4

INDIAN STATISTICAL INSTITUTE
Mid-Semesteral Examination : 2011-12
B. Stat. - First Year
Analysis I

Date : 01. 09. 2011 Maximum Score : 100 Time :3 1/2 Hours

1. Answer all the questions.

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. The paper carries a total of 133 marks. The maximum score on each question is shown against them.

(1) Is there a set X such that for every set Y there is a function from X onto Y? Give a full proof of your assertion.

[10]

(2) Show that the set of all infinite subsets of the set $\mathbb{N}$ of natural numbers is uncountable.

(You can take a hint predetermined by me. In case you take the hint the maximum score on this question will be 10.)

[15]

(3) Show that there is no rational number x such that $x^3 = 100$. [12]

(4) Show that every Cauchy sequence of real numbers converges to a real number. [10]

(5) Show that there is a unique real number x such that $e^x = 2$.

(You can take a hint predetermined by me. In case you take the hint the maximum score on this question will be 10)

[15]

(6) Show that for every non-zero integer m, e^m is irrational.

(You can take a hint predetermined by me. In case you take the hint the maximum score on this question will be 7.)

[15]

(7) For a sequence $\{a_n\}$ of positive real numbers show that

$$\liminf \frac{a_{n+1}}{a_n} \leq \liminf a_n^{\frac{1}{n}}.$$

[10]

P.T.

(8) For a sequence $\{a_n\}$ of real numbers, define

$$\sigma_n = \frac{a_1 + \cdots + a_n}{n}, \quad n \geq 1.$$

Show that

$$\limsup \sigma_n \leq \limsup a_n.$$

[12]

(9) Give a rearrangement of the series

$$\sum_1^\infty (-1)^{n+1} \frac{1}{n}$$

that converges to -10. [12]

(10) Let

$$\sum_0^\infty a_n x^n$$

be a power series which is convergent for a non-zero real number t. Show that for every real number x with $|x| < |t|$, the above series is uniformly convergent. [10]

(11) Show that for any two real numbers x and y,

$$\sin(x + y) = \sin x \cdot \cos y + \cos x \cdot \sin y.$$

[12]

INDIAN STATISTICAL INSTITUTE

Mid-Semester Examination – Semester I : 2011-2012
B.Stat. (Hons.) I Year
Probability Theory I

Date : 05.09.11 | **Maximum Score : 80 pts** | Time : $2\frac{1}{2}$ Hours

Note : This paper carries questions worth a **total** of **96 points**. Answer as much as you can. The **maximum** you can score is **80 points**.

1. If A and B are two events with $P(A) = 0.5, P(B) = 0.4$ and $P(A^c \cap B^c) = 0.2$, find the probability that (i) A occurs but B does not occur, (ii) **exactly one** of the two events A and B occur. (6+6)=[12]

2. In 10 throws of a die, the probability that exactly two distinct faces show up is given by $\alpha[(\frac{1}{3})^{10} - \beta]$. What are α and β? [10]

3. An urn contains eight balls, two of which are red and the rest are black. Balls are drawn one after another (without replacement) until both the red balls show up. What is the probability that (a) at least five draws will be needed, (b) exactly five draws will be needed? You need to simplify your answers for both problems. (10+10)=[20]

4. If 10 married couples are seated in a row at random, what is the probability that no husband sits next to his wife? It is not necessary to simplify your answer. [10]

5. Among 16 students participating in a doubles badminton tournament of ISI students, 8 are from B.Stat. classes and the remaining 8 are from the non-B.Stat. classes. If the 16 participants are paired up randomly to form 8 teams, what is the probability that there will be exactly 2 teams with a B.Stat.-non-B.Stat. combination? Please try to simplify your answer as much as you can. [10]

6. (a) With the notation of random walk, find the probabilities (i) $P(S_5 = 1, S_9 = -1)$; (ii) $P(S_{10} = 2, M_{10} \geq 3)$ and (iii) $P(S_{10} = 2, M_{10} = 3)$.

   (b) Using the technique of dual paths (or otherwise), show that, for any positive integer r and any integer $k \leq r$, $P(M_n = r, S_n = r) = P(M_n = r - k, S_n = -k)$. Hence deduce that $P(M_n = r) = P(M_n - S_n = r)$. ((6+6+6)+(8+8))=[34]

INDIAN STATISTICAL INSTITUTE

Mid-Semestral Examination: 2011-12

B. STAT 1st Year

STATISTICAL METHODS I

07.09.2011 Full Marks: 80 Duration: 3 hrs

1. Express diagrammatically the following data on employment status of people of India by their broad categories during 1989-90, justifying your choice of diagram(s): (10)

Employment status	Rural		Urban	
	Male	Female	Male	Female
Unemployment	961	708	977	826
Employment	11	9	7	8
Not in labour force	28	283	16	166

2. What is a frequency curve? Draw a rough sketch of three different types of frequency curves with a real-life example for each. Justify your examples and the corresponding frequency curves. (5+3×3=14)

3. The life of electric bulbs manufactured by Company A for 50 bulbs (in 100 hour unit) is given below.

9.5	15.2	25.1	0.1	4.5	15.1	10.3	1.5	5.5	3.3	8.5
10.6	7.0	8.4	4.0	7.7	23.4	10.3	5.7	3.7	2.7	1.2
4.3	0.1	15.2	21.6	0.6	6.6	21.2	19.2	4.5	5.9	3.9
21.4	5.9	15.0	10.2	23.3	5.2	12.6	5.9	3.5	14.2	22.7
16.5	1.8	26.2	3.7	17.3	15.4					

A frequency distribution of the same for a number of bulbs manufactured by Company B is also available and given in Table 1.

Table 1
Table showing frequency distribution of life of electric bulbs

Class intervals for life (in 100 hour unit)	Frequency
0.1 — 3.0	12
3.1 — 6.0	21
6.1 — 9.0	10
9.1 — 12.0	8
12.1 — 15.0	5
15.1 — 18.0	3
18.1 — 21.0	2
21.1 — 24.0	4
24.1 — 27.0	3
27.1 — 30.0	4

(a) Compare the two frequency distributions and write a short report. Each comment in your report must be justified properly with supporting measures and / or diagrams.

(b) Later it was discovered that during the collection of data, the information of 21 bulbs manufactured by Company B was not included in the dataset, because 16 bulbs were found to be defective and the life hours of 5 bulbs were found to be 42.5, 58.7, 37.6, 43.2, and 65.1. But a statistician wants to include this information in the second dataset. Assume that the life of each of 16 defective bulbs can be taken as 0. Do you want to modify your analysis and hence the report? If yes, do it; if not, justify your reason for not modifying this analysis and report. Supporting measures and / or diagrams must be provided. (12+16=28)

4. Develop a formula to calculate mode for a frequency distribution of a continuous variable where no two class-widths are equal. (5)

5. In presence of extreme values, define two new measures of central tendency as follows. Assume that the frequency distribution is symmetric and unimodal and the total frequency is an odd number.
   (i) Trimmed mean: Arithmetic mean of the remaining observations ignoring all values lower than the first quartile and higher than the third quartile.
   (ii) Winsorised mean: Arithmetic mean after replacing all values lower than the first quartile by the value of first quartile and all values higher than the third quartile by the value of third quartile.

   Then,
   (a) Compare these modified means with the usual arithmetic mean.
   (b) What happens when the distribution is skewed instead of symmetric?
   (c) Also compare the standard deviations based on these two concepts (trimming and winsorising) with the usual standard deviation based on all observations. (4+6+8=18)

6. Suppose the mean and the standard deviation of a set of n observations are $\bar{x}$ and s respectively. When a new observation is introduced the mean decreases but the variance remains the same. Express the new observation in terms of $\bar{x}$ and s. (6)

INDIAN STATISTICAL INSTITUTE
B. Stat. I: 2011–2012
Computational Techniques & Programming I
Mid Semester Examination

Date: 09. 09. 2011 Marks: 100 Time: 3 Hours

Answer any part of any question. The question is of 110 marks. The maximum marks you can get is 100. Please write all the part answers of a question at the same place.

1. Write a function in C that finds the GCD as well as the LCM of two unsigned integers. Show how the function executes when the two integers are 2310 and 96900. Will your functions work properly for all pairs of unsigned integers? Explain. $5 + 5 + 5 = 15$

2. Explain the parameter passing strategy (to functions) in C programming language with the example of "swap" function. 10

3. Explain in detail the implementation of a *stack* of integers using C programming language? 20

4. (a) What is the output of the following piece of C code?

```
int i, k = 1, n = 5;
for (i = 0; k < n+1; i = k-i)
        { printf("%d\n", k); k = k+i; }
```

   (b) Consider the following C program.

```
#include <stdio.h>
main(){
        int arr[10][10], i, j;
        for (i = 0; i < 10; i++)
        for (j = 0; j < 10; j++)
                arr[i][j] = i*j + 1;
        printf("%x\n", arr); printf("%x\n", arr+1);
        printf("%x\n", *arr); printf("%x\n", *arr+1);
        printf("%x\n", **arr);
}
```

   The first line of output of this program is fef56970. What are the rest of the outputs? Give proper justification to your answer. $10 + 10 = 20$

5. Explain how can you implement a Chess board in C. Consider a Chess board when a game is continuing. There can be m many pieces in the board $(2 \leq m \leq 32)$. Write a C program that will report the largest empty square given the current situation on the board. $5 + 15 = 20$

6. (a) Write a C program, that will accept a character string and inform you if there is any substring which is a palindrome of length at least 5 characters.

   (b) What is the prototype of the "strtok" library function in C? How can you implement it? $10 + (5 + 10) = 25$

INDIAN STATISTICAL INSTITUTE

Mid-semester Examination (2011–2012)

B STAT I

Vectors & Matrices I

12.09.11

Date : [redacted] Maximum Marks : 60 Time : 2 hrs.

This paper carries 65 marks. Maximum you can score is 60. Precisely justify all your steps. Carefully state all the results you are using.

1. Let $\mathcal{P}_n$ be the space of all polynomials of degree $\leq n$ with real coefficients.

   (a) Does there exist a basis of $\mathcal{P}_n$ consisting of polynomials each of degree n? [5]

   (b) Does there exist a basis of $\mathcal{P}_n$ consisting of polynomials each of degree $(n-1)$? [5]

2. (a) Does there exist a linear map $T : \mathbb{R}^2 \to \mathbb{R}^4$ such that [5]

$$\text{Range}(T) = \{(x_1, x_2, x_3, x_4) \in \mathbb{R}^4 : x_1 + x_2 + x_3 + x_4 = 0\}?$$

   (b) Does there exist a linear map $T : \mathbb{R}^2 \to \mathbb{R}^3$ such that [5]

$$\text{Range}(T) = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_1 + x_2 + x_3 = 0\}?$$

3. Let X, Y be vector spaces over $\mathbb{K}$ and $T \in \mathcal{L}(X, Y)$.

   (a) Assume X is finite dimensional. Show that there is a subspace $Z \subseteq X$ such that $Z \cap \mathcal{N}(T) = \{0\}$ and $\text{Range}(T) = \{Tz : z \in Z\}$. [5]

   (b) Assume X is finite dimensional. Show that T is onto if and only if there exists $S \in \mathcal{L}(Y, X)$ such that $TS = I_Y$, the identity map on Y. [10]

   (c) Assume Y is finite dimensional. Show that T is one-to-one if and only if there exists $S \in \mathcal{L}(Y, X)$ such that $ST = I_X$, the identity map on X. [10]

4. (a) Do there exist $n \times n$ matrices A and B such that $AB - BA = I_n$? Why? [5]

   (b) For $n \times n$ matrices A and B, show that $\rho \begin{bmatrix} A & I_n \\ I_n & B \end{bmatrix} = n$ if and only if $B = A^{-1}$. [15]

First Semester Examinations(2011-2012)
B Stat – 1st year
Remedial English
100 marks
1 ½ hours

Date: 8th November 2011

1. Write an essay on any one of the following topics. Five paragraphs ar expected:
   a) Good and Evil Effects of the Television
   b) Hostel Life
   c) Durga Puja Vacation – 2011

(60 marks

2. Fill in the blanks with appropriate prepositions :
   a) He was a contemporary ________ Rammohan Roy and was closely associated _________him.
   b) Moving _________one place _________another he landed _______ _____Lucknow.
   c) He had necessarily _____live away elsewhere and_____ retirement chose ____settle down ______ Ballygunge.
   d) I was too much ______ a child ____gain much ______contact ______the people who frequented the house.
   e) ______the opening day ______the conference father actually did this.
   f) I heard ______father what happened ______the conference.
   g) The meetings came_______ an abrupt end when he left _____Calcutta.

(20 marks

3. Fill in the blanks with appropriate words:

The attachment ____ ____ scientist ____ ____ poet ____ ____ more than ____ friendship. They ____ constantly ____ ideas. One ____ ____ of the ____ story ____ ____ written ____ ____ other of the remarkable _____ carried _____ _____ his ____ .

(20 marks

INDIAN STATISTICAL INSTITUTE
Semesteral Examination : 2011-12
B. Stat. - First Year
Analysis I

Date : 11. 11. 2011   Maximum Score : 100   Time :4 Hours

1. *Answer all the questions.*
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. *Each question carries* 10 *marks.*

(1) Let A be a non-empty set of positive real numbers such that

$\sup \cup_{n=1}^{\infty}\{a_1 + \cdots + a_n : a_1, \ldots a_n \in A$ and are distinct$\} < +\infty$.

Show that A is countable.

(2) Let x be a positive real number and $n > 1$ an integer. Show that there is a unique positive real number y such that $y^n = x$.

(3) For every real number x, show that

$$\lim_{n\to\infty} (1 + \frac{x}{n})^n = e^x.$$

(4) Stating an appropriate result on power series, show that

$$\lim_{x\to 0} \frac{\sin x}{x} = 1.$$

(5) Let $\sum_{n=0}^{\infty} a_n$ be a series of real numbers whose partial sums are bounded and $\{b_n\}$ be a decreasing sequence of real numbers converging to 0. Show that the series $\sum_{n=0}^{\infty} a_n \cdot b_n$ is convergent. Will such a series be always absolutely convergent? Justify your answer.

(6) Let $f : \mathbb{R} \to \mathbb{R}$ be a monotone function such that for every interval $I \subset \mathbb{R}$, $f(I)$ is an interval. Show that f must be continuous.

(7) Let $A \subset \mathbb{R}$ be such that for every real number x, there is an open set containing x such that $A \cap U$ is countable. Show that A is countable.

(8) Let $C \subset \mathbb{R}$ be a closed and bounded set and $\mathcal{U}$ an open cover of C. Show that there is an $\epsilon > 0$ such that whenever $x, y \in C$ and $|x - y| < \epsilon$ there an open set in $\mathcal{U}$ containing both x and y.

(9) Let $\{U_n\}$ be a sequence of dense open sets in $\mathbb{R}$. Show that $\cap_n U_n$ is dense in $\mathbb{R}$.

(10) Let f be continuous on $[0,1]$, differentiable on $(0,1)$ and there is a λ, $0 < \lambda < 1$ such that $|f'(x)| \leq \lambda \cdot |f(x)|$. Show that f is a constant.

(11) Let $f : \mathbb{R} \to \mathbb{R}$ be a differentiable function and $x < y$ be such that $f'(x)$ and $f'(y)$ are finite. Show that for every u between $f'(x)$ and $f'(y)$ there is a v, $x \leq v \leq y$ such that $f'(v) = u$.

(12) Let $f : \mathbb{R} \to \mathbb{R}$ be n-times differentiable and x, y are two real numbers. Show that there is a θ between x and y such that

$$f(y) = \sum_{k=0}^{n-1} \frac{f^{(k)}(x)}{k!} \cdot (y-x)^k + \frac{f^{(n)}(\theta)}{n!} \cdot (y-x)^k$$

INDIAN STATISTICAL INSTITUTE

First Semestral Examination (2011–2012)

B STAT I

Vectors & Matrices I

Date : 14.11.2011 Maximum Marks : 100 Time : $3\frac{1}{2}$ hrs.

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

For a matrix A, let $\rho(A)$, $\mathcal{R}(A)$ and $\mathcal{C}(A)$ denote respectively the rank, the row space and column space of A.

1. (a) Let $\mathcal{P}$ be the space of all polynomials with real coefficients. Let α, β, γ be distinct real numbers and let

$$\begin{aligned} P_1(x) &= (x-\alpha)(x-\beta) \\ P_2(x) &= (x-\beta)(x-\gamma) \\ P_3(x) &= (x-\gamma)(x-\alpha) \end{aligned}$$

Is the set $\{P_1, P_2, P_3\}$ linearly independent in $\mathcal{P}$? [5]

(b) Let $a, b, c \in \mathbb{R}$. Show that the matrix

$$\begin{bmatrix} 1 & 1 & 1 \\ ab & bc & ca \\ a+b & b+c & c+a \end{bmatrix}$$

is nonsingular if and only if a, b, c are all distinct. [10]

2. Let X be a finite dimensional vector space over $\mathbb{C}$ and $T : X \to X$ be a linear map. Suppose $x_1, x_2 \in X$ are two nonzero vectors such that $Tx_1 = \lambda_1 x_1$, $Tx_2 = \lambda_2 x_2$ for some $\lambda_1, \lambda_2 \in \mathbb{C}$ and $\lambda_1 \neq \lambda_2$. Show that x_1, x_2 are linearly independent. [5]

3. Let $a \neq b \in \mathbb{R}$ and let

$$A = \begin{bmatrix} a & 0 & b \\ 2a & 0 & 2b \\ 3a & 0 & 3b \end{bmatrix}.$$

Find a rank factorization of A. [5]

4. Find the inverse of the matrix [12]

$$A = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 3 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

5. Find conditions on $a, b, c \in \mathbb{R}$ to ensure that the following system is consistent, and in that case, find the general solutions : [18]

$$\begin{array}{rcrcrcl} x & - & 3y & - & 2z & = & a \\ -x & - & 5y & + & 3z & = & b \\ 2x & - & 8y & + & 3z & = & c \end{array}$$

6. Let $M_n(\mathbb{R})$ denote the vector space of all $n \times n$ matrices with real entries. Find all linear maps $f : M_n(\mathbb{R}) \to \mathbb{R}$ that satisfy the relation $f(AB) = f(BA)$ for all $A, B \in M_n(\mathbb{R})$. [12]

7. What is the reduced echelon form of a matrix with full column rank? [8]

8. Prove that $\rho \begin{bmatrix} A & B \\ 0 & C \end{bmatrix} \geq \rho(A) + \rho(C)$; and that the inequality can be strict. [10+5]

9. Show that $\rho(A + B) = \rho(A) + \rho(B)$ if and only if [15]

   (a) $\mathcal{C}(A) \cap \mathcal{C}(B) = \{0\}$; and

   (b) $\mathcal{C}(A) \subseteq \mathcal{C}(A + B)$.

10. If A and B are real matrices, show that $\mathcal{C}(B) \subseteq \mathcal{R}(A)$ implies $\rho(AB) = \rho(B)$, but the converse need not be true. [10+5]

INDIAN STATISTICAL INSTITUTE
B. Stat. I: 2011–2012
Computational Techniques & Programming I
Semestral Examination

Date: 16. 11. 2011 Marks: 100 Time: 3 Hours

Answer any part of any question. The question is of 110 marks. The maximum marks you can get is 100. Please write all the part answers of a question at the same place.

1. (a) Write the recursive as well as non-recursive version of a function in C that can evaluate $f(n) = f(n-1) + f(n-2)$ with initial condition $f(0) = 0, f(1) = 1$.
   (b) Write a variable argument function in C that takes the number of arguments n as the first parameter and then reports the standard deviation of n floating point numbers supplied to the function as the following n arguments.
   (c) Write a function in C to check whether a string contains at least five vowels.
   (d) Write a function in C to calculate a^b, where a, b are both positive integers. Do not use the 'pow' function available in C.

   (5+5)+5+5+5 = 25

2. (a) Write down the algorithm for heap sort.
   (b) Derive the time complexity to construct a heap with n many points.
   (c) Construct a heap with the data set 1201, 137, 166, x, 1210, 191, 981, 367, 485, where x is two least significant digits of your roll number.

   10+10+10 = 30

3. (a) Describe the data structure of a node in a binary tree.
   (b) Write down a C program for insertion of data in a binary search tree.
   (c) Execute your program for insertion of data (explain with proper figures of binary search tree) on the data set available in the question 2c.
   (d) Write down a C-function for the inorder traversal in a binary tree.

   5 + 10 + 5 + 5 = 25

4. (a) Explain an efficient data structure for implementing polynomials.
   (b) Write a function in C programming language to multiply two polynomials.
   (c) Show how your function works (step by step) when the polynomials $x^{100} + 2x^3 + 5$ and $x^{78} + x^9 + 17$ are multiplied.

   5+10+5 = 20

5. Write a C program that can read two 3×3 matrices from a file and write the multiplication result in the same file. 10

INDIAN STATISTICAL INSTITUTE
Semester Examination: 2011-12
B. STAT 1st Year
STATISTICAL METHODS I
Full Marks: 100
[This question paper carries 120 marks. You can score maximum 100 marks]

18.11.2011 Duration: 3 hrs

1. The lives of 60 electric bulbs (in 100 hour unit) manufactured by Company A is given below.

9.5	15.2	25.1	0.1	4.5	15.1	10.3	1.5	5.5	3.3	8.5	10.6	7.0
8.4	4.0	7.7	23.4	10.3	5.7	3.7	2.7	1.2	4.3	0.1	15.2	21.6
0.6	6.6	21.2	19.2	4.5	5.9	3.9	21.4	5.9	15.0	10.2	23.3	5.2
12.6	5.9	3.5	14.2	22.7	16.5	1.8	26.2	3.7	17.3	15.4	3.8	4.7
10.2	5.8	8.4	11.2	2.7	20.1	8.9	6.2					

A frequency distribution of the same for a number of bulbs manufactured by Company B is also available and given in Table 1.

Table 1
Table showing frequency distribution of life of electric bulbs manufactured by Company B

Class intervals for life (in 100 hour unit)	Frequency
0.1 — 3.0	13
3.1 — 6.0	22
6.1 — 9.0	11
9.1 — 12.0	7
12.1 — 15.0	4
15.1 — 18.0	3
18.1 — 21.0	2
21.1 — 24.0	3
24.1 — 27.0	3
27.1 — 30.0	2

Later it was discovered that during the collection of data, the information on 23 bulbs manufactured by Company B were not included in the dataset, because 17 bulbs were found to be defective and the life hours of 6 bulbs were found to be 42.5, 58.7, 37.6, 51.7, 43.2, and 65.

(a) A statistician argues that since the 17 bulbs are defective, they should not be included in the analysis. Another statistician is in favour of including these 17 defective bulbs in the analysis assuming their life hours to be 0. Which argument would you opt for appropriate analysis of the data set for Company B and why?

(b) In view of your idea, write a report after comparing the frequency distribution of bulbs manufactured by the two companies using (i) at least one suitable diagram and (ii) some appropriate measures for summarizing the data.

(c) Suppose that for some reason the two datasets are combined. Draw a suitable diagram to represent the combined data. Also obtain an appropriate measure of central tendency and dispersion for the combined data. (6+16+9=31)

2. Let there be k datasets consisting of n_i pairs of observations with r_i as the correlation coefficient for the i-th ($i = 1,2,\dots,k$) dataset. If the means and standard deviations for all the datasets are same, show that the correlation coefficient for the combined dataset would be of the form $\sum_{i=1}^{k} \lambda_i r_i$. Provide explicit expression for λ_i. (6)

3. Let (x_i, y_i), $i = 1,2,\dots,n$, be n pairs of observations. Derive the correlation coefficient if
   (a) x_n and y_n are very large compared to other observations
   (b) only x_n is very large compared to x_i, $i = 1,2,\dots,n-1$
   Also explain your result intuitively with proper justification. (5+5+3+3=16)

4. (a) What does skewness indicate? Suggest a measure of skewness with justification.
   (b) What does kurtosis indicate? Suggest a measure of kurtosis.
   (c) "Kurtosis of two symmetric frequency distributions are same, yet they differ in shape": is this statement true or false? Justify your answer.
   (d) If b_1 and b_2 are the measures of skewness and kurtosis respectively based on moments, show that $b_2 \geq b_1 + 1$. When does the equality hold? (7+7+5+6=25)

5. Some statisticians argue that a measure of dispersion can also be suggested that does not depend on any particular measure of central tendency. They emphasise to use the mutual differences $x_i - x_j$, $i, j = 1,2,\dots,n$ of the values of x. In view of this let us define two such measures as $\Delta_1 = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}|x_i - x_j|$ and $\Delta_2 = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - x_j)^2$. Show that,
   (a) $\Delta_1 \leq \sqrt{2\Delta_2}$
   (b) $\min(x_1, x_2) = \frac{1}{2}\{x_1 + x_2 - |x_1 - x_2|\}$
   (c) $\Delta_1 = 2\bar{x}(1 + \frac{1}{n}) - \frac{4}{n^2}\sum_{i=1}^{n} i x_i$, where $x_1 \geq x_2 \geq \dots \geq x_n$ [you can use (b)] (8+4+6=18)

6. A person was asked to calculate the mean and standard deviation (s.d.) of 50 observations. He noticed that 4 out of 50 observations were zeros and hence ignored them and calculated the mean and s.d. of the remaining observations as 42.96 and 13.8812. Find the correct values of mean and s.d. of the 50 observations. You have to prove any result that you use. (8)

7. The difference between upper face length (y) and nasal length (x), both measured in mm., is given for 15 Indian adult males:

   15, 13, 15, 15, 12, 19, 13, 12, 19, 14, 15, 19, 11, 15, 14

   (a) Calculate the correlation coefficient of x and y and the linear regression of y on x, given that for these people $\bar{x} = 49.34$ $mm.$, $\bar{y} = 64.07$ $mm.$, $s_x = 3.53$ $mm.$ and $s_y = 4.30$ $mm.$
   (b) Another statistician also collected data on another 20 individuals from the same population and the means, variances and correlation coefficient are found to be same as that in the first dataset. If we combine these two datasets, what would be the new regression line of y on x in the combined dataset? Comment on the old and the new regression lines with intuitive justification. (8+6+2=16)

INDIAN STATISTICAL INSTITUTE

Semestral Examination – Semester I : 2011-2012
B.Stat. (Hons.) I Year
Probability Theory I

Date : 22.11.11 **Maximum Score : 60 pts** Time : 3 Hours

Note : This paper carries questions worth a **total** of **80 points**. Answer as much as you can. The **maximum** you can score is **60 points**.

1. An auto-insurance company classifies drivers into three classes – low-risk, medium-risk and high-risk. Their records indicate that the probabilities of being involved in an accident over a 1-year span are $0.1, 0.3$ and 0.8 respectively for the low-risk, medium-risk and high-risk drivers. A randomly chosen driver, equally likely to be a low-, medium- or high-risk driver, is known not to have been involved in an accident during the past 1-year period. What is the probability that
(a) the selected driver is a medium-risk driver?
(b) the selected driver will have an accident during the next 1-year period?

(6+6)=**[12]**

2. Let X and Y be independent random variables having Poisson distributions with parameters λ and μ respectively.
(a) Derive the distribution of the random variable $X + Y$.
(b) Find the conditional probability that $X = 3$ given $X + Y = 8$. (6+6)=**[12]**

3. With usual notations for Random Walk, use equivalence of two sets of paths to show that **[6]**

$$P(S_1 \geq 0, \ldots, S_{2n-1} \geq 0, S_{2n} = 0) = 2f_{2n+2}.$$

4. If balls are randomly placed one after another in n cells, find
(a) the expected number of cells occupied after r balls have been placed;
(b) the expected number of balls placed until all the n cells are occupied.

(6+6)=**[12]**

5. Let X be a random variable taking non-negative integer values.
(a) Show that, for any $N > 1$,

$$\sum_{n=0}^{N-1} n(n+1)P(X = n) = 2\sum_{n=1}^{N} nP(X \geq n) - N(N+1)P(X \geq N).$$

(b) Show that the random variable $X(X + 1)$ has finite expectation if and only if the infinite series $\sum_{n=1}^{\infty} nP(X \geq n)$ converges, and in that case,

$$E[X(X+1)] = 2\sum_{n=1}^{\infty} nP(X \geq n).$$

(6+6)=**[12]**

6. A box has $m+n$ chips numbered $1, 2, \ldots, m+n$. A set of n chips are drawn at random from the box (without replacement).
Let X denote the count of chips among the ones drawn whose numbers exceed the numbers on all the chips that remain in the box and let Y denote the lowest number on the chips drawn.
(a) Find the probability mass function of X.
(b) Show that $E(X) = n/(m+1)$. [Hint: Use an identity involving binomial coefficients, which happens to be also a consequence of (a).]
(c) Find the joint distribution of (X, Y). (6+6+6) =[18]

7. A box has a total of n balls, some green and the others red, with at least one of each colour. A ball is drawn at random, its colour noted and it is discarded. After this, successive draws are made without replacement and the drawn balls discarded, as long as the balls are of the same colour as the first drawn ball. When a ball of colour different from the first drawn ball shows up for the first time, that ball is put back in the box, and, the whole experiment starts from the beginning with the remaining balls in the box. Show that the probability that the last drawn ball is green is $\frac{1}{2}$. [Hint: May try induction on n.] [8]

INDIAN STATISTICAL INSTITUTE
Semester Examination: 2011-12 (Back Paper)
B. STAT 1st Year
STATISTICAL METHODS I
Full Marks: 100

22.12.2011 Duration: 3 hrs

1. The word lengths of each of the 90 words in a poem are shown below:

5	4	3	5	8	6	6	3	4	3
4	4	5	8	2	6	7	6	4	5
6	4	9	6	4	2	2	2	9	2
3	3	3	2	4	7	7	2	4	4
4	3	4	4	2	4	4	9	3	7
4	5	12	6	3	5	2	5	10	3
5	7	3	3	3	6	2	5	3	3
3	2	4	5	8	5	3	4	4	6
7	2	3	5	5	5	3	2	4	5

(a) Construct a frequency table and also obtain the relative frequencies and the cumulative frequencies (of the 'less-than' type).
(b) Represent the data using an appropriate diagram.
Suppose that another frequency distribution for the words of another poem is as follows:

Word length	Frequency
1	6
2	7
3	7
4	10
5	9
6	6
7	4
8	3
9	0
10	1

(c) Compare these two frequency distributions using (i) at least one diagram, (ii) some appropriate measures of summarization of data and comment. (8+6+20=34)

2. Why do you use coefficient of variation? Give examples of two situations where coefficient of variation would be appropriate. (4+6=10)

3. Suppose that the correlation coefficient between two variables x and y is zero. Define a new variable $z = xy$. Then,
(a) Express s_z^2 in terms of means and variances of x and y
(b) If C_x, C_y and C_z are the coefficients of variation of the three variables, express C_z in terms of C_x and C_y (6+4=10)

4. Let (x_i, y_i), $i = 1,2,\ldots,n$, be n pairs of observations. Derive the correlation coefficient if
(a) x_n and y_{n-1} are very large compared to other observations
(b) explain your result intuitively with proper justification. (8+6=14)

P.T.O.

5. (a) Under what circumstance will you use correlation ratio e_{yx}?

(b) Show that the correlation ratio e_{yx} is the simple correlation coefficient between y and the array mean of y corresponding to x. (6+8=14)

6. Consider the following data:

x	-4	-3	-2	-1	0	1	2	3	4
y	0.1	2.5	3.4	3.9	4.1	3.8	3.5	2.8	0.3

Analyse the data and comment on your findings. (18)

INDIAN STATISTICAL INSTITUTE

First Semester Backpaper Examination (2011–2012)

B STAT I

Vectors & Matrices I

Date : 28.12.2011 Maximum Marks : 100 Time : 3 hrs.

Precisely justify all your steps. Carefully state all the results you are using.

Let $M_n(\mathbb{R})$ denote the vector space of all $n \times n$ matrices with real entries. For a matrix A, let $\mathcal{R}(A)$ and $\mathcal{C}(A)$ denote respectively the row space and column space of A.

1. Prove or disprove:

   (a) If A, B, C are pair-wise disjoint subsets of a vector space X such that $A \cup B$ and $A \cup C$ are both bases of X, then span(B) = span(C). [5]

   (b) There exists a linear map $T : \mathbb{R}^2 \to \mathbb{R}^4$ such that [5]

   $$\text{Range}(T) = \{(x_1, x_2, x_3, x_4) \in \mathbb{R}^4 : x_1 + x_2 + x_3 + x_4 = 0\}.$$

   (c) There exist a linear map $T : \mathbb{R}^2 \to \mathbb{R}^3$ such that [5]

   $$\text{Range}(T) = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_1 + x_2 + x_3 = 0\}.$$

   (d) If A is an $m \times n$ matrix with rank m and B is an $n \times m$ matrix with rank m, the AB has an inverse. [5]

2. In each of the following cases, decide whether the given subset V of $M_n(\mathbb{R})$ a subspace of $M_n(\mathbb{R})$? If yes, also find the dimension of V. [5+5]

   (a) V = the set of all singular matrices in $M_n(\mathbb{R})$.

   (b) V = the set of all matrices in $M_n(\mathbb{R})$ with trace zero.

3. Find the inverse of the matrix [10]

$$A = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 9 & 5 & 4 \end{bmatrix}.$$

4. Let

$$\begin{aligned} a_1x &+ b_1y &= c_1 \\ a_2x &+ b_2y &= c_2 \\ a_3x &+ b_3y &= c_3 \end{aligned}$$

be the equations of three straight lines in $\mathbb{R}^2$ that are not concurrent, but intersect pairwise. What can you say about the consistency of the system?

Also determine the rank of the matrix [5+5]

$$\begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{bmatrix}.$$

5. Find the general solution to the following system : [10]

$$\begin{aligned} x &- 3y &+ 4z &= 0 \\ 2x &- y &- 2z &= 5 \\ 5x &- 2y &- 3z &= -8 \end{aligned}$$

6. Let X, Y be vector spaces over $\mathbb{K}$ and $T \in \mathcal{L}(X, Y)$.

   (a) Assume X is finite dimensional. Show that T is onto if and only if there exists $S \in \mathcal{L}(Y, X)$ such that $TS = I_Y$, the identity map on Y. [10]

   (b) Assume Y is finite dimensional. Show that T is one-to-one if and only if there exists $S \in \mathcal{L}(Y, X)$ such that $ST = I_X$, the identity map on X. [10]

7. If A and B are square matrices of order n, show that [5]

$$\rho(AB - I_n) \leq \rho(A - I_n) + \rho(B - I_n).$$

8. If A and B are real matrices, show that $\mathcal{C}(B) \subseteq \mathcal{R}(A)$ implies $\rho(AB) = \rho(B)$, but the converse need not be true. [10+5]

INDIAN STATISTICAL INSTITUTE
B. Stat. I: 2011-2012
Computational Techniques & Programming I
Back Paper Examination

Date: 29.12.2011 | Marks: 100 | Time: 3 Hours

Answer any five questions. Each question is of 20 marks. Please write all the part answers of a question at the same place.

1. (a) Write the recursive as well as non-recursive version of a function in C that can evaluate $f(n) = 5f(n-1)+7f(n-2)$ with the initial conditions $f(0) = 10, f(1) = 11$.
   (b) Write a function in C to check whether a string contains at least three vowels.
   (c) Write a function in C to calculate x^y, where x, y are both positive integers. Do not use the 'pow' function available in C.

   (5+5)+5+5 = 20

2. (a) Write a function in C that finds the GCD as well as the LCM of two unsigned integers.
   (b) Show how the function executes when the two integers are 462 and 4845.
   (c) Will your functions work properly for all pairs of unsigned integers? Explain.

   10 + 5 + 5 = 20

3. (a) Write down the algorithm for heap sort.
   (b) Construct a heap with the data set 120, 37, 66, 210, 91, 98, t, 67, 45, where t is two least significant digits of your roll number.

   10+10 = 20

4. (a) Describe the data structure of a node in a binary tree.
   (b) Write down a C program for insertion of data in a binary search tree.
   (c) Execute your program for insertion of data (explain with proper figures of binary search tree) on the data set available in the question 2c.

   5 + 10 + 5 = 20

5. (a) Explain an efficient data structure for implementing polynomials.
   (b) Write a function in C programming language to add two polynomials.
   (c) Show how your function works (step by step) when the polynomials $x^{102} + 2x^{17} + 5$ and $x^{78} + x^{17} + 17$ are added.

   5+10+5 = 20

P. T. O.

6. (a) Explain the parameter passing strategy (to functions) in C programming language with the example of "swap" function.

   (b) Write a C program that can read two 3×3 matrices from a file and write the multiplication result in the same file.

$10+10 = 20$

7. (a) Explain how can you implement a Chess board in C.

   (b) Consider a Chess board when a game is continuing. There can be m many pieces in the board ($2 \leq m \leq 32$). Write a C program that will report the largest empty rectangle given the current situation on the board.

$5 + 15 = 20$

8. (a) What is the output of the following piece of C code?

```
int i, k = 1, n = 5;
for (i = 0; k < n+1; i = k-i)
        { printf("%d\n", k); k = k+i; }
```

   (b) Consider the following C program.

```
#include <stdio.h>
main(){
        int arr[10][10], i, j;
        for (i = 0; i < 10; i++)
        for (j = 0; j < 10; j++)
                arr[i][j] = i*j + 1;
        printf("%x\n", arr); printf("%x\n", arr+1);
        printf("%x\n", *arr); printf("%x\n", *arr+1);
        printf("%x\n", **arr);
}
```

The first line of output of this program is eef78650. What are the rest of the outputs? Give proper justification to your answer.

$10 + 10 = 20$

INDIAN STATISTICAL INSTITUTE

MID-TERM EXAMINATION (2011–12)

B. STAT. I YEAR

ANALYSIS II

Date : 20.02.2012 Maximum Marks : 80 Time : $2\frac{1}{2}$ hours

The question carries 90 marks. Maximum you can score is 80. Precisely justify all your steps. Carefully state all the results you are using.

1. A function f is defined on [0,1] by

$$f(x) = \begin{cases} x & \text{if} \quad x \text{ is rational} \\ 1 - x & \text{if} \quad x \text{ is irrational} \end{cases}$$

Compute $\underline{\int_0^1} f(x)\,dx$, $\overline{\int_0^1} f(x)\,dx$ and decide whether $f \in \mathcal{R}[0,1]$. [15]

2. Let $f : [a,b] \to \mathbb{R}$ be continuous. Show that for any partition P of $[a,b]$, there is a marking of P such that the Riemann sum $S(P,f) = \int_a^b f(x)\,dx$. [10]

3. Let $f : [a,b] \to \mathbb{R}$ be bounded. Prove that given $\varepsilon > 0$, there is a partition P_ε of $[a,b]$ such that for every refinement P of P_ε, there is a marking of P such that [10]

$$\left| \overline{\int_a^b} f(x)\,dx - S(P,f) \right| < \varepsilon.$$

4. Let $f \in \mathcal{R}[a,b]$. Show that for every $\varepsilon > 0$, there is a continuous function g on $[a,b]$ such that [10]

$$\int_a^b |f(x) - g(x)|dx < \varepsilon.$$

5. Test the convergence of the integral $\int_0^1 |\log x|\,dx$. [10]

6. Show that the improper integral

$$\int_0^\infty \frac{1}{x^p(1+x)^q}\,dx$$

converges if and only if $0 < 1 - p < q$. [15]

7. (a) Let $f, g \in \mathcal{R}[a,b]$ and $D = \{x \in [a,b] : f(x) \neq g(x)\}$ be a set of measure zero. Show that $\int_a^b f(x)\,dx = \int_a^b g(x)\,dx$. [15]
[Hint : Let $a < c < d < b$. Then $(c,d) \setminus D \neq \emptyset$.]

   (b) Let $f, g : [a,b] \to \mathbb{R}$ be bounded. If $f \in \mathcal{R}[a,b]$, and $D = \{x \in [a,b] : f(x) \neq g(x)\}$ is a set of measure zero, can you conclude that $g \in \mathcal{R}[a,b]$? [5]

INDIAN STATISTICAL INSTITUTE

Mid-semestral Examination – Semester II : 2011-2012

B.Stat. (Hons.) I Year

Probability Theory II

Date : 23.02.12 | **Maximum Score : 80 pts** | Time : $2\frac{1}{2}$ Hours

Note : This paper carries questions worth a **total of 98 POINTS**. Answer as much as you can. The **MAXIMUM** you can score is **80 POINTS**.

1. A Bernoulli trial with success probability p $(0 < p < 1)$ is repeated independently until n successes are obtained. For $m \leq n$, find the conditional distribution of the number of successes in the last m trials, given the total number of trials.

[**16**]

2. (a) In independent throws of a die, what is the expected number of throws needed to get two consecutive sixes? (Hint: You may consider conditioning on the number of throws needed to get the first six.)
(b) Using your answer in (a) or otherwise, find the expected number of throws needed to get three consecuive sixes.

(8+8)=[**16**]

3. A random variable X has density function $f(x) = \begin{cases} c(1+x)^{-3} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$
(a) Find c and find the distribution function of X.
(b) Find the probability $P[X/(3+X^2) < 1/4]$.
(c) If lightbulbs produced by a manufacturer has life-time with the above density, what is the probability that out of four such bulbs, at least one would survive for at least 4 units of time?

(8+8+8)=[**24**]

4. (a) Consider the function $F(x) = \begin{cases} 0 & \text{if } x < -1 \\ (2-x^2)/4 & \text{if } -1 \leq x < 0 \\ 1/2 & \text{if } 0 \leq x < 1 \\ (x+3)/6 & \text{if } 1 \leq x < 2 \\ 1 & \text{if } x \geq 2 \end{cases}$.
If X is a random variable with distribution function F, then find the following probabilities:
(i) $P[X > 3]$; (ii) $P[X \leq 17/3]$; (iii) $P[9/2 \leq X < 32/3]$; (iv) $P[X$ is an integer$]$.

(4×4)=[**16**]

5. Let X be a random variable with the $B(3,3)$-distribution.
(a) Find the density function of $Y = \log(1-X)$.
(b) Find the distribution function of $Z = \min\{X^2, 1-X^2\}$. Does Z have a density? If so, find it.

(8+8)=[**16**]

6. Show that if X is a random variable with a continuous distribution function F, then the random variable $Y = F(X)$ has uniform distribution on $(0,1)$.

[**10**]

INDIAN STATISTICAL INSTITUTE

Second Semestral Examination: 2011-12

Course Name : B Stat. I yr. Subject Name : **Vector and Matrices II**

Date : 27th February 2011 Maximum Marks: 30 Duration: 3 Hours

Note: Attempt all questions. Marks are given in brackets. State the results you want to use.

Problem 1 (3). To each permutation π over $\{1, \ldots, n\}$, we associate a matrix $P_\pi = ((p_{i,j}))$ such that $p_{i,j} = 1$ if $j = \pi(i)$, $1 \le i \le n$, otherwise we define $p_{i,j} = 0$. Which of the followings is true and justify your answer.

1. For all permutations π_1, π_2, $P_{\pi_1} P_{\pi_2} = P_{\pi_1 o \pi_2}$.
2. For all permutations π_1, π_2, $P_{\pi_1} P_{\pi_2} = P_{\pi_2 o \pi_1}$.
3. None of the above.

Problem 2 (4). Prove of disprove that for any matrices $B_{m\times n}, C_{n\times m}$ we have

$$\det(I - CB) = \det(I - BC).$$

Problem 3 (5). $A_{n\times n}$ is called skew symmetric matrix of size n if $A^{tr} = -A$. Find all positive integers n, for which there are non-singular skew symmetric matrices. Justify your answer.

Problem 4 (3+3 = 6). Define $\|x\|_\infty : \mathbb{R}^n \to \mathbb{R}^n$ as $\|x\|_\infty = \max_{1\le i\le n} |x_i|$ where $x = (x_1, \ldots, x_n)$. Prove that $\|\cdot\|_\infty$ is a norm. Moreover, prove that there is no inner product $\langle\cdot,\cdot\rangle$ such that $\|x\|_\infty = \sqrt{\langle x, x\rangle}$ for all $x \in \mathbb{R}^n$.

Problem 5 (3). Prove that $tr(A) > 0$ for any positive definite matrix A.

Problem 6 (4). Let A be a non-negative definite matrix then prove that $\det(I + A) \neq 0$.

Problem 7 (4). Prove or disprove the following:
There exists a constant c such that $adj(adj(A)) = cA$ for all real matrix A" where $adj(A)$ is the adjoint matrix of A.

Problem 8 (3). Let U be a subspace of W which is again a subspace of V. Let Proj_U (similarly Proj_W) denote the orthogonal projection onto U (and W respectively). Prove that for all $x \in V$,

$$\mathsf{Proj}_U(\mathsf{Proj}_W(x)) = \mathsf{Proj}_U(x).$$

Problem 9 (3). Let $T : V \to V$ be an isometry over a real inner product space V. Prove that the set of fixed points of T is exactly the set

$$\{x \in V, \langle T(x), x\rangle = \|x\|^2\}.$$

INDIAN STATISTICAL INSTITUTE
Mid-Semester Examination: 2011-12
B. STAT 1st Year
STATISTICAL METHODS II
Full Marks: 40
[This question paper carries 50 marks. You can score maximum 40 marks]

29.02.2012 Duration: 3 hrs

1. Based on 100 observations on four variables x_1, x_2, x_3 and x_4, it as found that $\bar{x}_1 = 4.91661$, $\bar{x}_2 = 1.702444$, $\bar{x}_3 = 14.106000$, and $\bar{x}_4 = 2.22895$ and the sample variance-covariance matrix is

$$S = \begin{pmatrix} 1.37771 & -0.19968 & 0.08646 & 0.06799 \\ & 1.70846 & 0.31959 & -0.05643 \\ & & 9.10327 & -0.07651 \\ & & & 0.45832 \end{pmatrix}$$

where the (i, j)th element represents $\text{cov}(x_i, x_j)$, $i, j = 1,2,3,4$.

(a) If $x_2 = 5.71295$, $x_3 = 22.11572$ and $x_4 = 4.12735$, predict the value of x_1 based on linear regression equation.

(b) What proportion of the variability of x_1 is not explained by the above regression?

(c) Explain if x_3 is useful in the above prediction of x_1.

(d) Can you comment on any possible modification of this prediction? Justify your answer? (8+3+3+3=17)

2. Suppose R 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) If $r_{1j} = \alpha$ for $j = 2,3,\ldots,p$ and $r_{ij} = \beta$ for $i, j = 2,3,\ldots,p; i \neq j$, obtain $r_{12.34\ldots p}$. (5+7=12)

3. Suppose you have a random number generator from U(0,1). Explain how you would generate two observations X and Y from $Bin(10, \frac{2}{3})$ such that the correlation between X and Y is 0.7. (7)

4. A DNA sequence comprises four types of nucleotides A, T, G, or C with same probability of occurrence at any specific position in a sequence. The proportion of the nucleotides A, T, G, or C in the population is p_A, p_T, p_G, and p_C ($p_A + p_T + p_G + p_C = 1$) and the population may be treated as of infinite size.

(a) What is the expected length of a DNA sequence one has to scan so as to observe all the four types of nucleotides?

(b) What is the chance that in a sequence of 100 nucleotides, all four will not be selected? (7+7=14)

INDIAN STATISTICAL INSTITUTE

Mid-Semestral Examination: 2011 – 12

Course Name: B. STAT. I YR.

Subject: Computational Techniques and Programming II

C2 – C3.

Date: 13. 02. 2012

Maximum Marks : 30 Duration: 1 hr.

Answer any 5 questions. All questions carry equal marks.

1. Consider the following nonlinear equation:

$$e^x - x^2 + 3x - 2 = 0$$

   i. Prove that the equation has a unique root in the interval [0,1].
   ii. Consider the bisection algorithm starting with the interval [-1,1]. Find the minimum number of iterations required to achieve an approximation with absolute error less than 10^{-5}.

2. We try to solve the following equation by Newton's Method as follows:

$$f(x) = 1 - \cos(x) \quad \text{on } [-\frac{\pi}{2}, \frac{\pi}{2}]$$

   (a) Write down the sequence of Newton's Method $\{x_n\}_{n=0}^{\infty}$.
   (b) If $x_0 \in \{x \in \mathbb{R} | 2x \sin x + \cos x - 1 = 0\}$ and $x_0 \neq 0$, show that the sequence, with such an initial value, does not converge.

3. For the following set of data,

x	0	1	2
$f(x)$	1	1	0

   compute a natural cubic spline function such that

$$S''(0) = S''(2) = 0.$$

4. Show how to use Newton's method to solve the equation below for x:

$$\int_{-x}^{x} e^{\sin t}\, dt = 1.$$

   Use the Composite Simpson's rule with $n = 4$ intervals for the integral.

   You do not have to perform any numerical calculations.

5. (a) Show that the fixed point iteration

$$p_n = \frac{p_{n-1}^2 + 3}{5}, \qquad n = 1, 2, \dots$$

converges for any initial $p_0 \in [0, 1]$.

(b) Estimate how many iterations n are required to obtain an absolute error $|p_n - p|$ less than 10^{-4} when $p_0 = 1$. No numerical value needed, just give an expression for n.

6. Find the Doolittle LU decomposition of A where the matrix A is

$$A = \begin{bmatrix} 8 & 1 & -1 \\ 2 & 1 & 9 \\ 1 & -7 & 2 \end{bmatrix}.$$

7. Solve the set of linear equations in the form $Ax = b$ by using the Gaussian Elimination procedure. Show the necessary calculations. Here

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 4 \\ 2 & 1 & 1 \end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix} 6 \\ 7 \\ 4 \end{pmatrix}$$

8. Let $x_0, x_1, \dots, x_n$ be distinct points and $l_j(x) = \frac{(x-x_0)(x-x_1)\cdots(x-x_{j-1})(x-x_{j+1})\cdots(x-x_n)}{(x_j-x_0)(x_j-x_1)\cdots(x_j-x_{j-1})(x_j-x_{j+1})\cdots(x_j-x_n)}$.

i. Show that

$$\sum_{j=0}^{n} x_j^k l_j(x) \equiv x^k \quad , (k = 1, \dots, n).$$

ii. Show that

$$\sum_{j=0}^{n} (x_j - x)^k l_j(x) \equiv 0 \quad , (k = 1, \dots, n).$$

9. Prove or disprove that in general the "Least Squares line" passes through the mean $\left(\frac{\sum x_i}{n}, \frac{\sum y_i}{n}\right)$ of the given set of points.

INDIAN STATISTICAL INSTITUTE

SECOND SEMESTER SEMESTRAL EXAMINATION (2011–12)

B. STAT. I YEAR

ANALYSIS II

Date : 23.04.2012 | Maximum Marks : 100 | Time : $3\frac{1}{2}$ 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. (a) Let $f : [a, b] \to \mathbb{R}$ be a bounded function. Define

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

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

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

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

2. Test the convergence of the integral

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

3. Let $f : \mathbb{R} \to \mathbb{R}$ be continuous. Define

$$f_n(x) = f\left(x + \frac{1}{n}\right), \quad x \in \mathbb{R}, n \geq 1.$$

Does $\{f_n\}$ converge uniformly to f on $\mathbb{R}$? If yes, prove it.

If not, will it work under some stronger assumption? Justify! [10]

4. 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]

5. Decide whether the power series $\sum_{n=1}^{\infty} \frac{n^3[\sqrt{2} + (-1)^n]^n}{3^n} x^n$ converges or diverges at the points $x = 1$ and $x = 2$. [8]

6. Starting from a geometric series and precisely justifying all your steps prove that $\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} = \log 2.$ [10]

7. Suppose $f : [-\pi, \pi] \to \mathbb{R}$ is defined by

$$f(x) = \begin{cases} 0 & \text{if} \quad x \in (-\pi, 0) \\ 2 & \text{if} \quad x \in (0, \pi) \end{cases}$$

(a) Determine the value of f at the points $-\pi$, 0 and π, so that the Fourier series for f converges to f at every $x \in [-\pi, \pi]$? Briefly justify your answer. [10]

(b) Compute the Fourier series of f and verify your result in (a). [15]

(c) With these values, is the convergence uniform on $[-\pi, \pi]$? [5]

(d) Show that

i. $\sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1} = \frac{\pi}{4}.$ [5]

ii. $\sum_{n=0}^{\infty} \frac{1}{(2n+1)^2} = \frac{\pi^2}{8}$ [10]

iii. $\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}.$ [5]

INDIAN STATISTICAL INSTITUTE

Second Semestral Examination: 2011-12

Subject Name : Vector and Matrices II (B Stat. I-year.) Date : 27/04/12

Maximum Marks: **60 (Attempt All)** Duration: $3\frac{1}{2}$ Hours

1. Answer of each question should start from a new page mentioning the question number.
2. **State results** and **provide clear explanations** which are used in your answers.

Problem-01. Prove for no integer matrix $H_{n\times n}$ with odd $n > 1$, $H^{\text{tr}}H = n \cdot \mathbf{I}_n$. [4]

Problem-02. Suppose for some $u, v \in V$, $\|2u - 3v\| = 2\|u\| + 3\|v\|$ where V is a f.d. real inner product space inducing the norm $\|\cdot\|$, then prove $u = \alpha \cdot v$ for some $\alpha \leq 0$. Is $\langle x, y\rangle := x_1(y_1 + 2y_3) + (x_2 + 2x_3)y_2 - x_3y_3$ an inner product over $\mathbb{R}^3$? Justify. [5 + 3 = **8**]

Problem-03. Prove that minimal polynomials of similar matrices are identical. [4]

Problem-04. Prove for any stochastic matrix: (1) 1 is an eigenvalue, (2) $|\lambda| \leq 1$, for all eigenvalues λ, (3) There exist real eigenvectors for real eigenvalues. [3+4+3 = **10**]

Problem-05. Denote $A \geq B$ (or $A > B$) for symmetric matrices A and B if $(A - B)$ is n.n.d (or p.d.) matrix. Prove that if $A \geq B > \mathbf{0}$ then $B^{-1} \geq A^{-1} > \mathbf{0}$. [6 + 4 = **10**]

Problem-06. Given an invertible matrix $A_{n\times n}$ and its characteristic polynomial $\chi_A(\lambda) = c_0 + c_1\lambda + \ldots + c_{n-1}\lambda^{n-1} + \lambda^n$, find A^{-1} in terms of the coefficients c_i's and non-negative integral powers of A? [5]

Problem-07. Prove or disprove: *For all matrices A with $|\lambda| \leq 1$ for its all eigenvalues λ, then all entries of the matrices A^n, $n \in \mathbb{N}$ are bounded above.* [5]

Problem-08. (1) Prove that a matrix is nilpotent if and only if zero is the only eigenvalue. (2) Prove that any non-zero nilpotent matrix is not diagonalizable. [6 + 4 =**10**]

Problem-09. Let $x_1, x_2, \ldots, x_n \in \mathbb{R}^n$ be n linearly independent vectors and we denote $n \times k$ matrix $A_k = [x_1\ x_2\ \ldots\ x_k]$. The recursive definition of volume of parallelepiped $\mathcal{P}_k$ with edges $x_1, \ldots, x_k$: $\text{vol}(\mathcal{P}_k) = \|\mathbf{n}_{A_{k-1}}(x_k)\| \times \text{vol}(\mathcal{P}_{k-1})$, $2 \leq k \leq n$, $\text{vol}(\mathcal{P}_1) = \|x_1\|$ where $\mathbf{n}_{A_{k-1}}(x_k) := x_k - \mathbf{proj}^{\perp}_{\mathcal{C}(A_{k-1})}(x_k)$ is the normal vector of x_k onto the column space of A_{k-1}.

(1) Prove that for all $2 \leq k \leq n$, $\det(A_k^{tr}A_k) = \det(B^{tr}B)$ where $B = [A_{k-1}\ \mathbf{n}_{A_{k-1}}(x_k)]$.
(2) Prove that $\text{vol}(\mathcal{P}_k) = \sqrt{\det(A_k^{tr}A_k)}$, $1 \leq k \leq n$ and
(3) $\text{vol}(\mathcal{P}_n) = |\det(A_n)|$. [4 + 4 + 2= **10**]

Problem-10. Let $P_i(x) = a_{i,1} + a_{i,2}x + \ldots + a_{i,n}x^{n-1} \in \mathbb{Z}[x]$ be integer polynomials, and $\alpha_i = i + \sqrt{2}$, $1 \leq i \leq n$. Show that $\det(P)$ is an integer where $P_{n\times n} = \big(\big(\ P_i(\alpha_j)\ \big)\big)$. [6]

INDIAN STATISTICAL INSTITUTE

Second Semestral Examination: 2011 - 12

Course Name: B. STAT. I YR.

Subject: Computational Techniques and Programming II

Date: 04. 05. 2012

Maximum Marks : 100 **Duration: 3 hrs.**

Answer any 5 questions. All questions carry equal marks.

1. a) Prove that for the computer representation of a real number x in base β, the value of the unit round is given by (t being the number of base β digits in the representation):

$$\delta = \begin{cases} \beta^{-t+1} & \text{for chopped definition of } fl(x) \\ \frac{1}{2}\beta^{-t+1} & \text{for rounded definition of } fl(x) \end{cases}$$ [8]

b) Show that the sequence generated by $x_n = \frac{1}{2}x_{n-1} + \frac{7}{2x_{n-1}}$ for $n \geq 2$ and for any $x_0 \geq 2$ converges to $\sqrt{7}$. [6]

c) It is known that Newton's iteration usually converges quadratically, but it is not always the case. Consider the following very simple equation:

$$x^2 = 0, \quad \text{on } [0,1]$$

Write down the sequence of Newton's iteration to solve the above equation with starting point $x_0 = \frac{1}{2}$. Now find the order of convergence of the sequence. [6]

2. (a) Find the quadratic polynomial $P(x)$ that interpolates $f(x) = \sin\left(\frac{\pi x}{2}\right)$ at $x = -1, 0, 1$.

Show that $|f(x) - P(x)| \leq \frac{\pi^3}{72\sqrt{3}}$ for $|x| \leq 1$. [12]

(b) A natural cubic spline S is defined as:

$$S(x) = \begin{cases} S_0(x) = 1 + B(x-1) - D(x-1)^3, & \text{if } 1 \leq x < 2, \\ S_1(x) = 1 + b(x-2) - \frac{3}{4}(x-2)^2 + d(x-2)^3, & \text{if } 2 \leq x < 3. \end{cases}$$

If S interpolates the data (1,1), (2,1), and (3,0), find B, D, b, and d. [8]

3. Let $A \in \mathbb{R}^{n\times n}$, and E is the perturbation of A and satisfies $\|A^{-1}\|\|E\| < 1$. $b \in \mathbb{R}^n$ and not equal to zero, and $\|I\| = 1$, if x and $\tilde{x}$ respectively satisfy the following linear systems:

$$Ax = b, (A+E)\tilde{x} = b + \delta b,$$

then, prove

$$\frac{\|\tilde{x} - x\|}{\|x\|} \leq \frac{\kappa(A)}{1 - \kappa(A)\frac{\|E\|}{\|A\|}}\left(\frac{\|E\|}{\|A\|} + \frac{\|\delta b\|}{\|b\|}\right).$$

where $\kappa(A)$ is the condition number of A and δb is perturbation of b. [20]

4. a) Find $\alpha_0, \alpha_1, \alpha_2$ and α_3 such that the quadrature formula

$$\int_{-1}^{1} g(t)dt \approx \alpha_0 g(-1) + \alpha_1 g(-\frac{1}{3}) + \alpha_2 g(\frac{1}{3}) + \alpha_3 g(1) \quad (1)$$

is exact for all polynomials of degree less than or equal to 3. [6]

b) Using the quadrature formula (1), derive the corresponding quadrature formula for computing $\int_a^b g(x)dx$. [5]

c) Compute the integral:

$$\int_0^2 e^{-x^2} dx$$

by using the quadrature formula of part 4. (a). [9]

5. a) Use the Gerschgorin theorem to determine the approximate location of the eigenvalues of the following matrix: $\begin{bmatrix} -2 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & -1 & 3 \end{bmatrix}$. [5]

b) Use the power method to calculate the dominant eigenvalue and the associated eigenvector of the following matrix: $\begin{bmatrix} 1 & 3 & -2 \\ -1 & -2 & 3 \\ 1 & 1 & 2 \end{bmatrix}$. Show the necessary calculation steps. [15]

6. a) Examine the convergence of the Gauss-Siedel iterative method with initial approximation $(x_1, x_2, x_3) = (0,0,0)$ for the following system of linear equations: [5]

$$2x_1 - 3x_2 = -7,$$
$$x_1 + 3x_2 - 10x_3 = 9,$$
$$3x_1 + x_3 = 13.$$

b) Solve the following differential equation by using fourth order Runge-Kutta method and show the necessary computational steps:

$$\frac{dy}{dx} = 2xy, \quad y(0) = 0.5.$$

Obtain solution for $1 \geq x \geq 0$. [15]

7. (a) Prove that for any two distinct vectors x and y in $\Re^n$, the Householder matrix $P = I - 2ww^t$, $w = (x-y)/\|x-y\|_2$ satisfies $Px = y$. [7]

(b) Tridiagonalize the following matrix by using the Householder's transform and indicate the necessary computational steps:
$$\begin{bmatrix} 3 & 1 & -2 & 1 \\ 1 & 2 & 0 & 2 \\ -2 & 0 & 4 & -2 \\ 1 & 2 & -2 & -1 \end{bmatrix}.$$
[13]

8. Find Doolittle's LU decomposition for the matrix A for which the diagonal elements of matrix A are all 1:

$$A = \begin{bmatrix} 4 & -2 & 0 & 0 \\ -2 & 2 & 2 & 0 \\ 0 & 2 & 8 & -6 \\ 0 & 0 & -6 & 10 \end{bmatrix}$$

Use the above LU factorization to find vector x such that:

$$Ax = \begin{bmatrix} 2 \\ 0 \\ -8 \\ 16 \end{bmatrix}$$
[20]

9. a) Explain whether Euler's method will give a stable solution for the following differential equation: $\dfrac{dy}{dx} + xy = 0$, $y(0) = 1$, from $x = 0$ to $x = 0.25$ with step-size $h = 0.05$. [5]

b) Find all eigenvalues of the following tridiagonal matrix and deduce the formula that you use: [15]

$$\begin{bmatrix} 10 & 10 & 0 & 0 & 0 & 0 & 0 & 0 \\ 4 & 10 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 4 & 10 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4 & 10 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 & 10 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 4 & 10 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 4 & 10 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 4 & 10 \end{bmatrix}$$

INDIAN STATISTICAL INSTITUTE
Semester Examination: 2011-12
B. STAT 1[st] Year
STATISTICAL METHODS II
Full Marks: 60

08.05.2012 Duration: 3 hrs

1. Suppose that we have a dependent variable Y and three independent variables X_1, X_2 and X_3 such that the correlation coefficient between Y and X_i (for all i=1,2,3) is 0.03 and the correlation coefficient between X_i and X_j is 0.21 for all $1 \le i \le j \le 3$. Compute the multiple correlation coefficient between Y and (X_1, X_2, X_3) and the partial correlation coefficient between Y and X_1 eliminating the effects of X_2 and X_3. (6+6=12)

2. Given three variable X_1, X_2, X_3, you are to find a linear regression equation for X_1. You need to decide whether to use (i) a simple regression with either X_2 or X_3 as a regressor, or (ii) a multiple regression. Which would you choose in the following situations? Justify your answer.

(i) $r_{23} = 1$ (ii) $r_{23} = 0$ (iii) $r_{23} = 0$, $r_{12} = r_{13}$ (iv) $r_{12} = 0.8$, $r_{13} = 0.4$, $r_{23} = 0.5$

(3+3+4+4=14)

3. Consider the bivariate density function

$$f(x,y) = \begin{cases} \frac{1}{2} & if \ \ |x|+|y| \le 1 \\ 0 & otherwise \end{cases}$$

Describe (with justification) how to generate (x,y) with p.d.f. *f(x,y)* using (u,v), where u and v are i.i.d. random variables having uniform distribution on [0,1]. (9)

4. Each of the variables X_1, X_2, and X_3 follows a normal distribution with $E(X_1)=E(X_2)=E(X_3)=1$, $V(X_1)=V(X_2)=2$, $V(X_3)=4$ and $\rho_{12}=0.2$, $\rho_{13}=0.8$, and $\rho_{23}=0$. You are given three independent uniform (on [0,1]) random observations 0.426, 0.238, and 0.927. From these three random numbers generate a set of values for X_1, X_2, and X_3 justifying your method. (9)

5. In an archery competition the target board has three concentric circles of radius $\frac{1}{\sqrt{3}}$, 1 and $\sqrt{3}$. An archer gets a score of 3 if the arrow hits the innermost circle, a score of 2 if it falls within the second and the first circles, a score of 1 if it falls within the second and the third circles and a score of 0 if it falls outside the outermost circle. Find the probability distribution of the score (X) given the probability density of the distance Y of hit from the centre of the target is

$$f(y) = \frac{2}{\pi}\frac{1}{1+y^2}, \qquad 0 < y < \infty$$

To move to the next round of the game, an archer has to hit the innermost circle three times out of six attempts. What is the probability that the archer will move to the next round? What is the expected number of hits the archer needs to hit the innermost circle three times? (7+4+5=16)

INDIAN STATISTICAL INSTITUTE

Semestral Examination Semester II : 2011-2012

B.Stat. (Hons.) I Year

Probability Theory II

Date : 02.05.12 Maximum Score : 120 pts Time : 3 Hours

Note : This paper carries questions worth a **total** of **130 POINTS**. Answer as much as you can. The **MAXIMUM** you can score is **120 POINTS**.

1. Let X be uniformly distributed on $(0,1)$. Find a function $G:(0,1)\to\mathbb{R}$ such that $Y=G(X)$ has a geometric distribution with parameter p $(0<p<1)$. You need to show that Y indeed has the required distribution. **[10]**

2. (a) Let G and H be two distribution functions on $\mathbb{R}$. Assuming usual notations, show that the function $F(z)=\int G(z+y)dH(y)$, $z\in\mathbb{R}$, is well-defined and that it defines a distribution function.

   (b) Show that if X and Y are independent random variables with distribution functions H and G respectively, then F in (a) is the distribution function of the random variable $Z=X-Y$.

   [Hint: Consider $Y_n=\frac{k}{2^n}$ if $\frac{k-1}{2^n}<Y\le\frac{k}{2^n}$.] (10+10)=**[20]**

3. A random point X selected from the interval $(0,1)$ according to a distribution F divides the interval into a shorter part U and a longer part V (assume $\Delta F(\frac{1}{2})=0$).

   (a) Find the distribution function of the random variable $Z=V/U$.

   (b) Show that if F has density f, then Z is absolutely continuous and find its density. (10+10)=**[20]**

4. Show that if X is any random variable, then $g(t)=E(\cos t^2X)$, $t\in\mathbb{R}$, defines a continuous function on $\mathbb{R}$. Show that if X has finite expectation, then g is continuously differentiable. (10+10)=**[20]**

5. (a) Show that if X is a random variable taking values in a bounded open interval (a,b), then $E(X)\in(a,b)$. Show that if X was assumed to take values in the closed interval $[a,b]$, the also $E(X)\in(a,b)$, unless $P(X=a)\vee P(X=b)=1$.

   (b) Let X be a random variable such that $E|X|^n\le C^n\ \forall\ n\ge 1$, for some real number $C>0$. Show that $P(|X|\le C)=1$. (10+10)=**[20]**

6. If X and Y are independent random variables with Gamma(α_1) and Gamma(α_2) distributions respectively, then show that the random variables $U=X+Y$ and $V=X/Y$ are independent and find the density of V. (10+10)=**[20]**

7. (a) Find the constant C such that the function $f(x,y)=Ce^{-(3x+y)}/(3x+y)$, if $x>0, y>0$ and $f(x,y)=0$ otherwise, is a bivariate density.

   (b) If (X,Y) has joint density f, find $\rho(X,Y)$. (10+10)=**[20]**

INDIAN STATISTICAL INSTITUTE
Compensatpry Back-paper Examination : 2011-12
B. Stat. - First Year
Analysis I

Date : 25. 06. 2012 Maximum Score : 100 Time :3 Hours

1. Answer all the questions.

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.

(1) For any set X, let $\mathcal{P}(X)$ denote the power set of X. Is there a function f from X onto $\mathcal{P}(X)$? Justify your answer [8]

(2) Show that the set of all rational numbers is dense in the space of all real numbers. [8]

(3) For a sequence $\{a_n\}$ of real numbers, define

$$\sigma_n = \frac{a_1 + \cdots + a_n}{n}, \quad n \geq 1.$$

Show that

$$\limsup \sigma_n \leq \limsup a_n.$$

[12]

(4) Let

$$\sum_0^\infty a_n x^n$$

be a power series not convergent for a real number t. Show that there is a real number $r \geq 0$ such that for every real number x with $|x| < r$, the above series is absolutelyly convergent and for every real number x with $|x| > r$, the above series is divergent. [12]

(5) Let $\{U_n\}$ be a sequence of dense open sets in $\mathbb{R}$. Show that $\cap_n U_n$ is dense in $\mathbb{R}$. [16]

(6) Let $f : \mathbb{R} \to \mathbb{R}$ be a monotone function such that for every interval $I \subset \mathbb{R}$, $f(I)$ is an interval. Show that f must be continuous. [10]

(7) Let f be continuous on $[0, 1]$, differentiable on $(0, 1)$ and there is a λ, $0 < \lambda < 1$ such that $|f'(x)| \leq \lambda \cdot |f(x)|$. Show that f is a constant. [10]

(8) Let $f : \mathbb{R} \to \mathbb{R}$ be a differentiable function and $\lim_{x\to 0} f'(x)$ exists. Show that this limit equals $f'(0)$. [12]

P.T.O.

(9) Let $f, g : \mathbb{R} \to \mathbb{R}$ be n-times continuously differentiable, the first $(n-1)$ derivatives at 0 of both f and g are 0 and the n-th derivative of g at 0 is non-zero. Show that

$$\lim_{x\to 0} \frac{f(x)}{g(x)} = \frac{f^{(n)}(0)}{g^{(n)}(0)}.$$

[12]

INDIAN STATISTICAL INSTITUTE
Backpaper Examination – Semester II : 2011-2012
B.Stat. (Hons.) I Year
Probability Theory II

Date : 27.6.12 | Total Marks : 100 | Time : 3 Hours

Note : This paper carries questions worth a **total** of **100 MARKS**. Answer as much as you can. The **MAXIMUM** you can score is **45**.

1. Let F be a probability distribution function on $\mathbb{R}$. Show that the function

$$G(x) = \sum_{k=4}^{7} \binom{7}{k} (F(x))^k (1 - F(x))^{7-k}$$

defines a probability distribution function on $\mathbb{R}$. **[10]**

2. Consider the function $f(x) = Cx^{\alpha-1}e^{-\lambda x^\alpha}$ for $x > 0$, and, $f(x) = 0$ for $x \leq 0$, where $\lambda > 0$, $\alpha > 0$, .

   (a) Find C such that f is a probability density function.

   (b) If X has density f, find the distribution of $X^{-\alpha}$.

   (c) Find a function $g : (0,1) \to \mathbb{R}$ such that, if Y is Uniform on $(0,1)$, then $X = g(Y)$ has density f. (8+8+8)=**[24]**

3. Let X be a random variable. Suppose for every integer $m \geq 1$, the random variable $X_m = [mX]$ has a geometric distribution with some parameter $0 < p_m < 1$. Show that X must be exponentially distributed with some parameter $\lambda > 0$. What is the relation between λ and the p_m-s? (7+3)=**[10]**

4. Prove that, $E[|X|^p] = p\int_0^\infty u^{p-1}P(|X| > u)\,du$, for any random variable X and any $p > 0$. [Hint: Do it first for the case when X is bounded, that is, $|X| \leq N$ for some positive integer N.] **[10]**

5. Show that if X is a positive random variable with finite first moment, then $E(\sqrt{X}) \leq \sqrt{E(X)}$ and $E(1/X) \geq 1/E(X)$. (6+6)=**[12]**

6. Show that if G is a univariate distribution function with density g, then for any $n \geq 2$, the function $f(x,y) = n(n-1)g(x)g(y)[G(y) - G(x)]^{n-2}$, $-\infty < x < y < \infty$ is a probability density function on $\mathbb{R}^2$. **[10]**

7. Let (X,Y) have joint density $f(x,y) = \frac{3}{4}[xy + (x^2/2)]$, $0 < x < 1, 0 < y < 2$.

   (a) Find the probabilities $P(X + Y < 1)$ and $P(XY \geq 1)$.

   (b) Find the marginal densities of X and Y

   (c) Find the covariance between X and Y.

   (d) Find the joint densities of $(X + Y, X - Y)$ and of (X, XY). (6+6+6+6)=**[24]**

INDIAN STATISTICAL INSTITUTE
Semester Examination: 2011-12 (Back Paper)
B. STAT 1st Year
STATISTICAL METHODS II
Full Marks: 100

28.06.12

Duration: 3 hrs

1. Given three variable X_1, X_2, X_3, you are to find a linear regression equation for x_1. You need to decide whether to use (i) a simple regression with either X_2 or X_3 as regressor, or (ii) a multiple regression. Which would you choose in the following situations? Just your answer.

(i) $r_{23} = 1$ (ii) $r_{23} = 0$ (iii) $r_{23} = 0$, $r_{12} = r_{13}$ (iv) $r_{12} = 0.8$, $r_{13} = 0.4$, $r_{23} = 0.5$

(4+4+6+6=20)

2. (a) From a data set with three variables, a student computed $r_{12} = 0.6$, $r_{13} = -0.4$, $r_{23} = 0.7$. Do you think these computations are internally consistent? Justify your answer.

(b) For three variables, let $|r_{ij}| = 1$ for all i,j = 1,2,3. How many of these r_{ij} values can be negative? Justify your answer.

(c) If $aX_1+bX_2+cX_3=0$ holds for all sets of values of X_1, X_2, and X_3, (a,b,c of same sign) what will be the values of the partial correlation coefficient $r_{12.3}$, $r_{23.1}$, and $r_{13.2}$? (9+7+9=25)

3. Consider the bivariate density function

$$f(x,y) = \begin{cases} \frac{1}{\pi} & if \;\; x^2 + y^2 \le 1 \\ 0 & otherwise. \end{cases}$$

Describe (with justification) how to generate (x,y) with p.d.f. *f(x,y)* using (u,v), where u and v are i.i.d. random variables having uniform distribution on [0,1]. (15)

4. For a binomial random variable X with m=2s and p=0.5, show that s is the most probable value of X and

$$P(X = s) = \frac{(2s-1)(2s-3) \times \ldots \times 3 \times 1}{2s(2s-2) \times \ldots \times 4 \times 2} \qquad (12)$$

5. Each of the variables X_1, X_2, and X_3 follows a normal distribution with

$E(X_1)=E(X_2)=E(X_3)=2.1$, $V(X_1)=V(X_2)=3.2$, $V(X_3)=1.6$ and $\rho_{12}=0.1$, $\rho_{13}=0.6$, and $\rho_{23}=0.3$.

You are given three independent uniform[0,1] random observations 0.537, 0.127, and 0.693.

From these three random numbers generate a set of values for X_1, X_2, and X_3 justifying your method. (15)

6. From a lot of N objects, of which Np are of a given kind, objects are drawn one by one without replacement till r objects of the given kind are obtained. Find the probability distribution of the number of drawings required (say X). Also calculate E(X) and V(X). (5+4+4=13)