	39	41	47

Analyze the above data.

ii) What is random effect model? Give an unbiased estimator of variance component in one way classification model assuming treatment effects random.

[10+5]

✠ ✠ ✠

Total No. of Questions : 5]

SEAT No. :

P583

[Total No. of Pages : 2

[5841]-304
Second Year M.Sc.
STATISTICS
ST 34 : Machine Learning
(2019 Pattern) (Semester - III) (4 Credits)

Time : 3 Hours] *[Max. Marks : 70*

Instructions to the candidates:

1) *All questions are compulsory:*
2) *Figures to the right indicate full marks.*
3) *Use of statistical tables and scientific calculator is allowed.*
4) *Symbols and abbreviations have their usual meaning.*

Q1) Attempt all questions: **[5×2=10]**

a) What is hard margin in support vector machine?

b) Define supervised machine learning with example.

c) Why the KNN Algorithm is known as lazy Lerner?

d) Define precision and recall.

e) What is classification?

Q2) Attempt any 3 questions out of 4 questions: **[3×5=15]**

a) What are advantages and disadvantages of machine learning?

b) What is kernel? And why would you use the kernel trick in SVM?

c) Explain Neural Network?

d) What is a confusion matrix explain all the term use in it.

Q3) Attempt any 3 questions out of 4 questions: **[3×5=15]**

a) What are the application of machine Learning?

b) What is Gini impurity an information gain in a decision tree?

c) What is Machine learning and explain reinforcement learning?

d) How does the KNN algorithm make the predictions on the unseen dataset?

P.T.O.

Q4) Attempt any 3 questions out of 4 questions: **[3×5=15]**

a) What are the advantages and disadvantages of backpropagation?

b) Briefly explain the properites of Gini Impurity.

c) Explain the CART algorithm for decision trees.

d) What do you mean by Bagging.

Q5) Attempt any 1 question out of 2 questions: **[1×15=15]**

a) i) List down the advantages and disadvantages of the random forest algorithm. **[8]**

ii) What do you mean by clustering? **[7]**

b) i) What is Hierarchical clustering algorithm? **[8]**

ii) Compare K - Nearest Neighbors (KNN) and Support Vector Machine. (SVM) **[7]**

✠ ✠ ✠

Total No. of Questions : 5]

SEAT No. :

P584

[Total No. of Pages : 2

[5841]-401
M.Sc.
STATISTICS
ST 41 : Asymptotic Inference
(2019 Pattern) (Semester - IV) (4 Credits)

Time : 3 Hours] *[Max. Marks : 70*

Instructions to the candidates:

1) ***All questions are complusory.***
2) ***Figures to the right indicate full marks.***
3) ***Use of statistical tables and scientific calculator is allowed.***
4) ***Symbols and abbreviations have their usual meaning.***

Q1) Answer the following. **[5×2=10]**

a) Define CAN estimator. Give one example.

b) State any four properties of MLE.

c) Define BAN estimator.

d) Define Marginal consistency.

e) Describe Wald test.

Q2) Attempt any 3 questions. **[3×5=15]**

a) Discuss the method of moments of obtaining consistent estimator for real valued parameter θ. Using the method of moments show that r^{th} sample raw moment is consistent for r^{th} population raw moment.

b) Let $X_1, X_2 ...., X_n$ be random sample of size n from pdf

$$f(x,\theta) = \theta x^{\theta-1} \qquad 0 < x < 1; \theta > 1$$
$$= 0 \qquad 0.w.$$

check whether it belongs to one parameter exponential family. Obtain MLE for θ. Also obtain asymptotic distribution of MLE.

c) State and prove invariance property of CAN estimator in one parameter setup.

d) Describe MP test with one example.

Q3) Attempt any 3 questions. **[3×5=15]**

a) Let $X_1, X_2 ...., X_n$ be random sample from U (0, θ). Verify whether $X_{(1)}$ is consistent estimator for θ.

P.T.O.

b) Describe Newton Raphson method to obtain MLE with one eample.

c) Let $X_1, X_2, ..., X_n$ be random sample from $f(x) = e^{-(x-\theta)}$ $x \geq \theta$. If $T_1 = X_{(1)}$ and $T_2 = \bar{X} - 1$ be two consistent estimator for θ. Which is more efficient?

d) Let $X_1, X_2, ..., X_n$ be random sample from Cauchy (μ, λ). Obtain asymptotic distribution of CAN estimator of (μ, λ).

Q4) Attempt any 3 questions. **[3×5=15]**

a) Let $X_1, X_2, ..., X_n$ be random sample from $P(\lambda)$. What is the distribution of $\frac{\bar{x}}{s^2}$?.

b) Let $X_1, X_2, ..., X_n$ be random sample from Exponential (θ_1) and Let $Y_1, Y_2, ..., Y_n$ be random sample from Exponential (θ_2). Suppose X and Y are independent. Find consistent estimator for $p(X < Y)$.

c) Let $X_1, X_2, ..., X_n$ be random sample from $N(\theta, 1)$. Obtain asymptotic distribution of LRT test Statistic for testing $H_0 : \theta = \theta_0$ Vs $H_1 : \theta \neq \theta_0$.

d) Let $X_1, X_2, ..., X_n$ be random sample from continuous pdf $f(x, \theta)$ then prove that $X_{(r)}$ is consistent for $\xi_{(p)}$ where r = [np].

Q5) Attempt any one of the following.

a) i) Explain method of scoring with example. **[7]**

ii) For the distribution belongs to one parameter exponential family show that MLE $\hat{\theta}$ is CAN for θ with asymptotic variance $\frac{1}{n\,I(\theta)}$. **[8]**

b) i) What is the variance stabilizing transformation? Based on random sample of size n from Exponential (mean (θ)), obtain 100 $(1-\alpha)$% confidence interval for θ based on variance stabilizing transformation. **[7]**

ii) State Crammer regularity conditions in one parameter setup. Give an example which satisfied regularity conditions. Justify your answer. **[8]**

Total No. of Questions : 5]

SEAT No. :

P585

[Total No. of Pages : 4

[5841]-402
M.Sc.
STATISTICS
Econometric & Time Series (4 Credits)
(2019 Pattern) (Semester - IV) (ST - 42 (A))

Time : 3 Hours] *[Max. Marks : 70*

Instructions to the candidates:

1) *All questions are compulsory.*
2) *Figures to the right indicate full marks.*
3) *Use of statistical tables and scientific calculator is allowed.*
4) *Symbols and abbreviations have their usual meaning.*

Q1) Attempt the following. **[5×2=10]**

a) Define endogenous and exogenous variables.

b) Let $\{X_t\}$ be iid N(0, 1) and define

$$\{X_t\} = Z_t \qquad \text{if t is even}$$

$$= \left(Z^2_{t-1}\right)/\sqrt{2} \qquad \text{if t is odd}$$

show that $\{X_t\}$ is WN(0, 1) but not iid(0, 1) noise.

c) Distinguish between strict and weak stationarity of time series.

d) Define SARIMA (p, d, q) × (P, D, Q) model with illustration.

e) Define the terms:

i) Forecast mean square error

ii) Optimum forecast function

Q2) Attempt any three questions of the following: **[3×5=15]**

a) Explain the concept of moving averages. Also mention its merits and demerits.

b) Find Auto Coveriance Function (ACVF) of MA (3) Process.

c) Define general linear process. Express ARIMA (1, 1, 1) process in the form of difference equation using the parameters ϕ & θ. Show that (1, 0, 1) is random walk.

d) Derive the partial autocorrelation function (PACF) of a MA(1) process and its two important properties.

P.T.O.

Q3) Attempt any three questions of the following: **[3×5=15]**

a) Describe Augmented Dickey-Fuller (ADF) test in AR(p) model.

b) Find Auto correlation function of AR (2) Model.

c) Identify the following models as ARMA (p, q) models & determine whether they are causal and/or invertible. $Y_t + 1.9Y_{t-1} - 0.88Y_{t-2} = Z_t + 0.2Z_{t-1} + 0.7Z_{t-2}$.

d) Show that the filter with coefficinet $[a_{-2}, a_{-1}, a_0, a_1, a_2] = \frac{1}{9}[-1, 4, 3, 4, -1]$ Passes third degree polynomial and eliminates seasonal components with period 3.

Q4) Attempt any three questions of the following: **[3×5=15]**

a) Discuss Durbin-Levinson algorithm.

b) Show that ARMA (1,1) process has unique solution when $|\phi| < 1$.

c) Describe Unit root test.

d) Describe Indirect least squares (ILS) method of consistent estimation of parameters.

Q5) Attempt any one of the following: **[1×15=15]**

a) i) Describe any two model selection criteria that are used in ARIMA modeling. **[8]**

ii) Discuss the maximum likelihood estimation of the parameters of the model $Y_t = \mu + \phi_1 Y_{t-1} + \phi_1 Y_{t-2} + Z_t,\ Z_t \sim WN(0, \sigma^2)$ **[7]**

b) i) For a causal AR(1) process $X_t = \phi X_{t-1} + Z_t$. Show that $P_n X_{n+1} = \phi X_n$ Where symbols have their usual meaning. Also obtain FMSE. **[8]**

ii) Suppose a certain hypothetical time series, two data points are observed as $X_2 = 2$, $X_1 = 4$. For the ARMA(1, 1) model, $X_{t,} - 0.1X_{t-1} = Z_t + 0.3Z_{t-1}, Z_t \sim WN(0,1)$. Forecast X_3 using the best linear predictor. **[7]**

Total No. of Questions : 5]

P585

[5841]-402
M.Sc.
STATISTICS
Operation Research (4 Credits)
(2019 Pattern) (Semester - IV) (ST - 42 (B))

Time : 3 Hours] *[Max. Marks : 70*

Instructions to the candidates:

1) *All questions are compulsory.*
2) *Figures to right indicate full marks.*
3) *Use of statistical tables and scientific calculator is allowed.*
4) *Symbols and abbreviations have their usual meaning.*

Q1) Attempt each of the following. **[5×2=10]**

a) Define the following term of linear programming problem (LPP):
   i) Basic feasible solution
   ii) Infeasible solution
b) Give any two application of dual simplex method.
c) Define dynamic programming problem.
d) Define the term simulation give any two application of it
e) What do you mean by degeneracy in a transportation problem? How to resolved it.

Q2) Attempt any three of the following. **[3×5=15]**

a) Prove that a collection of all feasible solution (if exist) of LPP constitute a convex set.
b) Prove the condition alternative optimal solution to exist for the LPP Max $Z = c'x$ subject to condition $AX \le b,\ x \ge 0$
c) Define an artificial variable. Discuss charnes Big M method of Solving LPP.
d) Solve the following problem by two phase method
   Maximize $Z = 2x_1 + 5x_2$
   Subject to constraints

$$-2x_1 + x_2 \le 0$$
$$x_1 + 3x_2 \le 14$$
$$x_1 + x_2 \le 8$$
$$x_1 \ge 0,\ x_2 \ge 0$$

Q3) Attempt any three of the following. **[3×5=15]**

a) State and prove weak law of duality.

b) State and prove the fundamental theorem of duality.

c) Define pure integer programming problem (IPP). Derive Gomary's cut for it.

d) Define a transhipment problem. Discuss method of solving transshipment problem.

Q4) Attempt any three of the following. **[3×5=15]**

a) Define a transportation problem (T.P). Prove that the number of basic variable T.P are at most M+ n – 1.

b) Define quadratic programming problem (QPP). Explain Beale's method of solving.

c) Explain the Hungarian method of solving an Assignment problem (A.P).

d) Explain the following terms of net work models.

i) Merge event

ii) Predecessar activity

iii) Dummy activity

Q5) Attempt any one of the following.

a) i) State and prove reduction theorem of Assignment problem (A.P). **[7]**

ii) Sole the following QPP by using Wolfe's method. **[8]**

Maximize $Z = 4x_1^2 + 5x_2^2 - 10x_1x_2 - 6x_1 - x_2$

Subject to constraint

$3x_1 + 11x_2 \le 16$

$4x_1 + 5x_2 \le 5$

$x_1 \ge 0,\ x_2 \ge 0$

b) i) Obtain a necessary and sufficient condition for existence of a feasible solution to a TP. **[6]**

ii) Sole the following pure integer programming problem by using Branch and Bound method. **[9]**

Maximize $Z = 3x_1 + 10x_2$

Subject to constraint

$x_1 + 5x_2 \le 12$

$x_1 \le 3$

$x_1 \ge 0,\ x_2 \ge 0$

Total No. of Questions : 5]

SEAT No. :

P586

[Total No. of Pages : 6

[5841]-403
M.Sc.
STATISTICS
ST-43 (A) : Survival Analysis
(2019 Pattern) (Semester - IV) (4 Credits)

Time : 3 Hours] ***[Max. Marks : 70***

Instructions to the candidates:

1) *All the questions are compulsory.*
2) *Figures to right indicate full marks.*
3) *Use of statistical tables and scientific calculator is allowed.*
4) *Symbols and abbreviations have their usual meaning.*

Q1) Attempt the following questions. **[2 each]**

a) Show that the empirical distribution function is an unbiased estimator of survival function. Also check the consistency of the same estimator.

b) Obtain Hazard rate for the Makeham family of life distribution.

c) Suppose 20 ceramic capacitors from exponential distribution are put on life test. In order to reduce the test time, the test is terminated after 12 ceramic capacitors fails. Following is lifetime in years of ceramic capacitors.

3.4, 3.7, 4.1, 4.8, 5.2, 6.2, 7.3, 8.5, 9.2, 10.1, 11.2, 13.

Find the m. l. e of parameter of the distribution.

d) If $T \rightarrow \text{Exp}(\lambda = 2)$, then find mean residual life function of T.

e) Define the following

i) Kernel and symmetric kernel

ii) U-statistic

P.T.O.

Q2) Attempt any three of the following questions. **[5 each]**

a) State and prove characteristic property of IFRA distribution.

b) Prove the implications IFRA $\Rightarrow NBU$ and IFR $\Rightarrow$ DMRL.

c) Suppose T follows exponential distribution with parameter λ then proof that scale TTT transform of random variable T is $\Psi_f(t) =, 0 \le t \le 1$.

d) Obtain Hazard rate for the Linear Hazard rate family and Pareto family and classify the family based on the Hazard rate.

Q3) Attempt any three of the following questions. **[5 each]**

a) Let $T_1, T_2, ...T_n$ be the random sample of size n from the exponential distribution with the parameter λ. In order to reduce the time of the experimentation experimenter has decided to terminate the experiment as soon as m (<n) failures occurs. Find 100 $(1-\alpha)$% confidence interval for the parameter λ.

b) If T is continuous non-negative random variable having distribution function F(t) and cumulative Hazard rate R(t) then show that distribution of R(t) is standard exponential.

c) If $T_1, T_2, ..., T_n$ are independent; $Ti \Rightarrow Exp(\lambda_i)$ for i = 1, 2, ... n and $T = \min \{T_1, T_2, ...T_n\}$, then show that $T \rightarrow Exp\left(\sum_{i=1}^{n} \lambda_i\right)$.

d) Explain the procedure to obtain the maximum likelihood estimators of parameters of Gamma distribution for complete data.

Q4) Attempt any 3 of the following questions. **[5 each]**

a) Explain the procedure to obtain the maximum likelihood estimators of parameters of weibull distribution for type II censored data.

b) Define the term is censoring. Hence discuss type I and type II censoring with an illustration.

c) Derive Green-woods formula for variance of acturial estimator of the survival function.

d) Explain redistribution to right algorithm to obtain estimator of the survival function with the help of follwoing data.

9, 13, 13+, 18, 23, 28+, 31, 34, 45+, 48, 161+

Q5) Attempt any one of the following questions.

a) i) Explain two graphical methods to test the exponentiality of the data. **[5]**

ii) Explain the Deshpande's test for testing exponentiality against IFRA class of life distribution. **[10]**

b) i) Explain the Gehan's test for testing whether two samples come from the population having same distribution functions in presence of right random censored data. **[10]**

ii) Explain the term covariates with illustration? Also explain model formulation in covariate analysis. **[5]**

Total No. of Questions : 5]

P586

[5841]-403
M.Sc.
STATISTICS
ST - 43 (B) : Categorical Data Analysis
(2019 Pattern) (Semester - IV) (4 Credits)

Time : 3 Hours] *[Max. Marks : 70*

Instructions to the candidates:

1) *All questions are compulsory.*
2) *Figures to the right indicate full marks.*
3) *Use of statistical tables and scientific calculator is allowed.*
4) *Symbols and abbreviations have their usual meaning.*

Q1) Attempt all of the following questions: **[5×2=10]**

a) Provide the general layout of a 2×2 contingency table and define joint and marginal probability.

b) Explain 'Base line Category Logits' for nominal response variables.

c) Define Relative risk for a 2×2 contingencytable.

d) Define 'Deviance' for a Generalized Linear Model (GLM).

e) State any two properties of Yule's Q-coefficient.

Q2) Attempt any three questions from the following questions: **[3×5=15]**

a) i) Discuss the need for Poisson regression model and its use.

ii) State the assumptions of Poisson Regression Model.

iii) Provide the tests for determining Statistical significance of regression coefficients.

b) Explain any four real life applications of binary logistic model.

c) Explain Negative Binomial Regression.

d) Explain Conditional Logistic Regression.

Q3) Attempt any three questions from the following questions: **[3×5=15]**

a) Explain 1×J Contingency table and explain the inference procedure for testing the dependency between two categorical variables with 1×J contingency table.

b) Explain multi nomial logistic regression.

c) Expalin any two association measures based on 2×2 contingency table.

d) Explain Fisher exact test for 2×2 contigency table.

Q4) Attempt any three questions from the following questions. **[3×5=15]**

a) Explain Binary Logistic regression in detail.

b) What is categorical data? Explain the types of categorical data.

c) Explain Log-Linear Analysis for analyzing dependency in contingency table.

d) Explain the components of GLM Modle in details.

Q5) Attempt any one question from the following questions. **[1×15=15]**

a) A study used logistic regression to determine characteristics associated with Y = whether a cancer patient achieved remission (1 = yes). The most important explanatory variable was a labeling index (LI) that measures proliferative activity of cells after a patient receives an injection of tritiated thymidine. It represents the percentage of cells that are "labeled." Software report for a logistic regression model using LI to predict $\pi = P(Y = 1)$.

		Standard	Likelihood	Ratio	
Parameter	Estimate	Error	95% conf.	Limits	Chi - square
Intercept	– 3.7771	1.3786	–6.9946	–1.4097	7.51
li	0.1449	0.0593	0.0425	0.2846	5.96

		LR Statistic	
Source	DF	Chi - square	Pr > chisq
li	1	8.30	0.0040

obs	li	remiss	n	pi_hat	lower	upper
1	8	0	2	0.06797	0.01121	0.31925
2	10	0	2	0.08879	0.01809	0.34010

. . . .

i) Show how software obtained $\pi^\wedge = 0.068$ when LI = 8.

ii) Show that $\pi^\wedge = 0.50$ when LI = 26.0.

iii) Show that the rate of change in $\pi^\wedge$ is 0.009 when LI = 8 and is 0.036 when LI=26.

iv) The lower quartile and upper quartile for LI are 14 and 28. Show that $\pi^\wedge$ increases by 0.42, from 0.15 to 0.57, between those values.

v) When LI increases by 1, show the estimated odds of remission multiply by 1.16.

b) Explain model for matched pairs also explain the types of matched pair.

Total No. of Questions : 5]

SEAT No. :

P587

[Total No. of Pages : 4

[5841]-404
M.Sc.
STATISTICS
ST-44(A) : Computer Intensive Statistical Methods
(2019 Pattern) (Semester - IV) (4 Credits)

Time : 3 Hours] ***[Max. Marks : 70***

Instructions to the candidates:

1) ***All questions are compulsory.***
2) ***Figures to the right indicate full marks.***
3) ***Use of statistical tables and scientific calculator is allowed.***
4) ***Symbols and abbreviations have their usual meaning.***

Q1) Answer the following questions in 1-2 lines each. **[10]**

a) Can jackknife be considered as a special case of k-fold cross validation? Justify.

b) MCMC stands for _______

c) What are the three types of missing data?

d) Can every PDF be a kernel function? Justify.

e) For acceptance-rejection sampling, which one needs to be large, the support of the proposal distribution or the support of the target distribution? Justify.

Q2) Attempt any three subquestions from the following:

a) The following table gives a small dataset. **[5]**

Obs	X	Y
1	5	y_1
2	10	35
3	15	55
4	20	y_2
5	25	y_3

If we use regression imputation, what will be the values of y_1, y_2,y_3? Provide all the calculations.

P.T.O.

b) Explain the similarities and differences between the bagging of trees and random forest models. In which way does the random forest algorithm improve the model based on bagging and how? **[5]**

c) Describe the connection between kernel density estimation and kernel regression estimation. **[5]**

d) Suppose that we obtain a bootstrap sample from a set of *n* observations. What is the probability that the j^{th} observation is not in the bootstrap sample? **[5]**

Q3) Attempt any three subquestions from the following.

a) Explain the procedure of boosting for regression trees. **[5]**

b) Explain the procedure for k-nearest neighbour regression. Enlist the situations (with justification) under which k-nearest neighbour regression is expected to give better results than the linear regression and vice-versa. **[5]**

c) What are the two key steps in EM algorithm? Explain in detail. **[5]**

d) Suppose that we want to estimate the probability of P[Z > 3] for a standard normal random variable Z. Roughly how many observations would be required if we try to get a reasonable estimate of this via Monte Carlo sampling? What would be the steps if we decide to use importance with N(4, 1) as the proposal density? **[5]**

Q4) Attempt any three subquestions from the following.

a) Explain the k- nearest neighbour estimation procedure of estimating a density function in the univariate setup. **[5]**

b) In a binary classification problem, after fitting an initial model with equal weights to all the observations, it is found that out of three observations, one observation is misclassified. What is the revised weight for the misclassified observation as per the Adaboost. M1 algorithm? Provide all the intermediate steps for calculations. **[5]**

c) Explain the working of acceptance - rejection sampling. **[5]**

d) What is single imputation and multiple imputation? What are the advantages of multiple imputation? **[5]**

Q5) Attempt any one subquestion.

a) Describe the Metropolis Hastings algorithm and justify its working by providing all the appropriate formulae. **[15]**

b) Derive the formulae for the estimates of missing proportions and other parameters in a finite mixture model via EM algorithm. **[15]**

✠ ✠ ✠

Total No. of Questions : 5]

P587

[5841]-404
M.Sc.
STATISTICS
ST-44(B) : Analysis of Clinical Trials
(2019 Pattern) (Semester - IV) (4 Credits)

Time : 3 Hours] ***[Max. Marks : 70***

Instructions to the candidates:

1) ***All questions are compulsory.***
2) ***Figures to the right indicate full marks.***
3) ***Use of statistical tables and scientific calculator is allowed.***
4) ***Symbols and abbreviations have their usual meaning.***

Q1) Attempt all questions: **[5×2=10]**

a) Define Clinical trial with examples.

b) Define safety and efficacy with example.

c) Explain Multicentre Clinical trial.

d) Explain Interim analysis

e) Write Advantages of Run in period

Q2) Attempt any three out of four questions: **[3×5=15]**

a) Write short note on Traditional design.

b) Write short note on Protocols of clinical trial.

c) Explain Log linear model.

d) Explain William design for more than two drugs.

Q3) Attempt any three out of four questions: **[3×5=15]**

a) Explain Fieller confidence interval approach for claiming bioequivalence.

b) For standard 2X2 cross over design, write down model and assumptions. Also develop a test for testing direct drug effects when there is no carry over effects.

c) Define Blinding and explain Types of Blinding's.

d) Define Bioavailability and Bioequivalence. Explain different regulatory rules to decide bioequivalence.

Q4) Attempt any three out of four questions: **[3×5=15]**

a) Derive the formula for sample size estimation for tests between two independent sample means.

b) Define parallel and cross over design with advantages and disadvantages.

c) Explain Design B in brief.

d) Explain Kaplan mier nonparametric tests in clinical trial.

Q5) Attempt any one out of two questions: **[1×15=15]**

a) i) The list of Primidone concentration (μg/ml versus time point (hrs) from subject over 32 hours period after administrated a 250 gm tablet of drugs. The blood sample were drawn immediately before and after at time points are as follows. **[9]**

Sr.No.	1	2	3	4	5	6	7	8	9	10	11	12	13
T_i	0	0.5	1	1.5	2	3	4	6	8	12	16	24	32
C_i	0	0	2.8	4.4	4.4	4.7	4.1	4	3.6	3.0	2.5	2	1.6

Find T_{max}, C_{max}, $AUC_{(0-32)}$, $AUC_{(0-\infty)}$, $\log_{10} C_i = 0.6713 - 0.01518t_i$

ii) Explain Carryover period, Direct drug effect and washout period. **[6]**

b) i) Write short note on Pharmacokinetic and Pharmacodynamics with suitable examples. Explain parameters of Pharmacokinetic. **[9]**

ii) A hospital administrator wishes to estimate the mean weight of babies born in her hospital. How large a sample of birth records should be taken if she wants a 99 percent confidence interval that is 1 pound wide? Assume that a reasonable estimate of σ is 1 pound and power is 90 percent. **[6]**

✠ ✠ ✠