INDIAN STATISTICAL INSTITUTE

MID-TERM EXAMINATION (2015–16)

B. STAT. I YEAR

ANALYSIS I

Date : 31.08.2015 Maximum Marks : 60 Time : 2 hours

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

1. In each of the following cases, decide whether the given set $A \subseteq \mathbb{R}$ is bounded above (and/or below) and if it is, find its supremum (and/or infimum).

   (a) $A = \{x \in \mathbb{R} : 3x^2 - 10x + 3 < 0\}$.

   (b) $A = \left\{\left|x + \frac{1}{x}\right| : x \neq 0\right\}$. [$2 + 3 = 5$]

2. Let $A \subseteq \mathbb{R}$ be bounded above. Show that there is a sequence $\{a_n\}$ taking values in A such that $\lim_{n\to\infty} a_n = \sup A$. [5]

3. Find the lim sup and lim inf of the following sequences : [$8 + 7 = 15$]

   (a) $a_n = (5^n + 7^n)^{1/n}$

   (b) $a_n = \frac{n}{7} - \left[\frac{n}{7}\right]$, where $[x]$ is the largest integer $\leq x$.

4. (a) For every $m \in \mathbb{N}$, show that $\frac{1}{\sqrt{m+1}} < 2(\sqrt{m+1} - \sqrt{m}) < \frac{1}{\sqrt{m}}$.

   (b) Deduce that for every $n \in \mathbb{N}$, $2(\sqrt{n} - 1) < \sum_{k=1}^{n} \frac{1}{\sqrt{k}} < 2\sqrt{n} - 1$.

   (c) If $a_n = 2\sqrt{n} - \sum_{k=1}^{n} \frac{1}{\sqrt{k}}$, show that $\{a_n\}$ converges. [$5 + 5 + 7 = 17$]

5. Let $0 < \alpha < 1$. Given $x_0, x_1 \in \mathbb{R}$, define $x_{n+1} = \alpha x_n + (1 - \alpha)x_{n-1}$ for all $n \geq 1$.

   Prove that $\{x_n\}$ is Cauchy and hence converges.

   Find $\lim_{n\to\infty} x_n$ in terms of x_0, x_1 and α. [$5 + 3 = 8$]

6. Find $\lim_{n\to\infty} \left(\frac{n^n}{n!}\right)^{1/n}$. [5]

7. Test the following series for convergence. [$5 + 5 = 10$]

   (a) $\sum_{n=1}^{\infty} n^{-1-1/n}$.

   (b) $\sum_{n=1}^{\infty} n^p(\sqrt{n+1} - \sqrt{n})$, $p \in \mathbb{R}$.

Indian Statistical Institute
Statistical Methods I
B. Stat. (hons.) 1st year
Midsemestral examination

Date: Sept 01, 2015 Duration: 2hrs.

Attempt all questions. The maximum you can score is 20. Justify all your steps. This is an open book, open notes examination. Notes may not be passed from one student to another. You may use your own calculator.

If copying is detected in the solution for any problem, all the students involved in the copying will get 0 for that problem. Also an additional penalty of 5 will be subtracted from the overall aggregate of each of these students.

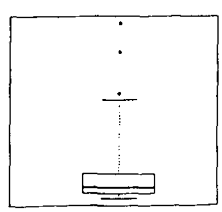
1. A new species of fish has been seen at certain points of the equatorial Pacific. The exact latitudes and longitudes of the locations have been recorded as $(x_1, y_1), ..., (x_n, y_n)$. All the points are close to each other and so it has been decided to aim a satellite at the centre of those points. Suggest how you may define a meaure of central tendency (x_*, y_*) suitable for this problem. Your measure should work in general for any such data set (not just in the equatorial Pacific region). Apply your formula to the following data.

lat	1.14	1.96	-1.25	0.67	-2.36	-2.24	-1.43	2.73
long	-179.52	177.04	177.10	-177.59	-180.00	-178.82	178.12	177.70

Here the latitudes and longitudes are given in degrees. A minus (or plus) sign means West (or East) for longitudes, and North (or South) for latitudes. [5+2]

2. Find the set of all points $a \in \mathbb{R}$ that minimises $\sum_{i=1}^{10} |x_i - a|$, where $x_1 < x_2 < \cdots < x_n$ are given numbers. Prove mathemtcally that your answer is correct. [10]

3. The following picture shows a boxplot.



Classify it as positively skewed, negatively skewed and symmetric. Justify your answer. Also roughly sketch a possible form of the (normalised) histogram based on the same data. [3]

4. A die is rolled 100 times and the average value of all the outcomes is computed. This entire exercise is repeated 1000 times, and a normalised histogram of all the 1000 averages is to be drawn. A student has written the following R program for this purpose. Find as many mistakes in it as possible.

```
a = numeric(100)
for(i in 1:1000) {
  x = sample(100,6)
  a = mean(x)
}
hist(a)
```

Bonus: Can you say how R is going to react to the above code? Will it give an error message? Will it give a wrong answer? Or blow up the computer? Or something else? [5]

Indian Statistical Institute

Vectors and Matrices I: B. Stat First Year: Mid-Semester Examination: 2015-16

Maximum Marks: 40 **Time: 2hrs** **Date: 03-09-2015**

1. Let V be a vector space of finite dimension. Show that every basis of V must have the same number of elements. [7]

2. Let V be a finite dimensional vector space over the field F. Let W be a subspace of V and f is a linear functional on W. Prove that there is a linear functional g on V such that $g(x) = f(x)$ for all $x \in W$. [7]

3. Prove or disprove: For the vector space $P_3(F)$ of all polynomials of degree less than or equal to three, with coefficients from F, there exists a basis without a polynomial of degree two. [5]

4. An arrangement is made of nk scalars from F in k rows and n columns such that each column and each row can be viewed as a vector in F^k and F^n respectively.

(a) Show that the maximal number of linearly independent columns is same as the maximal number of linearly independent rows. [8]

(b) If $k = n$ and F is a finite field with p elements, find the number of distinct arrangements such that all the columns are linearly independent. [9]

5. Let V be a n-dimensional vector space over the field F. Let W be a subspace of V with dimension $p < n$. Then show that W is the intersection of all $(n-1)$-dimensional subspaces of V which contains W. [9]

INDIAN STATISTICAL INSTITUTE
Mid Semestral Examination: 2015-16

Course Name: B. STAT. I YEAR

Subject Name: Introduction to Programming and Data Structure

Date: 04. 09. 2015

Maximum Marks: 60 Duration: 2 hours 30 Mins.

Answer as much as you can.

1. a) Let the function f be defined as follows:

```
int f (int n)
{
int t;
t = 100000 * (n % 10) + (n / 10);
return t / n;
}
```

What is the value returned by the call f(142857)? Briefly explain your answer. (5)

b) The following function expects a non-negative integer argument.

```
unsigned int doit ( unsigned int n )
{
unsigned int i, s, t;
s = 0; t = 1;
for (i=0; i<n; ++i) {
s = s + i + n;
t = t * 2;
}
return s * t;
}
```

What value does doit(4)return? Describe as a mathematical function of n the value returned by the function doit(n). (3+4 = 7)

c) What value will be stored in count after the execution of the following code and why?

```
int count, n = 100;
count = printf("\n\n:%d\n",n);
```

(3)

2. a) What will be the output of the following program? If you feel that there is any compile time or run time error, point out the same. Justify your answer. If your answer is yes, then can you correct this error by doing a minimal change in the program? (3+2 = 5)

P.T.O

```c
#include<stdio.h>
 main( )
   {
           int a[5]={1,2,3,4,5};
           for(i=0;i<5;i++) printf("%d\n", *a++);
   }
```

b) With example, discuss the differences between the *break* and *continue* statements in context to loop termination in C. (4)

c) Write an iterative C function void append (char *s, char *t) which appends the string t immediately after the end of the string s. For example, if s and t are passed respectively as "mumbo" and "humbo jumbo", then your function should change the string s to "mumbohumbo jumbo".

Do not use any string library functions. (6)

3. a) The mode of an array of numbers is the number *m* in that array that is repeated most frequently. If more than one number is repeated with equal maximum frequencies, there is no mode. Write a C function that accepts an array of numbers and returns the mode or an indication that the mode does not exist. (10)

b) Consider a *for* loop:

```c
for (i=1; i<100; ++i) i *= i+1;
```

i) How many times is the statement "i *= i+1" in the for loop executed?

ii) How many times is the loop condition i<100 checked in the loop?

iii) What is the value stored in the variable 'i' immediately after the loop terminates? (5)

4. a) Write a C program which reads a sequence of positive integers till the user types −1. It counts the lengths of the non-decreasing sub-sequences, and prints the maximum among them. For example, for input {6, 7, 2, 29, 17, 5, 5, 11, 6, 7, 8, −1} the non-decreasing sub-sequences are: {6, 7}, {2, 29}, {17}, {5, 5, 11} and {6, 7, 8}. Thus the answer should be 3. Assume that the first integer read is not −1 and a single integer is a sequence of length 1. (10)

b) What is printed by the following program? Briefly explain. (5)

```c
main()
{
  int x = 1, y = 0, z = 1, t;
  for (t = 0; t < 10; ++t)
    { y += (x) ? z : -z; z++;
      x = !x;
    }
  printf("y = %d", y);
}
```

5. a) What will be printed by the following code segment? (3)

```c
typedef struct {
                char *name;
```

P.T.O

```
        int age, ht;
    } human;

void hfn ( human A )
{
   A.name[0] += 'A' - 'a';
   A.age += 3; A.ht += 5;
}

void main ()
{
  char A[20] = "bob";
  human H = {A, 20, 180};
  hfn(H);
  printf("%s,%d,%d", H.name, H.age, H.ht);
}
```

b) For a real number x, the notation $\lfloor x \rfloor$ stands for the largest integer less than or equal to x. For example, $\lfloor \pi \rfloor = 3$ and $\lfloor 3 \rfloor = 3$. Write a program that reads a positive integer n and an integral base $b > 2$. The program computes and prints the value of $\lfloor \log_b n \rfloor$. For example, $\log_{23} 456789 = 4.1562752022$ and hence $\lfloor \log_{23} 456789 \rfloor = 4$. Therefore, upon input $n = 456789$ and $b = 23$, your program should print 4. You are not allowed to use any math library call (like log, log10 or floor). Do not make any floating point calculations. Do not write any function (other than main). You may, however, declare and use some additional *int* variables (but no arrays). (8)

c) With a suitable self made example, distinguish between static variables and automatic variables. (4)

6. a) The Fibonacci sequence is defined as:

$F(0) = 0$,

$F(1) = 1$,

$F(n) = F(n-2) + F(n-1)$ for $n > 2$.

The following function defines another sequence $G(n)$ for $n = 0,1,2,3,....$ How is the sequence $G(n)$ mathematically related to the Fibonacci sequence $F(n)$? (Express $G(n)$ in terms of $F(n)$. Your expression should hold for all integers $n > 0$.)

```
int G ( unsigned int n )
{
   if (n == 0) return 0;
   if (n == 1) return 1;
   return G(n-2) - G(n-1);
}
```

(4)

b) A positive integer is called square-free if it is not divisible by the square of any prime number. For example, $98 = 2\times7^2$, $99 = 3^2\times11$, $100 = 2^2\times5^2$ are not square-free, whereas 101 (a prime) and $102 = 2\times3\times17$ are square-free. Your task is to find the divisor m of a positive integer n supplied by the user, such that m is square-free and as large as possible. Indeed, m is the product of all the distinct prime factors of n, each taken only once. For example, for n = 98, 99,100,101,102, the values of m will be $14 = 2\times7$, $33 = 3\times11$, $10 = 2\times5$, 101, $102 = 2\times3\times17$, respectively. Write a program to solve this problem. (7)

c) Write a short C code segment that opens a data file named "my_file.txt" in write mode and stores the first 10 decimal digits in the file, each digit in a new line. (4)

INDIAN STATISTICAL INSTITUTE

Mid-Semestral Examination: 2015-16 (First Semester)
Bachelor of Statistics (B. Stat.) I Year
Probability Theory I

Teacher: Parthanil Roy

Date: 05/09/2015 **Maximum Marks: 30** **Duration: 2:30 - 5:00 pm**

Note:

- **Please write your roll number on top of your answer paper.**
- **There are three problems each carrying 10 marks with a total of 30 points. Solve as many as you can. Show all your works and write explanations when needed. Maximum you can score is 30 marks.**
- **You are NOT allowed to use class notes, books, homework solutions, list of theorems, formulas etc.**

1. Suppose there are two urns. The first urn contains 2 red balls and 4 green balls and the second one contains 3 red balls and 1 green ball. A fair coin is tossed. If head appears, then a ball is chosen at random from the first urn and if tail appears, then a ball is chosen at random from the second urn. The selected ball is returned to the urn it came from and two more balls of the same colour are added to that urn. If at the end of this urn scheme, it is observed that one of the urns has equal number of green and red balls, then what is the probability that the outcome of the toss was a tail? [10]

2. Ravi has decided to sit for the driving test, which costs (in Rupees) $i^3 - (i-1)^3$ for the i^{th} trial. If the probability that Ravi passes the test in one trial is $p \in (0,1)$ and all the trials are independent, then what is the expected total cost (in Rupees) for Ravi to pass the test? [10]

3. A box contains 3 different types of coupons. Suppose that the proportion of the coupons of Type i is $p_i \in (0,1)$, where $p_1 + p_2 + p_3 = 1$. Simple random sampling with replacement is performed until coupons of all types are obtained. Let N be the number of draws needed to achieve this.
   (a) For all positive integer n, compute $P(N > n)$. [7]
   (b) Using (a) or otherwise, calculate the probability mass function of N. [3]

Wish you all the best

INDIAN STATISTICAL INSTITUTE

First Semester Examinations: 2015-16

B. Stat. I Year Analysis-I

Date: 16/11/2015 Maximum Marks: 60 Duration: 3 Hours

Note: Give proper justification to all your answers. State clearly all the results you are using.

(1) (a) Find the supremum and infimum of the set $A = \left\{\frac{mn}{1+m+n} : m, n \in \mathbb{N}\right\}$.

(b) Let a, b and p be real numbers and $\{a_n\}_{n=1}^{\infty}$ be a sequence defined as follows:

$$a_1 = a, \quad a_2 = b, \quad a_{n+1} = pa_{n-1} + (1-p)a_n, \quad n = 2, 3, 4, \ldots.$$

Characterize all values of a, b and p for which the sequence $\{a_n\}_{n=1}^{\infty}$ converges. [5+5]

(2) (a) Let $f : \mathbb{R} \to \mathbb{R}$ be a function which satisfies

$$\lim_{h \to 0} [f(x+h) - f(x-h)] = 0$$

for every $x \in \mathbb{R}$. Does this imply that f is continuous?

(b) Let $f : \mathbb{R} \to \mathbb{R}$ be such that $|f(x) - f(y)| \leq |x-y|^2$ for all $x, y \in \mathbb{R}$. Show that f is a constant function.

(c) Let $r > 0$. Show that $\log x < \frac{x^r - 1}{r}$ for all $x > 1$. [3+3+3]

(3) (a) Construct a continuous function from the open interval $(0, 1)$ onto the closed interval $[0, 1]$.

(b) Let $g : (0, 1) \to [0, 1]$ be continuous and onto. Prove that it cannot be bijective.

[Hint. Consider $x_0 = g^{-1}(0)$ and $x_1 = g^{-1}(1)$. Apply intermediate value theorem.] [5+5]

(4) Suppose that $f : \mathbb{R} \to \mathbb{R}$ is continuous at 0 and satisfies

$$f(0) = 0 \quad \text{and} \quad f(s+t) \leq f(s) + f(t) \quad \text{for all } s, t \in \mathbb{R}.$$

Prove that f is uniformly continuous on $\mathbb{R}$. [8]

(5) Suppose that f is a twice-differentiable function on $(0, \infty)$. Let

$$M_0 = \sup_{x \in (0,\infty)} |f(x)|, \quad M_1 = \sup_{x \in (0,\infty)} |f'(x)| \quad \text{and} \quad M_2 = \sup_{x \in (0,\infty)} |f''(x)|.$$

Show that $M_1^2 \leq 4M_0M_2$.

[Hint. If $h > 0$, then by Taylor's theorem $f(x+2h) = f(x) + 2hf'(x) + (2h)^2\frac{f''(c)}{2}$ for some $c \in (x, x+2h)$.] [10]

(6) Let $f : \mathbb{R} \to [-1, 1]$ be such that f'' exists and is continuous on $\mathbb{R}$. Suppose that

$$[f(0)]^2 + [f'(0)]^2 = 4.$$

Prove that there exists $c \in \mathbb{R}$ such that $f(c) + f''(c) = 0$.

[Hint. Use Mean Value Theorem to find $a \in (-2, 0)$ and $b \in (0, 2)$ such that $|f'(a)| \leq 1$ and $|f'(b)| \leq 1$. Consider $F(x) = [f(x)]^2 + [f'(x)]^2$ and let $F(c) = \max\{F(x) : x \in [a, b]\}$. Show that $c \in (a, b)$ and $F'(c) = 0$. Prove that $f'(c) \neq 0$.] [10]

(7) Let $f : \mathbb{R} \to \mathbb{R}$ be the function

$$f(x) = \frac{1}{1+|x|} + \frac{1}{1+|x-1|}.$$

Find all local maxima and local minima of f. [10]

Indian Statistical Institute
Statistical Methods I
B-I, First Semestral Examination

Date: Nov 18, 2015 Duration: 2 hrs. 30 min.

This paper carries 55 marks. Attempt all questions. The maximum that you can score is 50. Justify all your steps. This is an open book, open notes examination. You may use your own calculator.

If copying is detected in the solution for any problem, all the students involved in the copying will get 0 for that problem. Also an additional penalty of 10 will be subtracted from the overall aggregate of each of these students.

1. A famous reality show used to broadcast the following puzzle on television:

   > The spectator is shown three closed doors. Behind one of the doors (selected randomly with equal probabilities) is a car. There is nothing behind the other two doors. The spectator knows that exactly one door hides a car, but is not told which door it is. If he can guess correctly, he wins. He picks a door at random with equal probabilities. The host (who knows the content behind each door) deliberately opens some *other* door which is *empty*. It is known to the spectator that the host will always open a door such that it is not the chosen door and not the door with the car. Then the spectator is given an option to change his choice of doors (*i.e.*, choose the other door that is still closed). The puzzle is whether the spectator should indeed change the door in order to improve his chance of winning.

   You are required to write an R program to simulate this set up such that we can run this program a large number of times to see how often we win with either strategy ("choose" and "don't choose"). Remember that you are *not* to solve the puzzle or even to simplify it by any probability computation. Just simulate it as it is, so that the output will indicate the solution. [15]

2. Suppose we have a data set $x_1, ..., x_n$ for some $n \geq 2$. Prove or disprove the following statement: the sample variance of this data set must be $\geq$ the sample variance of the absolute data set $|x_1|, ..., |x_n|$. [5]

3. We have three dice, A, B and C with the following probabilities:

Die	1	2	3	4	5	6
A	1/6	1/6	1/6	1/6	1/6	1/6
B	1/10	1/5	1/5	1/5	1/5	1/10
C	1/5	1/10	1/5	1/5	1/10	1/5

One of these dice is taken and this one die is rolled 50 times independently to produce the data

Number	1	2	3	4	5	6
Frequency	6	12	7	8	13	4

You are to guess which die was used. Cast this problem as an estimation problem clealry pointing out the parameter(s) and the parameter space. Then apply maximum likelihood estimation to find the answer. [3+7]

4. You are given 10 points $(x_1, y_1), ..., (x_5, y_5)$ and $(u_1, v_1), ..., (u_5, v_5)$. You are to fit two straight lines, one for the (x_i, y_i)'s, and the other for the (u_i, v_i)'s. The two lines may have different intercepts, but must have the same slope. Obtain the least squares solution to this problem. You may use the matrix form of normal equations without proof. [10]

5. Suppose that $x_i = i$ for $i = 1, ..., 10$.. Let $y_i = x_i/10$ and $z_i = 10x_i$. Let the correlation between the x's and the y's be r. Also let the correlation between the x's and z's be s. Then what would you expect: $r < s$ or $r = s$ or $r > s$? Now suppose that ϵ_i's are some iid realisations from $N(0, 0.5^2)$ distribution. Let $y_i' = y_i + \epsilon_i$ and $z_i' = z_i + \epsilon_i$. Let r' and s' be defined accordingly. Now what would you expect: $r' < s'$ or $r' = s'$ or $r' > s'$? Justify your answers. [3+7]

6. Answer any *one* of the following questions regarding the class projects. Each question gives you an idea whose feasibility you have to discuss. Feel free to say that an idea is infeasible, if that is what you think. Justify your answer intuitively. [5]

   (a) Consider the "chroma key" project where you have to classify each pixel on an RGB image as either foreground or background. Can you use the clasification and regression tree (C&RT) algorithm for this purpose?

   (b) In the project of "separating sound from silence" different students have prosposed different algorithms. Can you use cross-validation to compare them?

   (c) In the audio segmentation project the level of synchronisation between two audio tracks was measured by the correlation between them. Can you use the same idea to check the quality of the generated map in the map making project?

INDIAN STATISTICAL INSTITUTE

Semestral Examination: 2015-16 (First Semester)
Bachelor of Statistics (B. Stat.) I Year
Probability Theory I

Date: 23/11/2015 **Maximum Marks: 70** **Duration: 3 hours**

- **Please write your roll number on top of your answer paper.**
- **Justify ALL your steps. You can use any result proved in the class but you should clearly quote the result that you are using.**
- **You are NOT allowed to use class notes, books, homework solutions, list of theorems, formulas etc.**

1. Suppose r flags of different colours are displayed randomly on n poles arranged in a row. Assuming that there is no limitation on the number of flags on each pole, compute the probability that exactly $m(< n)$ poles are empty. [12]

2. Suppose n letters have to be sent to n different addresses and there are n envelopes with these addresses written on them. A person puts these letters into the envelopes (one letter in each) at random. Let X be the number of letters that are put into the correct envelope. Compute $\mathrm{E}(X)$ and $\mathrm{Var}(X)$. [6 + 8 = 14]

3. Let $\{S_n\}_{n\geq 0}$ be a simple symmetric random walk starting from 0 and
$$Q := \prod_{i=101}^{200} (S_i + 101).$$
Calculate the following probabilities:
(a) $\mathrm{P}\big(S_{101} \leq -100, S_{102} \leq -101, S_{103} \leq -101, \ldots, S_{200} \leq -101\big)$. [8]
(b) $\mathrm{P}\big(S_{101} \leq -100, S_{102} \leq -101, S_{103} \leq -101, \ldots, S_{200} \leq -101 \text{ and } Q = 0\big)$ [8]

4. Suppose that the number of customers visiting a restaurant on a particular day is $N \sim \mathrm{Poi}(\lambda)$. Assume that each customer orders fresh lime soda with probability p, independently of other customers, and independently of the value of N. Let L be the number of customers who orders fresh lime soda on that day. Find the probability mass function of L. [12]

5. A fair die is thrown two times independently. Let M denote the maximum of the two outcomes and S denote their sum. Calculate $\mathrm{E}(S|M)$ and $\mathrm{Var}(S|M)$. [6 + 10 = 16]

Indian Statistical Institute

First Semestral Examination (2015-2016)

B. STAT I: Vectors and Matrices –I

Date: 26.11.2015 **Maximum Marks: 60** **Time: 3 Hrs.**

This paper contains questions for 73 marks. The maximum you can score is 60.

1. Let T be a linear transformation from V to V. Prove that there exists a nonzero linear transformation S from V to V such that $TS=0$ if and only if there exists a nonzero vector $v \in V$ such that $T(v)=0$. [6]
2. Let V be the space of all polynomials with real coefficients up to degree n. Show that the derivative transformation is linear. Find its rank and nullity. [6]
3. Let T be a linear operator defined on a finite dimensional vector space V. If W is a T invariant subspace with $V = R(T) \oplus W$, then show that W must be $N(T)$. [7]
4. Let T be a linear transformation from V to W. Consider the restriction $T_{N(T)^C}$ of T on some complement of the null space of T. Show that $T_{N(T)^C} : N(T)^C \to R(T)$ is one-to-one and onto, where $R(T)$ is the range of T. [7]
5. Suppose that P and Q are $n \times n$ matrices such that $P^2 = P, Q^2 = Q$, and $I-(P+Q)$ is invertible. Show that P and Q have the same rank. [6]
6. Suppose that A and B are matrices of orders $m \times n$ and $s \times m$ respectively. Show that $R(A) = N(B)$ if and only if $R(B') = N(A')$, where $R(.), N(.)$ and $'$ denote range, null space and transpose respectively. [8]
7. Let A and B be two matrices of the same order. Prove that $\rho(A+B) = \rho(A) + \rho(B)$ if and only if $\rho(A\,|\,B) = \rho\begin{bmatrix} A \\ \hline B \end{bmatrix} = \rho(A) + \rho(B)$, where $\rho(.)$ denotes rank and $|$, $-$ denote column and row augmentation respectively. [8]
8. Show that if a non-singular matrix A has a standard LU factorization, then it is unique. [5]
9. Why do we do partial pivoting? Find $PA = LU$ factorization of the matrix A using partial pivoting, where, $A = \begin{bmatrix} 2 & 4 & 2 \\ 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix}$ [10]
10. Consider the system of linear equation $Ax = b$ in the form $(A|b)$ as
$$\begin{bmatrix} 1 & -4 & -1 & 1 & 3 \\ 2 & -8 & 1 & -4 & 9 \\ -1 & 4 & -2 & 5 & -6 \end{bmatrix}.$$
Using row reduced echelon form, find the solution. What other properties of the system can you tell from the row reduced echelon form? [10]

First Semester Examination (2015-2016)
B.Stat-1yr
Remedial English
100 Marks 1 1/2 hours

Date: 27.11.2015

1. Write an essay *on any one* of the following topics. Five paragraphs are expected.

a. Cameras

b. Different Modes of Transport

c. My Favourite Author

(60 marks)

2. Fill in the blanks with appropriate prepositions :

a. I will meet you ____ Monday.

b. I was born ___ 1986.

c. I will complete this assignment ____ evening.

d. The teacher pointed ___ the blackboard.

e. The school will remain closed ___ Monday.

f. The cat climbed ___ ___ the cupboard.

g. He sat ____ me.

h. Trees were planted ___ the road.

i. The geese flew ___ the houses.

(20 marks)

3. Fill in the blanks with appropriate words :

One day, Ali Baba was ____ bundles of firewood ___ ___ donkeys. He ___ then go ____ ___ market ____ sell ___ _____. Suddenly, ___ ____ cloud _____ dust ____ high ____ ____ air. He ___ ____ eyes to ___ ____ closer _____.

(20 marks)

Indian Statistical Institute
Statistical Methods I
B-I, First Backpaper Examination

Date: 25/01/2016 Duration: 3 hrs.

This paper carries 100 marks. Attempt all questions. The maximum that you can score is 45. Justify all your steps. This is an open book, open notes examination. You may use your own calculator.

1. Prove or disprove each of the following statements:

   (a) If $x_1, ..., x_n$ are some positive numbers and $h(\cdot)$ is a strictly increasing function, then the variance of $h(x_1), ..., h(x_n)$ must be $\geq$ the variance of $x_1, ..., x_n$. [6]

   (b) If $\{(x_i, y_i)\}_{i=1}^{5}$ and $\{(x_i, y_i)\}_{i=6}^{10}$ are two bivariate data sets each with correlation coefficient ≥ 0 then the correlation coefficient for the combined data set $\{(x_i, y_i)\}_{i=1}^{10}$ must also be ≥ 0. [7]

   (c) If $x_1, ..., x_{10}$ constitute a data set with variance 1, then it is possible to have numbers $x_{11}, ..., x_{10+n}$ for some $n \in \mathbb{N}$ such that the variance of the data set $x_1, ..., x_{10+n}$ is < 1. [7]

   (d) If $(x_1, y_1), ..., (x_5, y_5)$ is a bivariate data set with distinct x_i's, with least squares line $\hat{y} = 2.8 - 5.2x$. Then the least squares line passing through origin for the same data must be $\hat{y} = -5.2x$. [5]

2. A famous reality show used to broadcast the following puzzle on television:

   > The spectator is shown three closed doors. Behind one of the doors (selected randomly with equal probabilities) is a car. There is nothing behind the other two doors. The spectator knows that exactly one door hides a car, but is not told which door it is. If he can guess correctly, he wins. He picks a door at random with equal probabilities. The host (who knows the content behind each door) deliberately opens some other door which is empty. It is **known** to the spectator that the host will always open a door such that it is **not the chosen door** and **not the door with the car**. Then the spectator is given an option to switch his choice of doors. The puzzle is whther the spectator should switch in order to improve his chance of winning.

   You are required to write an R program to simulate this set up such that we can run this program a large number of times to see how often we win with either strategy. Remember that you are *not* to solve the puzzle or even to simplify it by any probability computation. Just simulate it as it is, so that the output will indicate the solution. [15]

3. Give a statistical set up for discrete data where MLE of a parameter is not unique. Justify your answer. [10]

4. 10 male and 9 female patients are used to test the effectiveness of a blood pressure pill to lower blood pressure. For each patient the initial blood pressures (both systolic and diastolic) are recorded before administering the pill. Then the blood pressures are measured again 4 times at intervals of 1 week as the medication continues. We also have 5 more males and 6 more female patients who are given a false pill (that looks like the actual pill, but has no active chemical in it). Their blood pressures are also measured like the others. Suggest how you should store this data in a suitable rectangular format for reading into R. Clearly mention the number of cases and the variable names you use. Also give an R command to read the data set into a variable called bp. [5+2+3]

5. You are given 10 points $(x_1, y_1), ..., (x_5, y_5)$ and $(u_1, v_1), ..., (u_5, v_5)$. You are to fit two straight lines, one for the (x_i, y_i)'s, and the other for the (u_i, v_i)'s. The two lines may have different slopes, but must have the same intercept. Obtain the least squares solution to this problem. You may use the matrix form of normal equations without proof. [10]

6. Suppose that $x_i = i$ for $i = 1, ..., 10$.. Let $y_i = x_i/10$ and $z_i = 10x_i$. Let the correlation between the x's and the y's be r. Also let the correlation between the x's and z's be s. Then what would you expect: $r < s$ or $r = s$ or $r > s$? Now suppose that ϵ_i's are some iid realisations from $N(0, 0.5^2)$ distribution. Let $y_i' = y_i + \epsilon_i$ and $z_i' = z_i + \epsilon_i$. Let r' and s' be defined accordingly. Now what would you expect: $r' < s'$ or $r' = s'$ or $r' > s'$? [3+7]

7. $X_1, ..., X_{10}$ are iid with $Unif[\theta, 0]$ distribution, where $\theta \in (-\infty, 0)$ is the unknown parameter. Find the MLE $\hat{\theta}$ of θ based on this data set. If $\hat{\theta}$ unbiased for θ? Justify your answer. [5+5]

8. Let $(x_1, y_1), ..., (x_n, y_n)$ be a bivariate data set to which we fit a straight line of the form $y = a + bx$ using least squares. Assume that not all x_i's are the same, and not all y_i's are the same. Let $\hat{y}_1, ..., \hat{y}_n$ be the fitted values. Assuming that not all $\hat{y}_i$'s are the same, express the correlation coefficient between y_i's and $\hat{y}_i$'s as a function of the correlation coefficient between x_i's and y_i's. [10]

Indian Statistical Institute

First Semestral Back Paper Examination (2015-2016)

B. STAT I: Vectors and Matrices –I

Date: ~~01/02/2016~~ 27/01/2016 **Maximum Marks: 100** **Time: 3 Hrs.**

1. Let $V = \{(x_1, x_2, x_3, x_4, x_5) \in R^5 : x_1 - 2x_2 + 3x_3 - x_4 + 2x_5 = 0\}$. Show that $S = \{(0,1,1,1,0), (1,0,0,1,0)\}$ is a linearly independent subset of V. Extend S to a basis for V. 2+7

2. $\mathcal{T}$ is a transformation from V to W. Show that for existence of a right inverse of $\mathcal{T}$, it is necessary that $\mathcal{T}$ is surjective. If $\mathcal{T}$ is surjective, show that any right inverse of $\mathcal{T}$ is injective. 3+3

3. Let $\mathcal{T}$ be a linear map from V to W. Show that there always exists a linear map S from W to V, such that $\mathcal{T}S\mathcal{T} = \mathcal{T}$. 5

4. Let $S = \{u_1, u_2, \ldots, u_n\}$ be a finite set of vectors. Prove that S is linearly dependent if and only if $u_1 = \theta$, the null vector, or $u_{k+1} \in span(\{u_1, u_2, \ldots, u_k\})$ for some $k (1 \le k < n)$. 10

5. Let $\mathcal{T}$ be a linear map from V to V. Show that V is the direct sum of $R(\mathcal{T}^k)$, the Range Space and $N(\mathcal{T}^k)$, the Null Space, for some positive integer k. 10

6. Show that a subset S of the vector space $\mathcal{F}^n$ is a subspace of $\mathcal{F}^n$ if and only if S is the null space of a matrix, where , $\mathcal{F}$ is the field of real numbers or complex numbers. 6

7. Suppose that the augmented matrix of a system $Ax = b$ is transformed into a matrix $(A'|b')$ in reduced row echelon form by a finite sequence of elementary row operations. Show that $Ax = b$ is consistent if and only if $(A'|b')$ contains no row in which the only nonzero entry lies in the last column. 6

8. Obtain the rank factorization of the matrix
$$\begin{bmatrix} -1 & 2 & 4 \\ 2 & -1 & 2 \\ 0 & 3 & 10 \end{bmatrix}.$$
10

9. Find a non-singular matrix P such that P^tAP is a diagonal matrix where
$$A = \begin{bmatrix} 2 & 2 & 3 \\ 2 & 5 & 6 \\ 3 & 6 & 10 \end{bmatrix}$$
12

[P.T.O]

10. Let V and W be finite dimensional vector spaces over a field F with $\dim V = n$ and $\dim W = m$. Let $\mathcal{T}$ be a linear map from V to W. Relative to two different pairs of ordered bases of V and W, let A and C be two matrix representations of $\mathcal{T}$. Show that there exists non-singular matrices P and Q such that $C = P^{-1}AQ$. 12

11. Let $\mathcal{R}$ be the set of all real numbers. Consider $\mathcal{R}$ as a vector space over the field Q of rational numbers. Show that $\sqrt{2}$ and $\sqrt{3}$ are linearly independent. 6

12. If f_1 and f_2 are two linear functionals on a vector space $\mathcal{V}$ over a field F satisfying $f_1(x) = 0$ whenever $f_2(x) = 0$ for $x \in \mathcal{V}$, show that $f_1 = \alpha f_2$ for some α in F. 8

INDIAN STATISTICAL INSTITUTE

Backpaper Examination: 2015-16 (First Semester)
Bachelor of Statistics (B. Stat.) I Year
Probability Theory I

Date: **Total Marks: 100** **Duration: 3 hours**

- **Please write your roll number on top of your answer paper.**
- **Justify ALL your steps. You can use any result proved in the class but you should clearly quote the result that you are using.**
- **You are NOT allowed to use class notes, books, homework solutions, list of theorems, formulas etc.**

1. A closet contains 25 different pairs of gloves. These fifty gloves are randomly arranged into 25 pairs.
   (a) Find the probability that each pair contains a left glove and a right glove. [6]
   (b) Find the probability that all the gloves are paired correctly. [6]

2. Roads A and B are the only escape routes from a state prison. Prison records show that, of the prisoners who tried to escape, 40% used road A, and 60% used road B. These records also show that 80% of those who tried to escape via A, and 70% of those who tried to escape via B were captured. Suppose that two prisoners have independently and successfully escaped from the prison. What is the probability that they have used the same road to escape? [8]

3. Suppose $X_1, X_2, \ldots, X_{100}$ are i.i.d. nonzero random variables defined on the same probability space. Compute
$$\mathrm{E}\left(\frac{\sum_{i=1}^{35} X_i^2}{\sum_{i=1}^{100} X_i^2}\right).$$
You need to justify every step of your answer. [6]

4. Suppose that the initial number M of bacteria in a bacteria colony follows binomial distribution with parameters $n = 100$ and $p = 0.5$. Assume that the bacteria behave independently of each other, they do not replicate and each of them die within an hour with probability 0.8 independently of M. Let Z denote the number of surviving bacteria in the colony after one hour. Find the probability mass function of Z. [12]

Please turn over

5. If $X \sim Bin(m,p)$ and $Y \sim Bin(n,p)$ are independent and $S := X + Y$, then compute the conditional distribution of X given S, $\mathrm{E}(X|S)$ and $\mathrm{Var}(X|S)$. [10 + 2 + 3 = 15]

6. A deck of 52 cards is shuffled in the following fashion: the topmost card is removed and put back into the deck in one of the 52 possible positions chosen uniformly at random. Let N be the number of such shuffles needed for the bottommost card to come to the top. Compute $\mathrm{E}(N)$ and $\mathrm{Var}(N)$. [7 + 7]

7. Suppose m and r are two positive integers. For the simple symmetric random walk $\{S_n\}_{n\geq 0}$ starting from 0, compute the following probabilities:

   (a) $P(S_1 > -m, S_2 > -m, \ldots, S_{n-1} > -m, S_n = r)$. [8]

   (b) $P(S_1 < m, S_2 < m, \ldots, S_{n-1} < m, S_n = r)$. [8]

8. Suppose 100 fair dice are rolled together and a one rupee coin is given to you. All the dice that produce 6 are removed, the rest of the dice are rolled together and another one rupee coin is given to you. Again, all the dice that produce 6 are removed, the rest of the dice are rolled together and another one rupee coin is given to you. This process goes on till all the dice get removed. Suppose Rs. R is your gain in this process. Calculate the probability mass function of R and $\mathrm{E}(R)$. [12 + 5 = 17]

INDIAN STATISTICAL INSTITUTE

First Semester Examinations: 2015-16 (Backpaper)

B. Stat. I Year **Analysis-I**

Date: 30/01/2016 Maximum Marks: 45 Duration: 3 Hours

Note: Give proper justification to all your answers. State clearly all the results you are using.

---

(1) Find the least upper bounds and greatest lower bounds of the following sets.

(a) $A = \{\frac{m}{n} : m, n \in \mathbb{N}, m < 2n\}$.

(b) $B = \{2^x + 2^{1/x} : x > 0\}$. [12]

(2) Suppose that the sequence $\{a_n\}$ satisfies the condition

$$0 < a_n < 1, \quad a_n(1 - a_{n+1}) > 1/4 \quad \text{for } n \in \mathbb{N}.$$

Show that the sequence $\{a_n\}$ converges and find its limit. [12]

(3) (a) Discuss the convergence of the series $\sum_{n=1}^{\infty} (\frac{n}{n+1})^{n(n+1)}$.

(b) Let $a \in \mathbb{R}$. Determine whether the series $\sum_{n=1}^{\infty} (-1)^n \sin(\frac{a}{n})$ is absolutely convergent, conditionally convergent, or divergent. [12]

(4) Let $r > 0$ and $f : (-r, r) \setminus \{0\} \to \mathbb{R}$ be such that

$$\lim_{x \to 0} \left(f(x) + \frac{1}{|f(x)|} \right) = 0.$$

Show that $\lim_{x \to 0} f(x)$ exists and is equal to -1.

[Hint. Show that there exists $s > 0$ such that $f(x) < 0$ for $x \in (-s, s) \setminus \{0\}$. For such an x, $|f(x) + 1| \leq \left| f(x) + \frac{1}{|f(x)|} \right|$.] [10]

(5) Let $a < b < c$. Show that if f is uniformly continuous on $(a, b]$ and on $[b, c)$, then it is uniformly continuous on (a, c). [10]

(6) Let $f : [a, b] \to \mathbb{R}$ be such that f' is continuous on $[a, b]$. Prove that for every $\epsilon > 0$, there exists a $\delta > 0$ such that

$$\left| \frac{f(t) - f(x)}{t - x} - f'(x) \right| < \epsilon \quad \text{whenever } x, t \in [a, b] \text{ with } 0 < |t - x| < \delta.$$

[10]

(7) Let $g : \mathbb{R} \to \mathbb{R}$ be such that $|g'(x)| \leq M$ for all $x \in \mathbb{R}$. Fix $\epsilon > 0$ and define $f(x) = x + \epsilon g(x)$. Prove that f is one-to-one if ϵ is small enough. [12]

(8) Let f be continuous on $[0, 2]$ and twice-differentiable on $(0, 2)$. Suppose that $f(0) = 0$, $f(1) = 1$ and $f(2) = 2$. Show that there exists $c \in (0, 2)$ such that $f''(c) = 0$. [10]

(9) Consider the function $f(x) = x^4 - 4x^3 + 10$.

(a) Find the local extrema of f.

(b) Find the intervals where f is increasing/decreasing.

(c) Find the intervals where f is convex/concave.

(d) Sketch a graph of the function. [12]

INDIAN STATISTICAL INSTITUTE
End Semestral (Back Paper) Examination: 2015-16

Course Name: B. STAT. I YEAR

Subject Name: Introduction to Programming and Data Structure

Date: 02/02/2016

Maximum Marks: 100

Duration: 3 hours

Answer as much as you can.

1 a) Consider the following way to compute the maximum in an array A of size n. First, divide the array in two equal (or almost equal) parts. Then, recursively compute the maximums in the two parts. Finally, return the larger of these two recursively computed maximum values.

(i) Write a function which uses the above idea for finding the maximum in an array.
(ii) Deduce the running time of the function on an array of size n. For simplicity, you may assume that n is a power of 2, that is, $n = 2^t$ for some integer $t > 0$. Express the running time in the Big-Oh notation (6+7=13)

b) Given an integer in decimal write a program to print the reverse of the number. For example, the reverse of 3481 is 1843. Do not use any array/string to store the number (7)

2 a) Distinguish between static and automatic variables using simple self-made examples. (5)

b) The absolute distance between two integers x_1 and x_2 is given by $| x_2 - x_1 |$. Write a function which sorts an array $x[\ ]$ of n integers in ascending order of their absolute distances with a given number z. For example, given $x[\ ] = \{9, 1, 12, 4, 2\}$ and $z = 6$, the sorted array will be $x[\ ] = \{4, 9, 2, 1, 12\}$. Note that 4 is closest to 6, and 12 is farthest from 6, in terms of absolute distances. The function will have the following prototype:

```
void dist_sort( int x[ ], int n, int z );
```
(15)

3 a) Define a C structure named customer to store the data of a customer in a bank. The data to be stored is *Account number* (integer), *Name* (character string having at most 50 characters), and *Balance* in account (integer). (3)

b) Assume data for all the 100 customers of the bank are stored in the array:

```
struct customer bank[100];
```

Write a function called `transaction` to perform a customer request for withdrawal or deposit to her account. Every such request is represented by the following three quantities: *Account number of the customer, request type* (0 for deposit and 1 for withdrawal) and *amount*. The `transaction` function returns 0 if the transaction fails and 1 otherwise. The transaction fails only when the account balance is less than the withdrawal amount requested. The array `bank` (defined above) is another input to the `transaction` function and is suitably updated after every request. In case of a failed transaction no change is made in the `bank` array. (7)

c) Write a C program that will read the first 10 characters from a file and store them in a new file in reverse order. Is it possible to write this program without using any character array for intermediate storage? Justify your answer. (6+4=10)

P.T.O

4.a) Express $f(n)$ as a function of n, where f is defined as follows. Show your calculations. (6)

```
unsigned int f ( unsigned int n )
{
  unsigned int s = 0, i, j;
  for (i=1; i<=n; ++i)
   for (j=i; j<=n; ++j)
     s += i + j;
   return s;
}
```

b) Provide a suitable **typedef** for representing a queue by using a linked list. Each node in the queue should contain one integer data field. Write C functions to implement the following operations on the queue:

init – which constructs and returns an empty queue,
empty – which returns a value indicating whether a given queue is empty or not,
enqueue – which adds a given integer to the rear of a queue,
dequeue – which deletes an element from the front of the queue. (2+2+2+4+4=14)

5. a) How can you use pointers to store a two-dimensional array with a user-specified number of rows and columns? Write a C code snippet to illustrate how memory can be allocated to such a dynamic two-dimensional array. (5)

b) Write a suitable typedef for the linked representation of a binary tree. Now write two recursive functions for the inorder and postorder traversal of the tree. The functions should have the following prototypes:

```
void in_trav (node *tree)         /* for inorder traversal*/
void post_trav (node *tree)       /* for postorder traversal*/
```

where *tree is a pointer to the root node of the binary tree. (2+4+4=10)

c) How many non-leaf nodes are there in a completely binary tree of level d and why? (5)

6. a) Write a simple recursive function to return the n-th Fibonacci number. Trace the recursion by showing how the stack memory changes with each function call for $n = 5$. (5+7=12)

b) Briefly describe the steps of a simple insertion sort algorithm. Comment on its best and worst case time complexities. (8)

INDIAN STATISTICAL INSTITUTE

Mid-Semester of Second Semester Examination : 2015–16

Course : Bachelor of Statistics (Hons.)

Subject : Numerical Analysis : BStat-I

Date : 22 February 2016 | Maximum Marks : 60 | Duration : 3 Hours

Attempt any two from the first three problems. In addition, attempt the bonus problem.

Problem 1 [25]

Given a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$, the QR decomposition of the matrix satisfies $\mathbf{A} = \mathbf{QR}$, where $\mathbf{Q} \in \mathbb{R}^{m\times n}$ is an orthogonal matrix, and $\mathbf{R} \in \mathbb{R}^{n\times n}$ is an upper-triangular matrix.

A. Describe an algorithm that orthogonalizes $\mathbf{A}$ through successive right multiplication by upper triangular matrices, such that $\mathbf{AR}_1\mathbf{R}_2\cdots\mathbf{R}_n = \mathbf{Q}$, and $\mathbf{R} = (\mathbf{R}_1\mathbf{R}_2\cdots\mathbf{R}_n)^{-1}$. [10]

B. Describe an algorithm that triangularizes $\mathbf{A}$ through successive left multiplication by orthogonal matrices, such that $\mathbf{Q}_n\cdots\mathbf{Q}_2\mathbf{Q}_1\mathbf{A} = \mathbf{R}$, and $\mathbf{Q} = (\mathbf{Q}_n\cdots\mathbf{Q}_2\mathbf{Q}_1)^{-1}$. [10]

C. Which of the above two algorithms will you prefer for the QR decomposition of an arbitrary $m \times n$ real matrix $\mathbf{A}$? Justify you answer. [5]

Problem 2 [25]

A. Prove that all eigenvalues of a real symmetric matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ are real. [4]

B. Suppose that $\mathbf{A} \in \mathbb{R}^{n\times n}$ is a real symmetric matrix, having n distinct eigenvalues, which satisfy the ordering $|\lambda_1| > |\lambda_2| > \cdots |\lambda_n| > 0$. Describe an *iterative* algorithm to compute the largest (in absolute value) eigenvalue λ_1 of $\mathbf{A}$. [7]

C. Propose a similar iterative method, by introducing a modification in the above algorithm, to compute the smallest (in absolute value) eigenvalue λ_n of $\mathbf{A}$. [7]

D. Given a real number σ, propose a similar iterative method, with a modification in the above algorithm, to compute the eigenvalue λ_k of $\mathbf{A}$ that is closest (in absolute value) to σ. [7]

Note: In this problem, do not use an algorithm that tries to find *all* eigenvalues of $\mathbf{A}$ simultaneously.

Problem 3

[25]

A. Suppose that the *full* Singular Value Decomposition of $\mathbf{A} \in \mathbb{R}^{m \times n}$ results in:

$$\mathbf{A} = \begin{bmatrix} | & & | & & | \\ \mathbf{u}_1 & \cdots & \mathbf{u}_r & \cdots & \mathbf{u}_m \\ | & & | & & | \end{bmatrix} \left[\begin{array}{ccc|c} \sigma_1 & & & \\ & \ddots & & 0 \\ & & \sigma_r & \\ \hline & 0 & & 0 \end{array}\right] \begin{bmatrix} | & & | & & | \\ \mathbf{v}_1 & \cdots & \mathbf{v}_r & \cdots & \mathbf{v}_n \\ | & & | & & | \end{bmatrix}^T$$

What are the dimensions of each matrix on the right-hand-side of this decomposition? Discuss the connection of these matrices with the RowSpace, ColSpace and NullSpace of $\mathbf{A}$. How can you determine the Rank of $\mathbf{A}$ given this SVD representation? [3 + 6 + 1]

B. As per the SVD of $\mathbf{A}$, determine the dimension and rank of each of the matrices $\mathbf{Z}_i = \sigma_i \mathbf{u}_i \mathbf{v}_i^T$, where $1 \leq i \leq r$. Is it possible to reconstruct $\mathbf{A}$ given the matrices $\mathbf{Z}_i$ for $1 \leq i \leq k$, where k is strictly less than r? If so, provide such a construction. If not, provide an *approximate* reconstruction $\mathbf{A_k}$ of $\mathbf{A}$ using the available matrices $\mathbf{Z}_i$ for $1 \leq i \leq k$. Provide an expression for $||\mathbf{A} - \mathbf{A_k}||_2 = \max_{||\mathbf{x}||_2=1} ||(\mathbf{A} - \mathbf{A_k})\mathbf{x}||_2$ in terms of the singular values σ_i. [2 + 3 + 5]

C. The *pseudo-inverse* of a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ is generally expressed as $\mathbf{A}^+ = (\mathbf{A}^\mathbf{T}\mathbf{A})^{-1}\mathbf{A}^T$. Provide a computationally simple construction of $\mathbf{A}^+$ from the singular value decomposition of $\mathbf{A}$. [5]

Bonus Problem

[10]

Represent a book in the form of an $m \times n$ real matrix $\mathbf{B}$, where m is the total number of sentences in the book and n is the total number of distinct words in the book, such that the entry $\mathbf{B}[i, j]$ in this matrix represents the frequency of occurrence of the j-th word W_j in the i-th sentence S_i.

Importance of the words and sentences are denoted by *scores*. Score u_i of sentence S_i is proportional to the weighted sum of scores of the words in it. Score v_j of word W_j is proportional to the weighted sum of scores of the sentences it is contained in. This may be expressed as follows.

$$u_i \propto \sum_{j=1}^{n} \mathbf{B}[i, j] \cdot v_j \quad \text{for } i = 1, 2, \ldots, m \qquad\qquad v_j \propto \sum_{i=1}^{m} \mathbf{B}[i, j] \cdot u_i \quad \text{for } j = 1, 2, \ldots, n$$

Devise an efficient strategy to identify 10 *keywords* (i.e., the most important words) from the book

Good luck! ☺

Indian Statistical Institute

B. Stat. (Hons) First Year: Mid Semester Examination: 2015-16

Vectors and Matrices II

Date: 23.02.2016 **Maximum Mark: 35** **Time: 3 Hours**

1. Prove or disprove: If U is an invariant subspace of V under every operator defined on a finite dimensional vector space V, then U must be either $\{0\}$ or U=V. [5]

2. T is an operator on F^n, given by
$T(x_1, x_2, \ldots x_n) = (x_1 + x_2 + \ldots + x_n, x_1 + x_2 + \ldots + x_n, \ldots, x_1 + x_2 + \ldots + x_n)$. Find all eigenvalues and eigenvectors of T. [5]

3. Let $u, v \in V$, a finite dimensional inner product space. Show that $<u,v>= 0$ if and only if $\|u\| \le \|u + \alpha v\| \ \forall \alpha \in F$. [5]

4. Let $V = R^3, u = (2,1,3), W = \{(x, y, z) : x + 3y - 2z = 0\}$. Find the element of W such that it has the minimum distance from u, justifying that it has the minimum distance. Find the distance. [9]

5. Write down the matrix $\begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}$ as a product of a unitary matrix and a positive semi definite matrix. Can you get two sets of such matrices? Explain. [9]

6. Find the Jordan canonical form and a basis of the following matrix, explaining all the steps clearly. [9]

$$\begin{pmatrix} -1 & 3 & 0 \\ 0 & 2 & 0 \\ 2 & 1 & -1 \end{pmatrix}$$

INDIAN STATISTICAL INSTITUTE

Mid-Semestral Examination: 2015-16 (Second Semester)
Bachelor of Statistics (B. Stat.) I Year
Probability Theory II

Teacher: Parthanil Roy

Date: 24/02/2016 | Maximum Marks: 40 | Duration: 2:30 - 5:00 pm

Note:

- Please write your roll number on top of your answer paper.
- Show all your works and write explanations when needed. If you are using a result stated and/or proved in class, please quote it correctly.
- You are NOT allowed to use class notes, books, homework solutions, list of theorems, formulas etc.

1. Suppose f is a nonnegative real valued integrable function defined on $\mathbb{R}$ and X is a random variable. State whether the following statement is true or false: *f is a probability density function of X if and only if for all $a \in \mathbb{R}$,*

$$P(X \geq a) = \int_a^\infty f(x)dx.$$

If it is true, give a detailed proof. On the other hand, if it is false, produce a counter-example. [12]

2. Suppose $X_1 \sim \text{Gamma}(\alpha_1, 1)$, $X_2 \sim \text{Gamma}(\alpha_2, 1)$ and X_1, X_2 are independent. Define $Y_1 = X_1 + X_2$ and $Y_2 = \frac{X_1}{X_1+X_2}$.

   (a) Find the joint probability density function of Y_1 and Y_2. [12]

   (b) Find the marginal probability density functions of Y_1 and Y_2, and identify their distributions (i.e., the name of the distribution and the parameter value(s)) if possible. [2 + 2]

   (c) Are Y_1 and Y_2 independent? Justify your answer. [2]

   (d) Suppose now that Z_1 and Z_2 are iid random variables following exponential distribution with parameter $\lambda = 2$. Find the probability density function of $V := \frac{4Z_1+3Z_2}{Z_1+Z_2}$. Can you identify the distribution of V? [9 + 1]

Indian Statistical Institute
Statistical Methods II
B-I, Midsem

Date: Feb 25, 2016 Duration: 2hrs.

Attempt all questions. The maximum you can score is 20. Justify all your steps. This is a closed book, closed notes examination. You may use your own calculator. No need to perform more than three steps of any iterative method.

If copying is detected in the solution for any problem, all the students involved in the copying will get 0 for that problem. Also an additional penalty of 5 will be subtracted from the overall aggregate of each of these students.

1. You have a fair coin that you can toss any number of times independently. Can you simulate an event having probability p, where p is any given rational number in $[0, 1]$? Justify your answer. [5]

2. We have data X_1, X_2, X_3 IID from $Unif(0, \theta)$, where $\theta \in \{1, 2\}$ is unknown. Consider testing $H_0 : \theta = 1$ against $H_1 : \theta = 2$ using the following test: Reject H_0 if and only if $\max\{X_1, X_2, X_3\} > 1.5$. Find P(type I error), P(type II error) and the power of this test. Is this test unbiased? [1+1+1+2]

3. Fingerprints are of two patterns: "Whorl" and "No whorl". We want to investigate if there is any association between the fingerprint patterns for a mother and her child. Data have been collected from n randomly selected (mother, child) pairs to produce the following table:

Mother \ Child	Whorl	No whorl	Total
Whorl	n_{11}	n_{12}	n_{1*}
No whorl	n_{21}	n_{22}	n_{2*}
Total	n_{*1}	n_{*2}	n

Remember that here n is fixed, everything else being random. Perform a χ^2 goodness-of-fit to test if the following model fits the data set:

> God has two coins M and C. He (or she?) tosses the M coin in IID manner for each mother to decide if she should be given

a whorl fingerprint. Then God does a similar process *independently* for the children using the C coin. Nothing is claimed about P(head)'s of the two coins.

Clearly mention the statistic as well as the degrees of freedom. You must *derive* these starting from the χ^2 goodness-of-fit test procedure done in class. You may use the MLE for Binomial probability without proof. [10]

4. Airlines used to sell more tickets than they had seats hoping for some last minute cancellations. If there are 100 seats and the airline has sold 120 tickets, then find the probability that at least one passenger will be unable to find a seat. Based on past data you may assume that the chance that a passenger will cancel his/her booking is 5%. **You must state all your assumptions clearly.** [5]

Indian Statistical Institute, Kolkata
Midterm examination: Second semester 2015 -'16
B. Stat I year

Subject	:	**Analysis II**
Time	:	3 hours
Maximum score (after scaling)	:	40

Attempt all the problems. Justify every step in order to get full credit of your answers. All arguments should be clearly mentioned on the answerscript. Points will be deducted for missing or incomplete arguments.

(1) Let $f : [a, b] \to \mathbb{R}$ be a bounded function.
 (i) If $f^2 \in \mathcal{R}[a, b]$, then does it follow $f \in \mathcal{R}[a, b]$?
 (ii) If $f^3 \in \mathcal{R}[a, b]$, then does it follow $f \in \mathcal{R}[a, b]$?
If your answer to any of the above questions is yes, prove it; otherwise, give a counter-example with proper justification.

[7 marks]

(2) Consider the function $f : [0, \pi] \to \mathbb{R}$ defined by $f(x) = \begin{cases} \sin x, & \text{if } x \in \mathbb{Q} \cap [0, \frac{\pi}{2}], \\ 1 - \sin x, & \text{otherwise.} \end{cases}$

Evaluate $\overline{\int_0^{\pi}} f(x)\mathrm{d}x$ and $\underline{\int_0^{\pi}} f(x)\mathrm{d}x$ and prove your answers.

[8 marks]

(3) Let f and g be two real-valued continuous functions defined on $[a, b]$ such that $\int_a^b f(x)\mathrm{d}x = \int_a^b g(x)\mathrm{d}x$. Show that there exists $c \in [a, b]$ such that $f(c) = g(c)$.

[6 marks]

(4) Let $f_0 : [0, a] \to \mathbb{R}$ be a continuous function. Then, prove that the sequence $\{f_n\}_{n\in\mathbb{N}}$ of functions defined iteratively by

$$f_{n+1}(x) = \int_0^a f_n(x)\mathrm{d}x \text{ for all } n \in \mathbb{N}$$

converges uniformly to 0.

[7 marks]

(5) Let $c \in (a, b)$ and f be a bounded real-valued function defined on $[a, c) \cup (c, b]$. If the improper integral of f over $[a, b]$ exists, then prove that $\tilde{f} \in \mathcal{R}[a, b]$ for any extension $\tilde{f} : [a, b] \to \mathbb{R}$ of f.

[8 marks]

(6) Let K be a compact subset of $\mathbb{R}$ and $f : K \to K$ be a function satisfying the conditi
$|f(x) - f(y)| = |x - y|$ for all $x, y \in K$. Show that f is surjective. (*Hint: Pick a element in the codomain and consider the sequence obtained by applying f multi times on that element.*)

[9 mark

INDIAN STATISTICAL INSTITUTE

Semestral Examination: 2015-16 (Second Semester)
Bachelor of Statistics (B. Stat.) I Year
Probability Theory II

Teacher: Parthanil Roy

Date: 29/04/2016 | **Total Marks: 60** | **Duration: 2:30 - 6:00 pm**

Note:

- **Please write your roll number on top of your answer paper.**
- **Show all your works and write explanations when needed. If you are using a result stated and/or proved in class, please quote it correctly.**
- **You are NOT allowed to use class notes, books, homework solutions, list of theorems, formulas etc.**

1. Suppose $X_1, X_2, \ldots, X_n$ is a random sample from exponential distribution with parameter 1. Find the joint probability density functions of the following random vectors:
   (a) $\left(\sum_{k=1}^n X_k/k,\ \sum_{k=2}^n X_k/k,\ \sum_{k=3}^n X_k/k,\ \ldots,\ X_n/n\right)$
   (b) $\left(X_{(1)},\ X_{(2)} - X_{(1)},\ X_{(3)} - X_{(2)},\ \ldots,\ X_{(n)} - X_{(n-1)}\right)$

   Fix $\lambda > 0$ and $i \in \{1, 2, \ldots, n\}$. Compute $\mathrm{E}\left(e^{-\lambda \sum_{k=i}^n X_{(k)}}\right)$. [8 + 8 + 6 = 22]

2. Suppose $(X_1, X_2, X_3, X_4) \sim \text{Dirichlet}(a_1, a_2, a_3, a_4; a_5)$. Fix $u, v > 0$ such that $u + v < 1$. Find the conditional probability density function of (X_1, X_2) given $(X_3, X_4) = (u, v)$. [8]

3. Suppose $\mathbf{X} \sim N_k(\mathbf{0}, I_k)$ and $\mathbf{Y} \sim N_k(\mathbf{a}, \Sigma)$, where Σ is a positive-definite matrix. Let B be a full row rank matrix of order $m \times k$.
   (a) Find the joint probability density function of $B\mathbf{X}$. [10]
   (b) Find the joint probability density function of $B\mathbf{Y}$. [6]
   (c) Suppose $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_k)^T$. Fix $1 \le i_1 < i_2 < \cdots < i_p \le k$. Find the joint distribution of $Y_{i_1}, Y_{i_2}, \ldots, Y_{i_p}$. [4]

4. Suppose X and Y are two independent and identically distributed random variables each having cumulative distribution function F. State whether the following statement is true or false: *If F is a continuous function, then $P(X = Y) = 0$.* If it is true, give a detailed proof. On the other hand, if it is false, produce a counter-example. [10]

Wish you all the best

Indian Statistical Institute

B. Stat. (Hons) First Year: Semester Examination: 2015-16

Vectors and Matrices II

Date: 02.05.2016 **Maximum Mark: 50** **Time: 3 Hours**

1. Consider the following 3×3 real matrix

$$A = \begin{pmatrix} 3 & 4 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix}$$

Without actual computation, find the eigenvalues of $(A+I)^{-1}$. Determine all real numbers α for which the limit $\lim_{n\to\infty} \alpha^n A^n$ is a nonzero matrix. State clearly the results you use. [2+5]

2. The minimal polynomial and the characteristic polynomial of a matrix A are given by $p(t) = (t-2)(t-3)$ and $f(t) = -(t-2)^2(t-3)$ respectively. Find the Jordan canonical form of A. Justify all your steps. [6]

3. Find the rational canonical form of the real matrix

$$A = \begin{pmatrix} 3 & 1 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 3 \end{pmatrix}$$

Justify your steps. [7]

4. Prove that if T is an operator defined on a finite dimensional inner product space V, then

$Trace(T^*T) = \|Tv_1\|^2 + \ldots + \|Tv_n\|^2$, where $v_1, \ldots, v_n$ is any orthonormal basis of V.

Use the above result to show that if $\|T^*v\| \le \|Tv\|$ for every $v \in V$, then T is normal. [5+5]

5. Without actual computation, determine the number of solutions of the following matrix equation.

$$\begin{pmatrix} 2 & 5 \\ 0 & 3 \end{pmatrix}\begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix} + \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix}\begin{pmatrix} -3 & 4 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$$

State the results used and justify. [5]

6. Consider the matrix

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

Prove or disprove that A can be uniquely decomposed as $A = LL'$, where L is a lower triangular matrix with positive entries in the diagonal. [7]

P.T.O

7. Consider the matrix $A = \begin{pmatrix} 3 & 1 & 4 \\ 1 & 4 & 3 \\ 4 & 3 & 1 \end{pmatrix}$. Without computing

(i) Find the largest eigenvalue. What is the multiplicity? What is known about its eigenvector?

(ii) What can you say about the other eigenvalues?

(iii) What can you say about $\|A\|$ where, $\|.\|$ is any matrix norm?

In every answer, state clearly the result you use.

[1+1+1+1+1]

8. Let $A = \begin{pmatrix} \alpha & 1 & 0 & 0 & 0 \\ 1 & \alpha & 1 & 0 & 0 \\ 0 & 1 & \alpha & 1 & 0 \\ 0 & 0 & 1 & \alpha & 1 \\ 0 & 0 & 0 & 1 & \alpha \end{pmatrix}$, where α is real.

Show that $\alpha - 1 \geq \lambda_{min}$ and $\alpha + 1 \leq \lambda_{max}$, where λ_{min} and λ_{max} are the smallest and the largest eigenvalue respectively. [6]

9. Let $W = \{(x,y,z) \in R^3 : 2x^2 + 6xy + 5y^2 - 2yz + 2z^2 + 16 = 0\}$.

Find an orthonormal basis for R^3, so that the defining equation of W takes a simpler form with respect to this basis. Use this new form to describe W geometrically. [7]

INDIAN STATISTICAL INSTITUTE

Second Semester Examination : 2015–16

Course : Bachelor of Statistics (Hons.)

Subject : Numerical Analysis : BStat-I

Date : [illegible] May 2016 | Maximum Marks : 100 | Duration : 3 Hours

[illegible] four (4) from the first eight (8) problems. In addition, attempt the bonus problem.

Problem 1 [25]

A. [illegible] the *fixed-point* and *floating-point* representations of numbers in a computer system. [illegible] the standard IEEE format for storing a float, as in C? What is the largest and the [illegible] positive number that can be represented using the floating-point format $1.\square\square\square \times 2^E$, [illegible] ≤ 2, and the empty boxes contain binary digits 0 or 1? [15]

B. [illegible] that you want to compute the matrix-vector product $\mathbf{Av}$. If the matrix is $\mathbf{A}$ fixed and [illegible] the vector $\mathbf{v}$ is an input prone to errors, describe the *relative forward error* and [illegible] *ward error* of this numerical computation. Hence compute the *relative condition* [illegible] matrix-vector product, and provide a bound for the same. [10]

Problem 2 [25]

[illegible] $\mathbf{A} \in \mathbb{R}^{n\times n}$, the *simple LU decomposition* of $\mathbf{A}$ may be defined as $\mathbf{A} = \mathbf{LU}$, where [illegible] triangular matrix, and $\mathbf{R} \in \mathbb{R}^{n\times n}$ is an upper-triangular matrix.

[illegible] algorithm that may decompose $\mathbf{A}$ into $\mathbf{L}$ and $\mathbf{U}$ through successive left multiplication [illegible] triangular matrices, such that $\mathbf{L}_n \cdots \mathbf{L}_2\mathbf{L}_1\mathbf{A} = \mathbf{U}$, and $\mathbf{L} = (\mathbf{L}_n \cdots \mathbf{L}_2\mathbf{L}_1)^{-1}$. [10]

[illegible] possible to obtain such a *simple LU decomposition* of $\mathbf{A} \in \mathbb{R}^{n\times n}$? If so, provide a [illegible] that your proposed algorithm (in Part A) achieves a *simple LU decomposition* [illegible] matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$. If not, state a case where the *simple LU decomposition* may [illegible] and modify the proposed algorithm (in Part A) to suit such a case? [15]

Problem 3

A. Given a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ and a vector $\mathbf{b} \in \mathbb{R}^m$, propose a (linear algebraic) numerical met to minimize $||\mathbf{Ax} - \mathbf{b}||_2$, where $\mathbf{x} \in \mathbb{R}^n$. Modify this method to minimize $||\mathbf{Ax} - \mathbf{b}||_2^2 + \lambda||$ where $\mathbf{x} \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$ is a given constant.

B. What is the output of the following algorithm, if it terminates? Justify your answer.

> **Input** : Matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ and threshold $\epsilon > 0$
> Choose $\mathbf{v}^{(0)} \in \mathbb{R}^n$ at random, with $||\mathbf{v}^{(0)}||_2 = 1$
> Compute $r^{(0)} = (\mathbf{v}^{(0)})^T \mathbf{A}\ \mathbf{v}^{(0)}$
> for $k = 1, 2, 3, \ldots$ do :
>   Solve $(\mathbf{A} - r^{(k-1)}\mathbf{I}_{n \times n})\ \mathbf{w} = \mathbf{v}^{(k-1)}$, for $\mathbf{w}$
>   Normalize $\mathbf{v}^{(k)} = \mathbf{w}/||\mathbf{w}||_2$
>   Compute $r^{(k)} = (\mathbf{v}^{(k)})^T \mathbf{A}\ \mathbf{v}^{(k)}$
>   **if** $||r^{(k)} - r^{(k-1)}||_2 < \epsilon$: **break**
>
> **return** $r^{(k)}$ and $\mathbf{v}^{(k)}$

Problem 4

Given n points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ in $\mathbb{R}^2$, there exists a univariate polynomial $p(x)$ degree at most $(n - 1)$, passing through all the points. That is, $y_i = p(x_i)$ for all $i = 1, 2, \ldots, n$

A. Suppose that $p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_{n-1}x^{n-1}$, expressed in the monomial basis. Describe algorithm to construct $p(x)$, that is, to construct the monomial basis coefficients $a_0, a_1, \ldots, a$ given the n points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$.

B. Suppose that $p(x) = b_0 + b_1(x - x_1) + b_2(x - x_1)(x - x_2) + \cdots + b_{n-1}(x - x_1)(x - x_2)\cdots(x-$ expressed in the Newton basis. Describe an algorithm to construct $p(x)$, that is, to const the coefficients $b_0, b_1, \ldots, b_{n-1}$, given the n points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$.

Problem 5

Given a function $f : \mathbb{R} \to \mathbb{R}$, one may find a root of the equation $f(x) = 0$ by converting the probl into a fixed-point problem $g(x) = x$, so that any fixed-point of $g(x)$ is a solution to $f(x) = 0$.

A. Describe an algorithm to solve the fixed-point problem $g(x) = x$, that is, an algorithm that fi out a fixed point of a given function $g(x)$. Discuss the convergence of this algorithm.

B. How does the Newton-Raphson method for finding a root of the equation $f(x) = 0$ choose g Describe the Newton-Raphson method, and discuss the convergence of this algorithm.

Problem 6

[25]

Given a function $f : \mathbb{R} \to \mathbb{R}$, one may want to estimate the values of the derivatives $f'(x)$ or $f''(x)$ at a specific point $x = x_0$ by evaluating the function $f(x)$ at the neighborhood of $x = x_0$.

A. From the definition of derivative, suggest an estimate for $f'(x_0)$ in terms of the functional values $f(x_0)$ and $f(x_0 + h)$, where h is adequately small. Comment on the accuracy of this estimate. Provide an *improved* estimate for $f'(x_0)$ when you are allowed to use three values of the function, $f(x_0)$, $f(x_0 + h)$ and $f(x_0 - h)$. Comment on the accuracy of this improved estimate. [15]

B. Extend the same logic (as in Part A) to suggest an estimate for $f''(x_0)$ in terms of the values of $f(x)$ at the neighborhood of $x = x_0$. How many functional values would you require, and what would be the accuracy of your estimate? [10]

Problem 7

[25]

Given a function $f : \mathbb{R} \to \mathbb{R}$, one may find the value of $\int_a^b f(x)dx$ by sub-dividing the interval $[a, b]$ into adequately small sub-intervals, and applying a quadrature rule on each small interval.

A. Suppose that $h = (b - a)/k$, and the small subintervals are $[x_i, x_{i+1}]$, where $x_i = a + (i - 1) \times h$ for $i = 1, 2, \ldots, k$. Describe mid-point rule and trapezoidal rule for computing $\int_a^b f(x)dx$. [10]

B. Suppose that $h = (b - a)/k$, and the small subintervals are $[x_i, x_{i+1}]$, where $x_i = a + (i - 1) \times h$ for $i = 1, 2, \ldots, k$. Describe Simpson's rule for computing $\int_a^b f(x)dx$. Compare the accuracy of Simpson's rule with that of Mid-Point rule and Trapezoidal rule (as in Part A). [15]

[25]

Problem 8

Suppose that we want to solve the first-order ordinary differential equation $y'(t) = F(t, y)$, subject to the initial value $y(0) = 1$, where the relation between t, y, y' is given by $F(t, y) = t + 2t \cdot y(t)$. To solve this problem, one may consider an equally-spaced time sequence $t_0 = 0, t_1, t_2, \ldots, t_k, t_{k+1}, \ldots,$ with $t_{k+1} = t_k + h$ for all $k \geq 0$, and provide a recurrence relation to compute $y(t_{k+1})$ given $y(t_k)$.

A. Describe the *forward* and *backward* Euler methods to compute $y(t_{k+1})$ given $y(t_k)$ in case of the ODE $\{y'(t) = t + 2t \cdot y(t);\ y(0) = 1\}$. Which method is more efficient, and why? [10]

B. Describe Runge-Kutta method for solving the ODE $\{y'(t) = t + 2t \cdot y(t);\ y(0) = 1\}$. Compare the accuracy of Runge-Kutta method with that of the Euler methods for solving ODEs. [15]

—— *End of the segment for regular problems. Please turn over for the Bonus problem.* ——

Bonus Problem

Suppose that G is an undirected graph consisting of 6 vertices $\{A, B, C, D, E, F\}$, and an unkn number of edges. You are provided with neither the adjacency matrix of the graph, nor with list of edges in it. However, you know that the following algorithm was executed step-by-step the graph G, and the corresponding output is as follows.

Algorithm

1. Construct the 6×6 adjacency matrix $\mathbf{A}$ of the graph G.
2. Construct a 6×6 matrix $\mathbf{D}$ that has the degrees of the six vertices in G as its diagonal elements. All other elements of $\mathbf{D}$ are zero.
3. Construct a 6×6 matrix $\mathbf{L}$ as follows: $\mathbf{L} = \mathbf{D} - \mathbf{A}$.
4. Compute the eigenvalues and eigenvectors of the matrix $\mathbf{L}$.

Output

```
Eigenvalues of L

4.561553e+00  3.000000e+00  3.000000e+00  3.000000e+00  4.384472e-01  5.313210e-16

Eigenvectors of L (in columns)
```

[v1]	[v2]	[v3]	[v4]	[v5]	[v6]
0.6571923	0.066054535	0.000000e+00	0.5735592	-0.2609565	0.4082483
0.1845241	-0.734936838	2.807975e-02	-0.2059434	0.4647051	0.4082483
0.1845241	0.668882304	-2.807975e-02	-0.3676157	0.4647051	0.4082483
-0.1845241	-0.060922638	-7.065490e-01	-0.2835670	-0.4647051	0.4082483
-0.6571923	0.066054535	7.979728e-17	0.5735592	0.2609565	0.4082483
-0.1845241	-0.005131897	7.065490e-01	-0.2899922	-0.4647051	0.4082483

Interpret the output eigenvalues and eigenvectors of $\mathbf{L}$ to take an educated guess about the struct of the graph G, and draw a regular vertex-edge layout of the graph to depict your guess.

Good luc

Indian Statistical Institute
Statistical Methods II
B-I, Second Semestral Examination

Date: May 06, 2016 Duration: 3 hrs.

This paper carries 55 marks. Attempt all questions. The maximum you can score is 50. Justify all your steps. This is an open book, open notes examination. You may use your own calculator. No need to perform more than three steps of any iterative method.

If copying is detected in the solution for any problem, all the students involved in the copying will get 0 for that problem. Also an additional penalty of 10 will be subtracted from the overall aggregate of each of these students.

1. Let X_1, X_2, X_3 be IID $N(\mu, \sigma^2)$. We want to estimate $\mu \in \mathbb{R}$ and $\sigma^2 > 0$. However we cannot observe X_1, X_2 directly. Instead, we can observe $U = X_1 + X_2$ and $V = X_2 + X_3$. Based on IID realisations $(u_1, v_1), ..., (u_n, v_n)$ of (U, V) obtain MLE of (μ, σ^2) using EM algorithm. Clearly state the incomplete and complete data you are using. [15]

2. State true or false. Justify your answers with proofs or counterexamples as appropriate.

   (a) Given any number $r \in [-1, 1]$ it is possible to have a bivariate data set $(x_1, y_1), ..., (x_{10}, y_{10})$ such that the sample correlation is equal to r. [5]

   (b) For a multivariate data set consisting of n cases of p variables $x_1, ..., x_p$, let the sample covariance matrix be S. Also let $\vec{x}_1, ..., \vec{x}_p$ be the columns of the data set. Then S is nonsingular if and only if the matrix $X = [\ \vec{1}\ \vec{x}_1 \cdots \vec{x}_p\ ]$ is full column rank. [5]

3. Let $x_1, ..., x_n$ be fixed numbers with distinct absolute values ($n \geq 4$). Consider the model

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i,$$

   where ϵ_i's are IID $N(0, \sigma^2)$ for some unknown $\sigma^2 > 0$.

   (a) Write the model in the form

$$\vec{Y} = X\vec{\beta} + \vec{\epsilon}.$$

   (b) Show that the MLE of σ^2 is

$$\frac{1}{n}\sum_1^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i - \hat{\beta}_2 x_i^2)^2,$$

   where $\hat{\beta}_i$'s are the least square estimators of β_i's.

(c) If the QR deocomposition of X is available, suggest how may compute the above MLE using this without explicitly finding $\hat{\beta}_i$'s.

(d) Find the bias of $\hat{\sigma}^2$ for estimating σ^2. [2+5+5+5]

4. The sample covariance matrix for X_1, X_2, X_3, X_4 based on 100 cases is

$$S = \begin{bmatrix} 54 & 22 & 11 & 20 \\ 22 & 108 & 108 & 2 \\ 11 & 108 & 174 & -45 \\ 20 & 2 & -45 & 78 \end{bmatrix}$$

What is the partial regression coefficient $b_{23\bullet 14}$?

[3]

5. Write an R code to simulate from the uniform distribution over the *surface* of the cube $[-1,1]^3$. Also explain your algorithm clearly (either as comments in the R code, or as a separate description). [10]

Indian Statistical Institute, Kolkata
Back paper examination: Second semester 2015 -'16
B. Stat I year

Subject	:	**Analysis II**
Time	:	3 hours
Maximum score	:	100
Date	:	28-07-2016

Attempt all the problems. Justify every step in order to get full credit of your answers. All arguments should be clearly mentioned on the answerscript. Points will be deducted for missing or incomplete arguments.

(1) Find a sequence of **continuous functions** on a **compact interval** converging to a **continuous function pointwise** but **not uniformly**. Stating a correct example without justification will fetch partial credit. So, justify every statement.

[16 marks]

(2) Let f be a Riemann integrable function over the interval $[a, b]$ not necessarily continuous. Prove that there exists a sequence $\{g_n\}_{n\in\mathbb{N}}$ of continuous functions on $[a, b]$ such that $\lim_{n\to\infty} \int_a^b |g_n(x) - f(x)|\, \mathrm{d}x = 0$.

[20 marks]

(3) Let $f : [a, b] \to \mathbb{R}$ be differentiable and f' be continuous. Show that there exists a sequence $\{f_n\}_{n\in\mathbb{N}}$ of twice differentiable functions such that it converges uniformly to f as well as its derivative converges uniformly to f'.

[16 marks]

(4) What can you say about $\displaystyle\int_0^\infty \frac{1}{1 + x^2 - e^{-x}} \mathrm{d}x$ and its existence? Justify.

[16 marks]

(5) Find the most general form of solutions to the differential equation

$$y'' - 7y' + 6y = \sin x.$$

Justify every step of your answer.

[16 marks]

(6) **Using the method of reduction of order**, find the most general form of solutions to the differential equation

$$y'' - 4xy' + (4x^2 - 2)y = 0.$$

Justify every step of your answer.

Hint: Check that $\varphi(x) = e^{x^2}$ is a solution.

[16 marks]

Indian Statistical Institute
Statistical Methods II
B-I, Second Backpaper Examination

Date: 29-07-16 Duration: 3 hrs.

This paper carries 100 marks. Attempt all questions. The maximum you can score is 45. Justify all your steps. No need to perform more than three steps of any iterative method.

1. Let X_1, X_2, X_3 be IID $N(\mu, \sigma^2)$. We want to estimate $\mu \in \mathbb{R}$ and $\sigma^2 > 0$. However, we cannot observe X_1, X_2 directly. Instead, we can observe $U = X_1 - X_2$ and $V = X_2 + 2X_3$. Based on IID realisations $(u_1, v_1), \ldots, (u_n, v_n)$ of (U, V) obtain MLE of (μ, σ^2) using EM algorithm. Clearly state the incomplete and complete data you are using. [15]

2. State true or false. Justify your answers with proofs or counterexamples as appropriate.

   (a) Given any *rational* number $r \in [-1, 1]$, it is possible to have a bivariate data set $(x_1, y_1), \ldots, (x_{10}, y_{10})$ such that the Spearman rank correlation is equal to r. [5]

   (b) If X_1 and X_2 are two variables, then the multiple correlation coefficient $r_{1\bullet 2}$ is the same as the product-moment correlation coefficient r_{12}. [5]

   (c) For a multivariate data set consisting of n cases of p variables $x_1, \ldots, x_p$, let the sample covariance matrix be S. Also let $\vec{x}_1, \ldots, \vec{x}_p$ be the columns of the data set. Then S is nonsingular if and only if the matrix $X_{n\times p} = [\ \vec{x}_1 \ \cdots\ \vec{x}_p\ ]$ is full column rank. [5]

3. Let $x_1, \ldots, x_n$ be fixed numbers with distinct absolute values ($n \geq 4$). Consider the model

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i,$$

   where ϵ_i's are IID $N(0, \sigma^2)$ for some unknown $\sigma^2 > 0$.

   (a) Write the model in the form

$$\vec{Y} = X\vec{\beta} + \vec{\epsilon}.$$

   (b) If the QR deocomposition of X is available, suggest how you may compute $\hat{\beta}_i$'s.

   (c) Obtain (with proof) an unbiased estimator of σ^2 based on x_i's and Y_i's. [2+8+10]

4. The sample covariance matrix for X_1, X_2, X_3, X_4 based on 100 cases is

$$S = \begin{bmatrix} 174 & -24 & 115 & 119 \\ -24 & 225 & 4 & 38 \\ 115 & 4 & 185 & 117 \\ 119 & 38 & 117 & 106 \end{bmatrix}$$

What is the partial regression coefficient $b_{12\bullet 34}$? Also find the partial correlation coefficient $r_{14\bullet 2}$. [5]

5. Write an R code to simulate from the uniform distribution over the unit sphere

$$\{(x, y, z) \in \mathbb{R}^2 \ : \ x^2 + y^2 + z^2 = 1\}.$$

Explain your algorithm clearly (either as comments in the R code, or as a separate description). [10]

6. A die is rolled 20 times and the outcomes are recorded as a frequency distribution. We want to perform a χ^2 goodness of fit test to check if the die is fair. However, the sample size is inadequate for the χ^2 approximation to hold. Suggest how you may use simulation to perform the test here. Write down your procedure clearly. [10]

7. Compute Fisher information, $I_n(\lambda)$, for a random sample of size n from the $Poisson(\lambda)$ distribution. Check whether $I_n(\lambda)$ increases or decreases with n and λ. Interpret your observation intuitively. [5+2+3]

8. Consider the simple linear regression set up

$$Y_i = \alpha + \beta x_i + \epsilon_i, \quad (1 \leq i \leq n),$$

where x_i's are fixed given numbers not all same, and ϵ_i's are IID with mean 0 and unknown variance $\sigma^2 > 0$ Let $\hat{Y}(x)$ be the least square predicted value of Y for some given value x. Obtain variance of $\hat{Y}(x)$. Sketch its graph as a function x. Interpret it intuitively. [15]

INDIAN STATISTICAL INSTITUTE

Backpaper Examination: 2015-16 (Second Semester)
Bachelor of Statistics (B. Stat.) I Year
Probability Theory II

Teacher: Parthanil Roy

Date: 01-08-16 **Total Marks: 100** **Duration: 3 hours**

Note:

- **Please write your roll number on top of your answer paper.**
- **Show all your works and write explanations when needed. If you are using a result stated and/or proved in class, please quote it correctly.**
- **You are NOT allowed to use class notes, books, homework solutions, list of theorems, formulas etc.**

1. Suppose f is a nonnegative real valued integrable function defined on $\mathbb{R}$ and X is a random variable. State whether the following statement is true or false: *f is a probability density function of X if and only if for all $a \in \mathbb{R}$,*
$$P(X < a) = \int_{-\infty}^{a} f(x)dx.$$
If it is true, give a detailed proof. On the other hand, if it is false, produce a counter-example. [12]

2. Let $Y_1 < Y_2 < Y_3$ be an order statistic of size 3 from the distribution with probability density function $f(x) = 2x$ for $0 < x < 1$. For $i = 1, 2$, define $U_i = Y_i/Y_{i+1}$. Compute the joint probability density function of U_1 and U_2, and show that $U_1 + U_2$ is independent of Y_3. [9 + 3]

3. Let X, Y and Z be three independent standard uniform random variables and $S = X + Y + Z$. Find (a) the conditional density of S given $Y = y$ and $Z = z$ and (b) the conditional density of (Y, Z) given $S = s$. [6 + 6]

4. Suppose U and V are two independent random variables with $U, V \sim Unif(0,1)$. Compute the probability density function of $Z = U - V$. [12]

5. Let R be the range of a random sample of size n from uniform distribution on $(0,1)$. Find the probability density function of R. [12]

6. Let $Z_1, Z_2, \ldots Z_{20}$ be independent and identically distributed random variables with $Z_1 \sim N(0,1)$. Find the joint probability density function of $\sum_{i=1}^{20} Z_i^2$ and $\frac{\sum_{i=1}^{8} Z_i^2}{\sum_{i=1}^{20} Z_i^2}$. Are they independent? [10 + 2]

7. Suppose (U, V) is a bivariate normal random vector with $E(U) = E(V) = 0$, $Var(U) = Var(V) = 1$ and $Cov(U, V) = \rho$. Compute the correlation coefficient between U^2 and V^2, and the probability $P(U > 0, V > 0)$. [3 + 7]

8. A continuous random vector (X, Y) has a joint probability density function given by

$$f_{X,Y}(x,y) = \begin{cases} 2/3 & \text{if } x > 0, y > 0, x + y < 1, \\ 4/3 & \text{if } x < 1, y < 1, x + y > 1, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Compute the marginal probability density function of X. [5]

(b) For $0 < x < 1$, find the conditional probability density function of Y given $X = x$. [5]

9. Suppose $T \sim Exp(\lambda)$. For any real number x, denote by $[x]$ the largest integer less that or equal to x. For example, $[2.3] = 2$, $[5] = 5$ and so on.

(a) Compute the probability mass function of $S = [T]$. [6]

(b) Exactly which distribution does S follow? [2]

Wish you all the best

Indian Statistical Institute

Vectors and Matrices II

B. Stat First Year: Semester II Back paper Examination: 2015-16

Maximum Marks: 100 **Time: 3hrs** **Date: 04 - 08 - 2016**

1. Prove or disprove: If x, y, and z are vectors in an inner product space such that $< x,y >=< x,z >$, then $y = z$. [7]
2. Prove or disprove: Every nonzero finite dimensional inner product space has an orthonormal basis. [8]
3. Prove or disprove: The adjoint of a linear operator is not unique. [7]
4. Prove or disprove: Every normal operator is diagonalisable. [8]
5. Suppose $u, v \in V$, a n dimensional inner product space V over a field F. Prove that $< v,u >= 0$ if and only if $\|u\| \le \|u + \lambda v\| \quad \forall \lambda \in F$. [8]
6. Prove that if P is a linear operator on a n dimensional inner product space V over a field F such that $P^2 = P$ and $\|P(v)\| \le \|v\| \; \forall v \in V$, then P is an orthogonal projection. [10]
7. Define a positive semi definite operator on a n dimensional inner product space V over a field F. Show that it has a unique positive semi definite square root. [10]
8. Show that any orthogonal operator on R^2 is either a rotation or a reflection about a line through origin. [10]
9. Let T be a linear operator on a n dimensional vector space V such that V is a T cyclic subspace of itself. Then show that the characteristic polynomial $f(t)$ and the minimal polynomial $p(t)$ of T have the same degree, and $f(t) = (-1)^n p(t)$. [10]
10. State and prove Sylvester's law of inertia. [10]
11. Find a singular value decomposition of the following matrix. [12]

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \end{pmatrix}$$