INDIAN STATISTICAL INSTITUTE

Mid-semester Examination: (2007-2008) I Semester

B. STAT. I year

Vectors and Matrices I

Date: 3 Sept. 2007. Maximum Marks: 30 Duration: 3 Hrs.

Note: (i) All the matrices and vectors considered are over complex field unless otherwise stated. (ii) Class room notation is used. (iii) This paper is for 40 marks. You can answer any part of any question. The maximum number of marks you can score is 30.

1.Prove the following:

(a) $r(AB) = r(B) - d\{C(B) \bigcap N(A)\}$.
(b) $d\{N(A)\} = n - r(A)$ where n is the number of columns of A.
(c) $r(AB) + r(BC) \leq r(ABC) + r(B)$.
(d) P and Q are square matrices of order $n \Rightarrow r(PQ) \geq r(P) + r(Q) - n$.
(e) $r(AB - I) \leq r(A - I) + r(B - I)$.
(f) $||Ax + By|| \geq ||Ax||$ for all x and $y \Rightarrow$ columns of A are orthogonal to columns of B.
(g) Every nonnull matrix has rank factorization.
(h) $A = A^2 \Leftrightarrow CB = I$ where $A = BC$ is any rank factorization of A.
(i) The following three statements are equivalent:
(i) G is a A^-, (ii) $AGA = A$ and (iii) GA is idempotent and $r(GA) = r(A)$.

[5 + 2 + 2 + 2 + 2 + 5 + 5 + 2 + 5 = 30]

2. Prove or disprove the following:

(a) $r(ABC) = r(AC)$ if B is a nonsingular matrix.
(b) $ABB^* = 0 \Leftrightarrow AB = 0$.
(c) $A^2 = A \Leftrightarrow Ax = x$ for all x in $C(A)$.
(d) $tr(A) = r(A) \Rightarrow A^2 = A$.
(e) $C(A) = N(A) \Rightarrow A = 0$.
(f) There exist matrices A and B such that $AB - BA = I$.
(g) $r(A^k) = r(A^{k+1})$ for some $k \Rightarrow r(A^k) = r(A^{k+2})$.
(h) $r(A - B) \geq r(A) - r(-B)$.
(i) Every idempotent matrix has an idempotent generalized inverse.
(j) The following system of three equations is inconsistent:
(i)$2\sqrt{2}x + y - z = \sqrt{3}$, (ii)$\sqrt{2}x + y - z = \sqrt{3}$ and (iii) $x + \sqrt{2}y - z = \sqrt{3}$.

[$10 \times 1 = 6$]

$\cdots xXx \cdots$

Compute Kendall's τ for the above data and comment on the association between the performances of the players in the French Open and the Wimbledon. [4+5+5]

4. The following data pertain to fasting glucose levels (FBS) and triglyceride levels (TRI) of 12 patients in a study on Type 2 diabetes:

Patient	FBS	TRI
1	85	157
2	115	535
3	109	96
4	145	87
5	210	388
6	107	154
7	75	136
8	235	185
9	81	202
10	188	247
11	264	119
12	173	206

(a) Draw the scatter plot of FBS and TRI. From the plot, comment on the suitability of a linear regression of TRI on FBS.

(b) Based on a least squares linear regression of TRI on FBS, predict the triglyceride level of a patient with fasting glucose level 150.

(c) Based on the fitted regression line, comment using both graphical and quantitative diagnostics, how good the linear fit is. [3+7+8]

INDIAN STATISTICAL INSTITUTE

Mid-Semester Examination : (2007 - 2008)

B.Stat First Year : Probability I

Date : 06-09-07. Maximum Marks : 40. Duration : Two hours.

Note: Justify your answers.
Quote precisely if you are using any result proved in class.
This paper is set for 45 marks. Maximum you can score is 40 marks.

1. I have n sticks. Each is broken into two pieces – one long and one short piece. These $2n$ pieces are paired at random to form n sticks.

   (a) What is the probability that they are joined to form original sticks?
   (b) Find the probability that all long parts are paired with short parts.

   [4 + 4 = 8]

2. (a) Take a sample of size 4 *with* replacement from $\{1, 2, \cdots, 100\}$. Find the probability that the product of the first three selected numbers is divisible by 5.

   (b) Suppose that I selected a sample of size 4 *without* replacement from the above set. Find the probability that the product of the first three elements of the sample is divisible by 5.

   [2 + 2 = 4]

3. I have a coin whose chance of heads in a single toss is p. I toss it independently. For $n \geq 1$, let a_n denote the probability of getting an even number of heads in n tosses (zero is an even number).

   (a) Show that $a_n = p(1 - a_{n-1}) + (1 - p)a_{n-1}$, for $n \geq 2$.
   (b) Show that $a_n = \frac{1}{2}[1 + (1 - 2p)^n]$, for $n \geq 1$.

   [3 + 3 = 6]

4. I have a sample space (Ω, p). For two events A and B, say that B gives positive information about A if $P(A|B) \geq P(A)$. If the following statement is correct prove it, otherwise, give an example to show that it is false.

P.T.O

(a) If B gives positive information about A, then A gives positive information about B.

(b) If C gives positive information about B and B gives positive information about A, then C gives positive information about A.

[2 + 4 = 6]

5. Let us pretend (for simplicity of calculations) that BI has 10 students, BII has 20 students, BIII has 20 students, MI has 25 students and MII has 25 students.

(a) I select a student at random and X denotes the number of students in his class. Find the distribution of X and $E(X)$.

(b) I select one of the five classes at random and Y denotes the number of students in the selected class. Find the distribution of Y and $E(Y)$.

[4 + 4 = 8]

6. Show that the number of paths of length $2n$ starting from $(0,0)$ and lying on or above X-axis is same as the number of paths of length $2n$ starting from $(0,0)$ and ending at $(2n,0)$.

[5]

7. I have a coin whose chance of heads in a single toss is p. I toss it independently till I get either *two consecutive* heads or *two consecutive* tails and stop.

(a) What are all the outcomes of this experiment and for each outcome what is its probability. Check that your probabilities add to one.

(b) If X is the number of tosses needed, show that $E(X) = \frac{2-pq-p^2q^2}{(1-pq)^2}$.

[3 + 5 = 8]

INDIAN STATISTICAL INSTITUTE

Mid Semester Examination, 1st Semester, 2007-08

Statistical Methods I, B.Stat 1st Year

Date: September 12, 2007 Time: 2 hours

This paper carries 35 marks.

1. (a) Explain the difference between a prospective study and a retrospective study with a suitable example in each case.

   (b) Consider a study on the effect of exposure to some environmental agent on a disease outcome. Based on the data collected, suppose we define a quantity OR (known as Odds Ratio):

$$\frac{\text{Prop}\{\text{disease/exposure}\}}{\text{Prop}\{\text{normal/exposure}\}} \div \frac{\text{Prop}\{\text{disease/no exposure}\}}{\text{Prop}\{\text{normal/no exposure}\}}$$

   If we define a quantity R as follows:

$$R = \frac{\text{Prop}\{\text{exposure/disease}\}}{\text{Prop}\{\text{no exposure/disease}\}} \div \frac{\text{Prop}\{\text{exposure/normal}\}}{\text{Prop}\{\text{no exposure/normal}\}}.$$

   show that $R = OR$. [6+4]

2. The following frequency distribution pertains to run rates in different innings of the warm-up and preliminary matches of ICC World Cup 2007:

Run Rate	Frequency
2-3	3
3-4	22
4-5	25
5-6	18
6-7	6
7-8	3
8-9	3

Draw a box and whisker chart for the above data. [10]

P. T. O

3. Consider a set of observations $x_1, x_2, \ldots, x_n$ such that the mean is $\bar{x}$ and the variance is s^2.

   (a) If x_n is the only observation greater than $\bar{x}$, show that the mean deviation of the observations about the mean is equal to $\frac{2}{n}(x_n - \bar{x})$.

   (b) Is $\frac{1}{n}\sum_{i=1}^{n} x_i^4$ necessarily greater than or equal to $\bar{x}^4$? Justify your answer.

   (c) When a new observation x_{n+1} is introduced to the dataset, the mean of the observations decreases but the variance remains the same. Express x_{n+1} in terms of $\bar{x}$ and s^2.

   (d) Assuming all the observations to be positive, determine which summary measure θ minimizes the expression $\sum_{i=1}^{n} \{\sqrt{x_i} - \frac{\theta}{\sqrt{x_i}}\}^2$?

   [3+3+6+3]

INDIAN STATISTICAL INSTITUTE
First Semester Examination (2007-2008)
B.STAT (FIRST YEAR)
Computing Techniques And Programming I
26 November 2007. Full Marks: 80. Duration: Three hours.

Note: Answer all the questions

1. (a) (i) What are the advantages of using floating point numbers over the fixed point numbers?
   (ii) With 8-bit exponent and 23-bit mantissa, estimate the range of numbers(take any case of sign bit positive or negative) in normalized representation.
   (iii) What is the utility of keeping exponents always positive? How do we achieve this? Explain.
   (b) What is a half adder? Starting from truth table derive a full adder circuit in terms of two half adders and an OR gate.

   [(1+2+2)+(2+3)]

2. (a) Starting from the truth table definition, derive the Boolean expressions of all the 2-variable non linear Boolean functions.
   (b) What is an ODD/EVEN function? Explain their significance.
   (c) Why NAND/NOR is considered to be a universal gate?

   [2 + 2 +6]

3. Design the SEC/DED (Single Error Correcting/ Double Error Detecting) code for 8-bit message. (You should illustrate your design with only four code-words instead of describing all the 256 code-words and your illustration should indicate the rectification of at least one single error and the detection of the occurrence of any double error). [8 +1+1]

4. (a) Distinguish between the following parameter passing techniques with examples:
   (i) Call by Value
   (ii) Call by reference
   (b) In two's complement arithmetic consider the following cases:
   (i) Two numbers are positive
   (ii) Two numbers are negative
   (iii) One positive and another negative
   In what situation overflow and underflow would take place? How these can be remedied?

[4 + 6]

5. Bring out the differences between the iterative and recursive programming techniques. Illustrate the above programming techniques with appropriate C code for:
   (i) Finding Fibonacci Sequence
   (ii) Finding an element in binary search [2 +4 + 4]

6. (a) Devise an algorithm for the conversion of infix notation into the corresponding postfix notation on using STACK data structure and illustrate it with the help of an example of dry run.
   (b) Give the corresponding C code with full documentation
[6 + 4]

7. A matrix is called a square matrix if it has the same number of rows and columns. Further a square matrix is a diagonal matrix if its only nonzero elements are on the diagonal from upper left to lower right. is called upper triangular if all elements below diagonal are zero anc lower triangular if all elements above the diagonal are zero. Write a program that reads a square matrix and determines if it is a diagonal upper triangular or lower triangular. (You should illustrate all the three cases with full documentation.)
[2 + 4 +4]

8. Write a C program for creating a single linked list of several intege and hence extend the C program (with full documentation) for reversing the direction of the pointers in the already created linked list. Your program should print both the linked lists. [5 + 5]

7.(a) X is a random variable with p.g.f. $Q(t) = e^{-\lambda(1-t^2)}$. Calculate $P(X = 20)$, $P(X = 19)$.

(b) Consider random walk starting from 0; probability of moving one step forward is p and probability of moving one step backward is $(1-p)$. Put $p_0 = 0$ and for $n \geq 1$, p_n be the probability that 1 is reached for the first time on n^{th} day. Calculate the pgf of the sequence (p_n).

[3+5]

8. The number of customers that enter a shop is a Poisson variable with parameter λ. Each customer, independent of others, buys an item with probability p (and does not buy anything with probability $(1-p)$. If Y is the number of customers who did not buy anything from the shop calculate the distribution of Y. Find $E(Y)$.

[7]

***

INDIAN STATISTICAL INSTITUTE
First Semester Examination: 2007-08
B. Stat. I Year
Probability Theory I

Date: 07.12.07 Maximum Marks: 60 Duration: 3 Hours

This paper is set for 65 marks. Maximum you can score is 60. Justify your answers.

1. n balls are distributed at random into r boxes. Let X be the number of empty boxes. Calculate Expectation and Variance of X.

[7]

2. Let $S = \{1,2,\ldots\ldots,n\}$ so that S has 2^n subsets (including empty set). Suppose two subsets A and B are selected at random (with replacement). Show that $P(A \subset B) = \left(\frac{3}{4}\right)^n$.

[8]

3. $X \sim B(n,p)$ and $Y \sim P(np)$. For $k \geq 0$ put $a_k = \dfrac{P(X=k)}{P(Y=k)}$. Show that a_k first increases and then decreases. For what value of k is a_k maximum?

[5]

4. A coin, whose chance of heads is p, is tossed independently. Let X be the length of the run started by the first trial and Y be the length of the second run.

Find the distributions of X and $E(X)$.
Find the distribution of Y and $E(Y)$.

[7+8]

5. Consider Polya Urn Scheme with r red balls and g green balls. Prove that the probability of a red ball at any draw is $\dfrac{r}{r+g}$.

[7]

6. The number of accidents in Madras, Calcutta, Bombay, Delhi are independent Poisson Variables with parameters 1,2,3,4. Given that a total of four accidents occurred, calculate the conditional probabilities of the following events.

a) One accident in each city.
b) One in Calcutta, one in Bombay and two in Delhi.

[8]

[P.T.O.]

(b) Consider a set of observations with mean $\bar{x}$ and standard deviation s. Show that the proportion of observations lying outside the interval $[\bar{x} - ks\ ,\ \bar{x} + ks]$ is at most β_2/k^4, where k is any positive real number and β_2 is the unadjusted measure of kurtosis.

(c) Consider two sets of bivariate data on (x,y) comprising m and n observations respectively. The means of x, the means of y, the variances of x and the variances of y are the same in the two sets of observations. The correlation coefficients between x and y in the two sets are r_1 and r_2, respectively. Show that the correlation coefficient between x and y in the pooled set comprising $(m+n)$ observations can be expressed as $\lambda r_1 + (1-\lambda) r_2$, for an appropriate real number $\lambda \in (0,1)$.

[3 + 4 + 4]

3. (a) Suppose X and Y are two binary variables, each assuming values 0 or 1. Suppose, based on a set of observations on (X,Y), a least squares regression of Y is performed on X. Show that the regression line has slope zero if and only if the Odds Ratio based on (X,Y) is 1.

(b) Based on a bivariate set of observations on (x,y), suppose you want to perform a linear regression of y on x by minimizing the expression: $\alpha_0 \sum \{y_i - a - bx_i\}^2 + (1-\alpha_0) \sum |y_i - a - bx_i|$ with respect to a and b, where α_0 is a fixed real number between 0 and 1. Formulate the above minimization as a weighted least squares minimization problem and describe an algorithm to estimate a and b. Show all your computational steps clearly.

(c) The following data pertain to the proportion of matches won in the French Open and the Wimbledon by seven women tennis players:

Player	French Open	Wimbledon
1	0.7	0.4
2	0.6	0.8
3	0.6	0.7
4	0.2	0.8
5	0.3	0.6
6	0.5	0.8
7	0.4	0.9

INDIAN STATISTICAL INSTITUTE

Final Examination, 1st Semester, 2007-08

Statistical Methods I, B.Stat 1st Year

Date: November 28, 2007 Time: $3\frac{1}{2}$ hours

This paper carries 55 marks. Attempt all questions. The maximum you can score is 50.

1. (a) A study showed that 88% of the people suffering from Hepatitis B who took Lamivudine recovered within a week. Do you think that the study design was appropriate to make conclusions on the effectiveness of Lamivudine? Explain.

   (b) Suppose the instrument to measure the maximum concentration of SO_2 in the overground air layer is calibrated only within the range 25-450 μ_g/m^3, and hence, levels lower than 25 μ_g/m^3 or greater than 450 μ_g/m^3 are recorded as the corresponding boundary values. Suggest a suitable measure of central location and a measure of spread which will summarize data on SO_2 levels on different days within a particular week. Justify your answer.

   (c) The following summary measures are reported for a set of observations on the scores of 10 students in an examination:

   mean = 71.6; variance = 32.5; median = 78.4.

   Comment on the feasibility of the above measures. You need to derive any result you may use. [3 + 3 + 6]

2. (a) Boyle's Law states that the pressure (P) of a gas at a given temperature is inversely proportional to the volume (V) of the gas at that temperature. Does this imply that if you collected bivariate data on (V,P) of a gas at 10C, the correlation coefficient between V and P will be -1? Justify your answer. Suggest suitable transformations on V and P such that the correlation coefficient between the transformed variables is -1.

P.T.O

INDIAN STATISTICAL INSTITUTE

First Semester Examination: (2007-2008) I Semester

B. STAT. I year

Vectors and Matrices I

Date: 30 Nov. 2007. Maximum Marks: 60 Duration: 3 Hrs.

Note: (i) All the matrices and vectors considered are over complex field unless otherwise stated. (ii) Class room notation is used ($R(A), C(A), N(A)$ and $r(A)$ denote the row space, the column space, the null space and the rank of the matrix A). (iii) This paper is for 70 marks. You can answer any part of any question. The maximum number of marks you can score is 60.

1 Prove that the following statements are equivalent:

(a) $r(A) = r(A^2) = r$.

(b) VU is a nonsingular matrix of order r where $A = UV$ is any rank factorization of A.

(c) $A = P\begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}P^{-1}$, for some nonsingular matrix P and a nonsingular matrix D of order r.

[10]

2 Prove the following:

(a) $r(A+B) = r(A) + r(B) \Leftrightarrow C(A) \bigcap C(B) = \{\phi\}$ and $R(A) \bigcap R(B) = \{\phi\}$.

(b) $R(A) \subseteq R(B) \Rightarrow r(B) = r(A) + r(B(I - A^{-}A))$.

(c) A is idempotent $\Rightarrow MA$ is hermitian for some positive definite (pd) matrix M.

(d) G is a $A_r^- \Rightarrow G$ is a A_{MN}^+ for some positive definite matrices M and N.

(e) A and B are square matrices such that $(A : B')$ is a g-inverse of $(-B' : A)' \Rightarrow A = 0$.

(f) If A and B are orthogonal projection operators then $A - B$ is a projection operator if and only if $AB = B$.

[$6 \times 5 = 30$]

3 Let A be a nonsingular matrix of order n and u and v be n-tuples. Obtain a necessary and sufficient condition for the matrix $(A + uv')$ to be nonsingular and when it is nonsingular, express the inverse of $(A + uv')$ interms of inverse of A.

[10]

4 Prove or disprove the following:

(a) G is a $A_m^- \Leftrightarrow G^*$ is a $(A^*)_\ell^-$.

(b) A has unique generalized inverse implies A is nonsingular.

(c) A is a generalized inverse of itself implies A is idempotent.

P.T.O.

(d) $r(APA^*) = r(A)$ where P is a positive definite matrix.
(e) $r(A) = r(A^2)$ and $C(A) = N(A) \Rightarrow A = 0$.
(f) The matrix T is triangular and also orthogonal $\Rightarrow T = I$.
(g) If A is real then there exists a solution of the consistent system $Ax = b$ in $C(A')$.
(h) Every real symmetric matrix has a rank factorization of the form $A = UU'$.
(i) A and B are real symmetric idempotent matrices such that $C(A) = C(B) \Rightarrow A = B$.
(j) A and B are matrices of same order such that $C(A) = C(B) \Rightarrow B = AZ$ for some nonsingular matrix Z.

[$10 \times 2 = 20$]

$\cdots xXx \cdots$

INDIAN STATISTICAL INSTITUTE

Semestral Examination

First semester 2007–2008

B. Stat (First year)

Analysis I

Date: 4 December, 2007

Maximum Marks:

Duration: 3 hou

Answer all questions.

State clearly any result that you use in your answer.

(1) Prove that every open subset of $\mathbb{R}$ can be expressed as countable union of closed set
Show by an example that the converse of this need not be true.

(2) Define a function $d : \mathbb{Z}^\times \times \mathbb{Z}^\times \longrightarrow \mathbb{R}$ by $d(m,n) = |\frac{1}{m} - \frac{1}{n}|$ for $m, n \in \mathbb{Z}^\times$, where $\mathbb{Z}^\times$ is
set of non-zero integers. Show that d is a metric on $\mathbb{Z}^\times$. Show that the sequence $x_n =$
$n = 1, 2, \ldots$ is a Cauchy sequence relative to the metric d but it is not convergent.

(3) Let $\mathbf{a} = (a_1, a_2) \in \mathbb{R}^2$ and $r > 0$. Define

$$S(\mathbf{a}; r) = \{(x_1, x_2) :\in \mathbb{R}^2 : (x_1 - a_1)^2 + (x_2 - a_2)^2 = r^2\}$$

(a) Determine whether $S(\mathbf{a}; r)$ is compact or not.
(b) Prove that $S(\mathbf{a}; r)$ is homeomorphic with $S(\mathbf{b}; s)$ for any $\mathbf{b} \in \mathbb{R}^2$ and $s > 0$.
(c) Can there be a continuous onto map $S(\mathbf{a}; r) \longrightarrow \mathbb{R}$? Justify your answer.

(4) (a) Let $f : [a, b] \longrightarrow \mathbb{R}$ be a continuous function which is monotonically increasir
Prove that $f([a, b]) = [f(a), f(b)]$.
(b) If $f : \mathbb{R} \longrightarrow \mathbb{R}$ is monotonic and image of f is an interval (may be unboundec
then show that f is continuous.

(5) Prove that the function

$$f(x) = \frac{\sin x}{x}$$

is a monotonically decreasing function and is bounded above by 1 on the interval $(0, \pi$

7+

(6) Use Taylor's theorem to prove that the function $\log(1 + x)$ has the following power serie
expansion for $0 \le x < 1$:

$$\log(1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots + (-1)^{n+1}\frac{x^n}{n} + \ldots$$

INDIAN STATISTICAL INSTITUTE

Second Semester Midterm Examination (2007–08)

B. Stat. I Year

Analysis II

Date : 18.02.2008 Maximum Marks : 100 Time : 3 hours

The question carries 120 marks. Maximum you can score is 100. Precisely justify all your steps. Carefully state all the results you are using.

1. Let $f : [0,1] \to \mathbb{R}$ be defined by

$$f(x) = \begin{cases} x^3 & \text{if} \quad x \text{ is rational} \\ x^2 & \text{if} \quad x \text{ is irrational} \end{cases}$$

Compute the upper and the lower Riemann integrals of f on $[0,1]$ and hence decide whether $f \in \mathcal{R}[0,1]$. [10]

2. (a) Without quoting the theorem on integrability of composite functions, show that $f \in \mathcal{R}[a,b] \implies |f| \in \mathcal{R}[a,b]$. Is the converse true? [10 + 3 = 13]

(b) Let $f : [a,b] \to \mathbb{R}$ be a bounded function. Define

$$f_+(x) = \begin{cases} f(x) & \text{if} \quad f(x) > 0 \\ 0 & \text{otherwise} \end{cases} \quad \text{and } f_-(x) = \begin{cases} -f(x) & \text{if} \quad f(x) < 0 \\ 0 & \text{otherwise} \end{cases}$$

Show that $f \in \mathcal{R}[a,b] \iff$ both f_+ & $f_- \in \mathcal{R}[a,b]$. Moreover,

$$\int_a^b f(x)dx = \int_a^b f_+(x)dx - \int_a^b f_-(x)dx.$$ [10]

(c) Let $f : [a,b] \to \mathbb{R}$ be a function with a continuous derivative. Show that f is the sum of a continuous increasing function and a continuous decreasing function. [7]

P.T.O

3. Test the convergence of the integral

$$\int_0^\infty \frac{1}{x^2+\sqrt{x}}\,dx. \qquad [15]$$

4. Let $\{f_n\}$ be a sequence of continuous functions which converges uniformly to a function f on an interval I. Prove that

$$\lim_{n\to\infty} f_n(x_n) = f(x)$$

for every sequence $\{x_n\} \subseteq I$ such that $x_n \to x \in I$.

Show that the conclusion may fail if the convergence is not uniform. [10+5 = 15]

5. Let $g : [0,1] \to \mathbb{R}$ be continuous. Let $f_n(x) = x^n g(x)$ for $x \in [0,1]$. Show that $\{f_n\}$ converges uniformly on $[0,1]$ if and only if $g(1) = 0$. [10]

6. Let $f_n(x) = n^\alpha x(1-x^2)^n$ for $x \in [0,1]$, $n \geq 1$.

(a) Show that $\{f_n\}$ converges pointwise on $[0,1]$ for any $\alpha \in \mathbb{R}$. [5]

(b) Find all α such that the convergence is uniform on $[0,1]$. [8]

(c) Find all α such that [7]

$$\lim_{n\to\infty}\left(\int_0^1 f_n(x)dx\right) = \int_0^1 \left(\lim_{n\to\infty} f_n(x)\right) dx.$$

7. Starting from a geometric series and precisely justifying all your steps prove that [20]

$$\int_0^1 \frac{\tan^{-1} x}{x} dx = \sum_{n=0}^{\infty} (-1)^n \frac{1}{(2n+1)^2}.$$

First Semester Examinations (2007-2008)
B Stat – 1 year
Remedial English
100 marks
1 ½ hours

Date: 5.12.07

1. Write an essay on any one of the following topics. Five paragraphs are expected.
   a. Winter in my city
   b. My hobby
   c. My views on the present education system.

(60 marks)

2. Fill in the blanks with appropriate prepositions:
   a. Industry is the key______ success.
   b. The customs were searching _______ drugs at the airport.
   c. Elizabeth knew how to inspire her soldiers ______ hope.
   d. Early rising is beneficial _______ health.
   e. We should live in a style suited _______ our condition.
   f. Examinations act as an incentive _______ diligence.
   g. He is too miserly to part _______ his money.
   h. He is a clever man but unfortunately diffident ______ his powers.
   i. Suddenly we were enveloped _______ dense fog.
   j. Many aspire _______ greatness but few attain it.
   k. His income is not adequate ______ his wants.
   l. The soil of Pune is favourable ______ roses.
   m. I am sick ______ the whole business.
   n. A car will be a great convenience ________ a busy man like him.
   o. Whoever acts contrary ________ nature does not go unpunished.
   p. The accident resulted _________ the death ________ five people.
   q. The mule was partially relieved _______ the load.
   r. Your wish is tantamount ________ a command.
   s. One is sure _______ what one sees.

(20 marks)

3. Fill in the blanks with appropriate words:

Ashok _________ over the top of his copy of the Times. The man in the _________ seat had drifted off to ________. His snores came out in intervals, coinciding _________ the rhythm of the train. His moustaches quivered with each breath, a silver mass of frost on the branch of a bare tree in the wind. Ashok smiled and folded his _________. Then, slipping out a small black ________, he opened it to an _______ page and hid behind it with a miniscule wood pencil. His strokes ______ light, quick and

P.T.O

clever and the yellowing white of the page was soon transformed into a man, a _______old man. snoring with his _______half open and his head tilted up against the _______.

There was humour _______ the lines of the drawing and Ashok could feel . The _______ was, people never saw his pictures the way he did. Every one of Ashok's _ was perfect in his eyes. Every _______ of expression on each of his subjects' s there for all the _______ to see, his compositions were clever, the colours just yet...he never seemed to _______ it big.

The carriage _______ a sudden heavy jolt and the man woke _______, . He shook his head to clear it and, staring hard at Ashok and his glance, eased s seat and went _______ for a smoke.

(20 marks)

INDIAN STATISTICAL INSTITUTE

Mid-semester Examination: (2007-2008)

B. STAT. I year

Vectors and Matrices II

Date: 26 Feb. 2008. Maximum Marks: 30 Duration: 3 Hrs.

Note: (i) All the matrices and vectors considered are over complex field unless otherwise stated. (ii) Class room notation is used.(iii) State clearly all the results used.

1 (a) Show that every square matrix is unitarily similar to a triangular matrix.

(b) Define algebraic multiplicity $am(\lambda, A)$ and geometric multiplicity $gm(\lambda, A)$ of an eigenvalue λ of a matrix A and show that $am(\lambda, A) \geq gm(\lambda, A)$.

[$2 \times 5 = 10$]

2 Let A be a real skew-symmetric matrix of order n and rank r. Then prove the following:

(a) A is singular if n is odd.

(b) A has r nonzero eigenvalues.

(c) If λ is a nonzero eigen value of A then $\lambda = i\mu$ for some real number μ.

(d) If $x + iy$, where x and y are real vectors, is an eigen vector of A corresponding to a nonzero eigen value then x and y are orthogonal and are of same norm.

[$2 + 2 + 2 + 4 = 10$]

3 State and prove Caley-Hamilton theorem.

[5]

4 Prove or disprove the following:

(a) $x^2 + 2x - 3$ is the minimal polynomial of A implies A is nonsingular.

(b) If A is symmetric then there exists a constant c such that $I + cA$ is positive definite.

(c) A is nilpotent if and only if all the eigenvalues of A are zero.

(d) For a triangular matrix T, $rank(T) = rank(T^2)$ if and only if T has r nonzero diagonal elements where r is rank of T.

(e) $gm(1, A) = rank(A)$ implies A is idempotent.

[$5 \times 2 = 10$]

INDIAN STATISTICAL INSTITUTE

Mid semester examination: (2007-2008)

B.Stat First Year   Pobability Theory II

Date: 22-02-08   Maximum Marks: 40   Duration: Two hours

1. We select a point at random in the unit square $[0,1] \times [0,1]$ of the plane and calculate $Z = X - Y$ where X and Y are the x-coordinate and y-coordinate respectively of the selected point. Find the density of Z.

[8]

2. Let X be a uniform $(0,1)$ random variable. Denote by U, the first digit in the decimal expansion of X and by V, the second digit in the decimal expansion. Calculate the distributions of U and V. Are they independent? [8]

3. Exhibit a funcion φ on the real line so that the following happens. If we take a random variable X wih density $f(x) = 2x$ for $0 < x < 1$, then $Y = \varphi(X)$ has density $g(y) = 4y^3$ for $0 < y < 1$. [5]

4. If $Z \sim N(0,1)$, then show that $P(|Z| < 21) = 2\,P(Z < 21) - 1$. [4]

5. Let X be a random variable with distribution function

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{x+1}{10} & \text{if } 0 \le x < 4 \\ \frac{1}{2} & \text{if } 4 \le x < 6 \\ \frac{x-5}{2} & \text{if } 6 \le x < 7 \\ 1 & \text{if } x \ge 7 \end{cases}$$

Calculate the following probabilities. $P(0 \le X \le 5)$; $P(0 < X < 5)$; $P(4 \le X < 6)$; $P(X$ is a rational number$)$ [8]

6. Let X be a random variable with density $f(x) = ax + bx^2$ for $0 < x < 1$. If $E(X^2) = 9/20$, calculate $P(X \le 1/2)$. [8]

7. If X is exponential with parameter λ, calculate density of the random variable $Y = (5 - X)/5$. [4]

****

INDIAN STATISTICAL INSTITUTE

Mid-semester Examination (2007- 08)
B. Stat. 1st Year

COMPUTATIONAL TECHNIQUES AND PROGRAMMING II

Date: February 20, 2008 Maximum marks: 50 Time: 2 hrs

1. For a function $g(x)$ tabulated at $x = x_0, x_1, \ldots, x_n$, where $x_i = x_0 + ih$, for $i = 0,1,2,\ldots n$, h being a positive real number, define finite forward differences of various orders, and express them in terms of the tabulated function values.
For the polynomial $p(x) = 4x^3 + x^2 - x + 99$, express finite forward differences of all orders as functions of h, taking $x_0 = 0$.
[5+5=10]

2. Derive the Newton Backward Difference Interpolation Formula from first principles, stating clearly the rationale behind the approach.
[10]

3. What are splines? Describe explicitly how the interpolation problem can be solved by using cubic natural splines. For what reasons is interpolation with splines superior to Newtonian interpolation?
[2+4+4=10]

4. Starting with the basic Newton interpolation formula, deduce the Simpson quadrature rule for evaluating a given definite integral, based on values of the integrand at eight equispaced values of the argument. Obtain an estimate of the error you can expect this rule to make, in approximating the definite integral.
[6+4=10]

5. Explain in a few sentences the essence of the Gaussian approach for obtaining an approximation to the definite integral

$$\int_a^b f(x)dx,$$

using values of $f(x)$ at $n+1$ points in the interval $[a,b]$.

If $n = 2$, deduce the explicit form of the Gauss quadrature rule for evaluating the above integral, and also the associated error, given that the Legendre polynomial of degree 3 is

$$\frac{5}{2}\left(x^3 - \frac{3}{5}x\right).$$

[5+5=10]

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

INDIAN STATISTICAL INSTITUTE

Mid Semester Examination, 2^{nd} Semester, 2007-08

Statistical Methods II, B.Stat 1^{st} Year

Date: February 29, 2008 Time: 2 hours

This paper carries 35 marks.

1. *This is an extremely diluted version of the Duckworth-Lewis method employed in rain interrupted one-day international cricket matches.*

   Based on 100 matches interrupted by rain during the run chase of the team batting second, data were collected on four variables:

   X_1 : the number of runs required to win the match at the time of interruption
   X_2 : the number of overs remaining at the time of interruption
   X_3 : the number of wickets lost by the team batting at the time of interruption
   X_4 : the number of overs that can be accommodated when it is possible to resume the match

   The mean-vector and the dispersion matrix of (X_1, X_2, X_3, X_4) were obtained as follows:

$$\begin{pmatrix} 142.58 \\ 23.06 \\ 3.27 \\ 12.94 \end{pmatrix}, \begin{pmatrix} 28.2 & 14.8 & 18.4 & 13.3 \\ & 9.9 & 9.6 & 7.2 \\ & & 14.4 & 8.2 \\ & & & 9.8 \end{pmatrix}$$

   (a) Obtain a linear prediction function of X_1 using X_2, X_3 and X_4 based on least squares regression.

   (b) What proportion of the variance in X_1 is explained by the above linear regression?

   (c) Explain whether X_3 is useful in the above linear prediction of X_1. [5 + 5 + 5]

2. (a) Suppose $\Sigma_{a,c} = aI + c1' + 1c'$ where a is a scalar and c is a vector. Under what necessary and sufficient conditions on a and c is $\Sigma_{a,c}$ a valid dispersion matrix?

P.T.O

(b) Suppose the correlation matrix of $(X_1, X_2, \ldots, X_p)$ is $R = ((r_{ij}))$ such that $r_{1j} = \alpha_1$; $j = 2, 3, \ldots, p$ and $r_{ij} = \alpha_2$; $i, j \geq 2$, $i \neq j$. Compute the multiple correlation coefficient between X_1 and $(X_2, X_3, \ldots, X_p)$.

[5 + 5]

3. (a) A DNA sequence comprises nucleotides A, T, G, C; each occurring independently with probabilities p_1, p_2, p_3 and p_4, respectively, where $\sum_{i=1}^{4} p_i = 1$. Given a DNA sequence comprising n nucleotides, obtain the correlation matrix of the proportions of the four nucleotides in the sequence.

(b) An internal component of a machine has a fixed chance of becoming inactive on any particular day. Suppose the second date in this month when the component became inactive was the 10^{th}. What is the probability that the component did not become inactive before the 4^{th}?

[5 + 5]

INDIAN STATISTICAL INSTITUTE

Seond Semester Examination: (2007-2008)

B. STAT. I year

Vectors and Matrices II

Date: 2 May, 2008. Maximum Marks: 60 Duration: 3 Hrs.

'otes: *(i) All the matrices and vectors considered are over complex field unless otherwise tated. (ii) Class room notation is used. (iii) State clearly the results used.*

1 (a) Let B be a real positive definite matrix of order n. Then show that $|B| \leq b_{11}b_{22}\cdots b_{nn}$.

(b) Let A be a real matrix of order n such that $|a_{ij}| \leq 1$ for all i, j. Then show that $|A| \leq n^{n/2}$ and equality occurs if and only if a_{ij} is $+1$ or -1 for all i, j, and rows of A are pairwise orthogonal.

[5 + 5 = 10]

2 (a) Given a pair of nonnegative definite matrices X and Y of same order, show that there exists a nonsingular matrix U such that both U^*XU and U^*YU are diagonal.

(b) Given any pair of matrices $A_{m\times n}$ and $B_{k\times n}$, show that there exist a pair of unitary matrices $P_{m\times m}$ and $Q_{k\times k}$ and a nonsingular matrix $Z_{n\times n}$ such that

$$A = P\begin{pmatrix} \Phi & 0 \\ 0 & 0 \end{pmatrix} Z \quad \text{and} \quad B = Q\begin{pmatrix} \Psi & 0 \\ 0 & 0 \end{pmatrix} Z,$$

where Φ and Ψ are diagonal matrices of order r with nonnegative diagonal elements satisfying $\Phi^2 + \Psi^2 = I_r$, r being the rank of the matrix $(A^* : B^*)$. (Assume that $r \leq min\{m, k\}$).

[5 + 10 = 15]

3 (a) Define rank and signature of a quadratic form and prove that a quadratic form $x'Ax$ can be written as the product of two linearly independent linear forms in x if and only if A has rank 2 and signature 0.

(b) Let A be a nonnegative definite matrix of order n and B be a matrix with n rows. Then show that the row space of $(0 : B)$ is contained in the row space of $\begin{pmatrix} A & B \\ B' & 0 \end{pmatrix}$.

[5 + 5 = 10]

P.T.O.

4 (a) For the matrix A given below, find a lower triangular matrix L such that $A = LL'$ using square root method. Also find $|A|$.

$$A = \begin{pmatrix} 4 & 2 & -2 & 0 \\ 2 & 2 & 1 & 3 \\ -2 & 1 & 6 & 7 \\ 0 & 3 & 7 & 14 \end{pmatrix};$$

(b) Give a nonsingular matrix B such that the transformation $y = Bx$ will reduce the quadratic form $x'Ax$ to diagonal form.

[5 + 5 = 10]

5 Prove or disprove the following:

(a) $r(A) = r(A^2) = r \Rightarrow A$ has a nonzero principal minor of order r.

(b) Every orthoganal matrix of order 2 is either symmetric or skew symmetric.

(c) Let $A' = B + C$ where B is real positive definite matrix and C is real skew symmetric matrix.Then $|A| = |B| \Rightarrow C = 0$.

(d) A and B are nonnegative definite matrices $\Rightarrow$ all the eigenvalues of AB are nonnegative.

(e) Every complex symmetric matrix is normal.

(f) A is a positive definite matrix $\Rightarrow B = \begin{pmatrix} A & I \\ I & A^{-1} \end{pmatrix}$ is also positive definite.

[6 × 2.5 = 15]

· · · xXx · · ·

INDIAN STATISTICAL INSTITUTE

Second Semestral Examination (2007- 08)

B. Stat. 1st Year

COMPUTATIONAL TECHNIQUES AND PROGRAMMING II

Date: 6 May, 2008 **Maximum marks: 100** **Time: 3 hrs**

Note: Answer any 5 of the following questions

1.

a. For a function $f(x)$ tabulated at $x = x_0, x_1, \ldots, x_n$, define divided differences of all orders, and show that if $f(x)$ is a polynomial of degree n, then divided differences of order k, where $k > n$, are equal to 0.

b. For a real-valued function $g(x)$ tabulated at $x = x_0, x_1, \ldots, x_n \in [a,b]$, deduce the error of the interpolating polynomial $p_n(x)$ of degree n, if $g(x)$ is $(n+1)$ times differentiable in $[a,b]$.

[10+10=20]

2.

a. Derive the general Newton-Cotes quadrature formula from first principles for evaluating numerically the definite integral

$$\int_a^b f(x)dx,$$

based on values of $f(x)$ at $(n+1)$ equispaced values of x in $[a,b]$. Give the explicit formula for $n = 2$.

b. Explain in detail the basic principle behind Gaussian quadrature. How does its error compare with that of Newton-Cotes quadrature?

[10+10=20]

3.

a. For the problem of solving numerically a differential equation

$$y'(x) = f(x,y), \quad y(x_0) = y_0,$$

at N equispaced points in the interval $[x_0, x_N]$, the step size being h, describe the Euler method and, making appropriate assumptions, show that its error is $O(h)$.

b. For solving the same problem as in 3(a), deduce the general Adams-Bashforth formula, and explain its advantages and disadvantages over the Euler approach.

[10+10=20]

4.

a. Describe the Newton-Raphson approach for solving numerically a nonlinear equation $f(x) = 0$ in a single variable x. Establish that the order of convergence of this method is 2.

P.T.O

b. Explain how the Newton-Raphson method can be adapted *efficiently* to obtain all the zeros of a polynomial of degree p.

[10+10=20]

5.

a. Consider the problem of solving n linear equations $Ax = b$ in n unknowns, where the coefficient matrix A is real, symmetric. Describe a computationally efficient (as compared to the basic Gaussian elimination-based approach) non-iterative method for obtaining the solution.

b. Discuss the rationale behind the Jacobi approach to the computation of the eigenvalues of a square matrix A.

[10+10=20]

6.

a. For the problem of solving a nonlinear equation in x, of the type $x = g(x)$, under what conditions does fixed-point iteration converge? Justify your answer.

b. In cases where fixed-point iteration does converge, explain how its convergence can be accelerated. Justify your answer.

[10+10=20]

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

INDIAN STATISTICAL INSTITUTE

Second Semester Semestral Examination (2007–08)

B. Stat. I Year

Analysis II

Date : 9.05.2008 Maximum Marks : 100 Time : $3\frac{1}{2}$ hou

The question carries 120 marks. Maximum you can score is 100. Precisely justify all your steps. Carefully state all the results you are using.

1. Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous periodic function of period π.

(a) Show that the integral of f over any interval of length π is the same, *i.e.*,

$$\int_a^{a+\pi} f(x)dx = \int_b^{b+\pi} f(x)dx \quad \text{for all } a, b \in \mathbb{R}.$$

(b) Show that $\int_0^{\pi} [f(x+a) - f(x)]dx = 0$ for all $a \in \mathbb{R}$.

(c) Show that given $a \in \mathbb{R}$, there exists $x \in [0, \pi]$ such that $f(x+a) = f(x)$.

[10 + 4 + 6 = 2

2. Test the convergence of the integrals

(a) $\displaystyle\int_{\frac{1}{2}}^{\infty} \frac{e^{-x}}{x} dx.$

(b) $\displaystyle\int_a^b \frac{1}{(x-a)(b-x)} dx.$

[7+8=1

3. Let $f(x) = \displaystyle\sum_{n\geq 1}^{\infty} \frac{1}{1+n^2x}$.

(a) For what values of $x \in \mathbb{R}$ does the series converge absolutely?

(b) On which intervals $I \subseteq \mathbb{R}$ does the series converge uniformly? On which interva $I \subseteq \mathbb{R}$ does the series fail to converge uniformly?

(c) Is f continuous wherever the series converges?

(d) Is f bounded?

[5 + 10 + 5 + 5 = 2

P.T.O

INDIAN STATISTICAL INSTITUTE

Mid-Semester Examination
First semester 2007–2008
B. Stat (First year)
Analysis I

Date: 10 September, 2007 — Maximum Marks: 60

Duration: 2 hours 30 minutes

Answer all questions.
State clearly any result that you use in your answer.

(1) Let X be any set and let $P(X)$ denote the power set of X. Prove that there exists a bijection between $P(X)$ and $\{0,1\}^X$. Hence show that the power set of the set of natural numbers is uncountable. 8+8

(2) Let $A = [1, 2]$ and $B = [-2, -1]$. Define the set AB by

$$AB = \{xy \in \mathbb{R} \mid x \in A \text{ and } y \in B\}.$$

Determine the supremum of the set AB. 8

(3) Let x be any real number. Prove that there exists a sequence of rational numbers $\{r_n\}$ which converges to x. Can we choose r_n strictly decreasing? Justify your answer. 5+5

(4) Let $\{x_n\}_{n=1}^{\infty}$ be a sequence of real numbers such that $\lim_{n\to\infty} x_{2n-1} = u$ and $\lim_{n\to\infty} x_{2n} = v$. Determine the limsup and the liminf of the sequence $\{x_n\}_{n=1}^{\infty}$. 8

(5) Discuss the convergence/divergence of the following power series:

$$\sum_{n=1}^{\infty} \frac{n^3}{3^n} z^n,$$

where z is any complex number. 8

(6) Let $\{a_n\}$ be a sequence such that $a_n > 0$ for all n. Suppose, the series $\sum_{n=1}^{\infty} a_n$ diverges. Show that

(a) $\sum_{n=1}^{\infty} \frac{a_n}{1+\not{n} a_n}$ diverges and

(b) $\sum_{n=1}^{\infty} \frac{a_n}{1+n^2 a_n}$ converges. 5+5

4. Let $\{f_n\}$ be a sequence of continuous functions on $[0, 1]$ decreasing pointwise to the constant function 0. Is it true that

$$\int_0^1 f_n(x)dx \longrightarrow 0 \text{ as } n \to \infty?$$

Briefly justify your answer. [10]

5. (a) Let $f(x) = (\pi - |x|)^2$, $x \in [-\pi, \pi]$. Compute the Fourier coefficients of f and show that

$$f(x) = \frac{\pi^2}{3} + \sum_{n=1}^{\infty} \frac{4}{n^2} \cos nx.$$

(b) Show, by any method, with full justification, that $\lim_{x\to 1^-} \sum_{n=1}^{\infty} \frac{x^n}{n^4} = \frac{\pi^4}{90}$.

[10 + 15 = 25]

6. Suppose $f : [-\pi, \pi] \to \mathbb{R}$ is such that the Fourier series for f converges to f at every $x \in [-\pi, \pi]$. Given that $f \equiv 0$ on $(-\pi, 0)$ and $f \equiv 2$ on $(0, \pi)$, can you determine the value of f at the points $-\pi$, 0 and π? Briefly justify your answer. [10]

7. Show that the function $f : [-1, 1] \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} x^2 \sin \frac{1}{x} & \text{if } x \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

is of bounded variation, while the function $g : [-1, 1] \to \mathbb{R}$ defined by

$$g(x) = \begin{cases} x \sin \frac{1}{x} & \text{if } x \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

is not of bounded variation. [7 + 8 = 15]

INDIAN STATISTICAL INSTITUTE
Second Semester Examination: 2007-08
B. Stat. I Year
Probability Theory II

Date: 12.05.08 Maximum Marks: 60 Duration: 3

Each question Carries 10 marks. Maximum you can score is 60.
You should give proper justifications for your arguments.

1. The life time, in some units, of an electronic item is a random vai density

$$f(x) = xe^{-x} \qquad x \geq 0$$

(i) What is the expected life time of the item.
(ii) What are the chances that the item works for at least 10 units?
(iii) Given that the item has been working for 10 units, what is the probability that it works for at least 10 more units?

2.(i) X_1, X_2 are independent uniform $(0,1)$. Put

$$Y_1 = \sqrt{-2\log X_1}.Cos(2\pi X_2) \qquad Y_2 = \sqrt{-2\log X_1}.Sin(2\pi X$$

Show that Y_1, Y_2 are independent standard normal.

(ii) X, Y are bivariate normal with Correlation ρ, means 0 and varianc that correlation between X^2 and Y^2 is ρ^2.

3.(i) X, Y are independent two sided exponential with parameter one. Find of $X - Y$.

(ii) Without any calculations, explain with justification, how to get densit from the above answer.

4. Show that the density of standardized χ^2 variable converges to the standard normal variable as the degrees of freedom increases to infinity

5. I pick a number p at random from $(0,1)$. Then I toss a Coin, with char p in a single toss, n times. Let X be the number of heads obtaine distribution of X. Given X, find the conditional distribution of p.

6. I pick a number X at random from $(0,1)$. Given $X = x$, you pick a number Y at random from $(x-1, x+1)$. Find the conditional distribution of X given $Y = y$, explain for which values y of Y this is defined. Find $E(X|Y = y)$. Verify $E(E(X|Y)) = E(X)$.

7. A bus travels between two cities A and B which are 100 miles apart. If the bus has a breakdown, the distance from Breakdown to city A has uniform distribution over $(0,100)$. There are bus service stations in city A, city B and in the center of the route between A and B. It is suggested that it would be more efficient to have the three stations located 25, 50 and 75 miles respectively from A. Do you agree? Justify your answer.

***

INDIAN STATISTICAL INSTITUTE

Final Examination, 2st Semester, 2007-08

Statistical Methods II, B.Stat 1st Year

Date: May 15, 2008 Time:3 hours

This paper carries 55 marks. Attempt all questions. The maximum you can score is 50.

1. (a) Consider data on three variables X_1, X_2 and X_3. Suppose X_1 is regressed on X_2 using least squares. When the residuals of this regression are regressed on X_3 using least squares, the slope of the regression line is β_1. If X_1 is regressed on X_2 and X_3 simultaneously using least squares, the regression coefficient corresponding to X_3 is β_2. Show that $|\beta_1| \leq |\beta_2|$. When does equality occur?

   (b) Suppose X_1, X_2 and X_3 are variables satisfying a linear relationship $2X_1 + 3X_2 + 4X_3 + 5 = 0$. Compute the partial correlation of X_2 and X_3 eliminating the linear effect of X_1. [6 + 6]

2. (a) Suppose the correlation matrix of $(X_1, X_2, \ldots, X_p)$ is $R = ((r_{ij}))$ such that $r_{ij} = \rho^{|i-j|}$; $i, j = 1, 2, 3, \ldots, p$. Compute the multiple correlation of X_1 and $(X_2, X_3, \ldots, X_p)$.

   (b) Suppose we wish to predict fasting glucose levels (X_1) using a linear function of HbA1C (X_2), BMI (X_3) and HDL (X_4) levels, but resources may not permit collection of data on all three factors. Hence, it is decided to collect data on the factors in the order of their importance in predicting X_1. From some preliminary data, the correlation matrix of (X_1, X_2, X_3, X_4) is found out to be:

$$\begin{pmatrix} 1 & 0.58 & 0.71 & 0.37 \\ & 1 & 0.43 & 0.35 \\ & & 1 & 0.76 \\ & & & 1 \end{pmatrix}$$

   Determine in which order you should collect data on the three factors? [4 + 4]

3. (a) Jet Airways uses five different aircrafts in the sector Kolkata-Mumbai. What is the expected number of flights one has to take in this sector so that one has flown on all the five aircrafts?

P.T.O

(b) Assuming that the duration of the television advertisements in between the overs of a twenty-twenty cricket match is distributed as normal with mean 38 seconds and standard deviation 6 seconds, obtain the variance of the duration of those advertisements with duration greater than 45 seconds.

(c) What is the probability that a particular unit in a population is included in a sample drawn from the population using SRSWR? Suppose two independent samples, of sizes n_1 and n_2, are drawn from a population using SRSWR such that the means of the two samples are $\overline{x_1}$ and $\overline{x_2}$, respectively. If a is any real number, show that $a\overline{x_1} + (1-a)\overline{x_2}$ is an unbiased estimator of the population mean. For what value of a does the above estimator have the minimum variance? [4 + 6 + 6]

4. (a) Using random observations from $U(0,1)$, explain, with suitable justification, how you can generate an observation from the density:

$$f(x) = \frac{1}{\lambda} exp\{-|x-\theta|/\lambda\}; \ \lambda > 0; \ -\infty < x < \infty.$$

(b) Suppose it is analytically very difficult to compute the distribution function of a continuous variable X with density g. Explain how you can estimate the value of $P(a \le X \le b)$ using simulations. [6 + 4]

5. The following data pertain to the number of credit cards owned by 100 randomly chosen individuals:

Number of credit cards	Number of individuals
0	17
1	29
2	28
3	14
4	7
5	3

Fit a suitable probability distribution to the above data and test for its goodness of fit. [9]