INDIAN STATISTICAL INSTITUTE
B.STAT-I (2008-09)
Theory of Probability and Applications - I (Mid-Smester Test)
Max. marks:35 Time:3 hours.

Date: 1.9.2008

Note: ***Answer as many questions as you can.The whole question paper carries 45 marks. The maximum you can score is 35.***

1. (a)Verify that

$$(\bigcup_1^\infty A_n)\Delta(\bigcup_1^\infty B_n) \subset \bigcup_1^\infty (A_n \Delta B_n).$$

(b) Show that if $A\Delta B = A\Delta C$, then $B = C$.

(c) For the subsets A_n, $n = 1, 2, \cdots$ of the real line :

$$A_n = \begin{cases} (-\frac{n-1}{n}, \frac{n+1}{n}] & \text{if } n \text{ is odd} \\ [-\frac{n+1}{n}, \frac{n-1}{n}) & \text{if } n \text{ is even} \end{cases}$$

find lim supA_n and lim infA_n.

(d) Prove that If $A_n \uparrow A$, then, $\lim_{n\to\infty} P(A_n) = P(A)$.

(e) Show that , for $m \geq n$, the number of functions $f : \{1, 2, \cdots, m\} \mapsto \{1, 2, \cdots, n\}$ which are onto is

$$\sum_{k=0}^{n} (-1)^{n-k} \binom{n}{k} k^m.$$

[3 × 5]

2. (a) A die is cast until <u>either</u> a ***three*** <u>or</u> a ***six*** appears at which point the rolling is stopped. Describe explicitly the sample space Ω of this random experiment. If the die is fair how should one attach probabilities to the different points of Ω ?

(b) Find the probability that in 8 tosses of a fair coin there are at least 4 consecutive heads.

(c) From an urn 10 balls numbered 1,2,⋯,10, five balls are drawn at random without replacement. Find the probability that the second largest number drawn is 8.

(d) An airport shuttle bus with 6 passengers makes 4 scheduled stops.Assuming that all possible distributions of 6 passengers in the 4 stops are equally likely find the probability that someone (at least one) gets down at each stop.

(e) Six symmetric dice are rolled. Find the probability that the number of 1's minus the number of 2's will be 3.

(f) In a sequence of independent Bernoulli trials with probability of success p find the probability a successes will occur before b failures.

(g) A hat contains 10 coins, 4 of which are fair coins and the other 6 are biased to land *heads* with probability $\frac{2}{3}$. A coin is drawn at random from the hat and tossed twice. The first time it lands *heads* and the second time *tails* . Given this information what is the conditional probability that the coin is a fair coin ?

(Please turn overleaf)

(h) At a certain hill-station there are 35 hotels. It is found by the tourist bureau that if there are k tourists arriving in a season, the 35 hotels are occupied as if k *indistinguishable* balls are being placed at random (i.e.all distributions equally likely) in 35 numbered boxes. If this theory is correct and 100 tourists arrive in a season what is the probability that no hotel is left completely vacant ?

(i) 3 players A, B, C toss coins independently.The coins tossed by A, B and C turn heads with probabilities $\frac{1}{3}$, $\frac{1}{2}$ and $\frac{2}{5}$ respectively. If one person gets a different outcome than the other two he is the odd man out. If there is no odd man out the first time the players toss again and continue to do so until they get an odd man out. Find the probability that A will be the odd man out. [3 × 9]

3. When do you say that the events $A_1, A_2, A_3, \cdots, A_n$ are independent. Prove that if $A_1, A_2, A_3, \cdots, A_n$ are mutually independent then :

$P(A_1^c A_3 A_7^c A_2) = (1 - P(A_1))P(A_3)P(A_2)(1 - P(A_7))$ [1+2]

INDIAN STATISTICAL INSTITUTE LIBRARY

★ Need membership card to check-out documents

★ Books are not to be loaned to anyone

★ Loans on behalf of another person are not permitted

★ Damaged or lost documents must be paid for

INDIAN STATISTICAL INSTITUTE

Mid Semester Examination, 1st Semester, 2008-09

Statistical Methods I, B.Stat 1st Year

Date: September 3, 2008 Time: 2 hours

This paper carries 35 marks. The figures given in parentheses after each question denote the marks alloted to the question.

1. (a) A study showed that 93% of the people suffering from hypertension who took Telma H had normal blood pressure levels within a week. Do you think that the study design was appropriate to make conclusions on the effectiveness of Telma H? Explain.

   (b) In order to validate the hypothesis that morning flights are delayed more often than evening flights, the performances of 10 morning flights and 12 evening flights were studied for a month. It was observed that the mean of the number of delayed flights was 8 in the morning and 3 in the evening. What should one be careful about in this study design before concluding that the hypothesis has been validated by the observed data?

   (c) Show that the variance of a set of observations $\{x_1, x_2, \ldots, x_n\}$ can be expressed as $\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j<i}(x_i - x_j)^2$. [3+3+3]

2. The following frequency distribution pertains to starting salaries of a randomly chosen set of call centre personnel:

Salary	Frequency
<3000	8
3000-4000	25
4000-5000	32
5000-6000	23
6000-7000	18
7000-8000	10
>8000	4

Compute a suitable measure of spread for the above data. [7]

P.T.O

3. Show that, no matter how much each student scores in this particular question, the absolute difference between the mean and the median scores for this question cannot exceed 3. State clearly any result you may use, but you need not prove it if done in class. [6]

4. (a) Consider a frequency distribution with $(2n + 1)$ class intervals: $a_0 - a_1, a_1 - a_2, \ldots, a_{2n} - a_{2n+1}$ each having the same width. Suppose the frequencies of these classes are $f_1, f_2, \ldots, f_{2n+1}$ such that $f_i = f_{2n+2-i}$ for $i = 1, 2, \ldots, n$; $f_1 < f_2 < f_3 < \ldots < f_{n+1}$ and f_{n+1} is even. Show that, for the above frequency distribution, mean = median = mode. Can you suggest minor modifications in the above conditions on the frequencies such that:

   i. mean = median $\neq$ mode

   ii. mean $\neq$ median = mode

   (b) Consider the durations of the men's singles matches in Wimbledon 2008. Explain, with suitable justification, whether both the mean and the variance of the durations of these matches can remain unchanged if:

   i. the final match is removed from the data

   ii. the two semi-final matches are removed from the data (5+2)+6]

INDIAN STATISTICAL INSTITUTE

Mid-semester Examination: (2008-2009)

B. STAT (FIRST YEAR)

Computing Techniques And Programming I

5th September 2008. Full Marks: 30. Duration: Two hours.

Note: Answer all the questions.

1. (i) Convert the following with the mathematical proofs/reasons for the conversion. [6]

   (a) (1101.1011) in binary to decimal.

   (b) (998.375) in decimal to binary.

   (c) (5423.76) in Octal to Hexadecimal.

   (ii) Design the General C language codes for (a) and (b) above. Provide the necessary flowcharts and documentations for each part of your programs for whole numbers as well as fractional parts. [6+6]

2. The set of three ways of organizing 96 bits of main memory is as follows:

   (a) There are six locations, the content of each location is 16 bits.

   (b) There are eight locations, the content of each location is 12 bits.

   (c) There are twelve locations, the content of each location is 8bits.

   Estimate the possible sizes of Memory Address Register and Memory Data Register in each of the above cases. [3]

3. Why 2's complement representation is preferred than using signed magnitude representation? You have to establish your reasons. [4]

4. Give the C program to test in all possible ways a given magic square 3 x 3 with full documentation. [2]

5. z is expressed as Boolean expression (complement a)(complement b)c + a(complement b)(complement c) + a(complement b)c +ab(complement c) + abc.

   (i) Simplify the expression (where + denotes OR) to the minimum extent possible.

   (ii) How many logic gates would you require to implement the minimum expression if you are given three input lines as noncomplemented a , noncomplemented b and noncomplemented c . [1 + 2]

INDIAN STATISTICAL INSTITUTE

Mid-semester Examination: (2008-2009)

B. STAT. I year

Vectors and Matrices I

Date: 8 Sep. 2008. Maximum Marks: 30 Duration: 3 Hrs.

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

. Let V be a vector space over the field of complex numbers. Then

(a) Define *innerproduct* and *norm* on V and show that the positive sqare root of inner product of a vector with itself is a valid norm. .

(b) Let S be a subspace of V and consider any subspace T of V such that $S \oplus T = V$. Then show that every vector v of V can be expressed uniquely as sum of two vectors s in S and t in T. Also show that the norm of the vector t is minimum when T is chosen as the orthogonal compliment of S. (Here the norm is the one induced by the inner product under which the orthogonality is considered.)

(c) State and prove $Cauchy - Schwarz\ inequality$.

$[4 + 4 + 2 = 10]$

Let A be a *nonnull* matrix of order $m \times n$ and *rank* r. Then

(a) Define and derive a $rank\ factorization$ of A.

(b) Define and derive $normal\ form$ of A..

(c) If B is another matrix of order $m \times p$ such that column space of B is same as column space of A then $B = AC$ for some full rank matrix C of appropriate order.

$[4 + 2 + 4 = 10]$

Prove or disprove the following:

(a) Let $A_{m \times n}$ and $B_{n \times m}$ be matrices such that $rank(AB) = rank(BA) = min\{rank(A), rank(B)\}$. Then AB is *idempotent* implies BA is also *idempotent*.

(b) A is an *idempotent matrix* implies A has a *generalized invese* G such that *row space* of G is same as *row space* of A.

(c) Let $A = BC$ be a $rank\ factorization$ of the matrix A. The CB is nonsingular implies rank of A is equal to rank of A^2..

(d) $rank(A^*) = rank(A)$.

(e) There exist matrices A and B such that $AB - BA = I$.

$[5 \times 2 = 10]$

$\cdots xXx \cdots$

INDIAN STATISTICAL INSTITUTE
Mid-semester Examination : 2008-09
B. Stat. - First Year
Analysis I

Date : 12. 09. 2008 | Maximum Score : 40 | Time :3 Hours

1. This paper carries questions worth a total of 45 marks. Answer as much as you can. The maximum you can score is 40 marks.
2. You are free to use any theorem proved in the class. However, you must state any theorem that you use at least once in the answer script.
3. We shall use $\mathbb{Z}$ to denote the set of all integers, $\mathbb{R}$ the set of all real numbers and $\mathbb{C}$ the set of all complex numbers.

(1) A real number $r \in \mathbb{R}$ is called *transcendental* if there is no polynomial $P(x) \in \mathbb{Z}[x]$ with integer coeficients such that $P(r) = 0$. Let Θ denote the set of all transcendental numbers. Show that

$$|\Theta| = |\mathbb{R}|.$$

[5]

(2) For an integer $p > 1$, set $\mathbb{Z}_p = \{0, 1, \ldots, p-1\}$. For $a, b \in \mathbb{Z}_p$, $a \neq 0$, let $r_a(b)$ denote the remainder term of $\frac{a \cdot b}{p}$. Show the following.
  (a) For each $a \in \mathbb{Z}_p \setminus \{0\}$, the map $r_a : \mathbb{Z}_p \to \mathbb{Z}_p$ is an injection if and only if p is prime.
  (b) If p is prime, show that for every $a \in \mathbb{Z}_p \setminus \{0\}$ there is a unique $b \in \mathbb{Z}_p \setminus \{0\}$ such that p divides $1 - a \cdot b$.

[3+ 2]

(3) For every real number a, show that there is a unique real number b such that $b^3 = a$.

[8]

(4) Assume that the sequence $\{n^{-1-1/n}\}$ is decreasing. Is the series

$$\sum n^{-1-1/n}$$

convergent or divergent? Justify your answer.

[6]

(5) Let $\{a_n\}$ be a sequence of positive real numbers such that $\lim \frac{a_{n+1}}{a_n} = l$. Show that $\lim a_n^{1/n} = l$.

[8]

P. T. O

(6) For any complex number $z \in \mathbb{C}$, set

$$Exp(z) = \sum_{0}^{\infty} \frac{z^n}{n!}.$$

Show the following.

(a) For every complex number $z \in \mathbb{C}$, the series

$$\sum_{0}^{\infty} \frac{z^n}{n!}$$

is absolutely convergent.

(b) For complex numbers, z_1 and z_2,

$$Exp(z_1 + z_2) = Exp(z_1) \cdot Exp(z_2).$$

(a) For any $m \in \mathbb{Z}$,

$$Exp(m) = e^m.$$

[5+5+3]

INDIAN STATISTICAL INSTITUTE

First-semester Examination: (2008-2009)

B. STAT. I year

Vectors and Matrices I

Date: 24 Nov., 2008. Maximum Marks: 60 Duration: 3 Hrs.

ote: (i) All the matrices and vectors considered are over complex field unless otherwise ated. (ii) Class room notation is used. $R(A)$ and $C(A)$ denote the row space and the lumn space of the matrix A. (iii) State clearly all the results used.

Let S and T be subspaces of R^n such that $S \oplus T = R^n$. Then

(a) Define projection of R^n into S along T.

(b) Show that the above projection is a linear transformation.

(c) Derive a set of necessary and sufficient conditions for a marix P to represent the above linear transformation.

(d) Show that the orthogonal projector of R^n into S is symmetric.

$[2 + 2 + 5 + 2 = 11]$

2 (a) Define a *minimum norm solution* of a consistent system of linear equations $Ax = b$.

(b) Define a *minimum norm g-inverse*, A_m^- of a matrix A.

(c) Show that G is a A_m^- if and only if $GAA^* = A^*$.

(d) Show that A_m^- is not unique.

(e) Show that minimum norm solution is unique.

$[1 + 2 + 4 + 1 + 2 = 10]$

3 (a) If A is a positive definite matrix then show that $|A| \leq a_{ii}A_{ii}$ for all i and equality occurs if and only if A is diagonal, where A_{ii} is the cofactor of the i^{th} diagonal element a_{ii} of the matrix A.

(b) For any real square matrix B of order n such that $|b_{ij}| \leq 1$ for all i and j, show that $|B|^2 \leq n^n$ and equality occurs if and only if $b_{ij} = \pm 1$ and the rows of B are pairwise orthogonal.

$[5 + 5 = 10]$

P.T.O.

4 Prove or disprove the following:

(a) If $ABC = ABD$ and $rank(AB) = rank(B)$ then $BC = BD$.

(b) If A and B are square matrices of order n then

$$rank \begin{pmatrix} A & I \\ I & B \end{pmatrix} = n \text{ if and only if } B = A^{-1}.$$

(c) If $C(A) \subseteq C(B)$ and $R(A) \subseteq R(D)$ then $A = BCD$ for some matrix C.

(d) Let b be a nonzero real vector of order $n \times 1$ and $M = bb'$.Then there exists an orthogonal matrix Q such that QMQ' is a diagonal matrix.

(e) Let $A = BC$ be a *rank factorization* of the matrix A. Then any reflexive g-inverse G of A is of the form $C^{-R}B^{-L}$ where C^{-R} is a right inverse of C and B^{-L} is a left inverse of B.

(f) Let T be a block triangular matrix given by

$$T = \begin{pmatrix} A & B \\ 0 & C \end{pmatrix}$$

where A is nonsingular. Then $rank(T) = rank(A) + rank(C)$.

(g) If $rank(A) = rank(A^2)$ then A has a nonzero principal minor of order r.

(h) If $C(A) = R(A)$ then $rank(A) = rank(A^2)$. However, the converse is not true.

(i) Let A be a square matrix of order n and b and c are vectors of order $n \times 1$. Then $|A + bc'| = |A| \; |1 + c'A^{-1}b|$.

(j) Let S is a subspace of R^n such that $d(S) = s$. Then AA' is the orthogonal projector of R^n into S where the columns of A form an orthonormal basis of S.

[10 × 3 = 30]

$\cdots xXx \cdots$

INDIAN STATISTICAL INSTITUTE
First Semestral Examination : (2008-2009)
B. Stat. - First Year
Analysis I

Date : 28. 11. 2008 | Maximum Score : 100 | Time :4 Hours

1. This paper carries questions worth a total of 116 marks. Answer as much as you can. The maximum you can score is 100 marks.
2. Unless otherwise stated, $\mathbb{R}$ will denote the space of all real numbers and all functions are real-valued.

(1) (a) Determine all reals x such that the series

$$1 + 2x + 3x^2 + \ldots + (n+1)x^n + \ldots$$

is convergent.

(b) A function $f(x)$ is called a *rational function* if it is of the form $\frac{P(x)}{Q(x)}$, P, Q polynomials (with real coefficients). Show that $x \to e^x$ is not a rational function.

[6 + 10]

(2) (a) Show that every one-to-one, continuous function defined on an interval is stricly monotone.

(b) Let $A \subset \mathbb{R}$ be countable. Show that there is a map $f : \mathbb{R} \to [0,1]$ such that f is continuous at x if and only if $x \in \mathbb{R} \setminus A$.

(c) Let $A \subset \mathbb{R}^n$ be such that $1 < |A| < |\mathbb{R}|$. Show that A is not connected.

[12 + 10 + 10]

(3) (a) Let $A \subset \mathbb{R}$ be compact and $\mathcal{U}$ an open cover of A. Show that there is an $\epsilon > 0$ such that for every $x, y \in A$,

$$|x - y| < \epsilon \Rightarrow \exists U \in \mathcal{U}(x, y \in U).$$

(b) Let A be a bounded subset of $\mathbb{R}^n$ such that for every $\overline{x} \in \mathbb{R}^n$ there is an open set $U \ni \overline{x}$ such that $A \cap U$ is finite. Show that A is finite.

[10 + 8]

(4) (a) Find

$$\lim_{x \to \infty} x(\sqrt{x^2 - 1} - x).$$

(b) Determine the right-angled triangles with parameter S which has the maximum area.

P.T.O

INDIAN STATISTICAL INSTITUTE
B.STAT-I (2008-09)
Probability Theory - I
Semestral-I examinations
Maximum marks: 65. Time: 3 hours.

Date : 1 Dec, 2008.

Note: ***Answer as many questions as you wish. The whole question paper carries 75 marks. The maximum you can score is 65.***

1. (a) Show that $S_1 \geq P(\bigcup_1^n A_i) \geq S_1 - S_2$, where $S_1 = \sum_1^n P(A_i)$ and $S_2 = \sum_{1 \leq i < j \leq n} P(A_i A_j)$.

(b) A sample of n tickets are drawn at random, one by one with replacement, from a box containing N tickets numbered $1, 2, \cdots, N$. Find the probability p_r that there are exactly r distinct ticket numbers in the sample, $r = 1, 2, \cdots, N$.

[5+5]

2. (a) A die is thrown until a *six* turns up. Given that the number of throws needed is even, what is the probability that the number of throws needed is 8?

(b)Consider 2 urns U_i containing r_i red and b_i black balls , $i = 1, 2$. A ball is drawn at random from U_1 , put into U_2 and then a ball is drawn from U_2 and put into U_1. After all this what is the probability that a ball drawn from U_1 now is red ? Verify that if $b_1 = r_1$ and $b_2 = r_2$, then this probability is the same as when no transfer has taken place.

(c) Let $A_1, A_2, \cdots, A_n$ be independent events each having the same probability p, $(0 < p < 1)$. Let B denote the event that ***exactly one*** among the A_i' s occur and C denote the event that ***exactly two*** among the A_i' s occur. If $P(B) = P(C)$, find p.

[5 × 3]

3. (a) Let X and Y be independent random variables having a common negative binomial distribution : $NB(\alpha, p)$. Show that

$$P[X = j \mid X + Y = k] = \frac{\binom{\alpha+j-1}{j}\binom{\alpha+k-j-1}{k-j}}{\binom{2\alpha+k-1}{k}}, \quad k = 0, 1, 2, \cdots, \text{ and } j = 0, 1, \cdots, k.$$

(b) Let $X_1, X_2, \cdots, X_m$ be independent non-negative integer-valued random variables all having the same distributions given by

$$P[X_1 = n] = p_n, \quad n = 0, 1, 2, \cdots.$$

Let $r_n = \sum_{j=n}^{\infty} p_j$ and let $Y = \text{Minimum}(X_1, X_2, \cdots, X_m)$. Show that

$$E(Y) = \sum_{n=1}^{\infty} r_n^m.$$

[3+5]

P. T.O

[8 + 6]

(5) (a) State and prove Rolle's Theorem.

(b) Establish the Taylor's theorem: *Let $f : (a, b) \to \mathbb{R}$ be a n-times differentiable theorem and $a < c < b$. Show that for every $a < x < b$ there is a ξ between c and x such that*

$$f(x) = f(c) + \sum_{i=1}^{n-1} \frac{f^{(i)}(c)}{i!} \cdot (x - c)^i + \frac{f^{(n)}(\xi)}{n!} \cdot (x - c)^n.$$

(c) Let $f : [a, b] \to \mathbb{R}$ be a continuous map which is twice differentiable in (a, b) and $f^{(2)}(x) < 0$ for every $a < x < b$. Show that f is strictly convex, i.e., for every $a \leq x < y \leq b$ and every $0 < t < 1$,

$$f(tx + (1 - t)y) < t \cdot f(x) + (1 - t) \cdot f(y).$$

[12 + 12 + 12]

INDIAN STATISTICAL INSTITUTE

Final Examination, 1^{st} Semester, 2008-09

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

Date: **9.1.09** Time: $3\frac{1}{2}$ hours

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

1. (a) Suppose that a set of observations $\{x_1, x_2, \ldots, x_n\}$ is represented by a vector $\mathbf{x}$. Show that the variance of this set can be expressed as $\frac{1}{n}\mathbf{x}'H\mathbf{x}$ where H is an idempotent matrix and $r(H) = n-1$.

   (b) For a set of observations, the coefficient of variation (CV) is defined as the ratio of the standard deviation to the absolute value of the mean expressed as a percentage. Can you suggest an advantage and a disadvantage that this measure has over the usual measures of spread? [5+3]

2. Consider a set of bivariate observations $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$. Obtain the limiting value of the correlation coefficient between x and y when:

   (a) $x_n \to \infty$ while other $x_i s$ are unchanged,

   (b) $x_n \to \infty$ and $y_n \to \infty$ while other $x_i s$ and $y_i s$ are unchanged,

   (c) $x_{n-1} \to \infty$ and $y_n \to \infty$ while other $x_i s$ and $y_i s$ are unchanged. [3+3+3]

3. (a) Based on a set of observations on x, suppose you want to obtain a measure of location by minimizing the expression: $\sum min\{(x_i-\theta)^2, |x_i-\theta|\}$ with respect to θ. Formulate the above minimization as a weighted least squares minimization problem and describe an iterative algorithm to estimate θ. Show all your computational steps clearly. A desirable property of your measure of location

P.T.O

4. In a sequence of independent Bernoulli trials, let X be the length of the run (of either S or F) started by the first trial and let Y be the length of the second run - for example if the first few outcomes are SFS then $X=1$ and $Y=1$, if $SSSFFS$ then $X=3$ and $Y=2$, and so on.

   (a) Find the joint p.m.f of (X,Y).

   (b) Find $E(X), E(Y), Var(X), Var(Y)$ and $Cov(X,Y)$. [3+9]

5. Let $X_1, X_2, \cdots$ be independent random variables with common distribution given by::

$$P(X_i = 1) = p \text{ and } P(X_i = -1) = 1-p,\ i = 1, 2, \cdots. 0 < p < 1.$$

Let $S_0 = 0$, and $S_k = X_1 + X_2 + \cdots + X_k$, $k = 1, 2, \cdots$. Let

$$\phi_0^{(r)} = 0 \text{ and}$$

$$\phi_n^{(r)} = P(S_1 \neq r, S_2 \neq r, \cdots, S_{n-1} \neq r, S_n = r),\ n = 1, 2, \cdots, \text{ and } r = 1, 2.$$

$$\Phi^{(r)}(s) = \sum_{n=0}^{\infty} \phi_n^{(r)} s^n,\ |s| < 1,\ r = 1, 2.$$

   (a) Show that (i) $\phi_n^{(2)} = \phi_1^{(1)}\phi_{n-1}^{(1)} + \phi_2^{(1)}\phi_{n-2}^{(1)} + \cdots + \phi_{n-1}^{(1)}\phi_1^{(1)}$, for $n = 1, 2, 3, \cdots$.

   (ii) $\phi_1^{(1)} = p$ and $\phi_n^{(1)} = q\phi_{n-1}^{(2)}$, for $n = 2, 3, \cdots$

   (b) Use (a) to show that $\Phi^{(1)}(s) = \frac{1-(1-4pqs^2)^{\frac{1}{2}}}{2qs}$, where $q = 1-p$. Compute $\phi_n^{(1)}, \forall n$. [3+12]

6. Let $X_1, X_2, X_3, \cdots$, be independent and identically distributed discrete random variables such that $E(|X_1|) < \infty$, and $E(X_1) = 0$. Let the random variables $U_i^{(n)}$ be defined as follows:

$$U_i^{(n)} = \begin{cases} X_i & , |X_i| \leq n \\ 0 & , |X_i| > n \end{cases}$$

$i = 1, 2, \cdots$, and $n = 1, 2, \cdots$.

Prove that, for every fixed $\epsilon > 0$, $P[|\frac{U_1^{(n)} + U_2^{(n)} + \cdots + U_n^{(n)}}{n}| > \epsilon] \to 0$, as $n \to \infty$. [15]

----

should be that it lies in between the lowest and the highest observations. Verify whether your algorithm ensures that this property is satisfied.

(b) Consider a set of bivariate observations on (x,y) such that there are no ties in either of the variables. If Kendall's τ for the above set of data is -1, show that Spearman's rank correlation is also -1. [6+4]

4. Consider two sets of bivariate data on (x,y). The least squares linear regression lines of y on x are identical for the two sets of data.

(a) If the correlation coefficients between x and y for the two sets of data are equal, explain whether the least squares linear regression lines of x on y will be identical for the two sets of data.

(b) If the two sets of data are pooled, explain whether the least squares linear regression line of y on x remains unchanged. [4+4]

5. The following data pertain to the time taken (T_1) for a particular chemical reaction (in minutes) and the final temperature (T_2) of the product (in degrees Fahrenheit):

T_1	T_2
49.0	101.6
53.8	107.3
55.5	115.1
43.1	91.5
49.3	102.4
44.6	94.3

(a) Comment on the skewness of T_2 in the above data.

(b) Based on a least absolute deviation linear regression of T_2 on T_1, predict the final temperature of the product if the chemical reaction took 52 minutes to complete.

(c) What proportion of the variance of T_2 is not explained by a least squares linear regression of T_2 on T_1? Explain whether this proportion is equal to the proportion of variance of T_1 not explained by a least squares linear regression of T_1 on T_2. [5+8+7]

INDIAN STATISTICAL INSTITUTE

Final-semester Examination

B.STAT (FIRST YEAR)

Computing Techniques And Programming I

03/12/2008 Full Marks: 100 Duration: Three hours

Note: Answer all the questions

1. (a)Considering the message 1101011011 and the generator polynomial as represented by the binary string 10011, calculate the CRC code which has to be appended with the message. [6]

(b) RAM chips of 256word x 8 bits capacity are used to design a memory segment of size 1kilo word x 16 bits.
   (i) How many such RAM chips are required for the segment?
   (ii) How many address bus lines are needed?
   (iii) What is the size of the data bus? [3]

(c) What is a half subtractor? Starting from truth table derive a full subtractor circuit in terms of two half subtractors and an OR gate. [2+4]

2. (a) Starting from the truth table definition, derive the Boolean expressions of all the 3-variable linear Boolean functions. (You should try to simplify all the expressions in terms of only EX-OR logic.).

(b) What is an ODD/EVEN function? Explain their significance. [8 + 7]

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). [10 + 5]

4. Illustrate the following programming techniques to swap two integer variables with appropriate C code :
   a. Using only two variables.
   b. Using three variables.
   c. Using 'Swap' function call
   d. Using Exclusive-OR Boolean Logic. [4 +3 +3 + 5]

P.T.O

5. (a) " Most often bubble sort can be speeded up by having successive passes in opposite directions"- Explain and illustrate the above statement.

(b) Give the C code for both the simple bubble sort and speeded up bubble sort with full documentations.

[7 +4 +4]

6. 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. It is called upper triangular if all elements below diagonal are zero and lower triangular if all elements above the diagonal are zero. Write a C program that reads a square matrix and determines if it is a diagonal or upper triangular or lower triangular. (You should illustrate all the three cases with full documentation.)

[5 + 5 +5]

7. Write the C program for creating a simple linked list of several integers 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]

INDIAN STATISTICAL INSTITUTE
B.STAT-I (2008-09)
Probability Theory - I
Semestral-I (supplementary) examinations
Maximum marks: 65. Time: 3 hours.

Date : 6.2.09

Note: ***Answer as many questions as you wish. The whole question paper carries 72 marks. The maximum you can score is 65.***

1. (a) Show that $\limsup_n (A_n \cup B_n) = (\limsup_n A_n) \cup (\limsup_n B_n)$.

Let

$$A_n = \begin{cases} (1 - \frac{1}{n}, 2) & \text{if } n \text{ is odd} \\ (1 + \frac{1}{n}, 2) & \text{if } n \text{ is even} \end{cases}$$

Find $\limsup A_n$ and $\liminf A_n$

(b) Let $A_1, A_2, A_3, \cdots, A_n$ are n events. Let

$$S_k = \sum_{1 \le i_1 < i_2 < \cdots i_k \le n} P(A_{i_1} A_{i_2} \cdots A_{i_k}) \qquad k = 2, 3, \cdots, n.$$

Let B_m denote the event that exactly m, $1 \le m \le n$, among the n events $A_1, A_2, A_3, \cdots, A_n$ occur. Show that

$$P(B_m) = S_m - \binom{m+1}{m} S_{m+1} + \binom{m+2}{m} S_{m+2} - \cdots + (-1)^{n-m} \binom{n}{m} S_n.$$

[(3+2)+5]

2. (a)Given that a throw of ten dice produced at least one ace, find the conditional probability of two or more aces.

(b) A throws six dice and wins if he scores at least one ace. B throws 12 dice and he wins if he throws at least two aces. Who, among A and B, has greater probability to win?

(c) Die A has four red and two white faces, whereas die B has two red and four white faces. A coin is flipped once. If it falls heads, die A is rolled 10 times and it falls tails die B is rolled 10 times. Given that the first two throws resulted in red what is the probability of red at the the third throw? Given that red turns up in all the ten throws what is the probability that die A is being used?

(d) Ten pairs of shoes are in a shoe box. Four shoes are chosen at random from the box. What is the probability that there is a [at least one] complete pair among the four chosen?

[4+4+(4+4)+4]

3. A number X is chosen at random from the set $\{1, 2, 3, 4, 5, 6, 7\}$, each choice being equally likely. After choosing X, a number Y is chosen at random from the set $\{1, 2, \cdots, X\}$.

(a) Find the joint probability mass function of X, Y.

(b) Find the conditional p.m.f of X given $Y = 4$.

(c) Are X and Y independent ? Give reasons.

[6+3+3]

P.T.O

4. An urn contains 1 ball numbered 1, 2 balls numbered 2, 3 balls numbered 3 and 4 balls numbered 4 — 10 balls in all with various numbers 1,2,3 and 4 on them. We draw 5 balls without replacement from the urn. Let S denote the sum of all the numbers from among $\{1,2,3,4\}$ that did not appear in the sample. Find $E(S)$ and $Var(S)$. [10]

5. Let X and Y be independent random variables having the same geometric distribution $\mathcal{G}(p)$ distribution. Let $U = Min\{X,Y\}$ and $V = X - Y$. Show that U and V are independent. [10]

6. Let $X_1, X_2, \cdots$ be independent random variables with common distribution given by::

$$P(X_i = 1) = p \text{ and } P(X_i = -1) = 1-p,\ i = 1, 2, \cdots .\ 0 < p < 1.$$

Let $S_0 = 0$, and $S_k = X_1 + X_2 + \cdots + X_k$, $k = 1, 2, \cdots$. Let $u_n = P(S_n = 0)$, $n = 0, 1, 2, \cdots$ and $f_n = P(S_1 \neq 0, S_2 \neq 0, \cdots, S_{n-1} \neq 0, S_n = 0)$ $n = 1, 2, \cdots$ and $f_0 = 0$. Find the generating functions of the sequences $\{u_n\}$ and $\{f_n\}$. [10]

7. State and prove Cauchy-Schwarz inequality. [10]

INDIAN STATISTICAL INSTITUTE

Mid-semester Examination: (2008-2009)

B. STAT. I year

Vectors and Matrices II

Date: 23 Feb. 2009. Maximum Marks: 30 Duration: 3 Hrs.

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

1 (a) Show that any set of eigenvectors of a matrx A corresponding to a set of distinct eigenvalues is independent.

(b) A matrix A is called *normal* if $AA^* = A^*A$. Show that a normal matrix is unitarily similar to a diagonal matrix.

(c) Show that if A is real and *nnd* matrix and B is a real symmetric matrix satisfying $C(B) \subseteq C(A)$ then there exists a nonsingular matrix R such that both $R'AR$ and $R'BR$ are diagonal.

[3 × 5 = 15]

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

(b) Find the set of all the eigenvalues of the matrix $A_{5\times5}$ whose rank factorization is $A = BC$ where

$$B' = \begin{pmatrix} 1 & -1 & 1 & 0 & 1 \\ 1 & 1 & 1 & -4 & -3 \\ 1 & -1 & 1 & -1 & 0 \end{pmatrix}, \text{ and } C = \begin{pmatrix} 4 & 2 & -2 & 0 & 1 \\ 2 & 2 & 1 & 1 & 0 \\ -4 & 2 & 0 & -5 & 6 \end{pmatrix}$$

[2 × 10 = 20]

Prove or disprove the following:

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

(b) A real symmetric matrix A is *nnd* if and only if it has an *nnd* g-inverse.

(c) $rank(A) =$ the number of nonzero eigenvalues of A implies $rank(A) = rank(A^2)$.

(d) For a real symmetric matrix the set of nonzero eigenvalues is same as the set of singular values.

(e) A is nilpotent implies all the eigenvalues of A are zero. (A matrix is called nilpotent if $A^k = 0$ for some positive integer k).

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

[6 × 2 = 12]

$\cdots xXx \cdots$

INDIAN STATISTICAL INSTITUTE
Mid-semester Examination : 2008-09
B. Stat. - First Year
Analysis II

Date : 27.2.09 Maximum Score : 40 Time :3 Hours

1. This paper carries questions worth a total of 48 marks. Answer as much as you can. The maximum you can score is 40 marks.
2. You are free to use any theorem proved in the class. However, you must state any theorem that you use at least once in the answer script.
3. We shall use $\mathbb{R}$ the denote the set of all real numbers.

(1) (a) Let $f : [0,1] \to \mathbb{R}$ be a continuous function. Define

$$F(x) = \int_0^x f(y)dy, \ x \in [0,1].$$

Show that F is differentiable and find its derivative.

(b) For every positive integer n, show that there is a real-valued function on $[0,1]$ that is differentiable n-times but not $(n+1)$-times.

[6 + 6]

(2) Let $\alpha : [0,1] \to \mathbb{R}^n$ be a bounded Riemann-integrable function. Show that $t \to |\alpha(t)|$ is integrable and

$$|\int_0^1 \alpha(t)dt| \leq \int_0^1 |\alpha(t)|dt.$$

[6]

(3) Prove or disprove: Every bounded monotone function on $[0,1]$ is Riemann-integrable.

[6]

(4) Let $\{f_n\}$ be a sequence of real-valued continuous function on $[0,1]$ and for each n, there is a real m_n, such that $\forall x \in \mathbb{R}(|f_n(x)| \leq m_n)$ and $\sum_n m_n < \infty$. Show that

$$g(x) = \sum_n f_n(x), \ 0 \leq x \leq 1,$$

is continuous,

[6]

P.T.O

(5) (a) Let $f : [-1,1] \to R$ be an even continuous function such that

$$\int_{-1}^{1} x^n f(x) dx = 0$$

for all even nonnegative integer n. Show that f is identically equal to 0.

(b) Show that there is no sequence of polynomials $P_n(x) \in \mathbb{R}[x]$ such that $P_n(x) \to \sin(x)$ uniformly on $\mathbb{R}$.

[6 + 6]

(6) Let $\{f_n\}$ be an equicontinuous family of functions on a compact set K such that for every $x \in K$, $\{f_n(x)\}$ is convergent. Show that $\{f_n\}$ is uniformly convergent.

[6]

INDIAN STATISTICAL INSTITUTE

B.STAT-I (2008-09)

Probability - II (Mid-Semestral)

Max. marks:40 Time:$2\frac{1}{2}$ hours

Date: 2.3. 2009

Note: ***Answer all questions. The maximum you can score is 40.***

1. Let F and G be Distributtion Functions (D. F). Determine which among the the following are also D. F. Give reasons for your answer.
   (a) $H(x) = F(3x-5)$
   (b) $H(x) = 1 - F(-x)$
   (c) $H(x) = F(x)G(x)$
   (d) $H(x) = 1 - (1-F(x))(1-G(x))$

[2 × 4]

2. Consider the following distribution function $F(x)$:

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{x}{9} & \text{if } 0 \le x < 1 \\ \frac{(x-1)^2+1}{4} & \text{if } 1 \le x < 2 \\ \frac{1}{2} & \text{if } 2 \le x < 3 \\ 1 & \text{if } 3 \le x < \infty \end{cases}$$

Roughly sketch the graph of $F(x)$ and find the discontinuity points of F, for each discontinuity point a, find $F(a) - F(a-)$.

[3+2]

3. Suppose X follows $N(0,\sigma^2)$ .Write down the density of X . Let

$$Y = \begin{cases} -2X & \text{if } X < 0 \\ X^2 & \text{if } X \ge 0 \end{cases}$$

Find the density of Y.

[1 + 6]

4. Let X and Y be independent random variables each having the uniform $U(0,1)$ distribution. Find the density of $Z = X - Y$.

[6]

5. The speed V of a molecule in a uniform gas at equilibrium is a random variable. Its density is given by:

$$f(v) = \begin{cases} Cv^2 e^{-\frac{v^2}{100}} & \text{for } v > 0 \\ 0 & \text{otherwise} \end{cases}$$

Find the value of C. Find the density of the Kinetic energy $U = \frac{1}{2}V^2$

[2 + 6]

Please Turn overleaf

6. Let X and Y be independent random variables such that $X \sim U(0,1)$ and $Y \sim Be(2,1)$.
(a) Show that the density $h(u)$ of the random variable $U = X/Y$ at a point $u > 0$ is given by :

$$h(u) = \int_0^{min(1,\frac{1}{u})} y f(uy) g(y) dy$$

where f and g are the densities of X and Y respectively. Show that

$$h(u) = \begin{cases} \frac{2}{3} & \text{for } 0 < u \leq 1 \\ \frac{2}{3u^3} & \text{for } u > 1 \end{cases}$$

[2+4]

7. X and Y are independent random variables, each having the exponential distribution with parameter $\lambda = 2$. Compute

$$P(X + Y \leq 2,\ \tfrac{X}{Y} \leq \tfrac{1}{2}).$$

[5]

INDIAN STATISTICAL INSTITUTE
B. Stat I
Computational Techniques and Programming II
(Mid Semester Exam)

Date: 4.3.09 Time: 2 hours

Answer as many questions as you like. Maximum you can score is 30.

1. a) Round off the following numbers to 4 significant digits:
(i) 38.46235 (ii) 0.70099 (iii) 0.0022218 (iv) 2.36425 (v) 19.235101

b) (i) Define absolute error and relative error. Define floating point number.
(ii) Let e_1 be the error in the sum $z = x+y$, on the floating point addition of floating point numbers x and y. Let e_2 be the error in the floating point product $w = v \cdot z$ where w, v and z are floating point numbers and z is the sum defined as above. Compute the error in the arithmetic expression $s = (z+w) \cdot z \cdot w$. (assume that a floating point number on the computer is represented as a 4 digit decimal normalized mantissa obtained by truncation.)

or

Find the limiting absolute and relative errors of the volume of a sphere $V = \frac{1}{6}\pi d^3$ if the diameter $d = 3.7 \text{ cm} \pm 0.05 \text{ cm}$ and $\pi \approx 3.14$.

c) It is required to obtain the roots of $X^2 - 2X + \log 2 = 0$ to four decimal places. To what accuracy should log 2 be given? [2+4+3=9]

2. a) What is meant by round off error?
Let x and y be two real numbers within the range of floating-numbers and let $*$ Denote any of the operators +, -, ×, /, then show that

$$fl\,(x * y) = \frac{x * y}{(1+\delta)}$$

where δ satisfies

$$|\delta| \leq \frac{1}{2}\beta^{1-n} \quad \text{in case of rounding operation}$$

$$|\delta| < \frac{1}{2}\beta^{1-n} \quad \text{in case of chopping operation}$$

and $fl(x)$, $fl(y)$ are n β-digit floating point number.

b) How many terms of the series $\arctan X = X - \frac{X^3}{3} + \frac{X^5}{5} - \dots\dots\dots$ are required to compute arctan X accurate to six decimal places? [5+4=9]

3. a) Obtain Newton's formula for polynomial interpolation using divided difference in case of n unequally spaced arguments.

P.T.O

b) Derive the error term for Newton's divided difference formula. Find a bound for the error in linear interpolation. [4+3=7]

4. a) Find the number and position of the real roots of the equation $x^4 - 6x^3 + 10x^2 - 6x + 1 = 0$ using Sturm's method for location of roots.

b) Give an algorithm for fixed point iteration scheme and find the condition required for the convergence of the functional iteration scheme. [3+4=7]

5. a) Give an algorithm for Newton-Raphson iteration scheme and explain its geometrical interpretation. Derive the order of convergence of the scheme also.

b) From the following table find $f'(1.1)$

x	f(x)
1.1	2.0091
1.2	2.0333
1.3	2.0692
1.4	2.1143
1.5	2.1667

[4+3=7]

INDIAN STATISTICAL INSTITUTE

Mid Semester Examination, 2^{nd} Semester, 2008-09

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

Date: March 6. 2009 Time: 2 hours

This paper carries 35 marks.

1. (a) Suppose you wish to perform a linear regression of X_1 on $X_2, X_3, \ldots, X_p$ by minimizing $\sum_j |X_{1j} - \beta_1 - \sum_{i=2}^{p} \beta_i X_{ij}|$ with respect to $\beta_1, \beta_2, \ldots, \beta_p$. Describe a suitable algorithm to estimate $\beta_1, \beta_2, \ldots, \beta_p$. Show all your computation steps clearly and discuss the problems you may face in your computations.

   (b) Suppose the correlation matrix of $(X_1, X_2, \ldots, X_p)$ is $R = ((r_{ij}))$ such that $r_{ij} = \rho^{|i-j|}$; $i, j = 1, 2, 3, \ldots, p$; $-1 < \rho < 1$. Obtain the multiple correlation coefficient of X_p and $(X_1, X_2, \ldots, X_{p-1})$. If all the variables have mean 1 and equal variances, find the least squares linear regression line of X_p on $X_1, X_2, \ldots, X_{p-1}$. [5+5]

2. (a) Suppose $\Sigma_{a,c} = a1' + 1a' + cI$ where a is a vector and c is a scalar. Under what necessary and sufficient conditions on a and c is $\Sigma_{a,c}$ a valid dispersion matrix?

   (b) Suppose X_1, X_2 and X_3 are variables satisfying a linear relationship $3X_1 - 4X_2 + 2X_3 + 1 = 0$. Compute $r_{12.3}$ and $r_{13.2}$. [5+5]

3. Suppose that the number of times a machine becomes non functional during a day is distributed as Poisson.

   (a) It intuitively seems that an odd number of failures of the machine is as likely as an even number of failures. Show that, contrary to intuition, the chance of an even number of failures is more than that of an odd number of failures. Can you provide a suitable explanation for this apparent contradiction?

   (b) Suppose the percentage efficiency of the machine on a particular day is defined as $\frac{100}{1+X}$, where X is the number of failures in a day. Due to shortage in manpower during the weekend, the mean number of failures during any of the weekdays is half of that on a Saturday or a Sunday. If the number of failures during a week is 10, what is the expected percentage efficiency on the Wednesday of the week? State your assumptions clearly. [5+5]

4. Find the variance of the number of rolls of a perfect die required to get all the six faces. [5]

INDIAN STATISTICAL INSTITUTE
B.STAT-I (2008-09)
Probability Theory - I
Semestral-I (Backpaper) examinations
Maximum marks: 100. Time: 3 hours.

Date : 8.4.09

Note: ***Answer as many questions as you wish. The whole question paper carries 105 marks. The maximum you can score is 100.***

1. (a) Show that $A\Delta B = C \Rightarrow A\Delta C = B$.
(b) Define $\limsup_n A_n$ and $\liminf_n A_n$. Let

$$A_n = \begin{cases} (1 - \frac{1}{n}, 2) & \text{if } n \text{ is odd} \\ (1 + \frac{1}{n}, 2) & \text{if } n \text{ is even} \end{cases}$$

Find $\limsup A_n$ and $\liminf A_n$.
[3+7]

2. (a) An airport shuttle bus with 7 passengers makes 6 scheduled stops.Assuming that all possible distributions of 7 passengers in the 6 stops are equally likely find the probability that someone (at least one) gets down at each stop.
(b) A pair of dice are rolled and the sum of the two faces noted each time. Find the probability that a sum of 5 appears before a sum of 7.
(c) 13 cards are selected at random from a standard deck of 52 playing cards. Assuming that all possible distributions are equally likely, find the probability that there is at least one card of each of the 4 suits among the 13 cards selected.
(d) N men arrive at a party on a rainy evening and each leaves his hat and umbrella in a closet. When the party is over each picks up a hat and an umbrella at random. What is the probability that exactly m persons pick up their own hats and umbrellas?
[5 × 4]

3.Let X, Y be independent random variables each having the same geometric $\mathcal{G}(p)$--distribution. Let $U = \boldsymbol{Minimum}(X, Y)$ and $V = |X - Y|$.
(a) Find the p.m.f of U.
(b) Find the joint p.m.f of U, V and verify that U, V are independent.
[3+12]

P.T.O

INDIAN STATISTICAL INSTITUTE

Backpaper Examination, 1^{st} Semester, 2008-09

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

Date: 9.4.09 Time: 3 hours

This paper carries 100 marks. Attempt all questions.

1. For a set of observations, arrange the following measures of spread in ascending order of magnitude: half range, mean deviation about mean, mean deviation about median, standard deviation. You need to derive any result you may use. Explain, with suitable justification, whether it is possible for these measures to be equal to each other. [20]

2. (a) 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 iterative algorithm to estimate a and b. Show all your computational steps clearly.

   (b) Explain whether it is possible that the correlation coefficient between x and y is equal to the proportion of variance in y not explained by a least squares linear regression on x? [15+5]

3. (a) Explain the terms: (i) retrospective study, (ii) double blind experiment, (iii) product binomial sampling in the context of statistical designs.

   (b) Consider a set of unequal observations. Suppose that the minimum observation is decreased and the maximum observation observation is increased. Explain how do the median, the mean deviation about the median and the mean deviation about the mean change?

P. T. O

4. Let $X_1, X_2, \cdots, X_n$ be independent random variables each having the same distribution given by

$$P[X_1 = 1] = \tfrac{1}{6}\ ,\ P[X_1 = 2] = \tfrac{2}{3} \text{ and } P[X_1 = 3] = \tfrac{1}{6}\ .$$

Define for $j = 1, 2, 3$,

$Y_j =$ number of indices i among $1, 2, \cdots, n$ such that $X_i = j$.

(a) Write down the joint p.m.f of (Y_1, Y_2, Y_3).

(b) By expressing each Y_j as sum of suitably defined indicator random variables find $E(Y_1)$, $Var(Y_1)$, $Cov(Y_1, Y_2)$. [5+10]

5. (a) Let X be a non-negative integer valued random variable with probability generating function (p.g.f) $P(s)$. Let $Y = 3X$. Express the p.g.f $Q(s)$ of Y in terms of $P(s)$.

(b) Find the p.m.f of the random variable whose p.g.f is $P(s) = e^{\lambda(s^3-1)}$, $(\lambda > 0)$. [5+10]

6. (a) Let X and Y be two random variables with $E(|X|^p) < \infty$, $E(|Y|^q) < \infty$ where $p > 1$ and $\frac{1}{p} + \frac{1}{q} = 1$. Prove that

$$E(|XY|) \le (E(|X|^p)^{\frac{1}{p}}(E(|Y|^q)^{\frac{1}{q}}.$$

(b) Give an example of two random variables X, Y such that $Cov(X,Y) = 0$, but X, Y are not independent. [10+5]

7. Let $X_1, X_2, X_3, \cdots,$ be independent and identically distributed discrete random variables such that $E(|X_1|) < \infty$. Let the random variables $V_i^{(n)}$ be defined as follows:

$$V_i^{(n)} = \begin{cases} 0 & , |X_i| \le n \\ X_i & , |X_i| > n \end{cases}$$

$i = 1, 2, \cdots,$ and $n = 1, 2, \cdots$.

Prove that $P[V_1^{(n)} + V_2^{(n)} + \cdots + V_n^{(n)} \neq 0] \to 0$, as $n \to \infty$. [15

(c) 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.

(d) Explain what is meant by "Simpson's Paradox" with a suitable example. [6+9+5+5]

4. The following data pertain to fasting glucose levels (FBS) and triglyceride levels (TRI) of 8 patients in a study on Type 2 diabetes:

Patient	FBS	TRI
1	85	157
2	115	335
3	109	96
4	145	87
5	210	185
6	235	288
7	81	202
8	188	185

(a) Comment on the kurtosis of TRI in the above data.

(b) Draw the scatter plot of FBS and TRI. Based on a least squares linear regression of TRI on FBS, predict the triglyceride level of a patient with fasting glucose level 130. Identify the high residual observations in the above regression.

(c) Compute Kendall's τ for the above data and comment on the association between fasting glucose levels and triglyceride. [10+15+10]

INDIAN STATISTICAL INSTITUTE

First Semestral Examination (Back Paper): 2008-2009

B. Stat. - First Year

Analysis I

Date : 15.4.2009. Maximum Score : 100 Time : 3 Hours

Answer all the questions.

(1) Assume that there is a one-to-one map $f : X \to Y$ and also a one-to-one map $g : Y \to X$. Show that there is a bijection $h : X \to Y$.

[16]

(2) (a) Show that every connected open subset of $\mathbb{R}^2$ is path connected.

(b) Show that there is no one-to-one continuous function from $(0, 1]$ onto $(0, 1)$.

[16 + 8]

(3) (a) Show that every real-valued continuous function f defined on a closed bounded subset of $\mathbb{R}$ is uniformly continuous.

(b) Let $A \subset \mathbb{R}$ be such that every real-valued continuous function on A is bounded. Show that A is compact.

[10 + 10]

(4) (a) Is the series $\sum_{n=1}^{\infty} \frac{1}{n(n+1)}$ convergent? Justify your answer. Further, if the series is convergent, determine its sum.

(b) Find the largest and the smallest value of $x^3 - 18x^2 + 96x$ in the interval $[0, 9]$.

[8 + 8]

(5) (a) Let $f : [a, b] \to \mathbb{R}$ be differentiable and d lies between $f'_+(a)$ and $f'_-(b)$. Show that there is a $c \in (a, b)$ such that $f'(c) = d$.

(b) Let $f, g : \mathbb{R} \to \mathbb{R}$ be differentiable maps such that for every real number x,

$$f'(x)g(x) - f(x)g'(x) \neq 0.$$

Show that between any two roots of $f(x) = 0$ lies a root of $g(x) = 0$.

[16 + 8]

INDIAN STATISTICAL INSTITUTE
Second Semestral Examination : (2008-2009)
B. Stat. - First Year
Analysis II

Date : 30. 04. 2009 Maximum Score : 100 Time : 4 Hours

1. This paper carries questions worth a total of 115 marks. Answer as much as you can. The maximum you can score is 100 marks.
2. Unless otherwise stated, $\mathbb{R}$ and $\mathbb{C}$ will respectively denote the space of all real and complex numbers.

(1) (a) Let $f, g : [0,1] \to \mathbb{C}$ be continuous functions. Show the following:

(i) $\int_0^1 |f(x) \cdot g(x)| dx \leq \sqrt{\int_0^1 |f(x)|^2 dx} \cdot \sqrt{\int_0^1 |g(x)|^2 dx}$.

(ii) $\sqrt{\int_0^1 |f(x) + g(x)|^2 dx} \leq \sqrt{\int_0^1 |f(x)|^2 dx} + \sqrt{\int_0^1 |g(x)|^2 dx}$.

(b) State Stone-Weierstrass theorem for real-valued continuous functions.

(c) Show that for every bounded Riemann-integrable $f : [0,1] \to \mathbb{R}$ and every $\epsilon > 0$, there is a continuous $g : [0,1] \to \mathbb{R}$ such that

$$\int_0^1 |f(x) - g(x)| dx < \epsilon.$$

(d) Show that there is a countable set $\mathcal{D}$ of continuous functions $f : [0,1] \to \mathbb{R}$ such that for every bounded real-valued Riemann integrable functions f defined on $[0,1]$ and every $\epsilon > 0$, there is a $g \in \mathcal{D}$ such that

$$\int_0^1 |f(x) - g(x)| dx < \epsilon.$$

[(6 + 4) + 3 + 7 + 8]

(2) (a) Show that

$$\gamma = \lim_{n \to \infty} (1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} - \log_e(n))$$

exists.

(b) Let $f : [0,1] \to \mathbb{R}^n$ be a continuous function. Define

$$\gamma(t) = \int_0^t f(x) dx, \ 0 \leq t \leq 1.$$

P.T.O

Determine if γ is a rectifiable curve or not. If γ is rectifiable, find its length.

[10 + 10]

(3) (a) State Stone-Weierstrass Theorem for complex-valued continuous functions.

(b) Let X be the set of all continuous functions $f : (-1,1) \to \mathbb{C}$ such that $\lim_{|x|\to 1} f(x)$ exists. Show that there is a countable $D \subset X$ such that for every $f \in X$ there is a sequence $\{g_n\}$ in D such that $g_n \to f$ uniformly on E.

[3 + 12]

(4) Let

$$f(z) = \sum_{n=0}^{\infty} a_n z^n, \quad |z| < R,$$

$$g(z) = \sum_{n=0}^{\infty} b_n z^n, \quad |z| < S$$

and $|f(z)| < S$ for every $|z| < R$. Define

$$h(z) = g(f(z)), \quad |z| < R.$$

Show that h has a power series expansion on $|z| < R$.

[15]

(5) (a) Find a trigonometric series

$$\sum_{0}^{\infty} (a_n \cos nx + b_n \sin nx)$$

that converges to x^2 uniformly on $[-\pi, \pi]$. Justify your answer also.

(b) Let $f : \mathbb{R} \to \mathbb{R}$ be a continuously differentiable, perodic function of period 2π. Show that the Fourier series of f converges to f uniformly on $\mathbb{R}$.

(c) Let $\varphi_0, \cdots, \varphi_n$ be an orthonormal set of continuous complex-valued functions on $[0,1]$. Show that there is a non-zero continuous function $f : [0,1] \to \mathbb{C}$ such that

$$\int_0^1 f(x) \cdot \overline{\varphi_k(x)} dx = 0$$

for every $k = 0, 1, \cdots, n$.

[10 + 12 + 15]

INDIAN STATISTICAL INSTITUTE

Second Semester Examination: (2008-2009)

B. STAT. I year

Vectors and Matrices II

Date: 4 May, 2009. Maximum Marks: 60 Duration: 3 Hrs.

otes: *(i) All the matrices and vectors considered are over complex field unless otherwise ated. (ii) Class room notation is used. (iii) State clearly the results used. (iv) 4 marks loted for neatness. (v) This paper contains 68 marks and maximum marks you can score 60.*

1 Prove the following:

(a) Every square matrix satisfies its characteristic equation.

(b) Algebraic multiplicity of any eigenvalue of a matrix is greater than or equal to its geometric multiplicity.

(c) Every square matrix A admits a decomposition as $A = B + C$ where C is nilpotent, $r(B) = r(B^2)$ and $BC = 0 = CB$.

(d) Every singular value of an idempotent matrix is 1 implies it is symmetric.

(e) H is idempotent matrix implies that there exists a positive definite matrix M such that HM is hermitian.

(f) Let $A = B + C$ and $r(A) = r(B) + r(C)$. Then A is nonnegative definite and B is symmetric implies B and C are nonnegative definite.

(g) A and B are nonnegative definite matrices implies all the eigenvalues of AB are nonnegative.

[$7 \times 6 = 42$]

(a) Define rank and signature of a quadratic form and prove that a real 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, B and $A - B$ be nonnegative definite matrices. Show that there exist nonnegative definite g-inverses A^- and B^- of A and B respectively such that $B^- - A^-$ is nonnegative definite.

[$6 + 6 = 12$]

(a) Consider the quadratic form $x'Ax$ where the matrix A is given below. If possible find a lower triangular matrix L such that $A = LL'$ using square root method. Also find $|A|$.

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

[P.T.O]

(b) For the given matrix A, reduce the quadratic form $x'Ax$ to diagonal form using a nonsingular linear transformation of the variables. Give also the matrix of the transformation.

[6 + 4 = 10]

$\cdots END \cdots$

INDIAN STATISTICAL INSTITUTE

Final Examination, 2^{nd} Semester, 2008-09

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

Date: May 8, 2009 Time: $3\frac{1}{2}$ hours

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

1. Based on the 2004 parliamentary elections, data were collected on four variables in 100 constituencies with three or more candidates:

   X_1 : the number of voters per unit area in the constituency
   X_2 : the number of candidates in the constituency
   X_3 : the mean age (in excess of 18) of voters in the constituency
   X_4 : the percentage margin of victory

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

$$\begin{pmatrix} 685.72 \\ 7.39 \\ 25.89 \\ 5.81 \end{pmatrix}, \begin{pmatrix} 84.6 & -14.8 & 55.2 & 20.1 \\ & 3.3 & -3.2 & -1.2 \\ & & 14.4 & 4.1 \\ & & & 4.9 \end{pmatrix}$$

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

   (b) What proportion of the variance in X_4 is explained by the above linear regression?

   (c) Explain whether X_2 is useful in the above linear prediction of X_4.

   [4 + 4 + 4]

2. Suppose $R = ((r_{ij}))$ is the correlation matrix of $(X_1, X_2, \ldots, X_p)$.

   (a) If $\bar{r}$ is the mean of the off-diagonal elements of R, show that $\bar{r} \geq \frac{-1}{p-1}$.

   (b) Show that $R^2_{1(23\ldots p)} - R^2_{1(34\ldots p)} \leq r^2_{12.34\ldots p}$. [3 + 3]

P.T.O

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

$$f(x) = \frac{cA^c}{x^{c+1}};\ x > A > 0,\ c > 0.$$

(b) Suppose you have a mechanism to generate Poisson random variables with any mean. Explain, with suitable justification, how you can generate two Poisson variables X_1 and X_2 such the mean of X_1 is 2.7, the mean of X_2 is 4.8 and the correlation coefficient between X_1 and X_2 is 0.6.

(c) Explain, with suitable justification, how you can draw a random sample of size n from a population of size N without replacement using a fair coin. [4 + 4 + 4]

4. (a) In genetic studies of rare diseases, families are often selected through an affected member in the family (known as a *proband*). Assuming that all individuals have the same chance of being affected, find the variance of the number of affected sibs in a family of 5 sibs selected through a proband.

(b) A mobile service provider records fractions of minutes of call time as complete minutes. If the actual call time of an individual is distributed as exponential with mean 5 minutes, what is the probability that the recorded call time is within 15 seconds of the actual call time? If the recorded call time is 7 minutes, what the probability of the above event? [4 + 4]

5. (a) Since it would cause a lot of inconvenience if an entire lane is closed down for repair, it is decided that a random stretch along the length of the lane will be kept open. If L_{max} and L_{min} denote, respectively, the maximum and the minimum of the random lengths of the stretch closed down and kept open, show that L_{max}/L_{min} cannot have finite expectation.

(b) Suppose a risk score for cardio-vascular disease (CVD) is defined by $(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)$ where $\mathbf{x}$ is a vector comprising body mass index (BMI) and homocysteine levels and is distributed as bivariate normal with mean vector μ and variance-covariance matrix Σ. What is the probability that the risk score of an individual is greater than 4? [4 + 4]

6. The following data pertain to triglyceride levels (TRI) of 200 patients in a study on Type 2 diabetes:

TRI values	Frequency
80-130	10
130-180	18
180-230	27
230-280	45
280-330	38
330-380	24
380-430	16
430-480	12
480-530	7
530-580	3

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

INDIAN STATISTICAL INSTITUTE
B. Stat I
Computational Techniques and Programming II
(End Semester Exam)

Date: 11.5.09 Time: 210 mins

Note: Answer as many questions as you like. Maximum you can score is 100.

1. (a) State the Lagrange form and Newton's divided difference form for polynomial interpolation for n unequally spaced arguments and compare the number of arithmetic computations required in both the methods.

(b) (i) Prove that if $p_n(x)$ is a polynomial of degree n with leading coefficient a_n, and x_0 is an arbitrary point, then

$$p_n(x_0) = a_n n!\, h^n$$

where h is the step length between two consecutive nodes.

(ii) What simplifications can you make in Newton's divided difference form of the interpolating polynomial when the data points are equally spaced.

(c) Describe Aitken's Nevielle's algorithm for iterated linear interpolation for (n+1) points and generate the triangular table.

[8+(3+5)+6=22]

2. (a) Write down the differentiation formula based on Newton's Forward formula and find $f'(1.2)$ and $f''(1.2)$ using numerical differentiation for the following table.

x	f(x)
1.2	1.5095
1.3	1.6984
1.4	1.9043
1.5	2.1293
1.6	2.3756

(b) (i) Write down the Gaussian Quadrature formula of numerical integration and hence find the three point quadrature rule.

(ii) Provide the geometrical interpretation of the three point quadrature rule and derive the expression for error.

[11+(5+6)=22]

P. T. 0

3. (a) Write the differences between one step method and multi-step method to solve an ordinary differential equation. Describe the 2nd order Runge-Kutta method and state the order of the error in this method.

(b) Describe Milne's Predictor-Corrector method and find the error terms in both the predictor and corrector formulae.

[11+11=22

4. (a) (i) Describe the Regula-Falsi method to approximate the solution of an equation in one variable. State all the needed assumptions for this method.

(ii) Provide the geometrical interpretation of the above method.

(iii) Show that the method indeed converges to the actual solution of the equation.

(b) Write down the algorithm for LU decomposition of the coefficient matrix A of a system of linear equations Ax = b.

[(4+3+4)+11=22

5. Let $A=\begin{bmatrix} 12 & -51 & 4 \\ 6 & 167 & -68 \\ -4 & 24 & -41 \end{bmatrix}$ and $b=\begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$.

Decompose A into an upper triangular matrix and an orthogonal matrix using each of

(a) Householder Transformation and

(b) Given's transformation

Hence solve for x in Ax = b.

[11+11=22

--------x------

INDIAN STATISTICAL INSTITUTE
B.STAT-I (2008-09)
Theory of Probability and its Applications - II
Semestral-II examinations
Maximum marks: 60. Time: 3 hours.

Date/15May,2009.

Note: Answer as many questions as you wish. The whole question paper carries 70 marks. The maximum you can score is 60.

1. (a) Show that if X is a random variable with a continuous Distribution Function $F(x)$ then the random variable $Y = F(X)$ has uniform distribution over the interval $(0, 1)$.

(b) Show that a bivariate Distribution Function $F(x, y)$ must satisfy the following condition:

$$F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2) \geq 0, \text{ for all } a_1 < b_1 \text{ and } a_2 < b_2.$$

(c) Is it possible that the function $F(x, y) = I_{(x+2y \geq 0)}(x, y)$ on $\mathbb{R}^2$ is a bivariate Distribution Function? Give reasons for your answer.

[10+3+2]

2. (a) Let X be a random variable having the density $f(x)$ where

$$f(x) = \begin{cases} 0 & \text{, when } x \leq -1 \\ \frac{1}{2} & \text{, when } -1 < x \leq 0 \\ \frac{1}{2}e^{-x} & \text{, when } 0 < x \end{cases}$$

Find the density of the random variable $Y = X^2$.

(b) Let Θ be a random variable having uniform distribution $U(0, 2\pi)$. Find the density of the random variable $X = Sin\Theta$. Do you think that the random variable $Y = Sin2\Theta$ has the same density as that of X? Give reasons.

[6+9]

3. Let (X, Y) have joint density

$$f(x, y) = \begin{cases} Cx^{\alpha-1}(y-x)^{\beta-1}e^{-\lambda x} & , 0 < x < y < \infty \\ 0 & , \text{otherwise} \end{cases}$$

(a) Find the constant C and the marginal densities of X and Y.

(b) Find the conditional density of Y given $X = x$.

(c) Find the joint density of (Z, Y), where $Z = \frac{X}{Y}$. Find the conditional density of Z given $Y = y$, $y > 0$. Compute $E(X|Y = y)$.

[5+2+8]

4. Let (X, Y) have bivariate Normal $N_2\big((0,0), 1, 1, \rho\big)$ – distribution, $-1 < \rho < 1$.

(a) Let $U = \frac{X-\rho Y}{\sqrt{(1-\rho)^2}}$, and $V = Y$. Show that U and V are independent $N(0, 1)$ random variables.

(b) Find the corelation coefficient between the random variables X^2 and Y^2.

[6+4]

(P.T.O.)

5. (a)Let the random variable $X \sim Beta(m,n)$, where $m > 0$, $n > 0$ are any real numbers. Compute $E(X^r)$, $r > 0$.
(b) Let $X \sim Beta(m,n)$, $Y \sim Beta(m+\frac{1}{2}, n)$ and X and Y are independent.
Let $Z = \sqrt{XY}$. Find $E(Z^k)$, for $k = 1,2,3,\cdots$. Check that these moments ae the same as that of a $Beta(2m, 2n)$ random variable. (Hint: First, prove the following dulpication formula for Gamma function: $\Gamma(2\nu) = \frac{1}{\sqrt{\pi}} 2^{2\nu-1}\Gamma(\nu)\Gamma(\nu+\frac{1}{2})$). /are

[3+12]

....

INDIAN STATISTICAL INSTITUTE

Second Semestral Examination (Back Paper): 2008-2009

B. Stat. - First Year

Analysis II

Date : 21.8.09 | Maximum Score : 100 | Time : 3 Hours

Answer all the questions.

(1) Let $\{a_{nm}\}$ be a double sequence of real numbers such that for each m, $\sum_n |a_{nm}| = t_m < \infty$ and $\sum_m t_m$ is convergent. Show that

$$\sum_n \sum_m a_{nm} = \sum_m \sum_n a_{nm}.$$

[16]

(2) (a) Show that for every bounded Riemann-integrable function $f : [0,1] \to \mathbb{R}$ and every $\epsilon > 0$, there is a continuous function $g : [0,1] \to \mathbb{R}$ such that

$$\int_0^1 |f(x) - g(x)| < \epsilon.$$

(b) Show that every C^1-function $\gamma : [0,1] \to \mathbb{R}$ is rectifiable. Also compute its length.

[12 + 16]

(3) (a) Let

$$f(z) = \sum_0^\infty a_n z^n, \ |z| < R,$$

with $a_0 \neq 0$. Show that there is a $\delta > 0$ and a sequence $\{b_n\}$ of complex numbers such that

$$\frac{1}{f(z)} = \sum_0^\infty b_n z^n, \ |z| < \delta.$$

(b) Let $f : [0, \infty) \to \mathbb{R}$ be a continuous function such that $\lim_{x\to\infty} f(x)$ exists. For every $\epsilon > 0$ show that there is a polynomial $P(x) \in \mathbb{R}[x]$ such that

$$|f(x) - P(e^{-x})| < \epsilon.$$

[16 + 12]

(4) (a) Let $\{\varphi_k\}$ be a maximal set of functions from $[a,b] \to \mathbb{C}$ satisfying the following two properties:

P. T. O

(i) $\int_a^b |\varphi_k(x)|^2 dx = 1$.

(ii) $l \neq k \Rightarrow \int_a^b \varphi_l(x) \cdot \overline{\varphi_k(x)} dx = 0$.

Show that for every $f : [a, b] \to \mathbb{C}$ with $|f|^2$ integrable,

$$\int_a^b |f(x)|^2 dx = \sum_k |\int_a^b f(x) \cdot \overline{\varphi_k(x)} dx|^2.$$

(b) Find the Fourier series expansion of $f(x) = x^2$.

[18 + 10]

INDIAN STATISTICAL INSTITUTE
B. Stat I
Computational Techniques and Programming II
(End Semester Exam, Back Paper)
Total Marks: 100

Date: 24.8.09 Time: 3 hours

Note: Answer any four questions

1. (a) Define absolute error and relative error in approximation of a number. An approximate number a=24,253 has a relative error of 1 %. How many correct digits does it have?

(b) The base of a cylinder has radius $R \approx 2m$, the altitude of the cylinder is $H \approx 3m$. With what absolute errors must we determine R and H so that the volume V may be computed to within 0.1 m^3?

(c) $Cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} + ... + (-1)^n \frac{x^{2n}}{(2n)!} + ...$

How many terms of the above series are to be computed so that the value of $Cos x$ is accurate upto 5 decimal places for each x where $|x| < \pi/2$?

[8+9+8=25]

2. (a) Obtain Newton's formula for polynomial interpolation using divided differences in case of n equally spaced data points when the unknown argument is situated
(i) at the beginning of the data set
and (ii) at the end of the data set.

Write down the difference table for each case.

(b) Let f(-2) = -50, f(-1) = 6, f(0) = 10, f(1) = 10, f(2) = 30, f(3) = 190. Find approximate values of $f'(-2)$ and $f'(3)$ by stating clearly the formulas you are going to use.

[(5+5+5)+10=25]

3. (a) Obtain the general quadrature formula for equidistant ordinates and hence find the two point quadrature rule. Explain why two point quadrature rule is called Trapezoidal rule?

(b) Find the error terms for both the two point and three point quadrature formulae.

[(5+5+5)+10=25]

P. T. O

4. (a) Apply Decarte's rule of signs for the number of positive real zeros of the equation $x^3 - 18 = 0$. State any formula for determining the range of zeros of a polynomial equation.

(b) Give an algorithm for fixed point iteration scheme and state the assumptions required for convergence of the sequence of numbers generated by it. Using those assumptions derive the order of convergence of the fixed point iteration scheme.

[10+15=25]

5. (a) Write an algorithm to solve the differential equation $y' = f(x,y)$ with the initial condition $y(x_0) = y_0$, numerically using Runge Kutta method of fourth order.

(b) Let $A = \begin{bmatrix} 3 & 2 & -4 \\ -1 & 5 & 2 \\ 2 & -3 & 4 \end{bmatrix}$ and $b = \begin{bmatrix} 12 \\ 1 \\ -3 \end{bmatrix}$.

Decompose A into an upper triangular matrix and an orthogonal matrix using Householder Transformation and hence solve for x in Ax = b.

[15+10=25]

-------x-----

INDIAN STATISTICAL INSTITUTE

Backpaper Examination, 2^{st} Semester, 2008-09

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

Date: 24. 8. 09 Time:3 hours

This paper carries 100 marks. Attempt all questions.

1. Define the multiple correlation coefficient of X_1 and $(X_2, X_3, \ldots, X_p)$. Show that it is a non-negative quantity. Obtain its expression in terms of the variance-covariance matrix of $(X_1, X_2, \ldots, X_p)$. Show that it is invariant under non-singular transformations on X_1 and $(X_2, X_3, \ldots, X_p)$. Show that for any $j = 2, 3, \ldots, p$, the multiple correlation between X_1 and $(X_2, X_3, \ldots, X_p)$ cannot be less than the partial correlation between X_1 and X_j eliminating the linear effects of $X_2, X_3, \ldots, X_{j-1}, X_{j+1}, \ldots, X_p$. You need to derive any result you may use. [2+3+6+6+7=24]

2. Suppose the correlation matrix of $(X_1, X_2, \ldots, X_9)$ is $R = ((r_{ij}))$ such that $r_{j9} = 0.3$; $j = 1, 2, \ldots, 8$ and $r_{ij} = 0.7$; $i, j \leq 8$, $i \neq j$. Compute the multiple correlation coefficient between X_9 and $(X_1, X_2, \ldots, X_8)$ and the partial correlation of X_1 and X_9 eliminating the linear effects of $X_2, X_3, \ldots, X_8$. [6+6=12]

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) In the classical "'capture-recapture"' method, show that the distribution of the number of tagged fish caught in the recapture stage is equivalent to the conditional distribution of a suitable Binomial variable conditioned on the sum of two independent Binomial variables.

   (c) What is the probability that two specific units in a population are included in a sample drawn from the population using SRSWR?

P.T.O

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 x_1 and x_2, respectively. If a is any real number, show that $ax_1 + (1-a)x_2$ is an unbiased estimator of the population mean. For what value of a does the above estimator have the minimum variance? [6+6+8=20]

4. (a) A machine has three components functioning independently of each other. The machine is active as long as at least two of the components function. If the lifetime of each of the components is distributed as exponential with mean 6 months, find the expected lifetime of the machine.

   (b) Assuming that the duration of songs in Bollywood movies is distrubted as normal with mean 4 minutes and variance 42 seconds, obtain the variance of the duration of those Bollywood songs with duration between 3 and 5 minutes. [9+9=18]

5. Using random observations from $U(0,1)$, explain, with suitable justification, how you would generate an observation from:

   (a) a Poisson distribution with mean 4.

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

   [8+8=16]

6. The following data pertain to the number of credit cards owned by 50 randomly chosen individuals:

Number of credit cards	Number of individuals
0	10
1	14
2	14
3	7
4	3
5	2

Fit a suitable distribution to the above data and test for its goodness of fit. [10]

INDIAN STATISTICAL INSTITUTE
B.STAT-I (2008-09)
Theory of Probability and its Applications - II
Backpaper Examination
Maximum marks: 100. Time: 3 hours.

Date : 26.8.09

Note: ***Answer as many questions as you wish. The whole question paper carries 105 marks. The maximum you can score is 100.***

1. Consider the following function $F(x,y)$ defined on $\mathbb{R}^2$:
$$F(x,y) = \begin{cases} 1 & \text{if } x+y \geq 0 \\ 0 & \text{otherwise} \end{cases}$$
Can $F(x,y)$ be possibly a bivariate Distribution Function ? Give reasons. [10]

2. Let $X \sim N(0,1)$. Let $Y = e^X$.
(a)Find the density function $g(y)$ of the random variable Y. Find $E(Y^k)$, $k = 1, 2, \cdots$.
(b) Let $h(y) = g(y)(1 + \sin(2\pi \log y))I_{(y>0)}$. Show that $h(y)$ is a density function. Also check that if U is a random variable with density function h, then for each k, $k = 1, 2, \cdots$, the kth moment of U, is the same as the corresponding moment of the random variable Y. [7+8]

3. Let X and Y be independent random variables each having $\mathcal{E}(\lambda)$ – distribution. Let $U = Min(X,Y)$, $V = Max(X,Y)$.
(a) Compute $P(U \geq u)$, $P(V \leq v)$, for $u > 0$ and $v > 0$. Hence find the densities of U and V.
(b) Find $P(U \geq u, V \leq v)$, for $0 < u < v < \infty$. Show that the joint density $g(u,v)$ of U and V is given by:
$$g(u,v) = \begin{cases} 2\lambda^2 e^{-\lambda(u+v)} & \text{, for } 0 < u < v < \infty \\ 0 & \text{, otherwise.} \end{cases}$$
(c) Let $W = V - U$. Find the joint density of (U, W). Check that U and W are independent random variables. [6+5+9]

4. Let
$$f(x,y) = \begin{cases} [(1+ax)(1+ay) - a]\, e^{-x-y-axy} & \text{if } x > 0 \text{ and } y > 0 \\ 0 & \text{otherwise} \end{cases}$$
(where $0 < a \leq 1$ is a constant).
i) Prove that $f(x,y)$ is a density .
ii) Show that both the marginal densities are exponential $\mathcal{E}(1)$. Find the conditional expectation $E(X \mid Y = y)$, for $y > 0$. [7+8]

(Please Turn overleaf)

5. Let (X, Y) have a bivariate normal density with $E(X) = E(Y) = 0$, $E(X^2) = E(Y^2) = 1$ and $E(XY) = \rho$.
(a) Find the conditional density of X given $Y = y$
(b) Consider the usual polar trasformation (R, Θ) where $X = R\cos\Theta, Y = R\sin\Theta$, $R > 0,\ 0 < \Theta < 2\pi$. Show that Θ has a density given by

$$f(\theta) = \begin{cases} \frac{\sqrt{(1-\rho^2)}}{2\pi(1-2\rho\sin\theta\cos\theta)} & \text{for } 0 < \theta < 2\pi \\ 0 & \text{otherwise.} \end{cases}$$

[5+10]

6. Let (X, Y) have the density function $f(x, y)$ given by :

$$f(x,y) = \begin{cases} Cx^{r-1}y^{s-1}(1-x-y)^{t-1} & \text{, for } 0 < x, y < 1 \text{ and } x + y < 1 \\ 0 & \text{, otherwise} \end{cases}$$

where $r > 0, s > 0, t > 0$ are fixed real numbers.
(a) Find the value of the constant C. Find the marginal densities of X and Y.
(b) Find the conditional density of $\frac{X}{1-Y}$ given $Y = y,\ y > 0$.

[7+8]

7. (a)Let X be a random variable with $M_X(t) = E(e^{tX})$ finite , $t \geq 0$. Show that

$$P(X \geq a) \leq e^{-ta} M_X(t).$$

(b) Let $X \sim \Gamma(\alpha, \lambda)$ - distribution . Find the set of all t in $\mathbb{R}$ such that $M_X(t)$ is finite. Find $M_X(t)$ for all such values of t. Prove that

$$P[X \geq \frac{2\alpha}{\lambda}] \leq 2^{\alpha} e^{-\alpha}$$

(Hint : Find $\underset{t \geq 0}{\text{Minimum}} \left(e^{-ta} M_X(t)\right)$, for $a = \frac{2\alpha}{\lambda}$.)

[5+10]
